

Л. М. ГИНДИЛИС, С. А. КАПЛАН
Н. С. КАРДАШЕВ, Б. Н. ПАНОВКИН
Б. В. СУХОТИН, Г. М. ХОВАНОВ

ВНЕЗЕМНЫЕ ЦИВИЛИЗАЦИИ

Проблемы
межзвездной связи

Под редакцией
проф. С. А. КАПЛАНА



ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1969

Введение. Экзосоциология — поиск сигналов внеземных цивилизаций (С. А. Каплан)	7
Теория развития цивилизаций (11). Проблема поиска сигналов внеземных цивилизаций (14). Дешифровочный аспект программы поиска сигналов внеземных цивилизаций (20).	
Литература	24
Глава I. Астрофизический аспект проблемы поиска сигналов внеземных цивилизаций (Н. С. Кардашев) . .	25
§ 1. Введение	25
§ 2. Основная дилемма	29
§ 3. Полнота и достоверность современных астрофизических данных	30
§ 4. Понятие цивилизации, главные закономерности и характер их развития	43
§ 5. Поиск проявлений деятельности сверхцивилизаций	52
Источники энергии (52). Твердое вещество (71).	
§ 6. Поиск передач информации	75
§ 7. О программе поиска сверхцивилизаций	97
Литература	100
Глава II. Влияние космической среды на распространение радиосигналов (Б. Н. Пановкин)	102
Литература	115
Глава III. Возможность радиосвязи с внеземными цивилизациями (Л. М. Гиндилис)	116
§ 1. Элементы общей теории связи	116
Структуры и основные характеристики системы связи (116). Количественное определение инфор-	

мации (119). Преобразование сообщения в сигнал. Виды модуляции (122). Физические характеристики сигналов (124). Зависимость между длительностью и шириной спектра импульсов. Количество импульсов, пропускаемых за единицу времени через канал с заданной полосой частот Δf (129). Передача непрерывных функций с помощью импульсных сигналов (132). Пропускная способность канала связи (139).

- § 2. Дальность и информативность межзвездной радиосвязи 145
 Оптимальный диапазон (145). Дальность связи (148). Дальность обнаружения (159). Дальность приема импульсных сигналов (165). Длительность передачи. Направленность и информативность (167).
- § 3. Позывные и критерии искусственности 172
- § 4. Методы обнаружения сигналов внеземных цивилизаций 182
 Мощность передатчика. Энергетический потенциал цивилизации (182). Радиосвязь между галактиками (199). Монохроматические сигналы. Поиск по частоте (200). Поиск сигналов по направлению (209). Широкополосные сигналы. Проведение обзоров неба (214).
 Литература 220

Глава IV. Методы дешифровки сообщений (Б. В. Сухотин) 222

- § 1. Введение 222
- § 2. Понятие сообщения, его понятности и разумности 225
 Определение понятия «сообщение» (225). Искусственность и естественность сообщений (227). Понятность сообщения, предсказуемость система, язык (229). Разумность сообщения (232).
- § 3. Традиционные методы военной и лингвистической дешифровки 234
 Военная дешифровка (234). Лингвистическая дешифровка (238).
- § 4. Порядок применения и структура дешифровочных алгоритмов 241
 Порядок применения алгоритмов. Уровни (241).
 Строение алгоритмов: множества возможностей,

оценочная функция, вычислительные процедуры. Типы алгоритмов (248).	
§ 5. Алгоритмы классификации (начало)	251
Признаки и классификации (251). Алгоритмы, опознающие гласные и согласные (255).	
§ 6. Алгоритмы склеивания (начало)	263
Алгоритм, выделяющий кодовые цепочки (263). Пример использования понятия разумности (276).	
§ 7. Алгоритмы дешифровки изображений	277
Язык изображений. Связность и детализирован- ность (277). Примеры оценочных функций. Неко- торые процедуры (284).	
§ 8. Алгоритмы, аналогичные алгоритмам, строящим двуязычные словари	294
Алгоритмы сравнения букв, использующие свой- ства близких окрестностей (294). Алгоритм, ис- пользующий далекие окрестности (308).	
§ 9. Алгоритмы классификации (окончание)	309
«Математически» правильный алгоритм выделе- ния гласных и согласных (309). Алгоритм, пере- водящий слоговую письменность в буквен- ную (313). Алгоритм, классифицирующий слова по смыслу (319).	
§ 10. Алгоритмы установления близости	324
Алгоритм, находящий граф синтаксической связи слов в предложении (324). Алгоритм, обнаружи- вающий типы синтаксической связи слов (330). Простейший алгоритм пословного машинного пе- ревода (333).	
§ 11. Алгоритмы склеивания (окончание)	336
Алгоритм, выделяющий морфемы (336). Алго- ритм, выделяющий буквы (343).	
§ 12. Заключение	351
Литература	352

Глава V. Темпы развития цивилизаций и их прогнози- рование (Г. М. Хованов) 353

§ 1. Значение проблематики темпов	353
§ 2. Аспекты развития цивилизаций	355
Язык и коммуникации (356). Демографические	

	характеристики цивилизации (359). Развитие индивидуальных способностей (362).	
§ 3.	Показатели технического прогресса	364
	О преимственности показателей (366). Математические функции, описывающие темпы роста (368).	
§ 4.	Темпы развития науки	370
§ 5.	Прогнозирование	376
	О способах классификации прогнозов (378). О точности прогноза (379). Прогноз темпов научно-технического прогресса (381). О прогнозе темпов развития земной цивилизации (384).	
	Литература	388
Глава VI. Некоторые общие вопросы проблемы внеземных цивилизаций (Б. Н. Пановкин)		
§ 1.	Введение	391
§ 2.	Методология «радиоастрономического» аспекта проблемы. «Энергетическая» гипотеза	393
§ 3.	Другая точка зрения. С. Лем и его «Сумма технологий»	407
§ 4.	Проблемы космических цивилизаций с точки зрения общей теории систем	417
	Литература	437

Экзосоциология — поиск сигналов внеземных цивилизаций

Поиск сигналов внеземных цивилизаций — одна из интереснейших проблем, поставленных современной наукой. Что представляют собой такие сигналы, как и где их искать и следует ли вообще этим заниматься? Подобные вопросы сначала поднимались в научно-фантастической литературе, а в последние несколько лет всерьез обсуждаются астрономами, физиками, биологами, лингвистами, философами на научных конференциях, в многочисленных книгах и статьях. Много вопросов, очень много различных гипотез и предположений — но еще очень мало научных исследований. Даже такой кардинальный для всей проблемы вопрос — что принесет человечеству установление контактов с внеземными цивилизациями — колоссальную пользу или огромный вред — и тот вызывает самые противоположные суждения. Достаточно хотя бы сравнить чрезвычайно радужные перспективы межзвездной связи, описанные в книге И. А. Ефремова «Туманность Андромеды», и весьма мрачную картину, нарисованную в книге Ф. Хойла и Ч. Эллота «Андромеда». Кстати, следует подчеркнуть, что обе эти книги написаны не просто писателями-фантастами, которые часто без достаточной ответственности относятся к существу проблемы, а серьезными учеными, вполне отдающими себе отчет в важности вопроса.

Выход человека в космическое пространство привел к возникновению и быстрому развитию ряда новых областей науки и техники. Одной из таких областей

является экзобиология — наука о зарождении и развитии жизни за пределами Земли. Весьма обширный круг вопросов, рассматриваемых экзобиологией, привлекает специалистов из самых различных областей науки и техники. Некоторые проблемы экзобиологии уже близки к разрешению, другие только поставлены.

Одной из основных проблем экзобиологии является чисто астрономическая задача: какова вероятность наличия у отдельно взятой звезды планеты, на которой имеются условия, допускающие возможность зарождения жизни. Иначе, необходимо найти вероятность существования планеты с массой, не слишком отличающейся от земной, обладающей соответствующими параметрами вращения, имеющей атмосферу и находящейся в «тепловом поясе жизни», т. е. на таком расстоянии от звезды, где не слишком жарко и не очень холодно. Несмотря на известную неопределенность даже в самой постановке задачи, на этот вопрос уже удалось получить более или менее однозначный ответ — вероятность существования такой планеты порядка нескольких процентов.

Однако наличие возможности для появления и развития жизни еще не означает, что она действительно на данной планете зарождается. К сожалению, именно в этом, уже чисто биологическом вопросе существует очень большая неопределенность. По-видимому, большая часть исследователей считает, что вероятность появления и развития жизни достаточно велика. Но есть и другая точка зрения, до сих пор полностью еще не опровергнутая, согласно которой вероятность зарождения жизни даже в самых подходящих условиях чрезвычайно мала. Трудность проблемы усугубляется еще и тем обстоятельством, что до сих пор не существует надежной теории происхождения жизни на самой Земле. Этот процесс не удается воспроизвести и в лабораторных условиях. Весьма важное значение для решения проблемы (наряду с лабораторными исследованиями) имело бы обнаружение хотя бы малейших следов жизни на планетах Венере или Марсе или доказательство многократности зарождения жизни на Земле. Поэтому вряд ли будет

преувеличением утверждать, что успехи экзобиологии в решении ее важнейшей проблемы — выяснении возможности зарождения жизни при определенных условиях — в значительной мере зависят от уровня космической техники. Быстрые успехи современной космонавтики вселяют надежду, что вопрос о вероятности зарождения жизни на планете с подходящими условиями будет решен уже в ближайшие годы.

Не менее сложен и также далек от разрешения вопрос о вероятности эволюции жизни от возникновения ее самых примитивных форм до появления разумных существ. И по этому вопросу имеются столь же противоречивые высказывания — от утверждений, что появление разумных существ есть однозначное следствие зарождения жизни, до не менее категорического суждения о том, что биологическая эволюция есть последовательность критических, неповторимых и непредсказуемых шагов и что любая случайность может оборвать процесс эволюции жизни в направлении усложнения организмов. Если принять последнюю точку зрения, то жизнь на Земле — единственная в пределах чуть ли не всей Метагалактики. Вероятно, это все же не так.

Автор этих строк, не будучи даже дилетантом в вопросах биологии, может ограничиться лишь выражением надежды, что и эта проблема экзобиологии будет решена не в столь отдаленном будущем и что вероятность эволюции от примитивной жизни до сознательных существ окажется не слишком малой.

И, наконец, интереснейшая проблема эволюции об-ществ разумных существ, живущих вне Земли, — вне-земных цивилизаций. Поскольку здесь идет речь имен-но о развитии общества и поскольку уже можно говорить о появлении, по существу, нового самостоятельного раздела науки, в котором должны изучаться эти пока еще не открытые общества, мы предлагаем по аналогии с экзобиологией использовать в качестве наименования этого раздела термин «экзосоциология». Предложенное И. С. Шкловским наименование «космософия» [1], т. е. «космическая мудрость», кажется

нам неудобным и не совсем соответствующим существу проблемы.

Ни одна наука, в том числе и экзосоциология, не может развиваться на основе чисто теоретических, «космософских» построений. Необходимы опыт, эксперимент, наблюдения. На данном этапе экзосоциология может использовать лишь опыт развития единственной известной нам цивилизации — земной. Но действительной, настоящей наблюдательной основой экзосоциологии должен стать анализ сигналов внеземных цивилизаций — если, конечно, их удастся обнаружить. Это основное положение и отражено в заголовке настоящего введения.

Предлагаемая книга посвящена экзосоциологии. Вопросы собственно экзобиологии, т. е. проблемы зарождения и развития жизни в космических условиях, здесь не затрагиваются вообще. По этим вопросам уже существует достаточно обширная литература (см., в частности, [1] и [2]).

У читателя, естественно, может возникнуть вопрос, нужна ли специальная книга по экзосоциологии в настоящее время, когда поиски сигналов внеземных цивилизаций, по существу, еще не ведутся и когда шансы на обнаружение этих сигналов не слишком велики? Нам кажется, нужна, и вот по каким причинам. Во-первых, систематические поиски сигналов внеземных цивилизаций рано или поздно будут организованы и необходимо заранее обсудить теоретические и практические возможности таких поисков. Во-вторых, исследование проблем экзосоциологии может дать в качестве «by product» немало интересных данных и для «земной» науки. Например, задача поиска источников космического радионизлучения, которые можно заподозрить в искусственном происхождении, вполне аналогична некоторым современным радиоастрономическим задачам, на первый взгляд не имеющим никакого отношения к экзосоциологии. Разработка проблемы дешифровки космических сообщений может дать много интересного и для исследования чисто лингвистических задач. Мы уже не говорим о проблеме прогнозирования развития цивилизаций. По-

этому, организовав поиски сигналов внеземных цивилизаций, исследуя различные проблемы экзосоциологии, мы не останемся в проигрыше даже в том случае, если эти сигналы и не будут обнаружены в ближайшем будущем. А выигрыш в случае удачи даже трудно себе представить.

Главы настоящей книги посвящены различным аспектам проблемы поиска сигналов внеземных цивилизаций. Чтобы облегчить чтение книги, мы постараемся описать общую схему этой проблемы и ее основные аспекты, как они представляются нам в настоящее время.

Теория развития цивилизаций. Наша земная цивилизация — единственный известный нам до настоящего времени образец общества разумных существ — существует, с точки зрения астрономических масштабов времени, весьма небольшой срок, всего несколько тысяч лет. Еще много меньше промежутков времени, в границах которого ее можно изучать. Между тем основной предмет экзосоциологии — изучение цивилизации на всем протяжении их эволюции, длительность которой, по крайней мере в принципе, может быть сравнима с астрономическими масштабами времени (миллионы и миллиарды лет). Во всяком случае сигналы будут приниматься только от цивилизаций, значительно превосходящих по уровню развития нашу земную цивилизацию.

Таким образом, экзосоциология должна найти подход к исследованию сверхцивилизаций, изучению эволюции обществ разумных существ за большие, астрономические промежутки времени.

Казалось бы, что решение этой задачи следует начинать с прогнозов развития нашей земной цивилизации. Однако при подобной попытке мы сразу же сталкиваемся с существенной трудностью. Любые прогнозы, как правило, основываются на экстраполяции закономерностей предыдущего развития. Но очевидно, такая экстраполяция допустима лишь на срок, в крайнем случае сравнимый, а обычно много меньший, чем тот промежуток времени, на основе изучения

закономерностей которого делается прогноз. Недаром прогнозирование будущего развития земной цивилизации (так называемая «футурология») чаще всего ограничивается 2000 годом (в редких случаях рискуют делать прогнозы до 2100 г.).

О несовершенстве метода экстраполяции свидетельствует хотя бы тот факт, что попытки его автоматического применения для прогноза дальнейшего развития земной цивилизации неизбежно приводят к так называемым «взрывам» — очень быстрым возрастаниям значения соответствующих показателей развития.

Общеизвестен «демографический взрыв»: вывод о том, что население земного шара около 2020—2030 гг. станет бесконечно большим. Другой пример: «энергетический взрыв». Подсчеты показывают, что примерно к 2100 г. на Земле будет производиться столько энергии, что температура планеты станет неограниченно повышаться. Наконец, совсем близок «информационный взрыв», при котором объем информации, накопленной наукой, сделается бесконечно большим уже около 1980 г.

Можно не сомневаться, что в действительности «взрывов» не будет, но каким именно образом будут пройдены «критические» моменты и как при этом будут меняться закономерности развития, пока не известно. Часто в качестве всеобщей панацеи предлагается расселение человечества в космосе. Вероятно, найдется и более простой выход из затруднений. Исследование смены ведущих показателей, характеризующих процесс развития, — одна из основных задач прогнозирования развития цивилизации (см. гл. V).

Таким образом, несмотря на чрезвычайный интерес, который представляют для нас прогнозы развития земной цивилизации, они пока мало что дают для экзосоциологии. Поэтому мы и не будем их здесь подробно обсуждать; пожалуй стоит только упомянуть о прогнозах известного футуролога А. Кларка [6] и еще более известной американской «Rend Corporation», согласно которым встреча с внеземными цивили-

зациями предсказывается лишь на вторую половину XXI в.

По-видимому, в настоящее время более целесообразно искать закономерности развития обществ разумных существ, цивилизаций, в каком-то более общем, абстрактном виде, опираясь на современные кибернетические понятия сложных систем. Стоит попытаться выработать общие определения понятия цивилизации и посмотреть, к каким тенденциям в развитии систем эти определения приведут. Так, в первой главе настоящей книги предлагается следующее определение цивилизации: «...высоко устойчивое состояние вещества, способное собирать, абстрактно анализировать и использовать информацию для получения максимума информации об окружающем и самом себе и для выработки сохраняющих реакций». В пятой главе отмечается еще одна общая закономерность: «простые системы используют эту информацию для сохраняющих реакций в данный момент, а более сложные системы обладают способностью и к прогнозирующим реакциям». Исходя из этих определений, можно ожидать, например, что цивилизации развиваются неограниченно и что они стремятся к установлению контактов между собой. Подробнее проблема кибернетического подхода к исследованию сверхцивилизаций обсуждается в гл. VI.

Разумеется, мы еще не только очень далеки от решения основной проблемы экзосоциологии — выяснения закономерностей развития цивилизаций — обществ разумных существ, но и, вероятно, еще толком не знаем, как вообще правильно поставить эту проблему. Однако авторы надеются, что соображения, приведенные в этой книге, особенно в гл. I, VI и отчасти V, будут бесполезны.

Остается заметить, что установление контактов с внеземными цивилизациями не только может привести к кардинальным изменениям наших представлений об обществе разумных существ, сколь бы «очевидными» ни казались нам эти представления до установления таких контактов, но, вероятно, изменит само дальнейшее развитие нашей собственной цивилизации.

Вступит в действие известный эффект «обратной связи», много раз обсуждавшийся, в частности, в связи с вопросом о полезности или вредности «межпланетных» контактов.

Проблема поиска сигналов внеземных цивилизаций. Несмотря на огромный объем информации, накопленный современной астрофизикой и радиоастрономией, такие сигналы до сих пор не обнаружены. Если принять во внимание, что большая часть великих открытий делается случайно тогда, когда их совсем не ожидают, то это обстоятельство не очень обнадеживает. Трудно сказать, в чем здесь дело. Может быть, в достаточно больших окрестностях нашей Солнечной системы вообще нет цивилизаций, способных посылать сигналы в космос? В то же время большинство исследователей считает, что сверхцивилизаций достаточно много. Однако какие-либо серьезные выводы можно будет делать только после того как будет проведена достаточно полная программа поиска сигналов других цивилизаций. Именно поэтому мы считаем, что такой поиск совершенно необходим.

Кстати, даже если внеземные цивилизации и не посылают специальных сигналов в космос, можно «подслушивать» их внутреннее радиовещание (телевидение). Уже на Земле имеется достаточно большая интенсивность искусственного радиоизлучения [1, 5], а у сверхцивилизаций она должна быть во много раз больше. Сочетание чувствительных приемников с интерферометрами большой базы (см. ниже), вероятно, сделает задачу «подслушивания» внеземных цивилизаций вполне выполнимой.

Как можно представить себе программу поиска сигналов внеземных цивилизаций? В настоящей книге этот вопрос подробно обсуждается в гл. I и III. По-видимому, в первую очередь следует осуществить радиообзор неба с целью выявления источников с минимальными угловыми размерами. В самом деле, антенны сверхцивилизаций, какую бы информацию они не передавали, должны быть очень малы по сравнению с астрономическими масштабами расстояний. Правда,

теоретически можно представить себе и передающее устройство планетарных размеров, но и эти масштабы крайне незначительны в сравнении с другими источниками космического радионизлучения. В последнее время достигнута разрешающая способность радионаблюдений порядка $0'',005$. Это было сделано с помощью радиоинтерферометра с независимой записью в каждом плече. Уже сейчас в принципе можно проводить наблюдения с базой порядка размеров Земли, а в будущем можно будет проводить наблюдения и с базой порядка 1 а. е. (разрешающая способность порядка $2 \cdot 10^{-9}$ угловой секунды!).

Существует еще целый ряд критериев, указывающих на возможное искусственное происхождение источников радионизлучения. Эти критерии подробно описаны и обсуждаются в гл. I и III, а более общий аспект определения искусственности сигналов и в гл. VI. Правильная переменность сигналов, определенная их поляризация, особенности в спектре и т. п. — подробный анализ этих особенностей всегда необходим. Об искусственности сигнала можно судить и по статистическим свойствам изменения электрического поля радиоволны. Однако, наиболее надежным критерием все же должен служить малый угловой размер.

Сложным является вопрос и о выборе длин волн, на которых следует искать источники искусственных сигналов. Как известно, считается, и вполне заслуженно, что идея о связи с внеземными цивилизациями перешла из области фантастики в область науки с тех пор, как в 1959 г. Д. Коккони и Ф. Моррисон предложили искать сигналы внеземных цивилизаций на природном стандарте длин волн — радиолинии свертонкой структуры атома водорода $\lambda = 21$ см. Впрочем, против этой идеи были и возражения; в частности, отмечалось, что для этой длины волны велико поглощение в межзвездной среде, в связи с чем предполагалась возможность использования гармоник этой радиолинии.

Но есть и другие природные стандарты длин волн, например, радиолинии так называемого Λ -удвоения

молекул гидроксила [ОН]. Собственно, здесь имеются четыре линии, определенные комбинацией Λ -удвоения и сверхтонкой структуры. Средняя длина волны четырех линий $\lambda = 18$ см. Во всех линиях гидроксила межзвездное поглощение значительно меньше, чем в линиях водорода, хотя еще довольно велико.

В последнее время радиолинии гидроксила стали объектом усиленного внимания радиоастрономов и астрофизиков благодаря тому, что на этих линиях обнаружен эффект «природного мазера»: наблюдаются очень узкие (с доплеровской шириной, соответствующей температурному профилю всего в несколько градусов Кельвина), но очень интенсивные (с яркостной температурой, большей 10^{13} град.), и сильно поляризованные линии гидроксила у ряда источников, расположенных вблизи горячих зон ионизованного межзвездного водорода. Необычность поведения этих линий стала причиной того, что они получили название «линий мистериума». Если к тому же учесть, что радионисточники линий «мистериума» обладают наименьшими известными угловыми размерами — порядка тысячных долей угловой секунды (при их расстояниях это соответствует линейным размерам порядка нескольких астрономических единиц), то не приходится удивляться, что на эти источники пало подозрение в их искусственном происхождении.

Мы далеки от утверждения, что источники «мистериума» являются внеземными цивилизациями, но этот пример наглядно показывает необходимость тщательного исследования и обсуждения всех «подозрительных» объектов.

Отметим также, что в сантиметровом и дециметровом диапазонах, наиболее выгодных с точки зрения космической радиосвязи (в этих диапазонах меньше мешающее влияние межзвездной среды; см. гл. I и II), есть и другие молекулярные линии, которые в принципе могут быть использованы для передачи сигналов внеземных цивилизаций. Наконец, такая передача возможна и даже весьма вероятна и в непрерывном спектре ($\lambda \approx 10-50$ см), допускающем значительно большую скорость передачи информации (гл. I и III).

В последнее время всеобщее внимание привлекли открытые 6 августа 1967 г. так называемые «пульсары» — источники радиоизлучения, обладающие удивительно правильной периодичностью повторения импульсов в непрерывном спектре.

История их исследования в первые месяцы после открытия оказалась связанной с проблемой поиска сигналов внеземных цивилизаций. Поэтому стоит рассказать здесь об этом немного подробнее. Пульсарами были названы некоторые объекты, дающие радиоизлучение в широком интервале непрерывного спектра в виде отдельных очень коротких импульсов (продолжительностью порядка нескольких сотых и даже тысячных долей секунды). В промежутках между импульсами излучение пульсаров пока не наблюдалось. Импульсы радиоизлучения разные как по форме, так и по величине, т. е. по мощности излучаемой энергии. Наблюдалась тонкая структура импульсов — у многих пульсаров они состоят из так называемых подимпульсов. Форма импульсов различна у разных пульсаров и даже меняется у одного и того же пульсара. Еще больше меняется величина пульсаров, и временами они вообще пропадают. Во многих, но не во всех случаях излучение импульсов поляризовано. Возможно, что по крайней мере некоторые пульсары дают импульсное излучение и в видимой области спектра. Все это было бы обычным для естественного излучения астрофизических объектов, может быть, даже и обычных звезд. Во всяком случае, некоторые особенности радиоизлучения пульсаров напоминают характеристики спорадического радиоизлучения Солнца. Но одна особенность пульсаров — самая важная особенность, которой они обязаны своим названием — показалась слишком необычной. Дело в том, что была обнаружена очень строгая периодичность повторения импульсов. У первых открытых пульсаров периоды повторения импульсов были близки к одной секунде, причем значение периода каждого пульсара сохранялось в течение года с точностью до седьмого-восьмого знака после запятой. Например, период наиболее известного пульсара CP 19019 равен $1,33730109 \pm 10^{-8}$ сек. Правда,

вскоре выяснилось, что периоды пульсаров систематически увеличиваются (в седьмом знаке в течение года). Именно эта строгая периодичность и привела А. Хьюша, под руководством которого в Кембридже (Англия) были открыты пульсары, к предположению о возможности искусственного происхождения пульсаров, представляющих собой сигналы «маленьких зеленых человечков», как тогда писали в прессе. Поэтому А. Хьюш держал это открытие в строжайшей тайне в течение полугода после обнаружения первого пульсара — крайне редкий случай при современном быстром развитии науки. И только тогда, когда в том же Кембридже были найдены еще три пульсара, это открытие было опубликовано. Почти одновременное обнаружение нескольких внеземных цивилизаций — очень маловероятное событие.

Надо заметить, что существование строгой периодичности в каких-либо естественных процессах, происходящих в телах астрономических масштабов, не столь уж и редкое явление. Очевидными примерами являются периоды вращения планет или двойных звезд. Известны переменные звезды (сравнительно немногочисленная группа звезд типа RR Лиры, принадлежащих к населению I типа), отличающиеся очень высокой стабильностью колебаний: их периоды не меняются сколько-нибудь заметным образом на протяжении миллионов пульсаций. Правда, до сих пор астрономы имели дело с периодами, измеряемыми в лучшем случае часами или сутками, а в пульсарах периоды равны секундам или долям секунд, но это различие, вероятно, не принципиально.

Кроме строгой периодичности у пульсаров решительно ничего не соответствовало ожидаемым признакам сигналов внеземных цивилизаций (см. гл. I и III настоящей книги). Гипотеза об искусственном происхождении пульсаров просуществовала всего несколько месяцев. К концу 1968 г. было известно уже 27 пульсаров с периодами от трех сотых до трех секунд. Свойства пульсаров оказались очень интересными и очень необычными — эти объекты стали отождествлять с вращающимися нейтронными звездами (здесь можно

объяснить и строгую периодичность и увеличение периода). Но к теме настоящей книги пульсары уже, по видимому, не имеют отношения.

Уже сейчас на основе современной теории связи можно обсуждать условия передачи сигналов через космическое пространство, рассмотреть требования, которые должны быть предъявлены к передающим и, что для нас сейчас особенно важно, приемным устройствам и антеннам. Такой анализ, подробно проведенный в гл. III, поможет при составлении программы поиска сигналов выбрать наилучшим образом параметры антенн, полосы частот приемников, периоды обзора. Отсылая читателя за подробностями к гл. III, заметим только следующее. В рассматриваемой проблеме можно различать две основные задачи — непосредственный поиск сигналов («обнаружение искусственных источников») и прием информации от внеземных цивилизаций. В первом случае уровень полезного сигнала может быть и много меньше уровня шумов. Прием таких сигналов возможен при помощи методов усреднений (это часто делается в радиоастрономии), но при этом, естественно, теряется часть информации. Оказывается, что в том случае, если мы рассчитываем на обнаружение сигналов внеземных цивилизаций с помощью современных технических средств, «мощность» цивилизаций может быть на несколько порядков меньше, чем в случае, если мы ставим задачу получения большого количества информации от этих цивилизаций (а расстояние соответственно больше). Это, между прочим, означает, что первые обнаружения сигналов внеземных цивилизаций во всяком случае не могут привести к слишком катастрофическим последствиям.

В настоящей книге не предлагается и не обсуждается какая-либо конкретная программа поиска сигналов внеземных цивилизаций. Выбор такой программы может быть сделан лишь после всестороннего обсуждения возможностей современной радиоастрономической аппаратуры, в частности, и с учетом того, сколько рабочего времени крупнейших радиотелескопов может быть отведено для ее выполнения. Необходи-

димость использования радиointерферометров с базой порядка радиуса земного шара неизбежно потребует международного сотрудничества в этой области.

Авторы, однако, надеются, что анализ проблемы поиска сигналов внеземных цивилизаций, проведенный в этой книге, при всей его неполноте поможет все же в разработке большой и полной программы поиска таких сигналов с большей надеждой на успех, чем у известной программы «Озма» по наблюдению двух близких к Солнцу звезд ϵ Эридана и τ Кита, проведенной в 1959 г. Ф. Дрейком.

Дешифровочный аспект программы поиска сигналов внеземных цивилизаций. Само собой разумеется, что до приема таких сигналов трудно обсуждать вопрос о том, какая информация и в какой форме может в них содержаться. Будут ли это телевизионные изображения (наиболее понятный язык, по крайней мере с нашей точки зрения) или сообщения, построенные по принципу формальной логики, вроде известного языка типа «Линкос», или же нечто совершенно иное, — нам сейчас абсолютно неизвестно, и не стоит на эту тему гадать.

Но вместе с тем возможно, а по нашему мнению, и необходимо уже сейчас сформулировать некоторые проблемы дешифровки неизвестных сообщений. Приведем пример. Допустим, что принято некоторое сообщение и предположим, что это текст, но составленный на неизвестном языке с неизвестным алфавитом, правилами разбиения на фразы и слова и даже с неизвестной кодировкой букв. Известно только, что принята некоторая последовательность сигналов, например, импульсов, имеющих явно искусственное происхождение. Можно ли этот текст расшифровать, понять его смысл? Для этого необходимо (хотя и недостаточно) определить кодировку букв, разбиение на слова и фразы, выяснить структуру грамматики языка, построить его словарь и даже выяснить произношение букв и слов.

Или другой пример. Принято отрывочное сообщение (например, с помехами), про которое можно пред-

положить, что оно является частью изображения (статической телевизионной картинки). Можно ли по этому сообщению восстановить общую картину, т. е. узнать число строк и элементов в каждой строке? Вероятно, многим известен пример с «космограммой Дрейка» (приведена в главе IV), когда последовательностью из 1271 знака (единиц и нулей) были зашифрованы изображения некоторых существ (удивительно напоминающих людей, но только более высоких), живущих на четвертой планете некоей планетной системы. Разгадка этой «космограммы» крайне упрощается тем, что число 1271 разбивается либо на 31 строку по 41 элементу, либо на 41 строку по 31 элементу. Поэтому существует всего две возможности для выбора и ответ находится почти сразу. А если бы несколько первых символов не было принято или экран оказался бы не прямоугольной формы, — как дешифровать сообщение тогда?

Конечно, вполне возможно, что сигналы внеземных цивилизаций несут в себе и понятный нам ключ для дешифровки передаваемых сообщений. Этот вопрос непосредственно связан и с проблемой позывных, которые должны указывать на искусственное происхождение сигналов. Здесь имеется большой простор для самых различных предположений и высказываний. В настоящей книге проблема позывных и простых ключей для дешифровки затрагивается в гл. I, III, IV и VI. Но нам представляется наиболее интересной и заслуживающей внимания проблема дешифровки непонятных сообщений при полном отсутствии декодировочных ключей. Эта проблема и рассматривается в гл. IV, написанной лингвистом.

Предлагаемый способ дешифровки, в сущности, сводится к известному в физике методу построения корреляционных функций (в гл. IV они называются оценочными функциями) для сообщений. В самом деле, существуют известные правила сочетаемости согласных и гласных букв, сочетания слов, принадлежащих различным частям речи — поэтому по корреляционным функциям, построенным для разных символов принятого сообщения, можно судить и о клас-

сах этих символов. Если сообщение представляет собой разбиение на элементы штрихового изображения, то и здесь корреляционные функции позволяют найти отдельные линии, а затем и всю картину. Разумеется, подобная расшифровка связана с большим объемом вычислительной работы и поэтому может проводиться лишь на электронно-счетных машинах. Следовательно, задача расшифровки заключается в построении алгоритма для расчета корреляционных функций и в сравнении их с некоторыми критериями (типа энтропийных), позволяющими выбрать наилучшее решение (для упорядоченных распределений энтропия минимальна). Очевидно и то, что поскольку рассматриваемая дешифровка основана на статистической обработке, для ее проведения в достаточно сложном случае необходим и большой объем материала, т. е. сообщение достаточной длины. Однако простые примеры можно провести и на коротких сообщениях.

Отсылая читателя за подробностями к гл. IV, необходимо, однако, отметить следующее. Здесь скорее рассматривается проблема дешифровки сообщений с лингвистической точки зрения. Читатель, который интересуется только основными принципами такой дешифровки, может ограничиться чтением лишь первых семи параграфов. Остальные четыре параграфа содержат усложненные формы алгоритмов, решающих более сложные задачи. Правда, и здесь мы еще очень далеки от того, чтобы, например, полностью расшифровать длинный текст на неизвестном языке. Но принципы такой расшифровки уже становятся ясными, остальное — дело техники.

* * *

Выше мы попытались кратко обрисовать новую проблему науки — поиск сигналов внеземных цивилизаций (экзосоциологию) и одновременно описали содержание предлагаемой книги. Теперь автор этих строк хотел бы сделать несколько замечаний уже как редактор книги.

По замыслу, каждая глава должна рассматривать вполне определенный аспект проблемы поиска сигнала

лов внеземных цивилизаций. Таким образом, предлагаемая книга не есть сборник статей, а скорее представляет собой своего рода монографию. Но трудность заключается в том, что в такой молодой и еще очень неопределенной области науки как экзосоциология нет еще ни единых мнений, ни установившихся понятий. Не удивительно поэтому, что и точки зрения авторов этой книги по ряду рассматриваемых вопросов не всегда тождественны. Редактор не считал нужным ни навязывать авторам свою точку зрения, ни выступать в роли третейского судьи. Это в свою очередь привело к тому, что ряд вопросов, например, определение понятия цивилизации, сроки энергетического кризиса и т. п., обсуждаются в нескольких главах, причем иногда с различных точек зрения. Читатель сам сможет судить, чьи аргументы покажутся ему более убедительными или может вообще предложить совершенно иные соображения.

Следует, однако, подчеркнуть, что подобных «разногласий» сравнительно немного и в целом книга имеет вполне определенную задачу изложить проблему поиска сигналов внеземных цивилизаций с научной точки зрения так, как она представляется нам в настоящее время, и тем самым стимулировать интерес к этой проблеме.

Предлагаемая книга рассчитана на широкий круг читателей, хотя и не является популярной в обычном смысле этого слова. Авторы старались по возможности выдержать современный научный уровень в каждой из рассматриваемых проблем, но вместе с тем не углубляться в изложение вопросов, понятных только узким специалистам (вероятно, единственным исключением в этом смысле является вторая часть четвертой главы). Трудность книги — в многообразии затронутых проблем: радиоастрономия, теория информации, лингвистика, кибернетические системы, аспекты цивилизации...

Вероятно, многим читателям покажется, что одни разделы изложены слишком бегло, другие, наоборот, чересчур подробно. Одни главы слишком просты, другие усложнены. Но читатель должен иметь в виду,

что в пределах одной книги при большом коллективе авторов трудно изложить на одном уровне столь различные вопросы.

На русском языке уже есть несколько книг, в которых проблема поиска сигналов внеземных цивилизаций занимает важное место (см. ниже). Предлагаемая книга сильно отличается от них. Книги [3] и [4] представляют собой сборники статей и докладов и поэтому не дают полного представления о проблеме, тем более, что они в значительной степени устарели.

Книга У. Салливана больше похожа на репортаж о событиях, связанных с проблемами экзобиологии, чем на сколько-нибудь связное изложение самих проблем, даже если и учесть ее популярный характер.

Наиболее интересна книга И. С. Шкловского [1], но, во-первых, она написана уже несколько лет назад и поэтому многие аспекты проблемы внеземных цивилизаций там не затронуты, во-вторых, она носит более популярный характер, чем настоящая.

Сказанное позволяет нам считать, что предлагаемая книга является первой в научной литературе монографией по проблеме поиска сигналов внеземных цивилизаций.

В заключение отметим, что все авторы считают своим долгом выразить благодарность акад. В. А. Котельникову, сделавшему ряд замечаний, способствовавших улучшению книги, и особенно Л. И. Гудзенко, прочитавшему всю рукопись и сделавшему много ценных замечаний, как по существу затронутых проблем, так и по общему содержанию книги.

ЛИТЕРАТУРА

- † 1. Шкловский И. С., Вселенная, жизнь, разум, изд. 2-е, «Наука», 1965
2. Салливан У., Мы не одни, «Мир», 1967.
- † 3. «Межзвездная связь», сб. статей, под ред. А. Камерона, пер. с англ., «Мир», 1965.
- † 4. «Внеземные цивилизации», Тр. совещания, Бюракан, 20—23 мая 1964 г., изд. АН Арм. ССР, 1965
- † 5. Каплан С. А., Элементарная радиоастрономия, «Наука», 1966.
6. Кларк А., Черты будущего, «Мир», 1966.

Астрофизический аспект проблемы поиска сигналов внеземных цивилизаций

§ 1. Введение

Вероятно, справедливо утверждение о том, что вопрос о поисках внеземных цивилизаций теснейшим образом связан с основными задачами астрофизики на современном этапе. Какая же часть возможной программы поисков полностью совпадает с астрофизическими исследованиями и что можно сказать о специфических для этих поисков наблюдениях?

Точно предсказать основные задачи и направления развития науки о космосе на сколько-нибудь длительный срок довольно затруднительно. Однако на ближайшие годы главная тенденция является совершенно очевидной.

По-видимому, в течение ближайших 5—10 лет будут обнаружены и в какой-то степени исследованы все источники излучения с наибольшим наблюдаемым потоком в каждом электромагнитном диапазоне (А)*. Постановка такой задачи стала реальной благодаря развитию методов регистрации электромагнитного излучения — радиоприемников, болометров, фоточувствительных приемников и материалов, а также счетчиков фотонов. По-видимому, чувствительность этих средств будет в ближайшие годы доведена до своего собственного природного предела (в некоторых диапазонах эта предельная чувствительность уже реализована, например, при

*) Основные положения статьи обозначены жирными буквами.

исследовании космического рентгеновского излучения современные счетчики фотонов регистрируют каждый квант, попадающий в их объем). Достижение предельной чувствительности открывает возможность исследовать различные космические объекты во всем электромагнитном диапазоне, т. е. получать из космоса практически всю содержащуюся в излучении астрофизическую информацию. Это делает реальным решение важной астрофизической задачи:

Выявление и изучение основных (по какому-то параметру) космических объектов (в первую очередь объектов наибольшей светимости (мощности излучения) в данном диапазоне объектов наибольшей массы и объектов, составляющих основную долю вещества в космосе) (Б).

Первоочередной задачей подобного исследования, по-видимому, можно считать получение функции светимости $N_L(L_v)$ и функции масс для всех объектов $N_M(M)$, где L_v — спектральная мощность излучения объекта. К сожалению, решение проблемы А не означает одновременного решения проблемы Б (скорее наоборот)*). Действительно, объекты самой высокой светимости (например, сверхновые звезды, квазизвездные радиогалактики — квазары) являются чрезвычайно редкими во Вселенной. Поэтому среднее расстояние между ними (а следовательно, и наиболее вероятное расстояние до ближайшего источника) очень велико и эти объекты могут не быть самыми яркими. Ближайший квазар 3С 273 имеет в оптическом диапазоне яркость 12,5 звездной величины. Звезд такой величины на небе более четырех миллионов и поэтому долгое время наблюдатели просто не обращали на него внимание; в то же время в радиодиапазоне этот квазар входит в первую сотню наиболее ярких источников и радиоастрономы сразу же его заметили.

*) N_L и N_M — число объектов в единице объема с мощностью излучения в данном диапазоне $L \pm \delta L$ или массой $M \pm \delta M$, соответственно $2\delta L_v$ и $2\delta M$ — единичные интервалы светимости и массы.

Оценим возможность наблюдений в данном диапазоне с помощью функции светимости. Предположим, что

$$N_L \propto L_v^{-n},$$

где степень n можно определить из наблюдений. Число источников в единице объема со светимостью в интервале от $\frac{1}{2} L_v$ до $\frac{3}{2} L_v$ тогда будет

$$N_L \propto L_v^{1-n}.$$

Наблюдаемый поток от ближайшего источника, светимость которого находится в интервале от $\frac{1}{2} L_v$ до $\frac{3}{2} L_v$, есть

$$F_v \propto \frac{L_v}{R^2},$$

где среднее расстояние между данными источниками $R \propto N_L^{-1/2}$.

Отсюда

$$F_v \propto L_v^{\frac{5-2n}{3}}. \quad (1.1)$$

Из этого соотношения следует, что если $n > \frac{5}{2}$, то $\frac{5-2n}{3} < 0$ и источники самой слабой светимости будут преобладать среди источников с наибольшим наблюдаемым потоком; если же $n < \frac{5}{2}$, то, наоборот, самые мощные источники будут преобладать среди самых ярких.

В частности, например, первый случай имеет место при сопоставлении среднего излучения нормальных галактик, радиогалактик и квазаров в оптическом диапазоне. Последних объектов мало, т. е. здесь n велико и поэтому нормальные галактики оказываются наиболее яркими среди всех внегалактических оптических источников.

С другой стороны, если рассмотреть оптическое излучение только нормальных галактик, то величина $n < \frac{5}{2}$ и поэтому здесь самыми яркими при наблюдениях являются наиболее мощные объекты. То же

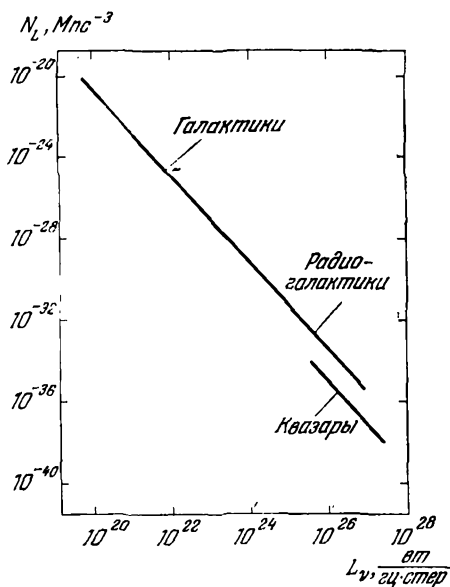


Рис. 1. Функция светимости внегалактических радионисточников.

самое наблюдается для внегалактических радионисточников. На рис. 1. приведена функция светимости радионисточников [46], которая показывает, что в большом интервале светимостей величина $n \sim 2,2$ и, следовательно, ярчайшими являются также наиболее мощные объекты, которые поэтому главным образом и изучаются в настоящее время. К сожалению, провести подобный анализ для функции масс пока не удастся, поскольку они известны еще очень плохо. Выяснен только закон распределения по массам для звезд [47]. Это распределение также хорошо представляется степенной зависимостью с показателем $n \approx 2,35$.

Проблема **Б** подчеркивает основную тенденцию развития астрофизических исследований на ближайшие годы. В частности, если деятельность внеземных цивилизаций определяет мощность излучения некоторых астрономических объектов, то эти цивилизации могут быть нами обнаружены. Исходя из того, что в ближайшие годы все диапазоны электромагнитного излучения станут доступными для космических исследований, можно наметить программу и оценить вероятность успеха поисков внеземных цивилизаций.

§ 2. Основная дилемма

Пожалуй, основным исходным моментом для рассматриваемой проблемы является следующая дилемма:

Большая вероятность полной цивилизации Вселенной и отсутствие в настоящее время каких-либо наблюдаемых проявлений космической деятельности разумных существ (**В**).

Действительно, современные данные о количестве планетных систем, условиях возникновения и развития жизни на планетах позволяют считать жизнь довольно распространенным и закономерным явлением во Вселенной. Подробный анализ указанных вопросов можно найти в книгах [1, 2, 3]*).

Согласно большинству оценок возраст нашей планетной системы и Солнца (отсчитываемый от момента их конденсации) составляет 4—6 млрд. лет. При этом существенно, что Солнце и наша планетная система являются объектами второго поколения, но так как возраст наиболее старых объектов наблюдаемой части расширяющейся Вселенной (точнее, ее объектов первого поколения), по-видимому, не менее 10 млрд. лет, то возможно существование планетных

*) В частности, недавно были получены новые палеонтологические данные, убедительно показывающие, что зарождение жизни на Земле около 3 миллиардов лет назад происходило многими независимыми путями [4].

систем с возрастом на миллиарды лет бóльшим, чем наша. Такие представления приводят к выводу о возможности существования во Вселенной цивилизаций, на миллиарды лет опередивших нас в своем развитии. Если исходить из современных темпов развития цивилизации на Земле, то мы вправе ожидать, что за такие космогонические сроки возможна едва ли не полная сознательная реорганизация вещества в нашей части Вселенной.

В то же время имеющиеся в нашем распоряжении астрономические данные на первый взгляд не дают никаких указаний на наличие такого рода деятельности. Нам кажется, что анализ альтернативы В может явиться основой для обсуждения программы поиска внеземных цивилизаций (ВЦ). Постараемся рассмотреть различные стороны этой альтернативы с целью уточнения вопросов, перечисленных в связи с ее обоснованием.

Возможны два решения альтернативы:

- 1) либо современные данные об отсутствии «сверхцивилизаций» неверны;
- 2) либо есть какая-то принципиальная причина, тормозящая развитие каждой цивилизации.

§ 3. Полнота и достоверность современных астрофизических данных

Рассмотрим обоснованность первого предположения, имеющего астрофизический характер.

Как уже отмечалось, серьезных сомнений в существовании большого количества планетных систем не существует (хотя и не следует забывать, что планеты с массой порядка массы Земли не могут непосредственно наблюдаться с помощью современных телескопов) (см. [1, 2, 3]). Оценки количества планет, обладающих подходящими для возникновения и развития жизни условиями, также не содержат никаких указаний на исключительное положение Земли (см. [1, 2, 3]).

Как было упомянуто выше, Солнце и наша планетная система являются объектами второго поколе-

ния, но если это не так, то имеется некоторая вероятность того, что Земля — старейший объект такого типа в наблюдаемой части Вселенной и, следовательно, наша цивилизация — также самая древняя.

На чем основаны современные представления о возрасте Солнечной системы?

Большинство звезд, близких по своим физическим параметрам к Солнцу, длительное время находятся в стационарном состоянии, которое характеризуется постоянством их светимостей и размеров. Излучаемая энергия компенсируется энерговыделением в результате ядерных реакций, происходящих внутри звезды. На основе этих представлений развита теория звездной эволюции, согласно которой Солнце может оставаться в стационарном состоянии около 13 млрд. лет, т. е. в течение всего времени эволюции Метагалактики. С другой стороны, возраст земных пород и метеоритов, определяемый по химическому анализу радиоактивных изотопов и продуктов их распада, составляет 4—5 млрд. лет. Эта цифра в настоящее время чаще всего и принимается за возраст нашей планетной системы и Солнца, так как современные теории образования планетных систем свидетельствуют об одновременной конденсации планет и центральной звезды из межзвездной газопылевой среды.

Правда, в последние годы начали появляться работы, в которых приводится существенно большее значение величины возраста Земли и метеоритов (см. [5]). Так, в работе Фишера [5] калий-аргоновым методом определен возраст некоторых железных метеоритов, который, как оказалось, доходит до 10 млрд. лет. Тем же методом для земных горных пород получен возраст до 10,8 млрд. лет [6]. Хотя эти и другие аналогичные данные пока еще нельзя считать сколько-нибудь достоверным доказательством более длинной шкалы эволюции нашей планетной системы, но, видимо, и нет оснований отбрасывать их как заведомо неверные.

Другая сторона этой проблемы связана с химическим составом планет. Поскольку для конденсации земноподобных планет необходимо достаточное

количество тяжелых элементов в межзвездной среде, то очень важное значение имеет еще не решенный однозначно вопрос об эволюции этой среды и образовании тяжелых химических элементов вообще.



Рис. 2. Квазар 3C 9.

В соответствии с современными данными об эволюции наблюдаемой части Вселенной предполагается, что все химические элементы образовались в результате ядерных реакций из первоначально чисто водо-

родной плазмы. До последнего времени считалось, что эти процессы протекают главным образом в недрах звезд, а наиболее тяжелые элементы образуются при реакциях во время вспышек сверхновых звезд. В дальнейшем эти элементы выбрасываются в межзвездную среду [1]. Такой механизм свидетельствует в пользу гипотезы, относящей земноподобные планеты к объектам второго поколения.

Однако за последние несколько лет обнаружен новый класс объектов первого поколения, также обладающих большим количеством тяжелых элементов — это «квазары». Квазары прежде всего замечательны тем, что обладают самой высокой мощностью излучения среди известных во Вселенной источников. Благодаря этому их удается наблюдать на гигантских расстояниях, а следовательно, учитывая скорость распространения света, изучать свойства Вселенной в отдаленном прошлом. На рис. 2 приводится фотография одного из самых далеких квазаров 3С 9. Вследствие наблюдаемого расширения Вселенной спектральные линии таких объектов смещены в сторону более длинных волн. Для 3С 9 величина этого смещения $z = \frac{\Delta\lambda}{\lambda} \approx 2$, так что все длины волн по сравнению с лабораторными увеличиваются в $1 + z = \frac{\lambda}{\lambda_0} \approx 3$ раза. Вычисление интервала времени между моментом испускания излучения и моментом наблюдения для удаленного объекта существенным образом зависит от принимаемой космологической модели. Так, например, для весьма вероятной модели Эйнштейна — де Ситтера (кривизна пространства $k=0$, параметр ускорения $q_0 = \frac{1}{2}$) время распространения светового луча равно

$$\tau = \frac{2}{3H_0} \frac{(1+z)^{3/2} - 1}{(1+z)^{3/2}}. \quad (1.2)$$

Здесь H_0 — постоянная Хаббла (для малых красных смещений z расстояние до объекта $\frac{cz}{H_0}$). Величина

$H_0 \sim 30$ км/сек · млрд. св. лет. В этой модели свет от источника 3С 9 идет до Земли около $t = 5,3$ млрд. лет. (Необходимые данные для расчетов по другим моделям можно найти в книге [7].)

Наиболее важным в рассматриваемом аспекте, как уже отмечалось, является обнаружение в этих объектах нормального химического состава [8]; другими словами, средняя распространенность химических элементов (по крайней мере, наиболее обильных) в квазарах близка к той, которая имеет место и в окрестностях Солнца. В то же время обнаружено, что квазары расположены в областях, где количество обычных галактик наименьшее (между скоплениями галактик). По-видимому, они образуются непосредственно из межгалактической среды. Возможно, что тяжелые химические элементы синтезируются в самих квазарах, так как вполне вероятно, что условия, возникающие при вспышках этих объектов, еще более благоприятны для ядерного синтеза, чем при взрывах сверхновых звезд. Поразительно, однако, сходство химического состава различных квазаров. Поэтому не менее вероятно предположение, что синтез тяжелых элементов произошел при еще более ранних стадиях развития Вселенной и межгалактическая среда, из которой образуются квазары, имеет такой же состав, как и межзвездная. Следовательно, время существования тяжелых элементов, пригодных для образования земноподобных планет, может быть сравнимо с возрастом наблюдаемой части Вселенной.

Все приведенные выше новые данные указывают на возможность существования планетных систем с возрастом, близким к возрасту старейших объектов во Вселенной. Но, пожалуй, наиболее уверенные данные в пользу того, что Земля не является самой старой планетой, дают некоторые наблюдения в свете современной теории звездной эволюции. Как уже отмечалось выше, звезды после их конденсации из межзвездной среды долгое время находятся в квазистационарном равновесии, поддерживая расход энер-

гии на излучение протекающими в их недрах ядерными реакциями. Длительность пребывания в этом равновесии тем больше, а светимость тем ниже, чем меньше масса звезды. После «выгорания» водорода ядро звезды начинает сжиматься, температура его повышается, а радиус звезды начинает расти. Совокупность звезд разной массы, находящихся в равновесии, поддерживаемом термоядерными реакциями

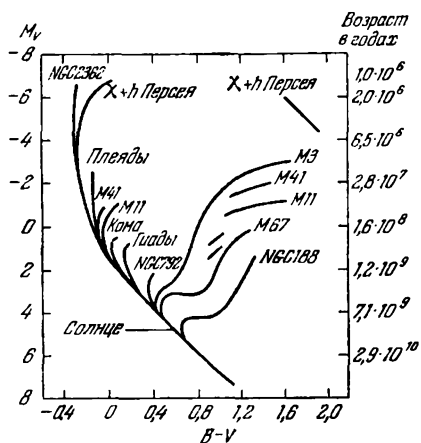


Рис. 3. Диаграмма Герцшпрунга—Рессела для некоторых звездных скоплений.

синтеза элементов (главным образом гелия), образует так называемую «главную последовательность». Звезды с выгоревшим водородом переходят из семейства звезд главной последовательности в семейство «красных гигантов». Время жизни на главной последовательности и наличие в какой-то группе звезд красных гигантов дают, очевидно, возможность определить возраст этой группы. На рис. 3 приведена так называемая диаграмма Герцшпрунга—Рессела для 11 звездных скоплений (зависимость цвет—светимость). По горизонтальной оси откладывается «цвет» звезд (разность между фотографической и визуальной

звездной величиной), по вертикальной оси — абсолютная визуальная звездная величина. Огибающая слева приведенного семейства кривых представляет собой главную последовательность и одновременно отражает распределение звезд по цвету и светимости для наиболее молодого из взятых скоплений звезд NGC 2362. Справа вдоль вертикальной оси указан возраст, соответствующий длительности нахождения звезды данной светимости на главной последовательности. Стрелкой показано положение Солнца на диаграмме. Ответвления от главной последовательности вправо — вверх соответствуют распределению по цвету и светимости звезд — красных гигантов для каждого скопления. Очевидно, что по точке ответвления можно определить возраст скопления. В частности, из рис. 3 видно, что точка ответвления для скопления NGC 188 лежит ниже Солнца, откуда следует, что возраст этого скопления больше, чем время нахождения Солнца на главной последовательности. Этот вывод подтверждается и некоторыми другими данными. По своему положению в Галактике и скорости движения относительно галактического центра Солнце относится к плоской или промежуточной подсистемам звездного населения, которые можно рассматривать как объекты второго поколения. Звезды первого поколения (сферическая подсистема), образовавшиеся сразу же при конденсации галактик, выбрасывают газ, обогащенный тяжелыми элементами. Этот газ, смешиваясь с межзвездным газом, оставшимся от первой конденсации, оседает к плоскости вращения галактики и, конденсируясь, образует звезды плоской и промежуточной подсистем — второе поколение. Спектральные исследования Солнца как раз свидетельствуют о характерном для второго поколения обилии тяжелых элементов.

Рассмотрим теперь второй тезис дилеммы — об отсутствии каких-либо проявлений деятельности суперцивилизаций. Насколько полны современные астрофизические данные в этой точки зрения?

Постараемся оценить процент уже известной астрофизической информации от полного количества

информации, которая может содержаться во всем электромагнитном диапазоне.

Современная астрофизика обладает необычайно богатой и разнообразной информацией. Оптические и радиокаталоги содержат тысячи звезд, галактик, туманностей. Для многих из этих объектов изучены химический состав, физическое состояние вещества; гипотезы относительно их эволюции удовлетворительно согласуются со статистическим анализом данных наблюдений. Наблюдательные средства астрономии стали настолько эффективными, что позволяют исследовать источники излучения на предельно далеких расстояниях в миллиарды световых лет. Может сложиться впечатление, что многообразие наблюдательного материала на базе имеющихся гипотез дает стройную картину эволюции Вселенной, причем практически все важнейшие объекты Вселенной уже известны и остается выяснить лишь некоторые детали.

Автору кажется, что эта весьма распространенная точка зрения неправильна, хотя состояние наблюдательной техники таково, что время, когда строение Вселенной в основных чертах станет ясным, вероятно, уже недалеко.

Можно привести много примеров очень важных, принципиальных открытий в астрофизике, сделанных за последние годы (например, обнаружение квазаров, реликтового фона, составляющего основную долю электромагнитного излучения, молекулярных генераторов линий гидроксила на волне 18 см, пульсаров). Весьма возможно, что некоторые из этих объектов могут иметь прямое отношение к проблеме поиска суперцивилизаций. С другой стороны, мы еще почти ничего не знаем о количестве и состоянии твердого вещества во Вселенной.

Как отмечалось в начале главы, получение качественно новой информации о космических объектах в ближайшие годы может быть связано с освоением новых диапазонов. Какой процент от всех диапазонов электромагнитного излучения уже освоен? Задача поиска основных источников излучения в каждом

диапазоне далеко не исчерпывает все многообразие информации, которое можно получить о космических объектах, но, как будет видно из дальнейшего, даже эта простейшая задача еще не решена. Поэтому процент освоенных диапазонов можно рассматривать как верхнюю оценку количества известной информации. Именно таким образом и следует понимать термин «освоенность диапазона».

Будем считать диапазон освоенным, если для данной длины волны проведен поиск источников, охватывающий более 30% от площади небесной сферы, причем в результате обзора было обнаружено более 100 космических объектов. Здесь придется различать два случая:

1. Поиск объектов, излучающих широкий спектр (ширина спектра $\Delta\nu \sim \nu$).

2. Поиск объектов, излучающих в узких спектральных линиях.

Очевидно, вторая задача несравненно более сложна, чем первая, так как требует перекрытия узкополосными фильтрами всего электромагнитного диапазона. Грубо говоря, во втором случае необходимо произвести в $\frac{\nu}{\Delta\nu}$ больше измерений. Например, линии радионизлучения межзвездного гидроксила OH на волне 18 см от некоторых объектов имеют $\frac{\nu}{\Delta\nu} = 3 \cdot 10^6$.

Задача такого узкополосного обзора представляет очень большой интерес и для астрофизики, и для проблемы поиска цивилизаций. Но в настоящее время ни одного узкополосного обзора неба ни в оптическом, ни в радиодиапазоне не сделано (исключение, пожалуй, составляет обзор всего неба в линии межзвездного водорода $\lambda = 21$ см в полосе около 1 МГц со спектральным разрешением ~ 10 кГц). Поэтому процент имеющейся информации о чисто монохроматических источниках пока еще очень мал.

Несравненно более простая задача — поиск широкополосных источников. Для таких источников число освоенных диапазонов (например, число октав) опре-

деляется выражением $l \ll \ln \frac{v_2}{v_1}$, где v_1 и v_2 — минимальная и максимальная частота поиска. Очевидно, что процент уже освоенных диапазонов будет

$$x = \frac{\ln \left(\frac{v_2}{v_1} \right)_{\text{рад}} + \ln \left(\frac{v_2}{v_1} \right)_{\text{опт}}}{\ln \frac{v_2}{v_1}}, \quad (1.3)$$

где $\left(\frac{v_2}{v_1} \right)_{\text{рад}}$ — отношение максимальной к минимальной частот обзоров, выполненных в радиодиапазоне, $\left(\frac{v_2}{v_1} \right)_{\text{опт}}$ — та же величина в оптическом диапазоне, а величина $\frac{v_2}{v_1}$ в знаменателе определяется максимальной и минимальной частотой для поиска астрономических объектов в будущем.

До настоящего времени обзоры в радиодиапазоне проводились в интервале частот от 40 до 400 Мгц, т. е. $\left(\frac{v_2}{v_1} \right)_{\text{рад}} = 10$.

В оптическом же диапазоне фотографирование и изучение отдельных источников охватывает интервал от 3000 до 6000 Å, т. е. $\left(\frac{v_2}{v_1} \right)_{\text{опт}} = 2$.

В остальных диапазонах имеются лишь отдельные наблюдения небольших участков неба, составляющие ничтожный процент от всей возможной информации.

Чему равен знаменатель (1.3)? Низкочастотный предел астрофизических исследований определяется довольно уверенно. Минимальная частота $v_1 \sim 1$ Мгц, так как для более низких частот межзвездная среда становится непрозрачной и на этих частотах возможны только наблюдения объектов, расположенных в ближайших окрестностях нашей планетной системы.

Труднее определить высокочастотный предел, который, по-видимому, связан с квантовой природой электромагнитного излучения. С ростом частоты энергии каждого регистрируемого кванта растет, а из-за ограниченности энергетических ресурсов астрономических источников с увеличением частоты количество

квантов, попадающих в приемник, уменьшается. Подробнее на оценке частоты ν_2 мы остановимся дальше. Пока примем $\nu_2 \sim 10^{18}$ гц (длина волны 3 Å). Тогда отношение $\frac{\nu_2}{\nu_1} = 10^{12}$, а процент освоенных диапазонов

$$x = \frac{\lg 10 + \lg 2}{12} = 11\%, \quad (1.4)$$

т. е. и для задачи поиска широкополосных источников мы обладаем пока лишь небольшой частью информации. Важно отметить, что из 89% недостающей информации 42% приходится на интервал 10^9 — 10^{14} гц (сантиметровые, миллиметровые, субмиллиметровые и инфракрасные волны) и 25% на интервал 10^{15} — 10^{18} гц (ультрафиолетовое и рентгеновское излучение).

Заметим, что границы всего интервала электромагнитных волн ν_1 и ν_2 определяются довольно неуверенно. Однако наиболее вероятно, что мы его заузили. Поэтому даже 11% могут оказаться несколько завышенной оценкой.

Рассмотрим теперь вопрос о количестве космических источников, которые могут быть обнаружены в том или ином диапазоне электромагнитных волн. Как уже указывалось ранее, чувствительность некоторых приемников радиации в настоящее время близка к своей предельной величине, определяемой квантовой природой электромагнитного излучения и фоном космического излучения. Поэтому при исследовании и поисках источников малых угловых размеров успех будет зависеть от количества поступающих на единицу площади приемника квантов и возможности пространственного разрешения источников друг от друга.

С длинноволновой стороны число источников будет определяться угловым разрешением используемой антенны. Число диаграмм направленности антенны, укладывающихся на небесной сфере, будет

$$N_1 \leq \frac{A_\nu}{\lambda^2} = \frac{A_\nu \nu^2}{c^2}, \quad (1.5)$$

где A_v — эффективная собирающая площадь телескопа. В радиодиапазоне для предельно возможных антенн $A_v \propto \lambda^2$. Это связано с тем, что относительная точность изготовления отражающих поверхностей примерно одинакова, т. е. $\frac{\varepsilon}{D} \sim \text{const}$, где D — диаметр зеркала, а ε — средняя ошибка поверхности, причем для работы зеркала в данном диапазоне необходимо, чтобы $\varepsilon \leq 0,1 \lambda$. Таким образом, оказывается, что в рассматриваемом диапазоне предельное число источников N_1 не зависит от длины волны.

Для задачи поиска, например, сотни ярчайших источников в каждом диапазоне предельные антенны, по-видимому, не потребуются. Однако и здесь, полагая $N_1 \sim 100$, мы должны при переходе от одного диапазона к другому изменять эффективную площадь $A_v \propto \lambda^2$.

Тогда соотношение (1.5) позволяет сделать важный вывод. Как при работе с предельными инструментами, так и для задачи поиска одного и того же числа ярчайших источников в каждом диапазоне количество ожидаемой информации пропорционально $\ln \frac{\nu_2}{\nu_1}$ и, следовательно, приведенные выше оценки по формуле (1.3) справедливы.

Для наблюдений с коротковолновой стороны (рентгеновский, γ -диапазоны) возможно создание аппаратуры, регистрирующей каждый приходящий квант и, кроме того, определяющей направление, откуда он пришел. Поэтому число источников, которые можно обнаружить за время наблюдения τ , не может превышать количества квантов от этих источников, попавших в приемник,

$$N_2 \leq \frac{c\rho A_v \tau}{h\nu}. \quad (1.6)$$

Здесь ρ — полная плотность электромагнитной энергии в данном диапазоне волн в единице объема от всех источников. Согласно данным измерений для длин волн, более коротких, чем в оптическом диапазоне, $\rho < 10^{-12}$ эрг/см³. Величина A_v (например,

поперечное сечение счетчиков γ -излучения) уже практически не меняется с длиной волны и поэтому N_2 убывает с ростом частоты. Очевидно, частота, для которой $N_1 \sim N_2$, и будет характеризовать ту определенную частоту ν_2 , выше которой может содержаться лишь небольшая часть информации. (Возможность увеличения A_ν с частотой, видимо, неприемлема из-за больших технических трудностей.)

Таким образом, приравнявая (1.5) и (1.6) и полагая A_ν величинами одного порядка, находим

$$\nu_2 \sim \left(\frac{\rho \epsilon^2 \tau}{h} \right)^{1/2}. \quad (1.7)$$

Учитывая слабую зависимость ν_2 от принимаемого значения параметров, положим $\rho \leq 10^{-12}$ эрг/см³, время обзора $\tau \sim 1$ год $\sim 3 \cdot 10^7$ сек и получим, что величина $\nu_2 \leq 5 \cdot 10^{17}$ гц.

Повторим кратко выводы, вытекающие из этого параграфа: несмотря на большие успехи астрофизики, мы пока еще располагаем недостаточной информацией, чтобы отрицать возможность существования суперцивилизаций на основании того, что до сих пор они не обнаружены. Далее будет рассмотрен вопрос о том, что даже некоторые из уже известных объектов (например, квазары) могут являться продуктом деятельности таких цивилизаций. С другой стороны, данные астрофизики весьма надежно свидетельствуют в пользу существования планетных систем, существенно более старых, чем наша. Это дает основания для серьезного обсуждения программы поиска внеземных цивилизаций.

Мы рассмотрели некоторые астрофизические стороны основной дилеммы (В) и пришли к выводу, что скорее всего она возникла как результат недостаточности данных наблюдений в настоящее время. Если это действительно так, то весьма важно было бы знать, какого рода астрофизические проявления деятельности суперцивилизации можно ожидать. Подобная проблема, вероятно, может быть решена путем исследования некоторых наиболее общих закономерностей развития цивилизаций в течение космогониче-

ских сроков. Следует отдавать себе ясный отчет в том, что в этой области мы еще очень мало знаем. С другой стороны, мы не можем сделать практически ни шагу, если не будем опираться на какие-то предположения. Мы не сомневаемся в возможности и целесообразности некоторой формализации закономерностей в любой области деятельности нашей цивилизации. По-видимому, этот метод может оказаться полезным и в рассматриваемом вопросе. Некоторые соображения об этом приведены в следующем параграфе *).

§ 4. Понятие цивилизации, главные закономерности и характер их развития

Основное, что нас интересует, — это максимальный уровень развития и общий характер деятельности цивилизаций, на которые можно рассчитывать, составляя первоначальную программу поиска. Исходя из этого, можно было бы сделать какие-то выводы относительно проявлений этой деятельности в космических масштабах и проанализировать возможность их обнаружения с помощью современных средств.

Пожалуй, главным фактором, надежно и безусловно установленным современной астрофизикой, является единство всех основных законов природы в масштабах всей наблюдаемой части Вселенной и в течение всего охватываемого наблюдениями времени ее эволюции. Поэтому мы вправе полагать, что известные нам физические законы известны и суперцивилизациям и используются ими. Безусловно, они могут знать и много больше, но их знания во всяком случае должны содержать то, что знаем мы. Более того, по-видимому, объем наших познаний является обязательным этапом в начальном развитии любой технологической цивилизации. Это делает возможным попытку сформулировать хотя бы в самом грубом приближении некоторые понятия, связанные с внеземными цивилизациями.

*) См. также гл. V и VI.

Для дальнейшего очень важно дать функциональное определение понятия цивилизации. В [1] (стр. 125—132) подробно обсуждается функциональное определение понятия жизни, предложенное А. А. Ляпуновым [9]:

«высокоустойчивое состояние вещества, использующее для выработки сохраняющих реакций информацию, кодируемую состояниями отдельных молекул» (Г).

Это определение довольно точно отражает сущность понятия, однако, на наш взгляд, оно обладает значительным недостатком — ничего не говорит о закономерностях, в силу которых происходит зарождение, развитие и усложнение различных форм жизни. По-видимому, жизненный путь каждого индивидуума можно рассматривать как некоторый случайный процесс, зависящий от взаимодействия с внешней средой и состояния живого объекта в каждый данный момент. Эволюция видов в этом случае представляет собой некоторую статистическую закономерность в развитии индивидуумов. Очевидным результатом эволюции является процесс накопления информации и возможностей ее использования. Поэтому, как нам кажется, основной статистической закономерностью живых организмов является стремление получить максимум информации об окружающем и самом себе (Д).

Для низших форм жизни эта закономерность реализуется благодаря естественному отбору. Эта же закономерность, по-видимому, является единственным стимулом развития для высших форм цивилизации.

Существенным отличием высших форм жизни является способность абстрактного анализа собираемой информации. С развитием форм жизни существенную роль начинают играть системы живых организмов. Однако конкретизация количества организмов и структуры цивилизации на высшем уровне вряд ли уместна в настоящее время. Таким образом, отталкиваясь от определения жизни, можно предложить

следующее функциональное определение цивилизации на высшем уровне:

высокоустойчивое состояние вещества, способное собирать, абстрактно анализировать и использовать информацию для получения максимума информации об окружающем и самом себе и для выработки сохраняющих реакций (Е).

Конкретизация метода кодирования информации в данном случае нам также кажется неуместной. Под информацией об окружающем и самом себе видимо следует понимать любые сведения о живой и неживой природе (включая цивилизации), науку, технику, культуру, искусство. (Вероятно, есть еще и другие области, о которых мы пока не имеем никакого понятия.)

Если принять определение Е, то важнейшими параметрами, характеризующими степень и характер развития цивилизации, являются количество информации и скорость ее прироста (например, время удвоения). В рамках современных представлений (в отличие от соображений, например, фон Хорнера [2], стр. 278) нам кажется, что определение Е обеспечивает неограниченную возможность развития цивилизаций. Основные гипотезы о пределе развития цивилизаций по фон Хорнеру: 1) полная гибель всего живого, 2) гибель сознательной жизни, 3) вырождение, 4) потеря интереса, — по-видимому, являются весьма существенными для каждой цивилизации на определенном этапе развития, но в то же время нет никаких доказательств того, что эти причины гибели являются принципиально неустраняемыми во всех случаях для всех цивилизаций. Единственной причиной прекращения развития с точки зрения определения Е могло бы быть конечное количество разнообразия информации во всех областях. Однако с точки зрения современных представлений это крайне маловероятное предположение.

Весьма существенным для программы поиска в то же время является конечность качественно

различающейся информации в отдельных областях (что, конечно, не означает конечности во всех областях). Возможно, что одной из таких областей с конечным количеством информации является наука о космосе в нашем современном представлении. Покажем это на следующем примере. Мы уже говорили, что современные методы астрофизики позволяют изучать свойства различных объектов во Вселенной, удаленных от нас на миллиарды световых лет. Для таких расстояний само понятие измерения длины и времени распространения света неоднозначно и самым существенным образом зависит от модели Вселенной. Основным методом оценки расстояний до очень далеких объектов является определение изменения длины волны спектральных линий их излучения (по отношению к лабораторным длинам волн), т. е. величина z . Как уже говорилось выше, в настоящее время получены спектры источников с красным смещением $z \sim 2$. В то же время наиболее слабые из наблюдающихся (по их непрерывному спектру) источников радионизлучения могут иметь существенно большие z . Если бы не было поглощения или рассеяния электромагнитного излучения при его распространении в межгалактической среде, то современные крупнейшие радиотелескопы могли бы обнаружить объекты типа квазаров до $z \sim 30$, а имеющиеся проекты радиотелескопов позволяют продвинуться еще дальше. Однако расчеты и статистический анализ наблюдений радиоисточников показывают, что это не так.

По-видимому, главным фактором, препятствующим исследованию очень далеких источников, является рассеяние электромагнитного излучения на свободных электронах межгалактической среды и среды галактик. Этот эффект, как показано в работе [10], приводит к наиболее вероятному предельному расстоянию, до которого могут наблюдаться источники, соответствующему значению $z \sim 5$ (величина, получающаяся для модели с положительной кривизной, $q_0 = 1$, $H_0 = 300 \text{ км/сек} \cdot \text{млрд. св. лет}$, и что соответствует средней плотности межгалактической среды в настоящее время $\rho_0 \sim 4 \cdot 10^{-29} \text{ г/см}^3$). Хотя прямого оп-

ределения плотности межгалактической среды пока еще не удается осуществить, статистические исследования источников радиоизлучения показывают, что количество слабых источников меньше, чем этого можно было бы ожидать при отсутствии эффекта рассеяния. Таким образом, вывод о достижении предельно возможного расстояния имеет некоторое экспериментальное подтверждение *).

В объеме сферы, характеризуемой предельным z , находится конечное количество вещества, т. е. конечное количество космических объектов. Учитывая также единообразие в строении небесных тел в разных участках Вселенной, о котором уже говорилось выше, весьма вероятным представляется вывод о возможности за конечное время изучить основные свойства этих объектов.

Весьма вероятно, что многие основные законы будут установлены в ближайшее десятилетие в связи с упомянутой выше тенденцией развития наблюдений (А). Таким образом, благодаря объективной конечности информации наук о космосе для высокоразвитых цивилизаций действительно возможна потеря интереса к этим наукам. Это, в частности, может быть дает ключ и к решению дилеммы (В) — отсутствие сплошной цивилизации по причине потери интереса к космическим исследованиям. Здесь мы, конечно, понимаем под космическими исследованиями исследования в смысле современной астрофизики. Может быть, имеются какие-то направления, связанные с наукой о космосе, о которых нам в настоящее время трудно догадываться (например, проблемы, связанные с универсальными физическими постоянными) и в которых имеется перспектива получить неограниченное количество информации.

Еще раз подчеркнем, что задача полного качественного познания законов космоса в существенной степени упрощается подобием космических объектов

*) Другая причина, затрудняющая наблюдения далеких источников — поглощение их излучения более близкими. Уже при $z \approx 2$ вероятность того, что на луче зрения находится более одного источника, близка к 1.

в различных областях Вселенной, которое следует из всех имеющихся в настоящее время астрономических данных. Пожалуй, единственным типом объектов, которые могут не подчиняться этой закономерности, могут быть сами цивилизации. В таком случае для обеспечения максимума прироста информации одним из важнейших каналов получения информации должен быть обмен информацией между цивилизациями. С точки зрения современных представлений обмен информацией через космическое пространство наиболее целесообразно осуществлять с помощью электромагнитного излучения. Кроме того, ясно, что наиболее общими факторами, сопровождающими деятельность суперцивилизации, должны быть использование в гигантских масштабах массы и энергии.

При попытках отличить деятельность цивилизаций от проявления естественных процессов в космосе, по-видимому, также следует руководствоваться приведенным выше определением понятия цивилизации.

Дать в настоящее время какую-либо количественную оценку максимального уровня развития суперцивилизаций кажется невозможным, однако, учитывая перспективу освоения всего электромагнитного диапазона и резкого увеличения благодаря этому количества астрономической информации, можно надеяться на возможность получения такой оценки.

Современные же данные астрофизики не противостоят никакому, пусть даже самому фантастическому уровню развития. В настоящее время можно даже обсуждать вопрос о том, не является ли факт расширения наблюдаемой части Вселенной результатом сознательной деятельности суперцивилизации? Согласно современным моделям расширяющейся Вселенной около 10 млрд. лет назад все вещество находилось в сверхплотном состоянии. Исключается ли тем самым непрерывное существование цивилизаций в более ранние этапы развития 20, 100, 1000 млрд. лет назад и могли ли они пережить момент, когда Вселенная находилась в сжатом состоянии? И только в том случае, если выяснится, что условия для появления и развития жизни во Вселенной до начала рас-

ширения всегда были неблагоприятны, мы будем вправе оценить возраст самых старых цивилизаций в несколько миллиардов лет.

Можно ли в общих чертах охарактеризовать развитие цивилизации за космогонические сроки? Известно, что многие основные параметры, характеризующие развитие цивилизации на Земле, увеличиваются по экспоненциальному закону (см. гл. V). Время удвоения научной и технической информации составляет около 10 лет, удвоения энергетики, сырья, населения — около 25 лет. Поэтому экстраполяция современных темпов развития общества даже на ближайшее будущее приводит нас к парадоксам.

Так, например, в книге известных американских специалистов по термоядерным реакциям [11] обращается внимание на то, что количество энергии, вырабатываемой на Земле, не может быть очень большим. Существует определенная верхняя граница. Земля ежесекундно поглощает (и излучает обратно) около $5 \cdot 10^{23}$ эрг солнечной радиации. Для того чтобы климат Земли существенно не изменился, необходимо ограничить вырабатываемую на Земле энергию примерно одним процентом этого количества. Если современная энергетика составляет около $4 \cdot 10^{19}$ эрг/сек, то при среднем приросте мощности 4% в год предельный уровень будет достигнут уже через 125 лет! Этот предел можно несколько увеличить, если непосредственно использовать солнечное излучение. Но для этого необходимо покрыть значительную часть земной поверхности преобразователями солнечной энергии, что не представляется реальным.

Подчеркнем принципиальную неустранимость указанной трудности, вытекающую из термодинамики. Действительно, вся расходуемая энергия неизбежно превращается в тепло. Как же развиваться дальше? Можно предположить два выхода: либо спустя 125 лет строго поддерживать энергетику на постоянном уровне, либо выносить все виды деятельности человека, связанные с большими затратами энергии (в первую очередь промышленность и крупные научные эксперименты) в космос. Нам кажется, что первый вариант

совершенно неприемлем, так как он фактически означает искусственное торможение развития. Вторая возможность представляется в настоящее время уже вполне реальной.

Аналогичный вывод об объективной необходимости выхода в космос можно проследить видимо и по другим показателям деятельности человека (рост населения, химическое и радиоактивное загрязнение океана, застройка поверхности Земли, расход неядерных видов топлива, сокращение биосферы и т. д.). Но энергетические трудности, вероятно, окажутся одними из первых. Если каждый год какой-то параметр P возрастает в α раз, то через t лет величина P_0 возрастет до $P = P_0 \alpha^t$, откуда

$$t = \frac{\lg(P/P_0)}{\lg \alpha} \text{ лет.} \quad (1.8)$$

Приведенную оценку 125 лет мы получили, пользуясь именно этим соотношением. Если величину $\alpha = 1,04$ поддерживать и на более долгие сроки, то через 240 лет будет превзойдено количество солнечной энергии, падающей на Землю, через 800 лет энергетика превысит всю энергию, излучаемую Солнцем, а через 1500 лет даже превзойдет излучение Галактики!

Отметим, что население растет так же экспоненциально, или даже быстрее, так что потребность в экспоненциальном развитии других параметров будет все время поддерживаться. Количество используемой массы в настоящее время $\sim 10^{17}$ г. При том же ежегодном приросте за 2000 лет оно составит 10^{51} г, что равно массе более десяти миллионов галактик! Количество информации, если экстраполировать современный прирост, составляющий 10% в год, возрастет за 2000 лет в 10^{80} раз и, следовательно, существенно превысит количество атомов во Вселенной (около 10^{80}), т. е. такое количество информации принципиально невозможно запомнить! Очевиден вывод: современное экспоненциальное развитие — явление временное и непременно должно быть ограничено естественными причинами.

Действительно, полагая среднюю плотность некоторой космической осваиваемой среды ρ , легко видеть, что даже двигаясь со скоростью света, можно через время t после начала деятельности ежесекундно осваивать количество массы, не большее

$$\frac{dM}{dt} \leq 4\pi (ct)^2 \rho c, \quad (1.9)$$

и количество энергии, не большее

$$\frac{dE}{dt} \leq 4\pi (ct)^2 \rho c^3. \quad (1.10)$$

Отсюда следует, что потребление массы и энергии (а следовательно, и рост информации, материальным носителем которой являются масса и энергия) может возрастать экспоненциально только на ограниченном интервале времени, а неограниченно долгий рост возможен лишь по закону, не более быстрому, чем t^3 . Для нашей цивилизации, как мы видели на примерах, приведенных выше, время предстоящего экспоненциального развития можно оценить интервалом порядка 1000 лет. Что будет потом? Поскольку развитие энергетики на Земле (и вообще в любом конечном объеме) ограничено из термодинамических соображений (эффект перегрева, о котором говорилось выше), то через 100—200 лет основой для развития экономики должно стать освоение космоса! Именно в этом мы видим объективную тенденцию и основную роль развития космических исследований в настоящее время.

Не явится ли неэкспоненциальное развитие свидетельством угасания цивилизации? На наш взгляд, даже линейный прирост информации не означает ее гибели. В самом деле, постоянная скорость накопления информации означает, что ежегодно будет поступать постоянное количество новых, очень важных и очень интересных сведений. Это несколько не препятствует действию основных закономерностей развития цивилизации. Для дальнейшего развития нашей цивилизации определяющим фактором, вероятно, будет так называемый «эффект обратной связи». Все

зависит от того, существуют суперцивилизации или нет. В случае положительного ответа процесс приема и усвоения информации может играть основную роль в дальнейшем развитии. Этот этап обучения может привести к быстрому, скачкообразному переходу цивилизации на высший уровень. Если предположить, что любая цивилизация на определенном этапе своего развития проходит через стадию подобного обучения, то мы придем к выводу, что цивилизаций, которые находились бы в промежуточном состоянии или состоянии развития, близком к нашему, практически нет.

Другая возможность — отсутствие суперцивилизаций, по-видимому, заставит пересмотреть все наши современные представления о перспективах неограниченного развития.

§ 5. Поиск проявлений деятельности сверхцивилизаций

Из общих соображений, приведенных в предыдущих параграфах, можно сделать некоторые выводы о тех видах деятельности сверхцивилизаций, для поиска которых можно было бы составить уже в настоящее время определенную программу.

По-видимому, наиболее общими параметрами такой деятельности должны быть источники энергии сверхвысокой мощности, использование гигантских масс твердого вещества и передача большого количества разного вида информации через космос. В этом параграфе мы рассмотрим первые два параметра, служащие непосредственно обеспечению деятельности цивилизации.

Источники энергии. Как уже отмечалось выше, вряд ли возможно, основываясь на современных астрофизических наблюдениях, указать какой-либо верхний предел энергетики сверхцивилизаций. Но это станет вполне реальным после проведения обзоров, охватывающих всю область электромагнитного диапазона от 10^6 до 10^{18} гц. К такому заключению можно прийти, исходя из термодинамических соображе-

ний: вся энергия, расходуемая сверхцивилизацией, неизбежно должна превращаться в тепло. Эта тепловая энергия не может (во избежание перегрева) накапливаться внутри замкнутого объема. Единственным способом отвода тепла является излучение в космос. Таким образом работа энергетических установок неизбежно должна сопровождаться излучением всей их мощности в космос. Если к. п. д. таких установок очень высок, то спектр и поверхностная яркость этого излучателя должны соответствовать спектру абсолютно черного тела с температурой, равной эффективной температуре всех видов электромагнитных излучений, приходящих из космоса (величина равновесной температуры в межгалактической среде составляет около 3°K). Однако весьма вероятно, что для удобства использования таких установок (а может быть, и благодаря излишку энергии) не всегда соблюдается принцип стопроцентного к. п. д. Тогда спектр излучения будет более сложным. Трудно предвидеть специфические особенности такого рода источников. Единственное, что разумно сделать в настоящее время, — это обратить внимание на природу источников излучения максимальной болометрической мощности. Из числа известных в настоящее время объектов такими являются квазары.

Перечислим вкратце основные закономерности, установленные в результате исследований этих замечательных образований [8].

Радиоизлучение квазаров было обнаружено уже более 10 лет назад, однако всеобщее внимание к этим радиоисточникам было привлечено совсем недавно. В 1960—1962 гг., благодаря существенному повышению точности измерения положения радиоисточников на небе, удалось показать, что некоторые из них совпадают по координатам с оптическими звездобразными объектами. До этого было широко распространено мнение, что большинство радиоисточников соответствует большим галактикам. Казалось, обнаружен новый класс звезд в нашей Галактике с аномально мощным радиоизлучением. Однако дальнейшие

исследования показали, что и эта гипотеза неправильна. По совокупности наблюдательных данных было установлено, что все указанные объекты вообще являются совершенно новым типом внегалактических образований, о котором раньше ничего не было известно. (Заметим, что это еще раз подтверждает соображение о необходимости обзора во всех электромагнитных диапазонах.)

При наблюдении в оптические телескопы квазары представляются звездобразными объектами в том смысле, что их видимый угловой диаметр существенно меньше предела разрешения астрономической оптики (долей секунды дуги). Около некоторых квазаров наблюдаются туманные волокна неправильной формы или в виде выброса, направленного по радиусу от звезды. На рис. 4 приведена фотография одного из самых близких и самых ярких квазаров — радиоисточника 3С 273, вправо вверх от которого заметен выброс. Выброс имеет ширину не более $1''$ — $2''$, начинается на расстоянии $11''$ и кончается на расстоянии $20''$ от звезды. Кроме этого объекта, выбросы и волокна наблюдаются у 3С 48, 3С 196 и 3С 279.

Одним из наиболее замечательных свойств оптических спектров квазаров является сильное смещение спектральных линий в красную сторону спектра. Величина z красного смещения составляет от 0,158 (для ближайшего квазара 3С 273) до изменения длины волны втрое (одним из таких объектов является 3С 9, о котором мы уже упоминали: см. рис. 2). Столь большая величина красного смещения, интерпретируемая как результат расширения Вселенной, говорит о гигантских расстояниях и колоссальных светимостях этих объектов.

Благодаря красному смещению в оптический диапазон попадают спектральные линии, лабораторная длина волны которых соответствует далекой ультрафиолетовой части спектра. Для обычных звезд и галактик эту область невозможно наблюдать с поверхности Земли из-за непрозрачности земной атмосферы для длин волн короче $3000 \text{ \AA}$. Спектры квазаров изучены в настоящее время вплоть до $1000 \text{ \AA}$, и, в

Рис. 4. Квазар 3С 273.

частности, от некоторых получен профиль линии водорода L_a — наиболее интенсивной линии большинства космических объектов. В табл. 1.1 приведен список элементов и стадий ионизации, обнаруженных по спектрам квазаров. В первом столбце помещено название элемента в порядке атомных номеров, во втором — наблюдаемые стадии ионизации. Отсутствие линий ряда элементов, по-видимому, связано с малым их обилием, соответствующим нормальной распространенности, или с отсутствием ярких линий в наблюдаемой части спектра. Логарифм нормальной космической распространенности (по числу атомов) приведен в последнем столбце таблицы.

Таблица 1.1

Элемент	Ионизация	Распространенность	Элемент	Ионизация	Распространенность
H	I	12,0	P	—	5,53
He	II	11,16	S	II	7,22
Li	—	3,0	Cl	—	5,4
Be	—	2,4	Ar	IV	6,62
B	—	2,8	K	—	4,88
C	II, III, IV	8,48	Ca	II	6,22
N	IV, V	7,93	Sc	—	2,91
O	I, II, III	8,83	Ti	III	4,82
F	—	5,4	V	—	3,78
Ne	III, V	8,44	Cr	III	5,38
Na	—	6,22	Mn	II, III	5,10
Mg	II, V	7,46	Fe	II	6,90
Al	II, III	6,28	Co	II	4,72
Si	II, III, IV	7,47	Ni	II	5,93

Видимо, условия возбуждения спектральных линий в квазарах очень разнообразны. Некоторые квазары обладают в основном эмиссионными линиями, большая часть которых отождествляется со спектрами указанных элементов. На рис. 5 приведена микрофотограмма спектра 3C 273 [12]. Кроме линий, в спектре имеются широкие эмиссионные полосы, происхождение которых пока не известно. На рис. 6 показан профиль линии H_β в спектре того же квазара

[13]. На профиле линии хорошо заметны детали, показывающие смещение по отношению к центру линии. Соответствующие этому смещению доплеровские скорости составляют несколько тысяч километров в секунду. Некоторые квазары обладают богатым спектром, состоящим также из абсорбционных линий и полос (например, ЗС 191). Некоторые (например, ЗС 682) не показывают вообще каких-либо линий.

Непрерывный оптический спектр квазаров также обладает рядом характерных особенностей. Прежде всего, распределение в нем энергии существенно отличается от распределения энергии в спектрах звезд. Это дает возможность выделять подобные объекты при массовых измерениях цвета звезд с помощью светофильтров. Распределение энергии в оптическом спектре квазаров скорее представляется степенным законом $F_\nu \propto \nu^{-\alpha}$, поэтому вероятным механизмом является излучение или рассеяние излучения релятивистскими электронами. На рис. 7 приведена

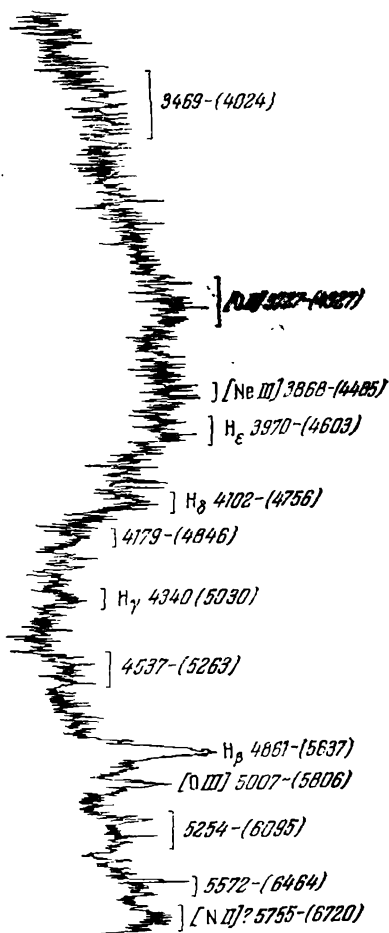


Рис. 5. Микрофотограмма спектра квазара ЗС 273.

диаграмма, построенная по измерениям цветов оптического излучения с светофильтрами U ($\lambda 3600 \text{ \AA}$),

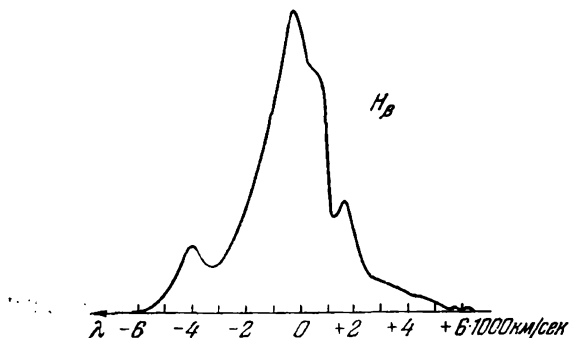


Рис. 6. Профиль линии H_{β} у квазара 3C 273.

B ($\lambda 4400 \text{ \AA}$) и V ($\lambda 5500 \text{ \AA}$). По оси абсцисс отложена разность звездных величин $B - V$, по оси ординат $U - B$ для того же объекта. Нижняя кривая

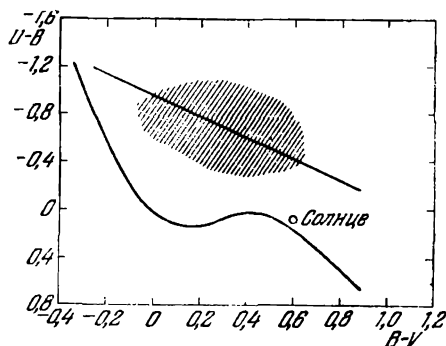


Рис. 7. Цвета звезд главной последовательности и квазаров.

показывает геометрическое место звезд главной последовательности, верхняя линия соответствует степенному распределению энергии. Штриховкой показана область, занимаемая квазарами.

Одним из наиболее удивительных свойств квазаров является переменность их излучения. До открытия квазаров вообще считалось, что внегалактические исследования имеют дело с наиболее стабильными источниками. Яркость галактик постоянна в течение миллиардов лет (за исключением моментов вспышек

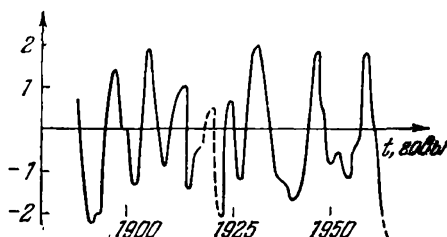


Рис. 8. Кривая блеска квазара 3С 273 после сглаживания и исключения векового убывания яркости.

сверхновых звезд). Наблюдения же первых квази-звездных объектов показали, что их светимость существенным образом меняется. Используя старые фотопластинки неба, удалось восстановить световые вариации некоторых таких объектов за довольно большой срок. На рис. 8 представлена сглаженная кривая колебаний яркости квазара 3С 273 за период с 1888 г. по 1963 г. [14]. Средний период для этого источника составляет около 9 лет. Кроме того, средняя фотографическая величина квазара 3С 273 убывает по закону

$$m_{pq} = 12^m,47 + 3^m,67 (T - 1900),$$

$$\pm 0,08 \quad \pm 0,47$$

(где T — год наблюдений), что дает время экспоненциального уменьшения яркости в e раз за 300 лет [15]. Наблюдаются и более быстрые вариации яркости, статистический характер которых пока еще не изучен. На рис. 9 приведены фотографические и фотоэлектрические измерения звездной величины с фильтром B для квазара 3С 446 [16]. По-видимому, временами у этого объекта происходят изменения

яркости почти вдвое за сутки! Это сразу же позволяет оценить размеры излучающей области — менее одного светового дня ($< 3 \cdot 10^{15}$ см), т. е. много меньше размеров галактик (десятки тысяч световых лет) и даже, вероятно, меньше размеров Солнечной

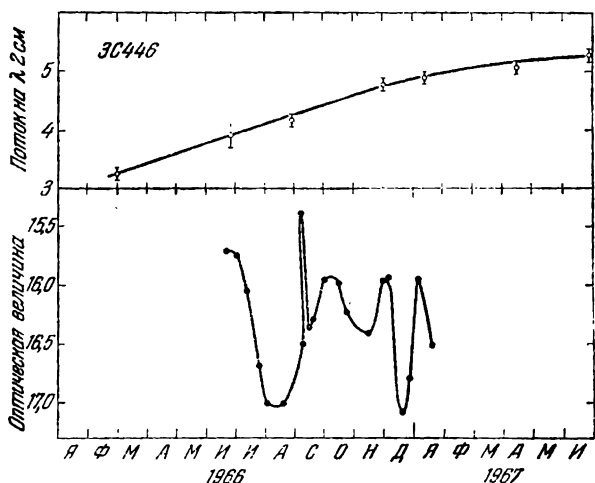


Рис. 9. Изменения блеска квазара 3С 446.

системы: диаметр орбиты Плутона — около $\frac{1}{2}$ светового дня.

По-видимому, существенно меняются и непрерывный и линейчатый спектры квазаров (причем изменяется ширина, интенсивность и длина волны линий [17]). Взаимосвязь всех этих вариаций изучена еще очень плохо.

Некоторые квазары обладают заметной линейной поляризацией оптического излучения. Так, тот же квазар 3С 446 имеет максимальное различие в яркости перпендикулярных компонент поляризации на $0^m,2$ [16]. Вероятно, степень поляризации и позиционный угол меняются со временем. Для квазара 3С 273 поляризация инфракрасного излучения $\lambda = 1,6$ мк достигает 40%.

Остановимся теперь на некоторых основных результатах радионаблюдений квазаров. Угловое разрешение современных радиотелескопов (благодаря использованию интерференционной методики, наблюдениям дифракционной картины при покрытии радионисточников Луной и изучению флуктуаций радиопотока, обусловленных распространением радиоволн

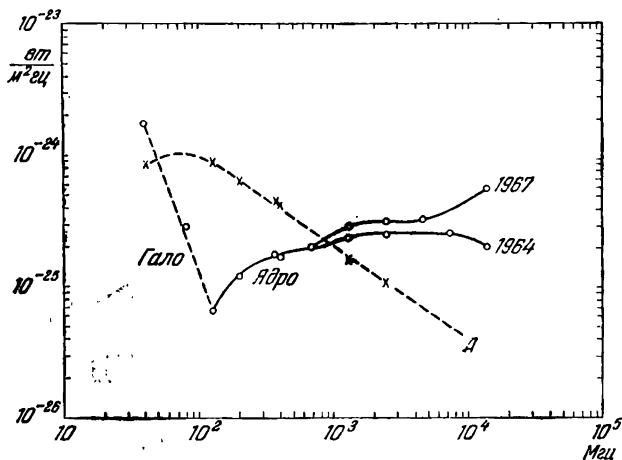


Рис. 10. Спектр компонент А и В квазара 3С 273.

в неоднородной межпланетной среде) доходит до $0'', 001$, что существенно превосходит возможности оптических телескопов при исследовании тех же объектов. Однако пока еще количество информации, полученной новыми радиометодами, очень невелико. Наблюдения квазара 3С 273 (наиболее изученного) показали наличие двух источников: источник А соответствует выбросу на фотографии этого объекта, источник В хорошо совпадает по координатам с самим квазаром. На рис. 10 показаны радиоспектры компонент А и В, которые заметно различаются между собой [18]. Источник А имеет вытянутую форму вдоль выброса и уярчается на внешнем конце, его угловые размеры $5'' \times 1'', 5$. Источник В в свою очередь состоит из сферического гало диаметром $\sim 6''$ и центрального ядра

[19]. Наблюдения с помощью радиоинтерферометра показывают, что большая доля энергии излучается из области $\leq 0''.002$ [20].

Спектры квазаров часто не соответствуют обычному степенному распределению, что, по-видимому, говорит о сложной структуре и разнообразии проис-

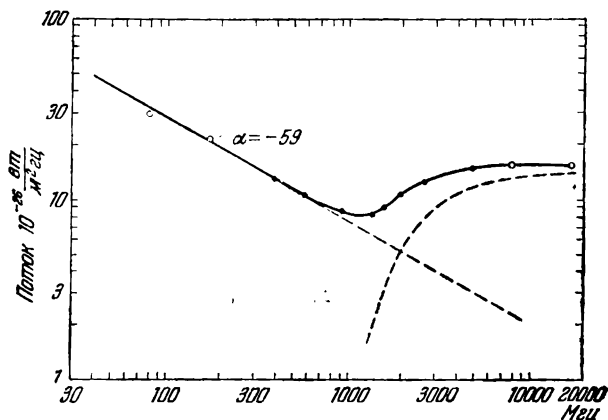


Рис. 11. Спектр квазара 3C 279.

ходящих в них процессов. Наиболее интересными свойствами обладают источники с особенностями в спектре в коротковолновой части радиоастрономического диапазона. На рис. 11 приведен спектр 3C 279; он, так же как и спектр 3C 273 В, имеет тенденцию к увеличению потока в сторону более коротких длин волн. Радиоизлучение такого типа объектов, как правило, является переменным. На рис. 12 и 13 представлено изменение радиопотока от этих двух источников со временем для различных длин волн [21]. Особенно сильные и быстрые вариации наблюдаются в миллиметровом диапазоне. В 1966 г. было принято решение о проведении международной программы систематических наблюдений избранных объектов во всем электромагнитном диапазоне для изучения переменности излучения квазаров. В качестве таких объектов выбраны

источники 3C 273, 3C 279, 3C 345, СТА-102 и другие. Кроме того, в этот же список включен источник 3C 84, являющийся ядром аномальной галактики NGC 1275. Свойства этого источника во многом совпадают со

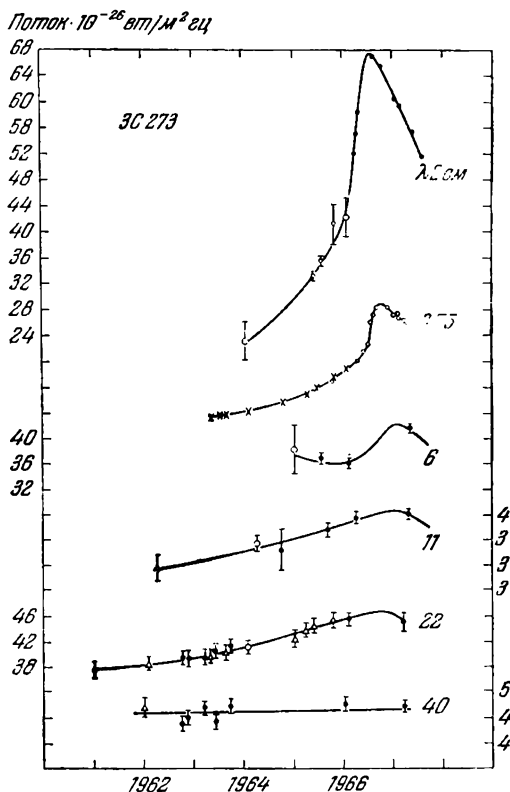


Рис. 12. Изменения радиопотока, степени поляризации и позиционного угла для квазара 3C 273 на волне 8000 МГц.

свойствами квазаров. Тщательные наблюдения покажут, сколь глубока эта аналогия.

Радиолинии от квазаров наблюдать пока еще не пытались, так как для каждого квазара требуется

разработка специальной приемной аппаратуры, учитывающей его красное смещение.

Как видно из краткого перечисления наблюдательных данных, наша информация о квазарах

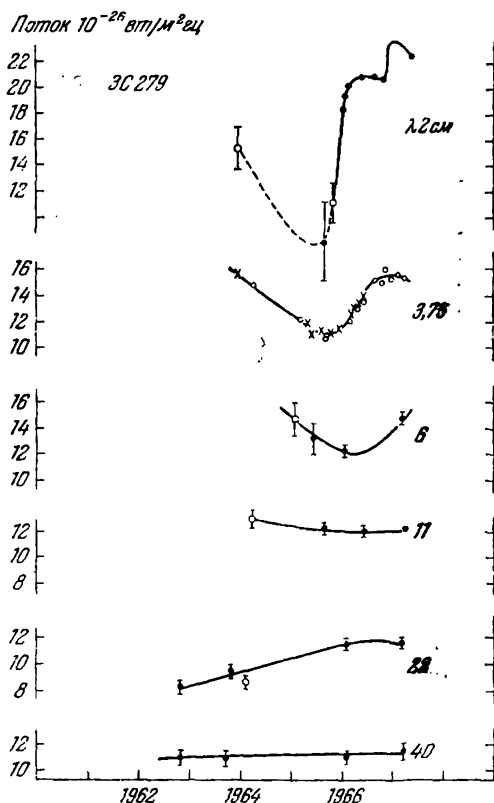


Рис. 13. Изменение радиопотока от квазара ЗС 279 на волне 8000 МГц.

крайне недостаточна даже в хорошо освоенных диапазонах и для наиболее ярких источников. Вполне возможно, что некоторые из них обладают весьма сильным излучением в промежуточной области спек-

тра (между радио- и оптической областями). Так, например, для квазара 3C 273 основная энергия заведомо излучается именно в этом диапазоне. На рис. 14 представлен объединенный спектр 3C 273 В с учетом радио- и оптических наблюдений, а также новых измерений в миллиметровом и инфракрасном диапазонах

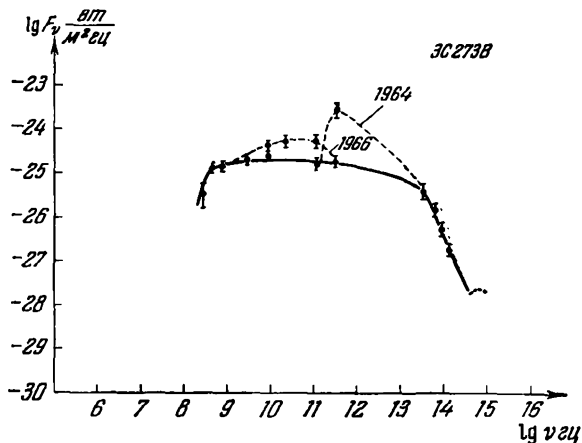


Рис. 14. Спектр компоненты В квазара 3C 273 по радио, инфракрасным и оптическим измерениям.

[23]. Более высокая и коротковолновая кривая относится к 1964 г., а более пологая к 1966 г. Вполне вероятно, что спектр квазаров имеет существенное продолжение в ультрафиолетовой и рентгеновской областях. Так, недавно было, по-видимому, обнаружено излучение 3C 273 в диапазоне $1-10 \text{ \AA}$ [24].

Полная болометрическая светимость квазаров необычно велика. Можно, например, подсчитать, что полный поток, излучаемый в инфракрасном и в субмиллиметровом диапазонах квазаром 3C 273 В, составляет около $4 \cdot 10^{-12} \text{ вт/м}^2$. При расстоянии до источника $R = \frac{cz}{H_0} = 1,5 \cdot 10^{27} \text{ см}$ полная энергия, излучаемая в этом диапазоне, составляет около 10^{47} эрг/сек . В оптическом диапазоне излучаемая энергия в 10 раз,

а в радиодиапазоне — более, чем в 100 раз, меньше этой величины. Таким образом, оказывается, что существуют квазары, являющиеся в основном инфракрасными объектами [25]. По своей болометрической мощности квазар 3C 273 в тысячи раз превосходит

мощность галактик-гигантов.

Исследования пространственного распределения квазаров выявили еще одну очень интересную особенность — они никогда не встречаются в скоплениях или вблизи отдельных галактик [26].

К очень интересным выводам приводят статистические исследования распределения квазаров по величине наблюдаемого потока излучения. Такая статистика отражает распределение источников вдоль луча зрения.

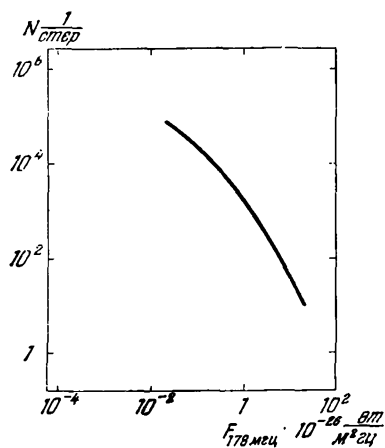


Рис. 15. Зависимость числа источников ярче данного потока от самой величины потока на волне 178 МГц.

Если учесть время распространения излучений, то это распределение может характеризовать количество и мощность квазаров в разные моменты развития Вселенной. На рис. 15 приведена зависимость $N(F_\nu)$, т. е. числа всех источников радиоизлучения ярче данного потока F_ν , от величины этого потока. Кривая построена по результатам наблюдений всех источников радиоизлучения на частоте 178 МГц до предельного потока $5 \cdot 10^{28}$ Вт/м² Гц [26].

Рассматривая случай однородного распределения источников в евклидовом пространстве при отсутствии расширения объема, найдем, что $N(F_\nu) \propto F_\nu^{-3/2}$.

Из теории следует, что красное смещение, обусловленное расширением Вселенной, должно приво-

дить к более пологой зависимости. Наблюдения же показывают, что зависимость более крутая:

$$N(F_\nu) \propto F_\nu^{-1,8}.$$

Недавно было обнаружено, что если все источники разбить на два класса — квазары и радиогалактики, то для каждого класса зависимость $N(F_\nu)$ оказывается различной. Для радиогалактик

$$N(F_\nu) \propto F_\nu^{-3/2}, \text{ а для квазаров } N(F_\nu) \propto F_\nu^{-2,2} [27].$$

Возможной причиной более крутого наклона кривой является быстрая эволюция радионисточников в процессе расширения Вселенной (либо убывание числа источников в каждом рассматриваемом ограниченном объеме с ростом диаметра этого объема либо убывание яркости источников, либо и то и другое). Весьма вероятно, что такой быстрой эволюцией обладают только квазары.

Если разделение источников на два класса справедливо, то среди источников с потоком 10^{-26} вт/м² гц количество квазаров сравнимо с числом других источников (по наблюдениям на частоте 178 Мгц). Кривая $N(F_\nu)$ показывает также заметное насыщение для малых потоков. О предельном количестве источников говорит и невысокая яркостная температура внегалактического фона (около 20° К на той же частоте [26]). Все это позволяет определить момент времени, в который началось образование квазаров. Этот момент соответствует примерно одному миллиарду лет от начала расширения Вселенной.

Объяснение природы квазаров встречается с очень большими трудностями. Открытие радиопеременности поставило под сомнение даже обычные механизмы радиоизлучения. Наличие довольно быстрых изменений радиопотока трудно совместить с предположением о том, что радиоизлучение обусловлено релятивистскими электронами, движущимися в магнитных полях. Другой возможностью является привлечение когерентных механизмов излучения (например, плазменные колебания [28] или когерентное вынужденное излучение тех же релятивистских электронов [29]).

Каково строение квазара по современным представлениям? Основой квазара является ядро, размеры которого $\leq 10^{15}$ см, масса около 10^8 масс Солнца. Ядро играет определяющую роль в поведении квазара. В частности, его излучение дает основной вклад в непрерывный спектр источника. Ядро представляет собой гигантскую звезду, равновесие которой может поддерживаться балансом между гравитационной энергией и энергией магнитной турбулентной плазмы или энергией вращения звезды вокруг осн. Расход энергии на мощное излучение ядра компенсируется за счет постепенного сжатия, т. е. за счет гравитационной энергии. Из теории гравитационного коллапса следует, что при сжатии массы до ее гравитационного радиуса $r_g = \frac{2CM}{c^2}$ может выделить-ся энергия, составляющая десятки процентов от Mc^2 , т. е. существенно бо́льшая, чем термоядерная энергия (около 0,5% от Mc^2). Для $M = 10^8 M_\odot$ $r_g = 3 \cdot 10^{13}$ см — порядка диаметра орбиты Земли. $\frac{1}{3} Mc^2$ составят резерв энергии около $6 \cdot 10^{61}$ эрг, что обеспечивает расход 10^{47} эрг/сек в течение 20 млн. лет. Однако существующие оценки массы квазара крайне неуверенны и скорее всего являются лишь нижней границей (принимается, что масса ядра больше масс окружающих его оболочек, которые определяются по эмиссионным и абсорбционным линиям). Активность ядра может быть обусловлена либо его пульсациями либо тем, что оно представляет тесную двойную систему сверхзвезд большой массы. Эта активность проявляется в выбросе оболочек ионизованного газа и потоков релятивистских частиц. Весьма вероятно, что в центре галактик, в частности, в центре нашей Галактики, имеются объекты подобные квазарам. Так, в центре нашей Галактики имеется область, дающая нетепловое радионизлучение.

Движение облаков ионизованного и нейтрального газа в центральных областях галактик также весьма напоминает квазизвездные объекты. В центре туманности Андромеды (M 31) обнаружено ядро с яркими

эмиссионными линиями. Мы уже упоминали о поразительном сходстве оптического и радиоспектра ядра галактики NGC 1275 и квазаров. Однако все подобные явления в галактиках на много порядков уступают по мощности процессам, происходящим в сверхзвездах.

Необходимо отметить, что среди известных радиоисточников имеется много объектов, природа которых изучена еще меньше, чем природа квазаров.

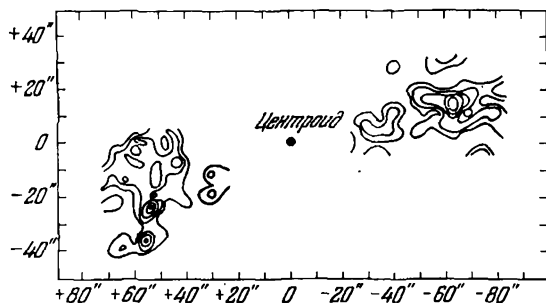


Рис. 16. Структура радиогалактики Лебедь А на волне 11 см.

Например, большую загадку представляют собой некоторые двойные радиогалактики, в том числе один из наиболее ярких радиоисточников Лебедь А. Оптическая туманность, находящаяся между радиоизлучающими компонентами, не является галактикой в обычном смысле слова. Она, по-видимому, целиком состоит из высокотемпературного газа. Недавно было обнаружено, что радиогалактики излучают очень большое количество энергии в рентгеновской области спектра. Радиогалактика Дева А излучает в 100 раз больше в рентгеновской области, чем в радио- и оптическом диапазонах [30].

На рис. 16 приведено современное радиоизображение галактики Лебедь А, полученное по наблюдениям на волне $\lambda = 11$ см на радиointерферометре. Видна очень сложная структура этого образования и наличие нескольких источников малых угловых размеров. На

рис. 17 приведена фотография неба в окрестностях двойного радиоисточника 3С 33 [31]. Компоненты радиоисточника показаны на фото в виде эллипсов, характеризующих размеры и уярчение к периферии.

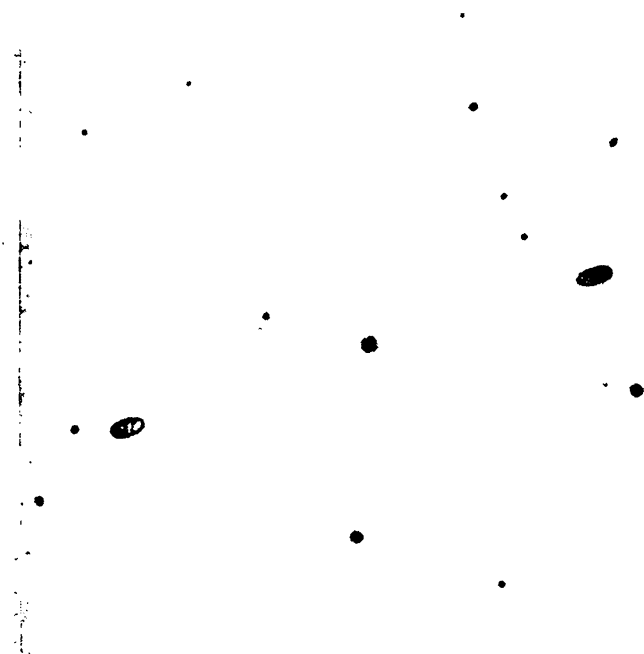


Рис. 17. Область радиоисточника 3С 33.

Посередине между источниками видна галактика, по-видимому, породившая эти объекты. Что «выстреливается» из этой галактики, почему выброшенные радиоисточники обладают удивительно идентичными радиоспектрами, что мешает расширяться этим образованиям в межгалактической среде, если, как многие считают, это облака релятивистского газа?

Наиболее поразительными среди объектов такого типа являются источники 3С 343 и 3С 343.1 [32].

Спектры их также совершенно идентичны, расстояние между компонентами $29''$, а угловые размеры каждого менее $0'',1$. Обнаружить галактику, которая породила эти образования, пока не удалось. Идентичность двух сложных космических объектов, находящихся друг от друга на расстоянии, во много раз превышающем размеры галактик, очень трудно объяснить естественным путем.

Подведем некоторые итоги. Конечно, слишком смело было бы утверждать, что квазары или некоторые радиогалактики являются генераторами энергии искусственного происхождения. Однако нам кажется, что такая гипотеза заслуживает внимания. Во всяком случае эта гипотеза за последние годы стимулировала открытие новых важных свойств квазаров, о чем будет идти речь в следующем параграфе. Пока наблюдениями установлено, что квазары являются наиболее мощными и в то же время наиболее компактными генераторами энергии среди всех известных астрофизических объектов (ядро квазара имеет размеры меньше размеров Солнечной системы и в то же время дает излучение более мощное, чем тысяча галактик!). Проведение обзоров в неосвоенных электромагнитных диапазонах покажет, существуют ли еще более мощные источники. Исследования наиболее мощных объектов, по-видимому, позволят установить верхнюю возможную границу энергетики цивилизаций.

Твердое вещество. С точки зрения современных физических представлений единственным состоянием вещества, способным неограниченно долго хранить большое количество информации, является твердое состояние. Существенная особенность этого состояния — постоянное расположение атомов один по отношению к другому. Именно на этом свойстве основана современная техника — обеспечивается неизменность заданных характеристик конструкций; это же свойство является основой, обеспечившей развитие биологических процессов на Земле.

Можно предполагать, что твердое вещество является также основой для технологии сверхцивилизации и, в частности, основой для систем поиска, обработки и накопления информации. Поэтому обнаружение твердых космических объектов может иметь весьма важное значение для рассматриваемой проблемы.

К сожалению, твердое состояние вещества наиболее трудно обнаружить, так как его собственное излучение весьма мало из-за низкой температуры. Поэтому в настоящее время о количестве и свойствах вещества в твердом состоянии в космосе почти ничего не известно.

Все имеющиеся в нашем распоряжении астрофизические данные относятся к изучению планет и их спутников, исследованиям (в том числе и ракетным) метеоритов и межпланетной пыли в Солнечной системе, а также к эффекту ослабления света звезд межзвездными пылинками и его рассеяния в отражательных туманностях. Оценка свойств межзвездной пыли делается в основном из теоретических соображений о количестве тяжелых элементов и о возможных процессах, приводящих к процессу образования и разрушения пылинок, их нагревания и остывания.

Твердые тела большой массы в космосе необычайно трудно обнаружить. Рассмотрим этот вопрос подробнее. Возможны два подхода к поиску таких объектов: обнаружение самых близких отдельных тел большой массы и обнаружение эффекта суммарного излучения (или поглощения) большого количества тел. Пусть d , δ и T — соответственно размер, плотность и поверхностная температура твердых образований, n — количество образований в единице объема, l — размер части Вселенной, заполненной этими образованиями. Тогда средняя плотность вещества, обусловленная твердыми образованиями, будет

$$\rho = n \delta d^3,$$

угловые размеры ближайшего тела

$$\varphi_{\max} = d n^{-1/3} = \left(\frac{\rho}{\delta} \right)^{1/3}, \quad (1.11)$$

наблюдаемая температура излучения многих тел (рассматриваемая как излучение фона)

$$T_B = Tnd^2l, \quad (1.12)$$

а оптическая толща для поглощения или рассеяния света

$$\tau = nd^2l. \quad (1.13)$$

Полагая, что количество твердого вещества с $\delta \sim 1 \text{ г/см}^3$ не превышает средней плотности Вселенной $\rho \sim 10^{-29} \text{ г/см}^3$ (для внегалактических твердотельных объектов) или плотности Галактики для галактических (и та и другая оценка сильно завышены), найдем, что угловые размеры ближайших тел не превосходят $4 \cdot 10^{-5}$ и $4 \cdot 10^{-3}$ секунд соответственно. Учитывая ограниченность поверхностной температуры этих объектов, мы убеждаемся в невозможности их индивидуального обнаружения.

Условием плохой наблюдаемости собственного суммарного излучения многих тел будет $T_B \leq T_p$, где $T_p \sim 3^\circ \text{ К}$ — равновесная температура всех видов электромагнитных излучений космоса. Условие плохой наблюдаемости эффектов, связанных с поглощением, будет $\tau \ll 1$. Так как поверхностная температура твердых тел, очевидно, должна быть больше (или равна) $T_p \approx 3^\circ \text{ К}$, то оба условия ненаблюдаемости приводят к требованию $nd^2l \ll 1$, а так как $n = \frac{\rho}{\delta d^3}$, то приходим к неравенству

$$d \gg \frac{\rho l}{\delta}; \quad (1.14)$$

полагая, что твердое вещество составляет заметную долю плотности Вселенной $\rho \sim 10^{-29} \text{ г/см}^3$ (эта оценка сильно завышена), плотность твердых образований $\delta \sim 1 \text{ г/см}^3$ и $l \sim \frac{c}{H_0} \sim 10^{28} \text{ см}$, получим, что даже при этих крайних условиях частицы с размерами $d \gg 1 \text{ мм}$ совершенно не обнаружимы.

Нам кажется, что данная трудность является почти непреодолимой в том смысле, что твердотельные

пазоне длин волн 8—13 мк с помощью современных высокочувствительных болометров и оптических телескопов. Здесь P — чувствительность имеющихся болометров, D — диаметр используемого телескопа и R — максимальное расстояние, с которого может быть обнаружено тепловое излучение сферы Дайсона в предположении, что минимальное отношение сигнала к шуму равно 9. Заметим, что такие наблюдения вполне могут быть проведены на имеющихся телескопах с Земли, так как диапазон 8—13 мк является окном прозрачности в земной атмосфере. Вообще же максимум теплового излучения тел, нагретых до температур в интервале от 3 до 300° К, расположен в диапазоне от 10 мк до 1 мм, что соответствует очень сильной непрозрачности атмосферы, обусловленной в основном поглощением водяным паром. Поэтому программа обзоров для поиска таких источников должна ориентироваться на внеатмосферные исследования.

§ 6. Поиск передач информации

В предыдущем параграфе мы говорили о поисках проявлений деятельности цивилизаций. Весьма вероятно также, что важнейшим элементом такой деятельности является передача и обмен информацией. С позиций нашей цивилизации эти передачи можно разбить на два типа: 1) обмен информацией между высокоразвитыми цивилизациями примерно одного уровня и 2) передача информации с целью поднять уровень менее развитых цивилизаций. В случае существования сверхцивилизаций первый вид передач для нас может оказаться в настоящее время принципиально недоступным (например, передачи могут вестись узконаправленными системами и линия передачи может не проходить через Солнечную систему). В то же время передачи второго вида по своему своему назначению должны быть максимально доступны и легко обнаружимы. Как можно ожидать, прием такого вида передач может оказать

принципиальное существенное влияние на развитие нашей цивилизации (эффект «обратной связи» по фон Хорнеру [2]), в результате которого станет возможным быстрый подъем до высшего уровня, существующего в настоящее время во Вселенной. Весьма вероятно

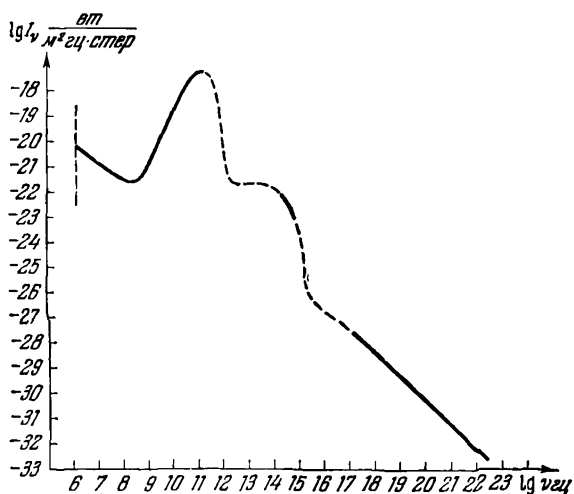


Рис. 18. Спектр фона космического электромагнитного излучения для наблюдателя в межгалактическом пространстве.

(хотя это и кажется фантастичным), что наиболее быстрый и простой путь такого подъема — объединение с ближайшей сверхцивилизацией.

Что можно сказать о возможных направлениях поиска такого рода передач?

Прежде всего, вряд ли можно ожидать из соображений дублирования, что источников таких передач очень много. Скорее всего надо искать один или несколько источников электромагнитного излучения, выделяющихся среди остальных своей интенсивностью и какими-то другими свойствами. Задачу поиска таких источников можно выполнить, проводя обзоры неба в оптимальном с точки зрения шумов диапазоне

длины волн. Как уже отмечалось выше, технические средства приема электромагнитного излучения становятся настолько совершенными, что при рассмотрении подобной задачи шумы аппаратуры можно не считать главным препятствием. По-видимому, в

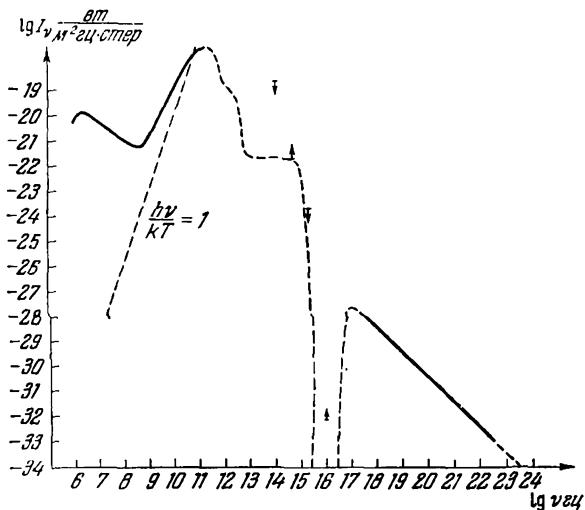


Рис. 19. Спектр фона космического электромагнитного излучения для наблюдения в Солнечной системе.

течение ближайших 5—10 лет будет достигнута предельная чувствительность приемников для астрофизических исследований, определяемая в каждом диапазоне волн величиной излучения фона и флуктуациями самого исследуемого сигнала. На рис. 18 приведен спектр интенсивности фона для наблюдателя, находящегося в межгалактическом пространстве, вдали от ярких галактик. Этот спектр построен по результатам измерений в радио-, оптическом и частично рентгеновском диапазонах. В промежутке между оптическим и радиодиапазонами и в большей части рентгеновского диапазона кривая проведена на

основе теоретических расчетов, учитывающих излучение межзвездной пыли в галактиках, суммарное излучение звезд галактик, а также путем экстраполяции результатов имеющихся наблюдений [35, 36]. Для наблюдателя с Земли к этому излучению добавляется излучение фона нашей Галактики. Суммарная интенсивность фона для земного наблюдателя представлена на рис. 19 для наиболее «яркой» (центр Галактики) и наиболее «холодной» (полюс Галактики) частей неба.

Оба спектра обладают глубокими минимумами интенсивности излучения, которые, по-видимому, и следует считать подходящими для космической связи. Весьма существенным для решения вопроса о выборе наиболее благоприятного диапазона является дискретная (квантовая) природа электромагнитного излучения. Существенная особенность спектров рис. 18 и рис. 19 состоит в том, что интенсивность фона всюду превышает интенсивность абсолютно черного тела при температуре 3°K . Диапазон с наименьшей эквивалентной чернотельной температурой (область, где преобладает излучение «реликтового» фона) занимает интервал длин волн от 3 м до 30 см для рис. 18 (несколько уже для случая рис. 19.) На обоих рисунках прерывистой линией показана граница, где для чернотельного излучения: величина $\frac{h\nu}{kT} = 1$ и со-

ответственно $I_\nu = B_\nu = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}} - 1} = \frac{2h\nu^2}{c^2} \frac{1}{e - 1}$. Спра-

ва от этой границы величина $\frac{h\nu}{kT} > 1$ и поэтому квантовые эффекты играют наибольшую роль.

Решение задачи оптимальной передачи сигнала на фоне шумов существенным образом зависит от того, какие параметры должны быть выбраны оптимальным образом. Учет квантовых и классических флуктуаций приводит к следующему выражению для предельного количества информации, которое можно принять в единицу времени и в единичном интервале

частот [37]:

$$c_v = \ln \left[1 + \frac{P_v}{h\nu} \left(1 - e^{-\frac{h\nu}{kT}} \right) \right] + \\ + \left[\frac{P_v}{h\nu} + \frac{1}{e^{\frac{h\nu}{kT}} - 1} \right] \ln \left[1 + \frac{1}{\frac{P_v}{h\nu} + \frac{1}{e^{\frac{h\nu}{kT}} - 1}} \right] - \frac{\frac{h\nu}{kT}}{e^{\frac{h\nu}{kT}} - 1}. \quad (1.15)$$

Здесь P_v — спектр мощности принимаемого сигнала на входе приемника, а $T = T(\nu)$ — эффективная тем-

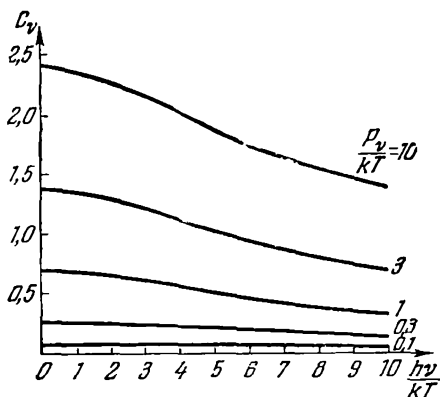


Рис. 20. Зависимость C_v от частоты.

пература всех шумов, определяемая при аппроксимации спектра шумов на входе соотношением

$$\epsilon_v = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}. \text{ При фиксированном } P_v \text{ и } T \text{ функция } c_v$$

монотонно убывает с ростом частоты (рис. 20).

В классическом случае $\left(\frac{h\nu}{kT} \ll 1 \right)$ формула (1.15) переходит в выражение

$$c_v = \ln \left(1 + \frac{P_v}{kT} \right) - \frac{1}{24} \left(\frac{h\nu}{kT} \right)^2 \left(\frac{P_v}{P_v + kT} \right)^2 + \dots \quad (1.16)$$

Если ограничиться разложением до первого члена, то соотношение (1.16) совпадает с известным выражением Шеннона для скорости передачи информации *).

В другом предельном случае $\frac{h\nu}{kT} \gg 1$ и при условии, что сигнал много больше шума $\left(P_v \gg \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}\right)$, получим

$$c_v = \ln \left(1 + \frac{P_v}{h\nu}\right) + \frac{P_v}{h\nu} \ln \left(1 + \frac{h\nu}{P_v}\right), \quad (1.17)$$

т. е. скорость передачи информации зависит только от количества квантов сигнала, приходящих за единицу времени $\frac{P_v}{h\nu}$. При больших спектральных интенсивностях сигнала $\left(\frac{P_v}{h\nu} \gg 1\right)$ формула (1.17) переходит в

$$c_v = \ln \left(1 + \frac{P_v}{h\nu}\right) + 1, \quad (1.18)$$

что также близко к выражению Шеннона, если ввести понятие «эквивалентной температуры» квантового шума $kT_{\text{экв}} = h\nu$.

В случае идеального приемника величина шумов ϵ_v определяется интенсивностью излучения фона неб I_v , т. е.

$$\epsilon_v = \frac{1}{2} I_v \Omega_v A_v = \frac{c^2 I_v}{2\nu^2}. \quad (1.19)$$

Здесь A_v и Ω_v — эффективная собирающая площадь и эффективный телесный угол приемника на частоте ν (антенны в случае радиодиапазона), а коэффициент $\frac{1}{2}$ учитывает одну компоненту поляризации интенсивности I_v , причем интенсивность от обеих компонент считается одинаковой. Последняя часть равенства (1.19) справедлива только, если $A_v \Omega_v = \lambda^2$, что не всегда выполняется. Например, при работе с оптическими телескопами минимальная величина Ω_v

*) См. также гл. III.

определяется обычно не дифракционным изображением точечного источника, а эффектом рассеяния на атмосферных неоднородностях. Благодаря этому угловые размеры источника редко бывают меньше $1''$ и, следовательно, диафрагма, используемая для уменьшения излучения фона, не может быть меньше этой величины. Таким образом, всегда величина $A_v \Omega_v \geq \lambda^2$ и равенство (соответствующее чисто дифракционному изображению) обеспечивает наилучшее отношение «сигнал/шум».

Спектр мощности сигнала на входе приемника связан с потоком излучения точечного источника F_v соотношением

$$P_v = \frac{1}{2} A_v F_v, \quad (1.20)$$

где фактор $\frac{1}{2}$ введен по той же причине, что и в (1.19).

Приведенные здесь общие соотношения позволяют нам оценить особенности сигналов искусственного происхождения.

Прием передач внеземных цивилизаций можно условно разбить на три этапа: 1) поиск и расшифровка позывных; 2) поиск и расшифровка ключа для приема информации; 3) прием и расшифровка информации.

Рассмотрим более подробно условия поиска позывных и выбор наиболее подходящего диапазона.

Под позывными понимаются сигналы, предназначенные для наиболее легкого обнаружения источника и обладающие некоторым минимальным количеством информации, достаточным для того, чтобы без сомнения считать источник позывных искусственным.

Таким образом, задача о нахождении оптимального диапазона позывных сводится к следующему. Определить частоту ν и режим работы передатчика, для которых может быть обеспечено максимальное отношение «сигнал/шум», если заданы F — полный поток энергии источника на единицу поверхности у Земли и время поиска источника t_0 .

Если мы располагаем идеальным приемником, то тогда среднее квадратичное значение мощности шумов, приведенное ко входу, будет определяться флуктуациями, обусловленными естественным фоном космического излучения. С учетом флуктуаций количества фотонов имеем

$$\sqrt{\Delta P_v^2} = [\epsilon_v^2 + \epsilon_v h\nu]^{1/2} \left(\frac{\Delta\nu}{\tau} \right)^{1/2}, \quad (1.21)$$

где ϵ_v определена формулой (1.19), а $\Delta\nu$ и τ — полоса и постоянная времени приемника соответственно. Учитывая, что мощность сигнала на входе $P_s = P_v \Delta\nu = \frac{1}{2} F A_v$, найдем отношение «сигнал/шум»

$$N = \frac{\frac{1}{2} F A_v \sqrt{\tau}}{[\epsilon_v^2 + \epsilon_v h\nu]^{1/2} \sqrt{\Delta\nu}}. \quad (1.22)$$

Время всего поиска t_0 определяется как поиском по частоте, так и поиском по направлению, так что требуется найти максимум величины N при заданной величине t_0 :

$$t_0 = \tau \frac{\nu}{\Delta\nu} \frac{4\pi}{\Omega_v}. \quad (1.23)$$

Как уже отмечалось выше, реальные условия приема приводят к тому, что произведение телесного угла Ω_v , из которого принимается излучение, на эффективную площадь A_v должно удовлетворять соотношению

$$\Omega_v A_v = k_v \lambda^2, \quad (1.24)$$

где численный коэффициент $k_v \geq 1$. Величина k_v из-за трудности высокоточного изготовления больших поверхностей увеличивается с ростом частоты. Учет условий распространения света и радиоволн в атмосфере Земли также приводит к росту коэффициента k_v , который особенно сильно сказывается при наблюдениях в оптическом диапазоне. При рассмотрении возможностей внеатмосферных наблюдений необхо-

димо учитывать увеличение угловых размеров за счет распространения радиоволн в межзвездной и межгалактической плазме. Этот вопрос разобран в гл. II и там показывается, что величина рассеяния весьма мала в диапазоне сантиметровых и дециметровых радиоволн, но может стать существенной в метровом диапазоне.

Выражение (1.22) с учетом (1.19), (1.23) и (1.24) приводит к формуле

$$N = \frac{F \sqrt{A_{\nu} \nu t_0}}{\sqrt{4\pi} c \sqrt{I_{\nu}^2 k_{\nu} + I_{\nu} \frac{2h\nu^3}{c^2}}}. \quad (1.25)$$

Из этого соотношения можно сделать следующие выводы. Во-первых, для получения максимального N необходимо предельно уменьшать k_{ν} , если определяющими являются классические флуктуации ($I_{\nu} k_{\nu} \gg \frac{2h\nu^3}{c^2}$). Это условие всегда накладывается в области, где спектр фона описывается эффективной температурой, удовлетворяющей неравенству $\frac{h\nu}{kT} \ll 1$.

Однако и в квантовой области ($\frac{h\nu}{kT} \gg 1$) в случае больших k_{ν} основную роль может играть первое слагаемое под корнем в знаменателе (1.25). Во-вторых, при условии $I_{\nu} k_{\nu} \ll \frac{2h\nu^3}{c^2}$, которое может удовлетворяться только в коротковолновой части (в области $\frac{h\nu}{kT} \gg 1$) величина N не зависит от k_{ν} . И, наконец, в-третьих, требование максимума N по (1.25) не накладывает никаких условий на ширину полосы $\Delta\nu$ и постоянную времени τ .

Из (1.25) для области, где интенсивность фона определяется выражением Рэля—Джинса ($\frac{h\nu}{kT} \ll 1$), следует:

$$N_c = \frac{F \sqrt{A_{\nu} \nu t_0}}{\sqrt{4\pi} c I_{\nu} \sqrt{k_{\nu}}}, \quad (1.26)$$

а в квантовой области ($\frac{h\nu}{kT} \gg 1$ и $I_\nu k_\nu < \frac{2h\nu^3}{c^2}$) имеем:

$$N_q = \frac{F \sqrt{A_\nu \nu t_0}}{\sqrt{4\pi} c \sqrt{I_\nu \frac{2h\nu^3}{c^2}}}. \quad (1.27)$$

Еще раз вернемся к рис. 18 и 19, где штриховой линией показана граница этих областей для случая $k_\nu \sim 1$. В общем случае в (1.25) первое и второе слагаемые в знаменателе будут равны при $I_\nu = \frac{2h\nu^3}{c^2 k_\nu}$. Если $k_\nu \gg 1$, то граница между квантовой и классической областями может существенно переместиться в коротковолновую сторону.

На основе количественных оценок по формулам (1.25)–(1.27) можно прийти к определенному выводу об оптимальном диапазоне: диапазоном, обеспечивающим наибольшее отношение «сигнал/шум» при обзоре неба в течение заданного времени t_0 является дециметровый диапазон — область минимума интенсивности фона в радиодиапазоне ($\lambda \sim 10\text{--}50\text{ см}$)*).

Кроме простоты обнаружения, позывные должны содержать в себе по крайней мере минимальное количество информации, необходимое для доказательства искусственности сигналов. Важный вопрос о принципиальном различии сигналов искусственного и естественного происхождения является в настоящее время еще не решенной задачей. Однако эта разница определяется не столько формой сигналов, сколько самим их содержанием. Содержание передач и, в частности, позывных, должно обладать некоторой инфор-

*) Последние исследования нового типа объектов — «пульсаров» указывают, что существует еще один вид помех при передачах на сверхдальние расстояния. Эта помеха, обусловленная флуктуациями коэффициента преломления межзвездной плазмы, приводит к пропаданию сигнала на большие интервалы времени. Это явление пока еще недостаточно исследовано. В отличие от фонового излучения, оно является мультипликативной помехой и ее учет, вероятно, переместит оптимальный диапазон в более коротковолновую область.

мацией, которая в принципе отсутствует в излучении, порождаемом естественными процессами.

С другой стороны, возникает вопрос, сможем ли мы понять сообщение цивилизации, условия и длительность развития которой существенным образом отличаются от наших? Видимо, понимание возможно, когда имеются условия для создания единого языка. Единство строения Вселенной и законов природы в разных ее областях и в разные моменты времени, вытекающее из наблюдательных данных, как нам кажется, и является такого рода языком.

На основе последних выводов можно сделать некоторые определенные заключения о характере позывных. Прежде всего из эмпирических соображений вытекает, что количество информации, необходимое для сообщения об искусственности сигнала, должно содержать более 10 и менее 100 двоичных единиц:

$$10 < J < 10^2. \quad (1.28)$$

Если исходить из универсальности законов природы, то весьма вероятно, что в качестве содержания позывных при столь малом количестве информации наиболее эффективно передавать некоторую комбинацию цифр. Например, для передачи шестизначным двоичным кодом первых восьми простых чисел, их суммы и интервала между передачами достаточно 60 бит: 000001 000010 000011 000101 000111 001011 001101 010001 111011 000000 ... что означает 1, 2, 3, 5, 7, 11, 13, 17, 59, 0, ... Периодически повторяющаяся передача такого вида, вероятно, не оставит никакого сомнения в искусственности сигнала.

Необходимо обратить внимание на разнообразие возможных видов позывных. При измерениях электромагнитного излучения регистрируются следующие параметры: две пространственные координаты источника, момент наблюдения, частота, интенсивность, степень и позиционный угол линейной поляризации, степень и направление вращения круговой поляризации. В принципе, изменение любого из этих параметров в зависимости от изменения любого другого мо-

жет являться источником информации. Видимо, все эти возможности целесообразно разбить на две группы: временные позывные и стационарные позывные. Временные позывные связаны с изменением какого-либо из перечисленных параметров со временем (например, для передачи двоичного кода может быть использовано изменение знака круговой поляризации излучения). Стационарные позывные проявляются в закономерном изменении одного измеряемого параметра от другого независимо от времени. Например, закон изменения того же знака круговой поляризации от частоты может содержать минимально необходимое количество информации (1.28).

В настоящее время неясно, какой из перечисленных способов передачи является наиболее эффективным. Поэтому пока нет каких-либо более определенных критериев, необходимо изучать все параметры излучения подозреваемых источников.

Рассмотрим еще одну возможность поиска позывных. Наиболее вероятно, что на специальную передачу позывных расходуется лишь небольшая доля мощности передатчика. Можно ли использовать некоторые общие свойства всей передаваемой информации в качестве позывных? Если передача ведется в очень широкой полосе, то за счет фактора осреднения чувствительность измерения параметров излучения может быть на много порядков выше, чем для приема без осреднения. Так, при работе в радиометрическом режиме при измерении средней мощности источника отношение «сигнал/шум» увеличивается в $n = \sqrt{\Delta\nu t}$ раз по сравнению с приемом без накопления. Поэтому для поиска источника, передающего информацию в полосе $\Delta\nu$, потребуется антенна с эффективной площадью в n раз меньшей, чем для приема информации от того же источника. Для примера, если передача информации ведется в полосе 10^{11} гц и время осреднения 10 сек, то $n = 10^6$.

Рассмотрим теперь вопрос об оптимальной передаче информации. Как распределить энергию передатчика по спектру, чтобы получить максимальную

скорость передачи информации? Заданными считаются спектр интенсивности космических шумов I_ν и полный поток энергии от передатчика на единицу площади у Земли F . Решение такого типа задач [37] проводится методом варьирования соотношения (1.15) при указанных условиях и приводит к следующему результату. Оптимальный спектр источника есть

$$F_\nu = \frac{\frac{2h\nu}{A_\nu}}{e^{\frac{2h\nu}{A_\nu}} - 1} - I_\nu \Omega_\nu, \quad (1.29)$$

где величина a определяется из условия $\int_L F_\nu d\nu = F$.

Меньшим величинам a соответствуют большие величины F . Учитывая, что из соотношения (1.24) следует $\Omega_\nu A_\nu \geq \lambda^2$, находим

$$F_\nu \leq \frac{2h\nu}{A_\nu} \left[\frac{1}{a \frac{2h\nu}{A_\nu} - 1} - \frac{1}{e^{\frac{h\nu}{kT}} - 1} \right], \quad (1.30)$$

откуда, принимая во внимание, что $F_\nu \geq 0$, получим

$$\frac{A_\nu}{2ak} > T, \quad (1.31)$$

где T , как и раньше, эффективная температура фона на частоте ν . Поэтому приходим к следующему выводу: оптимальный диапазон передач соответствует той области спектра, где эффективная температура фона минимальна*).

В случае фона космического электромагнитного излучения эта область соответствует диапазону, где преобладает так называемое «реликтовое излучение», т. е. излучение, описываемое формулой Планка с температурой $T \sim 3^\circ$. Этот диапазон охватывает часть спектра от субмиллиметрового до дециметрового с максимумом интенсивности фона у $\lambda \sim 1,7$ мм (см. рис. 18 и 19).

*) Возможно перемещение этого диапазона в сторону более коротких волн, что обусловлено причинами, которые отмечены в списке на стр. 84.

Более определенный вид спектра источника можно получить, задаваясь зависимостью A_ν и Ω_ν от частоты. Рассмотрим два возможных случая.

1. $A_\nu = A\nu^{-2}$, $\Omega_\nu = \Omega = \text{const}$. Как уже отмечалось в § 3, этот случай соответствует ограничениям, накладываемым на предельно большие антенны, при условии, что относительная точность изготовления поверхности для разных длин волн примерно одинакова. Вид спектра F_ν зависит от величины F (рис. 21). При

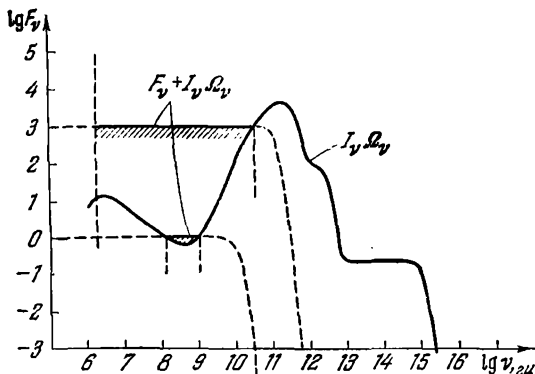


Рис. 21. Распределение энергии в спектре искусственного источника при $A_\nu \propto \nu^{-2}$, $\Omega_\nu = \text{const}$.

малых F (т. е. при малых мощностях передатчика) область максимума F_ν соответствует минимуму интенсивности фона в дециметровом диапазоне радиоволн, т. е. совпадает с диапазоном передачи позывных. При больших F спектр передатчика расширяется и, если излучением фона можно пренебречь, то

$$F_\nu = \frac{\frac{2h\nu^3}{A}}{e^{\frac{2h\nu^3}{A}} - 1}. \quad (1.32)$$

Этот спектр имеет плоскую область в низкочастотной части и экспоненциально падающую (без максимума) в области высоких частот. Завал обусловлен потерей информации за счет квантовых флуктуаций сигнала.

2. $A_v = A = \text{const}$, $\Omega_v = \frac{c^2 \nu^{-2}}{A}$. Этот случай соответствует условиям работы на одной антенне, обеспечивающей прием всего спектра сигнала:

$$F_v = \frac{2h\nu}{A} \left(\frac{1}{e^{\frac{2h\nu a}{A}} - 1} - \frac{1}{e^{\frac{h\nu}{kT}} - 1} \right). \quad (1.33)$$

При малых F удобно обозначить $\frac{A}{2h\nu a} = \frac{k(T + \Delta T)}{h\nu}$, $\Delta T \ll T$. Тогда

$$F_v = \frac{2k\Delta T}{A} \left(\frac{h\nu}{kT} \right)^2 \frac{e^{\frac{h\nu}{kT}}}{\left(e^{\frac{h\nu}{kT}} - 1 \right)^2}. \quad (1.34)$$

В области $\frac{h\nu}{kT} \ll 1$ спектр плоский, $F_v = \frac{2k\Delta T}{A}$; далее поток возрастает, достигая максимума при $\frac{h\nu}{kT} \sim 1,1$ (для $T = 3^\circ \text{K}$ этому соответствует $\lambda = 4,8 \text{ мм}$), а затем экспоненциально падает. F_v в максимуме в 2,7 больше, чем на плоском участке.

При увеличении F ограничения, накладываемые фоном, играют все меньшую роль, ширина спектра увеличивается. При очень больших мощностях это распределение плоское с низкочастотной стороны и экспоненциально (без максимума) убывает в области высоких частот:

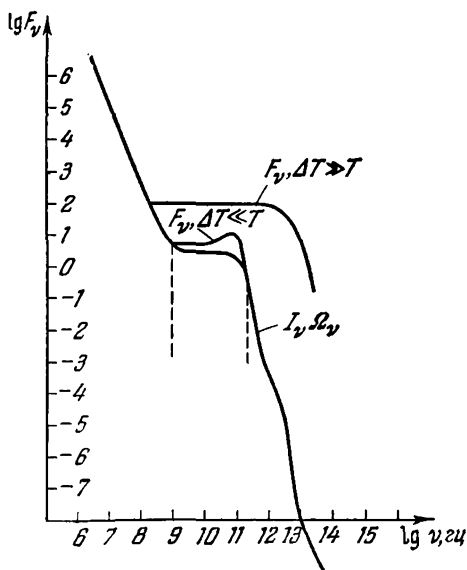
$$F_v = \frac{2h\nu}{A} \frac{1}{e^{\frac{2h\nu a}{A}} - 1}. \quad (1.35)$$

Ожидаемый вид спектра (качественная картина) для различных F приведен на рис. 22.

Остановимся теперь на общих свойствах передачи информации с целью использования их в качестве критериев для предварительного отбора источников радионизлучения и поиска позывных.

1. Во всех случаях существенная часть спектра искусственного источника находится в радиодиапазоне, обладая максимумом в районе минимума

излучения космического фона (коротковолновая часть дециметрового диапазона) или в миллиметровом диапазоне. Вид спектра с максимумом в дециметровом диапазоне, или уплощенный в этом диапазоне с максимумом в миллиметровом диапазоне волн может являться предварительным критерием при отборе подозреваемых объектов.



(Рис. 22. Распределение энергии в спектре искусственного источника при $A_\nu = \text{const}$, $\Omega_\nu \sim \nu^{-2}$.)

2. Предельно малые угловые размеры подозреваемого источника (радионизлучения) также могут являться весьма сильным предварительным критерием.

3. Исследования других астрофизических параметров источника во всех диапазонах также могут быть использованы для предварительной селекции (круговая поляризация, оптические и радиолинии, оптическое отождествление, рентгеновское излучение и т. д.).

В этом отношении поиск источников искусственного происхождения практически совпадает с основ-

ным направлением развития современной наблюдательной радиоастрономии. Вероятно, именно поэтому обсуждение возможных предварительных критериев искусственности источников оказало большое влияние на радиоастрономические исследования. Так, при обсуждении программы поиска внеземных цивилизаций в 1964 г. [38] предполагалось, что такого рода источники

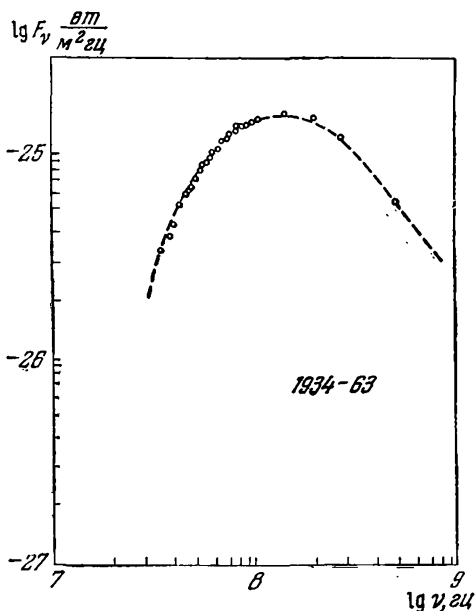


Рис. 23. Спектр источника радиоизлучения 1934—63.

должны обладать спектром с максимумом в дециметровом — сантиметровом диапазоне, предельно малыми угловыми размерами и переменностью во времени. В качестве вероятного источника такого типа предлагалось исследовать объект СТА-102. Замечательно, что, по-видимому, все эти свойства были обнаружены в последующие годы для целого ряда источников, в том числе и для СТА-102. Был обнару-

жен ряд источников, излучение которых в еще большей степени сконцентрировано в дециметровом диапазоне. Одним из наиболее интересных объектов такого вида является источник 1934-63 (координаты $\alpha = 19^{\text{h}}34^{\text{m}}48^{\text{s}}.9$, $\delta = -63^{\circ}49'42''$ (1950)) [39]. На рис. 23 приведен



Рис. 24. Область радионисточника 1934-63.

спектр этого объекта, имеющий максимум у $\lambda = 21$ см. На рис. 24 приведена фотография области неба, где находится источник. На месте источника заметна галактика с ярким звездоподобным ядром, которая соединяется едва заметной перемычкой с другим звездоподобным объектом.

Замечательным примером источника с плоским спектром и возможным максимумом в миллиметровом или субмиллиметровом диапазоне является уже упоминавшийся источник 3C 273 В (см. рис. 14). Этот и ряд других источников обладают крайне малыми угловыми размерами ($< 0''.002$), радиоизлучение их показывает сильную переменность. Как уже отмеча-

лось, эти результаты наблюдений противоречат общепринятому для источников радиоизлучения механизму синхротронного излучения. Расчеты показывают невозможность генерации наблюдаемой мощности при столь малых размерах источников с помощью этого механизма. Поэтому в настоящее время для объяснения наблюдаемых фактов приходится привлекать процессы, связанные с коллективным когерентным излучением (плазменные колебания [28], вынужденное излучение [29]). Заметим, что излучение искусственного передатчика также является типичным примером когерентного излучения.

Весьма важным для рассматриваемого круга вопросов представляется также обнаружение аномально сильного излучения некоторых туманностей в линиях на волне около 18 см [40]. Частоты линий 1612, 1665, 1667 и 1720 МГц соответствуют расщеплению (Λ -удвоению и сверхтонкой структуре) самого нижнего энергетического уровня молекулы гидроксила OH. Наблюдения показали наличие необычайно мощного излучения этих линий (особенно 1665 и 1667 МГц) из областей очень малых угловых размеров, находящихся внутри облаков ионизованного газа. На рис. 25 приведена фотография одной из таких туманностей (NGC 6334); квадратики указаны области аномально сильного монохроматического излучения [43]. Удивительным представляется вообще существование молекулы OH в областях H II, где температура $\sim 10\,000^\circ$. Это излучение, как оказалось, обладает совершенно необычными свойствами. Угловые размеры излучающих областей меньше $0''.002$ (линейные меньше четырех астрономических единиц). Поэтому можно дать лишь верхнюю оценку эффективной температуры в максимуме профиля линии — более 10^{13} градусов. В то же время необычайно малая ширина профиля (в отдельных случаях меньше 400 гц) показывает, что температура не более 10°K . Такое соотношение между интенсивностью и шириной возможно только при процессах нелинейного излучения, подобных механизму генерации в лабораторных молекулярных генераторах. Дальнейшее изучение

линий межзвездного гидроксила показало почти 100%-ную круговую поляризацию наиболее сильных компонент; в некоторых случаях наблюдается также сильная линейная поляризация излучения. Некоторые из линий

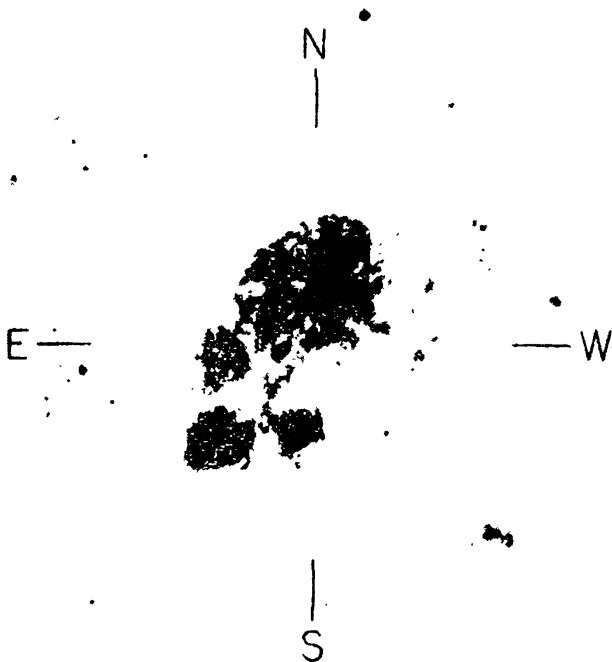


Рис. 25. Туманность NGC 6334 с указанием областей излучения в линиях OH.

показывают заметное изменение интенсивности компонент изо дня в день. На рис. 26 приведен профиль линии $1665\text{ M}\mu$ для туманности W49 для линейной, правой и левой круговой поляризации [41], а на рис. 27 — профиль той же линии для туманности NGC 6334 в различные дни [42].

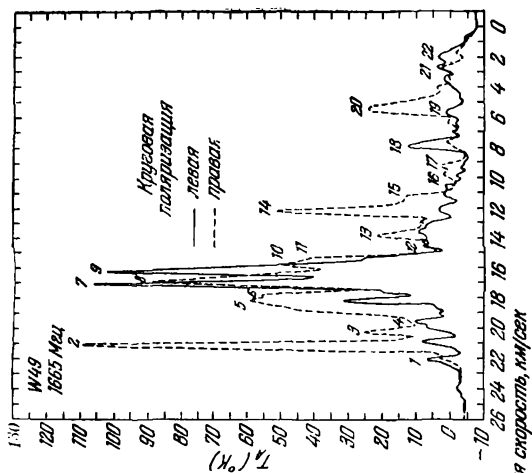
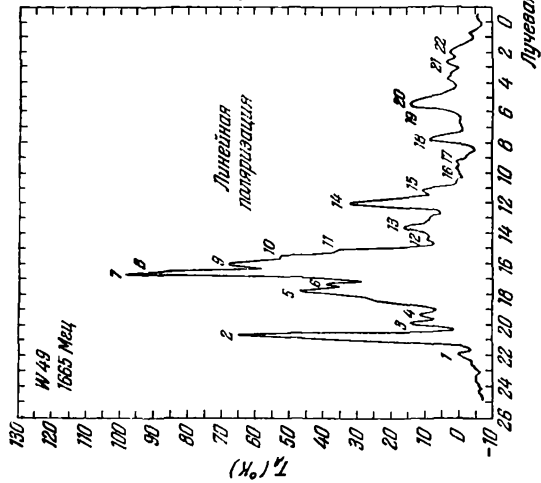
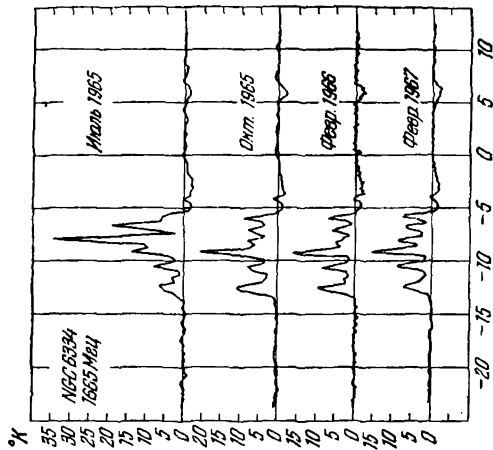
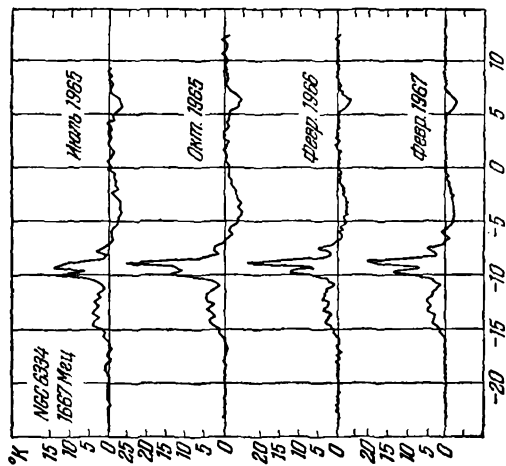


Рис. 26. Профиль линий 1665 Мгц для источника W 49 для линейной, правой и левой круговой поляризации.



Лучевая скорость, км/сек

Рис. 27. Профиль линии 1665 М μ для NGC 6334 по наблюдениям в различные дни.

Теоретические оценки возможности инверсной заселенности энергетических уровней атомов и молекул, соответствующих переходам в радиодиапазоне, показывают, что такая ситуация в естественных условиях межзвездной среды весьма вероятна [43]. Тем не менее полной ясности в механизме этого излучения нет.

Любопытно отметить, что для проблемы поиска позывных сигналов важен уже тот факт, что естественные условия в космосе могут существенным образом упростить задачу создания сверхмощных узкополосных генераторов радионизлучения.

§ 7. О программе поиска сверхцивилизаций

В предыдущих параграфах мы пытались обосновать точку зрения, согласно которой наиболее эффективным в ближайшие годы может оказаться поиск сигналов и проявлений деятельности только сверхцивилизаций. Это связывается с возможностью получения за время порядка 10 лет информации обо всех ярчайших источниках во всех диапазонах электромагнитного излучения. Такая задача совпадает с основным направлением астрофизических исследований. Однако, видимо, уже сейчас можно наметить и некоторую специфическую программу поиска искусственных источников. Есть основания предполагать, что для сверхцивилизаций передача информации является одним из основных условий существования. Исходя из этого, целесообразно на современном этапе связывать специальную программу с поиском позывных. Проведенное рассмотрение показывает, что можно с достаточной определенностью наметить частотный диапазон передачи позывных. Остальные же параметры (полоса, длительность излучения, поляризация и т. д.) определить заранее довольно трудно. Если объем передаваемой информации очень велик, то естественно ожидать очень широкополосного излучения и, следовательно, искусственные источники надо в первую очередь искать среди большого количества естественных радионисточников. Однако вполне вероятно, что вместе с широкополосным сигналом сла-

бой интенсивности, несущим основную информацию, генерируется мощное излучение позывных в очень узкой полосе на фиксированных частотах, или в виде очень коротких, но мощных импульсов.

Перечислим основные направления исследований, представляющие наибольший интерес для задач поиска позывных сигналов:

1. Проведение обзоров неба в диапазонах 3, 10, 30, 100 и 300 *мк* и особенно 1, 3, 10 *мм*, 3 и 10 *см* с целью обнаружения по крайней мере 100 ярчайших источников в каждом диапазоне.

2. Изучение свойств квазаров и других «подозрительных» объектов.

3. Поиск и изучение аномально мощных монохроматических источников радиоизлучения (типа излучения гидроксила) в дециметровом диапазоне.

4. Поиск импульсных сигналов космического происхождения в том же диапазоне *).

5. Поиск монохроматических сигналов перестраивающихся по частоте в том же диапазоне.

Отметим, что по каждому из перечисленных пунктов в настоящее время имеется лишь ничтожно малая информация по сравнению с той, которую можно было бы получить с помощью современных средств приема. Существенную помощь в отборе среди уже обнаруженных «подозрительных» источников могут оказать предварительные критерии, о которых говорилось выше. Отметим весьма важную роль критерия оценки угловых размеров источников. Измерение угловых размеров можно проводить с помощью радиointерферометров. Напомним, что уже в настоящее время созданы в сантиметровом и дециметровом диапазонах интерференционные системы с базой порядка диаметра Земли [44]. На очереди эксперимент с установкой одной из антенн на межпланетной станции и доведение базы интерферометра до размеров орбиты Земли. Другим перспективным направлением является оценка максимальных линейных размеров по времени

*) Поиск таких источников начал с открытием в 1967 г. «пульсаров». Однако среди обнаруженных источников этого типа пока нет никаких указаний на их искусственную природу.

изменения какого-либо параметра источника (например, потока или поляризации). Учитывая конечность скорости света, наблюдение одновременного изменения излучения всего объекта за время t возможно, если только $t > \frac{r}{c}$, где r — радиус объекта. Полагая, например, что квазар 3C 273 В представляет собой объект с массой $M \sim 10^8$ масс Солнца и радиус его больше критического (гравитационного) радиуса $r_g = \frac{2GM}{c^2} = 3 \cdot 10^{13}$ см (G — гравитационная постоянная), мы приходим к выводу, что в этом источнике невозможны изменения яркости за время менее, чем $T \sim \frac{r_g}{c} = 10^3$ сек. Обнаружение более быстрого изменения означало бы, что масса и размеры этого объекта существенно меньше.

Однако селекция источников по предварительным критериям является необходимым, но не достаточным этапом поиска. После того как будет отобрано достаточно большое число «подозрительных» объектов, основную роль начнет играть поиск «смыслового содержания» в информации об этих объектах. Эта работа и параллельное теоретическое рассмотрение различных возможностей поможет уточнить дальнейшую программу исследований. В частности, можно надеяться, что существенную информацию о параметрах квазаров и их изменении со временем в различных диапазонах удастся получить в результате осуществления международной программы их исследования, начатой в 1966 г. [45].

В настоящее время нет теории, на основе которой можно было бы оценивать наличие или отсутствие смыслового содержания принимаемой информации. Такую оценку в настоящее время может делать только человек. Поэтому неизбежно возникает вопрос о субъективности подхода к разработке программы поиска внеземных цивилизаций. Однако, как уже отмечалось выше, имеется уверенность в том, что такой подход не является полностью субъективным. Это связано с наблюдаемой одинаковостью законов природы и их

проявлений в пространстве и времени. Именно это обстоятельство и дает право надеяться на объективность понимания в настоящее время хотя бы части таких законов. Правильно понятые закономерности могут составить основу для взаимопонимания с другими цивилизациями и, в частности, для составления объективной программы поиска. В принципе, вероятно, можно разработать методику и построить анализирующую машину для сопоставления известных законов природы (в простейшем случае — математических соотношений) с содержанием любой принятой информации. Нам кажется, что по крайней мере при расшифровке позывных такая задача является вполне разрешимой.

ЛИТЕРАТУРА

1. Шкловский И. С., Вселенная, жизнь, разум, изд. 2-е, «Наука», 1965.
2. «Межзвездная связь», сб. статей, под ред. А. Камерона, пер. с англ., «Мир», 1965.
3. Shklovskii I. S., Sagan C., Intelligent life in the Universe, Holden Day, 1966.
4. Володин А. Г. Первые шаги эволюции, Лит. газета, 1 февр. 1967 г.
5. Баранов В. И., Астрон. ж. 43, 1074, 1966; Фишер Д. Е., Доклад на 20 конгрессе ЮПАК, Москва, 1965.
6. Герлинг Э. К., Маслеников В. А. и Морозова И. М., там же.
7. «Наблюдательные основы космологии» сб. статей, «Мир», 1965.
8. Бербидж Дж., Бербидж М., Квазары, «Мир», 1969.
9. Ляпунов А. А., Проблемы кибернетики, вып. 10, стр. 179, 1963.
10. Кардашев Н. С., Шоломицкий Г. Б., Астр. цирк. № 336, 1965.
11. Роуз Д., Кларк М., Физика плазмы и управляемые термоядерные реакции, гл. I, «Мир», 1963.
12. Andrillat Y., Andrillat M., Publ. de l'Observatoire de Haute Provence 7, № 11, 1964.
13. Дибай Э. А., Проник В. И., Астр. цирк. № 286, 1964.
14. Озерной Л. М., Чертопруд В. Е., Астрон. ж. 43, 20, 1966.
15. Geyer E., Zs. für Astroph. 60, 112—114, 1964.
16. Kinman T., Lamla E., Wirtanen C., Contr. from Lick Observatory, № 225, 1966.
17. Sandage A., Westphal J., Strittmatter P., Ap. J. 146, 332, 1966.

18. Hoerner S. von, *Ap. J.* **144**, 483, 1966.
19. Adgie R., Gent H., Slee O., Frost A., Palmer H., Rowson B., *Nature* **208**, 275, 1965.
20. Cohen M., Gunderman E., Hardebeck M., Sharol L., *Sky and Telescope* **34**, 143, 1967.
21. Kellerman K., Pauling-Toth I., *Ap. J.* **146**, 1966.
22. Berge G., Seielstad G., *Observations of the Owens Valley radio observatory*, № 9, 1966.
23. Low F., *Ap. J.* **142**, 1287, 1965; *Ap. J.* **150**, 1967.
24. Friedman H., Byram E., VII Cospar symposium, London, 24—28 July 1967.
25. Шкловский И. С., *Астрон. ж.* **42**, 893, 1965.
26. Sandage A., Muller W., *Ap. J.* **144**, 1240, 1966.
27. Veron P., *Ann. d'Astrophys.* **29**, 231, 1966.
28. Гинзбург В. Л., Озерной Л. М., *Ap. J.* **144**, 599, 1966; Зайцев В. В., Каплан С. А., *Астрофизика* **2**, 169, 1966.
29. Железняков В. В., *ЖЭТФ* **51**, 570, 1966; *Астрон. ж.* **44**, 1967; Каплан С. А., *Астрофизика* **2**, 409, 1966.
30. Byram E., Chabb T., Friedman H., *Science* **152**, 66, 1966.
31. Moffet A., *Annual Review of Astronomy and Astrophysics* **4**, 1966.
32. Williams P., *Observatory* **86**, 67, 1966.
33. Dyson F., *Science* **131**, 1667, 1960; *Perspectives in Modern Physics*, «Thoughts on the Search for Extra Terrestrial Technology», Interscience Publishers, N. Y., 1966.
34. Sagan C., Walker R., *Ap. J.* **144**, 1216, 1966.
35. Зельдович Я. Б., *УФН* **89**, 647, 1966.
36. Rocchia R., Rothenflug R., Boclet D., Dueros G., Labeyrie Y., Доклад на 7-м Международном симпозиуме по космическим исследованиям в Вене, 11—17 мая 1966 г.
37. Лебедев Д. С., Левитин Л. Б., Перенос информации электромагнитным полем, в сб. «Теория передачи информации», «Проблемы передачи информации», вып. 16, 1964.
38. «Внеземные цивилизации», Тр. совещания, Бюракан, 20—23 мая 1964 г., изд. АН Арм. ССР, 1965.
39. Kellerman K., *Austr. J. Phys.* **19**, 195, 1966.
40. Weaver H., Williams D., Dieter N. H., Lum W., *Nature* **208**, 29, 1965.
41. Palmer P., Zuckerman B., *HRAP* **124**, 1966 (препринт).
42. Dieter N. H., Weaver H., Williams D., *Sky and Telescope* **31**, 132, 1966.
43. Варшалович Д. А., Письма ЖЭТФ **4**, № 5, 180, 1966.
44. Кайдановский Н. Л., Смирнова Н. А., *Радиотехника и электроника* **10**, 1574, 1965; *Sky and Telescope* **34**, 143, 1967.
45. Симпозиум Международного астрономического союза № 29, май 1966 г., Бюракан.
46. Шоломницкий Г. Б., *Астрон. ж.* **44**, 939, 1967.
47. Reddish V. C., *Vistas in Astronomy* **7**, 173, 1966.

Влияние космической среды на распространение радиосигналов

Проблема поиска сигналов внеземных цивилизаций тесно связана с тщательным анализом радиоволн, приходящих от космических источников. Поэтому определенный интерес представляет изучение условий распространения радиосигналов в космических средах.

Космические среды (межпланетное, межзвездное и межгалактическое пространство) отличаются весьма большой разреженностью вещества. Эффекты влияния этих сред на распространяющиеся сигналы также весьма слабы. Однако благодаря огромным расстояниям, которые проходит сигнал, прежде чем достичь наблюдателя, возможно накопление действия слабых эффектов. Интегральный результат этого действия может внести довольно заметные коррективы в характеристики исследуемого сигнала.

Естественно, что для детального анализа условий распространения необходимо знать основные параметры космических сред: такие, как плотность вещества, структура неоднородностей среды, температура, величина магнитного поля. Эти данные (особенно для межгалактической среды) в настоящее время довольно неопределенны. Кроме того, проведение конкретного анализа необходимо связывать с учетом характеристик вполне определенных областей пространства, через которые проходят исследуемые радиоволны. Например, условия распространения радио-

волн в галактической плоскости значительно отличаются от условий распространения в направлении к полюсу Галактики. Наличие же на луче зрения тех или иных объектов (например, плотной области ионизованного водорода) может сразу внести существенное искажение в сигнал по сравнению со «средними» условиями распространения.

В связи с этим оказывается возможным дать в общем виде лишь предельные оценки влияния космической среды на распространяющиеся радиоволны.

К настоящему времени имеется лишь сравнительно небольшое число работ, в которых подобные оценки проводятся. Рассмотрим некоторые наиболее интересные выводы по «мешающему» влиянию космических сред.

Весьма важным фактором, сказывающимся на распространении радиоизлучения в среде, является поглощение. С классической точки зрения процесс поглощения радиоволн может быть описан как раскачка волной электронов, которые затем теряют приобретенную энергию при столкновении с протонами. Коэффициент поглощения на единицу расстояния будет выражаться соотношением

$$\mu = \frac{1 - n^2}{cn} v_{эф}^{ст}, \quad (2.1)$$

где n — показатель преломления среды, а $v_{эф}^{ст}$ — эффективное число соударений электронов с протонами. Формулу для определения показателя преломления можно записать в виде

$$n(\omega) = \sqrt{1 - \frac{4\pi e^2 N_e}{m\omega^2}}. \quad (2.2)$$

Здесь N_e — электронная концентрация среды, ω — частота радиоизлучения.

При распространении волн сантиметрового и дециметрового диапазонов в разреженных межзвездных и межгалактических средах, когда выполняется условие

$\frac{4\pi e^2 N_e}{m\omega^2} \ll 1$, коэффициент поглощения выражается формулой

$$\mu = \frac{0,58 N_e^2}{T^{3/2} \omega^2} \ln \left(\frac{4,3 \cdot 10^5 T}{\omega^{3/2}} \right), \quad (2.3)$$

где T — температура среды.

Оптическая толщина для поглощения (измеряющая величину поглощения) $\tau = \mu l$ (l — длина пути радиоволн в среде) пропорциональна мере эмиссии $N_e^2 \cdot l$, где N_e^2 — квадрат средней электронной концентрации среды на этом пути. Величина меры эмиссии для расстояний, сравнимых с размерами Галактики, заключена в пределах $6 \cdot 10^{18} \leq N_e^2 l \leq 6 \cdot 10^{20}$.

Значения оптических толщин τ в зависимости от величины меры эмиссии и частоты ω приведены в табл. 2.1.

Таблица 2.1
Оптические толщины τ (Галактика)

ω , гц \diagdown $N_e^2 \cdot l$	$6 \cdot 10^{20}$	$6 \cdot 10^{18}$	$6 \cdot 10^{16}$
10^{10}	$4 \cdot 10^{-5}$	$4 \cdot 10^{-7}$	$4 \cdot 10^{-9}$
10^9	$4 \cdot 10^{-3}$	$4 \cdot 10^{-4}$	$4 \cdot 10^{-5}$
10^8	$4 \cdot 10^{-1}$	$4 \cdot 10^{-2}$	$4 \cdot 10^{-3}$
10^7	40	4	0,4

Из данных таблицы следует, что практически без поглощения по любому из направлений в Галактике распространяются волны с частотой $\omega \geq 10^9$ гц. (Величина поглощения определяется фактором $e^{-\tau}$.) Передача на более низких частотах затруднена сильным поглощением, особенно в направлении плоскости Галактики, где значение $N_e^2 \cdot l$ максимально.

Величина показателя преломления (2.2) определяет также дисперсионные влияния, искажающие передаваемый сигнал. Здесь следует различать два

эффекта: эффект фазового сдвига спектральных составляющих сигнала за счет дисперсии в среде и эффект «запаздывания» групповых квазимонохроматических составляющих, переносящих энергию сигнала.

Чтобы проиллюстрировать эти два явления на простом примере, рассмотрим передачу через космическую среду последовательности импульсов длительности P и такой же «скважности» (рис. 28).

Как известно, импульс длительности P (а также и последовательность таких импульсов) можно разложить в спектр (интеграл Фурье). При этом последовательность импульсов, представленных на рис. 28,

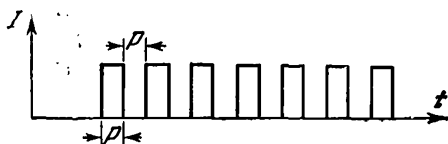


Рис. 28. Неискаженные импульсы.

будет иметь спектр составляющих сигнала в интервале частот шириной $\Delta\omega \sim \frac{2\pi}{P}$.

В средах с дисперсией ($n \neq 1$) фаза отдельных спектральных составляющих сигнала распространяется с различной скоростью $v_{\text{фаз}} = \frac{c}{n(\omega)}$. Поэтому фазы разных составляющих сдвигаются друг относительно друга и результирующее сложение в месте приема даст искаженную форму импульса [1]. В зависимости от характеристик среды может произойти как «сжатие», так и «расплывание» импульса.

Отметим, что в данном рассмотрении мы имели в виду распространение высокочастотных импульсов. Каждый импульс был «заполнен» монохроматическим излучением высокой частоты ω . Подобное рассмотрение корректно только для установившегося процесса передачи высокочастотного сигнала (тогда выполняется условие $\omega P \gg 1$). Это означает, что число периодов колебаний частоты ω , укладывающихся на длине импульса P , довольно значительно. В противном

случае ($\omega P < 1$) импульс будет представлять собой лишь макроскопическое изменение интенсивности низкочастотного поля в среде.

Предположим теперь, что последовательность импульсов (см. рис. 28) создается следующим образом: некий источник излучения с достаточно широким непрерывным спектром излучения периодически закрывается экраном. Тогда посылаемая в пространство последовательность импульсов будет иметь широкополосное «наполнение», возможно, шумового характера (тепловой «шум»).

Пусть полоса частот излучения равна $\Delta\omega$, а спектральная плотность $E(\omega)$. Тогда с энергетической точки зрения каждый импульс будет представлять собой набор квазимонохроматических групп $E(\omega)\delta\omega$, где $\delta\omega$ — очень узкая «квазимонохроматическая» полоска в спектре. Интегральный эффект сложения интенсивностей этих групп и будет определять величину и длительность импульса. Квазимонохроматические группы волн движутся в среде с групповой скоростью $v_{гр} = cn(\omega)$. Для космических ионизованных сред эта скорость тем меньше, чем меньше частота излучения. В результате на достаточно длинных путях распространения более высокочастотные группы волн начинают опережать низкочастотные группы и происходит «развертывание спектра по времени». На значение этого эффекта при возникновении радиовсплесков с дрейфом по частоте в солнечной атмосфере было обращено внимание в работе [6]. Подобный эффект при рассмотрении распространения радионизлучения в межзвездных средах обсуждался в [2]. В дальнейшем эта проблема рассматривалась в [3, 4]*).

Величину запаздывания группы волн частоты ω_1 относительно группы волн частоты ω_2 можно оценить из соотношения

$$\Delta t(\omega_1 - \omega_2) = \frac{l}{cn(\omega_1)} - \frac{l}{cn(\omega_2)}. \quad (2.4)$$

*) Этот эффект был обнаружен при исследовании импульсных источников радионизлучения — пульсаров.

В случае, если значения рассматриваемых частот ω_1, ω_2 далеки от критического значения частоты $\omega_{кр}$, при котором $n(\omega_{кр})=0$ (это условие можно записать как $1 - n \ll 1$), запаздывание определяется формулой

$$\Delta t(\omega_1 - \omega_2) \sim \frac{2\pi e^2}{mc} \frac{\omega_1^2 - \omega_2^2}{\omega_1^2 \omega_2^2} \bar{N}_e \cdot l. \quad (2.5)$$

Это выражение удобно преобразовать, представив ω_1 и ω_2 в виде

$$\omega_1 = \omega_0 + \frac{\Delta\omega}{2},$$

$$\omega_2 = \omega_0 - \frac{\Delta\omega}{2},$$

где ω_0 — средняя частота сигнала. Тогда

$$\Delta t(\Delta\omega) = \frac{2\pi e^2}{mc} \frac{\Delta\omega}{\omega_0^3} N_e \cdot l. \quad (2.6)$$

Величины запаздывания Δt для космических сред в зависимости от $\Delta\omega$ и ω_0 представлены в табл. 2.2. Как видно из данных расчетов, величина запаздывания может оказаться весьма значительной. К чему это приведет?

Для ответа необходимо рассмотреть условия приема сигнала, представленного на рис. 28. Пусть полоса приемного устройства $\Delta\omega_{\text{прием}} \geq \Delta\omega$, т. е. приемник способен принять полную интенсивность всего спектра сигнала $\Delta\omega$. Для простоты будем считать, что $E(\omega) = \text{const}$ во всей полосе $\Delta\omega$. Ясно, что если $\Delta t(\Delta\omega) \ll P$, то существенных искажений в принятом сигнале не будет. В случае $\Delta t \sim P$ (рис. 29, а) форма сигнала исказится. Раньше придут высокочастотные составляющие спектра импульса. Низкочастотные составляющие запоздадут и импульс «расползется».

Если $\Delta t \gg P$ (рис. 29, б), то импульсы полностью «замажутся» и будет наблюдаться только непрерывное (сильно ослабленное по сравнению с пиковой мощностью импульсной передачи) излучение от источника.

Разумеется, истинную форму в принципе можно реставрировать соответствующей коррекцией, введен-

Т а б л и ц а 2.2

Время запаздывания группы волн (сек)
(Галактика, Метагалактика)

$\omega_0, \text{гц}$ $\backslash$ $\bar{N}_e \cdot l$	$5 \cdot 10^{20}$ (предел для межзвездных расстояний)			
	$\Delta\omega = 0,5\omega_0$	$\Delta\omega = 0,1\omega_0$	$\Delta\omega = 10^{-3}\omega_0$	$\Delta\omega = 10^{-5}\omega_0$
10^{10}	0,25	$5 \cdot 10^{-2}$	$5 \cdot 10^{-4}$	$5 \cdot 10^{-6}$
10^9	25	5	$5 \cdot 10^{-2}$	$5 \cdot 10^{-4}$
10^8	$2,5 \cdot 10^3$	500	5	$5 \cdot 10^{-2}$
10^7	$2,5 \cdot 10^5$	$5 \cdot 10^4$	$5 \cdot 10^2$	5

$\omega_0, \text{гц}$ $\backslash$ $\bar{N}_e \cdot l$	$5 \cdot 10^{21}$			
	$\Delta\omega = 0,5\omega_0$	$\Delta\omega = 0,1\omega_0$	$\Delta\omega = 10^{-3}\omega_0$	$\Delta\omega = 10^{-5}\omega_0$
10^{10}	2,5	0,5	$5 \cdot 10^{-3}$	$5 \cdot 10^{-5}$
10^9	250	50	0,5	$5 \cdot 10^{-3}$
10^8	$2,5 \cdot 10^4$	$5 \cdot 10^3$	50	0,5
10^7	$2,5 \cdot 10^6$	$5 \cdot 10^5$	$5 \cdot 10^3$	50

$\omega_0, \text{гц}$ $\backslash$ $\bar{N}_e \cdot l$	$2 \cdot 10^{23}$ (предел для межгалактических расстояний)			
	$\Delta\omega = 0,5\omega_0$	$\Delta\omega = 10^{-3}\omega_0$	$\Delta\omega = 10^{-5}\omega_0$	$\Delta\omega = 10^{-7}\omega_0$
10^{10}	10^2	0,2	$2 \cdot 10^{-3}$	$2 \cdot 10^{-5}$
10^9	10^4	20	0,2	$2 \cdot 10^{-3}$
10^8	10^6	$2 \cdot 10^4$	20	0,2
10^7	10^8	$2 \cdot 10^6$	$2 \cdot 10^4$	20

ной в приемное устройство, или в процессе обработки. Но беда в том, что мы не знаем заранее конкретных параметров тех сред, через которые распространяется далекий космический импульс.

Можно «выловить» периодичность (при $\Delta t \gg P$), сузив полосу приемного устройства $\Delta\omega_{\text{пр}}$ до такой степени, чтобы $\Delta t(\Delta\omega_{\text{пр}}) < P$. Но при этом окажется су-

ущественной потеря в чувствительности приема (она пропорциональна $\left(\frac{\Delta\omega_{\text{пр}}}{\Delta\omega}\right)^2$), что имеет немаловажное значение при исследовании слабых космических излучений. Кроме того, из-за сужения полосы приемника получится большой проигрыш в скорости передачи информации (см. гл. III).

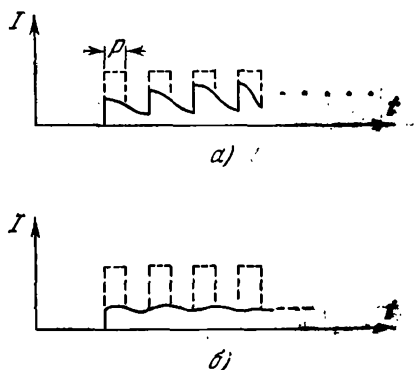


Рис. 29. Искажённые импульсы:
а) случай $\Delta t \sim P$; б) случай $\Delta t \gg P$.

Эффект группового запаздывания накладывает ограничения на возможную длительность импульсов P . Чтобы избежать крайне нежелательных существенных искажений, описанных выше, необходимо, чтобы $\Delta t(\Delta\omega) < P$. Вытекающие из этого условия минимальные значения P приведены в табл. 2.3.

Таблица 2.3

Минимальная длительность импульсов (в секундах)

$\bar{N}_e \cdot l$	$5 \cdot 10^{20}$	$5 \cdot 10^{21}$	$5 \cdot 10^{22}$	$2 \cdot 10^{23}$
ω , гц				
10^{10}	$2,5 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$2,5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$
10^9	$8 \cdot 10^{-4}$	$2,5 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1,5 \cdot 10^{-2}$
10^8	$2,5 \cdot 10^{-2}$	$8 \cdot 10^{-2}$	0,25	0,5
10^7	$8 \cdot 10^{-2}$	2,5	8	15

Следует отметить, что на больших расстояниях в межгалактической среде существенным может оказаться эффект «красного смещения» радионизлучения. «Красное смещение» действует в «неблагоприятную» сторону, понижая частоту идущего излучения. Поэтому на очень больших расстояниях рассмотренные эффекты должны усиливаться.

Специального изучения заслуживают характер искажений, вносимых космической средой в случае других форм радиосигналов (частотная, фазовая модуляция).

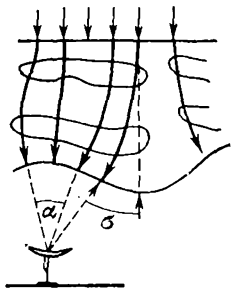


Рис. 30. Искажение фронта волны.

Анализ данных таблиц 2.2—2.3 подчеркивает выгоду использования для дальней космической связи наиболее коротковолновой части радиодиапазона.

Влияние условий распространения радиоволн в космической среде и атмосфере Земли на видимые угловые размеры источника радионизлучения было обсуждено в [5]. Если среда неоднородна, то фронт волны при прохождении через такую среду испытывает искажения (рис. 30). Степень искажения фронта волны определяется величиной отклонения фазы волны от невозмущенного значения. Исходя из некоторых модельных представлений (например, о размере неоднородностей, характере рассеяния волны на них), можно дать среднюю статистическую оценку интегрального искажения после прохождения волной неоднородного слоя определенной толщины. Подсчет среднего квадрата отклонения фазы от невозмущенного значения $\bar{\varphi}^2$ проводился в работе [5] по формуле

$$\bar{\varphi}^2 = \frac{4\pi^{5/2} l \cdot d}{\lambda^2} \overline{\Delta n^2}, \quad (2.7)$$

где l — длина пути волны в неоднородной среде, d — средний размер неоднородностей (считается $d \gg \lambda$),

λ — длина волны излучения, $\overline{\Delta n^2}$ — средняя квадратичная величина флуктуаций показателя преломления среды. Пользуясь приближением геометрической оптики, можно найти и выражение для отклонения луча, идущего от источника к наблюдателю, от невозмущенного направления на угол σ . Имеем

$$\bar{\sigma}^2 = 4\pi^{1/2} \frac{l}{d} \overline{\Delta n^2}. \quad (2.8)$$

Расширение угловых размеров источника до σ будет наблюдаться в дальней зоне рассеивающей области (она определяется расстоянием $R \approx l \geq \frac{d^2}{\lambda}$). Из табл. 2.4 и 2.5 видно, что всегда, когда $\bar{\varphi}^2 \gg 1$, это

Таблица 2.4

Параметры сред, используемые для расчетов в табл. 2.5
 $\bar{N}_e \cdot l$ и $\bar{N}_e^2 \cdot l$

Среда		$l, \text{ см}$	$d, \text{ см}$	$\sqrt{\overline{\Delta n^2}}$	$\bar{N}_e$
Тропосфера		$1,5 \cdot 10^6$	$6 \cdot 10^3$	$0,5 \cdot 10^{-6}$	—
Ионосфера		$4 \cdot 10^7$	$2 \cdot 10^3$	$4,5 \cdot 10^{-12} \lambda^2$	
Межпланетная среда	Эклиптическая плоскость	10^{11}	10^9	$4,5 \cdot 10^{-12} \lambda^2$	10^2
	По направлению к полюсу	$0,5 \cdot 10^{13}$	10^9	$0,5 \cdot 10^{-12} \lambda^2$	20
Межзвездная среда	Галактическая плоскость	$6 \cdot 10^{22}$	$3 \cdot 10^{18}$	$4,5 \cdot 10^{-14} \lambda^2$	$3 \cdot 10^{-2}$
	По направлению к полюсу Галактики	$6 \cdot 10^{20}$	$3 \cdot 10^{18}$	$4,5 \cdot 10^{-14} \lambda^2$	$3 \cdot 10^{-2}$
Межгалактическая среда		10^{28}	10^{22}	$4,5 \cdot 10^{-19} \lambda^2$	10^{-5}

условие на Земле не выполняется. Земной наблюдатель находится в ближайшей зоне рассеяния неоднородностями космической среды. В этом случае характер искажений, вносимых средой в размеры источника, определяется соотношениями между $\sqrt{\bar{\sigma}^2}$ угловыми размерами неоднородностей $\sqrt{\bar{\alpha}^2}$ ($\alpha = \frac{d}{l}$) и размерами антенного устройства D .

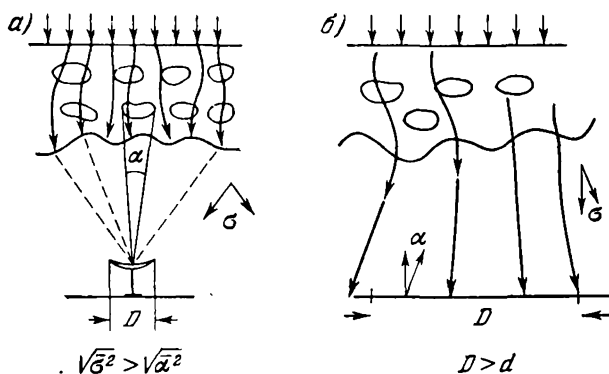


Рис. 31. Отклонение лучей при прохождении через неоднородную среду.

Величины $\sqrt{\bar{\sigma}^2}$, $\sqrt{\bar{\alpha}^2}$ для различных сред приведены в табл. 2.4.

В случае, если $\sqrt{\bar{\sigma}^2} > \sqrt{\bar{\alpha}^2}$, в антенну любого размера придут лучи, прошедшие расстояния больше радиуса корреляции (определяемого средними размерами неоднородностей), и источник расширится до размеров σ (рис. 31, а).

Если же $\sqrt{\bar{\sigma}^2} \ll \sqrt{\bar{\alpha}^2}$, то эффект будет различен в зависимости от соотношений между D и d .

В сплошную антенну размеров $D > d$ (рис. 31, б) флуктуации фазы на неоднородностях внесут потери в значения величины эффективной площади и приведут к расширению диаграммы направленности.

Для антенн размером $D \ll d$ (рис. 30) будет наблюдаться эффект рефракции, который приведет к

Искажение изображений точечного источника

Космическая среда		$\bar{\varphi}^2$, рад ²	$\sqrt{\bar{\sigma}^2}$, рад	$\sqrt{\bar{a}^2}$	Диапазон волн, где $\varphi^2 > 1$ $\sqrt{\bar{\sigma}^2} \ll \sqrt{\bar{a}^2}$	
Межпланетная среда	Плоскость эклиптики	$1,4 \cdot 10^2 \lambda^2$	$4 \cdot 10^{-9} \lambda^2$	10^{-5}	$\varphi^2 > 1$	$\sqrt{\bar{a}^2} \ll \sqrt{\lambda^2}$
	Полярная	$0,28 \lambda^2$	$1,7 \cdot 10^{-10} \lambda^2$	$2 \cdot 10^{-4}$	весь диапазон	$\lambda < 16 \text{ см}$
Межзвездная среда	Плоскость Галактики	$2,5 \cdot 10^{16} \lambda^2$	$1,7 \cdot 10^{-11} \lambda^2$	$5 \cdot 10^{-5}$	весь диапазон	$\lambda < 3,5 \cdot 10^2 \text{ см}$
	Полярная	$2,5 \cdot 10^{14} \lambda^2$	$1,7 \cdot 10^{-12} \lambda^2$	$5 \cdot 10^{-3}$	весь диапазон	весь диапазон
Межгалактическая среда		$1,4 \cdot 10^{15} \lambda^2$	$1,2 \cdot 10^{-15} \lambda^2$	10^{-6}	весь диапазон	весь диапазон

сдвигу видимого положения источника на угол рефракции. В итоге координаты источника будут определяться с ошибкой. При этом надо учитывать и возможные движения неоднородностей космической

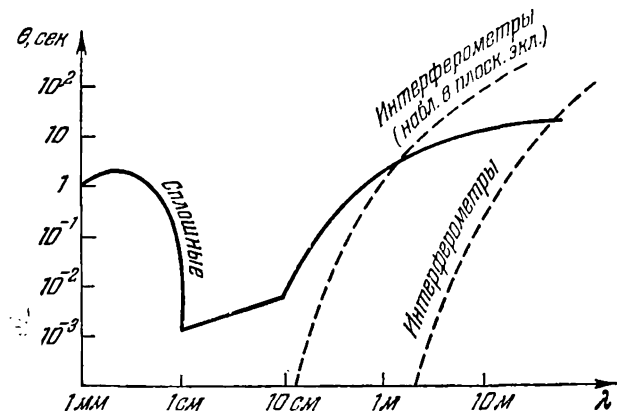


Рис. 32. Предельная разрешающая способность антенн.

среды. Если облака неоднородностей движутся с некоторой скоростью v , то мы будем наблюдать «дрожание» источника с периодом

$$t = \frac{d}{v}. \quad (2.9)$$

Эффекты рассеяния определяют, таким образом, предельную разрешающую способность антенных систем, так как либо любой источник будет казаться расширенным (в случае $\sqrt{\bar{\sigma}^2} > \sqrt{\bar{a}^2}$), либо рассеяние будет ухудшать коэффициент направленности антенны ($\sqrt{\bar{\sigma}^2} \ll \sqrt{\bar{a}^2}$, $D > d$).

В таблице 2.5 приведены диапазоны волн для различных космических сред, где выполняются условия $\varphi^2 \gg 1$, $\sqrt{\bar{\sigma}^2} \ll \sqrt{\bar{a}^2}$. Видно, что межпланетная среда может вносить существенные искажения в определенные угловых размеров источника.

На рис. 32 приведены графики предельной разрешающей способности антенных систем, которая может быть получена при учете влияния рассеивающих космических сред. Сплошные антенные системы имеют существенное ограничение в величине разрешающей способности. Максимальную разрешающую способность можно реализовать только с помощью интерференционных систем (для волн короче 10 см). Это еще раз подтверждает выгоду использования для наблюдений наиболее коротковолновой части радиодиапазона.

В заключение следует отметить, что для астрофизических целей весьма плодотворным может оказаться «обращение» задачи об учете искажений, вносимых средой. Если из каких-либо независимых соображений могут быть оценены истинные угловые размеры источников или характеристики переменного радиосигнала, идущего от источника, то изучение искажений, вносимых космической средой, может дать ценную информацию о характеристиках самой этой среды (плотности вещества, размерах неоднородностей и др.).

ЛИТЕРАТУРА

1. Гинзбург В. Л., Распространение электромагнитных волн в плазме, Физматгиз, 1960.
2. Гудзенко Л. И., Пановкин Б. Н., в сборн. «Внеземные цивилизации», Тр. совещания, Бюракан, 20—23 мая 1964 г., изд. АН Арм. ССР, 1965, стр. 68.
3. Haddock F. I., Sciana D. W., Phys. Sci. Let. **14**, 25, 1007, 1965.
4. Пановкин Б. Н., Доклад на V Всесоюзной конференции по радиоастрономии, Харьков, 1965.
5. Кайдановский Н. Л., Смирнова Н. А., Радиотехника и электроника, X, 1574, 1965.
6. Wild J. P., Sheridan K. V., Neylan A. A., Austr. J. Phys. **12**, 369, 1959.

Возможность радиосвязи с внеземными цивилизациями

- Начиная с пионерской работы Коккони и Моррисона [1], которые показали возможность осуществления связи с внеземными цивилизациями (ВЦ) с помощью электромагнитных волн, эта тема неоднократно обсуждалась в научной литературе [2, 3]. Нет сомнения в том, что организация связи с ВЦ в техническом отношении представляет собой беспрецедентную задачу, специфику которой невозможно оценить полностью, поскольку мы лишь приступаем к изучению этой проблемы. Вместе с тем можно полагать, что любая система связи, в том числе и система связи с ВЦ, должна удовлетворять некоторым определенным требованиям, которые вытекают из общих закономерностей передачи информации. Изучением этих закономерностей занимается теория информации или общая теория связи. Поэтому представляется полезным начать наш обзор с рассмотрения основных элементов общей теории связи, на которой базируется весь дальнейший анализ проблемы.

§ 1. Элементы общей теории связи

Структура и основные характеристики системы связи. Назначение всякой системы связи состоит в передаче некоторых сообщений. Сообщение может представлять собой текст, записанный с помощью букв некоторого алфавита (как при телеграфировании) или выраженный в звуковой форме (телефон, радиовещание). Оно может представлять собой также изо-

бражение некоторого объекта (фототелеграф, телевидение) или алгоритм, который необходимо передать в систему автоматического управления. Любое из этих сообщений может быть представлено как последовательность цифр или как некоторая непрерывная функция времени $x(t)$.

Сообщения передаются системой связи с помощью определенных сигналов. В настоящей главе мы будем рассматривать только такие системы связи, в которых для передачи сообщений используются электрические

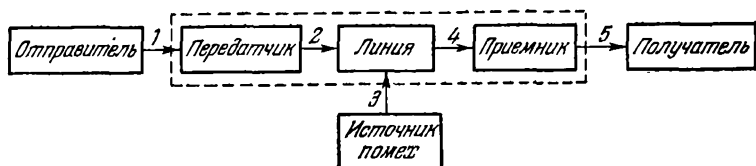


Рис. 33. Обобщенная система связи.

1 — сообщение, 2 — сигнал, 3 — шум, 4 — сигнал + шум, 5 — принятое сообщение. Часть системы в рамке из прерывистой линии — канал связи.

сигналы *). Электрический сигнал представляет собой изменяющуюся со временем электрическую величину (напряжение, ток, напряженность поля) и, следовательно, так же как и сообщение, может быть выражен некоторой функцией времени. Сигнал отображает сообщение в форме определенного электрического возмущения. По принятому сигналу необходимо восстановить переданное сообщение.

Блок-схема обобщенной системы связи изображена на рис. 33. Сообщение от источника информации, или отправителя, поступает в передатчик, который преобразует его в сигнал и направляет в линию связи.

*) При более широкой трактовке связи, когда рассматриваются системы типа биологической популяции, биологической эволюции и др., основные понятия — такие, как сообщение, сигнал, информация — нуждаются в уточнении. Впрочем, необходимо отметить, что после соответствующего обобщения основных понятий важнейшие положения теории электрической связи оказываются справедливыми и для более широкого класса систем связи.

Линия — это тракт электромагнитной волны, среда, используемая для передачи сигнала от передающего конца системы связи к приемному. Она может представлять собой пару проводов, коаксиальной кабель, волновод или просто зону пространства, в которой распространяются радиоволны. Так, для радиовещания на коротких волнах линия — это шаровой слой между поверхностью Земли и ионосферой. При направленной радиопередаче линия связи представляет собой область пространства внутри телесного угла, опирающегося на приемную антенну.

При распространении по линии связи сигнал может искажаться и на него могут налагаться помехи. Под искажениями обычно понимают такие изменения сигналов, которые обусловлены известными характеристиками системы. Подобные искажения, в принципе, могут быть скорректированы, и мы ими в дальнейшем интересоваться не будем *). Помехи же, как правило, имеют случайный характер и не могут быть полностью скомпенсированы. Случайные помехи часто называют шумами. Этот фактор имеет важнейшее значение для характеристики системы связи.

На приемном конце линии связи электрический сигнал поступает в приемник, который преобразует принятый сигнал и восстанавливает по нему переданное сообщение. Действие приемника математически обратно действию передатчика. Часть системы, включающая передатчик, линию и приемник, обычно называют каналом связи.

В идеальной системе связи, в которой не действуют помехи, принятое сообщение тождественно переданному. В реальных же системах из-за действия помех этого никогда не бывает. Степень соответствия

*) Искажения, которые испытывает сигнал при распространении в межзвездной среде, рассматривались в гл. II. Отметим, что при распространении в среде со случайно изменяющимися свойствами сигнал испытывает случайные искажения, которые невозможно скорректировать. Пример подобных искажений — хорошо известное явление мерцания звезд и источников радиоизлучения. Тот же эффект приводит к расширению угловых размеров источника (см. гл. II).

принятого сигнала переданному характеризует надежность связи. Надежность зависит от отношения мощности сигнала к мощности шума в канале связи. Как правило, надежность падает с расстоянием. Предельное расстояние, на котором еще достигается заданная надежность, определяет дальность связи. В системе связи с ВЦ этот параметр имеет первостепенное значение.

Другой важной характеристикой является пропускная способность канала связи, т. е. количество информации, которое может быть передано данной системой связи за единицу времени. Пропускная способность характеризует информативность связи. Однако система связи «не различает» ценности передаваемого сообщения, т. е. для передачи телеграммы, состоящей, например, из 100 знаков, к системе связи всегда предъявляются вполне определенные требования (время передачи, полоса частот, мощность сигнала и пр.), независимо от важности содержащегося в ней текста. Поэтому понятие информации, с которым имеет дело общая теория связи, лишено какого бы то ни было качественного оттенка и представляет собой чисто количественное понятие.

Количественное определение информации. Как определяется количество информации? Пусть, например, требуется передать ряд сообщений, представляющих собой четырехзначные десятичные числа. Это могут быть либо численные значения некоторой величины либо четырехбуквенные слова, записанные с помощью десятибуквенного алфавита. Пусть теперь мы передаем некоторое определенное число M . Каково информационное содержание нашего сообщения? Общее число четырехзначных чисел или возможных сообщений $N=10^4$. Передавая именно наше сообщение, т. е. заданное число M , мы делаем вполне определенный выбор из $N=10^4$ возможных выборов. Число N возможных выборов характеризует собой неопределенность ответа, которая существовала до передачи сообщения. Эта же величина N применяется для характеристики информативности данного сообщения.

Чем больше первоначальная неопределенность, которая существовала до передачи сообщения, тем больше количество информации, которое содержит в себе это сообщение и, наоборот, чем меньше первоначальная неопределенность, тем меньше количества информации содержится в переданном сообщении. Если обозначить количество информации через Q , то на основании сказанного можно написать

$$Q = Q(N), \quad (3.1)$$

причем Q — однозначная монотонно-возрастающая функция N . Из этого определения следует, что если $N_1 = N_2$, то

$$Q_1 \equiv Q(N_1) = Q(N_2) = Q_2. \quad (3.2)$$

Мы можем, например, записать наше четырехзначное десятичное число M в двоичной, трюичной или любой другой системе счисления с произвольным основанием a . Тогда $N_1 = a_1^{m_1}$, $N_2 = a_2^{m_2}$, где m — значность числа M (при целом m это просто количество цифр, которым изображается это число) в данной системе счисления. И наше условие (3.1) принимает вид

$$Q_1 = Q_2, \text{ если } a_1^{m_1} = a_2^{m_2}. \quad (3.3)$$

В таком виде оно просто означает, что количество информации, которое содержится в числе M , не зависит от способа его написания.

Следующее условие, которому должно удовлетворять наше определение информации, состоит в том, что если взять разные числа, записанные в одной и той же системе счисления, то информационно содержание каждого числа будет пропорционально его значности (т. е. шестизначное десятичное число 145 876 содержит вдвое больше информации, чем трехзначное число 963 и в три раза больше информации, чем двухзначное число 25). Итак,

$$Q = \gamma m, \quad (3.4)$$

где γ — коэффициент пропорциональности.

Применительно к передаче информации по каналу связи это означает, что количество переданной ин-

формации растет линейно со временем (действительно, для передачи шести знаков требуется вдвое больше времени, чем для передачи трех знаков). Таким образом, двухминутное сообщение, вообще говоря (при прочих равных условиях), вдвое информативнее одноминутного.

Можно показать, что условия (3.1) — (3.4)*) однозначно определяют вид функции

$$Q = \log_b N = m \log_b a. \quad (3.5)$$

Это определение принадлежит Хартли [4]. Оно было предложено в 1928 г. и с тех пор плодотворно используется в теории связи.

Основание b логарифма в выражении (3.5) является произвольным. Выбор этого основания означает выбор единицы измерения информации. Полагая $b=2$, получим количество информации Q в двоичных единицах или *битах*. Эту единицу, связанную с минимально возможным основанием системы счисления, можно считать основной единицей измерения информации. Она широко используется в технике.

Теперь мы можем сказать, сколько информации содержится в нашем четырехзначном десятичном числе M . Полагая $m=4$ и $a=10$, получим $Q=4 \log_2 10 \approx 13,3$ бита. Аналогично, количество информации в слове из пяти букв при 30-буквенном алфавите составляет $5 \log_2 30 = 24,6$ бита, а текст, состоящий из 100 слов при среднем количестве букв в слове, равном 5, содержит около 2460 двоичных единиц информации. При $a=b=2$ величина $Q=m$ бит, т. е. количество информации в битах, содержащееся в некотором числе M , равно количеству двоичных знаков, с помощью которых это число записывается в двоичной системе счисления (разумеется, при целом m)).

*) Поскольку условия (3.1) — (3.3) не являются независимыми, для определения вида функции Q достаточно выполнения какой-либо пары условий: (3.1) и (3.4), (3.2) и (3.4) или (3.3) и (3.4).

**) В случае, если m не является целым числом, для записи M используется m_1 двоичных знаков, где m_1 — ближайшее к m целое число: $m_1 > m = Q$.

Приведенное определение относилось к дискретным сообщениям. Но так как непрерывную функцию времени с любой степенью точности можно представить совокупностью дискретных величин, то это определение имеет в теории связи самое общее значение.

Перейдем теперь к рассмотрению способов преобразования сообщения в сигнал.

Преобразование сообщения в сигнал. Виды модуляции. Переносчиком сигнала может быть постоянный ток, электромагнитные колебания высокой частоты или периодическая последовательность импульсов. При передаче сигнала по линии связи один из параметров переносчика меняется в соответствии с функцией сообщения $x(t)$.

Постоянный ток характеризуется двумя параметрами: величиной и направлением тока. Изменяя одну из этих величин в соответствии с $x(t)$, получим электрический сигнал, который может распространяться по линии связи (как, например, при передаче телеграмм посредством азбуки Морзе). Но так как постоянный ток может распространяться только по проводам, этот способ передачи сигналов не имеет никакого значения для обсуждаемой нами проблемы.

В радиовещании переносчиком сигналов являются электромагнитные колебания высокой частоты, которые могут свободно распространяться в пространстве. Синусоидальное колебание характеризуется тремя параметрами: амплитудой, частотой колебания и начальной фазой. Изменяя один из этих параметров в соответствии с функцией сообщения, которую часто называют модулирующей функцией, мы получим электрический сигнал в виде модулированного колебания высокой (несущей) частоты. Соответственно трем параметрам несущего колебания различают три вида модуляции: амплитудную модуляцию АМ, частотную модуляцию ЧМ и фазовую модуляцию ФМ (рис. 34, а). Поступающие в приемник модулированные колебания подвергаются преобразованиям и по ним восстанавливается модулирующая функция $x(t)$, представляющая собой функцию сообщения.

Если переносчиком сигналов служит периодическая последовательность импульсов, то, изменяя высоту импульсов h , их длительность τ , частоту следования $\nu_0 = \frac{1}{T}$ (T — промежуток времени между двумя

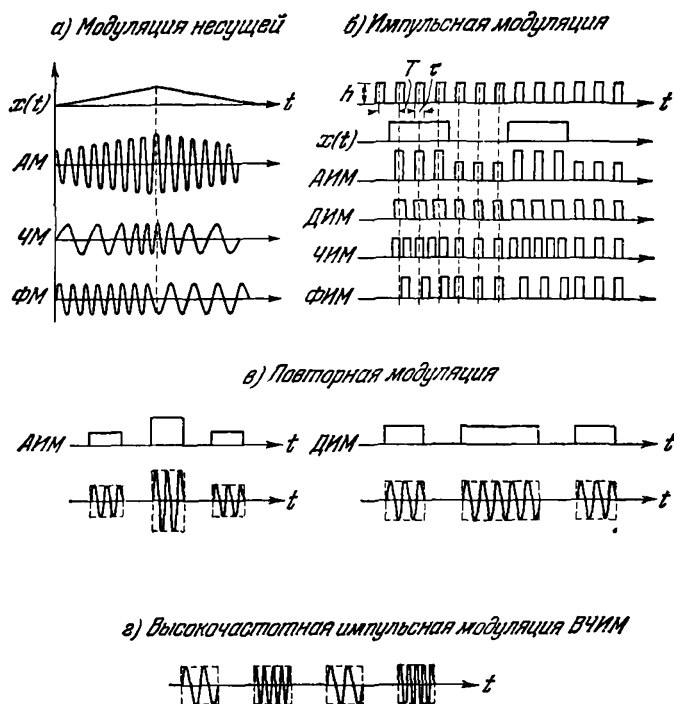


Рис. 34. Формирование электрических сигналов с помощью модуляции.

$x(t)$ — функция сообщения или модулирующая функция, АМ — амплитудная модуляция, ЧМ — частотная модуляция, ФМ — фазовая модуляция; АИМ — амплитудно-импульсная модуляция, ДИМ — модуляция импульсов по длительности или широтно-импульсная модуляция, ЧИМ — частотно-импульсная модуляция, ФИМ — фазово-импульсная модуляция.

последовательными импульсами) или положение импульсов во времени относительно немодулированной последовательности, получают четыре вида

импульсной модуляции ИМ (рис. 34, б): амплитудно-импульсную модуляцию АИМ, модуляцию импульсов по длительности ДИМ, частотно-импульсную модуляцию ЧИМ, и фазово-импульсную модуляцию ФИМ. В ряде случаев применяется повторная модуляция, когда последовательность импульсов модулируется в соответствии с функцией сообщения, а полученные модулированные импульсы используются для модуляции несущего колебания высокой частоты (рис. 34, в). При этом появляется возможность еще одного вида модуляции, — высокочастотной импульсной модуляции ВЧИМ, при которой высота, длительность, частота следования и фаза импульсов остаются постоянными, меняется лишь частота заполнения (рис. 34, г).

Важной разновидностью импульсной модуляции является передача сообщений посредством кода. Мы рассмотрим этот вид модуляции после ознакомления с некоторыми свойствами сигналов.

Физические характеристики сигналов. Для характеристики сигналов обычно употребляются следующие три параметра: длительность сигнала, динамический диапазон и ширина спектра (ширина полосы частот).

Длительность сигнала — наиболее простая характеристика. Практическое значение ее очевидно: чем длиннее сигнал, тем большее время требуется на его передачу, тем дольше занимает линия.

Динамический диапазон — это отношение максимальной мгновенной мощности сигнала (так называемой пиковой мощности) к минимальной мощности сигнала. Динамический диапазон измеряется логарифмической мерой и выражается в децибелах. Один децибел (1 дб) равен 0,1 логарифма; поэтому если n — отношение измеряемых величин в линейной мере, то это же отношение в децибелах равняется $10 \lg n$. Таким образом, превышению максимальной мгновенной мощности над минимальной в два раза соответствует динамический диапазон 3 дб, превышению в 10 раз соответствует 10 дб, превышению в 100 раз — 20 дб, в 1000 раз — 30 дб и т. д.

Выбор минимальной мощности связан с уровнем помех. Для уверенного приема сигнала необходимо, чтобы его минимальная мощность в определенное число раз превышала среднюю мощность шума $P_{\text{ш}}$ ($P_{\text{min}} = \alpha P_{\text{ш}}$). Например, для высококачественной передачи речи с помощью амплитудной модуляции требуется, чтобы P_{min} превышало среднюю мощность шума на 60—70 дБ. Величина $P_{\text{min}} = \alpha P_{\text{ш}}$ называется *пороговой* мощностью сигнала. Часто вместо динамического диапазона, связанного с определением пороговой мощности, пользуются просто отношением средних мощностей сигнала и шума $P_c/P_{\text{ш}}$. Это отношение для краткости называют отношением «сигнал/шум» или «сигнал/помеха». Обе величины — и динамический диапазон и отношение «сигнал/помеха» — характеризуют мощность сигнала, но не абсолютную, а относительную, отнесенную к мощности шума. Чем же определяется пороговая мощность?

Пусть требуется передать некоторое сообщение, представляющее собой значение функции $x(t)$ в момент времени t_0 . Мы можем использовать одну из систем импульсной модуляции, например, амплитудно-импульсную, и послать по линии связи импульс, высота которого равняется $x(t_0) = x_0$. При использовании идеального бесшумного канала этот импульс в искаженном виде будет получен на приемном конце линии связи и по его амплитуде x_0 будет восстановлено исходное сообщение $x(t_0)$. В реальном канале на сигнал накладывается помеха, поэтому принятое значение амплитуды будет $x_0 + \xi$, где ξ — амплитуда шума (положительная или отрицательная). Пусть мы хотим, чтобы наше сообщение было восстановлено с точностью до 0,001, т. е. относительная ошибка $\frac{\Delta x_0}{x_0} = 0,001$. Для этого необходимо, чтобы выполнялось условие

$$|\xi| < \frac{1}{2} \Delta x_0. \quad (3.6)$$

Если $|\xi| = \text{const}$, т. е. помеха постоянна, то для выполнения условия достаточно, чтобы $x_0 > 2000 |\xi|$ или

$P_c = x_0^2 > 4 \cdot 10^6 \xi^2 = 4 \cdot 10^6 P_{\text{ш}}$. Иными словами, необходимо четырехмиллионное (66 дб) превышение сигнала над шумом. Это и будет пороговое значение мощности при передаче мгновенного значения $x(t)$ по системе АИМ с ошибкой, не превышающей 0,001.

Рассмотренный случай связи с постоянной помехой является тривиальным: постоянные помехи обычно легко компенсируются. Трудность как раз состоит в том, что реальная помеха носит случайный характер и поэтому компенсировать ее невозможно. Случайная помеха может, вообще говоря, принимать любые, сколь угодно большие значения, однако вероятность такого события невелика. Для определения пороговой мощности при наличии случайных помех надо вычислить вероятность того, что помеха не превысит $\frac{1}{2} \Delta x_0$, т. е. вероятность выполнения условия (3.6). Если помеха представляет собой гауссов шум, т. е. ее амплитуда подчиняется нормальному закону распределения, то искомая вероятность будет

$$p_0 \left(|\xi| < \frac{1}{2} \Delta x_0 \right) = \Phi \left(\frac{\Delta x_0}{2 \sqrt{2} \sigma} \right) = \Phi(z), \quad (3.7)$$

а вероятность ошибки

$$p = 1 - p_0 = 1 - \Phi(z), \quad (3.8)$$

где σ — параметр нормального закона, $\sigma = \sqrt{\xi^2} = \sqrt{P_{\text{ш}}}$, а Φ — функция Лапласа или интеграл вероятностей. Для нее составлены таблицы, пользуясь которыми можно определить интересующую нас величину. При $\Delta x_0 = 10\sigma$ вероятность ошибки имеет порядок 10^{-6} , далее она быстро падает по мере роста $\frac{\Delta x_0}{\sigma}$. Для

большинства практических задач надежность, при которой вероятность ошибки равна 10^{-6} , считается достаточно хорошей. Таким образом, если мы хотим обеспечить надежную (в указанном выше смысле) передачу с точностью воспроизведения сигнала

$$\frac{\Delta x_0}{x_0} = 0,001, \text{ то } x_0 = 1000 \Delta x_0 = 10^4 \sigma \text{ и } \frac{P_c}{P_{\text{ш}}} = 10^8.$$

Мы рассмотрели определение пороговой мощности в простейшем случае передачи сообщения с помощью одного-единственного импульса. Но это определение остается в силе и при любом сколь угодно сложном электрическом сигнале $x(t)$. Только под величиной x_0 в подобном случае следует понимать минимальное значение амплитуды сигнала ($x_0^2 = P_{\min}$).

Отметим еще, что при заданном динамическом диапазоне и заданной минимальной мощности сигнала средняя мощность также имеет вполне определенное значение. При уменьшении средней мощности связь становится ненадежной. Таким образом, наряду с минимальной пороговой мощностью $P_{\min} = \alpha P_{\text{ш}}$, можно говорить также о *пороговом значении средней мощности сигнала*. Ниже (в разделе «Передача непрерывных функций с помощью импульсных сигналов») будет показано, как определяется пороговое значение средней мощности для определенных видов сигнала (ИКМ с произвольным основанием кода). А сейчас мы перейдем к спектральным характеристикам сигнала.

Хорошо известно, что всякую периодическую функцию $x(t)$ с периодом T можно представить в виде суммы гармонических колебаний с кратными частотами (разложение в ряд Фурье):

$$x(t) = \sum_{k=1}^{\infty} c_k \cos(\omega_k t + \varphi_k). \quad (3.9)$$

Каждая составляющая (гармоника) представляет собой синусоидальное колебание с частотой ω_k , амплитудой c_k и фазой φ_k . Частоты отдельных гармоник отличаются в целое число раз и связаны с периодом функции соотношением: $\omega_k = k \frac{2\pi}{T}$ ($k=1, 2, 3, \dots$).

Наименьшая частота равна $\omega_1 = \frac{2\pi}{T}$, на эту же величину отличаются между собой частоты двух соседних гармоник. Значения c_k и φ_k зависят от вида функции $x(t)$. Совокупность величин c_k образует спектр амплитуд, совокупность величин φ_k — спектр фаз. Такой

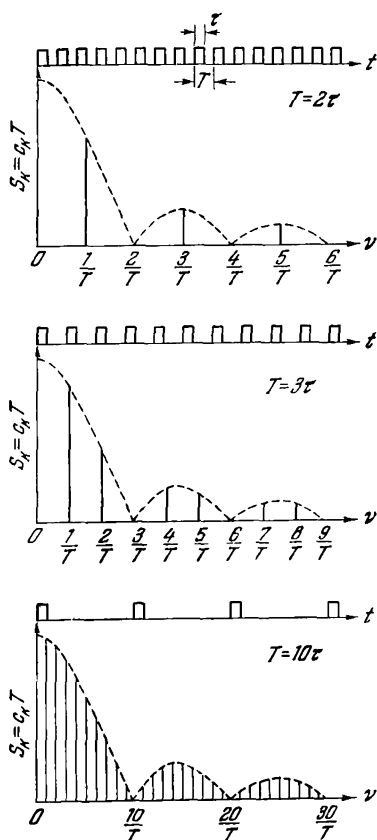


Рис. 35. Спектр периодической последовательности импульсов.

По оси ординат отложена величина $S_k = c_k T$ (произведение амплитуды соответствующей гармоники на период). Пунктирная линия изображает спектральную плотность амплитуды для одиночного импульса. С увеличением периода расстояние между спектральными линиями сокращается и в пределе при $T \rightarrow \infty$ получается сплошной спектр, совпадающий со спектром одиночного импульса.

спектр, состоящий из отдельных дискретных значений, называется линейчатым. С увеличением периода расстояние между линиями сокращается, и в пределе при $T \rightarrow \infty$ (т. е. для непериодической функции) получается сплошной спектр (рис. 35). Математически он описывается с помощью интеграла Фурье.

Задание спектра амплитуд и фаз полностью определяет функции $x(t)$. Поэтому всякий процесс можно описать либо заданием соответствующей функции времени либо заданием спектра, который является функцией частоты. Оба представления (частотное и временное) совершенно равносильны.

Все сигналы, с которыми приходится иметь дело на практике, представляются функциями с ограниченным спектром. Это значит, что они не содержат частот меньше некоторой минимальной частоты ν_1 и больше некоторой максимальной частоты ν_2 ,

т. е. занимают вполне определенную конечную полосу частот от ν_1 до ν_2 . Полоса частот, занимаемая сигналом, и определяет ширину его спектра $\Delta\nu = \nu_2 - \nu_1$. Это очень важная характеристика сигнала. При передаче по каналу связи занимаемая сигналом полоса частот может смещаться по спектру в сторону высоких или низких частот. Однако ширина спектра или ширина полосы частот $\Delta\nu$ должна оставаться постоянной*). Преобразование (сдвиг) частоты широко используется в радиотехнике, например, в приемниках супергетеродинного типа. Применение этого принципа в системах связи позволяет осуществлять одновременную передачу многих сообщений по одной линии, используя разные частоты.

Чем больше ширина полосы Δf линии связи, тем большее количество сигналов с заданной шириной спектра $\Delta\nu$ можно передать одновременно. Каждый сигнал связан с определенным сообщением, которому соответствует определенное количество информации. Следовательно, можно заключить, что количество информации, переданное за единицу времени по некоторому каналу связи, пропорционально его полосе частот.

Зависимость между длительностью и шириной спектра импульсов. Количество импульсов, пропускаемых за единицу времени через канал с заданной полосой частот Δf . Существует фундаментальное соотношение, связывающее продолжительность импульса с шириной его спектра $\Delta\nu$:

$$\tau \Delta\nu = \text{const.} \quad (3.10)$$

Из этого соотношения следует, что ширина спектра импульса обратно пропорциональна его длительности. Численное же значение постоянной зависит от формы

*) На отдельных этапах передачи сигнал может испытывать преобразования, при которых ширина его спектра $\Delta\nu$ меняется. Например, при ЧМ сигнал, распространяющийся в линии связи, имеет полосу в n раз шире полосы модулирующей функции (n — индекс частотной модуляции). Однако после преобразования в приемнике восстанавливается сигнал с полосой $\Delta\nu$, равной полосе частот модулирующей функции $x(t)$.

импульса. Однако во всех случаях эта константа — порядка единицы, а для некоторых видов импульсов, например для прямоугольного, ее можно с достаточной степенью точности положить равной единице.

Соотношение (3.10) является весьма общим соотношением, которое справедливо для любого временного процесса длительностью τ . Отсюда следует, что если имеется непрерывная функция времени $x(t)$ с полосой частот $\Delta\nu$ и длительностью $\Delta t > \Delta\nu^{-1}$, то она представляет собой совокупность отдельных импульсов разной длительности $\tau_i < \Delta t$, из которых самый короткий импульс имеет продолжительность τ порядка $\frac{1}{\Delta\nu}$.

Определим теперь, сколько импульсов можно передать в единицу времени по каналу с полосой частот Δf . Пусть импульс имеет длительность $\tau_1 = \frac{1}{\Delta f}$.

Ширина спектра такого импульса $\Delta\nu_1 = \frac{1}{\tau_1} = \Delta f$ (мы положили константу в выражении (3.10) равной единице). Ввиду равенства полосы частот рассматриваемого канала связи и импульса все частотные компоненты последнего пройдут через данный канал связи, и импульс может быть восстановлен в неискаженном виде. Общее количество импульсов, прошедших через канал за единицу времени, будет равно $\frac{1}{\tau_1} = \Delta f$. Пусть

теперь импульс имеет в 10 раз большую длительность $\tau_2 = \frac{10}{\Delta f}$. Ширина спектра такого импульса в десять

раз меньше: $\Delta\nu_2 = \frac{1}{\tau_2} = 0,1 \Delta f$. Применяя частотное разделение сигналов, мы можем разместить в нашей линии связи 10 частотных каналов шириной $0,1 \Delta f$. По каждому из таких каналов в единицу времени

можно передать $\frac{1}{\tau_2} = 0,1 \Delta f$ импульсов, а общее

число импульсов, прошедших через все 10 частотных каналов нашей линии связи, по-прежнему будет равно Δf . Пусть, наконец, продолжительность импульсов

$\tau_3 < \frac{1}{\Delta f}$. Тогда ширина спектра каждого импульса будет больше Δf . Компоненты импульса с частотами $\nu > \Delta f$ не смогут пройти через рассматриваемый канал связи, что приведет к искажению формы импульса. Поэтому может показаться, что величина Δf определяет максимальное число импульсов, пропускаемых без искажения за 1 сек через рассматриваемый канал связи. Тем не менее, это не так; более строгое рассмотрение показывает, что максимальное количество импульсов вдвое больше и равно $2 \Delta f$.

Действительно, пусть на входе линии связи стоит идеальный частотный фильтр с полосой пропускания Δf и пусть в момент $t=0$ на вход фильтра поступает короткий ($\tau < \frac{1}{\Delta f}$) импульс произвольной формы. После прохождения через фильтр импульс расплывается и приобретает форму, описываемую функцией

$$x(t) = x_0 \frac{\sin 2\pi \Delta f t}{2\pi \Delta f t}, \quad (3.11)$$

где x_0 — амплитуда исходного короткого импульса. Свойствами этой функции и определяется возможность пропустить в единицу времени через канал связи количество импульсов, вдвое превышающее его полосу частот. График функции (3.11) приведен на рис. 36. При $t=0$ $x=x_0$; при $t=1/2\Delta f$, $2/2\Delta f$, $3/2\Delta f$, ... $x(t)=0$. Если теперь мы пошлем серию коротких импульсов через равные промежутки времени $\Delta t = 1/2\Delta f$, то получим некоторый комбинированный сигнал, представляющий собой сумму сигналов вида (3.11), только смещенных относительно момента $t=0$ на величину $i\Delta t$ ($i=1, 2, 3 \dots$). Этот суммарный сигнал будет иметь вид

$$\tilde{x}(t) = \sum_i x_i \frac{\sin 2\pi \Delta f \left(t - \frac{i}{2\Delta f}\right)}{2\pi \Delta f \left(t - \frac{i}{2\Delta f}\right)}. \quad (3.12)$$

Так как каждое слагаемое этой суммы равно нулю в любой из моментов времени $t_j = j\Delta t$ для всех

$j=1, 2, 3 \dots$, за исключением $j=i$ (см. рис. 35), то суммарный сигнал в каждый момент посылки t_i будет определяться только амплитудой x_i соответствующего короткого импульса. Таким образом, несмотря на искажение коротких импульсов после прохождения через фильтр с полосой Δf , эти импульсы при скорости следования $2\Delta f$ импульсов в секунду могут быть полностью восстановлены, если изменение импульсов на приемном конце линии связи происходит с той же скоростью (через промежуток времени $\Delta t = 1/2\Delta f$).

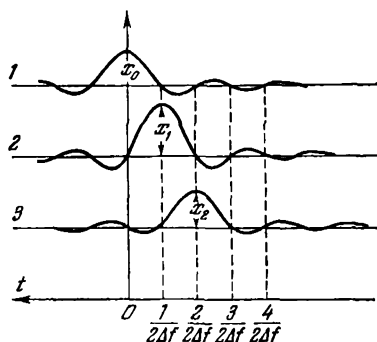


Рис. 36. К определению числа импульсов, проходящих за единицу времени через канал с заданной полосой частот.

После прохождения короткого импульса произвольной формы через идеальный фильтр нижних частот с полосой пропускания Δf импульс искажается и приобретает форму кривых, изображенных на этом рисунке: 1) $x(t) = x_0 \frac{\sin 2\pi \Delta f t}{2\pi \Delta f t}$.

$$2) x(t) = x_1 \frac{\sin 2\pi \Delta f \left(t - \frac{1}{2\Delta f}\right)}{2\pi \Delta f \left(t - \frac{1}{2\Delta f}\right)}, \quad 3) x(t) = x_2 \frac{\sin 2\pi \Delta f \left(t - \frac{2}{2\Delta f}\right)}{2\pi \Delta f \left(t - \frac{2}{2\Delta f}\right)}.$$

Кривые отно-

сятся к разным импульсам с амплитудой x_0, x_1, x_2 , поступающим на вход фильтра в моменты $t_0 = 0, t_1 = \frac{1}{2\Delta f}$,

$t_2 = \frac{2}{2\Delta f}$. В момент t_i амплитуда i -го импульса (после прохождения через фильтр) равна x_i , а для всех остальных импульсов амплитуда равна нулю. Поэтому суммарный сигнал в момент t_i определяется только амплитудой x_i исходного импульса.

Передача непрерывных функций с помощью импульсных сигналов. Непрерывная функция сообщения $x(t)$ продолжительностью $\Delta T = t_2 - t_1$ может быть представлена совокупностью дискретных значений $x(t_k)$, взятых через промежутки времени Δt_k . Очевидно, представление будет тем более точным, чем меньше промежутков времени Δt_k . Дискретные значения функции

можно передать по каналу связи, используя одну из систем импульсной модуляции. Если Δf — полоса ча-

стот используемой линии связи, то, как было показано, максимальное количество импульсов, которое может быть передано по ней за единицу времени, равно $2\Delta f$. Применяя импульсы, следующие с такой скоростью, мы получим на приемном конце линии связи функцию времени $\tilde{x}(t)$, которая определяется выражением (3.12). При отсутствии шума значения этой функции в моменты отсчета t_k , как мы только что видели, будут определяться только значениями исходной функции: $\tilde{x}(t_k) = x(t_k)$. Возникает вопрос, будут ли эти функции равны при любых значениях t , не обязательно равных t_k , т. е. будут ли они совпадать тождественно? Очевидно, совпадение будет тем более полным, чем медленнее меняется функция в промежутках времени между отсчетами t_k . А это значит, что она не должна содержать достаточно высоких гармоник. Согласно теореме В. А. Котельникова совпадение функций будет тождественным, если исходная функция $x(t)$ не содержит компонент с частотой ν , большей чем Δf , т. е. если она имеет ширину спектра $\Delta\nu$, равную полосе частот используемого канала связи. Теорема В. А. Котельникова имеет очень большое значение для теории и техники связи, так как позволяет свести передачу непрерывных функций к передаче некоторой последовательности дискретных величин. Эта теорема утверждает, что функция с ограниченным спектром $\Delta\nu$ полностью определяется своими значениями, отсчитанными через интервалы $\Delta t = 1/2\Delta\nu$. В частности, если функция имеет длительность Δt , т. е. отлична от нуля только на интервале $t_0 < t < t_0 + \Delta t$, то она определяется совокупностью $2\Delta t\Delta f$ дискретных значений. А если так, то к непрерывным функциям с ограниченным спектром можно применить определение количества информации, которое было получено для дискретных сообщений.

При передаче непрерывных функций с помощью импульсных сигналов приходится сталкиваться с тем, что функция может принимать какие угодно мгновенные значения, в частности, такие, которые могут выражаться иррациональными или трансцендентными

числами с бесконечным числом значащих цифр. Теоретически (по бесшумному каналу) эти числа можно передать совершенно точно, например, с помощью АИМ. В действительности, при наличии помех восстановление исходного импульса с достаточно высокой точностью (или передача достаточно большого числа значащих цифр) требует слишком высокого отношения сигнала к шуму в канале связи. Поэтому следующий шаг, к которому прибегают для передачи непрерывных функций, состоит в квантовании сообщения. Квантование заключается в том, что из всех значений функции $x(t)$ выбирается некоторое число N дискретных разрешенных уровней $x_1, x_2, \dots, x_N$, отстоящих друг от друга на величину Δx , называемую шагом квантования. Все остальные значения считаются запрещенными. Передаваться могут только разрешенные значения. Если истинное мгновенное значение функции попадает внутрь интервала (x_i, x_{i+1}) , т. е. принимает запрещенное значение, то при передаче по каналу связи оно заменяется ближайшим разрешенным значением, которое отличается от истинного на величину, меньшую половины шага квантования. Эта операция совершенно аналогична округлению чисел и по существу означает, что мы ограничиваемся передачей истинных значений функции с точностью до некоторого определенного числа значащих цифр.

На квантованное значение сигнала в канале связи налагается случайная помеха. Величину шага квантования можно выбрать таким образом, чтобы с заданной вероятностью p эта помеха не превосходила половины шага квантования. Тогда на приемном конце линии связи квантованный уровень сигнала может быть точно восстановлен, ибо в этом случае для искаженного помехой значения ближайшим будет тот же уровень сигнала, который поступает в линию связи. Вероятность ошибки при восстановлении уровня будет равна заданной величине p . Восстановленный сигнал снова можно послать в линию связи и эту процедуру можно повторять многократно, причем всякий раз будет обеспечено точное восстановление

исходного квантованного уровня. Передача квантованных значений вместо истинных равносильна наложению на них некоторой помехи δ_k , которая не превосходит половины шага квантования. Эта помеха называется шумом квантования. Таким образом, квантование не освобождает сигнал от помехи, а только заменяет одну помеху другой. Вместо случайной неконтролируемой помехи мы накладываем на сигнал искусственно созданную помеху — шум квантования. По интенсивности эта помеха не меньше естественной, которую мы стремимся устранить. Но преимущество системы связи с квантованием состоит в том, что в ней помеха контролируется нами и можно предотвратить накопление помех.

Другим важным преимуществом квантования является возможность передачи непрерывных функций сообщения с помощью импульсного кода. Каждое дискретное значение функции выражается определенным действительным числом, т. е. может быть записано в произвольной системе счисления в виде некоторой последовательности цифр. Точно так же соответствующей этому дискретному значению функции электрический сигнал может быть представлен комбинацией отдельных электрических импульсов. Все возможные комбинации импульсов, соответствующие всем возможным значениям функции сообщения, образуют определенный код. Каждая отдельная комбинация называется кодовой комбинацией. Различные элементарные сигналы (импульсы), с помощью которых строится кодовая комбинация, называются элементами кода, число элементов, из которых строится кодовая комбинация — основанием кода.

Телеграфная азбука Морзе является примером трюичного кода. Его элементы: короткая посылка — «точка», вдвое более длинный импульс — «тире» и отсутствие импульса — пауза, равная по длительности «тире» и представляющая собой разделительный знак между буквами. Число знаков в кодовых комбинациях азбуки Морзе различно — такой код называется неравномерным. Телеграфный код Бодо является пятизначным двоичным кодом; его элементы —

импульс и отсутствие импульса одинаковой длительности. Каждое сообщение (буква алфавита) изображается пятизначной кодовой комбинацией, состоящей из этих элементов (рис. 37).

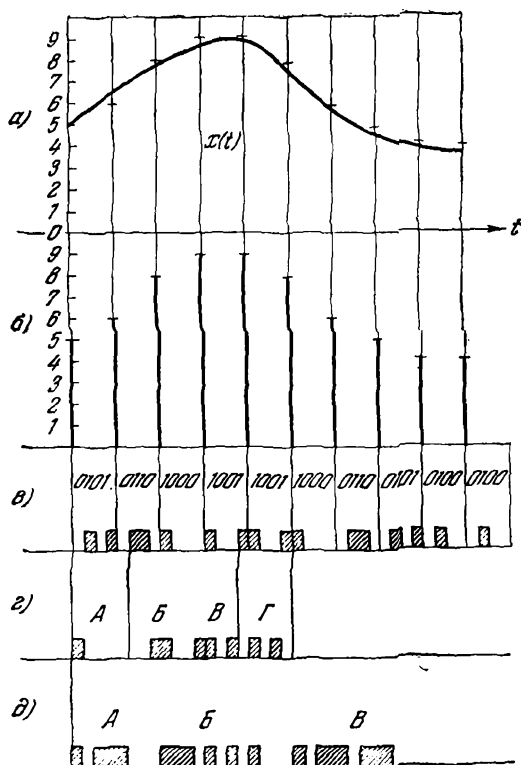


Рис. 37. Передача сообщений с помощью импульсного кода:

а) непрерывная функция сообщения $x(t)$; б) квантованные значения функции; в) передача квантованных значений с помощью двоичного кода; г) телеграфный код Бодо; д) код Морзе.

Если a — основание кода, m — число элементов в кодовой комбинации, то полное число кодовых комбинаций или различных значений, которые можно пе-

передать с помощью такого кода, равно $N = a^m$. Квантование позволяет передать значение функции с помощью кода с конечным числом m элементов в кодовой комбинации. При отсутствии квантования $N = \infty$, и для передачи истинных мгновенных значений функции $x(t_k)$ требуется, вообще говоря, код с любым, сколь угодно большим числом элементов в кодовой комбинации.

Передачу непрерывных функций с помощью импульсного кода, элементами которого являются импульсы, отличающиеся только своей высотой h , называют импульсно-кодовой модуляцией ИКМ*). Для передачи сообщений по системе ИКМ входной сигнал сначала ограничивают по спектру, так чтобы исключить все частоты выше некоторой частоты ν_0 . Затем с него снимают отсчеты со скоростью $2\nu_0$ отсчетов в секунду. Отсчеты квантуются и кодируются. Число знаков m в кодовой комбинации при выбранном основании кода a определяется необходимым числом N квантовых уровней. Так, для телефонной передачи установлено, что хорошее качество звучания достигается при $N = 100$. Следовательно, для передачи телефонного разговора с помощью ИКМ можно использовать семизначный двоичный код ($2^7 = 128$). Кодовые группы поступают в линию связи. На приемном конце линии связи импульсы, искаженные помехами, восстанавливаются, кодовые группы декодируются, и по ним формируется новая последовательность импульсов с амплитудой, пропорциональной исходным квантованным значениям $x_1, x_2, \dots, x_N$. Эти импульсы, следующие со скоростью $2\nu_0$ импульсов в секунду, пропускаются через фильтр нижних частот с граничной частотой ν_0 и, складываясь, образуют первоначальный сигнал.

Определим пороговую мощность ИКМ. (Под пороговой мощностью в данном случае мы будем

*) Обычно под ИКМ понимается передача сообщений с помощью двоичного импульсного кода. Этот способ широко применяется на практике. Однако теоретически мы можем рассматривать ИКМ при любом основании кода.

понимать не $P_{\min}$, а пороговое значение средней мощности.) Пусть применяется код с произвольным основанием a и пусть Δh — разность высот импульсов, соответствующих двум соседним элементам этого кода. Мощность i -го импульса равна $P_i = x_i^2 = (i \Delta h)^2$, а средняя мощность сигнала при условии, что все импульсы встречаются одинаково часто, составляет $P_c = \frac{1}{a} \sum P_i$. Эта мощность будет минимальна, если для образования кода применяются как положительные, так и отрицательные импульсы. В этом случае

$$P_c = \frac{(\Delta h)^2}{a} \sum_{i=-\frac{a-1}{2}}^{i=\frac{a-1}{2}} i^2 = \frac{(\Delta h)^2}{12} (a^2 - 1). \quad (3.13)$$

Для того чтобы импульсы, искаженные помехой, могли быть правильно восстановлены, необходимо, чтобы случайная помеха ξ не превосходила половины величины Δh ($|\xi| < \frac{1}{2} \Delta h$). Вероятность такого события, как мы видели выше, зависит от отношения $\frac{\Delta h}{\sigma}$.

При $\Delta h = 10\sigma$ вероятность ошибки (т. е. неправильного восстановления импульса) равна 10^{-6} . Подставляя это значение Δh в (3.13), получим пороговое значение мощности $P_c^0(a)$ для системы ИКМ с кодом при основании a

$$P_c^0(a) = \frac{100}{12} \sigma^2 (a^2 - 1) = \frac{100}{12} (a^2 - 1) P_{\text{ш}}. \quad (3.14)$$

При заданной мощности шума пороговая мощность сигнала возрастает вместе с основанием кода. Максимальная пороговая мощность имеет место при $a = N$, т. е. в случае обычной АИМ. Пороговая мощность при АИМ равна

$$P_{\text{АИМ}} = \frac{100}{12} (N^2 - 1) P_{\text{ш}} \approx N^2 \frac{100}{12} P_{\text{ш}}. \quad (3.15)$$

Из этого следует, что она возрастает пропорционально N^2 . При любом другом основании кода $a \neq N$ поро-

говая мощность не зависит от числа уровней квантования N , а определяется только основанием кода. При заданном a для кодирования квантовых уровней требуется m импульсов в каждой кодовой группе (согласно соотношению $N=a^m$). Уменьшая основание кода a , мы должны соответственно увеличить m , т. е. увеличить число импульсов, проходящих через линию связи в единицу времени. Следовательно, ИКМ позволяет уменьшить пороговую мощность сигнала за счет расширения полосы частот используемой линии связи. Минимальная пороговая мощность достигается при использовании двоичного кода. В этом случае $P_c^0(2) = 25P_{ш}$.

Пропускная способность канала связи. Теперь мы подошли вплотную к определению пропускной способности канала связи. Выше, на стр. 129, отмечалось, что пропускная способность пропорциональна полосе частот используемой линии связи. Помимо полосы существенную роль играет также отношение сигнала к шуму.

Предположим, что нам требуется передать сообщение, представляющее собой таблицу трехзначных десятичных чисел. Пусть в нашем распоряжении имеется канал с полосой 3 кГц и отношением сигнала к шуму $\frac{P_c}{P_{ш}} = 25$. Пользуясь формулой (3.14), получаем, что при таком отношении сигнала к шуму основание кода $a=2$, а из соотношения $N=10^3=a^m$ находим $m=10$, т. е. для передачи сообщения по такому каналу мы можем воспользоваться десятизначным двоичным кодом. Через канал с полосой 3 кГц можно пропустить 6000 импульсов в секунду или 600 кодовых групп по 10 двоичных импульсов в каждой группе. Количество информации, связанное с передачей каждой кодовой группы, $Q_1 = 3 \log_2 10 = 10 \log_2 2 = 10$ бит. Следовательно, пропускная способность составляет 6000 бит в секунду. Пусть теперь мы увеличили мощность передатчика в пять раз, так что отношение сигнала к шуму стало $\frac{P_c}{P_{ш}} = 125$. Если мы

по-прежнему будем передавать сообщение с помощью двоичного кода, то пропускная способность канала при заданной полосе 3 кгц , конечно, не изменится. Однако применяя двоичный код при таком высоком отношении сигнала к шуму, мы неэкономно используем предоставленный канал связи. Можно увеличить его пропускную способность, осуществив перекодировку сообщений. Из (3.14) находим, что при $\frac{P_c}{P_{\text{ш}}} = 125$ можно положить $a=4$. Далее, из условия $N=10^3=a^m$ получаем $m=5$. Итак, мы можем воспользоваться пятизначным четверичным кодом. Пропуская по-прежнему 6000 импульсов в секунду через канал связи, мы можем передать 1200 кодовых групп по пять четверичных импульсов в каждой группе. Количество информации, связанное с одной кодовой группой, будет по-прежнему составлять 10 бит ($3 \log_2 10 = 5 \log_2 4 = 10$) и пропускная способность составит $10 \times 1200 = 12\,000 \text{ бит}$ в секунду. При $\frac{P_c}{P_{\text{ш}}} = 825$ мы можем воспользоваться трехзначным десятичным кодом и пропускная способность составит $10 \times 2000 = 20\,000 \text{ бит/сек}$. Наконец, при отношении сигнала к шуму, равном $8 \cdot 10^6$, можно положить $a=10^3=N$, $m=1$, т. е. использовать для передачи обычную амплитудно-импульсную модуляцию АИМ (без кодировки). Каждый импульс, изображающий трехзначное десятичное число, будет содержать 10 бит информации. Через канал 3 кгц можно пропустить 6000 таких импульсов и, следовательно, пропускная способность канала связи будет составлять $6 \cdot 10^4 \text{ бит}$ в секунду. То же самое количество информации можно передать по каналу связи с отношением сигнала к шуму $\frac{P_c}{P_{\text{ш}}} = 25$, используя двоичный код и увеличив ширину полосы частот линии с 3 до 30 кгц.

Теперь совершенно ясно, какое значение имеет каждая из величин, от которых зависит пропускная способность канала связи. Полоса частот определяет количество импульсов, которое можно пропустить через канал в единицу времени. А отношение сигнала

к шуму определяет основание кода, которым можно пользоваться для передачи через данный канал связи, и, следовательно, информационное содержание каждого импульса. Так, при передаче двоичным кодом каждый импульс переносит 1 *бит* информации, при использовании троичного кода каждый импульс переносит 1,6 *бит*, при четверичном коде — 2 *бита*, десятичном — 3,3 *бита* и т. д. Уменьшая основание кода, мы понижаем пороговую мощность системы и одновременно снижаем количество информации, которое переносится каждым импульсом, так что за экономию мощности при неизменной пропускной способности канала приходится расплачиваться расширением полосы частот.

Определим выражение пропускной способности канала связи при ИКМ. Пусть линия связи имеет полосу частот Δf . Тогда по ней можно пропустить $2\Delta f = nm$ импульсов в секунду, где n — число кодовых групп, пропускаемых ежесекундно через канал связи, а m — число импульсов в кодовой группе. Информация Q_1 , связанная с передачей каждой кодовой группы, равна $Q_1 = m \log_2 a$, а общее количество информации, прошедшее через канал за 1 сек, будет

$$q = nQ_1 = nm \log_2 a = 2\Delta f \log_2 a = \Delta f \log_2 a^2. \quad (3.16)$$

Подставляя в это выражение значение a^2 из (3.14), получим

$$q = \Delta f \log_2 \left[1 + \frac{12}{100} \frac{P_c^0(a)}{P_{\text{ш}}} \right]. \quad (3.17)$$

Это и будет максимальная пропускная способность системы ИКМ. Если основание кода a выбрано так, что величина $P_c^0(a)$ равна средней мощности сигнала в линии связи, то $P_c^0(a)$ в формуле (3.17) можно заменить просто на P_c . При любом другом основании кода b ($2 \leq b \leq a$):

$$2\Delta f \leq q \leq \Delta f \log_2 \left(1 + \frac{12}{100} \frac{P_c}{P_{\text{ш}}} \right). \quad (3.18)$$

Для характеристики системы связи полезно вычислить величину $\frac{q}{\Delta f}$, которая характеризует пропускную способность, приходящуюся на 1 гц. Соответствующие значения для ИКМ приведены в табл. 3.1.

Таблица 3.1

Пропускная способность ИКМ на единицу полосы частот

Основание кода, a	Пороговая мощность, $P_c/P_{ш}$	Число бит на один импульс	Число двоичных единиц на 1 гц
			$\frac{q}{\Delta f} = \log_2 \left(1 + \frac{12}{100} \frac{P_c}{P_{ш}} \right)$
2	25	1,0	2,0
3	67	1,6	3,2
4	125	2,0	4,0
5	200	2,3	4,6
6	292	2,6	5,2
7	400	2,8	5,6
8	525	3,0	6,0
9	666	3,2	6,4
10	825	3,3	6,6

Система кодирования ИКМ не является оптимальной. Поэтому пропускная способность, которая дается формулой (3.17), не исчерпывает всех возможностей данной системы связи. Как показал Шеннон [5], существует некоторая, вообще говоря, достаточно сложная система кодирования, которая позволяет осуществить передачу информации со скоростью

$$q = \Delta f \log_2 \left(1 + \frac{P_c}{P_{ш}} \right). \quad (3.19)$$

Такая система кодирования называется идеальной.

Формула Шеннона (3.19) описывает максимальную пропускную способность, которую можно реализовать в канале с заданной полосой Δf и заданным отношением сигнала к шуму $\frac{P_c}{P_{ш}}$. Никакая система связи, как бы сложна и совершенна она ни была, не

позволяет передавать информацию с большей скоростью при тех же значениях Δf и $\frac{P_c}{P_{ш}}$. Таким образом, формула Шеннона устанавливает предельные соотношения между основными параметрами системы связи, соотношения, справедливые для любой системы связи, в том числе, конечно, и для связи с внеземными цивилизациями.

На рис. 38 приводится зависимость скорости передачи информации $\frac{q}{\Delta f}$, рассчитанной на 1 гц, от отношения сигнала к шуму в канале связи для идеальной системы Шеннона и для ИКМ. Как можно видеть, идеальная система обеспечивает выигрыш в мощности по сравнению с ИКМ приблизительно на 8—10 дб. Далее, ИКМ имеет вполне определенную пороговую

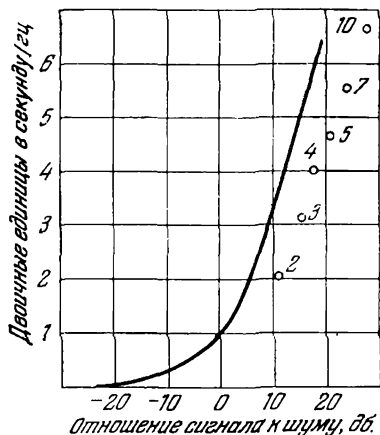


Рис. 38. Пропускная способность системы связи.

Сплошная кривая относится к идеальной системе Шеннона. Точки относятся к ИКМ с положительными и отрицательными импульсами при частоте ошибок 10^{-5} ; цифрами около точек указано основание кода.

мощность $P_c^0(a)$. Эта мощность определяется принятой вероятностью ошибки. С изменением принятой вероятности изменяется пороговая мощность и определяемый ею численный коэффициент при $\frac{P_c}{P_{ш}}$ в формуле (3.17). Для $P_c < P_c^0(a)$ передача информации с заданной надежностью (заданной частотой ошибок) невозможна. Идеальная система не имеет определенной пороговой мощности. Она может работать при любом значении P_c , обеспечивая скорость передачи информации в соответствии с выражением (3.19) со сколь угодно малой частотой ошибок. В частности,

при $\frac{P_c}{P_{\text{ш}}} = 3$ величина $q = 2 \Delta f$, т. е. идеальная система связи имеет пропускную способность, равную пропускной способности системы с двоичной ИКМ (при пороговом отношении сигнала к шуму $\frac{P_c}{P_{\text{ш}}} = 25$). При $\frac{P_c}{P_{\text{ш}}} = 1$ пропускная способность идеальной системы Шеннона $q = \Delta f$, затем она быстро убывает и при $\frac{P_c}{P_{\text{ш}}} = 0$ пропускная способность канала связи также равна нулю. Наконец, при $P_{\text{ш}} \rightarrow 0$ $\frac{P_c}{P_{\text{ш}}} \rightarrow \infty$ и q также стремится к бесконечности, т. е. скорость передачи информации по бесшумному каналу может быть сколь угодно велика. Это справедливо также и для ИКМ. Практически такую возможность можно реализовать (в случае ИКМ), применив код со сколь угодно большим основанием a . Действительно, любой текст можно представить в виде числа, содержащего достаточно много значащих цифр. По бесшумному каналу такое число может быть передано в виде импульса соответствующей высоты.

Рассмотрим зависимость максимальной пропускной способности канала связи от полосы частот. Величина $P_{\text{ш}}$, входящая в формулу Шеннона (3.19), зависит от полосы частот. В большинстве практических случаев можно положить

$$P_{\text{ш}} = P_{\text{ш. уд}} \Delta f. \quad (3.20)$$

Величина $P_{\text{ш. уд}}$, представляющая собой мощность шума, приходящуюся на единичный интервал частот, называется удельной мощностью шума. Подставляя значение $P_{\text{ш}}$ из (3.20) в формулу Шеннона, получим

$$q = \Delta f \log_2 \left(1 + \frac{P_c}{P_{\text{ш. уд}} \Delta f} \right). \quad (3.21)$$

Если положить $\Delta f_0 = \frac{P_c}{P_{\text{ш. уд}}}$, т. е. определить Δf_0 как полосу, при которой мощность шума равна мощности сигнала, то выражение (3.21) принимает вид

$$\frac{q}{\Delta f_0} = \frac{\Delta f}{\Delta f_0} \log_2 \left(1 + \frac{\Delta f_0}{\Delta f} \right). \quad (3.22)$$

На рис. 39 величина $\frac{q}{\Delta f_0}$ построена в функции $\frac{\Delta f}{\Delta f_0}$.

При увеличении полосы пропускная способность быстро растет, пока мощность сигнала не сравняется с мощностью шума (при $\Delta f = \Delta f_0$). После этого

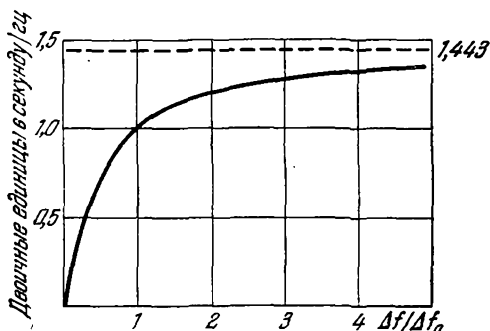


Рис. 39. Пропускная способность канала связи как функция полосы частот

Δf_0 — полоса, для которой $P_{\text{ш}} = P_{\text{с}}$.

рост пропускной способности замедляется и при $\Delta f \rightarrow \infty$ асимптотически стремится к величине, равной пропускной способности при $\Delta f = \Delta f_0$, умноженной на $\log_2 e = 1,443$.

§ 2. Дальность и информативность межзвездной радиосвязи

Оптимальный диапазон. Выше мы рассмотрели некоторые положения общей теории связи. Перейдем теперь к обсуждению проблемы связи с внеземными цивилизациями. Трудность организации связи с ВЦ состоит в том, что отдельные элементы системы связи принадлежат различным абонентам и об их устройстве заранее ничего не известно. Это требует от каждого из абонентов, как на передающем, так и на приемном конце линии связи, обеспечения таких способов передачи и приема сигналов, которые позволили бы установить надежную радиосвязь с учетом этой

неопределенности. В такой общей постановке эта задача включает в себя и проблему кодировки сообщений, и проблему позывных, и проблему обнаружения (в том числе критерии искусственного происхождения сигнала) и проблему дешифровки сигналов. Некоторые из этих проблем рассматриваются в других главах настоящей книги.

Схема межзвездной радиосвязи изображена на рис. 40. Сообщение отправителя ВЦ-1 поступает в передающее устройство, где оно преобразуется в сигнал,

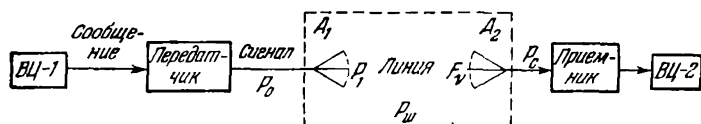


Рис. 40. Схема межзвездной радиосвязи.

A_1 — передающая антенна, A_2 — приемная антенна, P_0 — мощность передатчика, P_1 — мощность, излучаемая антенной, F_ν — спектральная плотность потока в точке наблюдения, P_c — мощность сигнала, поступающая на вход приемника.

и с помощью передающей антенны A_1 излучается в космическое пространство. На приемном конце линии связи радионизлучение улавливается приемной антенной A_2 , электрический сигнал направляется в приемное устройство, где он подвергается различным преобразованиям, в результате которых по полученному сигналу восстанавливается исходное сообщение.

Общим элементом системы связи для обоих абонентов является линия. В системе межзвездной радиосвязи линия представляет собой область космического пространства между передающей и приемной антеннами (межзвездная среда и планетные атмосферы), в которой распространяются радиоволны. Параметры линии определяются как сознательными действиями абонентов, так и некоторыми объективными факторами, не зависящими от их воли. Одним из таких факторов является поглощение радиоволн при распространении в межзвездной среде. Как указывалось в гл. II, коэффициент поглощения межзвездной среды

возрастает с убыванием частоты. При больших расстояниях (порядка размеров Галактики) межзвездная среда становится практически непрозрачной для метровых радиоволн. Это ограничивает диапазон волн для межзвездной связи: из-за сильного поглощения межзвездная связь на частотах ниже 1 Мгц становится невозможной.

Еще одним важным фактором являются шумы, действующие в линии связи. Их можно разделить на две группы: шумы, связанные с аппаратурой, и шумы фона. Шумы, связанные с аппаратурой, поддаются регулировке и могут быть сделаны сколь угодно малыми. Шумы фона определяются радиоизлучением атмосфер планет и космическим радиоизлучением. Атмосферные шумы в принципе можно преодолеть, вынося антенны с поверхности планеты в космическое пространство, например, на орбиты искусственных спутников планеты или звезды. Шумы, связанные с космическим радиоизлучением, преодолеть невозможно.

Еще одним источником принципиально неустраняемого шума являются квантовые флуктуации *), связанные с квантовой природой электромагнитного излучения.

Шумы фона и квантовые флуктуации определяют оптимальный диапазон электромагнитных волн для межзвездной связи. Этот вопрос подробно анализировался в гл. I. Там было показано, что оптимальный диапазон при поиске позывных ВЦ определяется областью, где интенсивность фона космического излучения минимальна ($\lambda \approx 10 \div 50 \text{ см}$), а в случае приема информативной передачи — областью, где минимальна яркостная температура фона. Последнее условие выполняется для очень широкой области спектра от дециметрового диапазона до субмиллиметрового.

*) Не путать с шумами квантования (§ 1).

Вопрос о выборе рабочей полосы частот внутри оптимального диапазона требует специального обсуждения. Этот вопрос также анализировался в гл. I, где было получено выражение для оптимального распределения энергии передатчика по спектру, которое требуется для обеспечения максимальной скорости передачи информации. При умеренном количестве информации скорость передачи не имеет существенного значения и поэтому полоса частот может быть достаточно узкой. В этом случае возникает проблема поисков сигнала по частоте. Хорошо известно предложение Коккони и Моррисона [1] использовать частоту радиолинии водорода $\lambda 21$ см ($\nu = 1420$ Мгц) или ее гармоник. Аналогичным образом можно использовать частоту радиолинии гидроксила OH $\lambda 18$ см. В. С. Троцкий [6] предложил проводить поиски около линий радиоизлучения отдельных молекул, применяемых в молекулярных генераторах и усилителях (линия аммиака на волне 1,25 см и линия формальдегида на волне 0,4 см).

Дальность связи. Важным параметром линии связи является ее протяженность. Так как в первом приближении можно полагать, что цивилизации равномерно распределены в пространстве, то число возможных абонентов и, следовательно, вероятность установления связи возрастают пропорционально третьей степени дальности связи. От чего зависит дальность межзвездной радиосвязи? Прежде всего следует уточнить, что понимается в данном случае под дальностью связи. Нас будут интересовать две задачи: обнаружение сигналов ВЦ и прием содержащейся в них информации. В соответствии с этим мы будем говорить о дальности обнаружения и о дальности приема информативной передачи или просто о дальности связи. Разумеется, прежде чем принимать информацию, необходимо обнаружить сигналы ВЦ. Однако методически более удобно вначале получить выражение для дальности связи, а потом перейти к дальности обнаружения.

Дальность связи определяется тем предельным расстоянием, при котором система связи позволяет осуществить передачу и прием информации с заданной надежностью (заданной частотой ошибок). При большем расстоянии мощность сигнала падает ниже порогового значения и восстановление исходного сообщения с заданной надежностью становится невозможным.

Определим выражение для дальности связи. Пусть P_0 — мощность передатчика ВЦ-1, Δf_1 — его полоса

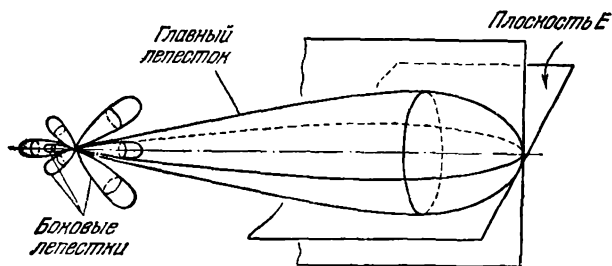


Рис. 41. Диаграмма направленности антенны.

частот, η — коэффициент полезного действия передающей антенны. Тогда излучаемая ею мощность $P_1 = \eta P_0$. Если эта мощность излучается изотропно, независимо от направления, то поток радионизлучения в точке наблюдения на расстоянии R будет

$$F_\nu \Delta f_1 = \frac{P_1}{4\pi R^2}. \quad (3.23)$$

Здесь F_ν — спектральная плотность потока энергии, или поток энергии, приходящийся на единицу полосы частот.

Реальные антенны обладают определенной направленностью. Для характеристики направленных свойств антенны используется диаграмма направленности. Диаграмма направленности передающей антенны — это кривая, которая характеризует изменение потока энергии, излучаемой антенной, в зависимости от направления. На рис. 41 изображена диаграмма

направленности зеркальной антенны. Практически вся энергия излучается такой антенной в пределах некоторого малого телесного угла внутри главного лепестка диаграммы направленности. Если антенна имеет прямоугольное сечение со сторонами l и h , то угловой размер главного лепестка в этих направлениях равен

$$2\theta_l = 2 \frac{\lambda}{l} \quad \text{и} \quad 2\theta_h = 2 \frac{\lambda}{h}, \quad (3.24)$$

где λ — длина волны радионизлучения. Для зеркальной антенны с круговым сечением (например, для параболоида вращения) ширина главного лепестка

$$2\theta_0 = 2 \times 1,22 \frac{\lambda}{D} = 1,22 \frac{\lambda}{r}, \quad (3.25)$$

где r — радиус, D — диаметр зеркала. Эту величину обычно называют шириной диаграммы по уровню нулевой мощности. Наряду с ней часто используют ширину диаграммы по уровню половинной мощности. Для антенны с круговым сечением ширина диаграммы по уровню половинной мощности равна

$$2\theta_{0,5} = 2 \times 0,51 \frac{\lambda}{D} \approx \frac{\lambda}{D}. \quad (3.26)$$

В первом приближении можно положить, что в пределах этого угла диаграмма постоянна и равна своему максимальному значению, а вне его равна нулю.

Для расчета излучаемой мощности полезно использовать коэффициент направленного действия (КНД). КНД передающей антенны равен отношению мощности, излучаемой антенной в данном направлении (например, в направлении оси антенны) в расчете на единичный телесный угол, к среднему по всем направлениям значению мощности, излучаемой антенной в единичном телесном угле. Другими словами, КНД равен отношению потока энергии, излучаемой антенной внутри малого телесного угла $d\omega$, к потоку энергии, излучаемой изотропным излучателем той же мощности в том же телесном угле $d\omega$. При использовании направленной антенны с коэффициентом на-

правленного действия g_1 поток радиоизлучения в точке наблюдения на расстоянии R будет

$$F_v \Delta f_1 = \frac{P_1 g_1}{4\pi R^2} = \frac{P_0 \eta g_1}{4\pi R^2} = \frac{P_0 \epsilon_1}{4\pi R^2}. \quad (3.27)$$

Величина $\epsilon_1 = \eta g_1$ называется коэффициентом усиления антенны. Если задана мощность передатчика P_0 , то, зная коэффициент усиления передающей антенны, легко определить поток радиоизлучения в любой произвольно взятой точке. В дальнейшем для упрощения формул мы будем полагать $\eta = 1$, $P_1 = P_0$, $\epsilon_1 = g_1$.

Приемные антенны, так же как и передающие, обладают определенной направленностью. В теории антенн доказывается, что в силу принципа взаимности направленные свойства антенны при приеме и передаче одинаковы. В частности, диаграмма направленности, КНД и коэффициент усиления приемной антенны совпадают с диаграммой направленности, КНД и коэффициентом усиления той же антенны, работающей в режиме передачи (когда на клеммы приемника присоединен передатчик).

Очевидно, мощность P , отдаваемая антенной в приемник, пропорциональна потоку радиоизлучения в точке приема. Поэтому можно записать

$$P_c = S F_v \Delta f. \quad (3.28)$$

Величина S , имеющая размерность см^2 , называется эффективной площадью приемной антенны. По своему значению она совершенно аналогична входному отверстию оптического телескопа. В частности, для зеркальной антенны с коэффициентом полезного действия $\eta = 1$ эффективная площадь совпадает с геометрической площадью зеркала. Между эффективной площадью антенны и ее коэффициентом усиления существует соотношение

$$S = \frac{\lambda^2}{4\pi} \epsilon. \quad (3.29)$$

Теперь мы можем получить выражение для дальности связи в зависимости от рассмотренных выше параметров передающей и приемной систем. Мощность сигнала P_c , поступающего в приемник,

существенно зависит от соотношения между полосой приемника и полосой передатчика. Здесь имеются две возможности.

а) Полоса приемника больше полосы передатчика ($\Delta f_2 > \Delta f_1$).

Этот случай может иметь место, например, при приеме узкополосных монохроматических сигналов. Применяя формулы (3.27) и (3.28) и обозначая величины, относящиеся к передающей системе, индексом 1, а величины, относящиеся к приемной системе, индексом 2, получим

$$P_c = S_2 F_v \Delta f_1 = S_2 \frac{P_1 g_1}{4\pi R^2}. \quad (3.30)$$

Отметим, что эта величина не зависит от полосы приемника Δf_2 и при заданной мощности передатчика P_1 не зависит также от его полосы частот. Что касается мощности шума, поступающего на вход приемника, то она, как известно, пропорциональна его полосе частот:

$$P_{ш} = P_{ш. уд} \Delta f_2 = k T_{ш} \Delta f_2. \quad (3.31)$$

Здесь k — постоянная Больцмана, равная $1,38 \cdot 10^{-16}$ эрг/град, $T_{ш}$ — шумовая температура, которая обычно вводится для характеристики мощности шума. Она равна температуре активной нагрузки (сопротивления), согласованной со входом приемника, которая, будучи включена вместо антенны, создавала бы ту же шумовую мощность. Когда речь идет о шумах фона, то $T_{ш}$ представляет собой эквивалентную яркостную температуру этого шумового излучения. В частности, если фон обусловлен тепловым радиоизлучением некоторой среды, то $T_{ш}$ совпадает с температурой этой среды.

Две последние формулы позволяют определить отношение сигнала к шуму на входе приемника:

$$\alpha = \frac{P_c}{P_{ш}} = \frac{P_1 g_1 S_2}{4\pi R^2 k T_{ш} \Delta f_2}. \quad (3.32)$$

В § 1 мы видели, что эта величина определяет надежность связи. Для надежной радиосвязи необходимо, чтобы α превышало некоторое пороговое значение,

определяемое способом кодировки. В обычных системах связи $\alpha > 1$. Формула (3.32) показывает, что надежность космической радиосвязи пропорциональна мощности передатчика, умноженной на коэффициент усиления передающей антенны, эффективной площади приемной антенны, и обратно пропорциональна шумовой температуре, полосе приемника и квадрату расстояния между цивилизациями.

При заданном α расстояние R , на котором можно получить требуемое отношение сигнала к шуму, определится из (3.32):

$$R = \left(\frac{P_1 g_1 S_2}{4\pi \alpha k T_{\text{ш}} \Delta f_2} \right)^{1/2}, \quad (3.33a)$$

или, используя (3.29),

$$R = \left(\frac{P_1 S_1 S_2}{\alpha k T_{\text{ш}} \Delta f_2 \lambda^2} \right)^{1/2}, \quad (3.33б)$$

$$R = \left(\frac{P_1 g_1 g_2 \lambda^2}{16\pi^2 \alpha k T_{\text{ш}} \Delta f_2} \right)^{1/2}, \quad (3.33в)$$

т. е. дальность радиосвязи возрастает с увеличением мощности передатчика, направленности или эффективной площади приемной и передающей антенн, с уменьшением шумовой температуры и сужением полосы приемника. Зависимость от λ в формуле (3.33б) связана с тем, что при данной площади S_1 передающей антенны направленность возрастает с убыванием длины волны, а зависимость от λ в (3.33в) является следствием того, что при заданной направленности g_2 приемной антенны ее площадь возрастает с длиной волны.

б) Рассмотрим теперь второй случай: полоса передатчика больше полосы приемника ($\Delta f_1 > \Delta f_2$). Этот случай имеет место при приеме широкополосных сигналов, например, когда распределение энергии передатчика определяется требованием максимальной информативности передачи (см. гл. I).

В этом случае спектр сигнала ограничен полосой приемника и его мощность P_c равна

$$P_c = S_2 F_v \Delta f_2 = \frac{P_1 g_1 S_2 \Delta f_2}{4\pi R^2 \Delta f_1}, \quad (3.34)$$

т. е. в отличие от первого случая, мощность сигнала пропорциональна полосе приемника Δf_2 и при заданной полной мощности передатчика обратно пропорциональна его полосе частот. Мощность шума по-прежнему выражается формулой (3.31), так что отношение сигнала к шуму (при заданной дальности) и дальность связи (при заданном отношении сигнала к шуму) будут равны соответственно

$$\frac{P_c}{P_{ш}} = \frac{p_1 g_1 S_2}{4\pi R^2 \Delta f_1 k T_{ш}}, \quad (3.35)$$

$$R = \left(\frac{P_1 g_1 S_2}{4\pi \alpha \Delta f_1 k T_{ш}} \right)^{1/2}. \quad (3.36)$$

Сравнивая эти выражения с выражениями (3.32) и (3.33), мы видим, что они отличаются только индексами при Δf . В первом случае отношение сигнала к шуму и дальность связи возрастают с уменьшением полосы приемника и при заданной мощности передатчика не зависят от его полосы частот. Во втором случае, напротив, отношение сигнал/шум и определяемая им дальность связи возрастают с уменьшением полосы передатчика и не зависят от полосы приемника. Таким образом, в общем случае можно написать

$$\alpha \propto \Delta f^{-1}, \quad R \propto \Delta f^{-1/2}, \quad (3.37)$$

где Δf — максимальная из величин Δf_1 и Δf_2 .

Рассмотрим зависимость дальности от параметров приемной и передающей антенн. Эта зависимость описывается выражением (3.33), где вместо величины Δf_2 следует подставить $\Delta f = \max(\Delta f_1, \Delta f_2)$. Полагая в (3.33в) $g_1 = g_2 = 1$, получим значение дальности в случае изотропной передачи и ненаправленного приема. Полагая $g_2 = 1$, получим значение дальности при направленной передаче и ненаправленном приеме. Наконец, полагая в (3.33а) $g_1 = 1$, получим значение

дальности при изотропной передаче и приеме на направленную антенну с эффективной площадью S_2 .

Остановимся на зависимости α и R от полосы частот. Формулы (3.32) и (3.33) удобно представить в логарифмическом виде:

$$\lg R = \frac{1}{2} \lg P_1 g_1 + \frac{1}{2} \lg \frac{S_2}{4\pi a k T_{\text{ш}}} - \frac{1}{2} \lg \Delta f. \quad (3.38)$$

Фактор усиления (ослабления) обычно выражается в децибелах. Аналогичным образом произведение $P_1 g_1$ можно выразить в децибелах на ватт. Пусть, например, $P_1 g_1 = 100 \text{ дб} \cdot \text{вт}$. Это значит, что передатчик мощностью 1 вт работает на антенну с коэффициентом усиления 100 дб или передатчик мощностью 1 кВт работает на антенну с усилением 70 дб, или, наконец, передатчик мощностью 1 Мвт работает на антенну с усилением 40 дб и т. д. Если в формуле (3.38) произведение $P_1 g_1$ выразить в дб·вт, R — в световых годах, S_2 в квадратных метрах, то эта формула принимает вид

$$\lg R = \frac{P_1 g_1}{20} + \frac{1}{2} \lg \frac{S_2}{4\pi a k T_{\text{ш}}} - \frac{1}{2} \lg \Delta f - 16. \quad (3.39)$$

Пусть, например, $P_1 g_1 = 200 \text{ дб} \cdot \text{вт}$, $S_2 = 10^4 \text{ м}^2$, $T_{\text{ш}} = 10^\circ \text{ К}$; тогда

$$\lg R = 6,4 - \frac{1}{2} \lg \alpha - \frac{1}{2} \lg \Delta f. \quad (3.40)$$

Аналогично, при тех же параметрах приемной и передающей систем, получим

$$\lg \alpha = 12,8 - 2 \lg R - \lg \Delta f. \quad (3.41)$$

На рис. 42 приведена зависимость α и R от полосы частот. Очевидно, что при $\Delta f_2 > \Delta f_1$ выгодно уменьшать полосу приемника. При этом мощность шума, поступающего на вход приемника, будет уменьшаться, а полезная мощность останется без изменения. В результате отношение сигнала к шуму на заданном расстоянии R , или дальность связи при заданном отношении сигнала к шуму, будут расти. Этот рост не сопровождается потерей информативности, так как полоса частот линии связи и определяемая ею

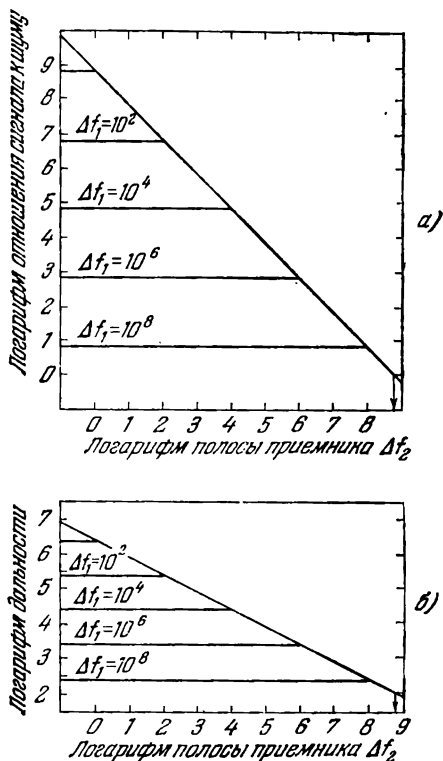


Рис. 42. а) Зависимость между отношением α сигнала к шуму и полосой приемника Δf_2 при различных значениях полосы передатчика Δf_1 .

Приняты следующие параметры системы связи: $R = 100$ световых лет, $P_1 g_1 = 200 \text{ дб} \cdot \text{вт}$, $S_2 = 10^1 \text{ м}^2$, $T_{\text{ш}} = 10^\circ \text{ К}$. Значения полосы выражены в герцах. При заданной полосе передатчика Δf_1 отношение сигнала к шуму возрастает с уменьшением Δf_2 до тех пор, пока не будет достигнуто равенство $\Delta f_1 = \Delta f_2$. Дальнейшее сужение полосы приемника не приводит к увеличению α . После того как равенство α достигнуто, рост отношения сигнала к шуму может быть обеспечен только за счет одновременного сужения полосы передатчика и полосы приемника.

б) Зависимость дальности от полосы приемника Δf_2 при различных значениях полосы передатчика Δf_1

($P_1 g_1 = 200 \text{ дб} \cdot \text{вт}$, $S_2 = 10^1 \text{ м}^2$, $T_{\text{ш}} = 10^\circ \text{ К}$, $\alpha = 1$). Значения дальности выражены в световых годах. На обоих рисунках стрелкой отмечена полоса Δf_2 , для которой при выбранных параметрах передающей и приемной систем сигнал равен шуму на расстоянии 100 световых лет.

пропускная способность канала ограничиваются в рассматриваемом случае полосой передатчика Δf_1 . Более того, на заданном расстоянии пропускная способность может быть увеличена в связи с ростом отношения сигнала к шуму. Максимальная дальность достигается при условии $\Delta f_2 = \Delta f_1$. Дальнейшее уменьшение полосы приемника нецелесообразно, так как теперь в одинаковой мере будут уменьшаться и шумы, и полезный сигнал. Вместе с тем сужение полосы приемника приведет теперь к ограничению полосы частот используемой линии связи.

При условии $\Delta f_2 < \Delta f_1$ сигнал, поступая в приемник, ограничивается по спектру и искажается (например, в случае импульсной модуляции ограничение полосы приведет к расплыванию и интерференции импульсов). Кроме того, вследствие сужения полосы уменьшается пропускная способность канала связи. Если на передающем конце линия с полосой Δf_1 используется полностью, то сокращение полосы частот приведет к потере определенного количества информации. Потеря информации в общем случае будет иметь место и при неполном использовании линии, так как характер сигнала, его распределение по времени и частоте заранее не известны. Отсюда вытекает, что рассмотренный нами случай $\Delta f_2 < \Delta f_1$ неблагоприятен для связи. Чтобы в этом случае исключить искажение сигнала и потерю информации, надо увеличить полосу приемника Δf_2 до величины Δf_1 . Этот рост полосы не скажется на дальности связи, так как он приведет к увеличению и полезного сигнала и шума. Однако расширение полосы приемника имеет свои технические ограничения. Например, при современной технике даже для специальных широкополосных приемников трудно получить полосу, превышающую 10% от частоты используемых электромагнитных волн. Другой способ согласования полос состоит в уменьшении полосы передатчика. Это может быть достигнуто двумя способами: а) при сохранении удельной мощности передатчика, т. е. мощности, приходящейся на единицу полосы частот, и б) при сохранении полной мощности передатчика P_1 .

В первом случае полная мощность и поток в точке наблюдения будут уменьшаться пропорционально сокращающейся полосе частот, но спектральная плотность потока F_ν останется без изменения. Вследствие этого будет возрастать доля полного потока или доля мощности передатчика, поступающая в приемник, так что отношение сигнала к шуму и дальность связи не изменятся. Во втором случае сокращение полосы частот будет сопровождаться ростом удельной мощности передатчика и спектральной плотности потока F_ν в точке наблюдения. Полный поток $F_\nu \Delta f_1$ останется при этом без изменения, однако доля его, перехватываемая приемником, возрастет с уменьшением Δf_1 . В результате сокращение полосы частот будет сопровождаться ростом отношения сигнала к шуму и увеличением дальности связи, как это следует из формул (3.35) и (3.36). Очевидно, максимальная дальность по-прежнему будет достигнута при условии $\Delta f_1 = \Delta f_2$. Дальнейшее сокращение полосы частот передатчика нецелесообразно, так как теперь весь поток в точке наблюдения будет целиком поступать в приемник и при заданной мощности передатчика P_1 отношение сигнала к шуму сохранится без изменения. Если же сужение полосы происходит при постоянной удельной мощности передатчика, то ее уменьшение ниже уровня Δf_2 приведет к уменьшению отношения сигнал/шум, а следовательно, и к уменьшению дальности.

Итак, максимальная дальность связи достигается при условии $\Delta f_1 = \Delta f_2$. После того как равенство полос достигнуто (путем сужения полосы приемника, когда $\Delta f_2 > \Delta f_1$, или путем сужения полосы передатчика, когда $\Delta f_1 > \Delta f_2$), дальнейшее увеличение дальности может быть обеспечено только за счет совместного сужения полосы передатчика и полосы приемника. Это сужение полосы приведет, конечно, к потере информативности, причем снижение пропускной способности канала связи на заданном расстоянии R между абонентами будет иметь место, несмотря на рост отношения сигнала к шуму, так как зависи-

мость пропускной способности от полосы более сильная.

Дальность обнаружения. Для обычных систем связи дальность связи ограничена условием $\alpha > 1$ *). Если же ставится более скромная задача обнаружения сигналов без расшифровки содержащейся в них информации, то от этого требования можно отказаться. Современная техника радиометрических измерений позволяет обнаружить сигналы намного более слабые, чем шум. Этим широко пользуются в радиоастрономии, которая имеет дело с измерением чрезвычайно слабых потоков от космических источников.

Возможность обнаружения таких сигналов основана на статистических свойствах шума. Заметим, прежде всего, что если бы мощность шума была строго постоянна, т. е. не испытывала бы никаких флуктуаций со временем, то в этом идеальном случае ее можно было бы легко компенсировать, введя в приемник соответствующее напряжение, равное по величине и противоположное по знаку напряжению шума. Таким путем в принципе можно было бы измерить сигнал сколь угодно малой мощности. По существу, в этом случае дело сводится к регистрации малого приращения, которое испытывает постоянное шумовое напряжение под действием полезного сигнала. Компенсация, строго говоря, не является необходимой, она лишь представляет собой один из удобных способов измерения.

Однако реальный шум представляет собой случайный процесс, в котором амплитуды напряжения (или тока) хаотически меняются со временем около нулевого значения. Если Δf_2 — полоса приемника, то средняя продолжительность отдельного шумового импульса Δt_2 (или время, в течение которого амплитуда затухающих колебаний, порожденных данной флук-

*) Для идеальной системы Шеннона это ограничение несущественно. В некоторых специальных системах связи также применяются методы, позволяющие осуществить прием информации при условии, когда сигнал много меньше шума.

туацией, остается постоянной) будет порядка $\frac{1}{\Delta f_2}$.
За период времени τ_2 будет наблюдаться $n = \frac{\tau_2}{\Delta t_2} =$

$= \tau_2 \Delta f_2$ независимых шумовых импульсов. Регистрирующее устройство с постоянной времени τ_2 осредняет эти импульсы, причем среднее значение

мощности $P_{ш\text{ ср}} = \frac{1}{n} \sum_{i=1}^n P_i$ испытывает флуктуации

(так называемые флуктуации записи) около теоретического значения средней мощности шума $P_{ш}$, представляющего собой математическое ожидание случайных значений мощности P_i отдельных шумовых импульсов. Очевидно, для обнаружения полезного сигнала необходимо, чтобы вызванное им приращение мощности шума $\Delta P_{ш} = P_c$, превышало среднее квадратичное отклонение величины $P_{ш\text{ ср}}$ от теоретического значения $P_{ш}$, т. е. чтобы выполнялось условие $P_c = \Delta P_{ш} > \sigma_{ср}(P)$ или

$$P_c = \beta \sigma_{ср}(P), \quad (3.42)$$

где β — некоторое безразмерное число, большее единицы ($\beta > 1$). Если мы имеем дело с гауссовым шумом, для которого распределение амплитуд подчиняется нормальному закону, то

$$\frac{\sigma_{ср}(P)}{P_{ш}} = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{\tau_2 \Delta f_2}}, \quad (3.43)$$

откуда

$$\alpha = \frac{P_c}{P_{ш}} = \beta \frac{\sigma_{ср}(P)}{P_{ш}} = \frac{\beta}{\sqrt{\tau_2 \Delta f_2}}. \quad (3.44)$$

Это выражение и определяет минимальный сигнал, который может быть зарегистрирован радиометром. При $\beta=1$ получаем предельную или теоретическую чувствительность радиометров. Фактическая чувствительность обычно значительно ниже, так как для уверенной регистрации сигнала необходимо, чтобы β было больше десяти. Величина $\sqrt{\tau_2 \Delta f_2}$ называется радиометрическим выигрышем. При $\sqrt{\tau_2 \Delta f_2} \gg 1$ отно-

шение сигнала к шуму может быть много меньше единицы. Например, при $\tau_2 = 1$ сек и $\Delta f_2 = 100$ Мгц радиометрический выигрыш равен 10^4 ; если $\beta = 10$, то $\frac{P_c}{P_{\text{ш}}} = 10^{-3}$, т. е. сигнал, в тысячу раз более слабый, чем шум, все же в 10 раз превышает среднее квадратичное флуктуации шума и поэтому может быть уверенно обнаружен.

Следует подчеркнуть, что возможность обнаружения слабых по сравнению с шумом сигналов связана с осреднением отдельных шумовых импульсов, которое выполняется регистрирующим устройством. При этом, конечно, происходит и усреднение полезного сигнала, так что мы получаем только среднее за время τ_2 значение мощности сигнала. Если характерное время модуляции τ_1 меньше постоянной времени τ_2 регистрирующего устройства, то все изменения, связанные с модуляцией сигнала, будут сглажены, и информация, содержащаяся в сигнале, будет полностью утеряна. Можно будет лишь констатировать присутствие некоторого полезного сигнала со средней мощностью P_c . Именно в этом смысле мы и говорим о дальности обнаружения, в отличие от дальности связи.

Для приема информации необходимо, чтобы $\tau_1 \geq \tau_2$. Учитывая соотношение между временем и полосой, это условие можно переписать в виде *)

$$\Delta f_2 \geq n \Delta f_1, \quad (3.45)$$

где n — число независимых шумовых импульсов, осредняемых регистрирующим устройством. Выше мы уже отмечали, что для приема информации полоса

*) Здесь мы предполагаем, что $\tau_1 = \frac{1}{\Delta f_1}$, т. е. применяется способ кодирования, который при заданной полосе передатчика Δf_1 обеспечивает максимальную пропускную способность. В общем случае, когда отправитель не полностью расходует полосу частот передатчика (сознательно сокращая пропускную способность канала связи, чтобы обеспечить более высокую надежность на заданном расстоянии или большую дальность при заданной надежности), характерное время модуляции τ_1 может быть больше $\frac{1}{\Delta f_1}$, и условие (3.45) не выполняется.

приемника не должна быть меньше полосы передатчика. Условие (3.45) является более сильным. Оно показывает, что при использовании осреднения простого превышения полосы приемника уже недостаточно; полоса приемника должна превышать полосу передатчика, умноженную на квадрат радиометрического выигрыша.

Чему же равна дальность обнаружения? Подставляя в формулы (3.33) и (3.36) значение α из (3.44), получим следующие выражения для дальности обнаружения:

$$1) \Delta f_2 > \Delta f_1,$$

$$R = \left(\frac{P_1 g_1 S_2}{4\pi\beta k T_{\text{ш}}} \right)^{1/2} \left(\frac{\tau_2}{\Delta f_2} \right)^{1/4}, \quad (3.46)$$

$$2) \Delta f_2 < \Delta f_1,$$

$$R = \left(\frac{P_1 g_1 S_2}{4\pi\beta k T_{\text{ш}} \Delta f_1} \right)^{1/2} (\tau_2 \Delta f_2)^{1/4}. \quad (3.47)$$

В рассмотренном выше примере, когда $\tau_2 = 1$ сек, $\Delta f_2 = 10^8$ гц, дальность обнаружения за счет радиометрического выигрыша может быть увеличена в 100 раз. При условии $\Delta f_2 > \Delta f_1$ дальность медленно убывает с расширением полосы приемника, при условии $\Delta f_2 < \Delta f_1$ она также медленно ($\propto \Delta f_2^{1/4}$) возрастает, так что в этом случае выгодно расширять полосу приемника. Максимальная дальность по-прежнему достигается при условии $\Delta f_2 = \Delta f_1$. При отсутствии радиометрического выигрыша ($\sqrt{\tau_2 \Delta f_2} = 1$) формулы (3.46) и (3.47), как и следовало ожидать, переходят в (3.33) и (3.36).

Пусть задана полоса приемника Δf_2 и пусть имеются два сигнала: узкополосный сигнал с полосой $\Delta f_{1\text{ узк}} < \Delta f_2$ и широкополосный сигнал с полосой $\Delta f_{1\text{ шир}} > \Delta f_2$. Обозначим через R_1 дальность обнаружения узкополосного сигнала, а через R_2 дальность обнаружения широкополосного сигнала. Из приведенных формул следует, что $R_1 \propto \Delta f_2^{-1/4}$, $R_2 \propto \Delta f_2^{1/4} \Delta f_1^{-1/2}$, поэтому

$$\frac{R_1}{R_2} = \left(\frac{\Delta f_{1\text{ шир}}}{\Delta f_2} \right)^{1/2} > 1, \quad (3.48)$$

т. е. при заданной полосе приемника дальность обнаружения узкополосного сигнала больше дальности обнаружения широкополосного сигнала. При этом обе величины, R_1 и R_2 , не являются максимальными. В первом случае дальность может быть увеличена путем сужения полосы приемника до величины $\Delta f_{1 \text{ узк}}$, во втором она может быть увеличена путем расширения полосы приемника до величины $\Delta f_{1 \text{ шир}}$. Тогда будем иметь:

$$R_{1 \text{ max}} = R_1 \left(\frac{\Delta f_2}{\Delta f_{1 \text{ узк}}} \right)^{1/4}, \quad R_{2 \text{ max}} = R_2 \left(\frac{\Delta f_{1 \text{ шир}}}{\Delta f_2} \right)^{1/4},$$

$$\frac{R_{1 \text{ max}}}{R_{2 \text{ max}}} = \left(\frac{\Delta f_{1 \text{ шир}}}{\Delta f_{1 \text{ узк}}} \right)^{1/4}. \quad (3.49)$$

Таким образом, несмотря на то что радиометрический выигрыш возрастает с полосой, максимальная дальность обнаружения узкополосных сигналов все же превышает максимальную дальность обнаружения широкополосных сигналов. Например, при $\Delta f_2 = 10^4 \text{ гц}$, $\Delta f_{1 \text{ узк}} = 1 \text{ гц}$, $\Delta f_{1 \text{ шир}} = 10^{10} \text{ гц}$ получим $R_1 = 10^3 R_2$; $R_{1 \text{ max}} = 10^4 R_2 = 316 R_{2 \text{ max}}$.

Зависимость между дальностью и полосой при использовании осреднения приведена на рис. 43. Здесь, как и выше, предполагается, что $\tau_1 = \frac{1}{\Delta f_1}$, т. е. отправитель стремится обеспечить максимальную пропускную способность при заданной полосе частот передатчика. Осреднение дает выигрыш в дальности при условии $\Delta f_2 > \frac{1}{\tau_2}$. Максимальная дальность достигается при условии $\Delta f_2 = \Delta f_1$. Эти значения полосы лежат правее линии $\Delta f_2 = \frac{1}{\tau_2}$, т. е. в области, где радиометрический выигрыш больше единицы, только при условии $\tau_1 < \tau_2$, когда происходит потеря информации. При $\tau_1 > \tau_2$ значения полосы, соответствующие максимальной дальности, лежат в области, где радиометрический выигрыш уже отсутствует. Таким образом, осреднение позволяет получить

выигрыш в дальности за счет полной потери информативности.

Это замечание справедливо только для максимальной дальности. Оно принимается во внимание при организации оптимального канала связи, когда полосы частот приемника и передатчика можно выбрать

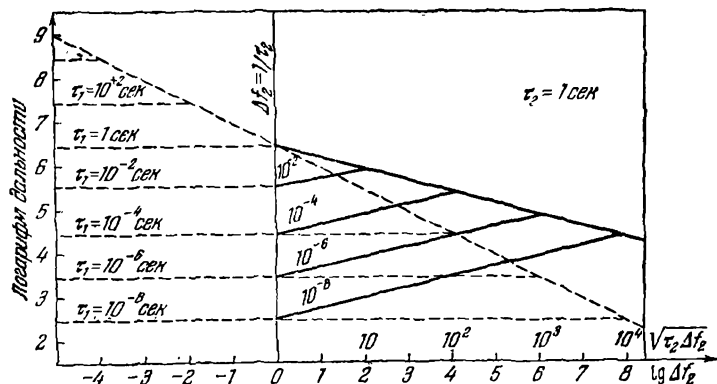


Рис. 43. Зависимость дальности от полосы приемника при использовании осреднения.

Дальность R выражена в световых годах, $P_1 g_1 = 200 \text{ дб} \cdot \text{вт}$, $S_2 = 10^4 \text{ м}^2$, $T_{\text{ш}} = 10^4 \text{ К}$, $\beta = 1$, $\tau_2 = 1 \text{ сек}$. При $\tau_1 < \tau_2$ дальность обнаружения с уменьшением полосы приемника вначале, когда $\Delta f_2 > \Delta f_1$, возрастет, а затем, после достижения максимума при $\Delta f_2 = \Delta f_1$, начинает убывать и при $\Delta f_2 = \frac{1}{\tau_2}$, когда радиометрический выигрыш отсутствует, падение дальности прекращается (как это и должно быть при отсутствии осреднения и при $\Delta f_2 < \Delta f_1$). Для значений $\tau_1 \geq \tau_2$ дальность связи возрастет с уменьшением полосы приемника пропорционально $\Delta f_2^{-1/4}$ вплоть до значения $\Delta f_2 = \frac{1}{\tau_2}$. Если при этом Δf_2 все еще больше Δf_1 , дальнейшее уменьшение полосы приемника сопровождается более быстрым ростом дальности (пропорционально $\Delta f_2^{-1/2}$), который прекращается при условии $\Delta f_2 = \Delta f_1$. Изменение дальности при отсутствии осреднения показано на рисунке пунктирными линиями.

одинаковыми. В проблеме связи с ВЦ дело обстоит иначе: мы не можем выбрать $\Delta f_2 = \Delta f_1$, так как полоса передатчика нам заранее неизвестна. Поэтому если ставится задача не только обнаружить сигналы ВЦ, но и принять содержащуюся в них информацию, по-

лосу приемника надо выбрать так, чтобы она была не меньше предполагаемой полосы передатчика, а постоянную времени желательно выбрать достаточно большой для обеспечения радиометрического выигрыша. Но при этом необходимо, чтобы τ_2 все же было меньше τ_1 , т. е. чтобы выполнялось условие $\frac{1}{\Delta f_2} < \tau_2 < \frac{1}{\Delta f_1}$. Пусть, например, имеются основания предполагать, что полоса передатчика Δf_1 порядка 0,1 гц. Чтобы не потерять информацию, выберем полосу приемника с некоторым запасом: $\Delta f_2 = 1$ гц. Тогда, чтобы обеспечить выигрыш за счет осреднения, постоянная времени τ_2 должна быть больше 1 сек, но не более 10 сек. Выберем $\tau_2 = 4$ сек, — радиометрический выигрыш будет равен 2. В данном случае мы получили выигрыш (правда, незначительный) не в дальности обнаружения, а в дальности связи, т. е. при приеме информации *).

Вопрос о предполагаемой полосе частот передатчика является весьма неопределенным. По этому поводу можно строить только более или менее правдоподобные гипотезы. Ширина полосы может быть оценена на основе соображений о предполагаемой скорости передачи информации. Если скорость передачи информации достаточно мала, то рабочая полоса частот ограничивается снизу только стабильностью передаваемого сигнала. В этом случае полоса может достигать нескольких герц или долей герца, а у молекулярных генераторов даже сотых долей герца.

Дальность приема импульсных сигналов. Одним из путей повышения дальности является применение импульсных сигналов. Если импульсы следуют с

*) При условии $\tau_1 \gg \frac{1}{\Delta f_1}$ (т. е. когда отправитель сам существенно сокращает количество информации, поступающей в канал связи в единицу времени) можно подобрать полосу приемника Δf_2 и постоянную времени τ_2 таким образом, чтобы обеспечить прием информации при достаточно большом радиометрическом выигрыше порядка $\sqrt{\tau_1 \Delta f_1}$, т. е. в этом случае информация при осреднении не теряется.

большими промежутками, то можно реализовать достаточно высокую мощность импульса при сравнительно небольшой средней мощности передатчика. Пусть Δt_1 — продолжительность импульсов, а t_1 — промежуток времени между ними. Отношение мгновенной или так называемой пиковой мощности импульса к средней мощности передатчика равно $\frac{t_1}{\Delta t_1}$. Чтобы импульсы при приеме не были сглажены регистрирующим устройством, постоянная времени τ_2 не должна превышать продолжительность импульса. Если это условие выполняется, мощность сигнала P_c пропорциональна мощности импульса. Каждая посылка длительностью Δt_1 может представлять собой простой или сложный импульс. В случае простых или так называемых видеопульсов (без высокочастотного наполнения) их продолжительность Δt_1 определяет полосу передатчика $\Delta f_1 = \frac{1}{\Delta t_1}$. В этом случае условие $\tau_2 \leq \Delta t_1$ совпадает с (3.45). Если $\tau_2 \geq t_1$, то мощность сигнала P_c пропорциональна средней мощности передатчика. Следовательно, в формулах дальности можно положить

$$P_1 = \begin{cases} P_1, & \text{если } \tau_2 \geq t_1, \\ \frac{P_1 t_1}{\Delta t_1}, & \text{если } \tau_2 \leq \Delta t_1. \end{cases} \quad (3.50)$$

Максимальная дальность связи достигается при условии $\Delta f_1 = \Delta f_2$, а так как при этом, как мы только что видели, $\tau_2 \Delta f_2 = 1$ (для $\tau_2 \leq \Delta t_1$), то

$$R_{\max} = \left(\frac{P_1 g_1 t_1 S_2}{4\pi \alpha \Delta t_1 \Delta f_1 k T_{\text{ш}}} \right)^{1/2}. \quad (3.51)$$

В частности, для простых импульсов, у которых $\Delta t_1 \Delta f_1 = 1$,

$$R_{\max} = \left(\frac{P_1 g_1 t_1 S_2}{4\pi \alpha k T_{\text{ш}}} \right)^{1/2}, \quad (3.52)$$

т. е. максимальная дальность связи при использовании простых импульсов и при заданной средней мощности передатчика не зависит от его полосы частот и возрастает вместе с продолжительностью промежутка времени между импульсами. Возможность

обеспечить высокую дальность связи при сравнительно скромной мощности передатчика*) за счет применения импульсных сигналов с большой временной скважностью делает этот способ передачи весьма привлекательным для межзвездной связи. Правда, введение временной скважности приводит к уменьшению пропускной способности системы связи. Однако в ряде случаев, например при передаче вызывных ВЦ, потеря информативности не имеет решающего значения.

Длительность передачи. Направленность и информативность. При увеличении направленности приемной и передающей антенн отношение сигнала к шуму и определяемая им дальность связи возрастают. Означает ли это, что всегда, при всех обстоятельствах, следует стремиться к увеличению направленности передающих антенн?

Рассмотрим зависимость между длительностью передачи и направленными свойствами антенны, когда положение абонента заранее не известно. Именно эта ситуация имеет место в проблеме связи с внеземными цивилизациями. Предположим, что передатчик с антенной находится на планете, которая подобно Земле вращается вокруг собственной оси с периодом T_p и пусть ось антенны фиксирована в системе координат, связанной с планетой. Вследствие вращения направленный луч радиоволн, излучаемых антенной, будет пересекать заданный участок неба в течение ограниченного промежутка времени Δt , пока направление на этот участок находится в пределах главного лепестка диаграммы направленности. Чем выше скорость вращения планеты и чем острее направленность антенны, тем меньше промежуток времени Δt . Пусть в некоторый момент времени t антенна направлена точно на абонента; за время $\frac{\Delta t}{2}$ она повернется на угол

$$\frac{\theta}{2} = \frac{d\theta}{dt} \frac{\Delta t}{2}, \quad (3.53)$$

*) Численные оценки см. в табл. 3.3—3.5 и на рис. 45.

где $\frac{d\theta}{dt} = \omega_p \cos \delta$ — скорость вращения диаграммы антенны, ω_p — угловая скорость вращения планеты, δ — угол между осью антенны и плоскостью экватора. Для того чтобы в рассматриваемый момент времени $t + \frac{\Delta t}{2}$ прием сигналов был возможен, угол поворота $\frac{\theta}{2}$ не должен превышать полуширины диаграммы направленности. Для определенности можно положить максимальный угол поворота, для которого еще возможен прием сигналов, равным полуширине диаграммы направленности по уровню половинной мощности ($\frac{\theta}{2} = \theta_{0,5}$). Тогда полное время передачи Δt (от момента $t - \frac{\Delta t}{2}$ до момента $t + \frac{\Delta t}{2}$) будет равно

$$\Delta t = \frac{2\theta_{0,5}}{\omega_p \cos \delta}. \quad (3.54)$$

Предположим, что используется зеркальная антенна с карандашной диаграммой направленности. Для такой антенны зависимость между шириной диаграммы и коэффициентом направленного действия дается соотношением

$$\theta^2 g = \text{const}. \quad (3.55)$$

Численное значение константы зависит от того, по какому уровню мощности отсчитывается θ . Если $\theta = 2\theta_{0,5}$, то $\text{const} = 10,2$; если $\theta = 2\theta_0$, т. е. полной ширине главного лепестка (или ширине диаграммы по уровню нулевой мощности), то $\text{const} = 59,2$. Учитывая эту зависимость, можно найти связь между длительностью передачи Δt и коэффициентом направленного действия антенны. Рис. 44, заимствованный из статьи Дж. Уэбба [7], дает представление о величине Δt при различных значениях g для случая, когда период вращения планеты равен 24^h , а ось антенны лежит в плоскости экватора.

Предположим, что после каждого оборота планеты ось антенны смещается по склонению на угол θ , равный ширине диаграммы, и таким образом произво-

дится облучение все новых и новых областей неба. Полное время обзора всего неба будет равно

$$\Delta T = T_p \frac{\pi}{\theta} \approx T_p \sqrt{g}, \quad (3.56)$$

а длительность облучения каждого абонента выразится через полное время обзора в виде

$$\Delta t = \frac{\Delta T}{2g \cos \delta}. \quad (3.57)$$

Минимальная длительность $\Delta t = \frac{\Delta T}{2g}$ получается для

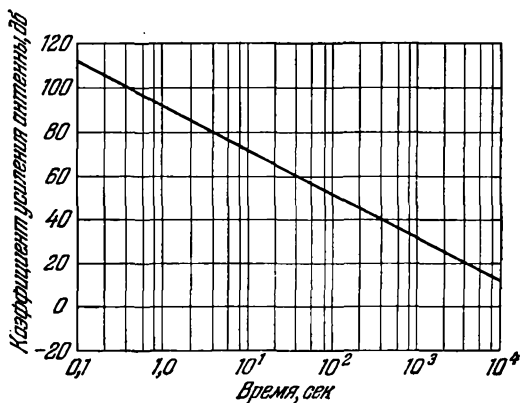


Рис. 44. Зависимость между временем передачи и направленностью передающей антенны.

Антенна неподвижна в системе координат, связанной с планетой. Ось антенны лежит в плоскости экватора, период обращения планеты равен 24 часам.

абонентов, расположенных в плоскости экватора. Похожее соотношение между полным временем обзора и длительностью облучения каждого абонента получается и в том случае, если обзор производится не за счет вращения планеты, а за счет движения самой антенны, когда она в течение времени Δt следит за данным участком неба, а затем скачком переводится на соседний.

Рассмотрим теперь зависимость между направленностью и информативностью. Пусть заданы мощность передатчика P_1 , продолжительность сеанса связи ΔT , и пусть P_c — мощность сигнала в точке приема на заданном расстоянии R в случае изотропной передачи. При изотропной передаче длительность облучения каждого абонента равна полной продолжительности сеанса связи. Поэтому максимальное количество информации Q_1 , которое может быть передано за это время, при изотропной передаче согласно теореме Шеннона равно

$$Q_1 = \Delta f \Delta T \log_2 \left(1 + \frac{P_c}{P_{\text{ш}}} \right). \quad (3.58)$$

Перейдем теперь от изотропной передачи к направленной при тех же значениях P_1 и ΔT . Тогда мощность сигнала в силу направленности возрастет в g раз, а время, затрачиваемое на облучение каждого абонента, уменьшится во столько же раз. Поэтому максимальное количество информации, которое может быть передано за время ΔT при направленной передаче, когда положение абонента заранее не известно, равно

$$Q_2 = \Delta f \Delta t \log_2 \left(1 + \frac{P_c}{P_{\text{ш}}} \right) = \Delta f \frac{\Delta T}{g} \log_2 \left(1 + \frac{gP_c}{P_{\text{ш}}} \right). \quad (3.59)$$

Сравнивая эту формулу с предыдущей, мы видим, что

$$Q_2 \leq Q_1. \quad (3.60)$$

Равенство имеет место при условии

$$\frac{P_c}{P_{\text{ш}}} < \frac{gP_c}{P_{\text{ш}}} \ll 1. \quad (3.61)$$

В этом случае, переходя от двончных логарифмов к натуральным и разлагая в ряд, получим

$$Q_2 = \frac{gP_c \Delta T \Delta f}{\ln 2 g P_{\text{ш}}} = \frac{P_c \Delta T}{\ln 2 P_{\text{ш. уд}}} = Q_1, \quad (3.62)$$

т. е. при условии $\frac{P_c}{P_{\text{ш}}} \ll 1$ количество информации растёт пропорционально продолжительности сеанса связи и отношению мощности сигнала к удельной

мощности шума и не зависит от полосы частот канала связи.

Выражение (3.60) показывает, что когда нужное направление передачи заранее не известно, то при переходе от изотропного излучения к направленному общее количество передаваемой информации при прочих равных условиях уменьшается. Этот вывод был получен В. И. Сифоровым [8]. Уменьшение информативности передачи связано с неопределенностью направления, в котором следует искать абонента. С уменьшением этой неопределенности положение меняется коренным образом. Пусть, например, на основе каких-то соображений (астрономических или других данных) передающая ВЦ может отобрать из всех возможных направлений в пространстве такие направления, в которых вероятность присутствия других цивилизаций достаточно велика. Обозначим через $\frac{1}{\gamma}$ отношение суммарного телесного угла Ω этих направлений к полному телесному углу 4π . Использование направленных антенн будет более выгодным при условии

$$\frac{\gamma}{g} \log_2 \left(1 + \frac{gP_c}{P_{\text{ш}}} \right) > \log_2 \left(1 + \frac{P_c}{P_{\text{ш}}} \right). \quad (3.63)$$

В частности, если $P_c = P_{\text{ш}}$ и $g \gg 1$, то

$$\gamma > \frac{0,3g}{\lg g}, \quad \frac{\Omega}{(2\theta_{0,5})^2} < 4 \lg g. \quad (3.64)$$

При $g = 10^6$ размер области, подлежащей обследованию с помощью направленных передающих антенн, не должен превышать площадь диаграммы (по уровню половинной мощности) более чем в 24 раза. Для слабых сигналов неопределенность в направлении поисков может быть увеличена. Пусть $\alpha = \frac{P_c}{P_{\text{ш}}} \ll 1$, но $\alpha g \gg 1$. Тогда

$$\gamma > \frac{0,43\alpha g}{\lg \alpha g}, \quad \frac{\Omega}{(2\theta_{0,5})^2} < \frac{3 \lg \alpha g}{\alpha}. \quad (3.65)$$

При $g=10^6$ и $\alpha=10^{-2}$ область неопределенности (область поисков) может превышать площадь диаграммы в 1000 раз.

§ 3. Позывные и критерии искусственности

Для установления связи с внеземными цивилизациями необходимо прежде всего обнаружить источник искусственных сигналов в космическом пространстве. При этом трудность состоит не только в том, что эти сигналы надо выявить на фоне шумов космического радионизлучения (с подобной ситуацией приходится встречаться и в обычных системах радиосвязи и радиолокации), сколько в том, что источник таких сигналов надо с уверенностью отличить от огромного числа естественных источников радионизлучения, таких как галактики, радиогалактики, квазары, облака ионизованного и нейтрального водорода, остатки вспышек сверхновых и, наконец, отдельные звезды.

Чтобы выявить радиосигналы из того излучения, которое поступает на вход приемника, его надо соответствующим образом преобразовать. Характер этого преобразования зависит от способа модуляции, применяемой нашими корреспондентами. Современная техника располагает богатым арсеналом средств для анализа радионизлучений, но было бы нелепо применять их к источникам, естественная природа которых не вызывает сомнения. Прежде чем приступить к такому анализу, необходимо убедиться в том, что мы встретились с искусственным источником радионизлучения, или по крайней мере имеем достаточно оснований подозревать его искусственное происхождение. Поэтому можно предположить, что излучение искусственного источника должно обладать свойствами, предназначенными для того, чтобы облегчить предполагаемым абонентам его обнаружение, т. е., должны существовать своего рода космические позывные.

Можно высказать несколько предположений о характере позывных ВЦ. Прежде всего они должны удовлетворять требованию высокой надежности обна-

ружения. Выполнение этого условия легче всего обеспечить путем непрерывного во времени (хотя, вообще говоря, переменного) радиоизлучения. Если положение абонента заранее не известно, излучение должно быть изотропным, так как поочередное облучение различных участков неба при высокой направленности передающей антенны приводит к очень малому времени облучения (см. рис. 44). Можно, далее, полагать, что позывные содержат информацию об искусственной природе источника излучения, указания на частоту и полосу, в которой ведется основная информативная передача, а также некоторые другие сведения, которые можно рассматривать как «ключ» к основной программе. Общее количество этой информации невелико. Поэтому в качестве позывных могут быть использованы узкополосные, квазимонохроматические сигналы. Это весьма благоприятное обстоятельство, так как подобный способ, с одной стороны, позволяет обеспечить высокую дальность связи, а, с другой, — облегчает выявление искусственного источника. Действительно, подавляющее число естественных источников радиоизлучения обладает очень широким, практически сплошным, непрерывным спектром. Даже монохроматическое излучение межзвездного водорода в линии 21 см занимает довольно широкую полосу порядка $5 \cdot 10^4$ гц. Еще более узкополосное излучение гидроксила ОН в диапазоне 18 см, для объяснения которого предлагается механизм естественного мазера [9], обладает полосой порядка нескольких сотен герц. Эти рекордно узкие естественные линии все же значительно уступают, в указанном смысле, излучению искусственных генераторов сигналов, которые позволяют получать полосу шириной в несколько герц или даже долей герца, а молекулярные генераторы вплоть до сотых долей герца. Обнаружение столь узкополосных сигналов уже само по себе явилось бы веским основанием для предположений об искусственной природе источника. Следует, однако, отметить, что применение узкополосных сигналов приводит к определенным трудностям, связанным с поисками сигнала по частоте. ▲

Эта сложная проблема не является, впрочем, непреодолимой. Мы рассмотрим ее в следующем параграфе.

Наряду с узкополосными, квазимонохроматическими сигналами можно ожидать применения в качестве позывных импульсных сигналов с большой временной скважностью. Это также позволяет обеспечить высокую дальность связи и отличить искусственный источник от естественных, излучающих, как правило, непрерывно *). Для обнаружения таких сигналов нужна специальная аппаратура и методика.

Хотя применение специальных (узкополосных, импульсных или иных) сигналов в качестве позывных ВЦ представляется наиболее разумным и вероятным, нельзя априори исключать и другую возможность, когда излучение ведется непрерывно в широкой полосе (для обеспечения большой скорости передачи информации), а роль позывных выполняют не специальные сигналы, а некоторые признаки самого источника и особые свойства его излучения.

Мы подходим, таким образом, к проблеме критериев искусственного происхождения источника радиоизлучения. Этот вопрос был впервые рассмотрен Н. С. Кардашевым [10]. Впоследствии он анализировался В. И. Слышем [11], Л. И. Гудзенко, Б. Н. Пановкиным [12] и другими. Предложенные критерии удобно разбить на две группы: 1) критерии или признаки, непосредственно вытекающие из искусственной природы источника; 2) специальные свойства излучения, намеренно заложенные передающей ВЦ для обеспечения связи и для облегчения задачи обнаружения. К первой группе можно отнести: угловые размеры; спектр; статистические свойства сигналов; переменность, связанную с возможным вращением системы. Ко второй группе — круговую поляризацию; переменность, связанную с модуляцией излучения; информацию об искусственности и «ключ».

Угловые размеры являются одним из наиболее показательных критериев первой группы. Для источников искусственного происхождения видимый угло-

*) Замечательным исключением из этого правила являются пульсары.

вой размер не может быть больше некоторой, вообще говоря, достаточно малой величины. Это связано с одной стороны, с ограничением пространственных масштабов деятельности цивилизаций (например, размерами планетных систем), а с другой стороны, с конечной скоростью передачи информации. Действительно, пусть t — промежуток времени между двумя последовательными импульсами; для того чтобы различные части передающей системы излучали согласованно, расстояние между ними, а значит, линейный размер всей системы L не должен превосходить величины ct (c — скорость света). Если R — расстояние до передающей системы, то ее видимые угловые размеры

$$\varphi < \frac{ct}{R} = \frac{c}{Rq}, \quad (3.66)$$

где q — скорость передачи информации. Для расстояния в 1 *кпс* при $q = 3 \cdot 10^{-4}$ (что соответствует передаче одной двоичной единицы информации в час) получаем $\varphi < 0'',007$. С увеличением скорости передачи информации предельные угловые размеры источника соответственно уменьшаются. Для скорости 1 *бит/сек* $\varphi < 0'',000002$. Между тем угловые размеры естественных источников, как правило, значительно больше. Даже наименее протяженные из них (источники линии гидроксила OH) имеют угловые размеры порядка нескольких тысячных долей секунды дуги. Правда, следует иметь в виду, что с увеличением чувствительности и разрешающей способности радиотелескопов, когда им станет доступна регистрация излучения отдельных звезд, этот критерий утратит абсолютную роль, но в сочетании с другими признаками (мощность, ширина полосы и т. п.) он, по-видимому, сохранит определенное значение.

Если передатчик ВЦ работает в достаточно широкой полосе спектра, его излучение должно походить на непрерывное излучение естественных источников. Однако можно полагать, что спектральное распределение мощности передатчика отличается от распределения мощности естественных источников радиои излучения. Этот вопрос подробно рассматривался в

гл. I. Если ставится задача обеспечить максимальную скорость передачи информации, то спектр искусственного источника должен иметь вид кривых, изображенных на рис. 21 и 22. Таковую форму кривой можно рассматривать как один из критериев искусственного происхождения источника. Правда, этот критерий недостаточно сильный. Во-первых, условие максимальной скорости передачи информации не является строго обязательным. Более того, чрезмерная насыщенность информацией даже нежелательна, так как затрудняет дешифровку сигналов. Во-вторых, похожую форму кривой спектрального распределения мощности в некоторых случаях могут иметь и естественные источники излучения. Тем не менее, и этот критерий не следует совершенно сбрасывать со счета. В сочетании с другими свойствами радионизлучения он может оказаться весьма полезным для суждений о природе источников.

То же самое относится к переменности, связанной с возможным вращением системы. В этом случае продолжительность облучения определяется периодом обращения и направленностью передающей антенны, а полный период изменения мощности — только периодом обращения. Здесь можно ожидать изменений с периодом от нескольких часов до нескольких суток, если передатчик расположен на планете, вращающейся вокруг собственной оси, или изменений с периодом от нескольких месяцев до нескольких лет, если планета или какое-то другое небесное тело, на котором расположен передатчик, обращается только вокруг своей звезды, на некотором расстоянии от нее, в пределах ее «зоны жизни». Долгое время считалось, что естественные источники радионизлучения отличаются высокой степенью постоянства излучаемой мощности. Это следовало из теоретических расчетов и подтверждалось непосредственными измерениями. После открытия Г. Б. Шоломицким [13] переменности источника СТА-102 положение изменилось коренным образом, так как вслед за этим была обнаружена переменность радионизлучения квазаров на различных частотах и с различным характерным вре-

менем изменения (от нескольких дней до нескольких лет). Таким образом, этот критерий также потерял абсолютное значение, но так же, как и другие критерии, его не следует совершенно упускать из виду.

Пожалуй, наиболее сильным критерием первой группы является критерий, связанный со статистическими свойствами излучения. Этот вопрос рассматривался М. Голеем [14], В. И. Слышем [11], Л. И. Гудзенко и Б. Н. Пановкиным [12], В. И. Сифоровым [8]. Суть дела состоит в том, что радиоизлучение естественных источников является чисто случайным, некоррелированным шумом, так как оно складывается из очень большого числа независимых элементарных актов излучения. В генераторах же сигналов, применяемых в радиотехнике, отдельные элементарные акты не являются полностью независимыми. Поэтому искусственное излучение по своим статистическим свойствам (например, по распределению амплитуд) отличается от шумового излучения. Следовательно, поиски искусственных радионисточников должны включать возможно более полное исследование статистических свойств сигналов*). Анализ подобного рода в условиях наблюдения очень слабых космических источников радиоизлучения представляет собой нелегкую задачу и требует создания специальной аппаратуры, отличной от той, которая применяется обычно в радиоастрономии. Следует отметить, что хотя на необходимость разработки статистических критериев неоднократно обращалось внимание, вопрос этот до сих пор исследован очень слабо.

Перейдем к рассмотрению критериев, относящихся ко второй группе. При распространении плоскополяризованного радиоизлучения в межзвездной среде плоскость поляризации может испытывать сильное вращение в межзвездных магнитных полях вследствие эффекта Фарадея. Это явление хорошо

*) С этой целью можно исследовать моменты функции распределения, автокорреляционную функцию, функцию корреляции спектра и др. [11].

известно в радиоастрономии и используется для оценки расстояния до радиоисточников по наблюдаемому повороту плоскости поляризации. Но если в радиоастрономии оно приносит определенную пользу, позволяя получить дополнительную информацию об источнике, то в проблеме связи с ВЦ оно может принести только вред, искажая поступающую информацию. Дело в том, что эффект вращения в сильной степени зависит от частоты. Поэтому при распространении широкополосного сигнала в межзвездной среде различные спектральные составляющие сигнала поворачиваются на различные углы. В результате при приеме на антенну, которая регистрирует только одно направление поляризации, разные спектральные составляющие испытают различное ослабление. Это приведет к искажению спектра, а следовательно, и временных характеристик сигнала. Чтобы избежать этого неприятного явления, радионизлучение, посылаемое передатчиком ВЦ, с самого начала должно быть поляризованным по кругу и его прием надо вести на антенну с круговой поляризацией. Именно так поступают в современных системах дальней космической связи (в пределах Солнечной системы).

Переменность, связанная с модуляцией излучения, является самым надежным признаком искусственного происхождения радиосигналов. Трудность обнаружения этой переменности состоит в том, что совершенно не известно характерное время возможных изменений мощности. Если речь идет о модуляции, связанной непосредственно с кодированием информации, время модуляции, скорее всего, очень мало. Если, например, передача ведется двоичным кодом, то при скорости передачи информации 1000 бит/сек , которую трудно считать высокой (для телевидения необходима в тысячу раз более высокая скорость), характерное время изменения мощности, совпадающее в этом случае с продолжительностью двоичного импульса, составляет 10^{-3} сек . Регистрация таких кратковременных вариаций требует применения регистрирующей аппаратуры с очень малой постоянной времени $\tau_2 \leq 10^{-3} \text{ сек}$. Для обеспечения высокой чув-

ствительности при такой малой постоянной времени необходимы приемные антенны с очень большой эффективной площадью.

Что касается позывных, то для них скорость передачи информации может быть значительно ниже. Вероятно, скорости 1 бит/сек с избытком достаточно, чтобы за разумное время передать те несколько десятков или несколько сотен бит информации, которые, как можно полагать, содержатся в позывных ВЦ. О характере этой информации, предназначенной для того, чтобы сообщить об искусственной природе источника излучения, можно только строить различные догадки. Предполагалось, например, что для этого могут быть использованы несколько первых чисел натурального ряда, или несколько простых чисел, или, наконец, комбинация известных математических констант, вроде чисел e и π .

Заметим, что помимо специальных монохроматических сигналов функцию позывных могут также выполнять переменные широкополосные сигналы, полученные модуляцией несущих информацию кратковременных импульсов. Переменность может, в частности, соответствовать перерывам в основной программе, например, началу или концу определенного сеанса связи. Эти медленные изменения мощности широкополосного радиоизлучения можно обнаружить уже при помощи существующей радиоастрономической аппаратуры. Однако при этом очень важно знать величину ожидаемого периода изменения: будут ли это секунды, минуты или годы? В настоящее время этот вопрос совершенно не ясен. Можно указать только очень грубую нижнюю границу для величины характерного времени изменения в позывных ВЦ. Если передача ведется на частоте ν , то время модуляции τ в позывных ВЦ должно удовлетворять условию

$$\tau > \tau_1 = \Delta \nu^{-1} > \nu^{-1}. \quad (3.67)$$

Для радиодиапазона это дает $\tau > 10^{-11} \div 10^{-9}$ сек. Более точную оценку можно получить из условия стабильности импульса при распространении в

межзвездной среде. Как было показано в гл. II, эффект группового запаздывания, связанный с различием групповой скорости для разных квазимонохроматических групп волн, из которых состоит широкополосный импульс, накладывает определенные ограничения на его продолжительность τ . Так, для галактического источника, работающего в диапазоне дециметровых волн, продолжительность импульса должна быть много больше 10^{-6} сек (если источник расположен вне плоскости Галактики) и много больше 10^{-5} сек (если источник расположен в галактической плоскости). Для внегалактического источника предельная продолжительность импульса может достигать 10^{-4} сек. Во всяком случае, мы можем написать

$$\tau > 10^{-6} \text{ сек.} \quad (3.68)$$

Задача обнаружения позывных сигналов ВЦ существенно упростилась бы, если бы можно было указать какой-то стандартный период модуляции, которым пользуются все ВЦ. Разумеется, этот период должен удовлетворять условиям (3.67) и (3.68). Можно попытаться найти подход к решению этого вопроса, подобрав подходящую комбинацию фундаментальных мировых констант, имеющих размерность времени, или взяв за основу характерное время определенных универсальных для Вселенной процессов, скажем, атомных или космологических. Например, мы можем воспользоваться атомной единицей времени, равной периоду обращения электрона по первой бордовской орбите, или так называемым йордановским элементарным временем, равным классическому радиусу электрона, деленному на скорость света. Первая из этих величин равна $2,4 \cdot 10^{-17}$ сек, вторая $9,4 \cdot 10^{-24}$ сек. Однако ни одна из них не удовлетворяет условию (3.67) и тем более (3.68). Рассмотренные единицы времени определяют временной масштаб явлений в микромире. Их можно было бы назвать микроскопическими единицами времени. С другой стороны, нам известен совершенно иной космический масштаб времени, связанный с расширением Вселенной, — масштаб, который характеризуется постоянной

Хаббла H — универсальной константой мегамира. Можно думать, что период модуляции в позывных ВЦ лежит где-то «посередине» между микроскопическими и мегаскопическими единицами времени; например, он может представлять собой среднее геометрическое из указанных величин. Беря те же атомное и йордановское элементарное время и пользуясь мегаскопической единицей $H^{-1} = 3 \cdot 10^{17}$ сек, мы можем получить две макроскопические единицы времени

$$\tau_a = \sqrt{2,4 \cdot 10^{-17} \times 3 \cdot 10^{17}} = 3 \text{ сек}$$

и

$$\tau_z = \sqrt{9,4 \cdot 10^{-24} \times 3 \cdot 10^{17}} = 0,002 \text{ сек.}$$

Обе величины удовлетворяют условиям (3.67) и (3.68). В качестве микроскопической единицы времени довольно естественно пользоваться величиной ν^{-1} , где ν — частота посылаемого сигнала. Для оптимального диапазона межзвездной связи $\nu^{-1} = 10^{11} \div 10^{-9}$ сек, и соответствующие макроскопические величины заключены в пределах от получаса до 4,5 часа. Разумеется, приведенные примеры носят только иллюстративный характер. В частности, трудности, связанные с точным определением постоянной Хаббла, делают ее мало удобной для основной единицы времени. Впрочем, можно думать, что внеземные цивилизации, приступившие к межзвездной связи, располагают достаточно точными знаниями о величине H .

Вероятно, в первую очередь следовало бы искать изменения с периодом в несколько часов. Такие изменения проще обнаружить, для их регистрации не нужно разрабатывать специальную аппаратуру. Подобные вариации мощности не искажаются эффектом мерцания, возникающим при распространении электромагнитных волн в межзвездной, в межпланетной среде и в ионосфере. Даваемый ими масштаб временных явлений достаточно характерен для макромира, к которому, надо полагать, принадлежат наши партнеры (во всяком случае, если речь идет об антропоморфных цивилизациях). Наконец, сам факт обнаружения

периодических изменений мощности с периодом τ , который связан с частотой излучения соотношением $\tau = \sqrt{\nu^{-1} H^{-1}}$, несомненно, заставил бы отнестись к такому явлению с повышенным интересом и вниманием.

Заканчивая этот параграф, отметим, что затронутая в нем проблема позывных и критериев искусственности почти совершенно не разработана. В этих вопросах еще очень много неясного, неопределенного, что открывает обширное поле для исследований. Следует подчеркнуть необходимость разработки строгих и совершенно однозначных критериев искусственного происхождения источника, основанных, например, на анализе статистических свойств сигнала или на общих теоремах теории информации и теории сложных систем. Некоторый подход к решению последней задачи намечен в гл. VI настоящей книги.

§ 4. Методы обнаружения сигналов внеземных цивилизаций

Мощность передатчика. Энергетический потенциал цивилизаций. При попытках обнаружения сигналов ВЦ мы сталкиваемся с неопределенностью двоякого рода: неизвестно, на какой частоте следует искать эти сигналы и в каком направлении. Аналогичная неопределенность возникает и для передающей ВЦ. Наиболее простым выходом из этого положения является непрерывное излучение достаточно широкополосных сигналов по всем направлениям в пространстве. Это позволяет одновременно обслуживать все цивилизации, расположенные в сфере действия передатчика, и дает возможность к подключению все новых абонентов по мере того как они достигнут соответствующего уровня развития. Если при этом сигналы обладают достаточно большой мощностью, а приемник достаточно высокой чувствительностью, то для приема сигналов можно использовать малонаправленные или совсем не направленные антенны, что позволяет избежать поисков по направлению на первом этапе обнаружения. Однако для такого «простого» способа свя-

зи требуются гигантские мощности. В таблице 3.2 приведены значения минимальной мощности передатчика, необходимой для обнаружения и связи при непрерывной изотропной передаче и ненаправленном приеме на волне 3 см. При этом для обнаружения принималось десятикратное превышение уровня полезных сигналов над средней квадратической флуктуацией шума ($\beta=10$), а для приема информации — сто-кратное превышение сигнала над шумом ($\alpha=100$); шумовая температура принималась равной 10°K , а постоянная времени (при обнаружении) $\tau_2=100 \text{ сек.}$ Наконец, в соответствии с условиями минимальной мощности положено $\Delta f_1=\Delta f_2=\Delta f$. Значения полосы Δf выражены в герцах, расстояние R — в световых годах, мощность передатчика P_1 — в ваттах. Как можно видеть на основании чисел табл. 3.2, необходимые значения мощности не только лежат далеко за пределами возможности наших современных передающих устройств, но и значительно превышают потенциальные энергетические возможности человечества.

Современное человечество потребляет в год около $1,5 \cdot 10^{27} \text{ эрг}$ энергии всех видов, что соответствует мощности порядка $5 \cdot 10^{12} \text{ вт.}$ Вся используемая энергия в конечном счете переходит в тепло и излучается в мировое пространство. Принципиально ее можно было бы преобразовать в радиоизлучение (такой процесс не противоречит термодинамике) и использовать для космической радиосвязи. Однако при рассматриваемом способе связи (непрерывная изотропная передача — ненаправленный прием) излучаемой мощности все равно не хватило бы для передачи информации даже в пределах десяти световых лет, т. е. на расстояние ближайших звезд. Означает ли это, что данный тип связи является абсолютно бесперспективным? Разумеется, нет. Коль скоро мы допускаем, что наша цивилизация не является единственной во Вселенной, то отсюда с неизбежностью вытекает, что должны существовать цивилизации как нашего и менее развитого уровня, так и более развитые, чем наша. Последние могут располагать гигантскими энергетическими мощностями, совершенно недоступными

Таблица 3.2

Минимальная мощность передатчика при непрерывной изотропной передаче и ненаправленном приеме на волне 3 см

$$(g_1 = g_2 = 1; \Delta f_1 = \Delta f_2 = \Delta f)$$

		Обнаружение ($\beta = 10$; $\tau_2 = 100$ сек; $T_{ш} = 10^\circ \text{K}$)					Связь (прием информации) ($\alpha = 100$; $T_{ш} = 10^\circ \text{K}$)						
Δf		10^7	10^8	10^9	10^{10}	10^{11}	1	10	10^2	10^3	10^5	10^6	10^8
R													
10	$2 \cdot 10^{17} 2 \cdot 10^{18}$	$2 \cdot 10^{19} 2 \cdot 10^{20}$	$2 \cdot 10^{21} 2 \cdot 10^{22}$	$2 \cdot 10^{23} 2 \cdot 10^{24}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{19} 2 \cdot 10^{20}$	$2 \cdot 10^{21} 2 \cdot 10^{22}$	$2 \cdot 10^{23} 2 \cdot 10^{24}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$
10^2	$2 \cdot 10^{19} 2 \cdot 10^{20}$	$2 \cdot 10^{21} 2 \cdot 10^{22}$	$2 \cdot 10^{23} 2 \cdot 10^{24}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{21} 2 \cdot 10^{22}$	$2 \cdot 10^{23} 2 \cdot 10^{24}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$
10^3	$2 \cdot 10^{21} 2 \cdot 10^{22}$	$2 \cdot 10^{23} 2 \cdot 10^{24}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{23} 2 \cdot 10^{24}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$
10^4	$2 \cdot 10^{23} 2 \cdot 10^{24}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$
10^5	$2 \cdot 10^{25} 2 \cdot 10^{26}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$
10^6	$2 \cdot 10^{27} 2 \cdot 10^{28}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$	$2 \cdot 10^{41} 2 \cdot 10^{42}$
10^7	$2 \cdot 10^{29} 2 \cdot 10^{30}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$	$2 \cdot 10^{41} 2 \cdot 10^{42}$	$2 \cdot 10^{43} 2 \cdot 10^{44}$
10^8	$2 \cdot 10^{31} 2 \cdot 10^{32}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$	$2 \cdot 10^{41} 2 \cdot 10^{42}$	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$	$2 \cdot 10^{41} 2 \cdot 10^{42}$	$2 \cdot 10^{43} 2 \cdot 10^{44}$	$2 \cdot 10^{45} 2 \cdot 10^{46}$
10^9	$2 \cdot 10^{33} 2 \cdot 10^{34}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$	$2 \cdot 10^{41} 2 \cdot 10^{42}$	$2 \cdot 10^{43} 2 \cdot 10^{44}$	$2 \cdot 10^{35} 2 \cdot 10^{36}$	$2 \cdot 10^{37} 2 \cdot 10^{38}$	$2 \cdot 10^{39} 2 \cdot 10^{40}$	$2 \cdot 10^{41} 2 \cdot 10^{42}$	$2 \cdot 10^{43} 2 \cdot 10^{44}$	$2 \cdot 10^{45} 2 \cdot 10^{46}$	$2 \cdot 10^{47} 2 \cdot 10^{48}$

человечеству на современном этапе его развития. Энергетический потенциал цивилизации, в конечном счете, определяет мощность ее передатчиков. А это один из важнейших параметров межзвездной связи, от которого зависят дальность обнаружения и связи, объем передаваемой информации, характер сигналов, а следовательно, и методы их обнаружения. Поэтому вопрос о возможном энергетическом потенциале цивилизаций заслуживает более подробного обсуждения.

Основные черты роста важнейших показателей технического развития цивилизаций анализируются в главах I, V и VI. Мы остановимся здесь только на рассмотрении энергетического развития цивилизаций.

Ежегодный прирост энергопотребления на земном шаре за последние 100 лет составляет приблизительно 3%. Если и в дальнейшем прирост будет идти такими же темпами, то уже через 300 лет ежесекундное потребление энергии на Земле достигнет 10^{17} вт, т. е. сравняется с притоком энергии от Солнца. Дальнейшее увеличение энергопотребления станет невозможным, так как это поведет к значительному нарушению радиационного баланса планеты*). Это очень важное обстоятельство, значение которого обычно недооценивается. Следует подчеркнуть, что ограничение производства энергии не связано с недостатчей энергетических ресурсов, а вытекает из необходимости сохранить равновесие основных процессов в атмосфере и на поверхности земного шара.

С аналогичным положением должна столкнуться любая цивилизация, обитающая на какой-либо планете. Так как величина энергии, которую получает от своей звезды планета, находящаяся в «зоне жизни», не может изменяться в очень широких пределах, то для любой планетной цивилизации расход энергии должен быть ограничен величиной порядка 10^{17} вт. По достижении этого предела дальнейшее развитие

*) Фактически увеличение производства энергии на земном шаре необходимо прекратить значительно раньше, когда оно достигнет нескольких процентов от величины потока энергии, ежесекундно поступающей от Солнца (т. е. через 150—200 лет).

может быть обеспечено только на пути активного освоения и заселения околозвездного пространства, размещения в нем мощных энергетических установок и производства. В свете этих обстоятельств не является ли начавшееся освоение космоса, о котором часто пишут, что оно преждевременно, жизненно важным делом для человечества? Активное освоение и заселение космического пространства в конечном итоге должно привести к созданию искусственной биосферы вокруг Солнца (сферы Циолковского — Дайсона). Цивилизация такого типа, обитающая на искусственной биосфере в окрестностях своей звезды, может располагать гораздо большей мощностью, вплоть до $3 \cdot 10^{26}$ вт. При экспоненциальном характере развития время, которое требуется для перехода к цивилизации дайсоновского типа, относительно невелико. Действительно, если предположить, что ежегодный прирост энергии составляет только 1%, то для перехода от цивилизации планетного типа с энергопотреблением 10^{17} вт к цивилизации дайсоновского типа с энергетическим потреблением порядка $3 \cdot 10^{26}$ вт потребуется всего 2200 лет. А еще через 2500 лет ежесекундное потребление энергии при том же приросте достигнет 10^{37} вт, что равно энергии излучения всех звезд Галактики *). Этот процесс экстраполяции развития цивилизаций может быть продлен и дальше, но более разумно где-то остановиться. В 1964 г. Н. С. Кардашев [10] предложил разделить все цивилизации на три типа по величине потребляемой ими энергии. К I типу он относит цивилизации, по техническому уровню развития близкие к нашей земной цивилизации (ежесекундное потребление энергии $10^{12} \div 10^{13}$ вт), к II типу — цивилизации с энергопотреблением порядка $3 \cdot 10^{26}$ вт и, наконец, к III типу — цивилизации, овладевшие энергетическими ресурсами в масштабе целой галактики с энергопотреблением 10^{37} вт. Мы воспользуемся терминологией Н. С. Кардашева, но будем понимать

*) При более высоком ежегодном приросте эти величины энергопотребления будут достигнуты значительно раньше (см., например, гл. I, стр. 50.)

под цивилизациями I типа все планетные цивилизации, как с уровнем энергопотребления, близким к нашему, так и с более высоким, вплоть до 10^{17} вт. Разумеется, предположение о существовании сверхцивилизаций с энергетическим потреблением порядка $10^{26} \div 10^{37}$ вт является только гипотезой. Но, строго говоря, гипотезой является и предположение о существовании внеземных цивилизаций вообще. Лучшим способом проверки этой гипотезы будут экспериментальные исследования по обнаружению сигналов от таких сверхцивилизаций. Поэтому представляется полезным оценить открывающиеся в этом направлении возможности.

Вернемся к табл. 3.2. При полосе больше 1 гц обнаружение сигналов и прием информации от цивилизаций I типа невозможны даже с расстояния ближайших звезд. Следовательно, при непрерывной изотропной передаче и ненаправленном приеме может иметь место связь только с цивилизациями II или III типов. При полосе 1 Мгц обнаружение сигналов от цивилизаций II типа возможно на расстоянии 1000 световых лет, а прием информации от них только на расстоянии ближайших звезд. При полосе $\Delta f \leq 100$ гц обнаружение сигналов от цивилизаций II типа возможно повсюду в пределах Галактики, а прием информации от них на расстоянии более 1000 световых лет. Что касается цивилизаций III типа, то обнаружение сигналов от них возможно практически во всей наблюдаемой области Вселенной. При полосе 10 Мгц они могут обеспечить передачу информации только на расстояние ближайших галактик, а при достаточно узкой полосе $\Delta f \leq 100$ гц — в пределах Метагалактики. Отсюда следует, что если где-нибудь в нашей Галактике существует хотя бы одна цивилизация II типа, или где-нибудь во Вселенной хотя бы одна цивилизация III типа, которые, используя свой гигантский энергетический потенциал, непрерывно излучают по всем направлениям мощные монохроматические сигналы с полосой $\Delta f \leq 100$ гц, мы

можем обнаружить эти сигналы, даже не зная, где находится их источник.

Рассмотренный способ связи с энергетической точки зрения является наименее выгодным. При других способах передачи и приема требуются значительно меньшие мощности. Пусть, например, в нашем распоряжении имеется первоклассный приемник с шумовой температурой 10°K и полосой 100 МГц , работающий в диапазоне 3 см . Предположим, что приемник вынесен за пределы атмосферы, так что полная шумовая температура определяется только собственными шумами приемника и равна $T_{\text{ш}} = 10^\circ \text{K}$. Определим, какая мощность требуется для обнаружения сигналов и для приема информации на расстоянии 1000 световых лет при различных способах передачи и приема сигналов. Расстояние 1000 световых лет выбрано нами потому, что согласно некоторым современным оценкам [2, 3, 15] оно соответствует среднему расстоянию до ближайшей ВЦ. В табл. 3.3 приведены значения мощности (в ваттах), необходимой для обнаружения сигналов и связи на расстоянии 1000 световых лет при условии $\Delta f_1 < \Delta f_2$. Для приемной антенны принята площадь 10^4 м^2 , что соответствует площади крупнейших современных радиотелескопов, работающих в сантиметровом и дециметровом диапазонах. Для направленной передачи приводится произведение мощности на коэффициент усиления антенны в децибелах на ватты, а также значения мощности при коэффициенте усиления $g_1 = 10^9$. Прочерки в двух последних столбцах таблицы связаны с тем, что формально условие обнаружения импульсных сигналов ($\tau_2 < \Delta t_1$) совпадает с условием для приема информации (см. § 2). Заметим, что для проблемы связи с ВЦ основное значение имеет не относительная продолжительность импульсов $\frac{\Delta t_1}{t_1}$, а величина t_1 — промежутка времени между ними, которая определяет скорость передачи информации в случае импульсных сигналов. Так как в нашем случае $\Delta t_1 = \Delta f_1^{-1} > 10^{-8}$, то приведенным в таблице значениям относительной продолжительности

Таблица 3.3

Мощность передатчика, необходимая для обнаружения узкополосных сигналов и связи на расстоянии 1000 световых лет
 $\lambda = 3 \text{ см}$, $\Delta f_1 < \Delta f_2 = 100 \text{ МГц}$, $T_{\text{ш}} = 10^\circ \text{ К}$

Передача	Связь; $\alpha = 100$		Обнаружение; $\beta = 10$, $\tau_2 = 100 \text{ сек}$	
	ненаправленный прием $T_{\text{ш}} = 10^\circ \text{ К}$	направленный прием $S_2 = 10^4 \text{ м}^2$; $T_{\text{ш}} = 10^\circ \text{ К}$	ненаправленный прием $T_{\text{ш}} = 10^\circ \text{ К}$	направленный прием $S_2 = 10^4 \text{ м}^2$; $T_{\text{ш}} = 10^\circ \text{ К}$
Изотропная непрерывная	$2 \cdot 10^{31}$	$1,5 \cdot 10^{23}$	$2 \cdot 10^{25}$	$1,5 \cdot 10^{17}$
Изотропная импульсная, $\frac{\Delta t_1}{t_1} =$				
10^{-1}	$2 \cdot 10^{30}$	$1,5 \cdot 10^{22}$		
10^{-2}	$2 \cdot 10^{29}$	$1,5 \cdot 10^{21}$		
10^{-4}	$2 \cdot 10^{27}$	$1,5 \cdot 10^{19}$		
10^{-6}	$2 \cdot 10^{25}$	$1,5 \cdot 10^{17}$	—	—
10^{-8}	$2 \cdot 10^{23}$	$1,5 \cdot 10^{15}$		
10^{-10}	$2 \cdot 10^{21}$	$1,5 \cdot 10^{13}$		
Направленная непрерывная	$(P_1 g_1) \frac{P_1}{\text{дб} \cdot \text{вт}; \text{при } g_1 = 10^9}$	$(P_1 g_1) \frac{P_1}{\text{дб} \cdot \text{вт}; \text{при } g_1 = 10^9}$	$\frac{P_1 g_1}{\text{дб} \cdot \text{вт}; \text{при } g_1 = 10^9} \frac{P_1}{\text{дб} \cdot \text{вт}; \text{при } g_1 = 10^9}$	$(P_1 g_1) \frac{P_1}{\text{дб} \cdot \text{вт}; \text{при } g_1 = 10^9}$
	313 $2 \cdot 10^{22}$	232 $1,5 \cdot 10^{14}$	253 $2 \cdot 10^{16}$	172 $1,5 \cdot 10^8$
Направленная импульсная, $\frac{\Delta t_1}{t_1} =$				
10^{-1}	303 $2 \cdot 10^{21}$	222 $1,5 \cdot 10^{13}$		
10^{-2}	293 $2 \cdot 10^{20}$	212 $1,5 \cdot 10^{12}$		
10^{-4}	273 $2 \cdot 10^{18}$	192 $1,5 \cdot 10^{10}$		
10^{-6}	253 $2 \cdot 10^{16}$	172 $1,5 \cdot 10^8$	—	—
10^{-8}	233 $2 \cdot 10^{14}$	152 $1,5 \cdot 10^6$		
10^{-10}	213 $2 \cdot 10^{12}$	132 $1,5 \cdot 10^4$		

импульсов соответствует следующая скорость передачи информации *):

$$\frac{\Delta t_1}{t_1} = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}$$

$$t_1 > 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-2}, 1, 10^2 \text{ сек}$$

$$q = \frac{1}{t_1} < 10^7, 10^6, 10^5, 10^4, 10^2, 1, 10^{-2} \text{ бит/сек}$$

Как можно видеть на основании данных табл. 3.3, в случае изотропной непрерывной передачи и ненаправленного приема обнаружение сигналов и прием информации от цивилизаций I типа на расстоянии 1000 световых лет невозможны. Обнаружение сигналов от цивилизаций II типа при этих условиях возможно, но для приема информации мощности не хватает. Чтобы обеспечить прием информации, надо перейти либо к импульсной либо к направленной передаче, либо, наконец, к направленному приему. В случае изотропной непрерывной передачи при направленном приеме возможно обнаружение сигналов от цивилизаций I типа и прием информации от цивилизаций II типа. Изотропная импульсная передача при относительной длительности импульсов $\left(\frac{\Delta t_1}{t_1}\right) < 10^{-5}$ позволяет осуществить связь с цивилизациями II типа с использованием ненаправленной приемной антенны. Переход к направленной приемной антенне с эффективной площадью 10^4 м^2 позволяет осуществить связь с цивилизациями II типа практически при любой относительной продолжительности импульсов. При $\left(\frac{\Delta t_1}{t_1}\right) \leq 10^{-6}$ становится возможным прием информации от цивилизаций I типа (со скоростью меньше 100 бит/сек); в частности, при $\left(\frac{\Delta t_1}{t_1}\right) = 10^{-10}$, когда скорость передачи информации меньше одного бита в

*) В предположении, что используются двоичные импульсы. Если применяется импульсный код с некоторым основанием $a \neq 2$, то числа, характеризующие скорости передачи информации, следует умножить на величину $\log_2 a$.

минуту, требуется мощность порядка 10^{13} вт, что уже сравнимо с современным расходом энергии человечеством. В случае непрерывной направленной передачи и ненаправленного приема возможно обнаружение сигналов от цивилизаций I типа и прием информации от цивилизаций II типа. При направленных приеме и передаче для приема информации требуется 10^{11} квт, а для обнаружения сигналов 150 Мвт. Направленная импульсная передача позволяет осуществить связь с цивилизациями II типа при приеме на ненаправленную антенну при любой относительной продолжительности импульсов. При относительной продолжительности импульсов $< 10^{-5}$ возможна связь с цивилизациями I типа при приеме на ненаправленную антенну, а при приеме на направленную антенну с эффективной площадью 10^4 м² связь с цивилизациями I типа возможна при любой продолжительности импульсов (в том числе при непрерывной передаче). Наконец, в случае направленной импульсной передачи с относительной длительностью импульсов меньше 10^{-8} и приеме на направленную антенну с эффективной площадью 10^4 м² требуется уже совсем скромная мощность: менее 1,5 Мвт.

В табл. 3.4 приведены значения мощности, необходимой для обнаружения сигналов при тех же параметрах системы связи, но при условии, что $\Delta f_1 > \Delta f_2$. Следует отметить, что в данном случае, строго говоря, речь может идти только об обнаружении сигналов, так как связь при условии $\Delta f_1 > \Delta f_2$ неизбежно сопряжена с искажением сигналов и потерей информации. Это надо иметь в виду в отношении левой половины таблицы, где приводятся значения мощности, которая требуется для такой неполной связи. В отличие от рассмотренного выше примера, когда $\Delta f_1 < \Delta f_2$, в данном случае значения мощности при импульсной передаче зависят не от относительной длительности импульсов, а от промежутка времени между ними. При сравнении данных этой таблицы с данными предыдущей табл. 3.3 следует иметь в виду, что так как здесь $\Delta t_1 = \Delta f_1^{-1} < 10^{-8}$, то приведенным в табл. 3.4

Таблица 3.4

Мощность передатчика, необходимая для обнаружения широкополосных сигналов и связи (с потерей информации) на расстоянии 1000 световых лет
 $(\lambda = 3 \text{ см}, \Delta f_1 > \Delta f_2 = 10^8 \text{ гц}, T_{\text{ш}} = 10^\circ \text{ К})$

№	Передача	Связь (с потерей информации) $\alpha = 100$	Обнаружение; $\beta = 10; \tau_2 = 100 \text{ сек}$	
		ненаправленный прием, $T_{\text{ш}} = 10^\circ \text{ К}$	направленный прием, $S_2 = 10^4 \text{ м}^2; T_{\text{ш}} = 10^\circ \text{ К}$	ненаправленный прием, $T_{\text{ш}} = 10^\circ \text{ К}$
1	Изотропная непрерывная $\Delta f_1 = 10^9$ $\Delta f_1 = 10^{10}$	$2 \cdot 10^{32}$ $2 \cdot 10^{33}$	$1,5 \cdot 10^{24}$ $1,5 \cdot 10^{25}$	$2 \cdot 10^{26}$ $2 \cdot 10^{27}$
2	Изотропная импульсная $t_1 =$ 1^s 1^m 1^h 24^h 1 месяц 1 год	$2 \cdot 10^{23}$ $3 \cdot 10^{21}$ $6 \cdot 10^{19}$ $2 \cdot 10^{18}$ $8 \cdot 10^{16}$ $7 \cdot 10^{15}$	$1,5 \cdot 10^{15}$ $2,5 \cdot 10^{13}$ $4 \cdot 10^{11}$ $2 \cdot 10^{10}$ $6 \cdot 10^8$ $5 \cdot 10^7$	—
3	Направленная непрерывная $\Delta f_1 = 10^9$ $\Delta f_1 = 10^{10}$	$(P_1 g_1)$ дб, ат; при $g_1 = 10^9$ 323 333	$(P_1 g_1)$ дб, ат; при $g_1 = 10^9$ 242 252	P_1 дб, ат; при $g_1 = 10^9$ 182 192

Продолжение табл. 3.4

№	Передача	Связь с потерей информации $\alpha = 100$		Обнаружение: $\beta = 10$; $\tau_2 = 100$ сек	
		ненаправленный прием, $T_{ш} = 10^6$ К	направленный прием, $S_2 = 10^4$ мк; $T_{ш} = 10^6$ К	ненаправленный прием, $T_{ш} = 10^6$ К	направленный прием, $S_2 = 10^4$ мк; $T_{ш} = 10^6$ К
4	Направленная импульсная $t_1 =$ 1^s 1^m 1^h 24^h 1 месяц 1 год	$(P_1 g_1)$ дб. вт; при $g_1 = 10^9$ 233 215 197 184 169 158	P_1 дб. вт; при $g_1 = 10^9$ $2 \cdot 10^{14}$ $3 \cdot 10^{12}$ $6 \cdot 10^{10}$ $2 \cdot 10^9$ $8 \cdot 10^7$ $7 \cdot 10^6$	$(P_1 g_1)$ дб. вт; при $g_1 = 10^9$ 152 134 116 103 88 77	P_1 дб. вт; при $g_1 = 10^9$ $1,5 \cdot 10^{16}$ $2,5 \cdot 10^{14}$ $4 \cdot 10^{12}$ $4 \cdot 10^{10}$ $6 \cdot 10^{-1}$ $6 \cdot 10^{-2}$

Таблица 3.5

Минимальная мощность, необходимая для обнаружения сигналов и приема информации на расстоянии 1000 световых лет
 $(\Delta f_1 = \Delta f_2 = \Delta f; \lambda = 3$ см)

№	Передача	Прием информации: $\alpha = 100$		Обнаружение: $\beta = 10$; $\tau_2 = 100$ сек	
		ненаправленный прием, $T_{ш} = 10^6$ К	направленный прием, $S_2 = 10^4$ мк; $T_{ш} = 10^6$ К	ненаправленный прием, $T_{ш} = 10^6$ К	направленный прием, $S_2 = 10^4$ мк; $T_{ш} = 10^6$ К
1	Изотропная непрерывная $\Delta f =$ 1 10^2 10^4 10^6 10^8	$2 \cdot 10^{23}$ $2 \cdot 10^{25}$ $2 \cdot 10^{27}$ $2 \cdot 10^{29}$ $2 \cdot 10^{31}$	$1,5 \cdot 10^{15}$ $1,5 \cdot 10^{17}$ $1,5 \cdot 10^{19}$ $1,5 \cdot 10^{21}$ $1,5 \cdot 10^{23}$	$2 \cdot 10^{21}$ $2 \cdot 10^{22}$ $2 \cdot 10^{23}$ $2 \cdot 10^{24}$ $2 \cdot 10^{25}$	$1,5 \cdot 10^{13}$ $1,5 \cdot 10^{14}$ $1,5 \cdot 10^{15}$ $1,5 \cdot 10^{16}$ $1,5 \cdot 10^{17}$

Продолжение табл. 3.5

№	Передача	Прием информации; $\alpha = 100$		Обнаружение; $\beta = 10$; $\tau_2 = 100$ сек	
		ненаправленный прием, $T_{ш} = 10^6$ К	направленный прием; $S_2 = 10^2$ м ² ; $T_{ш} = 10^6$ К	ненаправленный прием, $T_{ш} = 10^6$ К	направленный прием; $S_2 = 10^4$ м ² ; $T_{ш} = 10^6$ К
2	Изотропная импульсная $t_1 = 1^s$	$2 \cdot 10^{23}$	$1,5 \cdot 10^{15}$	—	—
	1^m	$3 \cdot 10^{21}$	$2,5 \cdot 10^{13}$	—	—
	1^h	$6 \cdot 10^{19}$	$4 \cdot 10^{11}$	—	—
	24^h	$2 \cdot 10^{18}$	$2 \cdot 10^{10}$	—	—
	1 месяц	$8 \cdot 10^{16}$	$6 \cdot 10^8$	—	—
3	1 год	$7 \cdot 10^{15}$	$5 \cdot 10^7$	—	—
	Направленная непрерывная, $\Delta f = 1$	P_1 $g_1 = 10^9$	P_1 $g_1 = 10^9$	P_1 $g_1 = 10^9$	P_1 $g_1 = 10^9$
	10^2	233	152	213	132
	10^4	253	172	223	142
	10^6	273	192	233	152
4	10^8	293	212	243	162
	Направленная импульсная $t_1 = 1^s$	313	232	253	172
	1^m	333	152	—	—
	1^h	215	134	—	—
	24^h	197	116	—	—
	1 месяц	184	103	—	—
	1 год	169	88	—	—
		158	77	—	—

значениям t_1 соответствуют следующие относительные продолжительности импульсов:

$t_1 = 1^s$	1^m	1^h	24^h	1 мес.	1 год
$\frac{\Delta t_1}{t_1} < 10^{-8}$	$2 \cdot 10^{-10}$	$3 \cdot 10^{-12}$	10^{-13}	$4 \cdot 10^{-15}$	$3 \cdot 10^{-16}$

В табл. 3.5 приводятся значения минимальной мощности, необходимой для обеспечения связи на расстоянии 1000 световых лет, которая достигается при условии равенства полос приемника и передатчика. Рассматривая эту таблицу, нетрудно убедиться, что при направленном приеме и передаче для обнаружения сигналов и связи на расстояние 1000 световых лет требуется весьма скромная мощность, особенно при импульсной передаче. С энергетической точки зрения это наиболее выгодный тип связи. Однако обнаружение таких сигналов, когда направление передачи и приема заранее не известны, крайне маловероятно. Для поиска новых абонентов, как уже отмечалось выше, целесообразнее использовать изотропную передачу. При этом, если излучение непрерывно во времени, для обнаружения лучше всего применять направленные антенны, последовательно обследуя различные области неба. В этом случае требуется мощность порядка $10^{13} \div 10^{17}$ вт, что доступно цивилизациям I типа. В случае импульсных сигналов, особенно когда импульсы следуют с большими интервалами, целесообразнее использовать ненаправленную антенну, так как это позволит осуществить непрерывное слежение за всем небом и значительно увеличит вероятность обнаружения. При таком способе связи требуется мощность $10^{15} \div 10^{23}$ вт. Правда, наименьшее из этих чисел соответствует столь медленной передаче информации (одна двоичная единица в год), что здесь уместнее говорить только об обнаружении сигналов, а не о приеме информации. (Впрочем, не исключено, что для позывных скорость передачи информации порядка 1 бита в час или даже в год является приемлемой.) Еще выгоднее для поиска импульсных сигналов использовать систему направленных приемных антенн, которые в совокупности

перекрывают своими лучами все небо. Если каждая антенна имеет площадь $S_2 = 10^4 \text{ м}^2$, то для обнаружения импульсных сигналов потребуется мощность $10^7 \div 10^{15} \text{ вт}$.

После того как цивилизации обнаружат друг друга, между ними может быть установлена взаимно направленная радиосвязь *). В этом случае для передачи информации на расстояние 1000 световых лет со скоростью $1 \div 10^8 \text{ бит/сек}$ требуется мощность порядка $10^6 \div 10^{14} \text{ вт}$, т. е. значительно меньше, чем для обнаружения сигналов. Это на первый взгляд парадоксальное обстоятельство легко может быть понято: высокая мощность, которая требуется для обнаружения сигналов, является своеобразной платой за незнание адреса абонентов.

Попробуем оценить стоимость связи при межзвездных переговорах. Пусть между цивилизациями, расположенными на расстоянии 1000 световых лет, установлена взаимно направленная радиосвязь по высокочастотному каналу с полосой 1 гц. Из табл. 3.5 находим мощность передатчика, которая требуется для такой связи; она составляет 1000 квт, что вполне доступно цивилизации нашего уровня. При этом в точке приема будет обеспечено стократное превышение сигнала над шумом, что вполне достаточно для надежной радиосвязи с помощью двончной ИКМ. Скорость передачи информации по ИКМ с полосой 1 гц составляет 1 бит/сек. Определим, сколько стоит передача сообщения из 100 слов по такому каналу связи. Будем считать, что используется 30-буквенный алфавит и каждое слово состоит в среднем из пяти букв. Тогда наше сообщение содержит $2,5 \cdot 10^3 \text{ бит}$ информации (см. § 1) и для его передачи со скоростью 1 бит/сек потребуется $2,5 \cdot 10^3 \text{ сек}$. При мощности 1000 квт за это время будет израсходовано около

*) Говоря о взаимно направленной радиосвязи, мы имеем в виду не диалог между цивилизациями (такой диалог при межзвездных расстояниях, когда на ответ требуются тысячи лет, вряд ли возможен), а два односторонних «монолог», передаваемых по взаимно направленному каналу.

700 квт-час электроэнергии, стоимость которой 28 руб. (при цене 4 коп. за киловатт-час, значительно превышающей себестоимость). Как видно, связь с внеземными цивилизациями не такое уж дорогое дело. Главное — обнаружить абонента!

На рис. 45 изображена зависимость между минимальной мощностью передатчика и дальностью связи для различных способов приема и передачи. В верхнем левом углу рисунка расположена область изотропного излучения широкополосных сигналов при ненаправленном приеме. Внизу, справа — область, соответствующая направленному излучению узкополосных или импульсных сигналов при условии направленного приема. Между ними — все промежуточные случаи.

Мы предоставляем читателю выбрать на оси ординат то значение мощности, которое соответствует мере его скептицизма и его воображению, и определить оптимальный способ передачи и приема для любой дальности в пределах Вселенной.

Представляет интерес оценить возможность обнаружения ВЦ по их радиоизлучению, если они не посылают специальных сигналов («подслушивание» внутренних радиопередач). В этом случае можно ожидать, что используются мощности порядка $10^5 \div 10^6$ вт. При такой мощности всенаправленные широкоэмиттерные передачи на УКВ невозможно обнаружить даже с расстояния ближайших звезд. Обнаружение остронаправленных радиопередач с коэффициентом направленного действия порядка 10^9 , которые могут применяться внеземными цивилизациями для некоторых специальных целей (например, для межпланетной связи), возможно на расстоянии в несколько сотен и даже тысяч световых лет. Однако при такой высокой направленности вероятность попасть в конус излучения передающей антенны чрезвычайно мала. Вращение планеты увеличивает эту вероятность, но она все-таки остается малой, если учесть, что на расстоянии $100 \div 1000$ световых лет находятся в лучшем случае несколько передающих ВЦ. Таким образом, обнаружение внеземных цивилизаций по их радиоизлучению практически

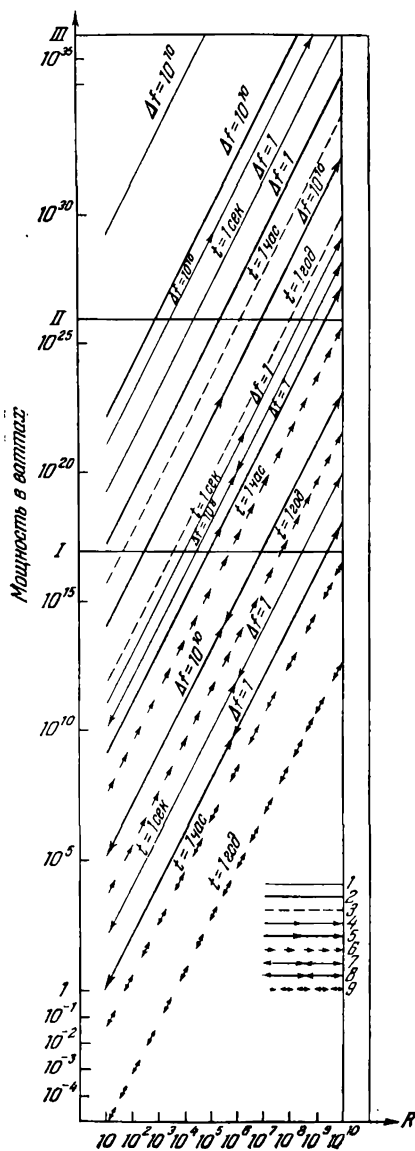


Рис. 45. Зависимость между минимальной мощностью передатчика (при условии $\Delta f_2 = \Delta f_1 = \Delta f$) и дальностью связи в сантиметровом диапазоне радиоволн для различных способов приема и передачи.

невозможно, если они не передают специальных сигналов в виде мощного изотропного излучения или в виде высоконаправленного излучения со сканирующей по всем направлениям антенной *).

Радиосвязь между галактиками. Рассмотрим некоторые особенности межгалактической радиосвязи. Так как можно полагать, что в каждой галактике находится по крайней мере одна цивилизация, способная к передаче и приему информации, то, следовательно, имеет смысл исследование отдельных галактик при помощи направленных приемных и передающих антенн. При этом направленность антенны разумно выбрать таким образом, чтобы ее диаграмма целиком покрывала галактику, т. е. чтобы выполнялось условие $\varphi = \theta$, где φ — угловой размер галактики, а θ — ширина диаграммы направленности антенны. Принимая во внимание зависимость (3.55) между шириной диаграммы и КНД, а также зависимость между расстоянием до галактики и ее угловым размером, перепишем это условие в виде

$$\frac{L^2}{R^2} = \frac{10}{g}, \quad (3.69)$$

где L — средний линейный размер галактики. Пусть, например, цивилизация, расположенная в некоторой галактике, ведет изотропную передачу ($g_1 = 1$), а для обнаружения ее в соседней галактике используется направленная антенна с коэффициентом усиления g_2 , удовлетворяющим условию (3.69). Подставляя эти значения g_1 и g_2 в формулу (3.33в), получим выражение мощности, необходимой для такой связи:

$$P_1 = 16\alpha k T_{\text{ш}} \Delta f \left(\frac{L}{\lambda} \right)^2. \quad (3.70)$$

Парадоксальность этой формулы состоит в том, что потребная мощность не зависит от

*) С точки зрения приема и необходимой мощности эта ситуация равносильна изотропной импульсной передаче.

расстояния! То же самое будет иметь место, если некоторая ВЦ ведет направленную передачу в направлении другой галактики, целиком охватывая ее диаграммой своей антенны ($g_1 = 10 \frac{R^2}{L^2}$), а цивилизация, расположенная в этой галактике, принимает излучение на ненаправленную антенну ($g_2 = 1$). Подсчитаем мощность, которая требуется при таком типе межгалактической радиосвязи. Полагая в (3.70) $\alpha = 100$, $T_{\text{ш}} = 10^\circ \text{ K}$, $\lambda = 3 \text{ см}$, $\Delta f = 1 \text{ гц}$, получим $P_1 = 2 \cdot 10^{26} \text{ вт}$.

При взаимно направленной связи между галактиками коэффициенты усиления приемной и передающей антенн удовлетворяют условию (3.69). Поэтому мощность, необходимая для связи, будет равна

$$P_1 = \frac{16akT_{\text{ш}} \Delta f L^4}{10\lambda^2 R^2}. \quad (3.71)$$

Получается еще более странный вывод: чем больше расстояние между галактиками, тем меньше мощность, которая требуется для связи между ними! На эти любопытные особенности межгалактической связи обратил внимание В. С. Троицкий [6]. Они связаны с условием (3.69), т. е. выполняются лишь постольку, поскольку направленность антенн изменяется в соответствии с изменением угловых размеров Галактики (или площадь каждой антенны возрастает пропорционально расстоянию между ними). Какие параметры антенн и какая мощность требуется для направленной межгалактической радиосвязи, можно судить на основе данных таблицы 3.6, где приводятся КНД, площадь приемной антенны и необходимая мощность для связи с двумя соседними галактиками: Большим Магеллановым Облаком и туманностью Андромеды, а также некоторой типичной галактикой с размерами порядка 10^5 световых лет. При этом, как и ранее, принималось $\alpha = 100$, $T_{\text{ш}} = 10^\circ \text{ K}$, $\Delta f_1 = \Delta f_2 = \Delta f = 1 \text{ гц}$.

Монохроматические сигналы. Поиск по частоте. Мы рассмотрели проблему связи с энергетической точки зрения. Остановимся более подробно на значе-

Таблица 3.6

КНД приемной антенны и необходимая мощность
при направленной связи между галактиками

Галактики	Б. Магелланово Облако	Туманность Андромеды	Типичная галактика с размерами $L = 10^5$ световых лет			
Расстояние в световых годах	$2 \cdot 10^5$	$2 \cdot 10^6$	10^7	10^8	10^9	10^{10}
Угловые размеры	9'	3° 5'	34'	206''	21''	2''
КНД приемной антенны . . .	360	$3 \cdot 10^3$	10^5	10^7	10^9	10^{11}
Площадь (m^2) для $\lambda = 3,5$ см . .	0,04	0,3	10	10^3	10^5	10^7
Мощность передатчика, вт . .	$6 \cdot 10^{25}$	10^{-1}	$2 \cdot 10^{22}$	$2 \cdot 10^{20}$	$2 \cdot 10^{18}$	$2 \cdot 10^{16}$

нии полосы сигнала. Здесь следует различать два случая: широкополосные сигналы с практически непрерывным спектром ($\Delta\nu \sim \nu$) и узкополосные, монохроматические сигналы с полосой много меньшей, чем частота ($\Delta\nu \ll \nu$).

Важнейший этап при обнаружении узкополосных сигналов состоит в поисках сигналов по частоте. Как уже отмечалось выше, при малой скорости передачи информации (например, в случае позывных) ширина полосы частот определяется только стабильностью передаваемого сигнала и может составлять доли герца. Между тем ширина оптимального диапазона спектра, где следует ожидать сигналов ВЦ, порядка $10^{10} \div 10^{11}$ гц. Таким образом, задача состоит в том, чтобы в этом широком диапазоне обнаружить узкую линию излучения с относительной шириной $10^{-11} \div 10^{-10}$. Так как направление поиска тоже не известно, то задача, по остроумному выражению Парселла, похожа на попытку встретиться с нужным человеком в Нью-Йорке, не договорившись о месте встречи. Тем не менее, эта сложная задача принципиально и технически выполнима. Остановимся сначала на поисках сигнала по частоте, а затем рассмотрим поиск по направлению.

Существует два метода поисков сигнала по частоте: одноканальный поисковый приемник с автоматически перестраиваемой частотой настройки и многоканальный приемник с узкополосными фильтрами, каждый из которых настроен на определенную частоту, и которые в совокупности перекрывают весь подлежащий исследованию диапазон спектра. Возникает вопрос, какой из этих методов лучше? Чтобы ответить на него, надо воспользоваться каким-то разумным критерием оценки. В. И. Сифоров [8] предложил следующий критерий. Для обнаружения сигналов ВЦ

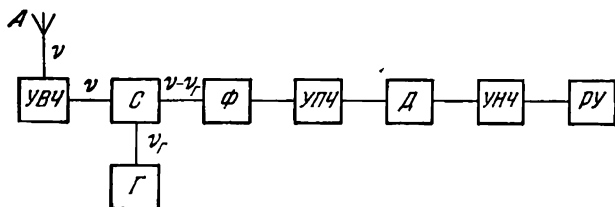


Рис. 46. Блок-схема одноканального приемника для поиска сигналов по частоте.

A — антенна, $УВЧ$ — усилитель высокой частоты, $Г$ — гетеродин, $С$ — смеситель, $Ф$ — фильтр, $УПЧ$ — усилитель промежуточной частоты, $Д$ — детектор, $УНЧ$ — усилитель низкой частоты, $РУ$ — регистрирующее устройство.

в конечном итоге необходимо принять определенное количество информации, на основании которой с высокой достоверностью можно было бы заключить об искусственной природе обнаруженного источника радиоизлучения. Естественно поэтому отдать предпочтение тем методам обнаружения сигналов, которые позволяют за минимальное время получить минимально необходимое количество информации. Рассмотрим с этой точки зрения указанные методы поиска сигнала по частоте.

Пусть в нашем распоряжении имеется одноканальный поисковый приемник с непрерывно изменяющейся частотой настройки, блок-схема которого изображена на рис. 46. Частота гетеродина ν_r изменяется в соответствии с изменением настройки усилителя высокой частоты. В смесителе происходит преобразование те-

кущей высокой частоты ν в постоянную промежуточную частоту $\nu_{\Pi} = \nu - \nu_{\Gamma}$. После смесителя сигнал фильтруется по промежуточной частоте, усиливается, детектируется и через усилитель низкой частоты поступает на регистрирующее устройство. Полоса приемника Δf_2 определяется фильтром промежуточной частоты: $\Delta f_2 = \Delta f_{\Pi}$. Обозначим через Δt промежуток времени, в течение которого текущая переменная частота ν проходит интервал частот Δf_2 . В результате прохождения этого интервала на вход фильтра поступает некоторый сигнал $x(t)$ длительностью Δt , полоса которого $\Delta f = \Delta t^{-1}$. Для того чтобы этот сигнал мог без искажения пройти через фильтр, необходимо выполнение очевидного условия: $\Delta f \leq \Delta f_2$ или $\Delta t \geq \Delta f_2^{-1}$. Отсюда следует, что скорость изменения частоты зависит от полосы приемника Δf_2 :

$$\frac{d\nu}{dt} = \frac{\Delta f_2}{\Delta t} \leq (\Delta f_2)^2. \quad (3.72)$$

Эта скорость ограничена сверху, поэтому время поиска при заданной полосе частот нельзя сделать произвольно малым. Максимальная скорость изменения частоты

$$\left(\frac{d\nu}{dt} \right)_{\max} = (\Delta f_2)^2, \quad (3.73)$$

а соответствующее ей минимальное время прохождения интервала частот Δf_2 равно

$$\Delta t_{\min} = \Delta f_2^{-1}. \quad (3.74)$$

Обозначим через Δf_0 диапазон частот, в котором решено производить поиск. Тогда полное время поиска, необходимое для прохождения текущей частотой настройки всего обследуемого диапазона, будет

$$\tau_{\Pi} = N \Delta t_{\min} = \frac{\Delta f_0}{(\Delta f_2)^2}, \quad (3.75)$$

где N — число элементарных полос, т. е. полос приемника Δf_2 , которое содержится в исследуемом диапазоне спектра. Мы получили, что время поиска по

частоте обратно пропорционально квадрату полосы частот одноканального поискового приемника. Это вполне понятно, так как с уменьшением полосы приемника, с одной стороны, увеличивается число элементарных полос, на которые разбивается подлежащий исследованию диапазон спектра, а с другой, увеличивается время обследования каждой из элементарных полос (потому что скорость изменения частоты уменьшается пропорционально квадрату сокращающейся полосы частот). Выразим время поиска в зависимости от расстояния до источника. Так как $\Delta f_2 \propto R^{-2}$, то (3.75) можно переписать в виде

$$\tau_{\text{п}} \propto \Delta f_0 R^4. \quad (3.76)$$

Итак, при выбранном методе приема, когда используется одноканальный поисковый приемник, время поиска монохроматических сигналов ВЦ оказывается прямо пропорциональным полной полосе частот Δf_0 , в которой решено производить поиск, и четвертой степени расстояния до цивилизации.

Если используется многоканальный приемник, состоящий из N каналов с полосой Δf_2 в каждом канале, которые в совокупности перекрывают весь подлежащий исследованию диапазон спектра ($N\Delta f_2 = \Delta f_0$), то полное время поиска по частоте равняется времени поиска внутри одного канала. Из формулы (3.74) следует, что в этом случае оно будет обратно пропорционально полосе частот каждого канала или, учитывая зависимость между дальностью и полосой, пропорционально квадрату расстояния между цивилизациями.

Пусть, например, поиск производится в диапазоне 3 см, расстояние до цивилизации равно 1000 световых лет, мощность передатчика 150 Мвт, коэффициент усиления передающей антенны 90 дб, эффективная площадь приемной антенны 10^4 м^2 , шумовая температура 10° К и пусть для приема информации необходимо обеспечить стократное превышение полезного сигнала над шумом. Из табл. 3.5 находим, что в этом случае полоса приемника должна составлять 100 гц. Минимальное время поиска внутри этого интервала по-

рядка 0,01 сек. Если используется одноканальный поисковый приемник, то для обследования всего диапазона спектра в районе 3 см ($\Delta f_0 = 10^{10}$ гц) потребуется несколько лет. С многоканальным приемником, в состав которого входят 10^8 каналов с полосой 100 гц в каждом канале, время поиска будет порядка 0,01 сек. Если расстояние до цивилизации увеличить в 10 раз, время поиска с многоканальным приемником возрастет в 100 раз и достигнет 1 сек, а при использовании одноканального поискового приемника оно увеличится в 10^4 раз и составит около 30 000 лет.

Таким образом, время поиска, которое требуется при использовании одноканального приемника с непрерывным смещением частоты, много больше, чем время поиска с многоканальным приемником. Кроме того, с увеличением расстояния между цивилизациями время поиска с одноканальным приемником растет во много раз быстрее ($\tau_{\text{п}} \propto R^4$), чем возрастает время поиска при использовании многоканальных систем ($\tau_{\text{п}} \propto R^2$).

Все это делает нецелесообразным применение одноканального поискового приемника для обнаружения монохроматических сигналов ВЦ. Применяя многоканальные приемники с большим числом узкополосных фильтров, эту задачу можно решить более успешно. При этом, однако, следует иметь в виду, что сокращение времени поисков достигается за счет усложнения приемной аппаратуры, причем сложность аппаратуры (число каналов в приемнике) возрастает пропорционально R^2 . Тем не менее, можно полагать, что такое усложнение целесообразно [8], так как полосовые фильтры (и другие элементы, входящие в состав многоканальных систем) являются простыми и дешевыми элементами.

Пример многоканальной системы, предназначенной для обнаружения монохроматических сигналов ВЦ, предложен В. А. Котельниковым [16]. Блок-схема приемника приведена на рис. 47. Здесь A — антенна, $У$ — усилитель, в котором происходит усиление и, если нужно, преобразование частоты принимаемых

сигналов, Φ — фильтры с полосой Δf , перекрывающие в совокупности весь исследуемый диапазон спектра, D — детекторы, I — интеграторы, суммирующие энергию, прошедшую через фильтр за время посылки длительностью τ , Π — пороговые устройства, которые дают сигнал на выходе только в том случае, когда энергия, прошедшая через фильтр за время τ , превысит некоторое пороговое значение.

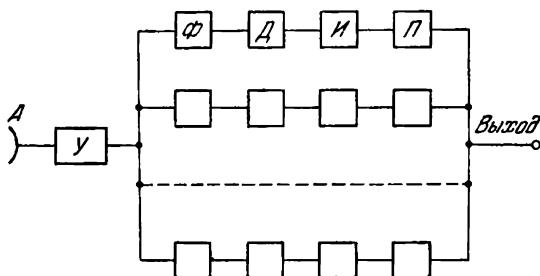


Рис. 47. Многоканальный приемник В. А. Котельникова для поиска сигналов по частоте.

A — антенна, $У$ — усилитель, Φ — узкополосные фильтры, D — детекторы, I — интеграторы, Π — пороговые устройства.

Очевидно, такой приемник невозможно использовать для приема информации, если при передаче применяется одна из систем амплитудной модуляции. Можно, однако, передавать информацию, меняя частоту сигнала от посылки к посылке. Тогда в одном из каналов приемника будет появляться сигнал, причем с каждой новой посылкой появление сигнала будет регистрироваться в новом канале.

Появление сигнала в данном канале можно рассматривать как определенное сообщение. Так как при этом делается выбор из N возможностей, то каждое сообщение будет содержать $\log_2 N$ единиц информации (см. § 1). Следовательно, скорость передачи информации в данной системе связи равна

$$q = \frac{\log_2 N}{\tau}. \quad (3.77)$$

Определим теперь дальность связи. При взаимно направленной радиосвязи дальность приема сигналов с помощью приемника В. А. Котельникова равна

$$R = \left(\frac{P_1 g_1 S_2 \tau}{4\pi k T_{\text{ш}} \Psi} \right)^{1/2}. \quad (3.78)$$

Величина Ψ , стоящая в знаменателе этой формулы, зависит от числа N каналов в приемнике, полосы Δf каждого канала и от установки порога, который в свою очередь определяется вероятностью ложного срабатывания системы $p_{\text{л.с.}}$ (вследствие случайных помех в приемнике) и вероятностью пропуска сигнала $p_{\text{пр}}$. Дальность связи возрастает с уменьшением Ψ . Минимальное значение Ψ достигается при условии $\Delta f = \frac{1}{\tau}$; оно равно

$$\Psi_{\text{min}} = \left(\sqrt{\ln \frac{N}{p_{\text{л.с.}}} - 2} + \sqrt{\ln \frac{1}{p_{\text{пр}}} - 2} \right)^2. \quad (3.79)$$

Этому значению Ψ при заданном τ соответствует максимальная дальность связи

$$R_{\text{max}} = \left(\frac{P_1 g_1 S_2}{\Psi_{\text{min}} 4\pi k T_{\text{ш}} \Delta f} \right)^{1/2}. \quad (3.80)$$

Полученная формула аналогична формуле (3.33). Разница состоит только в том, что в знаменателе вместо α стоит величина Ψ_{min} , имеющая тот же смысл, что и α , так как она определяется порогом системы. Из двух последних формул следует, что максимальная дальность связи падает с увеличением числа каналов в приемнике и расширением полосы частот каждого канала. Если задан интервал спектра Δf_0 , в котором решено производить поиск, то $\Delta f = \frac{\Delta f_0}{N}$. С ростом N полоса частот каждого канала сокращается быстрее, чем растет Ψ_{min} ; в результате максимальная дальность связи с ростом числа каналов увеличивается. При условии $\Delta f \gg \frac{1}{\tau}$

$$\Psi = (2\Psi_{\text{min}}\tau\Delta f)^{1/2} \quad (3.81)$$

и при заданном интервале спектра Δf_0 дальность связи равна

$$R = \left(\frac{P_1 g_1 S}{4\pi k T_{ш}} \sqrt{\frac{\tau N}{2\Psi_{\min} \Delta f_0}} \right)^{1/2}. \quad (3.82)$$

Рассмотрим следующий пример. Пусть передатчик ВЦ мощностью 100 Мвт посылает монохроматические сигналы в диапазоне 3 см в виде импульсов продолжительностью 100 сек, меняя частоту импульсов от посылки к посылке. Предположим, что для передачи используется антенна с коэффициентом усиления 10^9 , а для приема антенна с эффективной площадью 10^4 м² и приемник В. А. Котельникова с $N=10^9$ каналами, шумовой температурой $T_{ш}=10^\circ$ К, временем накопления $\tau=100$ сек, полосой каждого канала 1 гц, причем вероятность ложного срабатывания и вероятность пропуска сигнала равны 10^{-5} . Подставляя эти значения N , τ , Δf , $\rho_{л.с}$ и $\rho_{пр}$ в (3.79) и (3.81), найдем $\Psi=120$. После этого по формуле (3.78) получаем расстояние $R=7 \cdot 10^4$ световых лет. Информативность же данной системы связи будет $0.01 \log_2 10^9 = 0,3$ бит/сек. С уменьшением τ дальность связи медленно убывает, а информативность довольно быстро растет. Если в нашем примере положить $\tau=1$ сек, то $\Psi=\Psi_{\min}=73$, $R=9 \cdot 10^3$ световых лет, а информативность будет равна 30 бит/сек.

Интересно сравнить эти числа с соответствующими данными для одноканального приемника, работающего на заданной частоте. Пользуясь рис. 45, найдем, что при $P_1=10^8$ вт, $g_1=10^9$, $S_2=10^4$ м², $T_{ш}=10^\circ$ К, $\Delta f=1$ гц и $\alpha=100$ дальность взаимно направленной радиосвязи составляет $8 \cdot 10^3$ световых лет, информативность равна 1 бит/сек, т. е. в 30 раз меньше, чем в системе В. А. Котельникова при той же дальности. Дальность обнаружения при рассмотренных параметрах системы и $\beta=10$, $\tau_2=100$ сек составляет $8 \cdot 10^4$ световых лет, т. е. она одного порядка с дальностью связи, которую обеспечивает многоканальный приемник при скорости передачи информации 0,3 бит/сек и прочих равных условиях.

Итак, мы можем заключить, что многоканальный приемник В. А. Котельникова является оптимальной системой для поиска монохроматических сигналов, когда полоса частот исследуемой области спектра не слишком велика. Однако следует признать, что число каналов, требующееся, чтобы обнаружить линию шириною менее 1 гц в полосе частот 10^{10} — 10^{11} , все же непомерно велико. Чтобы избежать этой трудности, В. С. Троицкий предложил использовать оригинальный комбинированный метод. Суть его состоит в том, что с помощью специального спектроанализатора исследуется сразу достаточно широкий интервал спектра с полосой Δf_0 порядка 1 Мгц. При этом выясняется, присутствует или нет в данном широком интервале спектра монохроматический, синусоидальный сигнал. Частота сигнала не определяется, фиксируется лишь широкий интервал частот Δf_0 , внутри которого он содержится. После того как этот интервал частот зафиксирован, для точного определения частоты сигнала и для приема информации можно применить многоканальный приемник. При полосе канала 1 гц для перекрытия интервала $\Delta f_0 = 1$ Мгц, внутри которого был обнаружен сигнал, потребуется 10^6 каналов. Это число, по-видимому, можно считать приемлемым. Кроме того, сооружение многоканального приемника, перекрывающего данный интервал частот, будет оправдано обнаружением сигнала внутри этого интервала.

Целесообразно в первую очередь исследовать таким способом участки спектра вблизи линии водорода $\lambda = 21$ см, вблизи ее гармоник, вблизи линий гидроксила ОН $\lambda = 18$ см, а также, возможно, вблизи линий аммиака $\lambda = 1,25$ см и формальдегида $\lambda = 0,4$ см, применяемых в молекулярных генераторах.

Поиск сигналов по направлению. Перейдем к задаче поиска сигналов по направлению. Предположим, что расстояние до ближайшей цивилизации, посылающей радиосигналы в пространство, не

превышает некоторой величины R . Тогда можно обнаружить эти сигналы, исследуя звезды, расположенные около Солнца в сфере радиуса R . Сколько звезд придется исследовать таким способом? Среднее расстояние между звездами в окрестностях Солнца приблизительно равно 2,2 парсека (т. е. около семи световых лет). Следовательно, звездная плотность составляет здесь 0,1 звезды на кубический парсек или 0,003 звезды на кубический световой год. Пусть $R = 1000$ световых лет. В сфере такого радиуса содержится 10 млн. звезд. Число звезд, подлежащих исследованию, можно сократить, если принять во внимание, что лишь небольшая их доля (не более 1 %) может иметь планетные системы с подходящими для жизни условиями. Возникает сложная и, главное, весьма неопределенная задача: выбрать из 10 млн. звезд несколько сотен тысяч таких, около которых можно подозревать наличие высокоразвитых цивилизаций. Как ни странно, значительно проще дело обстоит с поисками цивилизаций в других галактиках; (этот вопрос мы рассматривали выше). Вернемся, однако, к звездам.

Оптимальной системой для поиска сигналов, проходящих с неизвестного направления, является система направленных неподвижных антенн, которые своими лучами охватывают весь небосвод. При этом, если передающая ВЦ излучает сигналы по всем направлениям в пространстве, их обнаружение не составит труда. Однако, как мы видели выше (см. табл. 3.5), при такой изотропной передаче требуется огромная мощность передатчика и высокая направленность приемных антенн. Например, при расстоянии до цивилизации, составляющем 1000 световых лет, и полосе передатчика, равной 1000 $M\mu$, требуется мощность порядка 10^{24} вт (что доступно только цивилизациям II типа) и приемная антенна с эффективной площадью 10^4 m^2 при $T_{\text{ш}} = 10^\circ K$. Такая антенна в сантиметровом диапазоне имеет направленность 10^8 , поэтому для перекрытия всего небесного свода понадобится 100 млн. антенн. Несомненно, создание подобной системы обнаружения в перспективе ближай-

ших 100 лет возможно. Однако в настоящее время осуществление этого проекта лежит за пределами финансовых возможностей человечества.

Впрочем, требования к площади и числу приемных антенн, так же как и к мощности передатчика, могут быть значительно снижены, если при передаче применяется направленная антенна. Следуя В. А. Котельникову, рассмотрим две цивилизации, A и B , расположенные на расстоянии R друг от друга. Цивилизация A ведет передачу с помощью остронаправленной антенны, а цивилизация B работает на прием. Очевидно, цивилизации A заранее не известно, куда направить сигналы. Поэтому она должна лучом своей антенны последовательно «обшарить» все небо. Пусть τ — длительность посылки сигнала, а ω — телесный угол антенны ($\omega = \frac{4\pi}{\rho}$). Тогда для обхода небесной сферы потребуется время

$$t_0 = \frac{4\pi}{\omega} \tau = g\tau. \quad (3.83)$$

Допустим, что цивилизация B располагает системой обнаружения из совокупности направленных антенн, перекрывающих небесную сферу. Одна из них направлена на цивилизацию A . Приемник, связанный с этой антенной, регистрирует сигнал в момент, когда передающая антенна цивилизации A окажется направленной на цивилизацию B . Эксперимент по обнаружению сигнала должен длиться в течение времени t , значительно превышающего t_0 . Тогда за время проведения эксперимента сигнал будет зарегистрирован несколько раз через равные промежутки времени t_0 . Это позволит уверенно отличить его от случайных помех. Время обнаружения можно значительно сократить, если цивилизация A вместо обхода всего небесного свода выберет подходящие звезды, расположенные в сфере радиуса R (среди которых есть, конечно, и цивилизация B), и будет посылать сигналы только в направлении этих звезд, быстро переводя антенну с одной звезды на другую.

Пусть, например, цивилизация А расположена на расстоянии 1000 световых лет, мощность передатчика 10^{17} вт (что доступно цивилизации I типа), $\Delta f = 1000$ Мгц, направленность передающей антенны $g = 10^9$, $\lambda = 3,5$ см, длительность посылки $\tau = 3$ сек. Тогда по формуле (3.83) получим, что для обхода небесной сферы требуется $3 \cdot 10^9$ сек или 100 лет. Время обхода всех звезд, расположенных в сфере радиусом 1000 световых лет, равно $3 \cdot 10^7$ сек = 1 году. Если допустить, что облучаются только подходящие звезды, у которых можно ожидать присутствия цивилизаций, и принять, что доля таких звезд составляет 1%, то полное время облучения составит $3 \cdot 10^5$ сек или около 3,5 суток. При мощности передатчика 10^{17} вт и полосе 1000 Мгц для обнаружения сигнала на расстоянии 1000 световых лет требуется довольно скромная приемная антенна с эффективной площадью около 100 м² (при $T_{ш} = 10^\circ$ К). Чтобы одновременно охватить всю небесную сферу, потребуется 1 млн. таких антенн. В. А. Котельников предложил применять многолучевые антенны (технически это возможно, так как каждая антенна неподвижна), тогда число необходимых антенн можно сократить по крайней мере на порядок. Далее, для сокращения числа антенн он рекомендует, если позволяет время, разбить небесную сферу на несколько частей и исследовать их последовательно. Так, в нашем примере, когда время обхода всех подходящих звезд около 3,5 суток, можно разбить небесную сферу на 10 частей и обследовать каждый участок в течение 36 дней (за это время сигнал должен появиться не менее десяти раз), тогда весь эксперимент будет закончен за 1 год. При этом для охвата одного участка потребуется 10^4 десятилучевых антенн. Если $R = 100$ световых лет, то время обхода всех подходящих звезд составляет 300 сек. Мы можем разделить небосвод на 10^4 частей и, обследуя каждую часть в течение 3000 сек (за это время сигнал появится 10 раз), закончить эксперимент за 1 год. Площадь приемной антенны, необходимая для того, чтобы обнаружить сигнал на расстоянии 100 световых лет, равняется 1 м², а КНД составляет 10^4 , т. е.

для охвата каждого участка достаточно только одной антенны! Если $R=10^4$ световых лет, то время обхода подходящих звезд составит $3 \cdot 10^8$ сек или около 10 лет. Исследование небесного свода по частям в этом случае сильно удлинит эксперимент. Для полного же охвата неба при данном расстоянии до цивилизации потребуется 10^8 антенн по 10^4 м² каждая. Таким образом, при заданной мощности и полосе передатчика, направленности передающей антенны и длительности облучения τ существует оптимальное расстояние до цивилизации, при котором разумно создавать многоантенную систему обнаружения. В нашем примере это расстояние порядка 1000 световых лет.

Сказанное выше справедливо как для поисков широкополосных сигналов, так и монохроматических сигналов. Разница состоит только в том, что в случае монохроматических сигналов с каждой антенной системы обнаружения должен быть связан многоканальный приемник. Пусть, например, приемник В. А. Котельникова имеет 10^9 каналов, $T_{ш}=30^\circ \text{ К}$, $\Delta f=0,3$ гц, а передающая цивилизация, обладая передатчиком мощностью 10^9 вт и антенной площадью 10^5 м², работает на волне 10 см, посылая монохроматические сигналы длительностью $\tau=3$ сек. Тогда для

Таблица 3.7

Параметры системы обнаружения монохроматических сигналов

ВЦ ($P_1=10^9$ вт, $S_1=10^5$ м², $\lambda=10$ см, $\tau=3$ сек, $N=10^9$, $\Delta f=0,3$ гц; $T_{ш}=30^\circ \text{ К}$) согласно В. А. Котельникову [16]

Расстояние в световых годах	Число звезд в сфере радиуса R	Время обхода всех звезд	Время обхода подходящих звезд (1% всех звезд)	Площадь приемной антенны, м ²	Число прием- ных каналов на весь небо- свод	Число частей, на которые можно раз- бить небо- свод	Число прием- ных каналов при делении небосвода
2000	10^8	10 лет	36 дней	400	480 000	1	480 000
1000	10^7	1 год	4 дня	100	120 000	10	12 000
500	10^6	36 дней	9 час	25	30 000	100	300
200	10^5	4 дня	1 час	4	4 800	1000	5

обнаружения сигналов потребуется многоантенная система, параметры которой приведены в табл. 3.7.

На основе этих данных В. А. Котельников пришел к выводу, что обнаружение радиосигналов от цивилизаций нашего (или несколько более высокого уровня) вполне реально, если одна такая цивилизация приходится на 10^6 звезд. Если одна цивилизация приходится на 10^7 звезд, то обнаружить ее значительно труднее, но при определенных усилиях все же можно. Если одна цивилизация приходится на 10^8 звезд, то обнаружить ее современными средствами крайне затруднительно.

Широкополосные сигналы. Проведение обзоров неба. Рассмотрим задачу поиска широкополосных сигналов. При ширине полосы порядка излучаемой частоты искусственный сигнал похож на радиоизлучение естественных источников. Отсюда можно сделать два вывода. Во-первых, для обнаружения широкополосных сигналов можно использовать обычную радиоастрономическую аппаратуру. Во-вторых, для обнаружения широкополосных сигналов необходимо установить, чем должен отличаться искусственный сигнал от естественного. Ведь прежде чем пытаться получить информацию путем расшифровки полученных сигналов, мы должны убедиться в том, что имеем дело с искусственным источником, суметь выделить его из множества естественных источников радиоизлучения. Это уже знакомая нам проблема критериев искусственного сигнала, которая рассматривалась в § 3.

Систематические поиски искусственных источников должны включать в себя прежде всего выявление всех источников радиоизлучения и последующий отбор согласно ожидаемым критериям искусственности. С этой целью необходимо проведение полных обзоров неба в радиодиапазоне. В настоящее время сравнительно хорошо изучен диапазон метровых и дециметровых радиоволн. В этом диапазоне составлены достаточно подробные каталоги, содержащие все источники с потоками вплоть до 10^{-26} вт/м² гц. Гораздо

хуже обстоит дело в сантиметровом диапазоне спектра. Здесь до сих пор не проводилось ни одного полного обзора неба, а между тем именно этот диапазон является оптимальным для связи между цивилизациями. Поэтому проведение полного обзора неба в сантиметровом диапазоне спектра с помощью чувствительной радиоастрономической аппаратуры надо рассматривать как первоочередную задачу при поисках искусственных источников.

Каким требованиям должен удовлетворять радиотелескоп, предназначенный для такого обзора? Совершенно естественно стремиться к тому, чтобы полное время обзора было минимальным. Вместе с тем радиотелескоп должен иметь предельно высокую чувствительность и, следовательно, максимально возможную площадь приемной антенны. Для обозрения неба удобно использовать суточное вращение Земли. Пусть в нашем распоряжении имеется радиотелескоп, у которого ось антенны фиксирована в плоскости меридиана. За счет суточного вращения Земли в диаграмму антенны будут попадать различные участки небесной сферы, расположенные на одной суточной параллели. За 24 часа телескоп обследует кольцевую полосу неба шириной $2\theta_h$, где $2\theta_h$ — вертикальный размер диаграммы направленности по уровню половинной мощности. Переместим теперь ось антенны на величину $2\theta_h$ по склонению, тогда в следующие сутки будет обследована новая кольцевая зона, примыкающая к первой и имеющая ту же ширину $2\theta_h$. Очевидно, что полное время обзора всего неба будет равно

$$T_0 = \frac{\pi}{2\theta_h} \approx \frac{\pi h}{\lambda} \text{ суток,} \quad (3.84)$$

где h — размер отражающей поверхности радиотелескопа в вертикальной плоскости. Итак, время обзора сокращается с увеличением θ_h или с уменьшением вертикального размера h отражающей поверхности. При заданной площади поверхности для обеспечения минимального времени

обзора вертикальный размер отражающей поверхности должен быть много меньше горизонтального, а вертикальный размер диаграммы направленности много больше горизонтального ($\theta_h \gg \theta_l$), т. е. радиотелескоп должен иметь «ножевую» диаграмму направленности.

Пусть, например, имеется радиотелескоп с антенной в виде параболоида вращения диаметром 50 м (площадь поверхности 2000 м²). Такая антенна имеет симметричную диаграмму, размер которой на волне 1 см равен $2\theta_h = 2\theta_l = 2\theta_{0.5} = 2 \cdot 10^{-4}$ радиан = 40". Подставляя это значение θ_h в (3.84), получим полное время обзора $T_0 = 43$ года. Рассмотрим теперь антенну в виде параболического цилиндра с горизонтальным раскрытием $l = 400$ м и высотой $h = 5$ м. Такая антенна при той же геометрической площади 2000 м² будет иметь «ножевую» диаграмму с размерами $2\theta_l = 5''$ и $2\theta_h = 7'$ (на волне 1 см), а полное время обзора составит 4,3 года. Из этого примера видно, что радиотелескоп с «ножевой» диаграммой не только значительно сокращает время обзора, но и дает возможность реализовать высокую разрешающую способность (по одной координате). При рассмотрении предельно больших радиотелескопов, для которых существенную роль начинают играть эффекты, связанные с флуктуациями радиояркости метагалактического фона и атмосферы, выявляется еще одно важное преимущество телескопов такого типа: они дают возможность реализовать максимальную чувствительность при заданной площади.

В настоящее время известны два типа радиотелескопов с «ножевой» диаграммой направленности: радиотелескоп системы Н. Л. Кайдановского и С. Э. Хайкина с антенной переменного профиля АПП и радиотелескоп системы Краусса. На рис. 48 приведена фотография большого Пулковского радиотелескопа БПР с антенной переменного профиля. Телескоп состоит из отдельных щитов, установленных по дуге окружности. Каждый щит может перемещаться вдоль радиуса окружности, поворачиваться по ази-

муту и по углу места. Соответствующим перемещением отдельных щитов можно перестроить отражающую поверхность таким образом, чтобы направить радиотелескоп в заданную точку неба. Горизонтальный размер диаграммы направленности определяется горизонтальным раскрывом антенны (длиной хорды,



Рис. 48. Большой Пулковский радиотелескоп с антенной переменного профиля.

стягивающей рабочий сектор); вертикальный размер при работе вблизи горизонта определяется высотой щитов. С увеличением угла места вертикальный размер диаграммы сокращается и в зените (при замкнутом круге АПП) становится равным горизонтальному; диаграмма из «ножевой» превращается в «карандашную». Этот эффект приводит к увеличению времени обзора. С точки зрения обзора более выгодной оказывается система Краусса (рис. 49). Радиотелескоп системы Краусса состоит из двух отражающих поверхностей: неподвижного параболического рефлектора, оптическая ось которого лежит в плоскости меридиана, и подвижного плоского рефлектора, который может вращаться вокруг горизонтальной оси, обеспечивая проведение наблюдений при

различных углах места в меридиане. Этот радиотелескоп имеет «ножевую» диаграмму, вертикальный размер которой определяется высотой параболического рефлектора и не зависит от угла места источника. Несколько модифицированный радиотелескоп такого

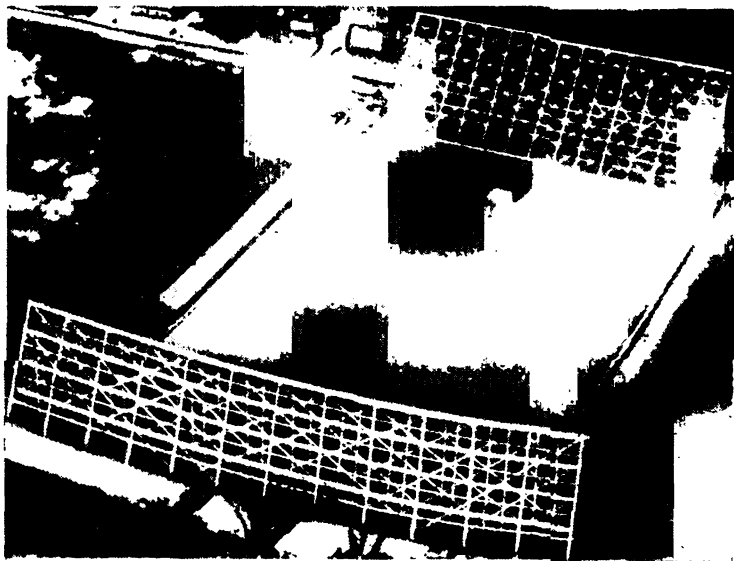


Рис. 49. Радиотелескоп системы Краусса (США).

типа, работающий на волне 21 см, построен недавно во Франции (рис. 50).

После того как обзор неба будет закончен, необходимо произвести отбор источников по каким-либо признакам. По-видимому, наиболее удобно осуществить отбор по признаку угловых размеров. Можно, например, поставить задачу выявления всех источников с угловыми размерами меньше $0'',1$ и потоком до 10^{-27} вт/м²гц (отбросив как заведомо естественные источники больших угловых размеров). Эта задача может быть решена с помощью чувствительных



Рис. 50. Большой радиотелескоп в Нанси (Франция), работающий на волне 21 см. Эффективная площадь антенны 7000 м^2 , размер диаграммы по горизонтали $3'5$, по вертикали $20'$. Приемник радиотелескопа имеет 15 каналов с полосой 280 кГц в каждом канале.

радиоинтерферометров, состоящих из больших антенн с площадью $10^3 \div 10^4 \text{ м}^2$, разнесенных на расстояние порядка $10^6 \div 10^7$ длин волн (в сантиметровом диапазоне).

Отбор источников с угловыми размерами меньше $0'',1$ можно рассматривать только как предварительный этап работы. Использование радиоинтерферометров со сверхдлинной базой на основе существующей сети радиотелескопов (глобальная система интерферометров) позволяет реализовать разрешающую способность порядка $0'',001$ (разрешающая способность $0'',005$ уже достигнута). В перспективе можно говорить о создании на той же основе радиоинтерферометра в системе Земля—ракета. Это позволит использовать базы порядка 1 а. е. и добиться разрешающей способности порядка 10^{-8} сек дуги в сантиметровом диапазоне. Отобранные источники необходимо тщательно изучить в соответствии с ожидаемыми критериями искусственного сигнала. Здесь открывается обширное поле для исследований.

С точки зрения радиоастрономии искусственный источник радионизлучения должен обладать какими-то особыми необычными свойствами, т. е. искусственный источник заведомо является пекулярным. Задача обнаружения и исследования пекулярных источников космического радионизлучения — важнейшая задача радиоастрономии. Таким образом, проблема обнаружения внеземных цивилизаций при всей своей специфичности смыкается в этой части с актуальными задачами радиоастрономии.

ЛИТЕРАТУРА

1. Cocconi G. and Morrison P., Nature **184**, 844, 1959.
2. «Межзвездная связь», сб. статей, под ред. А. Камерона, пер. с англ., «Мир», 1965.
3. «Внеземные цивилизации», Тр. совещания, Бюракан, 20—23 мая 1964 г., изд. АН Арм. ССР, 1965.
4. Hartley L. V. L., Transmission of Information, BSTJ **7**, 3, 535—563, 1928. Русский перевод: Р. Л. В. Хартли, Передача информации, в сб. «Теория информации и ее приложения», Физматгиз, 1959, стр. 5—35.

5. Shannon C. E., Communication in the presence of noise, PIRE 37, 1, 10—21, 1949. Русский перевод: К. Шеннон «Связь при наличии шума», в сб. «Теория информации и ее приложения», Физматгиз, 1959, стр. 82—112.
6. Троицкий В. С., Некоторые соображения о поисках разумных сигналов из Вселенной В сб. [3], стр. 97—112.
7. Уэбб Д. ж., Обнаружение разумных сигналов из космического пространства. В сб. [2], стр. 195—211.
8. Сифоров В. П., Некоторые вопросы поиска и анализа радиоизлучений от других цивилизаций. В сб. [3], стр. 121—128.
9. Шкловский И. С., Излучение «мистернума» как лазерный эффект. Астр. цирк. № 372, 1966, стр. 1—8.
10. Кардашев Н. С., Передача информации внеземными цивилизациями. Астрон. ж. 41, 282 1964.
11. Слыш В. И., Радиоастрономические критерии искусственности радиоисточников. В сб. [3], стр. 61—67.
12. Гудзенко Л. И., Пановкин Б. Н., К вопросу о приеме сигналов внеземной цивилизации. В сб. [3], стр. 68—71.
13. Шоломицкий Г. Б., Флуктуации потока СТА-102 на волне 32,5 см. Астрон. ж. 42, 673, 1965.
14. Голей М., Когерентность разумных сигналов. В сб. [2], стр. 212—219.
15. Шкловский И. С., Вселенная, жизнь, разум, изд. 2-е, «Наука», 1965.
16. Котельников В. А., Связь с внеземными цивилизациями в радиодиапазоне. В сб. [3], стр. 113—120.

Методы дешифровки сообщений

§ 1. Введение

В цепи задач, объединяемых темой «связь с внеземными цивилизациями», важное место занимает проблема дешифровки сигналов космического происхождения.

Конечно, каждый астроном, анализируя сигналы любых космических объектов, по-своему их дешифрует, добывая сведения о строении этих объектов. Однако в данном случае имеется в виду получение информации, не специфичной для данного источника излучения, информации очень богатой и в принципе отражающей устройство всего мира, включая устройство некоторого общества «разумных существ», т. е. охватывающей примерно тот же круг знаний, что и «земная» литература.

Характерной особенностью проблемы дешифровки сообщений от космических цивилизаций является практически полное отсутствие каких бы то ни было знаний об этих цивилизациях. Это обстоятельство, по существу, уравнивает ее с проблемой дешифровки произвольного текста.

Следует отметить, что в лингвистике до недавнего времени проблема дешифровки произвольного текста не привлекала к себе серьезного внимания. Тем не менее, некоторые методы дешифровки, использующие малую исходную информацию о тексте, все же существуют. Общие идеи, лежащие в их основе, кажутся интересными, а результаты эксперименталь-

ных исследований — многообещающими. Можно думать, что «космическая» проблематика будет в значительной степени способствовать развитию этого направления языкознания.

Разумеется, нельзя отрицать, что какая-нибудь случайность поможет понять принятые сообщения. Можно также думать, что эти сообщения окажутся настолько хорошо организованными, что их понимание будет максимально упрощено. Однако осторожнее полагать, что дешифровка космических посланий будет представлять значительные трудности, не меньшие, например, чем те, с которыми сталкивается исследователь надписей на забытых языках. Такая точка зрения важна еще и тем, что она призывает готовиться к необходимости разбирать «космические послания», а не ограничиваться их ожиданием. Для лингвиста же построение методов дешифровки является увлекательнейшим занятием, требующим проникновения в природу и строение языка.

В лингвистической проблематике, связанной с межзвездной связью, есть и другая основная задача (помимо дешифровки). Речь идет о создании наиболее эффективного языка для космического общения. Эта задача заманчива тем, что каждый исследователь выполняет ее целиком, т. е. создает некоторый язык, тогда как вряд ли найдется человек, который создаст полный набор приемов дешифровки. Однако практическая актуальность этой задачи совсем другая, поскольку космическая связь не может быть организована по принципу диалога. Ожидание реплики в масштабах космоса в лучшем случае будет занимать сотни лет. Если же космические цивилизации будут отвечать на наши вопросы медленнее, чем мы сами будем находить на них ответы (или терять к ним интерес), то космическая связь сможет найти оправдание лишь в праздном любопытстве.

Космическое общение, по-видимому, сходно с литературной деятельностью: сообщения посылаются цивилизацией — автором по всем направлениям (подобно тому как книги поступают в различные библиотеки и книжные магазины); посылающие не ждут

ответа, так же как земные авторы обычно не надеются дожидаться рецензий на свои труды. Наградой за посланные сообщения является возможность знакомиться с сообщениями, присланными из других миров.

Естественно считать, что человечество вступит в космический культурный мир сначала как читатель и затем уже как автор. Поэтому проблема дешифровки представляется гораздо более настоятельной (по сравнению с проблемой создания космического языка), по крайней мере в настоящее время.

Данная глава имеет целью познакомить читателя с новыми методами лингвистической дешифровки сообщений.

Эти методы ориентированы на широкое использование современной вычислительной техники и поэтому в основном представляют собой алгоритмы, т. е. вычислительные инструкции для счетной машины. Для удобства читателей алгоритмы излагаются обобщенно, с пропуском многих несущественных деталей.

Дешифровочные методы имеют две цели. Практически они предназначены для дешифровки непонятных сообщений. Теоретически алгоритмы дешифровки представляют собой определения тех лингвистических явлений, которые они опознают в сообщениях на неизвестных языках. С этой точки зрения они особенно интересны специалисту по языку. Если рассматривать алгоритмы с последней точки зрения, то их главный интерес в том, что они представляют собой общие приемы исследования, допускающие многократное использование. Их общность ограничивает роль интуиции и произвола исследователя.

Эти две цели алгоритмов выдвигают к ним разные требования; с одной стороны, они должны давать по возможности правильные результаты, с другой, — в их строении должно быть как можно меньше произвола и логической неясности. Например, мы старательно избегали использования числовых констант, находимых «из опыта». Там же, где произвол был неизбежен, мы старались использовать простые решения. Сюда относится построение по возможно-

сти простых «оценочных функций». В редких случаях мы приводили алгоритмы, о которых даже известно, что они дают неудовлетворительные результаты, но обладают интересной «схемой», могущей помочь в дальнейшей работе.

Читатель увидит, что не все сведения, приводимые здесь, имеют одинаковый интерес. Мы хотели бы обратить внимание на «основные» алгоритмы — алгоритм выделения двух классов букв, семантический, алгоритм поиска графа предложения, алгоритмы выделения кодовых цепочек и морфем, алгоритм дешифровки изображений и алгоритм сравнения букв.

Возможно, многим покажется, что алгоритмы, обнаруживающие, например, гласные и согласные, имеют отдаленное отношение к космической проблематике. Мы хотим заранее обратить внимание (далее об этом будет идти речь) на то, что все приводимые алгоритмы имеют и более общую интерпретацию.

§ 2. Понятие сообщения, его понятности и разумности

Определение понятия «сообщение». Цель данного параграфа — более точная формулировка основных понятий и проблем, возникающих при дешифровке. Обычная практика справедливо предполагает, что дешифровщики имеют дело с сообщениями, которые они должны понять и перевести на известный язык. В случае космической дешифровки требуется к тому же предварительно выяснить, является ли данное сообщение предназначенным (интересным) для дешифровки, или, как мы будем говорить, разумным. Об этих понятиях и пойдет речь.

Под сообщением мы понимаем систему M трех множеств: $A = \{a_i\}$ — алфавит (множество букв); $L = \{l_j\}$ — множество мест; $T = \{a_i l_j\}$ — текст или множество вхождений букв в множество мест, т. е. множество пар вида $a_i l_j$, где $a_i \in A$, $l_j \in L$.

При самом общем понимании сообщения на все три множества не следует налагать никаких ограничений; эти множества могут быть как конечными, так

и бесконечными, могут быть в математическом смысле группами, кольцами, пространствами и т. д. Конечно, такое понимание очень общо и поэтому для многих целей понятие сообщения приходится ограничивать.

Например, множество A обычно считают конечным или по крайней мере счетным, а на множестве мест задается метрика или топология. Наконец, T — множество вхождений — можно характеризовать обычно как одно-многозначное отображение букв в места, т. е. считать, что для каждого места l_j найдется лишь одна стоящая на нем буква a_i , но для каждой буквы a_i может найтись сколь угодно большое число мест в L , на которых стоит эта буква.

Последнее условие позволяет очень легко ввести важное понятие «абсолютной частоты» буквы a_i .

«Абсолютной частотой» буквы a_i (или $\varphi(a_i)$) мы будем называть мощность множества $\{a_i l_x\}$, т. е. множества всех пар из текста, содержащих букву a_i .

Метрика множества мест легко переносится на текст: текстуальным расстоянием между вхождениями $a_i l_t$ и $a_k l_s$ естественно считать расстояние между l_t и l_s .

Впрочем, на множестве мест близость может определяться и каким-нибудь другим способом. Например, может быть задано отношение непосредственной близости, т. е. указываться, какие пары мест являются соседними, а какие нет.

Пусть, например, $A = \{a_i\}$ — буквы русского языка, L состоит из отрезков двух колец. Кольца могут перемещаться друг относительно друга случайным образом, текст выглядит так, как показано на рис. 51.

В данном случае задать расстояния между местами на одном и другом кольцах нельзя, но можно указать для каждого места двух ближайших «соседей».

Возможно, такое определение сообщения может вызвать упреки в чрезмерной общности. Почему бы не считать сообщение, как обычно, цепочкой букв?

Однако легко привести примеры, когда обычное представление выглядит неестественным. Например,

рисунок не может считаться сообщением; однако сообщением будет линейная развертка этого же самого рисунка. Аналогичным недостатком страдает и определение текста как отображения в граф *).

В то же время мы не исключаем целесообразности дальнейшего сужения понятия сообщения для определенных целей.

Так, сообщение в обычном смысле характеризуется следующими дополнительными особенностями:

1. Множества A , L и T конечны. 2. На множестве L

задано отношение непосредственной близости v

со следующими свойствами: а) если $l_i v l_j$, то $l_j v l_i$;

б) для каждого места l_x , кроме двух (l_p и l_k),

имеется по два «соседа», т. е. имеются такие

места l_y и l_z ($l_y \neq l_z$), что $l_x v l_y$, $l_x v l_z$, ($l_x \neq l_y$, $l_x \neq l_z$);

в) для l_p и l_k существует только по одному

соседу; г) для любого разбиения L на две части найдутся «соседи», принадлежащие к разным частям.

По-видимому, можно считать, что сообщение всегда задается текстом. Некоторое сомнение в этом могут вызвать случаи, когда «не все буквы алфавита встретились в тексте». Во всяком случае, в значительной степени слова «сообщение» и «текст» взаимозаменяемы, и мы будем этим широко пользоваться.

Теперь перейдем к вопросу о том, какие же сообщения интересны для космической дешифровки.

Искусственность и естественность сообщений. Если хотят отличить сигналы, идущие от «обычных»

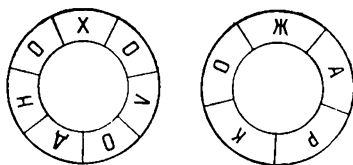


Рис. 51. Пример несвязного текста.

*) Графом называется совокупность двух множеств: множества «вершин» и множества «ребер», находящихся в дву-однозначном соответствии (т. е. каждое ребро соединяет две вершины). Графы допускают наглядное представление. Типичным примером графа является, например, сеть авиалиний: «вершины» — города, «ребра» — маршруты.

звезд, от сигналов, посылаемых разумными обитателями космоса, то говорят, что в первом случае мы имеем дело с «естественными» сообщениями, а во втором — с «искусственными». Нередко предполагается, что искусственные космические сигналы должны резко отличаться от естественных каким-нибудь необычным свойством, не поддающимся объяснению из физических соображений (см. гл. I и III).

Можно, однако, заметить, что многие неожиданные явления часто получают вполне «естественные» объяснения; в то же время можно построить примеры искусственных сообщений, сколь угодно мало отличающихся от естественных.

Допустим, например, что создано высококачественное объемное кино. Если мы не имеем доступа к экрану, то у нас нет никаких средств для отличения пейзажа, наблюдаемого через окно (естественное сообщение), от того же пейзажа, изображаемого на экране. Заметим, кстати, что изобретение голограммы дает надежду на создание киноизображений, обладающих именно такими свойствами.

Другой пример может показаться менее реальным (но он ближе к условиям поиска сигналов в космосе): пусть переменная звезда наблюдаема в точке А, но не наблюдаема в точке Б. В то же время точка А находится в пределах видимости из точки Б. Наблюдатель, находящийся в точке А, может сообщить в точку Б о поведении звезды, построив модель, копирующую ее поведение. При высоком качестве модели сигналы, исходящие от нее, не будут отличаться от сигналов, исходящих от самой звезды, но тем не менее сигналы модели искусственные, а сигналы звезды естественные.

Приведенные примеры говорят о том, что пытаться формализовать понятие искусственности и использовать его как всеобщий критерий отбора сигналов, предназначенных для дешифровки, не имеет смысла.

Постараемся показать, что даже вполне искусственные «осмысленные» сообщения могут иметь такой вид, который исключает возможность дешифровки. Если существуют «космические» сообщения такого

типа, то, несмотря на их возможную искусственность, они не могут быть нами поняты.

Дело в том, что любое сообщение можно «запутать» таким образом, что оно будет понятно лишь тому, кто обладает заранее полученным способом распутывания, «ключом» к шифру. В ряде случаев оказывается возможным обнаружить способ распутывания, даже если он не сообщался заранее. Однако если объем записи ключа сравним с величиной закодированного сообщения, то имеется возможность так запутать это сообщение, что его нельзя будет дешифровать никаким мыслимым способом. На этот факт указывает Шеннон [8]. Примеры таких способов запутывания нетрудно построить.

Пусть пересылается русский текст длиной в N букв. Положение каждой буквы в этом тексте можно охарактеризовать ее порядковым номером i от начала текста. Приготовим карточки, на которых записаны номера i ($1 \leq i \leq N$), перемешаем их и разложим в случайном порядке. В цепочке C полученной таким образом карточка с номером i будет стоять на j -м месте от начала цепочки. Если заменить эти номера буквами, стоящими на соответствующих местах в исходном сообщении, то мы получим его зашифровку. Чтобы ее расшифровать, необходимо знать цепочку цифр C (ключ). При дешифровке букву, стоящую в зашифрованном сообщении на j -м месте, следует поставить на место, обозначенное номером, стоящим на j -м месте в ключе.

Если ключ неизвестен, то расшифровать подобное сообщение, очевидно, нет никакой возможности: ведь последовательность букв, полученная при зашифровке, ничем не отличается от случайной последовательности букв. Правда, относительные частоты отдельных букв будут соответствовать частотам букв русского языка, но и этот факт легко скрыть, добавив в текст столько более редких букв, сколько необходимо, чтобы уравнивать все частоты.

Понятность сообщения, предсказуемость система, язык. Итак, нас интересуют не просто сообщения,

посылаемые разумными существами, но такие сообщения, которые постижимы или, иначе говоря, допускают понимание.

Можно ли найти признаки, отличающие постижимые сообщения от непостижимых?

Пусть текст некоторого сообщения предъявлен не полностью; иначе говоря, дано разбиение текста сообщения на предъявленную часть и не предъявленную. Если мы можем более или менее точно предвидеть, какова не предъявленная часть, на основе изучения предъявленной, то мы будем говорить, что сообщение понятно относительно данного разбиения. Если для любого разбиения текста на предъявленную и не предъявленную части мы можем предвидеть, какова не предъявленная часть, то мы будем говорить, что данное сообщение понятно сплошь *).

Далее мы постараемся показать на примерах, что такое представление о понятности не противоречит общепринятому значению этого слова.

Действительно, фраза «Пушкин родился в 18 параграфе» непонятна потому, что если в качестве предъявленной части взято «Пушкин родился в 18...», то никак не можем угадать, что следующая часть текста есть слово «параграфе». Мы, конечно, ожидаем появления слова (или группы слов), обозначающих время, например «... веке».

Непонятна и фраза «Пушкин родился в 18 веком», потому что ожидается предложный падеж слова «век», а никак не творительный. Равным образом непонятна и фраза «Пушкин родился в 18 century», поскольку мы не ожидаем появления английского слова в русской фразе. (Если фраза непонятна, но ясно, как ее надо изменить, чтобы она стала понятной, то обычно говорят, что фраза неправильная.)

Изображение человека, у которого вместо правой руки нарисована левая нога, непонятно; предмет, внешне неотличимый от бревна, но тонущий в воде,

*) Понятность относительно данного разбиения должна представлять собой числовую функцию от этого разбиения. Формального определения такой функции мы сейчас предложить не можем.

ведет себя непонятно; случайная последовательность букв непонятна сплошь.

Приведем примеры понятных сообщений. Бесконечная цепочка «... аааа ...» понятна для любого разбиения, поскольку здесь имеет смысл лишь прогноз вида «на i -м месте неосмотренной части стоит буква а», и этот прогноз верен всегда.

Сообщение вида «...*abc abc abc* ...» понятно для любого разбиения, содержащего достаточно крупную осмотренную часть, чтобы убедиться в периодичности повторений букв с периодом 3. Изображение бесконечной прямой линии понятно сплошь. Фраза «Пушкин родился в 18 веке» понятна для образованного человека, знающего русский язык, т. е. овладевшего особой системой предвидения появлений русских слов в русских фразах. Изображение человека понятно для разумных существ, видевших людей в жизни или на других изображениях. Любой периодический процесс понятен для разбиений, содержащих достаточно крупную осмотренную часть.

Покажем, что умение «более или менее точно предвидеть» опирается на знание особых свойств текста или его частей. Допустим, что после предъявления слов «Пушкин родился в 18 ...» мы предположили, что опущено слово «веке», тогда как в действительности в тексте стояло слово «столетии». Несомненно, наш прогноз следует считать весьма удачным, тогда как предположение о появлении слова «параграфе» было бы ошибочным. Однако совершенно очевидно, что слово «веке» внешне ничуть не более походит на слово «столетии», чем слово «параграфе». Сходство между ними смысловое, т. е. слово «веке» и слово «столетии» входят в некоторый класс «близких по смыслу» слов, или, говоря иначе, некоторое «смысловое расстояние» между ними достаточно малó.

Далее, в текстах могут быть более мелкие части (например, слова) и более крупные (предложения). Если мы знаем типичные наборы слов, составляющих предложения, то по некоторой части предложения сможем предвидеть слова, стоящие в остальной его

части (при этом типичными являются, собственно говоря, наборы так называемых грамматических классов, например, таких, как «именительный падеж», «глагол в личной форме» и т. д.). Следовательно, оказывается, что для правильного предсказания необходимо знать некоторое разбиение текста на части, аналогичные предложениям.

Такое разбиение может иметь весьма запутанный вид; например, в обычных языках нередко бывает, что одно предложение включено в другое. Поэтому оказывается необходимым оценивать близость «слов» в сообщении не по их «непосредственной близости», а как-то иначе.

Разумность сообщения. Сведения, необходимые для эффективного прогнозирования элементов текста, могут быть отражены специальной его перекодировкой, при которой части, сходные по смыслу, должны иметь сходную запись, а несходные — существенно отличную; части текста, объединяющиеся в более крупные части, должны отграничиваться символами типа скобок; части текста, «связанные по смыслу», должны быть текстуально близкими. Подобную перекодировку текста мы будем называть его интерпретацией.

Очевидно, наилучшей интерпретацией является та, для которой имеет место наивысшее значение понятности. Нахождение наилучшей интерпретации можно считать дешифровкой сообщения в узком смысле слова или частичной дешифровкой. Соответствие (отображение) между элементами сообщения и его наилучшей интерпретацией мы будем называть «предсказуемой системой» сообщения или его полной грамматикой. Предсказуемая система должна, по видимому, напоминать, с одной стороны, обычные грамматики, а с другой, толковые словари.

Языком естественно называть множество сообщений, предсказуемые системы которых совпадают. Другими словами, сообщения на одном языке строятся «одинаковым образом». Если между элементами наилучших интерпретаций двух сообщений на

разных языках может быть указано соответствие, то некоторое соответствие может быть установлено и между элементами сообщений. В этом случае один из текстов можно считать переводом другого.

Построение перевода дешифруемого сообщения можно рассматривать как конечную цель дешифровки. Как мы увидим в § 8, легче отыскивать соответствие именно между элементами наилучших интерпретаций, а не непосредственно между элементами сообщений. Поэтому для дешифровки следует изучать предсказуемые системы не только неизвестных, но и известных языков.

Примеры сообщений, понятных с нашей точки зрения, представляются до известной степени разочаровывающими. Очевидно, понятность не исчерпывает всех свойств сообщения, имеющих значение для его успешной дешифровки. Попробуем использовать то обстоятельство, что космическое сообщение должно быть, по-видимому, составлено таким образом, чтобы максимально облегчить его дешифровку.

В самом деле, в этом случае наилучшая интерпретация должна быть легко выделяема, т. е. сильно отличаться от ненаилучших. Если качество интерпретации оценивается ее понятностью, то под выделяемостью наилучшей интерпретации можно понимать разность между понятностью наилучшей и некоторой ненаилучшей (например, наихудшей) интерпретации. Выделимость наилучшей интерпретации представляется основным свойством сообщений, предназначенных для дешифровки: именно ее мы и будем называть разумностью.

Легко видеть, что сообщения, не имеющие достаточно понятных интерпретаций, не обладают высокой степенью разумности; с другой стороны, сообщения, все интерпретации которых понятны, также не обладают достаточной степенью разумности. Тем самым проясняется тривиальность примеров на стр. 231: для них невозможно построение малопонятных интерпретаций. Можно заметить, что сообщения на обычных языках (специально не зашифрованные) в некотором

смысле «предназначены для дешифровки» и весьма разумны.

Представляет некоторый интерес то, что можно назвать «внешней разумностью». Пусть у нас имеются два разбиения R_i и R_j текста T на осмотренную и неосмотренную части: $R_i = T_i^{\text{осм}}, T_i^{\text{неосм}}$; $R_j = T_j^{\text{осм}}, T_j^{\text{неосм}}$; $T_i^{\text{осм}} \subset T_j^{\text{осм}}$ (символ $T_i^{\text{осм}}$ означает осмотренную часть текста, символ $T_i^{\text{неосм}}$ — неосмотренную). Пусть каждому из этих разбиений соответствует значение понятности Π в наилучшей и в наихудшей интерпретации: $\Pi(R_i)_{\text{наил}}$, $\Pi(R_i)_{\text{наих}}$, $\Pi(R_j)_{\text{наил}}$, $\Pi(R_j)_{\text{наих}}$. Приращение $\Delta\rho$ разумности при переходе от разбиения R_i к R_j будет следующим:

$$\Delta\rho = [\Pi(R_j)_{\text{наил}} - \Pi(R_j)_{\text{наих}}] - [\Pi(R_i)_{\text{наил}} - \Pi(R_i)_{\text{наих}}].$$

Значение $\Delta\rho$ можно считать внешней разумностью сообщения, текст которого равен $T_j^{\text{осм}} \setminus T_i^{\text{осм}}$ (т. е. осмотрен в R_j и неосмотрен в R_i). В частности, если T есть текст сообщения обо всем мире, доставляемого органами чувств в течение человеческой жизни, то $\Delta\rho$ есть приращение разумности мира, раскрываемое сообщением с текстом $T_j^{\text{осм}} \setminus T_i^{\text{осм}}$.

Конкретный пример применения изложенных принципов можно найти в § 6.

§ 3. Традиционные методы военной и лингвистической дешифровки

Военная дешифровка. Существуют две области деятельности, в которых приходится иметь дело с дешифровкой непонятных сообщений: с одной стороны, — лингвистам и историкам (при попытках чтения текстов на забытых языках), с другой, — военным и дипломатам, сталкивающимся с намеренно запутанными сообщениями на известных современных языках.

Насколько можно судить по литературе [см., например, [14)], методы военной шифровки имеют довольно ограниченные традиционные формы (хотя, как мы знаем, возможности запутывания сообщений в принципе весьма велики).

Раскрыть тот или иной военный шифр бывает трудно, а иногда и невозможно. Однако устройство этих шифров, как правило, весьма простое в сравнении со сложностью предсказуемой системы (грамматики) реального языка. Методы шифровки обычно сводятся к преобразованиям цепочек букв без какой бы то ни было связи со смыслом передаваемого текста.

Рассмотрим некоторые из этих методов [14].

Шифр простой подстановки. В этом шифре каждая буква сообщения заменяется некоторым символом (обычно тоже буквой).

Транспозиция с фиксированным периодом t . В этом случае все сообщение делится на цепочки, длиной в t букв, и к каждой цепочке применяется одна и та же подстановка.

Шрифт Виженера и его варианты. Ключ задается набором из t букв. Эти наборы подписываются с повторением под исходным сообщением и полученные две последовательности складываются по модулю n , где n — количество букв в алфавите сообщения *).

Например:

исходный текст:	ЖИЛБЫЛУБАБУШКИСЕРЕНЬ- КИПКОЗЛИК
ключ:	КОЗЛИККОЗЛИККОЗЛИККОЗ- ЛИККОЗЛИК
зашифровка:	РЦТМГХЭПЗМЫВФЦШРШПЧ- КСУСФШХТУТ

Если подписывается одна и та же буква, шифр носит название шифра Цезаря; если кодирование производится с помощью непериодической цепочки, то шифр называется шифром Вернама (шифр Вернама

*) Сложение букв по модулю n производится так: пусть $i(a)$, $i(b)$ и $i(c)$ — номера букв в алфавите. Тогда c определяется из соотношения $i(a) + i(b) = i(c)$, если $i(a) + i(b) \leq n$; $i(a) + i(b) - n = i(c)$, если $i(a) + i(b) > n$.

При прочтении зашифрованного текста пользуются вычитанием по модулю n , т. е. $i(a)$ определяют так: $i(a) = i(c) - i(b)$, если $i(c) - i(b) \geq 0$; $i(a) = i(c) - i(b) + n$, если $i(c) - i(b) < 0$.

дает недешифруемый текст). t -граммные подстановки имеют место, когда буква заменяется последовательностью символов длиной в t символов. В так называемых «кодах» слова, группы слов или слоги заменяются различными последовательностями букв.

Существуют две основные идеи, на которых основывается дешифровка: статистический метод и метод характерных слов. При первом из них сравниваются частоты букв зашифрованного сообщения и частоты букв предполагаемого языка сообщения, подсчитанные на более или менее обширном материале. Применяют также сравнение частот двух, трех и т. д. буквенных сочетаний. Если, например, частоты букв языка похожи на некоторые элементы зашифрованного сообщения, то эти элементы и есть образы соответствующих букв.

При втором методе в изучаемом тексте отыскиваются последовательности, состоящие из более мелких частей, повторяющихся так же, как буквы некоторых характерных слов, наличие которых предполагается в зашифрованном сообщении. Эти принципы используются и нами в алгоритмах § 8.

В отдельных случаях оказывается возможным перебрать все возможные шифры некоторого класса и привлекать исследуемый текст по мере надобности для установления вероятности того, что в данном случае принят тот или иной шифр. Например, если известно, что применялся шифр Цезаря, то вероятность конкретного шифра изменяется в зависимости от объема N перехваченной части криптограммы, и это изменение можно вычислить.

Допустим, перехвачена криптограмма, содержащая часть английской фразы «...creases to...» (creases — конец слова increases) и зашифрованная шифром Цезаря. Если принята лишь зашифровка одной буквы «с», то возможной расшифровкой будет любая буква английского алфавита. Вероятность каждой из этих расшифровок будет просто вероятностью данной буквы английского текста.

Если приняты две буквы (сг), то опять имеются ровно 26 возможных расшифровок сообщения (если

оно зашифровано шифром Цезаря). Вероятность каждой из них есть вероятность соответствующей пары букв английского языка, и так далее.

В табл. 4.1, составленной Шенноном, проводятся эти вычисления, проделанные вплоть до пятибуквен-

Таблица 4.1

Расшифровки	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$
CREAS	0,028	0,0377	0,1111	0,3673	1,0000
DSFBT	0,038	0,0314			
ETGCU	0,131	0,0881			
FUNDV	0,029	0,0189			
GVIEW	0,020				
HWJFX	0,053	0,0063			
IXKGY	0,063	0,0126			
JYLHZ	0,001				
KZMIA	0,004				
LANJB	0,034	0,1321	0,2500		
MBOKC	0,025		0,0222		
NCPLD	0,071	0,1195			
ODQME	0,080	0,0377			
PERNF	0,020	0,0818	0,4389	0,6327	
QFSOG	0,001				
RGTPH	0,068	0,0126			
SHUQI	0,061	0,0881	0,0056		
TIVRJ	0,105	0,2830	0,1667		
UJWSK	0,025				
VKXTL	0,009				
WLYUM	0,015		0,0056		
XMZVN	0,002				
YNAWO	0,020				
ZOBXP	0,001				
APCYQ	0,082	0,0503			
BQDZR	0,014				
H (десятичных единиц)	1,2425	0,9686	0,6034	0,2850	0

ных цепочек. Пусть зашифровка проводилась подписанием буквы a_i . Цепочки букв в графе «расшифровки» — это последовательности, полученные в результате вычитания из криптограммы

последовательностей

$a_i,$	$a_i,$	$a_i,$	$a_i,$	$\dots,$
$a_{i-1},$	$a_{i-1},$	$a_{i-1},$	$a_{i-1},$	$a_{i-1}, \dots,$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$a_1,$	$a_1,$	$a_1,$	$a_1,$	$a_1, \dots,$
$a_n,$	$a_n,$	$a_n,$	$a_n,$	$a_n, \dots,$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$a_{i+1},$	$a_{i+1},$	$a_{i+1},$	$a_{i+1},$	$a_{i+1}, \dots$

Первая колонка цифр ($N=1$) содержит вероятности цепочек английских букв длины 1. Эти цепочки являются расшифровками в том случае, если принята только одна буква (зашифровка буквы «с»). Вторая колонка содержит вероятности расшифровок двух первых букв (сг), т. е. сг, ds, et и т. д., колонка $N=5$ содержит вероятности расшифровок всех пяти букв текста creas, т. е. creas, dsibt и т. д. В этой колонке вероятность цепочки creas приближенно равна 1, остальные вероятности близки к нулю.

Пустые клетки таблицы соответствуют весьма малым числам. Вероятности вычислялись на основе данных о частотах двух и трехбуквенных цепочек, приведенных в [8].

Строка H дает энтропию распределений вероятностей для указанных пяти случаев. Энтропия H ($H = -\sum_i p_i \log p_i$) вычислялась по значениям p в этой таблице с помощью десятичных логарифмов.

Лингвистическая дешифровка. По-видимому, гораздо больший интерес для «космической лингвистики» могут представить способы дешифровки древних текстов. Исследователю древнего языка приходится раскрывать его весьма сложное устройство. Кроме того, эти тексты обычно не запутаны намеренно и поэтому, как правило, не похожи на случайно разбросанные буквы. Однако сложность естественных языков и обусловила тот факт, что, несмотря на целый ряд успешных дешифровок, общих методов не создано.

Успехи же, о которых идет речь, в большой степени обязаны счастливым случайностям, а также

удачным, но неповторимым для других случаев догадкам исследователей.

Так, знаменитая дешифровка египетских иероглифов обязана находкам значительного количества двуязычных надписей, т. е. таких, где наряду с непонятным текстом имелся и его перевод; дешифровка хеттского языка — удачной догадке о том, каким языкам родствен хеттский, дешифровка крито-микенских надписей — верному предположению о том, что язык этих надписей греческий.

Приведем цитату из статьи Б. Грозного (дешифровавшего хеттскую клинопись) о том, как ему удалось сделать первые успехи. Заметим при этом, что было заранее известно (на материале исследования других языков), как следует произносить клинописные знаки и что значат отдельные идеограммы — символы, обозначающие понятия, а не звуки.

«Способ моей работы лучше всего виден по следующему предложению, одному из первых, смысл которых мне удалось установить, и в котором я распознал три хеттских слова индоевропейского происхождения*). Это клинописное слово я читал фонетически**)

nu  — an ezateni vādar—ma ekuteni.

*) Родственные языки — это языки, возникшие из одного «исходного языка». В родственных языках имеются слова со сходным значением, которые и звучат сходным образом.

Например, близко родственными языками являются русский, украинский, польский, чешский, болгарский, сербский (так называемые славянские языки); более отдаленное родство у русского языка с немецким, латышью, греческим, некоторыми индийскими языками, составляющими так называемую индоевропейскую семью. Сравните, например.

русск. мать (род. п. матери)
немецк. Mutter
латинск. mater
греч. μήτηρ
древнеиндийск. mātār.

**) То есть знал, как произносить. Клинописный знак есть идеограмма (символ понятия). Звуковой состав соответствующего слова не был известен.

Когда я впервые наткнулся на это хеттское предложение, я знал в нем только значение идеограммы, которая часто, но не всегда, означает «хлеб». По другим местам можно было судить, что окончание означает винительный падеж единственного числа. Несмотря на наличие и других возможностей, было легко предположить, что в предложении, говорящем о хлебе, можно ожидать также глагол «кушать». Поэтому я, сначала, конечно, чисто гипотетически, предположил, что слова «эзатени» означает понятие еды. Вскоре я заметил, что хеттский корень «эза» также во многих других местах означает «кушать» и что вместе с этим встречается и корень «ад» с тем же значением, например, в форме «аданзи» — они едят, который, вероятно, тождествен с «эза». Тогда я сравнил, опять-таки только гипотетически, эти хеттские корни «ад», «эз» — «кушать» с латинским «edo», немецким «essen» — и т. д. По другим местам я видел, что «тэни» — окончание второго лица множественного числа настоящего и будущего времени, так что первое предложение я мог перевести «хлеб будете есть». Следующее предложение казалось параллельным с первым: «vadar» — имя существительное, «та» — предлог, «ekuteni» — глагольная форма с окончанием «тэни». Так как слово «вадар» было параллельно слову «хлеб», то нетрудно было предположить в нем название какого-нибудь простого кушанья. Здесь можно было допустить значение «вода», да и все остальные данные говорили в пользу этого слова. В то же время мне вспомнилось английское слово «water» — вода, англосаксонское «watar» и т. д.

Если я здесь, таким образом, имел слово «вода», то в следующем глаголе «экутэни», который был параллелен глаголу «эзатэни» — «вы будете кушать», я должен был видеть выражение «вы будете пить». Затем я нашел, что наряду с корнем «эку» — «пить», встречается также близкий ему корень «аку» — «пить», например, в слове «акуванна» — «пить». Сравнение слова «акуванна» — «пить» с латинским «aqua» — «вода» получалось само собой. Поэтому я

перевел все предложение следующим образом: «хлеб будете есть и воду будете пить».

Из этого отрывка видно, что никаких последовательных методов у Грозного не было: путеводной нитью ему служило предположение о том, что неизвестный язык родствен некоторым известным (в родственных языках слова со сходным значением звучат похоже); остальные соображения довольно темны: сюда относится, например, утверждение о параллельности двух фраз, а также частые напоминания о том, что данное слово и в других местах значит то же самое.

Если в неизвестном языке не установлено и произношение букв, то о нем можно судить благодаря наличию в текстах собственных имен, которые до известной степени интернациональны; раскрыть смысл помогают идеограммы и нероглифы (значение последних в какой-то мере соответствует изображенному рисунку), картинки, встречающиеся в текстах или предметы, на которых сделаны эти надписи. Обычно бывают также известны некоторые сведения о соответствующей исторической эпохе, о войнах, которые происходили в период написания текстов, о правителях. Иногда известны и словари (тангутская письменность, письменность майя). Таким образом, дешифровка древних текстов в решающей степени обеспечивается особым чутьем исследователя, его интуицией, использующей знание огромного количества сведений, которые могут оказаться полезными; но она не является собранием методов, основанных на общелингвистических закономерностях. Иначе говоря, эта дешифровка скорее является искусством, неповторимой и не вполне понятной способностью добиваться практических результатов.

§ 4. Порядок применения и структура дешифровочных алгоритмов

Порядок применения алгоритмов. Уровни. Разумеется, при дешифровке конкретного космического сообщения также придется пользоваться определенными догадками и более или менее случайной попутной

информацией. Однако сейчас, до получения космических сообщений, у нас есть только один способ готовиться к их разгадыванию: разработка наиболее общих методов дешифровки, пригодных для возможно большего числа случаев.

Нам представляется, что эти методы будут иметь ценность лишь в том случае, если они допускают ясные формулировки, исключающие возможность кривотолков. Такой естественной формой для дешифровочных методов являются алгоритмы — точные инструкции по обработке текстов, настолько ясные и понятные, что их выполнение можно поручить вычислительной машине. Последнее обстоятельство особенно желательно ввиду того, что при дешифровке приходится обрабатывать весьма большие массивы информации, что зачастую почти неосуществимо при «ручной» работе.

При полностью построенной системе таких алгоритмов процесс дешифровки можно представить себе следующим образом: в вычислительную машину подается дешифруемый текст, а из нее выходит готовый перевод этого текста на какой-нибудь из известных языков. Однако такая идеальная ситуация, по-видимому, нереальна.

В таком прогрессе, вообще говоря, и нет необходимости — достаточно, если алгоритмы будут решать некоторые «ключевые» проблемы дешифровки. Тогда на долю догадок, интуитивной деятельности дешифровщика, останутся сравнительно простые задачи или такие задачи, решать которые человеку проще, чем машине.

Как мы уже упоминали, под дешифровкой можно понимать, во-первых, деятельность, направленную на обнаружение «предсказуемой системы», во-вторых, деятельность, имеющую целью получение перевода оригинала на какой-нибудь из известных языков. Если бы мы обладали мощными способами дешифровки во втором смысле, то нам незачем было бы искать «предсказуемую систему» оригинала.

Однако практика показывает, что все же целесообразнее вначале отыскивать предсказуемую си-

стему, а затем уже строить способы перевода. В этом смысле показательны эксперименты, проведенные по алгоритму, излагаемому на стр. 299. Кроме того, следует учесть соображения о возможной «неинтерпретируемости» космических сообщений, приводимые в заключение.

Мы рассмотрим некоторые алгоритмы, важность которых для построения предсказуемой системы более или менее очевидна.

Традиционная лингвистика учит распознавать лингвистические явления с помощью двух приемов: один из них опирается на те образы из действительности, которые языковые выражения вызывают в сознании человека, владеющего данным языком, другой использует наше умение отличать выражения, правильные для данного языка, от неправильных.

Например, глаголы от существительных отличают благодаря тому, что глаголы обычно выражают некоторое действие или процесс, а существительные — предмет или абстрактное понятие. В то же время глагол можно опознать благодаря тому, что в конце слов, входящих в этот класс, могут находиться особые «окончания» — характерные цепочки букв, например, «л», «ла», «ло» (окончания прошедшего времени мужского, женского и среднего рода).

Современная прикладная лингвистика, к которой прежде всего относятся работы по машинному переводу, использует лишь сведения второго рода, т. е. считает заранее известными некоторые явные признаки лингвистических явлений.

Разумеется, при дешифровке космических сообщений мы не можем опираться ни на способность обращаться к образам из действительности, ни на знание того, как выглядят в буквенной записи опознаваемые лингвистические явления.

При построении дешифровочных алгоритмов мы должны отталкиваться лишь от самых общих свойств опознаваемых явлений. В этом заключается их особый интерес для лингвиста.

Дело в том, что конкретный дешифровочный алгоритм является, собственно говоря, определением того •

явления, которое он должен опознавать. Точнее, определениями можно было бы считать фразы следующего вида: данное лингвистическое явление есть то, что обнаруживается в произвольном тексте данным дешифровочным алгоритмом.

Подобные определения привлекательны тем, что они годятся и для неизвестных языков (а значит, они более общие), что они ясны (допускают машинную реализацию) и практичны, т. е. могут служить средством опознавания определяемых явлений, причем интерес к таким алгоритмам может возникать независимо от дешифровочных целей.

Рассмотрим подробнее устройство и порядок применения дешифровочных алгоритмов. Очевидно, дешифровочные алгоритмы могут использовать сведения, полученные в результате работы других дешифровочных алгоритмов. Если некоторый алгоритм В использует умение распознавать лингвистическое явление, определяемое алгоритмом А, то мы будем говорить, что уровень алгоритма В выше уровня алгоритма А. Разумеется, было бы очень плохо, если бы уровень алгоритма А в свою очередь был выше уровня алгоритма В, так как в этом случае неизвестное явление определялось бы с помощью неизвестного же. Исключение составляют алгоритмы, попеременно улучшающие результаты своей работы. В этом случае старшинство алгоритмов определяется их старшинством в первой итерации.

Отсюда видно, что должен существовать алгоритм нулевого уровня, не использующий никаких сведений, полученных каким-либо другим дешифровочным алгоритмом. Нулевые алгоритмы должны быть различными в зависимости от того, какое воздействие оказывает субстанция символов на органы чувств человека или дешифрующего устройства. Они должны быть связаны с минимальными различиями, улавливаемыми этими органами чувств.

Если мы строим алгоритмы, предназначенные для изучения письменных языков, то, очевидно, нулевой алгоритм должен находить алфавит данного языка на основании осмотра некоторого текста на этом языке.

Используемая информация, по-видимому, должна сводиться к умению отличить черный квадратик от белого, если считать, что текст покрыт настолько мелкой сеткой, что каждая ячейка сетки может рассматриваться либо как черная, либо как белая. Кроме того, следует считать заданным умение определять, где находится данный квадратик.

Если дешифруется звучащая речь, то нулевой алгоритм должен использовать минимальные акустические различия. Разнообразие нулевых алгоритмов можно, по-видимому, сократить, преобразуя сигналы с помощью физических приборов; например, звучащую речь можно преобразовать в записанную на бумаге кривую, показывающую изменения давления воздуха.

На первый взгляд задача обнаружения алфавита языка очень проста и мы всегда можем решить ее на глаз. Однако и для исследователя — человека при неразборчивости текста возникают затруднения; для машины же задача нелегка и в случае сравнительно четких надписей. Быть может, представление о трудности обнаружения алфавита звуков даст приводимый отрезок осциллограммы русской речи (рис. 52).

Конечно, не приходится ждать особой «разборчивости» от кривой, изображающей некоторый космический сигнал; вполне вероятно, что он будет сильно искажен помехами. Кроме того, кривая не является таким наглядным материалом, с которым обычно человек имеет дело, поэтому нельзя ожидать, что выделение алфавита «элементарных» сигналов будет несложной задачей.

Итак, нулевой алгоритм для письменных языков должен находить буквы как особые сочетания темных и светлых квадратиков. Следует заметить, что хотя возможность построения хорошо работающего алгоритма, решающего эту задачу, не вызывает сомнений, в настоящее время такого алгоритма не имеется. В § 11 мы приведем весьма несовершенный алгоритм подобного рода, который имеет скорее принципиальное значение, чем практическое.

Когда выделено множество «элементарных сигналов», можно переходить к обнаружению и изучению более крупных элементов сообщения. Если мы имеем

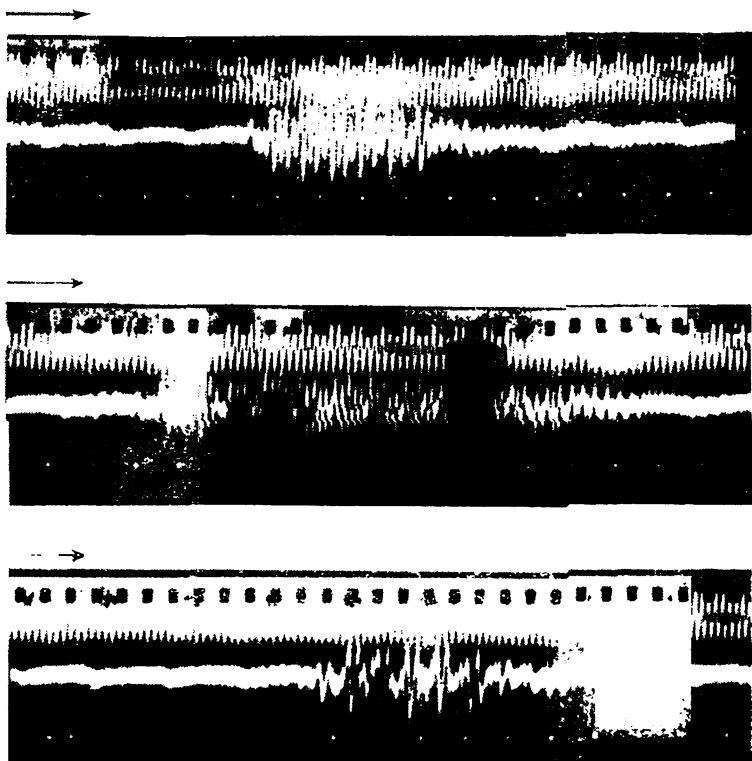


Рис. 52. Осциллограмма слогов русской речи: «ту», «та», «па».

Все эти слоги — безударные, взятые из записи отдельных фраз. Плоская часть кривых — момент молчания (рот закрыт), затем следует «взрыв», обеспечивающий слышимость звуков «т» и «п», далее колебания с большой амплитудой — гласные «у» и «а».

дело с языком, похожим на человеческий, то следующими алгоритмами (первого уровня) мы можем получить различные классы букв, содержащие буквы со сходным произношением, а также осмысленные це-

почки букв, не состоящие из других осмысленных цепочек, так называемые морфемы^{*)}). Этот алгоритм должен находить разбиение сообщения на морфемы и в том случае, когда не указаны заранее границы между словами (пробелы), так как слова — более сложные единства по сравнению с морфемами. В ряде орфографий границы между словами не отмечаются.

Алгоритмы второго уровня должны находить границы между словами, а также классы морфем (например, морфемы «знаменательные», такие, как «дом», «автомат», и служебные, например, «у», «на» — предлоги, «-ой» — окончание творительного падежа первого склонения). Алгоритмы третьего уровня должны искать классы слов и находить границы предложений. Алгоритмы следующих уровней должны анализировать предложения и исследовать смысл (семантику).

Следует отметить, что многие алгоритмы разных уровней обнаруживают большое сходство, а иногда и полное тождество, отличаясь лишь тем материалом, который они обрабатывают. Так, алгоритмы, находящие классы букв с похожим произношением, можно без всяких изменений использовать для обнаружения классов морфем; алгоритмы, находящие предложения, могут сильно напоминать алгоритмы, находящие границы между слогами (т. е. «благозвучными» сочетаниями звуков, например «та+ра+кан»), алгоритмы, находящие границы между морфемами — алгоритм, выделяющий буквы и т. д.

Излагая конкретные алгоритмы, мы будем указывать, к каким задачам различных уровней данный алгоритм может быть применен. Впрочем, у сходных, по существу, алгоритмов может быть и некоторая специфика, как правило, связанная с различием объемов перерабатываемых множеств символов.

Например, алгоритм, отыскивающий классы морфем, в качестве ответа должен выдавать множество, раз в сто превышающее множество, выдаваемое тем же алгоритмом при работе над буквами. Эта

^{*)} Например, слово «благодарность» состоит из следующих морфем: благ+о+дар+н+ость.

специфика приводит к тому, что эксперименты и программирование алгоритмов, предназначенных для исследования высших уровней, значительно сложнее, чем для невысоких.

В то же время, конечно, более высокие уровни интереснее: они в большой степени определяют понятность текстов, позволяя осуществить «далекое прогнозирование».

Здесь рассматриваются алгоритмы, предназначенные для изучения невысоких уровней, например, для изучения произношения букв. Сами по себе далеко не все они будут иметь значение для дешифровки космических сообщений. Однако мы советуем относиться к ним как к моделям: обычно эти алгоритмы на простом материале решают те же задачи, которые на более высоких уровнях приобретают большую актуальность.

Можно думать, что существует ограниченное количество различных типов дешифровочных алгоритмов, поэтому на первых порах интересно не столько забираться в высокие уровни, сколько попытаться построить по одному алгоритму каждого из возможных типов.

Строение алгоритмов: множества возможностей, оценочная функция, вычислительные процедуры. Типы алгоритмов. В устройстве различных дешифровочных алгоритмов имеется много общего. Ранее мы говорили, что алгоритмы, опознающие в неизвестном тексте лингвистические явления, могут служить определением этих явлений. Однако оказывается возможным предлагать и определения несколько более общего вида. Для этого достаточно четко указывать те признаки лингвистического явления, с помощью которых оно опознается. Дело в том, что вычислительные процедуры, предназначенные для опознавания (собственно, алгоритмы), могут быть различными даже для одного и того же набора опознавательных признаков. При изложении дешифровочных алгоритмов мы вначале будем описывать опознавательные

признаки, а затем уже саму опознавательную процедуру.

Опознавательные признаки в свою очередь делятся на два сорта: одни из них не требуют выполнения каких бы то ни было вычислений с помощью текста, другие требуют. Первые носят так называемый двончный характер, т. е. должны присутствовать у определяемого явления и отсутствовать у остальных. Вторые отражают свойства, которые лишь в большей мере присутствуют у определяемого явления, чем в каком-нибудь другом, имеющем те же «двончные» опознавательные признаки.

Мы будем говорить, что признаки первого рода определяют множество возможностей (множество интерпретаций), а вторые — качество или правдоподобие этих возможностей. Другими словами, качество есть числовая функция, определенная на множестве возможностей. Множество возможностей мы будем также называть множеством допустимых решений, а качество — оценочной функцией.

Следует подчеркнуть крайне важное значение оценочных функций. До недавнего времени лингвисты строили определения, пользуясь лишь двоичными (в крайнем случае, конечнозначными) признаками. Однако эти определения оказывались чрезмерно сложными, содержали большое количество «исключений» и были мало пригодны для автоматического опознавания.

Как правило, при этом не было возможности давать определения, общие для сходных явлений разных языков. Использование оценочных функций позволяет предельно упростить «двончную», т. е. логическую часть определений, в силу чего они приобретают большую общность. Как увидит читатель, оценочные функции оказались удобными и практически, поскольку программирование алгоритмов, использующих эти функции, обычно не слишком сложно.

Целью опознавательной процедуры (собственно, алгоритма) является поиск такого допустимого решения, для которого оценочная функция принимает максимальное (иногда минимальное) значение.

Коль скоро задано множество допустимых решений и определена оценочная функция, нахождение допустимого решения, для которого оценочная функция имеет максимальное или минимальное значение, становится уже чисто математической задачей.

Строгое решение математических задач такого рода, к которым сводятся дешифровочные алгоритмы, в большинстве случаев неизвестно. Мы старались давать по возможности практичные решения, т. е. достаточно точные, чтобы получать приемлемые результаты и достаточно простые, чтобы выполнение необходимых вычислений на вычислительных машинах было реально.

Вернемся еще раз к вопросу о возможных типах алгоритмов. Имеющиеся в настоящее время алгоритмы, предназначенные для изучения «предсказуемой системы» можно разбить на следующие группы:

Алгоритмы классификации. К ним относятся алгоритмы, разбивающие множество изучаемых единиц на непересекающиеся подмножества, например, алгоритмы, разбивающие множество букв на классы, содержащие сходно читающиеся буквы; алгоритмы, разбивающие множество морфем на классы морфем с одинаковыми «грамматическими» свойствами (служебность — знаменательность), алгоритмы, находящие классы слов, близких по значению.

Алгоритмы склеивания. Так мы назовем алгоритмы, образующие более крупные лингвистические единицы из более мелких: например, алгоритм выделения морфем, алгоритм выделения букв, алгоритм выделения предложений, алгоритм выделения слогов.

Алгоритмы установления близости. Смысл этих алгоритмов несколько менее очевиден. Дело в том, что визуальная близость слов в тексте не всегда соответствует некоторому более интересному представлению о близости слов. Например, в фразе «буревестник с криком реет, черной молнии подобный» слова «буревестник» и «подобный», несомненно, теснее связаны по смыслу, чем слова «буревестник» и «молнии», хотя

первые два слова дальше отстоят друг от друга, чем вторые.

Аналогичным образом в линейной развертке двумерного изображения разумно считать близкими не только символы, расположенные рядом, но и символы, отделенные интервалом, равным строке.

К алгоритмам этого типа относятся приводимые на стр. 329 алгоритмы, отыскивающие так называемый граф предложения. Заметим, что знание «истинной близости» символов необходимо для правильной работы алгоритмов склеивания.

Алгоритмы, предназначенные для построения перевода, отражены в этой статье в меньшей степени. По-видимому, для дешифровки можно ограничиться алгоритмами, составляющими различного рода двуязычные словари. В практике машинного перевода большую роль играют алгоритмы, синтезирующие (строящие) фразы на языке, на который осуществляется перевод. В случае дешифровки вполне возможно оставить эту задачу дешифровщику — человеку.

Расположение описаний алгоритмов в главе не соответствует приведенному порядку. Более простые и очевидные алгоритмы, снабженные примерами, даны раньше, в § 5 и 6, остальные даны в § 7—11.

§ 5. Алгоритмы классификации (начало)

Признаки и классификации. Алгоритмы классификации нужны для того, чтобы судить о сходстве и различии лингвистических явлений. Принято считать, что лингвистическая единица характеризуется набором свойств или признаков, которые есть у данного явления, но отсутствуют у прочих.

Если набор таких свойств задан, то данное лингвистическое явление можно описать, сопоставив ему некоторый вектор из нулей и единиц, в котором i -я координата соответствует i -му признаку; она равна единице, если признак у данного объекта имеется, и нулю в противоположном случае.

Иногда считается, что признак может принимать и другие значения, помимо нуля и единицы. Вообще,

под признаком можно понимать некоторую числовую функцию, определенную на множестве исследуемых объектов.

Если признак может принимать k значений, то он разбивает множество объектов самое большее на k непересекающихся классов. Наоборот, если у нас имеется классификация (разбиение) множества объектов на k непересекающихся классов, то мы можем ввести признак, принимающий k значений. Этот обратный путь более характерен для дешифровки.

Если два объекта имеют описания — векторы, то сходство этих объектов можно оценить, вычислив по некоторой формуле расстояния*) между ними, как между точками n -мерного пространства.

Приведем, например, табл. 4.2 [13], характеризующую звуки русского языка.

Столбцы этой таблицы озаглавлены символами русских звуков. Штрих у согласных выражает мяг-

	й	т	д	т'	д'	н	н'	ц	с	з	с'	з'	р	б	р'	б'	м	м'
	j	t	d	t'	d'	n	n'	s	s	z	s'	z'	p	b	p'	b'	m	m'
Вокальность	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Консонантность	—	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Диффузность	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Компактность	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Низкая тональность	0	—	—	—	—	—	—	—	—	—	—	—	+	+	+	+	+	+
Напряженность	0	—	—	—	—	+	+	+	+	+	+	+	—	—	—	—	—	—
Назальность	0	—	—	—	—	+	+	0	0	0	0	0	—	—	—	—	+	+
Протяженность	0	0	0	0	0	0	0	—	+	+	+	+	0	0	0	0	0	0
Звонкость	0	—	+	—	+	0	0	0	—	+	—	+	—	+	—	+	0	0
Палатальность	0	—	—	+	+	—	+	0	—	—	+	+	—	—	+	+	—	+
Ударность	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

*) Расстояние есть функция от пар элементов некоторого множества, обладающая следующими свойствами: 1) $\rho(a, a) = 0$ (невырожденность), 2) $\rho(a, b) = \rho(b, a)$ (симметричность), 3) $\rho(a, b) + \rho(b, c) \geq \rho(a, c)$ (неравенство треугольника).

кость, штрих у гласных — их безударность. Строки озаглавлены названиями признаков. Например, вокальность (т. е. гласность) принимает либо значение + (да), либо — (нет) для каждой буквы, причем «й», по мнению автора, ни гласный, ни согласный, «р», «р'», «л», «л'» — и гласные и согласные сразу. Поэтому консонантность (согласность) не определяется значением вокальности. Ударность — признак, характеризующий гласные, поэтому для согласных он принимает значение 0 (неприменимость).

Значения + и — данного признака характеризуют более похожие звуки, чем + и 0 или — и 0, поэтому в качестве расстояния можно использовать функцию, описанную на стр. 321.

Если бы мы захотели использовать евклидово расстояние $\rho_{\text{евк}} = \sqrt{(x_i - x'_i)^2}$, то предварительно

Таблица 4.2

ф	в	ф'	в'	ч	ш	ж	к	к'	д	х	е	е'	о	о'	а	а'	и	и'	н	н'	р	р'	л	л'
f	v	f'	v'	č	š	ž	k	k'	d	x	e	e'	o	o'	a	a'	i	i'	u	u'	g	g'	l	l'
—	—	—	—	—	—	—	—	—	—	—	—	—	+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	—	—	—	—	—	—	—	—	—	—	—	—	+	+	+
0	0	0	0	0	0	0	0	0	0	0	—	—	—	—	—	—	+	+	+	+	0	0	0	0
—	—	—	—	+	+	+	+	+	+	—	—	—	—	+	+	0	0	0	0	0	0	0	0	0
+	+	+	+	—	—	—	+	+	+	+	—	—	+	+	0	0	—	—	+	+	0	0	0	0
+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	—	+	+	—	—	—	+	0	0	0	0	0	0	0	0	0	0	0	0	—	+
—	+	—	+	0	—	+	—	—	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
—	—	+	+	0	0	0	—	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	+
0	0	0	0	0	0	0	0	0	0	0	—	+	—	+	—	+	—	+	—	+	0	0	0	0

следовало бы сопоставить каждому значению (0, + или —) некоторое число. Если измерять расстояние между двумя звуками по формуле (4.5) (стр. 321), то расстояние между а и в будет равно 20, а расстояние,

например, между a и o равно 3. Это хорошо соответствует интуитивному представлению.

Если у нас есть набор так называемых грамматических классов слов (например, «именительный падеж», «мужской род», «единственное число»), то мы могли бы построить аналогичную таблицу, выражающую грамматические свойства слов, а если у нас имеется набор классов слов, содержащих нечто общее в значении (например, одушевленность, великость, разумность и т. д.), то можно было бы дать смысловое или, как еще говорят, семантическое описание слов.

Кроме возможности оценивать сходство объектов, векторы — описания могли бы быть полезны тем, что вместо огромного количества объектов мы могли бы оперировать цепочками, состоящими из очень малого количества признаков. Так, если пользоваться двузначными признаками, то множество из n объектов мы могли бы описать с помощью всего $[\log_2 n] + 1$ *) признаков. Если учесть огромное количество возможных слов и понятий, то это свойство описаний следует признать весьма ценным.

Рассматривая табл. 4.2, можно заметить, что признаки выбраны самые разнообразные: некоторые из них связаны со способом произнесения, другие — с акустическими свойствами.

Если бы мы могли получить таблицу аналогичного вида на основании изучения способов комбинации звуков друг с другом в потоке речи, то мы смогли бы и по письменному тексту восстановить звучание букв. Ведь письмо в малой степени искажает способность звуков (которые изображаются буквами) сочетаться друг с другом. Если же подобные таблицы мы могли бы построить для слов, притом таким образом, чтобы классы слов, соответствующие некоторому значению определенного признака, содержали бы слова, имеющие нечто общее в значении, то мы могли бы на

*) Здесь $[\log_2 n]$ означает целую часть логарифма.

основании изучения текстов «угадывать» значения слов.

На пути решения этой проблемы лежат значительные трудности. Однако подобная общая схема делает более понятными те частные алгоритмы классификации, которые мы будем приводить ниже.

Алгоритмы, опознающие гласные и согласные. Первый алгоритм этого класса служит восстановлению произношения букв по их сочетаемости в тексте. Он соответствует делению букв на два класса всего по одному бинарному признаку.

Если этот алгоритм работает над буквами, то он должен указать, какие буквы гласные, а какие согласные; работая над морфемами — указать, какие из них являются знаменательными (полнозначными), а какие служебными. Если бы этот алгоритм работал над текстами типа математических, то он, вероятно, мог бы установить, какие символы являются предикатными (+, —, = и т. д.), а какие — предметными (например, π , x , 10, 2^{32}). Если бы этот алгоритм работал над словами, то, возможно, он мог бы выделить классы именных и глагольных слов.

Конечно, установить, какие буквы изображают гласные, а какие — согласные, еще не значит установить произношение букв. Однако для практической дешифровки бывает полезно и это *).

Если считать, что этот алгоритм работает над буквами, то в соответствии со сказанным выше, можно считать, что он дает определение того, что такое гласные и согласные.

Это определение выгодно отличается от обычных (имеющих акустическую или физиологическую природу) тем, что оно годится и для таких букв, для которых традиционные соображения не имеют силы.

Множество возможностей. Итак, мы считаем, что гласные и согласные представляют собой некоторое

*) Дешифровка надписей на так называемом карийском языке, успешно проводящаяся В. В. Шеворошкиным, началась как раз с установления гласных и согласных.

разбиение множества букв на два класса: класс гласных и класс согласных. Иначе говоря, мы считаем, что эти два подмножества не пересекаются и исчерпывают алфавит.

Можно, однако, заметить, что такое ограничение не вполне соответствует реальному положению дел. В самом деле, английская буква «у» иногда передает гласный звук, иногда согласный («*vegu*» и «*ueag*»). Не следует думать, что подобное положение вызывается исключительно несовершенством письма: один и тот же звук тоже может иногда быть гласным, иногда согласным (например, звук «г» в чешском слове «*Прага*» — согласный, в слове «*prst*» (палец) — гласный). Если же не ограничиваться рассмотрением букв, то мы увидим, что такое явление крайне распространено; в частности, одно и то же слово обычно имеет много значений (это явление носит название омонимии).

Однако нам не ясно, каким образом можно было бы построить алгоритм поиска гласных и согласных без принятых ограничений.

Конечно, сказать, что гласные и согласные есть классы в некотором разбиении, значит, сказать очень мало. Ведь если в данном алфавите имеется n букв, то для него можно построить 2^n различных разбиений! Однако и такое определение есть шаг вперед: ведь до этого множество возможностей вообще ничем не ограничивалось.

Оценочная функция. Оценочную функцию будем строить из следующих соображений: очевидно, в любом тексте гласные не слишком охотно сочетаются с гласными, а согласные — с согласными. Наоборот, гласные с согласными сочетаются очень охотно.

Если мы возьмем произвольное разбиение алфавита на два класса, то такого свойства мы почти наверное не заметим. Пусть, например, согласной объявлена буква П, а гласными — все остальные буквы. Очевидно, при таком разбиении «гласные» могут в текстах весьма часто встречаться в непосредственной близости.

Будем изучать сочетаемость букв данного языка с помощью таблицы, строки и столбцы которой озаглавлены буквами данного алфавита. В клетку строки i и столбца j будем вносить числа, указывающие, сколько раз буква a_i и буква a_j встретились рядом в исследуемом тексте (порядок следования букв во внимание не принимается).

Пусть у нас есть некоторое разбиение алфавита на два класса. Сдвинем все строки и столбцы, озаглавленные «гласными буквами», в левый верхний угол таблицы и отделим от остальных линией. Таблица примет следующий вид:

	гласные	согласные
гласные	1	2
согласные	4	3

В блоке 1 окажутся числа, показывающие, как гласные сочетаются с гласными, в блоке 3 — как согласные с согласными, а в блоках 2 и 4 — как гласные с согласными. Если разбиение правдоподобное, то с точки зрения нашего предположения числа в блоках 1 и 3 будут маленькими, а в блоках 2 и 4 большими. Поэтому качество разбиения мы сможем оценить, например, суммой чисел в блоках 1 и 3.

Если число букв в алфавите равно n , а число гласных m , то эту функцию мы сможем записать следующим образом:

$$K_1 = \sum_{i=1}^m \sum_{j=1}^m \varphi(a_i, a_j) + \sum_{k=m+1}^n \sum_{l=m+1}^n \varphi(a_k, a_l). \quad (4.1)$$

В этой формуле $\varphi(a_i, a_j)$ означает число появлений букв a_i и a_j рядом, без различия порядка. На последнее обстоятельство будет указывать наличие запятой: так, $\varphi(a_i, a_j) = \varphi(a_i a_j) + \varphi(a_j a_i)$. Буквы a_i и a_j принадлежат к одному из классов, буквы a_k и a_l к другому.

Разбиение тем лучше, чем меньше значение K_1 , а наилучшее разбиение то, для которого эта функция принимает минимальное значение.

Приведенная оценочная функция для обнаружения гласных является далеко не единственно возможной. Имеется, в частности, большое количество оценочных функций, эквивалентных данной, т. е. имеющих экстремум для того же допустимого решения, что и K_1 . Существуют также интересные оценочные функции, не эквивалентные K_1 . Такова, например, функция

$$K_2 = \sum_{i=1}^m \sum_{j=1}^m p(a_i, a_j) p^2(a_x) + \\ + \sum_{k=m+1}^n \sum_{l=m+1}^n p(a_k, a_l) p^2(a_y) - \\ - 2 \sum_{i=1}^m \sum_{k=m+1}^n p(a_i, a_k) p(a_x) p(a_y). \quad (4.2)$$

Символ $p(a_x)$ означает $\frac{\sum_{x=m+1}^n \varphi(a_x)}{N}$, где a_x относится к тому же классу, что a_k и a_l . Обозначение основано на том, что появление произвольной буквы данного класса можно считать появлением некоторой

буквы a_x ; аналогично, $p(a_y)$ означает $\frac{\sum_{y=1}^m \varphi(a_y)}{N}$. Число $p(a_y)$ есть относительная частота одного из классов, $p(a_x)$ — относительная частота другого.

Функция K_2 похожа на функцию K_3 , эквивалентную K_1 :

$$K_3 = \sum_{i=1}^m \sum_{j=1}^m p(a_i, a_j) + \\ + \sum_{k=m+1}^n \sum_{l=m+1}^n p(a_k, a_l) - 2 \sum_{i=1}^m \sum_{k=m+1}^n p(a_i, a_k), \quad (4.3)$$

но отличается от нее коэффициентами $p(a_x)$ и $p(a_y)$. В целом функция K_2 представляет собой корреля-

ционный момент последовательности чисел 1 и -1 , полученной в результате замены каждой гласной исследуемого текста на 1 и каждой согласной на -1 . Функция K_2 хороша тем, что отражает неравномерность частот гласных и согласных. Сочетания согласных обычнее, поэтому менее существенны, что и выражено взвешиванием появлений согласных частотой гласных, и наоборот. Впрочем, эти рассуждения носят умозрительный характер, так как эксперименты проводились лишь с использованием функции K_1 .

Опознавательные процедуры. Простейшая процедура, основывающаяся на описанных признаках, тривиальна. Достаточно построить таблицу частот парных сочетаний, перебрать все возможные разбиения и для каждого подсчитать значение оценочной функции. Затем следует выбрать то из разбиений, для которого оценочная функция минимальна. Однако выполнение вычислений по этому алгоритму превышает возможности даже очень хорошей вычислительной машины, если размеры алфавита обычны (т. е. около 30 букв).

Поиск эффективного способа нахождения минимума подобной функции связан с большими математическими трудностями. Поэтому приходится выбирать между непрактичными и неправильными способами.

Ниже мы приведем алгоритм, во многих случаях находящий минимум K_1 быстро и просто. Для некоторых таблиц он дает ошибочные ответы, хотя, по всей видимости, не слишком отличные от наилучших. Во всяком случае эксперименты, проводившиеся с этим алгоритмом, ни разу не дали ошибок, которые можно было бы приписать несовершенству алгоритма. Это несовершенство обнаруживается лишь на специально подобранных таблицах.

Изложим опознавательную процедуру по пунктам:

1. Для данного текста построить таблицу чисел $\varphi(a_i, a_j)$, где $\varphi(a_i, a_j)$ означает количество появлений пары букв a_i, a_j в безразличном порядке,

2. Вычеркнуть числа вида $\varphi(a_i, a_i)$ *).

3. Подсчитать сумму чисел в каждой строке таблицы.

4. Поставить на первое место (от левого верхнего угла) строку и столбец с максимальной суммой.

5. Отделить пограничными линиями перенесенные строки и столбцы от неперенесенных.

6. Для строк, лежащих ниже горизонтальной границы, подсчитать сумму чисел, расположенных правее вертикальной границы, и сумму чисел, расположенных левее вертикальной границы; вторую сумму вычесть из первой. Полученные числа назовем решающими разностями.

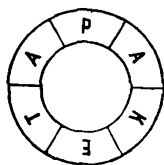


Рис. 53. Слово «ракета».

7. Определить, имеются ли положительные решающие разности. Если да, то перейти к команде 9.

8. Конец. Строки, лежащие выше горизонтальной границы, озаглавлены буквами первого класса (это обычно гласные), лежащие ниже — буквами второго класса.

9. Выбрать строку с максимальной положительной решающей разностью и перенести ее через горизонтальную границу, а соответствующий столбец — через вертикальную границу. Перейти к выполнению пункта 6.

Можно привести небольшой пример, показывающий, как работает этот алгоритм. Отметим, что для работы этого алгоритма не нужно знать, где проходит граница между словами. В частности, будем считать, что мы не знаем и то, с какой буквы текст начинается, а какой кончается, как если бы он был записан по кругу (рис. 53).

*) Эти числа представляют собой частоты пар одинаковых букв. Легко видеть, что они входят в сумму $\sum_i \sum_j \varphi(a_i, a_j) + \sum_k \sum_l (a_k, a_l)$ при любом разбиении алфавита на два класса, и поэтому не влияют на качество данного разбиения.

Таблица чисел $\varphi(a_i, a_j)$ для этого текста имеет следующий вид:

	Р	А	К	Е	Т
Р		2			
А	2		1		1
К		1		1	
Е			1		1
Т		1		1	

По пункту 2 таблица не претерпевает изменений. Выполняем пункт 3:

	Р	А	К	Е	Т	
Р		2				2
А	2		1		1	4
К		1		1		2
Е			1		1	2
Т		1		1		2

Далее выполняем пункты 4 и 5:

	А	Р	К	Е	Т
А		2	1		1
Р	2				
К	1			1	
Е			1		1
Т	1			1	

Пункт 6:

	А	Р	К	Е	Т	
А		2	1		1	
Р	2					-2
К	1			1		0
Е			1		1	2
Т	1			1		0

Пункт 7 требует выполнять девятую команду. Выполнение пунктов 9 и 6 дает:

	А	Е	Р	К	Т	
А			2	1	1	
Е				1	1	
Р	2					
К	1	1				
Т	1	1				

Пункт 7 требует перейти к пункту 8 и окончить процедуру. Ответ: первый класс: а, е; второй р, к, т.

В табл. 4.3 даются результаты подобного машинного эксперимента по обработке русского, английского и французского текстов объемов в 10 000 букв каждый.

Легко видеть, что эти результаты для русского и французского языка практически безошибочны. Заметим, что мягкий и твердый знаки изображают некогда существовавшие, но исчезнувшие из живой речи гласные звуки; французская буква «k» встречается весьма редко, в данном тексте, по-видимому, в сокращениях (например, в инициалах при нефранцузских фамилиях).

Что же касается английского языка, то ошибка здесь не случайна, а вызвана использованием буквы t в сочетании, передающем отдельный звук. В работе [12] излагается алгоритм, исправляющий ошибки такого рода, и приводятся результаты успешных экспериментов.

На стр. 312 помещен один из «математически правильных» алгоритмов, минимизирующих K_1 . Далее, на стр. 313 приведен примыкающий к данному алгоритм, переводящий так называемое «слоговое письмо» в обычное, а на стр. 319 алгоритм, находящий классы слов с общим смыслом.

§ 6. Алгоритмы склеивания (начало)

Алгоритм, выделяющий кодовые цепочки. Алгоритмы, предназначенные для обнаружения более крупных единиц, когда более мелкие известны, представляют, по-видимому, один из важнейших классов алгоритмов.

Мы начнем разбор этих алгоритмов с одного из самых простых: с алгоритма, выделяющего буквы, закодированные цепочками символов равномерной длины.

Значение этой задачи для космического случая очевидно. Не исключено, что «элементарный сигнал» сообщения ВЦ будет иметь простую форму, в частности, допускающую представление в виде одного из двух возможных символов, например, 0 и 1. Естественно, что для изображения более длинного алфавита в этом случае придется прибегать к кодированию букв последовательностями элементарных сигналов, и вполне возможно, что эти последовательности будут одинаковой длины для любых букв алфавита.

Множество допустимых решений (множество интерпретаций) в этом случае описывается несложно. Пусть длина кодовых цепочек равна m , а длина текста, выраженная в элементарных символах, равна N . В этом случае число допустимых решений равняется m и определяется числом сдвигов текста на i разрядов ($i=0, 1, \dots, m-1$). При этом следует

Русский

		Гласные															
		О А Е И У Ъ Ь Я Ы Э Ю												Б	В	Г	Д
		О	А	Е	И	У	Ъ	Ь	Я	Ы	Э	Ю					
Гласные	О	28	15	39	32	7	4	3	16	2	3	7	72	167	136	107	
	А	15	6	10	22	10	2	1	30	1	2	13	35	127	43	99	
	Е	39	10	28	57	7	5	15	5	19	4	4	46	83	37	93	
	И	32	22	57	52	13	9	2	36	4	2	2	23	95	25	60	
	У	7	10	7	13	0	3	2	4	1	3	19	20	20	26	42	
	Ъ	4	2	5	9	3	0	0	11	0	3	6	5	21	2	12	
	Ь	3	1	15	2	2	0	0	1	0	0	0	26	27	1	8	
	Я	16	30	5	36	4	11	1	8	0	0	0	8	31	6	15	
	Ы	2	1	19	4	1	0	0	0	0	0	0	14	13	1	11	
	Э	3	2	4	2	3	3	0	0	0	0	0	0	4	0	0	
	Ю	7	13	4	2	19	6	0	0	0	0	2	5	3	0	2	
Согласные	Б	72	35	46	23	20	5	26	8	14	0	5	0	2	0	1	
	В	167	127	83	95	20	21	27	31	13	4	3	2	8	5	24	
	Г	136	43	37	25	26	2	1	6	1	0	0	0	5	0	19	
	Д	107	99	93	60	42	12	8	15	11	0	2	1	24	19	2	
	Ж	24	40	44	26	12	0	0	4	0	0	0	0	1	0	13	
	З	30	90	35	37	11	4	6	6	4	0	3	1	20	5	10	
	И	56	10	34	31	0	0	11	1	10	1	0	2	5	4	2	
	К	143	147	43	83	41	20	4	10	3	3	2	7	17	3	4	
	Л	192	173	166	148	24	42	22	30	13	2	9	12	16	16	9	
	М	118	67	87	102	26	9	31	10	15	0	1	0	16	6	5	
	Н	181	228	202	143	43	40	43	46	37	0	2	3	16	7	39	
	П	152	46	71	42	27	8	7	14	8	1	4	0	17	1	4	
	Р	121	120	156	68	42	9	14	11	8	0	1	9	21	19	24	
	С	133	103	129	89	34	70	7	60	5	1	5	55	44	1	7	
	Т	192	110	124	95	53	59	15	31	14	23	13	2	21	2	10	
	Ф	6	11	0	5	0	0	0	1	0	0	0	0	0	8	0	
	Х	44	42	16	28	10	0	12	4	6	0	0	4	8	1	0	
	Ц	4	9	19	18	2	2	4	0	3	0	0	0	4	0	8	
	Ч	31	60	63	43	16	10	1	1	0	0	1	0	3	0	0	
	Ш	25	50	22	39	16	7	5	1	4	0	0	0	21	0	0	
	Щ	6	4	27	21	8	1	0	1	0	0	14	2	0	0	0	

Таблица 4.3

Я З Ы К

Согласные

Ж З Й К Л М Н П Р С Т Ф Х Ц Ч Ш Щ

24	30	56	143	192	118	181	152	121	133	192	6	44	4	31	25	6
40	90	10	147	173	67	228	46	120	103	110	11	42	9	60	50	4
44	35	34	43	166	87	202	71	156	129	124	0	16	19	63	22	27
26	37	31	83	148	102	143	42	68	89	95	5	28	18	43	39	21
12	11	0	41	24	26	43	27	42	34	53	0	10	2	16	16	8
0	4	0	20	42	9	40	8	9	70	59	0	0	2	10	7	1
0	6	11	4	22	31	43	7	14	7	15	0	12	4	1	5	0
4	6	1	10	30	10	46	14	11	60	31	1	4	0	1	1	1
0	4	10	3	13	15	37	8	8	5	14	0	6	3	0	4	0
0	0	1	3	2	0	0	1	0	1	23	0	0	0	0	0	0
0	3	0	2	9	1	2	4	1	5	13	0	0	0	1	0	14

0	1	2	7	12	0	3	0	9	5	2	0	4	0	0	0	2
1	20	5	17	16	16	16	17	21	44	21	0	8	4	3	21	0
0	5	4	3	16	6	7	1	19	1	2	8	1	0	0	0	0
13	10	2	4	9	5	39	4	24	7	10	0	0	8	0	0	0
0	2	0	12	4	0	17	0	10	1	0	0	0	0	0	0	0
2	0	3	6	10	14	20	3	5	2	6	0	2	1	0	0	0
0	3	0	11	6	4	14	12	8	23	13	0	3	1	0	3	0
12	6	11	4	19	8	11	10	40	57	15	0	5	0	3	5	0
4	10	6	19	2	10	13	21	4	43	8	0	9	0	5	12	0
0	14	4	8	10	16	23	10	12	21	5	0	2	1	3	0	0
17	20	14	11	13	23	80	7	12	32	20	0	9	8	12	2	2
0	3	12	10	21	10	7	0	59	17	8	0	6	0	2	0	1
10	5	8	40	4	12	12	59	0	16	48	0	20	1	2	4	0
1	2	23	57	43	21	32	17	16	16	160	0	6	0	12	1	0
0	6	13	15	8	5	20	8	48	160	4	1	3	0	22	0	1
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	2	3	5	9	2	9	6	20	6	3	0	0	0	0	2	0
0	1	1	0	0	1	8	0	1	0	0	0	0	0	1	0	0
0	0	0	3	5	3	12	2	2	12	22	0	0	1	0	1	0
0	0	3	5	12	0	2	0	4	1	0	0	2	0	1	0	0
0	0	0	0	0	0	2	1	0	0	1	0	0	0	0	0	0

Французский

	Г л а с н ы е												
		е	а	о	і	и	у	к	б	с	д	ф	г
Г л а с н ы е	е	96	67	6	114	166	6	0	40	168	283	33	45
	а	67	16	0	116	90	22	1	27	63	81	34	37
	о	6	0	0	118	150	3	0	35	73	19	14	2
	і	114	116	118	4	68	1	4	13	45	43	17	12
	и	166	90	150	68	4	0	0	9	25	57	3	10
	у	6	22	3	1	0	0	0	0	4	1	0	0
	к	0	1	0	4	0	0	0	0	1	1	0	0
С о г л а с н ы е	б	40	27	35	13	9	0	0	0	0	0	0	0
	с	168	63	73	45	25	4	1	0	16	2	0	0
	д	283	81	19	43	57	1	1	0	2	12	1	1
	ф	33	34	14	17	3	0	0	0	0	1	10	0
	г	45	37	2	12	10	0	0	0	0	1	0	0
	и	46	25	22	8	3	0	0	0	56	0	0	0
	ј	30	8	10	5	7	0	0	0	0	1	0	0
	л	318	173	43	137	70	11	0	21	15	1	11	2
	т	219	72	71	55	10	11	0	4	1	1	0	0
	п	302	166	253	100	92	3	0	2	55	83	11	22
	р	127	127	60	17	17	7	0	0	3	5	0	0
	к	27	5	0	5	94	0	0	0	0	2	0	1
	г	428	194	118	85	111	0	1	13	27	64	7	18
	с	457	144	88	147	146	7	1	11	37	59	5	3
	т	283	118	64	140	83	2	0	5	36	37	3	3
	в	109	89	43	39	20	0	0	0	0	2	0	0
	ш	0	0	0	0	0	0	0	0	0	0	0	0
	х	14	7	1	10	33	0	3	2	3	5	1	0
	з	12	1	0	1	0	0	0	0	2	0	0	0

Продолжение табл. 4.3

ЯЗЫК

Согласные

h	j	l	m	n	p	q	r	s	t	v	w	x	z
46	30	318	219	302	127	27	428	457	283	109	0	14	12
25	8	173	72	166	127	5	194	144	118	89	0	7	1
22	10	43	71	253	60	0	118	88	64	43	0	1	0
8	5	137	55	100	17	5	85	147	140	39	0	10	1
3	7	70	10	92	17	94	111	146	83	20	0	33	0
0	0	11	11	3	7	0	0	7	2	0	0	0	0
0	0	0	0	0	0	0	1	1	0	0	0	3	0
0	0	21	4	2	0	0	13	11	5	0	0	2	0
56	0	15	1	55	3	0	27	37	36	0	0	3	2
0	1	1	1	83	5	2	64	69	37	2	0	5	0
0	0	11	0	11	0	0	7	5	3	0	0	1	0
0	0	2	0	22	0	1	18	3	3	0	0	0	0
0	0	8	0	4	2	0	1	8	1	0	0	2	0
0	0	1	0	2	0	0	1	5	4	0	0	0	2
8	1	74	3	24	28	7	36	88	44	3	0	2	2
0	0	3	62	5	19	0	12	9	11	0	0	1	0
4	2	24	5	60	14	14	16	100	170	12	0	0	0
2	0	28	19	14	22	0	58	49	28	0	0	7	2
0	0	7	0	14	0	0	8	20	6	0	0	1	0
1	1	36	12	16	58	8	96	58	160	10	0	0	0
8	5	88	9	100	49	20	58	152	112	11	0	1	0
1	4	44	11	170	28	6	160	112	54	10	0	1	0
0	0	3	0	12	0	0	10	11	10	0	0	2	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	2	1	0	7	1	0	1	1	2	0	0	0
0	2	2	0	0	2	0	0	0	0	2	0	0	0

Английский

		Г л а с н ы е											
		е	о	а	і	т	и	у	б	с	д	ф	г
Г л а с н ы е	е	116	48	112	63	108	8	52	70	65	204	57	56
	о	48	114	15	20	157	115	17	29	39	71	97	41
	а	112	15	6	48	138	4	44	24	40	89	27	64
	і	63	20	48	0	102	9	8	5	26	73	37	39
	т	108	157	138	102	104	57	32	8	13	60	35	20
	и	8	115	4	9	57	0	0	32	9	13	7	34
	у	52	17	44	8	32	0	2	18	5	17	8	5
С о г л а с н ы е	б	70	29	24	5	8	32	18	4	0	15	1	3
	с	65	39	40	26	13	9	5	0	0	5	1	1
	д	204	71	89	73	60	13	17	15	5	22	11	12
	ф	57	97	27	37	35	7	8	1	1	11	16	4
	г	56	41	64	39	20	34	5	3	1	12	4	6
	х	430	91	97	140	352	12	20	6	46	49	12	45
	ж	1	0	0	0	1	3	0	1	0	2	0	0
	к	35	38	16	23	7	0	2	1	11	0	2	1
	л	135	119	104	88	43	28	42	7	8	63	27	2
	м	109	70	87	48	16	14	16	6	1	20	2	4
	н	154	125	214	203	119	63	13	11	16	155	5	106
	р	61	30	34	64	19	14	5	1	1	5	4	1
	к	1	0	1	8	0	12	1	0	0	1	0	1
	г	312	139	105	119	77	31	33	18	14	44	18	31
	с	170	80	170	121	139	52	25	13	23	39	20	25
	в	111	14	30	23	1	0	3	0	0	1	0	0
	w	106	91	90	53	28	0	19	1	1	17	7	9
	х	7	0	1	0	0	0	0	0	1	0	0	0
	z	1	0	1	1	0	0	1	0	0	0	0	0

предполагать, что за последней буквой текста следует первая.

Если N не делится на m без остатка, то остаток мы будем отбрасывать. Отсюда общее число возможных решений равно $\sum_{m=1}^{N/2} \left[\frac{N}{m} \right]$, где $\left[\frac{N}{m} \right]$ означает целую часть от деления N на m . Это число не больше

$$\sum_{m=1}^{N/2} \frac{N}{m} < \frac{N^2}{2} \int_1^{N/2} \frac{dm}{m} = \frac{N^2}{2} \ln \frac{N}{2}.$$

Приступим к обсуждению оценочной функции. Пусть некоторый текст длины N закодирован цепочками цифр длины m . Чем такой текст отличается от случайной последовательности цифр, разбитой на такие же блоки длины m ? Кажется очевидным, что частоты m -ок во втором случае должны быть значительно равномернее, чем в первом. Ведь если вторая последовательность получена выбором цифр «наугад», то это означает, что мы не оказывали предпочтения ни одной из цифр в сравнении с другими.

В то же время выбор тех или иных букв в обычном сообщении далеко не случаен. Например, существуют звуки и последовательности звуков, более и менее удобные для произношения; если алфавит сообщения — это множество слов, то разные слова тоже резко различны по своей употребительности как в зависимости от своей «обычности», так и в зависимости от смысла сообщения.

Если закодированный текст содержит блоки длины m , а мы пытаемся интерпретировать его как содержащий блоки длины p ($p \neq m$) или хотя бы блоки длины m , но со смещением на i разрядов (где i не делится нацело на m), то сообщение становится более похожим на последовательность символов, полученную путем «выборки с возвратом». Это утверждение равносильно тому, что элементы цепочки, кодирующей букву, связаны теснее, чем символы, принадлежащие к разным кодовым цепочкам,

Это выглядит правдоподобным, потому что при неправильной «нарезке на блоки» уже отсутствуют те «критерии предпочтительности», которые ограничивали сочетаемость букв.

Поэтому естественно в качестве оценочной функции выбрать функцию, имеющую экстремальное значение для равномерного распределения частот символов и в то же время для некоторого «очень» неравномерного с другой.

К сожалению, интуитивных соображений недостаточно для априорного выбора какой-нибудь определенной функции, оценивающей разнообразие.

Существует целый ряд традиционных способов такой оценки. Сюда, например, относится вычисление среднего квадратичного отклонения, среднего отклонения по модулю, энтропии.

Подсчеты, проводившиеся нами на небольшом тексте, показали пригодность функции

$$V = \sum_i (\varphi(c_i) - \overline{\varphi(c)})^2.$$

Здесь c_i означает некоторую цепочку заданной длины, $\overline{\varphi(c)}$ — среднюю абсолютную частоту цепочки этой длины, равную $\frac{N}{m \cdot |A|^m}$, где m — длина цепочки, N — длина текста в единичных символах, $|A|$ — количество букв в алфавите единичных символов, $|A|^m$ — количество букв в алфавите цепочек длины m , $\frac{N}{m}$ — число цепочек длины m в данном тексте (округленно); таким образом, V есть сумма квадратов отклонений от среднего значения абсолютной частоты цепочки.

Для данной длины цепочек функция V принимает минимальное значение — нуль, когда все частоты цепочек равны, и максимальное, когда один символ повторяется на протяжении всего текста.

Чтобы можно было сравнивать результаты, полученные для разных длин цепочек, выражение

$$\sum_i (\varphi(c_i) - \overline{\varphi(c)})^2$$

умножается на нормирующий множитель v . Этот множитель можно вычислить, если исходить из предположения, что наилучшее (максимальное) значение V не должно зависеть от m . Так как в наилучшем (с точки зрения данной функции) решении встречается лишь один символ, то он имеет частоту $\frac{N}{m}$, а остальные — нуль. При этом V равно

$$\left(\frac{N}{m} - \frac{N}{m \cdot |A|^m}\right)^2 + \left(\frac{N}{m \cdot |A|^m}\right)^2 (|A|^m - 1) = \\ = \frac{N^2}{m^2} \left[\left(1 - \frac{1}{|A|^m}\right)^2 + \frac{|A|^m - 1}{|A|^{m \cdot 2}} \right] = \frac{N^2 (|A|^m - 1) \cdot |A|^m}{m^2 |A|^{m \cdot 2}},$$

так как обычно $|A|^m$ велико, то $|A|^m - 1 \approx |A|^m$, что дает $V_{\max} \approx \frac{N^2}{m^2}$.

Пусть в некотором другом решении длина цепочек равна l ; максимум V будет тогда примерно равен $\frac{N^2}{l^2}$.

Нормирующий множитель вводится так, чтобы уравнивать наилучшие значения

$$\frac{N^2}{m^2} = v \frac{N^2}{l^2}.$$

Отсюда

$$v = \frac{l^2}{m^2}.$$

В приводимом ниже примере за «основу сравнения» принимается разбиение на блоки длиной $m=3$; для разбиений на цепочки длины 3 коэффициент v будет равен 1.

Короткий английский текст *) закодирован последовательностью троичных чисел по следующему

*) Первое четверостишие из стихотворения Р. Л. Стивенсона «Сумерки»:

The twilight is sad and cloud,
The wind blows wild and free,
And as the wings of sea birds
Flash the white caps of the sea.

правилу:

$a = 000$	$j = 100$	$s = 200$
$b = 001$	$k = 101$	$t = 201$
$c = 002$	$l = 102$	$u = 202$
$d = 010$	$m = 110$	$v = 210$
$e = 011$	$n = 111$	$w = 211$
$f = 012$	$o = 112$	$x = 212$
$g = 020$	$p = 120$	$y = 220$
$h = 021$	$q = 121$	$z = 221$
$i = 022$	$r = 122$	

Последняя возможная цепочка длины 3 не используется. Текст будет выглядеть так:

201 021 011 201 211 022 020 021 102 022 020 021
 201 022 200 200 000 010 000 111 010 002 102
 112 022 010 220 201 021 011 211 022 111 010
 001 102 112 211 200 211 022 102 010 000 111 010
 012 122 011 011 000 111 010 000 200 201 021 011
 211 022 111 020 200 112 012 200 011 000 001 022
 122 010 200 012 102 000 200 021 201 021 011 211
 021 022 201 011 002 000 120 200 112 012 201 021 011 200 011 000

В табл. 4.4 приведены абсолютные частоты троек, для разбиений, начинающихся соответственно с первой, второй и третьей букв текста и обозначенных R_1^3 , R_2^3 и R_3^3 :

Таблица 4.4

R_1^3	200 011 000 021 010 022 201 211 102 111 012 112 020 001	10 10 9 9 8 8 8 6 5 5 4 4 3 2
R_2^3	110 010 000 001 002 100 112 210 020 122 220 221 021 202	15 8 7 6 6 6 6 6 5 4 4 4 3 3
R_3^3	102 101 020 000 201 001 011 120 220 100 121 211 002 202	14 11 9 7 6 5 5 5 5 4 4 4 3 3
R_1^3	002 122 120 202 220 100 101 110 121 210 212 221 222	2 2 1 1 1 0 0 0 0 0 0 0 0
R_2^3	102 120 121 200 212 222 011 012 211 022 101 111 201	2 2 2 2 2 2 1 1 1 0 0 0 0
R_3^3	210 110 200 212 012 021 022 221 010 111 112 122 222	3 2 2 2 1 1 1 1 0 0 0 0 0

Множитель ν принят равным 1. Получаем:

$$V(R_1^3) = 340,30; \quad V(R_2^3) = 277,34, \quad V(R_3^3) = 321,37.$$

Абсолютные частоты для блоков длины 2 и для разбиений, начинающихся с первой буквы текста (R_1^2) и со второй, (R_2^2) таковы:

Т а б л и ц а 4.5

R_1^2	10	00	20	01	11	02	21	22	12
	26	24	19	17	15	14	12	11	9
R_2^2	02	00	01	11	10	20	21	12	22
	25	23	21	17	16	16	12	11	6

Принимая нормирующий множитель равным $\nu = \frac{2^2}{3^2} = 0,44$, получим $V(R_1^2) = 268,00 \cdot 0,44 = 117,92$, $V(R_2^2) = 295,44 \cdot 0,44 = 129,99$. Оба значения заметно меньше, чем $V(R_1^3)$. Абсолютные частоты для $m=1$ следующие:

$$\varphi(0) = 124; \quad \varphi(1) = 94; \quad \varphi(2) = 76.$$

Нормирующий множитель равен $\frac{1}{9}$; получаем $V(R_1^1) = 1176 \cdot \frac{1}{9} = 130,67$. Это также значительно меньше, чем $V(R_1^3)$.

Рассмотрим абсолютные частоты для блоков длины 4. В данном случае возможны четыре разбиения, начинающиеся соответственно с первой, второй, третьей и четвертой букв текста (обозначения: R_1^4 , R_2^4 , R_3^4 , R_4^4).

Мы изобразим частоты символов в таблице (4.6).

Нормирующий множитель равен $\frac{4^2}{3^2} = \frac{16}{9} = 1,67$ и значения V таковы: $V(R_1^4) = 165,77$; $V(R_2^4) = 169,02$; $V(R_3^4) = 141,38$; $V(R_4^4) = 125,88$.

Чтобы убедиться в том, что V принимает максимальное значение именно для правильного разби-

Таблица 4.6

R_1^4	Абсол. частота		R_2^4	Абсол. частота		R_3^4	Абсол. частота		R_4^4	Абсол. частота	
	Число блоков			Число блоков			Число блоков			Число блоков	
	5	1		5	2		4	1		4	2
	4	2		3	5		3	9		3	7
	3	4		2	12		2	9		2	13
	2	12		1	24		1	24		1	27
	1	24		0	38		0	38		0	35
	0	38									

ния, следовало бы вычислить значения этой функции и для всех m ($m = 1, 2, \dots, \frac{N}{2}$).

Но, по-видимому, в этом нет нужды: очевидно, что с ростом m квадраты разностей $(\varphi(c_i) - \overline{\varphi(c)})^2$ сильно убывают, а v возрастает умеренно.

Можно также думать, что вообще никогда нет смысла рассматривать кодовые цепочки длиннее, например двадцати символов; ведь даже для двухэлементного множества элементарных символов мощность алфавита цепочек такой длины равна 2^{20} , т. е. с лихвой удовлетворит потребности изображений самого сложного алфавита (включая алфавит типа китайского).

Интересно сравнить значения энтропии, вычисленные нами для некоторых из этих разбиений.

Энтропия $H = - \sum_i p_i \log p_i$ для равномерного распределения принимает максимальное значение; она равна нулю, если вероятность одного символа равна 1, а остальных — нулю. Вместо p мы брали $p = \frac{\varphi(c_i)}{\frac{N}{m}}$,

где $\frac{N}{m}$ — длина текста в цепочках длины m .

Имеем: $H(R^3)=1,165$; $H(R_2^3)=1,248$; $H(R_3^3)=1,226$, что тоже вполне удовлетворительно. Для прочих значений длины кодовых цепочек энтропию следует нормировать делением на $\log m$. Любопытно отметить, что почти все значения V от неправильных разбиений близки: это свидетельствует об удовлетворительности нормировки.

Пример использования понятия разумности. На основе приведенного алгоритма оказывается возможным проиллюстрировать соображения о разумности, изложенные в § 2. Для этого отождествим множество интерпретаций текста с множеством допустимых решений данного алгоритма, а разумность с разностью функций V или H для наихудшего и наилучшего разбиения.

Результаты, которые могут быть получены при исследовании такого рода, можно описать на примере. Допустим, алфавит элементарных сигналов содержит два символа: 0 и 1, а длина кодовой цепочки равна двум. В этом случае множество интерпретаций содержит всего два разбиения: R_1^2 , при котором цепочки начинаются с нечетных символов, и R_2^2 , при котором цепочки начинаются с четных символов.

Как выглядят эти интерпретации в наилучшем случае, когда $\Delta(H) = H(R_1^2) - H(R_2^2)$ максимально? Охарактеризуем эти разбиения распределениями вероятностей кодовых цепочек:

	Разбиения	
	R_1^2	R_2^2
распределения	$p'(00)$	$p''(00)$
	$p'(01)$	$p''(01)$
	$p'(10)$	$p''(10)$
	$p'(11)$	$p''(11)$

Эти вероятности не независимы. Их можно выразить через вероятности цепочек длины 4, начинающихся, например, с символов, номера которых в тексте кратны четырем:

$$\begin{aligned}
p'(00) &= p(0000) + p(0001) + p(0010) + p(0011) \\
p'(01) &= p(0100) + p(0101) + p(0110) + p(0111) \\
p'(10) &= p(1000) + p(1001) + p(1010) + p(1011) \\
p'(11) &= p(1100) + p(1101) + p(1110) + p(1111) \\
p''(00) &= p(0000) + p(0100) + p(1000) + p(1100) \\
p''(01) &= p(0001) + p(0101) + p(1001) + p(1101) \\
p''(10) &= p(0010) + p(0110) + p(1010) + p(1110) \\
p''(11) &= p(0011) + p(0111) + p(1011) + p(1111)
\end{aligned}$$

Если мы подставим в выражение для $\Delta(H)$ значения p' и p'' , выраженные через $p(\alpha\alpha\alpha\alpha)$, то мы получим функцию от 16 переменных, максимум которой позволит вычислить оба распределения. Распределения $\{p'_i\}$ и $\{p''_i\}$ будут характеризовать текст, в некотором смысле максимально разумный! Взяв средние

вида $p_i^{(2)} = \frac{p'_i + p''_i}{2}$, мы получим распределение вероятностей пар элементарных символов этого текста, а суммируя $p(01) + p(00)$ и $p(01) + p(11)$ — распределение вероятностей одиночных символов $\{p_i^{(1)}\}$. Эти распределения дают возможность вычислить условную энтропию второго порядка данного текста по формуле $H_2 = H(p_i^{(2)}) - H(p_i^{(1)})$ и, приближенно, избыточность (вернее, нижний предел избыточности) по формуле $1 - \frac{H_\infty}{H_0}$, где вместо H_∞ подставляется H_2 , а $H_0 = \log 2$.

Весьма интересные результаты можно было бы получить, если бы «понятность» оценивалась энтропией бесконечного порядка. Вычисления, проделанные при этом предположении, позволили бы оценить уровень энтропии и избыточности, характерные для разумных текстов.

§ 7. Алгоритмы дешифровки изображений

Язык изображений. Связность и детализированность. Предложенное нами представление о понятности не является ни общепринятым, ни даже обычным. Как правило, текст считается понятным, если он вызывает в мозгу человека некоторые образы реальной действительности.

Правда, подобный подход не помогает ответить на вопрос, почему сами явления действительности могут быть как понятными, так и непонятными. Однако известная доля правды в обычном представлении все же имеется. Дело в том, что между «предсказуемыми системами» человеческого языка и действительности, несомненно, имеется существенное сходство, иначе говоря, слова сочетаются друг с другом примерно так же, как и обозначаемые ими явления действительности.

Так, если человеку свойственно говорить, передвигаться, воздействовать на окружающие предметы, то и слову «человек» естественно сочетаться со словами, обозначающими речь, передвижение, действие. Весьма примечательно соответствие между предложениями или фразами, обычно состоящими из набора слов, обозначающих предметы (существительные), действия (глаголы) и свойства (прилагательные), и типичными явлениями действительности (ситуациями), складывающимися из предметов, их взаимодействий и свойств.

Очевидно, это соответствие далеко не просто, хотя и не настолько сложно, чтобы для каждого сообщения на одном из языков приходилось задавать годные только для него правила перевода.

Разумеется, особенно сложно переводить с языка действительности на язык типа человеческого и обратно; однако существует язык, употребляемый людьми, для которого подобный перевод осуществляется сравнительно просто. Мы имеем в виду язык изображений.

Конечно, такой язык страдает значительными недостатками. Он весьма неэкономичен: достаточно, например, сравнить высказывание «человек идет» с рисунком, сообщающим об этом факте. Правильное изображение непременно должно содержать много деталей, которые в данном случае могут быть несущественными. Далее, возможны сообщения, которые нельзя перевести на язык изображений, не нарушив простоты преобразования, связывающего изображение с действительностью (например, сообщения типа

«терпение и труд все перетрут» и даже «в 1963 году была засуха»).

Поэтому нет оснований думать, что язык изображений может быть единственным средством космического общения.

Однако его бесспорным положительным свойством является понятность, причем не только вызванная легкостью перевода на привычный нам язык действительности, но и тем, что соседние участки изображения сильно прогнозируют друг друга.

Наглядность дешифровочных задач, связанных с анализом изображений, является весьма благоприятным обстоятельством, позволяющим рассматривать эти задачи как модели более трудных задач по дешифровке обычных языков.

Одна из типичных дешифровочных задач при изучении языка изображений заключается в следующем: допустим, у нас имеется последовательность сигналов: требуется перекодировать ее в двухмерный рисунок так, чтобы образовалось осмысленное сообщение — изображение. Эта задача характерна тем, что она может вызвать наибольшие сомнения в целесообразности вводить формальное понятие «разумности». Дело в том, что человек обладает способностью практически мгновенно отличать осмысленное изображение от неосмысленного.

Допустим, последовательность сигналов представляет строки прямоугольной развертки изображения, состоящего из черных и белых точек (изображаемых 1 и 0 соответственно), идущих одна за другой. Дешифровать эту последовательность мы могли бы следующим образом: изменять некоторый параметр d от 1 до N (N — длина последовательности), резать последовательность на строки длины d и располагать эти строки одну под другой.

Нет никакого сомнения, что человек, осматривающий получающиеся изображения, легко выберет из них самое правильное.

Читатель может без труда проделать такой эксперимент с известным рисунком («космограммой») Кейда:

```

1 0 0 1 1 1 1 1 0 0 1 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1
0 0
1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0
1
1 0 0 1 0 1 1 1 1 1 0 1 0 0 1 0 1 1 1 1 1 0 1 0 0 1 0 1 1 1 1 1
0
1 1 1 1 1 0 1 0 0 1 0 1 1 1 1 1 0 1 0 0 1 0 1 1 1 1 1 0 1 0 0 1
0
1 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 0 1 1 0 1 1
0
1 1 0 1 1 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 0 0 1 1 0 1 1 0 0 0 0 0
0
0 0 1 0 0 1 1 0 1 1 0 0 1

```

Расшифровка космограммы:

```

1 0 0 1 1 1 1 1 0 0 1
0 0 1 1 1 1 1 1 1 0 0
0 0 0 1 0 1 0 1 0 0 0
0 0 0 1 1 1 1 1 0 0 0
0 0 0 0 1 1 1 0 0 0 0
0 1 1 1 1 1 1 1 1 1 0
0 1 0 1 1 1 1 1 0 1 0
0 1 0 1 1 1 1 1 0 1 0
0 1 0 1 1 1 1 1 0 1 0
0 1 0 1 1 1 1 1 0 1 0
0 1 0 1 1 1 1 1 0 1 0
0 0 0 1 1 0 1 1 0 0 0
0 0 0 1 1 0 1 1 0 0 0
0 0 0 1 1 0 1 1 0 0 0
0 0 0 1 1 0 1 1 0 0 0
0 0 0 1 1 0 1 1 0 0 0
0 0 0 1 1 0 1 1 0 0 0
0 0 0 1 1 0 1 1 0 0 1

```

Единицы образуют изображения человека в шляпе.

Очевидно также, что такой способ дешифровки изображений на прямоугольном экране будет и самым эффективным, если возложить задачу получения множества возможных решений (т. е. нарезку текста на куски длины l и расположение их друг под другом) на машину. Однако положение резко меняется, если допустить возможность сообщений, являющихся развертками экрана произвольной формы. В этом случае начинает играть роль практически неограниченное возрастание числа допустимых решений.

При такой ситуации и машина не сможет осмотреть и оценить все допустимые решения. Однако если машина обладает формальным критерием осмысленности, то она сможет пользоваться сокращенным способом перебора, опираясь, например, на переход от несколько менее осмысленного изображения или части изображения к более осмысленному. При этом степень различия осмысленности сравниваемых изображений может быть и совсем неочевидной, незаметной на глаз.

Предположим, что развертка изображения в последовательность сигналов происходит следующим образом: изображение покрывается двумя системами x и y координатных линий, причем линии каждой системы отстоят друг от друга на некоторое расстояние Δr , вычисляемое по заданному правилу. Участок изображения, находящийся между двумя соседними линиями системы x , назовем строкой, участок, расположенный между соседними линиями системы y , — столбцом. Участок текста, расположенный между соседними линиями системы x и y , назовем точкой. Если дана некоторая классификация точек, то каждую точку мы сможем заменить номером класса точек, пронумеровать строки, теперь уже содержащие цифры, и расположить строки одну за другой.

Предварительной гипотезой при построении критерия качества изображения является предположение о том, что системе координатных линий можно задать некоторым наилучшим для данного изображения способом.

При этом мы будем опираться на одно свойство текстов, которое, по-видимому, является довольно общим. Можно думать, что разумные сообщения распадаются на некоторые части, напоминающие высказывания, в свою очередь состоящие из более мелких частей (нечто вроде слов). При этом «квазислова», составляющие «квазивысказывания», должны быть в некотором смысле по возможности различными;

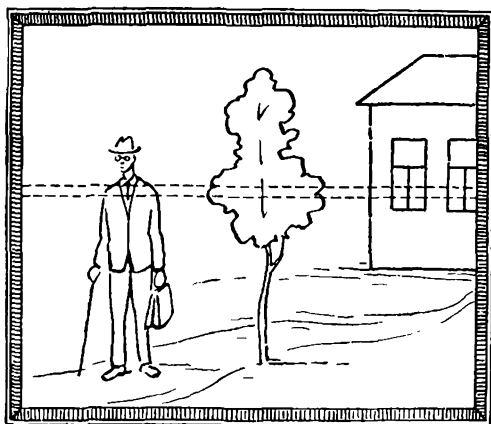


Рис. 54. В этом изображении, богатом вертикальными линиями, «горизонтальные строки» похожи.

в то же время соседние «квазивысказывания» должны быть в некотором смысле похожими.

Так, предложения русского текста состоят из слов, выражающих весьма разнообразные вещи: из существительных, означающих предметы, глаголов, являющихся именами процессов или состояний, прилагательных, выражающих свойства. Весьма различны и образы действительности, обозначаемые различными словами предложения. В то же время соседние предложения не только имеют похожую структуру, но нередко содержат и слова со сходным значением, что придает тексту ощущаемую нами «связность». На пример, фразы «Из-за горизонта показался корабль

Это был пароход с широкой белой трубой», имеют «общую тему», выражаемую словами «пароход», «корабль», и выглядят «связными».

Эти свойства для изображений кажутся еще нагляднее: если рис. 54 разрезать на горизонтальные строки, то внутри строки черные и белые точки будут часто чередоваться; в то же время соседние строки будут очень похожими и в столбцах смена черной и белой точек будет нечастой.

Нетрудно заметить, что это предположение будет справедливым лишь тогда, когда разбиение на строки



Рис. 55. Изображения, богатые горизонтальными, радиальными и concentрическими линиями.

и столбцы производится определенным образом: так, для изображения крокодила строки должны быть вертикальными, а столбцы — горизонтальными; для изображения цветка строки должны идти по окружностям, а яблока — по радиусам (рис. 55).

Аналогичное свойство характерно для сообщений на формальных языках и на языках типа линкоса: в текстах на этих языках рядом расположены высказывания, связанные отношением «логического следования». Эти высказывания обычно похожи графически.

По-видимому, для довольно сложных изображений выбор правильных координатных линий играет меньшую роль, поскольку при любом выборе направления строк две соседние строки будут иметь сходный вид и при этом содержать частые смены черных и белых точек. Мы будем говорить, что сходство соседних строк развертки обеспечивает связность изображений, сходство более отдаленных — гладкость, а разнообразие внутри строк — детализированность.

Примеры оценочных функций. Некоторые процедуры. Можно предложить простую функцию, оценивающую качество изображений путем измерения связности и детализированности. При этом детализированность будет оцениваться количеством переходов от темной точки к светлой в строке, а связность — количеством переходов, расположенных в соседних строках друг под другом. Обозначим темную точку единицей, а светлую нулем. Тогда функцию $u_{i, i+1}$, оценивающую сходство соседних строк, можно будет записать следующим образом:

$$u_{i, i+1} = \varphi(|01) + \varphi(|10),$$

где $\varphi(|01)$ означает количество переходов от светлой точки к темной в $i+1$ -й строке, находящихся под таким же переходом в i -й строке, а $\varphi(|10)$ — количество таких же двойных переходов от темной точки к светлой. Качество изображения, вычисленное по строкам, будет выражаться формулой

$$U_{\text{стр}} = \sum_i u_{i, i+1}.$$

Поскольку заранее не известно, строки или столбцы следует считать «квазивысказываниями», качество можно измерять и по столбцам, определяя сходство соседних столбцов по формуле

$$u'_{j, j+1} = \varphi(\overline{01}) + \varphi(\overline{10}),$$

где $\varphi(\overline{01})$ означает количество переходов от светлой точки к темной по вертикали, находящихся рядом и правее такого же перехода в соседнем столбце. Качество изображения по столбцам $U_{\text{столб}}$ выразится тогда формулой

$$U_{\text{столб}} = \sum_j u'_{j, j+1},$$

а полное качество U формулой

$$U = \sum_i u_{i, i+1} + \sum_j u'_{j, j+1}. \quad (4.4)$$

Суммирование можно проводить по одним строкам ввиду равенства суммы $\sum_i u'_{i,j+1}$ сумме $\sum_i u'_{i,i+1}$, где $u'_{i,i+1} = \varphi(0\bar{1}) + \varphi(\bar{1}0)$ в соседних строках.

Не очень ясно, как относиться к первому и последнему символам в соседних строках. Нам представляется, что разбивка на строки эквивалентна

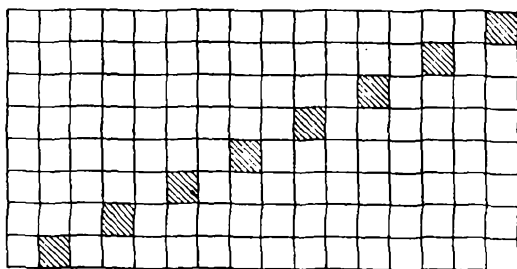


Рис. 56. Изображение наклонной непрерывной линии с помощью квадратной сетки.

внесению особых «граничных символов». Конечно, переход к граничному символу несет некоторую информацию и должен был бы влиять на качество изображения, если бы он сообщался вместе со всем изображением, но поскольку мы предполагаем, что в данном дешифруемом сообщении таких символов нет, эта информация «примыслена» и ее следует минимизировать.

Ради простоты мы будем считать, что изображение не ограничено ни по вертикали, ни по горизонтали, т. е. нанесено на тор (баранку); за последней строкой идет первая, за крайним правым столбцом — крайний левый. Если последняя строка оказывается неполной, то она дополняется более частым символом, например, нулями.

Приведем несколько первых расшифровок текста

00000010100101001110000100001000000

Правильной интерпретацией является расшифровка с длиной строки в пять символов (изображение четверки). Этой длине соответствует и наибольшее значение $U=10$.

Приведенная оценочная функция пригодна для изображений, богатых сплошными и тонкими вертикальными или горизонтальными линиями. Однако она будет давать ошибки в случае разрывных изображений, а также в случае изображений с преобладающими наклонными линиями. Оба случая взаимосвязаны; в самом деле, при использовании квадратной сетки непрерывное изображение наклонной линии невозможно (при сохранении толщины линии). Рассмотрим рис. 56.

В этом изображении все центры черных квадратиков лежат на прямой $y = \frac{1}{2}x + \frac{1}{4}$. Легко видеть, что невозможно зачернить еще хотя бы один квадратик при соблюдении этого условия. Оценочная функция, устраняющая эти трудности, использует то, что можно назвать гладкостью изображения. Над парой строк λ_i и λ_j мы определим особую операцию, которую мы назовем «линейный прогноз», ставящую в соответствие строкам λ_i и λ_j строку λ_p .

Эта операция выполняется следующим образом: единицы строки λ_i соединяются с единицами строки λ_j отрезками с соблюдением трех условий:

1. Каждая единица строки λ_i соединена по крайней мере с одной единицей строки λ_j .
2. Каждая единица строки λ_j соединена по крайней мере с одной единицей строки λ_i .
3. При соблюдении условий 1 и 2 сумма длин отрезков минимальна.

Далее отрезки продолжаются на произвольные расстояния. Если отрезки продолжены до $j+k=t$ -й строки, то мы будем говорить, что максимальная глубина прогноза равна k . Прогноз глубины k осуществляется так: квадраты, в центры которых попали проведенные линии, считаются черными (единицами), остальные — белыми (нулями). В случае прямоугольного

экрана длина строки q_p та же, что и длина строк q_1 и q_j (в приводимом примере $k=1$) (рис. 57).

В противном случае вначале определяется местоположение граничных точек (рис. 58).

Затем полученная строка сравнивается с реальной строкой λ_i , обнаруженной на этом месте, с помощью функции $u_{r,t}$.

Видимо, практически почти всегда достаточно сравнивать две соседние строки и делать прогноз глубины 1, т. е. о соседней строке.

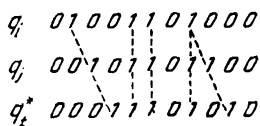


Рис. 57. Прогноз глубины 1 на прямоугольном экране.

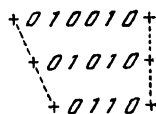


Рис. 58. То же, что и на рис. 57, на произвольном экране.

Если обозначить число $u_{r,t}$, полученное в результате прогноза по строкам λ_i и λ_{i+1} , через $u_{r,t}^*(\lambda_i, \lambda_{i+1})$, то качество изображения можно оценить функцией, вычисляемой по формуле

$$U_{\text{стр}}^* = \sum_{i=1}^n u_{r,t}^*(l_i, l_{i+1}).$$

Пример. Для изображения

0 0 0 0 0 0 0
 0 1 0 0 0 1 0
 0 0 1 0 1 0 0
 0 0 0 1 0 0 0
 0 0 1 0 1 0 0
 0 1 0 0 0 1 0
 0 0 0 0 0 0 0

$U_{\text{стр}}^*$ имеет значение 10, что гораздо лучше значения $U_{\text{стр}}=0$.

Аналогично «линейному прогнозу» можно было бы определить операцию нелинейного прогноза, определяющую по трем строчкам четвертую. Например,

$$\begin{array}{cccccccc} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}$$

Здесь сохраняется разность между двумя соседними смещениями по горизонтали.

Изложенный способ преодоления трудностей, вызванных разрывностью, с логической точки зрения не вызывает серьезных нареканий и не выглядит чересчур произвольным. Более простой способ приводится ниже.

Наличие наклонных линий, составленных из квадратов, соприкасающихся углами, можно учесть, если в формуле (5.4) строки заменить диагоналями, идущими снизу вверх, а столбцы — диагоналями, идущими сверху вниз. Обозначив пару переходов по диагональным «строкам», расположенных рядом в диагональных «столбцах» через /01 или /10, а пару таких же переходов в диагональных столбцах через \01 или \10, мы сможем записать сходство диагоналей формулами $u_{i, i+1}^{\text{диаг}} = \varphi(/10) + \varphi(/10)$ и $u'_{j, j+1} = \varphi(\backslash 01) + \varphi(\backslash 10)$ соответственно и «диагональное качество» $U^{\text{диаг}}$ формулой $U^{\text{диаг}} = \sum_i u_{i, i+1}^{\text{диаг}} + \sum_j u'_{j, j+1}$. Полное качество изображения выразим суммой обоих «качеств»:

$$U^{\text{полн}} = U + U^{\text{диаг}}.$$

Впрочем, для более или менее сложных изображений достаточно, вероятно, пользоваться функцией u .

На рис. 59 и в табл. 4.7 приводится «космограмма» Дрейка. Рядом даны значения $u_{i, i+1}$; $u'_{i, i+1}$ и соответствующие суммы. Заметна существенная разница в значении u для правильной и неправильной интерпретации. Оба рисунка считаются размещенными на торе.

Для сколько-нибудь крупных изображений прямой перебор всех значений длины строки становится затруднительным, особенно при работе вручную.

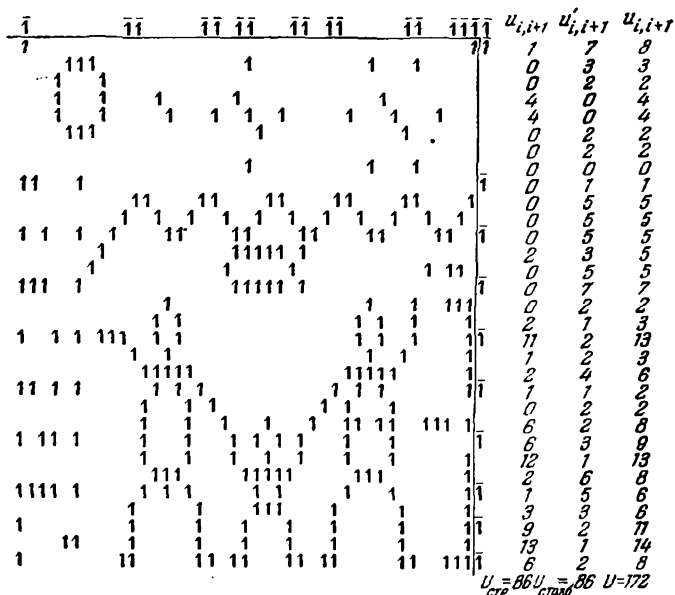


Рис. 59. Космограмма Дрейка. Правильная интерпретация (длина строки равна 41). Перед первой строкой — последняя строка, за последним столбцом — первый столбец.

Поэтому мы предложим менее надежный, но зато более быстрый способ. Этот же способ положен в основу приемов дешифровки изображений на экране произвольной формы, где перебор возможных значений длины строки вообще невозможен.

Если вставить между двумя символами некоторый граничный символ μ , то для него можно будет указать номер i места от некоторой начальной точки μ_0 . Будем говорить, что точка μ_i обладает U -окрестностью, если по обе стороны от граничного символа, находящегося на месте i от начальной точки, найдены отрезки длины d такие, что при расположении

правого отрезка под левым мы получим две строки λ_1, λ_2 , для которых $u_{1,2} + u'_{1,2} \geq U$.

Обозначим множество точек, обладающих U -окрестностью, через $\{\mu_i(U)\}$. Простейшая процедура, использующая свойства U -окрестностей, основывается на предположении, что существует $U_{\max}$ такое, что $|\{\mu_i(U_{\max})\}| = 1$ (мощность множества точек, обладающих U -окрестностью, равна единице). Это истолковывается так: в изображении имеется наилучшая пара строк, сильно сходных между собой и проходящих через часть изображения, богатую деталями. Таким образом, $\{\mu_i(U_{\max})\}$ содержит лишь одну точку $\mu_{\text{наил}}$. Длина U -окрестности для $\mu_{\text{наил}}$ и будет длиной строки; $\mu_{\text{наил}}$ укажет точку отсчета.

Поэтому оказывается применимой следующая процедура: задаются значениями $U = 1, 2, 3 \dots$ и определяют, имеется ли более одной точки, для которой U принимает данное значение. Если да, то значение U увеличивается на единицу и поиск повторяется. Если нет, то, следовательно, имеется одна точка, обладающая U -окрестностью. При наличии такой точки ($\mu_{\text{наил}}$) длина ее окрестности определит длину строки, а расположение укажет на начало (конец) строк. Если такой точки нет, то длиной строки объявляется длина U -окрестности какой-нибудь из точек, обладающих $(U - 1)$ -окрестностью. В этом случае возможна многозначность ответа.

Пример. Дешифровать изображение

010100101001110000100001.

1) Определим, имеется ли более одной точки μ_i , для которой $U \geq 1$? Да; например, μ_2, μ_7 .

2) То же для $U \geq 2$? Да; например, те же μ_2, μ_7 .

3) То же для $U \geq 3$? Нет; имеется лишь одна точка μ_5 , для которой $U_{\max} = 4$; длина строки равна 5. Ответ: четверка на прямоугольном экране.

При дешифровке изображения на произвольном экране для нахождения $\mu_{\text{наил}}$ будем действовать по тем же правилам; затем снизу и сверху к полученной паре строк будем добавлять по строке, определяя ее длину и расположение, исходя из требования макси-

мальности прироста U . Затем, производя изменения начал и концов соседних строк, не затрагивающие значения U , добьемся максимальной гладкости границ, для чего используем какую-нибудь из предложенных функций.

Пример. Дано сообщение

```
000000111110000000000010001000000000111110000000101010000
10101000111110000000000.
```

Все точки $\mu_{\text{наил}}$ расположены вплотную между 53-м и 58-м местами (это — особенность сообщений на произвольном экране). Вот пара строк, соответствующая одной из точек:

```
0000010101
0000010101
```

Для этих строк $U_{1,2}=6$.

Наилучшее расположение следующей строки таково:

```
0000010101
0000010101
00011111
```

Здесь $U_{2,3}=2$.

Построение следующих строк никак не изменяет значения U . Сверху строки пристраиваются следующим образом:

```
00000011111
0000000000010001
000000000111110
00000010101
```

и в целом

```
00000011111
0000000000010001
000000000111110
0000010101
0000010101
00011111
```


что дает правильный ответ: изображение «окошка». Уточнение границ экрана необходимо, так как исходное изображение таково:

```

00000011111000000
000001000100000
0000111110000
00010101000
001010100
0111110
00000
000
0

```

т. е. «окошко» на треугольном экране. Однако, как мы видим, оно уже не затрагивает «существа» изображения.

§ 8. Алгоритмы, аналогичные алгоритмам, строящим двуязычные словари

Алгоритмы сравнения букв, использующие свойства близких окрестностей. В предыдущих параграфах мы познакомились с различными алгоритмами, анализирующими текст на неизвестном языке.

Мы уже упоминали, что целью такого анализа является построение наилучшей интерпретации, содержащей сведения, которые позволяют наиболее эффективным способом прогнозировать еще неосмотренные части текста при его произвольном разбиении на осмотренную и неосмотренную части.

Допустим, что мы получили такую интерпретацию. Теперь, осмотрев часть исследуемого текста, мы сможем предвидеть, что последует дальше. Может быть, мы научимся строить «правильные фразы», а, возможно, и «правильные тексты» на этом языке.

Однако мы вправе спросить себя: представляют ли полученные сведения именно то, к чему мы стремились, принимаясь за дешифровку неизвестного текста? Ответ очевиден: конечно, нет. Ведь мы не научились переводить полученное сообщение на русский

язык, не умеем переводить его и на «язык действительности», например, создать механизм, описываемый дешифруемым сообщением.

Чтобы научиться переводить с неизвестного текста, надо узнать, какие элементы русского текста соответствуют некоторым элементам дешифруемого. Сопоставляемые элементы могут в принципе выбираться в том и другом языке множеством различных способов; например, можно составлять списки предложений неизвестного языка и их переводов на русский; можно составлять списки слов и их переводов; однако, по-видимому, естественнее и проще сопоставлять некоторые лингвистические явления, на основе которых строится предсказуемая система обоих языков. Мы уже упоминали, что основой изучения смысла текста должны быть «семантические классы слов». Двухязычный словарь, содержащий вместо слов названия семантических классов, был бы и короче и лучше обыкновенного двухязычного словаря; что же касается словаря, содержащего вместо слов фразы, то его создание и вообще нереально даже для пары известных языков.

Итак, для перевода с одного языка на другой нам нужно уметь составлять двухязычные словари некоторых простейших явлений, из которых состоит текст. Однако, очевидно, что одного этого мало. Нужно, кроме того, уметь сопоставлять правила соединения элементов того и другого языка. Ведь из одних и тех же слов можно составлять фразы с совершенно разным смыслом.

Иначе говоря, нужно уметь определять «соотношения близости» в обоих языках и, строя перевод, следить, чтобы слова-переводы связывались теми же соотношениями близости, что и слова-оригиналы.

Впрочем, при дешифровке, очевидно, не следует добиваться того, чтобы создаваемые алгоритмы выдавали готовый русский текст. Хорошо, если они дадут достаточно сведений для построения перевода человеком. Поэтому создание «словаря» и «сопоставительной грамматики» можно считать последней целью дешифровочных алгоритмов.

Далее мы приведем алгоритм, составляющий нечто вроде словаря, но не для слов или семантических элементов, а, как это обычно принимается в нашей работе, для букв. Предполагается, что некоторый язык типа «человеческого» (т. е. произносимый) неизвестен, а другой язык, тоже человеческий, известен. Мы знаем, как читаются буквы известного языка, но не знаем, как читаются буквы неизвестного. Ставится задача: описать произношение букв неизвестного языка буквами известного. В простейшем случае это можно сделать, указав «правильную» взаимно однозначную подстановку, переводящую буквы неизвестного языка в буквы известного.

Мы предложим простой алгоритм, находящий подстановку. Будут также рассмотрены некоторые средства для отыскания ответа в более общем виде.

Итак, множество допустимых решений в разбираемом случае будет представлять собой множество взаимно однозначных отображений букв неизвестного языка в буквы известного. Если число букв в сравниваемых алфавитах неравное, то алфавиты можно уравнивать, добавив в меньший алфавит буквы, которые считаются «случайно не появившимися» в исследуемых текстах.

Гипотезой, используемой при построении критерия оценок подстановок, является предположение о том, что буквы, передающие сходные звуки, должны иметь похожую «сочетаемость». Конкретный вид алгоритма зависит от того, что понимается под сочетаемостью.

Сходное предположение естественно сделать и при составлении словарей, содержащих «семантические множители»*) и слова. Иными словами, мы предполагаем, что и элементы, имеющие сходные значения, должны сочетаться сходным образом. Хотя в некоторых случаях это допущение кажется абсолютно не очевидным, оно все же представляет единственную мыслимую основу для дешифровки.

*) То есть названия классов слов, сходных по смыслу.

В данном алгоритме сочетаемость букв мы будем представлять таблицей T частот появления пар букв в непосредственном соседстве. Число $P(a_i, a_j)$, стоящее в строке i и столбце j , определяется по формуле $P(a_i, a_j) = \frac{\varphi(a_i a_j) + \varphi(a_j a_i)}{2N}$, где a_i и a_j — буквы данного алфавита, $\varphi(a_i, a_j)$ — количество появлений пары букв a_i и a_j в данном тексте, N — длина текста.

Под сочетаемостью буквы a_i мы будем понимать строчку чисел вида $P(a_i, a_x)$ из таблицы T .

Для букв a_i и a_j одного и того же языка мы можем измерить несходство по одной из формул, описывающих расстояние между точками n -мерного пространства.

Введем следующую меру несходства букв a_i и a_j :

$$\bar{\sigma}(a_i, a_j) = \sum_{k=1}^n |P(a_i, a_k) - P(a_j, a_k)|,$$

где a_i, a_j, a_k — буквы данного алфавита.

Пусть теперь задана некоторая подстановка $\pi: A \rightarrow B$ алфавита неизвестного языка A в алфавит B известного языка. Обозначим через a'_i образ буквы a_i при подстановке π . Тогда величина

$$\bar{\sigma}(a_i, a'_i, \pi) = \sum_{k=1}^n |P(a_i, a_k) - P(a'_i, a_k)|$$

будет характеризовать несходство (по сочетаемости) буквы a_i и ее образа a'_i при подстановке π .

Далее, качество подстановки π можно определить как сумму несходств пар букв, входящих в эту подстановку, если их несходство определять по этой же подстановке

$$\rho = \sum_{i=1}^n \bar{\sigma}(a_i, a'_i, \pi) \quad (Z)$$

(ρ — качество подстановки).

Итак, следует искать такую подстановку $\pi_{\text{наил}}$ (наилучшую подстановку), что

$$\rho(\pi_{\text{наил}}) = \min.$$

В принципе $\pi_{\text{наил}}$ мы могли бы найти, перебрав все возможные подстановки и оценив качество р каждой из них по формуле (Z). Однако число подстановок очень велико ($n!$), поэтому мы предложим некоторый неполный перебор, который, как мы надеемся, практически равноценен полному.

Будем говорить, что π_j есть транспозиция от π_i или $\pi_j (\pi_i)$, если некоторым парам $a_r \rightarrow a'_r$, $a_s \rightarrow a'_s$ подстановки π_i соответствуют пары $a_r \rightarrow a'_s$, $a_s \rightarrow a'_r$ подстановки π_j , а остальные пары обеих подстановок идентичны. Как известно, для двух произвольных подстановок π_u и π_v всегда можно построить цепочку подстановок $\pi_x, \dots, \pi_i, \dots, \pi_1$, где π_1 есть транспозиция от π_v ; π_x есть транспозиция от π_u и для любой пары $\pi_i \pi_{i-1} (i \leq x; i > 1)$ π_i есть транспозиция от π_{i-1} . Поэтому и искомая подстановка достижима из произвольной, если переходить от произвольной к некоторой транспозиции, и т. д.

Транспозиция π_j некоторой подстановки π_i может быть наилучшей подстановкой лишь в том случае, если $\rho(\pi_j) \leq \rho(\pi_i)$. Если для некоторой π_i и любой $\pi_j (\pi_i)$ имеет место $\rho(\pi_j) \geq \rho(\pi_i)$, мы будем говорить, что в точке π_i функция ρ имеет локальный минимум. Очевидно, что если ρ имеет в точке $\pi_{\text{наил}}$ абсолютный минимум, то в той же точке $\pi_{\text{наил}}$ функция ρ имеет и локальный минимум.

Поэтому мы нашли бы абсолютный минимум $\pi_{\text{наил}}$, перебрав все возможные локальные минимумы.

Такого способа в настоящее время мы предложить не можем, однако мы дадим способ, отыскивающий, по-видимому, довольно глубокий локальный минимум.

Существо этого способа сводится к следующему. Для произвольной подстановки π_i определяется множество транспозиций $\{\pi_j (\pi_i)\}$. Для каждой транспозиции $\pi_j (\pi_i)$ определяется приращение функции $\Delta\rho$, равное $\rho(\pi_i) - \rho(\pi_j)$. Затем выбирается такая π_q , что $\Delta\rho(\pi_q)$ максимально и положительно в множестве $\{\Delta\rho(\pi_j (\pi_i))\}$. Аналогичная процедура повторяется для $\{\pi_t (\pi_q)\}$ и далее. Процедура останавливается, ко-

гда найдена подстановка, для которой не существует транспозиций с положительными приращениями.

Для вычисления $\Delta\rho$ необязательно определять $\rho(\pi_i)$ и $\rho(\pi_j)$. Если π_i отличается от π_j тем, что в π_i имеются пары $a_r \rightarrow a'_r, a_s \rightarrow a'_s$, а в π_j — пары $a_r \rightarrow a'_s, a_s \rightarrow a'_r$, то приращение $\rho(\pi_j) - \rho(\pi_i)$ можно вычислить по следующей несколько громоздкой формуле:

$$\begin{aligned} 2 \sum_x |P(a_r, a_x) - P(a'_r, a'_x)| + 2 \sum_x |P(a_s, a_x) - P(a'_s, a'_x)| - \\ - |P(a_r, a_r) - P(a'_r, a'_r)| - |P(a_s, a_s) - P(a'_s, a'_s)| - \\ - 2 \sum_x |P(a_r, a_x) - P(a'_s, a'_x)| - 2 \sum_x |P(a_s, a_x) - P(a'_r, a'_x)| + \\ + |P(a_r, a_r) - P(a'_s, a'_s)| - |P(a_s, a_s) - P(a'_r, a'_r)|. \end{aligned}$$

В качестве начальной подстановки целесообразно выбирать, например, такую, которая получается при выстраивании обоих алфавитов рядом в порядке убывания чисел

$$\sum_i P(a_i, a_x) \quad \text{и} \quad \sum_x P(a'_i, a'_x)$$

соответственно.

В начальную подстановку будут входить пары букв $a_i \rightarrow a'_i$, имеющие один и тот же номер i в соответствующих последовательностях.

Изложим еще раз этот алгоритм в виде системы обобщенных команд.

1. Для текста на неизвестном языке составляется таблица T_1 чисел

$$P(a_i, a_j) = \frac{\varphi(a_i a_j) + \varphi(a_j a_i)}{2N},$$

где $f(a_i a_j)$ — количество появлений цепочки букв $a_i a_j$, N — длина текста.

2. Аналогичную таблицу T_2 составить для текста на известном языке.

3. Одинаково ли число рядов в 1-й и 2-й таблицах? Если нет, то дополнить меньшую таблицу

рядами, состоящими из одних нулей до равенства размеров.

4. Строки и столбцы обеих таблиц расположить в порядке убывания чисел $\sum_x P(a_i, a_x)$ и $\sum_x P(a'_i, a'_x)$.

5. Построить все $\frac{n(n-1)}{2}$ возможных пар вида a'_i, a'_j и для каждой из них подсчитать число $\Delta p(a'_i, a'_j) = \Delta p$.

6. Имеются ли $\Delta p(a'_i, a'_j)$, большие нуля? Если нет, то выполнять пункт 7, если да, то пункт 9.

7. Выдать ответ: множество пар вида $a_i \rightarrow a'_i$.

8. Конец.

9. Поменять местами строку и столбец a'_i со строкой и столбцом a'_j во второй таблице, для которых $\Delta p(a'_i, a'_j)$ максимально; выполнять пункт 5.

Вполне аналогичным образом можно было бы поступить и в том случае, когда сочетаемость представляется несимметричной матрицей, заполненной числами вида $P(a_i, a_j) = \frac{f(a_i, a_j)}{N}$.

В этом случае единственное отличие состояло бы в том, что числа Δ определялись бы по другой формуле.

Теперь вкратце изложим идею построения алгоритма, использующего преобразование более общего вида.

Будем отыскивать преобразование Π , удовлетворяющее следующим требованиям:

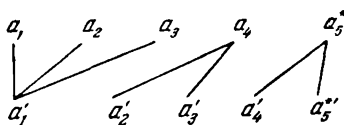
1. Для любой буквы a_i ($a_i \in A$) найдется пара $a_i, a'_i \in \Pi$.

2. Для любой буквы a'_j ($a'_j \in B$) найдется пара $a_j, a'_j \in \Pi$.

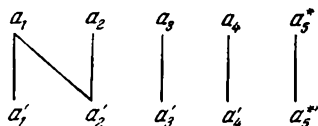
3. В каждый из алфавитов добавляется по меньшей мере одна «фиктивная» буква a_i^* такая, что $p(a_i^*, a_x) = 0$ для любой буквы a_x .

4. Сумма чисел $P(a_i, a'_i)$ при соблюдении условий 1—3 минимальна.

Поэтому, в частности, преобразование



имеет допустимый вид, а преобразование



недопустимый: в нем можно выбросить пару $a_1 a'_2$ при соблюдении условий 1—3.

Отличие алгоритма, использующего этот тип преобразований, в следующем: элементарными изменениями будут не только обмены местами рядов второй таблицы, но и удвоение ряда как в табл. T_1 , так и в табл. T_2 .

Элементарные изменения будут распадаться на вынужденные и свободные. В самом деле, применение некоторого элементарного изменения может дать преобразование недопустимого вида. Чтобы придать преобразованию допустимый вид, следует произвести еще некоторую последовательность элементарных изменений.

Элементарные изменения, не направленные на возвращение цепочки допустимого вида, назовем свободными. Истинной выгодой свободного элементарного изменения назовем выгоду самого свободного элементарного изменения плюс сумму выгод наивыгоднейших элементарных изменений, вынужденных данным свободным элементарным изменением и приводящих преобразование к допустимому виду.

На каждом шаге процедуры должна определяться истинная выгода всех свободных элементарных изменений. Затем сразу совершается вся цепочка элементарных изменений, соответствующих наиболее истинно выгодному свободному элементарному изменению,

Англий

	e	t	o	h	a	n	s	d	r	i	w
e	440	774	160	2173	574	794	1059	1059	1484	230	530
t	774	520	749	2018	804	530	704	295	325	759	175
o	160	749	440	305	85	624	350	270	654	195	385
h	2173	2018	305	70	659	120	385	215	145	530	235
a	574	804	85	659	20	1054	829	505	644	224	589
n	794	530	624	120	1054	10	250	819	125	1109	110
s	1059	704	350	385	829	250	500	200	220	709	190
d	1059	295	270	215	505	819	200	90	145	355	90
r	1484	325	654	145	644	125	220	145	120	395	60
i	230	749	195	530	225	1109	709	355	395	0	215
w	530	175	385	235	589	110	190	90	60	215	60
c	420	80	205	215	340	135	145	35	90	215	0
f	265	245	640	105	165	35	140	65	150	155	40
l	604	125	405	40	455	35	140	215	75	435	30
m	515	95	315	5	325	15	140	70	85	260	20
u	75	315	410	60	95	175	280	40	290	40	5
g	230	80	110	260	190	530	50	35	70	300	10
b	330	40	175	10	110	35	60	125	115	40	0
p	270	55	140	25	290	15	130	30	120	125	5
k	170	35	130	30	55	110	75	20	15	95	65
v	345	0	75	5	130	30	25	20	15	135	0
y	230	120	100	45	195	40	90	40	90	40	50
z	25	5	10	0	10	10	0	0	0	15	0
x	75	20	0	0	0	5	0	0	0	35	0
j	15	5	5	0	10	0	10	0	5	0	0
q	5	0	0	0	0	0	5	0	0	0	5

С К И Й Я З Ы К

[illegible]

[illegible]

СКИНЬ ЯЗЫК

с	р	т	у	q	b	g	f	h	j	x	z	y	w	k
949	574	884	654	115	160	215	175	190	175	85	165	40	10	0
185	145	95	75	80	60	35	70	15	45	10	5	20	0	0
270	514	415	290	25	145	155	180	135	55	30	10	80	0	0
225	65	400	100	40	15	145	50	25	30	5	0	10	5	0
70	65	175	140	90	100	95	85	65	5	85	15	0	0	0
195	255	50	15	60	30	15	20	0	25	25	5	15	0	0
185	80	100	105	575	70	40	30	5	85	80	5	5	0	0
205	205	80	95	50	110	90	35	0	15	10	0	40	0	0
50	175	15	15	65	85	10	60	15	5	25	20	35	0	0
450	245	305	285	5	240	55	85	70	60	10	5	25	5	0
0	15	15	0	5	0	5	10	25	0	20	30	5	0	0
10	15	15	5	15	0	10	0	229	0	30	15	5	0	0
15	70	145	0	0	0	0	0	30	0	15	5	0	0	0
15	145	145	0	5	65	10	0	0	0	10	20	10	0	0
5	0	0	0	0	0	0	0	0	0	10	25	5	0	0
15	0	5	0	0	0	9	0	0	0	09	10	0	0	0
0	0	65	0	0	0	0	0	0	0	5	0	0	0	0
5	0	10	0	5	0	5	0	0	0	15	0	0	0	0
0	0	0	0	0	0	0	90	0	0	0	0	0	0	0
220	30	0	0	0	0	0	0	0	0	0	5	0	0	0
0	0	0	0	0	0	0	0	0	0	5	0	0	0	0
30	15	10	10	5	5	15	0	0	5	0	0	0	0	0
15	5	20	25	10	0	0	0	5	0	0	0	0	0	0
5	0	10	5	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

если эта выгода положительна. Процесс обрывается, когда истинные выгоды всех возможных свободных элементарных изменений становятся отрицательными.

По приведенному алгоритму были проведены эксперименты на машине М-20. Для этого выбирались тексты на английском и французском языках по 5000 букв.

Были получены следующие соответствия (рис. 60).

Таблицы 4.8 и 4.9 содержат относительные частоты двух буквенных цепочек английского и французского языков. Для удобства обозрения все числа умножены

<i>e</i>	<i>t</i>	<i>o</i>	<i>h</i>	<i>a</i>	<i>n</i>	<i>s</i>	<i>d</i>	<i>r</i>	<i>i</i>	<i>w</i>	<i>c</i>	<i>f</i>	<i>l</i>	<i>m</i>	<i>u</i>	<i>g</i>	<i>b</i>	<i>p</i>	<i>k</i>	<i>v</i>	<i>y</i>	<i>z</i>	<i>x</i>	<i>j</i>	<i>q</i>
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
<i>e</i>	<i>s</i>	<i>a</i>	<i>n</i>	<i>i</i>	<i>t</i>	<i>u</i>	<i>r</i>	<i>l</i>	<i>o</i>	<i>d</i>	<i>c</i>	<i>p</i>	<i>m</i>	<i>v</i>	<i>q</i>	<i>b</i>	<i>g</i>	<i>f</i>	<i>h</i>	<i>j</i>	<i>x</i>	<i>z</i>	<i>y</i>	<i>w</i>	<i>k</i>

Рис. 60. Подстановка, полученная в результате эксперимента по сопоставлению английских и французских букв.

на 10 000. Порядок следования букв в цепочках во внимание не принимался.

Строчки (и столбцы) с одинаковыми по порядку номерами должны быть похожими, так как таблицы уже обработаны изложенным алгоритмом. Следует признать, что сходство соответствующих рядов не представляется значительным.

Очевидно, что результаты сопоставления совершенно неудовлетворительны. Они во всяком случае не лучше тех, которые могли бы получиться при сопоставлении букв просто по частоте.

Заслуживает интереса обсуждение того, почему результаты оказались настолько плохими.

Нам представляется, что основная причина в искажениях, вносимых орфографией. Так, англ. *th* передает один звук, а англ. *y* — иногда гласную, иногда согласную. Впрочем, заслуживает удивления и то, что эти искажения не сыграли существенной роли в эксперименте по выделению гласных и согласных.

Неудовлетворительные результаты проведенного эксперимента показывают трудность непосредственного сопоставления элементов сравниваемых языков.

Однако мы покажем, что предварительный анализ весьма сильно улучшает качество сопоставления. Например, разобьем предварительно буквы того и другого языка на гласные и согласные (этот алгоритм, как мы знаем, работает практически безошибочно) и будем сопоставлять таблицы, озаглавленные классами Г (гласные) и С (согласные).

	Г	С
Г		
С		

Так как числа в блоках ГС (СГ) значительно больше, чем в блоках ГГ и СС, то для правильного сопоставления достаточно подсчитать суммы в блоках ГГ и СС. Имеем *)

		ГГ	СС
французский	текст	1972	4302
английский	текст	1330	5261

Отсюда мы можем получить, что первый класс французских букв (гласные) соответствует первому классу английских букв. Далее, разобьем гласные того и другого языка на подклассы:

французский текст:

	е, а, о	і, и, у
е, а, о	256	785
і, и, у	785	146

английский текст:

	е, о	а, і, и, у
е, о	326	385
а, і, и, у	385	234

*) Вычисления проводились по табл. 4.3.

Сопоставление этих табличек дает соответствие между французскими *e, a, o*, английскими *e, o*, французскими *i, u, y* и английскими *a, i, u, y*. Разбиение на подклассы согласных дает для французского языка

s, r, n, l, m, h, x, z (1-й класс)

b, d, g, p, t, k, c, q, f, v, w, j (2-й класс),

для английского

s, r, n, l, m, h, x, q, j (1-й класс)

b, d, g, p, t, k, c, f, w, v (2-й класс)

Диагональные квадраты соответствующих таблиц равны:

	1 кл., 1 кл.	2 кл., 2 кл.
французский язык	1174	416
английский язык	936	372

И здесь мы получаем правильное соответствие. Дальнейшее разбиение на подклассы в принципе могло бы дать более детальные сведения о соответствии букв двух сравниваемых языков, однако (при небольшом тексте) используемая информация (о совместной встречаемости) становится слишком скудной.

Возможно, следует использовать предыдущий алгоритм с учетом предварительного разбиения на классы и с ограничениями, запрещающими сопоставление букв не соответствующих классов.

Приведенные расчеты убеждают в том, что для правильного сопоставления необходимо предварительное исследование «предсказуемости системы» данного текста. В частности, теперь представляется очевидным, что прямое сопоставление слов сравниваемых языков обречено на неуспех. Следует, конечно, сравнивать предварительно полученные «семантические классы».

Алгоритм, использующий далекие окрестности. Алгоритмы, разобранные выше, конкретизируют так называемый «статистический» принцип военной дешифровки. Возможно и построение алгоритма, использующего вторую идею военной дешифровочной практики:

метод «характерных слов». Рассмотрим множество допустимых решений и оценочную функцию такого алгоритма.

Этот алгоритм также находит соответствие в форме подстановки. Оценим вначале качество пары a_x, a'_x , входящей в подстановку π . Обозначим букву a_x , входящую в неизвестный текст на месте p , через $a_x(p)$, а букву a'_x , находящуюся на месте t известного текста, через $a'_x(t)$. Назовем пару цепочек букв допустимой u, v -окрестностью пары $a_x(p), a'_x(t)$, если для каждого $k (k < u)$ и $m (m \geq v)$ пара $a_y(p+k), a'_y(t+k)$ и пара $a_y(p-m), a'_y(t-m)$ входят в π . Разность $u - v$ назовем длиной допустимой u, v -окрестности. Длину максимально длинной допустимой u, v -окрестности пары a_x, a'_x назовем качеством пары a_x, a'_x или $q(a_x, a'_x)$. Качество Q подстановки π можно определить суммой качеств составляющих ее пар:

$$Q(\pi) = \sum_x q(a_x, a'_x).$$

Подстановка тем лучше, чем больше $Q(\pi)$. Для наилучшей подстановки $\pi_{\text{наил}} Q(\pi_{\text{наил}})$ минимальна.

Если предыдущие алгоритмы этого параграфа использовали свойства ближайших окрестностей, то данный алгоритм использует сходство длинных цепочек, содержащих рассматриваемые буквы. Вероятно, возможно объединение обоих принципов.

§ 9. Алгоритмы классификации (окончание)

«Математически» правильный алгоритм выделения гласных и согласных. Ниже мы приведем один из «математически» правильных алгоритмов, минимизирующих K_1 .

Будем говорить, что для разбиения $R = k_1, k_2$ алфавита A на два класса функция K_1 принимает локально минимальное значение, если перенос какой-нибудь из букв класса k_2 в класс k_1 , или наоборот, не приводит

к разбиению k_1^*, k_2^* , для которого значение K_1 меньше, чем для $R = k_1, k_2$.

Если вообще не существует разбиения, для которого значение K_1 меньше, чем для R , то мы будем говорить, что для разбиения R функция K_1 принимает абсолютно минимальное значение.

Очевидно, что если для R функция принимает абсолютно минимальное значение, то для этого же разбиения она принимает и локально минимальное значение. Поэтому если бы мы нашли все локальные минимумы (разбиения, для которых K_1 локально минимальна), то мы среди них нашли бы и абсолютный минимум.

Назовем подмножество $\mathfrak{R}$ ($\mathfrak{R} \subset A$) допустимым, если для каждого a_i ($a_i \in \mathfrak{R}$) имеет место соотношение

$$\sum_{i=m+1}^n \varphi(a_i, a_k) - \sum_{i=1}^m \varphi(a_i, a_j) \geq 0.$$

Здесь $m = |\mathfrak{R}|$, $a_i \in \mathfrak{R}$; $a_j \in \mathfrak{R}$; $a_k \in A \setminus \mathfrak{R}$.

Имеют место две следующие теоремы:

Теорема 1. Если в разбиении k_1, k_2 оба класса допустимы, то для k_1, k_2 функция K_1 принимает локально минимальное значение.

Допустим, что это неверно, т. е. существует буква a_x , перенос которой, например, из класса k_1 в класс k_2 уменьшает значение K_1 . Обозначим классификацию, получаемую переносом a_x из k_1 в k_2 , через k'_1, k'_2 , так что $k'_1, k'_2 = k_1 \setminus a_x, k_2 \cup a_x$. Значение K_1 для k'_1, k'_2 равно

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^m \varphi(a_i, a_j) + \sum_{k=m+1}^n \sum_{l=m+1}^n \varphi(a_k, a_l) - \\ - 2 \sum_{i=1}^m \varphi(a_x, a_i) + 2 \sum_{k=m+1}^n \varphi(a_x, a_k). \end{aligned}$$

Значение K_1 для k_1, k_2 равно

$$\sum_{i=1}^m \sum_{j=1}^m \varphi(a_i, a_j) + \sum_{k=m+1}^n \sum_{l=m+1}^n \varphi(a_k, a_l).$$

По предположению, имеем

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^m \varphi(a_i, a_j) + \sum_{k=m+1}^n \sum_{l=m+1}^n \varphi(a_k, a_l) - \sum_{i=1}^m \sum_{j=1}^m \varphi(a_i, a_j) - \\ - \sum_{k=m+1}^n \sum_{l=m+1}^n \varphi(a_k, a_l) + 2 \sum_{i=1}^m \varphi(a_x, a_i) - \\ - 2 \sum_{k=m+1}^n \varphi(a_x, a_k) > 0 \end{aligned}$$

или

$$\sum_{i=1}^m \varphi(a_x, a_i) - \sum_{k=m+1}^n \varphi(a_x, a_k) > 0,$$

что неверно, так как a_x принадлежит к допустимому классу и для нее должно соблюдаться неравенство

$$\sum_{k=m+1}^n \varphi(a_x, a_k) - \sum_{i=1}^m \varphi(a_x, a_i) \geq 0.$$

Теорема 2. Если подмножество $\mathfrak{R}_r$ недопустимо, то любое подмножество $\mathfrak{R}_t$, содержащее $\mathfrak{R}_r$, также недопустимо.

Из того, что $\mathfrak{R}_r$ недопустимо, следует, что найдется такая буква a_x , для которой

$$\sum_{k=m+1}^n \varphi(a_x, a_k) - \sum_{i=1}^m \varphi(a_x, a_i) < 0,$$

где a_i, a_x входят в $\mathfrak{R}_r$, а a_k не входит.

Составим следующее выражение:

$$\sum_{r=l+1}^n \varphi(a_x, a_r) - \sum_{s=1}^l \varphi(a_x, a_s),$$

где $|\mathfrak{R}_t| = l$; a_x , конечно, входит и в $\mathfrak{R}_t$, $a_s \in \mathfrak{R}_t$, $a_r \in A \setminus \mathfrak{R}_t$. Легко видеть, что

$$\sum_{r=l+1}^n \varphi(a_x, a_r) = \sum_{k=m+1}^n \varphi(a_x, a_k) - \sum_{p=m+1}^l \varphi(a_x, a_p),$$

где $a_p \in \mathfrak{R}_l \setminus \mathfrak{R}_r$ и, следовательно,

$$\sum_{r=l+1}^n \varphi(a_x, a_r) \leq \sum_{k=m+1}^n \varphi(a_x, a_k);$$

В то же время

$$\sum_{s=1}^l \varphi(a_x, a_s) = \sum_{i=1}^m \varphi(a_x, a_i) + \sum_{p=m+1}^l \varphi(a_x, a_p),$$

следовательно, $\sum_{s=1}^l \varphi(a_x, a_s)$ больше или равно

$$\sum_{i=1}^m \varphi(a_x, a_i). \text{ Разумеется, и подавно,}$$

$$\begin{aligned} \sum_{r=l+1}^n \varphi(a_x, a_r) - \sum_{s=1}^l \varphi(a_x, a_s) &\leq \\ &\leq \sum_{k=m+1}^n \varphi(a_x, a_k) - \sum_{i=1}^m \varphi(a_x, a_i), \end{aligned}$$

а следовательно, и меньше нуля.

Доказанные теоремы позволяют построить алгоритм, который перебирает все допустимые классы. При этом теорема 2 позволяет сократить перебор. Затем будут отыскиваться все возможные пары, состоящие из допустимых классов, которые не пересекаются и исчерпывают алфавит. Каждая такая пара будет локальным минимумом в силу теоремы 1. Абсолютный минимум можно будет найти среди локальных непосредственным перебором. Изложим этот алгоритм по пунктам:

1. Составить все возможные подмножества алфавита, содержащие по t букв каждая.

2. Вычеркнуть из списка те подмножества, которые не являются допустимыми.

3. Составить список всех подмножеств, содержащих по $t+1$ букве и не содержащих недопустимых подмножеств, имеющих t букв.

4. Пуст ли этот список? Если да, то выполнять пункт 5. Если нет, то изменить t на $t+1$, выполнять пункт 2.

5. Найти все пары непересекающихся допустимых подмножеств, исчерпывающих алфавит, подсчитать для каждой пары значение K_1 , выбрать пару, для которой оно минимально: данная пара классов есть ответ.

6. Конец.

Начальное значение параметра t равно 2. Сокращение перебора, обеспечиваемое теоремой 2, осуществляет пункт 3: строятся не все возможные подмножества, содержащие по $t+1$ буквы, а только те, которые состоят из допустимых подмножеств объемом в t букв. Проверка на допустимость производится с помощью неравенства

$$\sum_{k=m+1}^n \varphi(a_x, a_k) - \sum_{i=1}^m \varphi(a_x, a_i) \geq 0,$$

которое должно выполняться для каждой буквы допустимого подмножества.

Алгоритм, переводящий слоговую письменность в буквенную. Если алгоритм выделения гласных и согласных сопоставлял элементу текста значение одного двончного признака, то алгоритм, излагаемый далее, сопоставляет каждому элементу значения двух признаков, причем эти признаки многозначны.

Разумеется, этому алгоритму далеко до такого, который сам бы определял число нужных признаков и их возможные наборы значений. Однако он имеет по крайней мере двоякий практический смысл. Во-первых, он может помочь установить произношение букв в случае так называемой слоговой письменности, во-вторых, аналогичный алгоритм был бы полезен для выделения морфем. Далее мы поясним практический смысл этого алгоритма подробнее.

Существует много образцов письменности, в которых отдельный значок изображает не один звук, а цепочку звуков. Если нужно дешифровать такой текст, то представляется желательным перевести его в запись обычного типа. Для обычной письменности проще установить и чтение, и грамматику.

Очень часто цепочка звуков, соответствующая слоговому значку, не представляет собой настоящего слога, а имеет следующую стандартную структуру: первый звук цепочки согласный, второй — гласный.

В частности, так обстоит дело в крито-микенской письменности.

Например,

ϕ -tu, 7-to, 7-te, 7-pu, 7-po,
 ϕ -ti, 7-ta, 7-va, 7-ve и т.д.

Если исследуемая письменность именно такого рода, то приводимый алгоритм должен переводить каждый значок указанного типа в пару символов, причем второй символ должен быть одинаковым для значков, содержащих один и тот же гласный звук, а первый для значков, содержащих один и тот же согласный.

Множество символов дало бы алфавит «обычного типа», который вводится вместо слогового.

О морфологической интерпретации алгоритма мы расскажем позднее.

Идея алгоритма состоит в следующем. Для алфавита S слоговых значков строятся две классификации, k_1 и k_2 . По первой из них в одном классе должны оказаться значки с общей гласной частью, например ta , pa , ka и т. д.; по другой — с общей согласной частью (ta , tu , to , ... и т. д.).

Обозначим классы из k_1 символами $\alpha_1, \alpha_2, \dots$, а классы из k_2 символами $\beta_1, \beta_2, \dots$. Если k_1 есть гласная классификация, то каждому слоговому значку мы сумеем сопоставить цепочку вида β_i, α_j . Обычный алфавит A , соответствующий слоговому, будет совокупностью символов $\alpha_1, \alpha_2, \dots, \alpha_m, \beta_1, \dots, \beta_k$. В алфавите A могут находиться и символы для «нулевых» гласных и согласных. «Нулевой гласный» (т. е. символ, обозначающий отсутствие гласного) нужен для записи слогового значка, изображающего отдельный

согласный, а нулевой согласный — для записи значка, соответствующего гласному.

Оказывается, k_1 и k_2 нельзя задавать произвольно. Искомая пара классификаций должна обладать некоторым свойством. Если данная пара обладает этим свойством, то мы будем называть ее допустимой.

Определив множество допустимых пар, мы будем отыскивать в нем наилучшую пару, т. е. такую, которая соответствует экстремальному значению некоторой оценочной функции.

Определим допустимую пару классификаций. Нетрудно заметить, что если в одной из классификаций m классов, то в каждом классе другой классификации не может быть больше, чем m элементов. Допустим, что в k_1 имеется всего два класса — класс значков, содержащих гласную a , и класс значков, содержащих гласную i . Рассмотрим в k класс значков, содержащих, например, согласную p . Он может содержать только значки pa и pi или один из этих значков (в силу того, что в исследуемом тексте могут не встретиться некоторые значки из S). Третьего значка, отличного от pa или от pi , в этом классе не может быть, так как нет третьего гласного.

Возможны, конечно, письменности, обладающие различными слоговыми значками, передающими одну и ту же последовательность звуков, но излагаемый алгоритм на такой случай не рассчитан.

Кроме того, предполагается, что больше шансов оказаться правильными у таких пар классификаций, для которых можно ожидать отсутствие в тексте меньшего числа слоговых значков.

Рассмотрим применяемый критерий качества допустимой пары. Пусть дана таблица $TI = \|f(s_i s_j)\|$, у которой в строке i и столбце j стоит количество упорядоченных цепочек вида $s_i s_j$, встреченных в исследуемом тексте. В строке i этой таблицы стоят, следовательно, числа, указывающие, какие значки из S и в каком количестве следуют за s_i . Пусть s_i декодируется как β_h, α_l . Естественно думать, что вид строки i больше зависит от α_l , чем от β_h , поэтому строки, соответствующие значкам, кончающимся на

α_i , должны быть похожи. Наоборот, среди столбцов должны быть похожими те, которые соответствуют цепочкам, начинающимся на одну и ту же букву.

Сходство строк и столбцов можно оценивать каким-нибудь расстоянием между точками n -мерного пространства ($n = |S|$).

Мы пользуемся, например, следующими простыми формулами:

$$\bar{\sigma}_1(s_i, s_j) = \sum_{k=1}^n |\varphi(s_i s_k) - \varphi(s_j s_k)| \text{ для строк и}$$

$$\bar{\sigma}_2(s_i, s_j) = \sum_{k=1}^n |\varphi(s_k s_i) - \varphi(s_k s_j)| \text{ для столбцов.}$$

Здесь $\bar{\sigma}_1$ и $\bar{\sigma}_2$ означают несходство (расстояние), а s_k пробегает весь алфавит слоговых значков.

Теперь мы можем построить две таблицы T_2 и T_3 расстояний по строкам и по столбцам таблицы T_1 .

Пусть в T_2 и T_3 ряды, соответствующие одному и тому же классу, соприкасаются, вследствие чего таблицы можно разбить на полосы, внутри которых имеются лишь ряды данного класса. Таблицы примут такой вид, как показано на рис. 61.

На рисунке заштрихованы квадраты, в которых размещаются расстояния от значков некоторого класса до значков того же класса. В силу нашего предположения эти расстояния должны быть невелики и сумма чисел во всех этих квадратах тоже не должна быть большой, если классификация правдоподобная.

Обозначим сумму чисел в квазидиагональных квадратах таблицы T_2 через Σ_1 , а таблицы T_3 — через Σ_2 . Чтобы построить оценочную функцию, мы воспользуемся тем очевидным соображением, что в хорошей паре классификаций даже самая плохая классификация должна быть достаточно хороша. Поэтому примем

$$K_4(k_1, k_2) = \max(\Sigma_1, \Sigma_2).$$

Здесь K_4 означает качество допустимой пары. К сожалению, функция K_4 имеет ясный смысл лишь

в том случае, когда количество элементов в каждом классе одной из классификаций в точности равно числу классов в другой. В принятом (более общем случае) приходится либо усложнять функцию, либо изменять определение допустимой пары. Мы пошли по второму пути. Поэтому определение допустимой пары окончательно будет звучать так:

Пара классификаций k_1, k_2 допустима, если:

1. При наличии в одной из классификаций m классов количество элементов в каждом классе другой не превышает m .

2. Сумма расстояний от элемента s_i до элементов того же класса меньше суммы расстояний от s_i до элементов любого другого класса.

3. При соблюдении условий 1 и 2 не существует пары классификаций k'_1, k'_2 , также удовлетворяющей условиям 1 и 2 и для которой

$$|k'_1| = |k_1| \quad \text{и} \quad |k'_2| < |k_2|.$$

Простейший алгоритм поиска наилучшей допустимой пары классификаций состоит в переборе всех возможных пар классификаций, отбрасывании недопустимых и выборе среди оставшихся той, для которой K_4 минимальна. Но, разумеется, такой алгоритм был бы практически абсолютно непригодным.

Поэтому мы предложим идею сокращенного способа получения наилучшего решения, который, как мы надеемся, практически равноценен полному перебору.

Мы будем строить две последовательности классификаций: $k_{11}, k_{12}, k_{13}, \dots$ и $k_{21}, k_{22}, \dots$. Одной из последовательностей будет соответствовать таблица T_2 , другой — T_3 .

При этом по одной из таблиц строятся все более дробные классификации, по другой — все более грубые. Числа Σ_{1i} будут уменьшаться, числа Σ_{2i} возрастать.

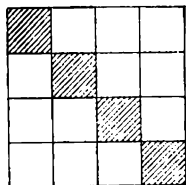


Рис. 61. Разбиение таблицы на полосы, соответствующие классам сложных значков.

Значение K_4 сначала будет определяться величиной Σ_{1i} , затем величиной Σ_{2i} , так, как изображено на рис. 62.

Жирной линией выделен график функции K_4 . Из рисунка видно, что минимум функции приходится на точку x , т. е. построение последовательности пар классификаций следует прекращать тогда, когда в следующей паре классификаций, построенной по правилам 1—4, Σ_{2i} станет больше, чем Σ_{1i} .

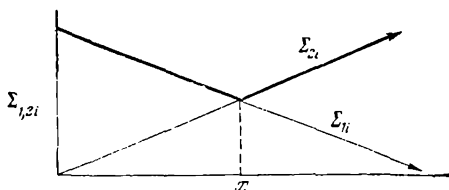


Рис. 62. Функции $\Sigma_{1,2}$ и K_4 .

Подробное изложение вычислительной процедуры мы опустим. Укажем только, что она широко использует алгоритм, изложенный на стр. 259.

Этот алгоритм может иметь интересную интерпретацию, если он будет обрабатывать не слоговые значки, а слова. Каждое слово можно считать входящим в класс некоторой «грамматической» классификации и в класс некоторой «лексической» классификации. Например, слово «домой» входит в тот же грамматический класс, что слово «столом», «водой», «решением», т. е. в творительный падеж единственного числа. С другой стороны, это слово входит в класс слов «дом», «дома», «домах» и т. д., т. е. в класс слов с основой «дом». Принадлежность к тому или другому грамматическому классу обычно подчеркивается наличием так называемого аффикса, например, «-ом», «-ем», «-ой» для творительного падежа единственного числа. Однако так бывает не всегда. Например, именительный падеж обычно не выражается никаким аффиксом. Будем говорить, что в таких случаях мы имеем дело с нулевым аффик-

сом. Основа тоже может отсутствовать, например, в слове «вынуть».

Вся эта ситуация полностью аналогична предположениям о структуре слоговых значков. Аналогичные предположения можно сделать и для построения оценочной функции.

Таблицу T_1 может заменить таблица условных вероятностей того, что если появилось слово λ_i , то, скажем, в пределах простого предложения появится и слово λ_j .

«Лексически» сходные слова должны сходным образом «управлять» другими словами. Так, например, личные формы одного и того же глагола должны требовать одних и тех же падежей. Грамматически сходные слова должны, наоборот, сходным образом «управляться», например, винительные падежи обусловлены наличием во фразах так называемых переходных глаголов. Все это полностью аналогично исходным предположениям, использовавшимся для построения оценочной функции приведенного алгоритма.

Такой способ анализа довольно интересен благодаря тому, что он позволяет обнаруживать «нулевые» морфемы. Далее мы приведем алгоритм выделения морфем, не использующий умение выделять слова. Однако этот алгоритм не обладает способностью выделять нулевые морфемы.

Можно представить себе следующую последовательность применения алгоритмов: вначале мы обнаруживаем «ненулевые» морфемы, затем слова как некоторые сочетания этих ненулевых морфем, затем слова вторично разлагаются на морфемы алгоритмом типа приведенного.

Алгоритм, классифицирующий слова «по смыслу». Исключительно важное значение алгоритма классификации имеют для дешифровки смысла текста.

Мы уже упоминали, что из внешнего различия облика слов совсем не следует, что существенно различен и смысл этих слов. Например, слова «иметь»

и «обладать» внешне совсем непохожи, а смысл их, несомненно, схож. Обычно трудно предсказать, какое слово появится: автомобиль или машина, но очевидно, подстановка одного из них вместо другого не будет грубой ошибкой.

Установление классификаций, объединяющих близкие по смыслу слова неизвестного текста, позволяет резко увеличить понятность этого текста. Поэтому подобная операция является необходимым этапом дешифровки. В частности, она может существенно помочь поиску иноязычно-русских словарей. Как мы увидим, наличие «смысловых» классификаций позволяет оценить смысловую близость слов и количественно.

Мы приведем способ, вырабатывающий систему классификаций, предложенный Ю. А. Шрейдером. В этом способе предполагается, что все классификации — бинарные или тернарные, т. е. делят все множество слов на два или на три класса. Если классификация бинарна, то один из ее классов содержит все слова, обладающие некоторым общим смысловым свойством, а второй — все остальные слова. Если классификация тернарная, то в ней еще имеется и класс слов, обладающий противоположным свойством. Так, один класс бинарной классификации могут составлять слова, связанные с понятием «пространства», другой класс — слова, не связанные с этим понятием. Понятие «одушевленность» может лежать в основе классификации на три класса: одушевленных, неодушевленных и таких слов, к которым такой признак неприменим (например, слово «постановление»).

Приводится, например, следующий список возможных признаков (соответственно, классификаций): 1) мышление, 2) элементарность (свойство быть единичным), 3) действие, 4) одушевленность, 5) положительность (хорошо — плохо), 6) великость (большой — малый), 7) пространство, 8) время, 9) упорядоченность, 10) свойство быть границей, 11) восприятие, 12) изменение, 13) свойство быть частью.

Обозначая наличие у слова данного свойства номером признака, а наличие противоположного свойства — номером с чертой, можно привести следующие примеры смысловых кодов слов:

алгоритм 3, 9, 12,
 вычислительный процесс 3, 9, 12,
 мгновение $\bar{6}$, 8, 13,
 дурак 1, 2, 4, $\bar{5}$,
 отдых $\bar{3}$, 5, 8, 13,
 французский артикль множ. числа 2.

Вместо этих «смысловых кодов» можно использовать векторы, в данном случае тринадцатимерные, у которых на i -м месте стоит 1, -1 или 0 в зависимости от того, обладает данное слово данным свойством, противоположным свойством или не обладает ни тем, ни другим; например:

дурак 1, 1, 0, 1, -1 , 0, 0, 0, 0, 0, 0, 0, 0.

Для принятой системы значений признаков смысловая близость оценивается по следующей формуле:

$$\rho_K(a, b) = \sum_{i=1}^n l(\epsilon_i, \eta_i),$$

где ϵ_i — i -я координата слова a , η_i — i -я координата слова b ,

$$l(\epsilon_i, \eta_i) = \left\{ \begin{array}{ll} 0, & \text{если } \epsilon_i = \eta_i, \\ 1, & \text{если } \epsilon_i = \pm 1, \eta_i = \mp 1, \\ 2, & \text{если } \epsilon_i = 0, \eta_i = \pm 1, \\ & \text{или } \epsilon_i = \pm 1, \eta_i = 0, \end{array} \right\} \quad (4.5)$$

Здесь расстояние выбирается таким образом, чтобы слова, различающиеся противоположными свойствами, были ближе, чем слова, различающиеся присутствием и отсутствием свойства (например, слова «великан» и «карлик» ближе, чем слова «великан» и «мыслитель»). Расстояние $\rho_K(a, b)$ не зависит от текста, а зависит от выбора смысловых категорий.

Далее предлагается некоторое расстояние, вычисляемое по тексту. С помощью этого расстояния оказывается возможным установление списка категорий (набора классификаций).

Назовем удалением $L(a, b)$ слов a и b во фразе количество слов от a до b включительно. (Как мы знаем, визуальное удаление не всегда соответствует степени смысловой связанности. Поэтому удаление измеряется не по тексту, а по так называемому графу предложения. Способы нахождения графов мы будем рассматривать в следующем параграфе.) Так, $L(a, a) = 0$, а для соседних слов $L(a, b) = 1$. Величина $\rho_T^1(a, b)$ определяется как среднее значение величин $L(a, b)$ для всех фраз, в которых одновременно встречается a и b . Если a и b в тексте T не встречаются в одной фразе, то считается, что $\rho_T^1(a, b) = \infty$.

Величина $\rho_T^1(a, b)$ характеризует близость между «сочетаемыми» словами, для которых смысловое родство проявляется в совместном появлении, но для «вытесняемых» слов, таких, как «дом» и «изба», величина $\rho_T^1(a, b)$ будет очень большой, вопреки их очевидной смысловой близости. Математически этот недостаток величины $\rho_T^1(a, b)$ выражается в том, что она не обладает одним из важных свойств расстояния: для нее не выполнено «неравенство треугольника» ($\rho(a, b) + \rho(b, c) \geq \rho(a, c)$). Поэтому вводится расстояние $\rho_T(a, b)$.

Пусть

$$\left. \begin{aligned} \rho_T^2(a, b) &= \min_c [(\rho_T^1(a, c) + \rho_T^1(c, b))], \\ \rho_T^k(a, b) &= \min_c [\rho_T^{k-1}(a, c) + \rho_T^{k-1}(c, b)]. \end{aligned} \right\} \quad (4.6)$$

Текстуальным расстоянием $\rho_T(a, b)$ называется

$$\rho_T(a, b) = \lim_{k \rightarrow \infty} \rho_T^k(a, b).$$

Эта величина удовлетворяет всем свойствам расстояния.

Если у нас уже имеется некоторое множество классификаций, то с помощью $\rho_T(a, b)$ путем обследования

некоторого текста исходное множество категорий можно улучшить. Для этого используется предположение о согласованности $\rho_T(a, b)$ и $\rho_K(a, b)$. Например, можно считать, что из малости $\rho_T(a, b)$ должна вытекать малость $\rho_K(a, b)$ и наоборот.

Схема действия алгоритма такова: задаются начальные смысловые векторы для некоторого списка слов из данного текста (Ю. А. Шрейдер считает, что смысловую кодировку следует давать только так называемым предикатным словам, но не предметным, в частности, не именам собственным, таким, как «Иван», «Москва», «БЭСМ»). При дешифровке все координаты начальных векторов должны быть нулевыми.

Затем для этих слов определяются расстояния $\rho_T(a, b)$ и $\rho_K(a, b)$ и эти расстояния сравниваются. Если критерий согласия выполняется, то на этом алгоритм завершает свою работу. Если нет, то изменяется система смысловых категорий.

Возможны два случая:

1. Для некоторого количества пар слов $\{a, b\}$ $\rho_T(a, b) < \rho_K(a, b)$. Тогда выбираются координаты, различные для многих пар слов из $\{a, b\}$ и исключаются из списка координат (векторов-столбцов). При этом понижается размерность смысловых векторов за счет отбрасывания векторов-столбцов, различающих близкие в данном тексте слова.

2. Для некоторых пар из $\{a, b\}$ $\rho_T(a, b) > \rho_K(a, b)$. В этом случае вводятся новые смысловые координаты с помощью следующего приема: отыскивается слово c такое, что

$$d = \rho_T(a, c) - \rho_T(c, b) = \max. \quad (4.7)$$

Этому слову c соответствует новая смысловая категория (координата), которая принимает значение 0 для слов d , удовлетворяющих неравенству

$$\rho_T(c, d) < \rho_T(c, b) + \frac{\alpha}{2}, \quad (4.8)$$

и значение 1 для всех остальных слов (поэтому для слова b эта координата равна нулю, а для a единице).

Иначе говоря, если слово *с* есть «стол», то введение новой категории соответствует различению слов по признаку наличия или отсутствия свойства «столовости».

Так получается новая система координат, отличная от прежней. Она проверяется с помощью критерия согласия и, в зависимости от его выполнения, осуществляется переход к новой системе координат или происходит остановка.

§ 10. Алгоритмы установления близости

Алгоритм, находящий граф синтаксической связи слов в предложении. Мы уже упоминали, что расстояния, наблюдаемые визуально между элементами сообщения, не соответствуют интуитивно ощущаемой «смысловой близости». Иначе говоря, понятность сообщения (т. е. возможность предсказывать его неосмотренные части на основании просмотренных), относительно малая в случае использования визуально наблюдаемых соотношений близости, может сильно возрасти, если перейти к оптимальным соотношениям.

Например, если сообщение есть линейная последовательность сигналов, то можно было бы считать максимально близкими сигналы, непосредственно следующие друг за другом; однако, если эта последовательность есть развертка двумерного изображения, то максимально близкими следовало бы считать и сигналы, разделенные интервалом в строку.

Поэтому если данный сигнал представляет темную деталь изображения, то можно ожидать появления сигнала, также представляющего темную точку не только рядом, но и через интервал.

Вполне аналогично, слова «обычного» языка, связанные по смыслу, не обязательно располагаются в тексте рядом. В приведенном предложении (рис. 63) слова, соединенные дугами, «связаны» теснее, чем не соединенные.

Пары слов, соединенные дугами, кажутся осмысленными, хотя иногда и незаконченными: «буреве-

стник реет», «реет с», «с криком», «буревестник подобный», «подобный молнии», «молнии черной».

Прочие пары являются неосмысленными: «буревестник черной» «с подобный» и т. д.

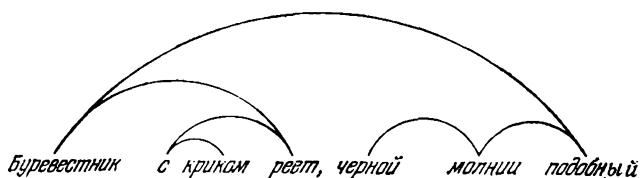


Рис. 63. Схема синтаксической связи слов в предложении «Буревестник с криком реет, черной молнии подобный».

В некоторых текстах (например, в латинских стихах) расположение далеко друг от друга связанных по смыслу слов производит даже впечатление злонамеренного запутывания (рис. 64).

Сохраняя построение, эту фразу можно передать словами: «В новые влечет душу измененные сказать

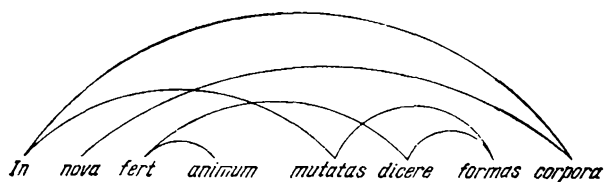


Рис. 64. Пример схемы синтаксической связи слов в латинской фразе.

формы тела», т. е. хочется рассказать о том, как тела превращаются в новые тела. Знание истинных соотношений близости необходимо при обнаружении единиц более высокого уровня, поскольку эти единицы должны состоять «из близких» единиц более низкого уровня.

Впрочем, при изложении основного варианта алгоритма, приводимого ниже, мы будем считать известными и единицы более низкого уровня — слова, и

единицы более высокого уровня — предложения, состоящие из слов.

Этот алгоритм должен устанавливать «истинную», смысловую близость слов в предложении, а точнее, в простом предложении, т. е. в предложении, не содержащем в себе других предложений.

Мы будем считать нашу задачу выполненной, если сможем указать пары слов простого предложения, непосредственно связанных по смыслу.

Пару слов, непосредственно связанных по смыслу, можно изображать отрезком, концы которого озаглавлены данными словами или какими-нибудь условными

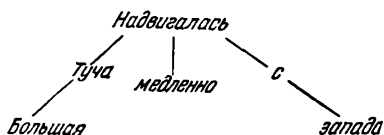


Рис. 65. Пары слов, соединенные отрезками прямых, осмыслены.

обозначениями этих слов. Тогда все множество слов предложения, непосредственно связанных по смыслу, изобразится некоторым рисунком, или, как говорят, графом, например, вида, изображенного на рис. 65.

Можно заранее предположить, какая форма графа характерна для простого предложения. Наиболее естественно думать (такое мнение сейчас преобладает), что граф простого предложения есть «дерево», иначе говоря, связный граф без циклов. Граф называется связным, если по составляющим его отрезкам (ребрам) можно пройти из произвольной вершины (узла) графа в любую другую; связный граф называется деревом, если это прохождение можно совершить единственным образом.

Высказанное выше предположение кажется разумным, потому что все слова в предложении представляются пусть не прямо, но все же связанными по смыслу. Обычно к тому же эти связи недвусмысленны: действительно, граф, изображенный на рис. 66, двусмыслен: неясно, следует ли эту фразу понимать как

«решение части собрания» или как «решение собрания части» (например, воинской части).

Основное предположение, на котором будет строиться оценочная функция, заключается в том, что непосредственно связанные в предложении слова сильно прогнозируют друг друга. Однако для того чтобы этот принцип был практичен, нам потребуется составить некоторую классификацию слов. Во-первых, слова вообще встречаются весьма редко, а их пары — еще реже. Введение классов позволяет не различать слова, принадлежащие к одному и тому же классу, и изображать любое слово символом «своего» класса. Этот символ в тексте, соответствующим образом перекодированном, будет уже встречаться часто. Во-вторых, алфавит слов весьма велик. Если его каким-нибудь образом не сжимать, то постановка машинных экспериментов по проверке алгоритмов будет практически невозможна.

Разумеется, классы следует вводить не какие попало, а «хорошие», с помощью специального дешифровочного алгоритма. Таких алгоритмов мы сейчас предложить не можем, но некоторое представление о том, как они должны выглядеть, у нас имеется.

Можно думать, что следует использовать так называемые «грамматические» классы слов. Для небольших подсчетов, производившихся на материале русского языка, мы использовали следующий набор классов — падежи существительного (порознь) и прилагательного, глагол в личной форме, наречие, деепричастие, инфинитив, предлог, союз, частица.

Теперь каждому слову можно поставить в соответствие символ класса, куда оно входит, и закодировать текст этими символами.

Мы будем изучать, как слова (вернее, классы слов) прогнозируют друг друга в предложении. Для этого вычислим условные вероятности того, что если в простом предложении появилось слово класса k_i , то в

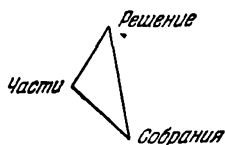


Рис. 66. Пример графа с неопределенным смыслом.

нем окажется и слово класса k_j . Эту величину можно вычислить по формуле

$$p(k_i/k_j) = \frac{p(k_i, k_j)}{p(k_i)} \approx \frac{P(k_i, k_j)}{P(k_i)} = \frac{\varphi(k_i, k_j)}{\varphi(k_i)}.$$

Здесь $p(k_i, k_j)$ означают вероятности совместного появления символов класса k_i и k_j в пределах простого предложения, $p(k_i)$ — вероятность появления в простом предложении символа k_i . Буквы P и φ означают соответственно относительную и абсолютную частоту.

Для оценки «взаимной прогнозируемости» символов k_i и k_j мы возьмем среднюю из $p(k_i/k_j)$ и $p(k_j/k_i)$, обозначаемую $\bar{p}(k_i, k_j)$.

Частичной предсказуемой системой для данного текста мы назовем квадратную таблицу чисел $\bar{p}(k_i, k_j)$.

Если у нас есть некоторое дерево предложения, то каждому его ребру, соединяющему слова λ_u и λ_v , мы можем сопоставить вес — число $\bar{p}(k_i(\lambda_u), k_j(\lambda_v))$, зависящее от классов k_i и k_j слов λ_u и λ_v , стоящих на концах этого ребра.

Так как предполагается, что слова, непосредственно связанные в графе, сильно прогнозируют друг друга, то мы можем в качестве функции, оценивающей граф, предложить сумму чисел $\bar{p}(k_i(\lambda_u), k_j(\lambda_v))$ для всех ребер данного графа.

Итак, множеством допустимых решений у нас является множество всех возможных деревьев, построенных из символов слов данного простого предложения, а оценочной функцией величина

$$D = \sum_u \sum_v \bar{p}(k_i(\lambda_u), k_j(\lambda_v)),$$

где λ_u, λ_v есть символы, стянутые ребром данного дерева. Для наилучшего решения функция D должна принимать максимальное значение.

Аналогичная задача решалась в математической экономике для нахождения кратчайшей телефонной сети, связывающей города. Существуют две разновид-

ности алгоритма, находящего экстремум функции D . Мы приведем более простой способ, изложенный в [6]. Этот способ, как большинство алгоритмов так называемого дискретного анализа, нестрог, но практически вполне приемлем. Конечно, в принципе возможен и абсолютно строгий алгоритм, построенный путем перебора всех возможных деревьев, подсчета значения D и выбора дерева, максимизирующего D . Однако для сколько-нибудь больших предложений этот способ практически непригоден из-за своей громоздкости.

Множество символов слов данного предложения делится на осмотренные и неосмотренные. Первоначально множество осмотренных символов пусто.

На первом шаге выбираются две вершины, λ_1 и λ_2 , такие, что

$$\bar{p}(k(\lambda_1), k(\lambda_2)) \geq \bar{p}(k(\lambda_i), k(\lambda_j)),$$

где λ_i и λ_j — слова данного предложения, не совпадающие соответственно с λ_1 и λ_2 соответственно. Вершины λ_1 и λ_2 объявляются осмотренными и соединяются ребром.

Если имеются и осмотренные и неосмотренные вершины, то отыскивается неосмотренная вершина λ_n , такая, что

$$\bar{p}(k(\lambda_u), k(\lambda_n)) \geq \bar{p}(k(\lambda_v), k(\lambda_w)),$$

где λ_u — некоторая осмотренная, λ_w — любая неосмотренная вершина, отличная от λ_n , а λ_v — произвольная осмотренная вершина. Вершины соединяются ребром, а λ_n вносится в список осмотренных вершин. Если неосмотренных вершин нет, то процесс заканчивается.

Этот алгоритм имеет ряд более интересных вариантов. Так, мы можем указать нечто вроде порядка наилучшего считывания предложения, пути осмотра, обеспечивающего наискорейшее понимание. При этом в первую очередь будут «схватываться» главные мысли», точнее, главные слова. Другими словами мы можем указать для слов отношение подчинения.

С этой целью достаточно проставить направления на ребрах графа непосредственной связи слов. Эти

направления можно проставить из следующих соображений: подчиненное слово сильнее прогнозирует подчиняющее, «сильнее в нем нуждается», чем наоборот; поэтому если

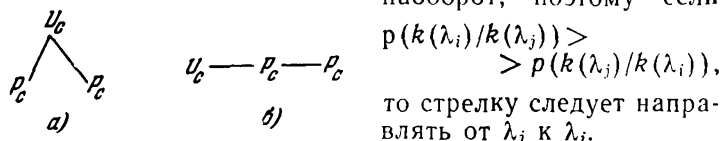


Рис. 67. Неправильный (а) и правильный (б) граф фразы «строение атома водорода».

$$p(k(\lambda_i)/k(\lambda_j)) > p(k(\lambda_j)/k(\lambda_i)),$$

то стрелку следует направлять от λ_j к λ_i .

Если бы считывание шло в обратном направлении, то мы дольше ощущали бы незавершен-

ность, непонятность фразы, так как уже осмотренные подчиненные слова сильно прогнозируют своих еще не появившихся хозяев.

Это ощущение слишком хорошо знакомо каждому, изучавшему немецкий или латинский язык!

Алгоритм, обнаруживающий «типы синтаксической связи» слов. Алгоритм, излагавшийся выше, проверялся вручную лишь на небольших примерах. Однако до проведения серьезных экспериментов можно предвидеть некоторые из ошибок, которые может совершить этот алгоритм.

По-видимому, большинство ошибок будет связано с тем, что слова в принятой записи описываются весьма неточно: все слова одного и того же грамматического класса не различаются между собой. Например, фраза «строение атома водорода» получит кодировку «и_с, р_с, р_с» *). Если $\bar{p}(и_с, р_с) > \bar{p}(р_с, р_с)$ (как этого следует ожидать), то алгоритм построит для этой фразы дерево (а) (рис. 67), тогда как правильный граф этой фразы есть граф (б).

Нетрудно заметить, что иногда связанность слов вообще не зависит от их грамматической формы; например, для куска некоторой фразы: «строением атома водорода» граф очевидным образом не зависит от падежа первого слова и совпадает с графом фразы «строение атома водорода».

*) То есть именительный существительного, родительный существительного, родительный существительного.

Вариант алгоритма, предлагаемый ниже, служит для устранения этих ошибок; кроме того, он представляет и самостоятельный интерес.

Предположим, что слова описываются с помощью двух классификаций, а не одной, как это было ранее. Одну из них назовем грамматической и будем отождествлять ее с классификацией для уже изложенного варианта алгоритма. Другую классификацию назовем лексической. Будем считать, что классы этой классификации содержат слова с общей основой. Теперь каждое слово будет описываться двумя символами: символом грамматического класса, к которому принадлежит данное слово, и символом его лексического класса.

«Частичную предсказуемую систему» для такого описания словаря можно представить в виде таблицы

	Γ_1	$\dots$	Γ_m	Λ_1	$\dots$	Λ_k
Γ_1						
$\vdots$						
$\vdots$						
Γ_m						
Λ_1						
$\vdots$						
$\vdots$						
Λ_k						

в которой ряд, обозначенный символом Γ_i , содержит числа, характеризующие грамматический класс Γ_i , а ряд, озаглавленный символом Λ_j , — числа, характеризующие лексический класс Λ_j . Число в клетке i, j этой таблицы есть условная вероятность того, что если в предложении появилось слово класса i , то в нем появится и слово класса j . Числа в клетках находятся в результате обследования некоторого текста.

После того как вычислена эта таблица, для двух произвольных слов λ_i и λ_j мы сможем вычислить так называемую связующую функцию σ_1 :

$$\sigma_1(\lambda_i, \lambda_j) = p(\Gamma(\lambda_j)/\Gamma(\lambda_i)) + p(\Gamma(\lambda_j)/\Lambda(\lambda_i)) + \\ + p(\Lambda(\lambda_j)/\Gamma(\lambda_i)) + p(\Lambda(\lambda_j)/\Lambda(\lambda_i)).$$

Смысл этой функции тот же, что и условной вероятности, но ее введение позволяет использовать преимущества описания слов с помощью классификаций.

Мы уже обращали внимание на то, что лексические грамматические классы сходны с классификациями слоговых значков на гласные и согласные классы. Поэтому данную пару классификаций в принципе можно было бы обнаруживать алгоритмом, сходным с упомянутым на стр. 318.

Как и в первом варианте алгоритма, таблицу условных вероятностей мы заменим таблицей взаимных прогнозируемостей, т. е. чисел вида $\bar{p}(k_i, k_j) = \frac{1}{2} p(k_i/k_j) + p(k_j/k_i)$; связующая функция σ_2 , вычисленная с помощью таблицы $\|\bar{p}(k_i, k_j)\|$, примет следующий вид:

$$\sigma_2(\lambda_i, \lambda_j) = \bar{p}(\gamma(\lambda_i), \gamma(\lambda_j)) + \bar{p}(\gamma(\lambda_i), \lambda(\lambda_j)) + \\ + \bar{p}(\lambda(\lambda_i), \gamma(\lambda_j)) + \bar{p}(\lambda(\lambda_i), \lambda(\lambda_j)).$$

Теперь, если дано предложение, каждое слово которого закодировано парой символов указанного типа, то мы сможем построить дерево этого предложения, сопоставляя ребрам числа $\sigma_2((\lambda_i, \lambda_j))$ и применяя ранее изложенную процедуру.

Ошибки, подобные той, которая предполагалась во фразе «строение атома водорода», могут быть исправлены, потому что для слов «атом водорода» связующая функция будет учитывать их частое появление в одной фразе, безразлично в каких падежах (это выразится большой величиной слагаемого $\bar{p}(\gamma(\lambda_j)/\lambda(\lambda_i))$). Направления ребер в этом графе могут быть приписаны из тех же соображений, что и ранее.

Данный алгоритм позволяет получить и новый вид информации о предложениях. Речь идет о том, что мы можем определить четыре типа связи в зависимости от того, какое из слагаемых вносит наибольший вклад в значение связующей функции $\sigma_2(p(\gamma(\lambda_j)/\gamma(\lambda_i)), p(\gamma(\lambda_j)/\lambda(\lambda_i)), p(\lambda(\lambda_j)/\lambda(\lambda_i))$ или $p(\lambda(\lambda_j)/\gamma(\lambda_i))$. Их ин-

терпретация легко воспринимается, если мы вспомним типы «подчинительной связи», изучавшиеся в средней школе. Знание типов связи увеличивает понятность текста, поскольку дает возможность конкретнее прогнозировать появляющиеся слова. Рассмотрим эти типы.

Тип $г \rightarrow г$, как явствует из его названия, есть вид связи, при котором грамматический класс слова λ_i сильно предсказывает грамматический класс слова λ_j , т. е. то, что обычно называют «согласованием» (примеры: «красного дома», «Андроны едут»).

В типе $л \rightarrow г$ лексический класс слова λ_i сильно прогнозирует грамматический класс слова λ_j . В традиционной грамматике этот тип носил название «управления» (примеры: «вижу корову», «направляюсь тебе», «восхищаюсь картиной»).

Тип $л \rightarrow л$ (лексический класс сильно прогнозирует лексический же класс), традиционное название — примыкание. Согласно традиционному учению этот тип характерен для связи неизменяемых слов, например, «очень давно» (не совсем ясно до какой степени предложенное определение согласуется с обычным; это, впрочем относится и к другим типам).

Тип $г \rightarrow л$. В традиционной грамматике не имеет аналогий. Не исключено, что его в действительности почти не бывает. Очевидные примеры неизвестны.

Простейший алгоритм пословного машинного перевода. Далее мы сделаем последнее обобщение алгоритма поиска графа предложения, имеющее, пожалуй, наибольший интерес, но зато и требующее наибольшего количества предварительных сведений о тексте.

Мы покажем, что этот вариант алгоритма может быть полезен (хотя бы в какой-нибудь модификации) для машинного перевода. При этом имеется в виду самый простой вид перевода, так называемый пословный, когда каждому знаменательному слову оригинала сопоставляется некоторое слово перевода.

Трудность такого перевода заключается в том, что одно и то же слово оригинала может переводиться

весьма различным образом. Если у нас имеется словарь, сопоставляющий словам языка оригинала все возможные переводы, то задача алгоритма перевода в основном сводится к тому, чтобы выбросить лишние варианты, оставив самый удачный. Алгоритм будет тем интересней, чем большее количество вариантов отбрасывается для получения наилучшего.

Очень хорошо, например, было бы, чтобы английскому слову *hand* словарь сопоставил не только русские «кисть» (руки), «стрелка» (часов), но и кисти, кистью и т. д. В этом случае выбор наилучшего варианта означал бы и постановку переводящего слова в должной грамматической форме. Для работы требуется система смысловых классификаций, возможно, сходных с теми, которые описывались на стр. 320, а также некоторых «грамматических классификаций». Мы должны уметь описывать слова языка перевода с помощью этой системы классификаций, т. е. сопоставлять каждому слову перевода его вектор — описание (см. стр. 321). Список символов смысловых классов представит некоторый смысловой («семантический») алфавит.

Теперь для языка перевода мы сможем построить квадратную таблицу, ряды которой озаглавлены символами классов, а в клетках стоят условные вероятности $p(k_j/k_i)$ (эта таблица подсчитывается на материале большого текста на языке перевода). Связующие функции можно ввести уже знакомым нам способом:

$$\sigma_3(\lambda_i, \lambda_j) = \sum_{u=1}^n \sum_{v=1}^n p(k_u(\lambda_j)/k_v(\lambda_i)),$$

где n — общее число классов;

$$\sigma_4(\lambda_i, \lambda_j) = \sum_{u=1}^n \sum_{v=1}^n \bar{p}(k_u(\lambda_i), k_v(\lambda_j)).$$

Функция σ_4 будет принимать большие значения для тех пар слов, которые сильно прогнозируют друг друга в текстах на языке перевода.

Чтобы представить остальные этапы алгоритма, изобразим схематически некоторую фразу на языке оригинала и варианты переводов, выписанные с помощью двуязычного словаря:

фраза на языке оригинала

	«1-е слово	2-е слово	3-е слово»
варианты	+		+
	+		
	+	+	+
	+		
	+		

Из каждой колонки следует выбрать по одному слову. Если каждое слово каждой колонки мы соединим линией с каждым словом из других колонок, то получим граф, содержащий все возможные графы всех возможных предложений (рис. 68).

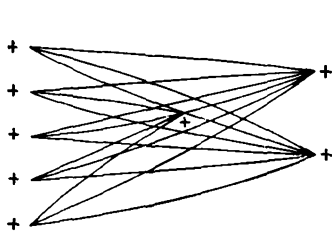


Рис. 68. Граф, содержащий все возможные графы предложений-переводов.

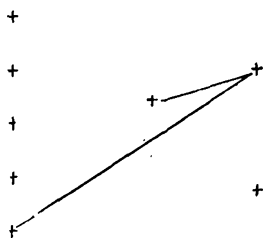


Рис. 69. Возможный вид графа, получаемого в конце работы алгоритма.

Однако из этого графа следует выбрать всего одно дерево, имеющее по одной вершине в каждой колонке. Нахождение такого графа можно будет провести обычным способом (с несущественными изменениями); отличие будет в том, что мы не будем искать граф, связывающий все наличные вершины: в каждой колонке все вершины, кроме одной, должны оказаться изолированными, как, например, на рис. 69. Изолированные вершины и будут представлять собой отброшенные варианты.

§ 11. Алгоритмы склеивания (окончание)

Алгоритм, выделяющий морфемы. Как мы уже упоминали, морфемами являются мельчайшие осмысленные единицы человеческого языка, т. е. множество морфем служит алфавитом элементарных сигналов для семантических уровней.

Например, слово «усатый» делится на морфемы так «ус+ат+ый». Некоторые цепочки морфем образуют слова. Слова во многих орфографиях, но далеко не всегда, отделяются друг от друга особыми знаками — пробелами. Морфемы, как правило, друг от друга не отделяются. Поэтому нередко системы автоматического анализа текста, в которых знание границ слов предполагается известным, а границы между морфемами отыскиваются специальным алгоритмом.

Однако мы не можем ожидать, что космическое сообщение будет строиться по принципу земных систем письменности. Поэтому следует ориентироваться на самый трудный случай — когда неизвестны ни границы слов, ни границы морфем, и начинать надо с выделения кодовых цепочек. Основное отличие заключается в том, что цепочки, образующие морфему, неравной длины. Поэтому множество допустимых решений является множеством всех возможных разбиений текста, которых, как нетрудно вычислить, $(N - 1)!$ Имеется еще одно существенное отличие: число морфем в человеческих языках значительно превышает число букв, изображающих звуки. Это отличие не принципиальное, но в случае декодирования неравномерного кода, изображающего буквы, вероятно, можно было бы использовать более простую процедуру.

Алгоритм, излагаемый ниже, использует оценочную функцию, отличную от функции V , а также другую опознавательную процедуру. Эти отличия позволяют находить и более интересные цепочки букв, чем морфемы. Цепочки морфем также могут представлять интерес, если они являются словами или группами слов. Предлагаемый далее алгоритм должен находить наряду с морфемами и слова, по крайней мере некоторые, и отдельные группы слов.

Правда, не исключено, что некоторые цепочки, являющиеся сочетаниями морфем, при этом не будут разбиты на морфемы; для выполнения этой работы потребуется специальный алгоритм, который мы здесь не приводим.

В связи с изменением цели алгоритма множество допустимых решений будет описываться иначе. Назовем правильным разбиением текста круглыми скобками исходный текст с вставленными в него символами двух видов — правыми и левыми скобками, при этом правые и левые скобки размещаются в соответствии со следующими определениями:

1) назовем правильной цепочкой нулевого порядка цепочку букв исходного текста, ограниченную слева левой скобкой, справа — правой;

2) назовем правильной цепочкой i -го порядка цепочку из правильных цепочек $i-1$ -го порядка.

Будем говорить, что у нас имеется правильное разбиение текста круглыми скобками, если скобочные символы, вставленные в исходный текст, превращают его в правильную цепочку.

Из этого определения следует, что правильные цепочки одного порядка не могут «зацепляться», т. е. начало одной правильной цепочки не может совпадать с концом другой, если одна из них не входит в другую целиком. Так, в тексте «еговинадоказана» не могут быть одновременно правильными цепочки «вина» и «надо». Это условие соответствует представлению о том, что морфемы отделены одна от другой, а не переплетаются, как слова в чайнвордах.

Простые цепочки соответствуют морфемам; однако мы не учитываем, что возможны случаи, когда одна морфема вставлена в другую (англ. man — человек, men — люди, русск. «вёз» и «воз»), а также изменений, происходящих на границах морфем (нога, ножка). Однако учет этих явлений невозможен без существенного усложнения алгоритма.

Оценочную функцию мы будем строить из следующих соображений. Рассматривая предыдущий алгоритм, мы уже упоминали, что в случае равенства длин

кодовых цепочек у сравниваемых решений разница между цепочками правильного и неправильного решения заключается в том, что части правильных цепочек «связаны» между собой теснее, чем части неправильных цепочек. Именно это представление и лежит в основе предлагаемой оценочной функции.

Мы будем считать, что цепочка состоит из «сильно связанных» частей, если появление некоторой части цепочки сильно прогнозирует появление остальной части. Например, если в каком-либо русском тексте появилась цепочка «челове», то вслед за ней почти обязательно появится буква «к». Подобное утверждение, с разной степенью основательности, можно сделать и для других разбиений цепочки «человек» на две части. Степень прогнозирования естественно вычислять по формуле условной вероятности: если дано i -е разбиение слова λ на правую и левую части, то само слово можно записать так: $\lambda = \Pi_i \Lambda_i$ (где Λ_i — левая часть, Π_i — правая часть при i -м разбиении); степень прогнозирования левой частью правой можно определить дробью $\frac{\varphi(\Lambda_i \Pi_i)}{\varphi(\Lambda_i)}$; то, как правая часть прогнозирует левую, определяется дробью $\frac{\varphi(\Lambda_i \Pi_i)}{\varphi(\Pi_i)}$.

В целом качество цепочки определяется следующим образом:

1. Цепочка разбивается на две части всеми возможными способами.

2. Для каждого разбиения определяется степень прогнозирования одной части другой.

3. Вычисляется средняя из всех прогнозируемостей, соответствующих проделанным разбиениям.

Так как в цепочке длины d можно осуществить $d-1$ разбиений, то общее число прогнозируемостей (дробей вида $\frac{\varphi(\Lambda_i \Pi_i)}{\varphi(\Lambda_i)}$ или $\frac{\varphi(\Lambda_i \Pi_i)}{\varphi(\Pi_i)}$) равно $d-1$. Поэтому качество цепочки, «устойчивость» определяется следующей формулой:

$$Y_\lambda = \frac{1}{2(d-1)} \sum_{i=1}^{d-1} \left(\frac{\varphi(\Lambda_i \Pi_i)}{\varphi(\Lambda_i)} + \frac{\varphi(\Lambda_i \Pi_i)}{\varphi(\Pi_i)} \right).$$

Разумеется, эта формула неприменима к оценке цепочек, состоящих из одной буквы: устойчивость таких цепочек принимается равной нулю. Нулю приравнивается также устойчивость цепочек, не встретившихся в тексте ни разу или встретившихся всего один раз; последнее условие объясняется тем, что любая, в том числе и неправильная, цепочка текста встречается не реже одного раза.

Далее, качество разбиения $Y(R)$ определяется как сумма устойчивостей всех правильных цепочек, входящих в это разбиение.

Разумеется, мы не смогли бы перебрать все возможные правильные разбиения текста круглыми скобками и вычислить для каждого значение функции $Y(R)$. Мы, как обычно, предложим процедуру, которая, вероятно, позволяет получить довольно высокое значение $Y(R)$.

Для работы нам понадобится частотный словарь всех встретившихся в исследуемом тексте цепочек, т. е. список этих цепочек с указанием того, сколько раз встретилась данная цепочка в тексте.

Самый простой способ составления такого словаря сводится к следующему: вначале создаются списки возможных цепочек длины 1, 2, ... и т. д., а затем осматривается текст и подсчитывается количество вхождений в текст данной цепочки. Этот способ, однако, не годится для сколько-нибудь значительных длин цепочек.

Построение частотного словаря существенным образом облегчается ввиду того, что нам неинтересны цепочки, встретившиеся реже двух раз (их устойчивости равны нулю). Поэтому можно предложить следующий способ: составляется список букв и подсчитываются их частоты; буквы, встретившиеся реже двух раз, выбрасываются из списка; частоты оставшихся букв вносятся в список. Полученный фрагмент частотного словаря назовем фрагментом первого порядка.

Если построен фрагмент $i-1$ -го порядка, то фрагмент i -го порядка строится так: для каждой цепочки λ

из фрагмента i —1-го порядка строятся все различные цепочки вида λa_x , где a_x — некоторая буква алфавита: затем определяются частоты этих цепочек и выбрасываются цепочки, встретившиеся реже двух раз. Список оставшихся цепочек и их частоты и есть фрагмент i -го порядка.

Процесс построения фрагментов оборвется тогда, когда вновь построенный фрагмент окажется пустым. Этот момент наступит тогда, когда любая цепочка, более длинная, чем какая-нибудь из построенных, встречается реже двух раз.

Полученный частотный словарь можно сильно сократить, выбросив все цепочки, содержащиеся в некоторой более длинной цепочке с той же частотой. Если некоторая цепочка λ_u встречается k раз, то любая цепочка λ_v , содержащаяся в λ_u , встречается не менее k раз, поэтому если частоты λ_u и λ_v равны, излишне вписывать цепочку λ_v в словарь.

Имея сведения о частотах цепочек, мы можем вычислить их устойчивости. Вычисленные устойчивости также впишем в частотный словарь наряду с частотами.

Дальнейшими действиями мы преобразуем частотный словарь цепочек в частотный словарь правильных цепочек.

Предположение, обеспечивающее начало процедуры, состоит в следующем: мы полагаем, что самая устойчивая цепочка λ_1 всегда входит в текст правильно, т. е. не зацепляется ни с какой другой правильной цепочкой.

Чтобы пояснить это утверждение, заметим, что некоторые цепочки в одном месте текста образуют морфемы или сочетания морфем, а в другом нет. Например, в фразе «еговинадоказана» цепочка «надо» морфемы не образует, а в фразе «надоучиться» — является морфемой.

В силу этого предположения мы должны отыскивать в частотном словаре самую устойчивую цепочку λ_1 , а затем все ее вхождения в текст окружить круглыми скобками.

По той же причине ни одна цепочка, зацепленная с каким-нибудь вхождением цепочки λ_1 , не может быть правильной. Поэтому мы должны уменьшить частоты тех цепочек, находящихся в частотном словаре, вхождения которых зацеплены с вхождениями λ_1 , и притом на столько раз, сколько раз вхождения данных цепочек зацепляются с вхождениями λ_1 .

Цепочки, частоты которых нужно уменьшать, можно найти, рассматривая вхождения λ_1 по следующей схеме:

$$a_i (a_{i+1} \dots a_{i+l-2} a_{i+l-1} [\dots [[a_{i+l}] a_{i+l+1}] a_{i+l+2}] \dots) \dots$$

Здесь цепочка $a_{i+1} \dots a_{i+l}$ изображает вхождение правильной цепочки в текст, i — номер от начала текста.

В частотном словаре отыскиваются цепочки, заключенные в квадратные скобки в порядке возрастания их номеров; $k+1$ есть номер первой цепочки, которой уже нет в частотном словаре.

Частоты найденных цепочек уменьшаются на единицу; затем серия квадратных скобок строится вправо от промежутка между a_{i+l-1} и a_{i+l-2} , затем — от промежутка между a_{i+l-2} и a_{i+l-3} и т. д., до промежутка между a_{i+1} и a_{i+2} . Потом аналогичные серии скобок строятся влево от цепочки $a_{i+1} \dots a_{i+l}$.

Так осматриваются все вхождения цепочки λ_1 . В результате описанной процедуры частотный словарь становится больше похожим на список правильных цепочек, поскольку в нем уменьшается количество цепочек, зацепляющихся друг с другом в тексте; если в результате уменьшения частот частота некоторой цепочки станет меньше двух, то данная цепочка выбрасывается из словаря.

Вообще, частотный словарь можно рассматривать как приближение к списку правильных цепочек, устойчивость которых больше нуля. Можно оценить качество этого списка по формуле

$$Y' = \sum_i Y(\lambda_i) \varphi_{\text{пр}}(\lambda_i).$$

Здесь λ_i — некоторая цепочка, $\varphi_{\text{пр}}(\lambda_i)$ — количество правильных вхождений q_i в текст.

Если бы частотный словарь был действительно списком правильных цепочек с ненулевыми устойчивостями, то значение Y' было бы просто значением оценочной функции Y , оценивающей разбиение, состоящее из правильных цепочек, представленных списком, и из некоторого количества цепочек с нулевой устойчивостью.

Однако частотный словарь содержит и неправильные цепочки. Чтобы превратить его в список правильных цепочек, мы должны выбросить все неправильные.

Это выбрасывание мы можем произвести различными способами. Наши действия направлены на то, чтобы в процессе выбрасывания минимально уменьшалось число Y' .

После того как мы произвели уменьшение частот цепочек, зацепленных с λ_1 , следует пересчитать устойчивости цепочек, содержащихся в словаре, так как нет смысла при вычислении устойчивостей учитывать частоты неправильных цепочек.

Цепочка λ_1 переходит в список так называемых выбранных цепочек.

В оставшемся списке снова выбирается максимально устойчивая цепочка, например, λ_i . Относительно нее делается предположение, что она не зацепляется ни с одной правильной цепочкой, если эта цепочка не является выбранной. Поэтому все невыбранные цепочки, зацепленные с λ_i , не являются правильными, и их частоты должны быть уменьшены. Если λ_i зацеплена с некоторыми выбранными цепочками, то эти вхождения λ_i были выброшены на предыдущих этапах работы алгоритма.

Затем должным образом уменьшаются частоты и пересчитываются устойчивости. Остальные действия аналогичны; процесс заканчивается, когда все цепочки частотного словаря окажутся либо выброшенными, либо выбранными. Одновременно в тексте появляются скобки, окружающие правильные цепочки.

Правда, полученная разметка текста не будет правильным разбиением круглыми скобками, хотя бы потому, что цепочки, имеющие нулевую устойчивость, не выделялись в течение процедуры.

Эту неправильность можно устранить, добавив некоторое количество круглых скобок. Так как от этой процедуры значение Y' не меняется, то при добавлении скобок мы будем следить лишь за тем, чтобы количество добавленных скобок было минимальным.

Впрочем, вряд ли представляет интерес добавление скобок более, чем нулевого порядка; скобки же нулевого порядка добавляются по следующему правилу: всякий участок вида $\dots(\dots(\dots)\dots)$ или $\dots)(\dots(\dots)$ заменяется на $((\dots)(\dots)(\dots))$ или $(\dots)(\dots)(\dots)$ соответственно.

Окончательно список правильных сочетаний получается добавлением в измененный частотный словарь цепочек с нулевой устойчивостью, окруженных скобками на заключительном этапе.

Изложенная процедура основывалась на том, что некоторые вхождения цепочки λ_i считались правильными, и в соответствии с этим цепочки, зацепленные с λ_i , неправильными. Большее значение функции Y' обеспечивала бы процедура, при которой вместо цепочки λ_i рассматривалась бы пара граничащих цепочек $\lambda_i\lambda_k$, а выбрасывались цепочки, зацепленные хотя бы с одной из цепочек λ_i или λ_k ; еще более точный результат получался бы при рассмотрении троек, четверок, ..., n -ок, поскольку в этом случае процедура приближалась бы к прямому перебору разбиений текста круглыми скобками.

Алгоритм, выделяющий буквы. На стр. 248 мы говорили о том, что лингвистические явления целесообразно характеризовать набором существенных признаков, но как именно выбрать такое множество признаков в настоящее время, неизвестно. Можно представить себе такую ситуацию: существует некоторое множество «элементарных признаков». Затем по некоторым правилам из этих элементарных признаков

наилучшим образом получают сложные признаки. Для оценки качества построенных признаков должен существовать некий критерий (например, оценочная функция).

В этой программе действий, пожалуй, самым простым является вопрос о выборе множества элементарных признаков. Можно даже думать, что в значительной степени этот выбор несуществен.

Рассмотрим, например, вопрос об элементарных признаках для сообщений, которые воспринимаются зрительно. Сюда относятся письменные тексты, рисунки и вообще весь видимый мир. Очевидно, что элементарные признаки должны составлять нечто вроде простейших актов восприятия и быть связанными с минимальными различиями, замечаемыми глазом (воспринимающим инструментом).

Если считать, что любой цвет образуется путем комбинации трех основных цветов — красного, синего и желтого, то видимый мир можно считать пространством, каждая точка которого описывается шестью координатами: $\alpha_k \cdot \epsilon_k$, $\alpha_c \cdot \epsilon_c$, $\alpha_{ж} \cdot \epsilon_{ж}$, $\alpha_d \cdot \epsilon_d$, $\alpha_{ш} \cdot \epsilon_{ш}$, $\alpha_v \cdot \epsilon_v$, где ϵ_k , ϵ_c , $\epsilon_{ж}$ суть минимальные приращения интенсивности трех основных цветов, $\epsilon_{ш}$, ϵ_d , ϵ_v — минимальные воспринимаемые изменения ширины, длины и высоты, выраженные, например, в угловой мере. Символы, содержащие букву ϵ , суть единичные векторы или «элементарные признаки», символы, содержащие букву α , означают степень признака у данного объекта.

Для передачи явлений слышимого мира следовало бы ввести другие элементарные признаки, связанные с минимальными акустическими различиями явлений, но это не обязательно: ведь возможно перекодирование звуков в запись на бумаге (например, осциллограмма).

Вообще все многообразие мира мы воспринимаем посредством так называемых «чувств». Перекодирование непосредственно невоспринимаемых сигналов (если речь идет об «элементарных» сигналах) осуществляется с помощью «физических приборов».

В то же время переход от микрособытий (точек в пространстве элементарных признаков) к более сложным единицам, например, к буквам в обычном смысле, представляет собой в общем случае весьма сложную задачу.

Далее мы рассмотрим, в весьма приблизительном виде, задачу выделения букв (типа букв русского алфавита) при условии, что мы обладаем прибором, умеющим отличать черный квадратик от белого (на листе бумаги, покрытом мелкой сеткой) и расположение этих квадратиков.

Разбираемая задача относится к дисциплине, занимающейся так называемой проблемой распознавания образов. По этой проблеме существует в настоящее время обширная литература.

Следует, однако, заметить, что подавляющая часть этой литературы рассматривает проблему как обучение распознаванию образов. Процедуры, основанные на этом принципе, строятся так: существует некоторое множество объектов, предъявляемых машине; машина обладает некоторым числом ответных реакций.

Машина должна сопоставить каждому предъявлению некоторого объекта какую-нибудь из ответных реакций (т. е. фактически дать некоторую классификацию предъявляемых ответов). Если машина ошибается, то «учитель» — человек, обучающий машину, сообщает ей об этом и «наказывает», в противном случае «поощряет». По окончании обучения машина распознает объекты самостоятельно.

Разумеется, в дешифровочной ситуации обучение невозможно; кроме того, в нашей задаче недостаточно классификации объектов — необходимо еще и указать границы, отделяющие буквы одну от другой.

Если отвлечься от задачи объединения «образов» одной и той же буквы в классы, то задача выделения букв станет очень похожей на задачу выделения простых морфем. С некоторой степенью точности можно считать, что буквы занимают неперекрывающиеся куски плоскости, и сочетания точек, занимающих эти куски, являются в некотором смысле устойчивыми.

Не надо думать, что буквы следует выделять как разумные изображения. Буквы совсем не обязательно должны быть гладкими и связными, а могут представлять собой случайный набор точек; достаточно, чтобы все вхождения одной и той же буквы представляли собой примерно один и тот же набор точек.

Яркой специфической особенностью задачи выделения букв по сравнению с задачей выделения морфем является то, что не существует двух одинаковых вхождений одной и той же буквы, а только более или менее похожие. При этом очень трудно уловить, на чем основано это сходство: порой очень незначительное изменение очертания переводит одну букву в другую (например: И и Й, Ш и Щ), а более значительные изменения оказываются несущественными (например, σ и z , d и d'). Обычно размеры букв, наклоны и растяжения не играют роли, но латинское рукописное e и l отличаются только размерами, E и 3 — поворотом, P и b — поворотом в плоскости и отражением.

Существуют системы письма, например, стенография, где осмысленными являются еще менее броские черты — приподнятость над уровнем строки, толщина линий и т. д.

По-видимому, наиболее эффективный алгоритм должен был бы содержать правила, по которым из элементарных признаков строились существенные признаки букв. Возможно, что некоторые из этих признаков можно было бы опознать лишь в присутствии других букв (такие, как приподнятость над строкой, наклоны, размер).

Поэтому распознавание букв без учета соседних букв следует признать невозможным. Можно только надеяться, что в большинстве случаев достаточно осматривать весьма недалекую окрестность распознаваемой буквы.

Приводимая ниже идея алгоритма не преследует серьезных целей. Тем не менее она представляет известный интерес благодаря использованию весьма малой исходной информации (в частности, применяемые приемы позволяют обходиться без задания размеров

букв) и определению понятия частоты, в данном случае неочевидного.

Опишем вначале, как определяется так называемое остаточное сходство. Пусть знак a расположен на некотором куске K_1 площади текста. Будем считать, что его описывает функция $\delta(x, y)$, где x и y — прямоугольные координаты точек этого куска, а $\delta(x, y)$ принимает два значения: значение 1 для черной точки и значение 0 — для белой.

Пусть на втором куске, K_2 , расположен знак b , описываемый функцией $\delta_1(x_1, y_1)$, где x_1, y_1 — координаты точек куска в своей системе координат. Эти координаты также прямоугольные.

Совместим начало координат первого куска и второго, а также направления координатных осей. Будем считать, что остаточное несходство $\frac{1}{S}$ определяется выражением

$$\frac{1}{S} = 1 + \frac{1}{|K_1 \cap K_2|} \int \int_{(K_1 \cap K_2)} (\delta_1(x_1, y_1) - \delta(x, y))^2 dx dy.$$

Символ S здесь обозначает «остаточное сходство». Пересечение кусков обозначено символом $K_1 \cap K_2$.

Разрешим производить над координатами куска K_2 следующие преобразования: 1) параллельный перенос, 2) преобразование подобия, 3) сжатие вдоль каждой из осей, 4) изменение угла между осями, 5) поворот в плоскости, 6) зеркальный поворот.

Будем преобразовывать функцию δ_1 и координаты x_1, y_1 так, чтобы значения новой функции в новых координатах совпадали со значениями старой функции в старых координатах, т. е.

$$\delta_1^1(x_1^1, y_1^1) = \delta_1(x_1, y_1),$$

где $\delta_1^1(x_1^1, y_1^1)$ означает новую функцию в новых координатах. При этих условиях изменение функции будет вполне определено указанием преобразования координат. Если новую функцию мы выразим в старых координатах и определим остаточное сходство с $\delta(x, y)$,

то увидим, что оно изменилось по сравнению с остаточным сходством до преобразования. Таким образом, каждому преобразованию исходных координат можно поставить в соответствие некоторую величину остаточного сходства.

Пусть шесть типов преобразований определяют шесть осей в пространстве преобразований. Вдоль каждой оси отметим «значения значений», которые может принимать данное преобразование (предполагается, что для каждого преобразования определена одномерная оценочная функция). Точки этого пространства, определяемые наборами значений оценочных функций, назовем допустимыми точками. В каждой допустимой точке определена скалярная функция — остаточное сходство, и векторная функция — градиент, указывающая направление быстрее́го роста остаточного сходства. Двигаясь вдоль градиента, можно получить точку, в которой значение остаточного сходства максимально для двух данных кусков и знаков. Это значение мы будем считать истинным сходством знаков a и b . Обозначим его также через S ; этот символ в дальнейшем будем понимать именно в таком смысле.

Возьмем произвольный кусок K площади текста и назовем его эталоном. Контур, ограничивающий этот кусок, будем смещать параллельно себе минимальными шагами по вертикали и по горизонтали. Для каждого положения контура определим сходство куска, ограниченного данным положением, с куском, ограниченным исходным положением контура. S в данном случае является функцией положения контура. Количество максимумов этой функции мы будем считать частотой знака, содержащегося в исходном куске текста. Пусть дан кусок текста, состоящий из двух квадратиков, K_1 и K_2 , имеющих общую сторону. Абсолютные частоты первого и второго квадрата и частоту прямоугольника, состоящего из них, будем обозначать $\varphi(K_1)$, $\varphi(K_2)$ и $\varphi(K_1, K_2)$.

Назовем прогнозируемостью K_2 по K_1 , или $p(K_2/K_1)$ величину $\frac{\varphi(K_1 K_2)}{\varphi(K_1)}$ и, соответственно, прогнозируе-

мостью K_1 по K_2 величину $\frac{\varphi(K_1 K_2)}{\varphi(K_2)}$, или $p(K_1/K_2)$, а среднюю из этих величин — взаимной прогнозируемостью K_1 и K_2 или $p(K_1, K_2)$.

Теперь будем увеличивать размеры K_1 и K_2 , замечая прогнозируемость при каждом увеличении размеров на минимальный шаг (постоянны при этом: прямая, на которой лежит граница, и положение центра границы). Тем самым мы представим взаимную прогнозируемость как функцию от изменения размеров квадратов. График этой функции будет, по-видимому, иметь примерно такой вид, как на рис. 70, где D — размер прямоугольника.

Действительно, все значения взаимной прогнозируемости будут меньше единицы: ведь знаки, содержащиеся в парах квадратов, более индивидуальны, более богаты деталями, чем знаки, содержащиеся в отдельных квадратах; поэтому знаки, максимально похожие на эти последние, должны встречаться чаще чем знаки в парах квадратов.

При малых размерах квадратов значения взаимной прогнозируемости близки к единице, поскольку знаки в квадратах и в парах квадратов еще похожи друг на друга и их функции сходства имеют максимумы примерно в тех же местах текста.

При возрастании размеров взаимные прогнозируемости сначала убывают, а затем растут и становятся равными единице, поскольку знаки в больших квадратах, равно как и парах квадратов (прямоугольнике), встречаются только один раз.

Найдем такое минимальное значение b переменной D , что при любом выборе в тексте места границы между квадратами K_1 и K_2 (при размерах квадратов, соответствующих b) взаимная прогнозируемость

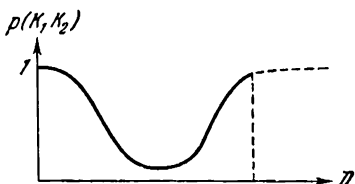


Рис. 70. Возможный вид графика функции $\bar{p}(K_1, K_2)$ в зависимости от размеров квадратов K_1 и K_2 .

$p(K_1, K_2)$ равна единице. Назовем центральностью границы квадратов K_1 и K_2 величину

$$C = \int_0^b p(K_1, K_2)(D) dD. \text{ Центральность характеризует}$$

взаимную прогнозируемость квадратов независимо от их размеров и служит, тем самым, характеристикой места границы.

Основной гипотезой нашего алгоритма является предположение о том, что центральность принимает



Рис. 71. Возможный график функции C в зависимости от перемещения места границы.

минимальное значение (по сравнению с соседними положениями границы) тогда, когда граница расположена между буквами и ориентирована вдоль истинной границы между буквами. Дальнейшие действия таковы: начнем перемещать границу в некотором направлении, замеряя значение C

через каждый шаг. Мы получим функцию, график которой будет выглядеть примерно так, как на рис. 71.

Расположим границу в точке, для которой зафиксирован некоторый минимум, и будем вращать границу вокруг своего центра, замеряя C через каждый минимальный поворот. Выберем угол α , соответствующий минимальному значению C , и сделаем в этом направлении минимальный шаг. В новом положении границы мы снова выберем наилучшее направление дальнейшего следования тем же способом. В том случае, когда есть несколько наилучших направлений, будем выбирать то, которое правее. Если мы действительно идем вдоль границы буквы, то опишем замкнутую кривую, окружив эту букву.

Знак, содержащийся внутри кривой, мы можем взять за эталон и найти все максимумы S по данному эталону в исследуемом тексте.

С точки зрения предложенного нами определения сходства, мы не вполне вправе называть выделенные знаки буквами исследуемого алфавита. Например,

русские буквы Р и Ъ заведомо не будут различены друг от друга, поэтому назовем полученный алфавит каркасом истинного.

Не вдаваясь в детали, можно наметить идею получения алфавита из его каркаса. Рядом с каждым вхождением знака каркаса запишем вектор пространства преобразований, переводящий его, например, в первое вхождение этого знака. Можно заметить, что некоторые значения, составляющие этот вектор, устойчивы, т. е. сильно прогнозируются соседними знаками. Например, буквы слова «гора» будут предсказывать положение знака «р» палочкой вниз, а в слове «весьма» будут предсказывать положение того же знака палочкой вверх. Случайные изменения знаков не будут прогнозироваться сколько-нибудь устойчиво.

§ 12. Заключение

В заключение нам хочется обратить внимание читателей на любопытный аспект дешифровочной проблематики.

Речь идет о том, насколько чужда может оказаться нам действительность, о которой расскажут космические послания. Сможем ли мы представить себе сущность этих посланий?

Следует сразу сказать, что понять сообщение и представить то, о чем там написано — разные задачи. Если понимание основано на возможности предсказывать неосмотренные части сообщения или непроизшедшие события, то представление связано с нашей способностью переводить сообщения на язык образов явлений действительности.

Конечно, далеко не все то, что понято, может быть представимо. Мы не можем представить ощущения существа, воспринимающего радиоволны, но это не умаляет понятности его поведения. Поэтому даже если «иная» действительность окажется чудовищно непохожей на нашу, у нас все-таки не будет оснований опасаться ее непознаваемости.

ЛИТЕРАТУРА

1. Апресян Ю. Д., Идеи и методы современной структурной лингвистики, М., 1966.
2. Грозный Б., Хеттские народы и языки. Вестник древней истории 2 (3), 1938.
3. Добрушин Р. Л., Математические методы в лингвистике. «Математическое просвещение» 6, М., 1961.
4. Каплан С. А., Элементарная радиоастрономия, «Наука», 1966.
5. Pratt F., Secret and Urgent, N. G., 1942
6. Прим Р. К., Кратчайшие связывающие сети и некоторые обобщения. Кибернетический сборник 2, ИЛ, 1961.
7. Романов В. П., Интегральные методы опознавания.читающие устройства. М., 1962.
8. Сухотин Б. В., Алгоритмы лингвистической дешифровки. Проблемы структурной лингвистики, М., 1963.
9. Сухотин Б. В., Экспериментальное выделение классов букв с помощью электронной вычислительной машины. Проблемы структурной лингвистики, М., 1962.
10. Сухотин Б. В., Алгоритм, сравнивающий буквы двух различных языков. «Научно-техническая информация» 2, 1966.
11. Сухотин Б. В., Исследование языка дешифровочными методами. Русский язык в школе 6, 1966.
12. Harris Z., From phoneme to morpheme, Language, 1952, т. 28, № 1.
13. Халле М., Фонологическая система русского языка. Новое в лингвистике, М., 1962.
14. Шеннон К., Теория связи в секретных системах (Шеннон К., Избранные работы по теории информации и кибернетике, ИЛ, 1963).
15. Шрейдер Ю. А., Машинный перевод на основе смыслового кодирования текстов «Научно-техническая информация» 1, 1963.
16. Яглом И. М. и Яглом А. М., Вероятность и информация, Физматгиз, 1960.

Темпы развития цивилизаций и их прогнозирование

§ 1. Значение проблематики темпов

Существенное влияние на характер и интенсивность попыток установления космических связей могут оказать существующие в науке представления о возможных темпах развития жизни и цивилизаций на других планетных системах.

От этих представлений непосредственно зависит оценка распространенности жизни и разума во Вселенной, а также количества цивилизаций, которые по своему техническому развитию значительно обогнали человечество и, в частности, способны вести радиопередачи на расстояния в тысячи, миллионы и миллиарды световых лет.

Прежде всего возникает вопрос о том, насколько темпы истории земной цивилизации (а также ее биологической предыстории) являются типичными для цивилизаций вообще. Например, мнение, согласно которому условия, существовавшие на Земле, стимулировали биологическую эволюцию в большей степени, чем условия на других планетах, где могла возникнуть жизнь, заставляло бы отнести земную цивилизацию к одной из авангардных в нашем районе Галактики. Некоторые доводы в пользу такого мнения будут приведены ниже. Впрочем, наряду с фактором благоприятствующих эволюции условий следует также учесть фактор сроков возникновения жизни и степень детерминированности темпов эволюции заданными условиями (можно предположить, что при одних и

тех же начальных условиях и внешних воздействиях квантовые флуктуации могут вызывать различия в сроках появления важных мутаций, а следовательно, в сроках и направлении эволюции).

Для радиосвязи с удаленными цивилизациями не менее важен прогноз собственных темпов развития человечества в будущем. Различные прогнозы возможностей разработки средств передачи и приема сигналов могут повлечь за собой разные решения о допустимых затратах на их разработку. Эти решения могут зависеть от представления о внутренней детерминированности темпов и направления развития (т. е. от того, насколько эти темпы способны возрасти в результате получения сигналов извне, от более продвинутых цивилизаций) и от представления о важности сохранения высоких темпов научно-технического прогресса.

Во многих научных и научно-популярных публикациях допускается большое разнообразие форм жизни и цивилизаций на разных планетах, но молчаливо предполагается, что в темпах эволюции большого разброса нет. Например, Мак-Гован и Ордуэй [1] приводят три возможных графика распределения длительности периода от возникновения жизни до появления цивилизации.

На каждом из графиков среднее квадратичное отклонение составляет 10—20% средней длительности. Если, следовательно, средняя длительность составляет 3—4 млрд. лет, то почти ни на какой планете жизнь не будет эволюционировать до разума менее одного или более 6 млрд. лет. Такая точка зрения заранее лишает космос богатства и разнообразия.

Даже на Земле темпы эволюции в различных зоогеографических районах весьма неодинаковы: малым изолированным зонам (Австралия, Мадагаскар) соответствуют архаичные формы животных. Земные аналогии заставляют ожидать гораздо меньшей скорости эволюции на планетах размером с Марс или Луну или на планетах с малой площадью суши. С этой же точки зрения на более крупных планетах

темпы выше, так как там при прочих равных условиях имеется большее число особей.

Скорость эволюции должна снизиться при однообразии климатических условий (отсутствие ледниковых периодов) или, наоборот, их слишком резкой изменчивости. Отсутствие наклона плоскости экватора к плоскости орбиты также может снять многие из задач эволюции или приостановить ее. Перпендикулярность же плоскостей экватора и орбиты приведет к дальним годичным переселениям многих животных, что помешает изоляции отдельных зоогеографических зон и уменьшит возможности разветвлений в эволюции. Количество крупных спутников планеты определяет сложность колебаний ночного освещения и степень гибкости поведения ночных животных. Все сказанное, правда, справедливо при сходстве органического мира планеты с земным, особенно в отношении экологии и механизмов наследования и изменчивости. Но из сказанного следует, что темпы эволюции как жизни, так и ее социальных проявлений, цивилизаций, могут быть весьма различными. К сожалению, пока нельзя сколько-нибудь уверенно судить о темпах эволюции и развития внеземных цивилизаций [2]. Ниже мы будем исходить только из данных о темпах эволюции нашей земной цивилизации.

§ 2. Аспекты развития цивилизаций

Если такие понятия, как переработка, хранение и передача информации полезны для характеристики развития общества как единого целого, то для конкретного описания какого-либо этапа в этом развитии таких понятий недостаточно. Характеризуя темпы развития цивилизации, мы рассматриваем их в таких аспектах, как социально-политическое и экономическое развитие, эволюция языка, искусства, развитие науки и техники, роль религии и т. д. Мы не можем, однако, утверждать, что наше развитие всегда будет рассматриваться в этих аспектах, а тем более предполагать, что и другие цивилизации также развиваются подобным образом.

Можно ли быть уверенным, например, что стадия религиозного аспекта имеется у всех цивилизаций? С одной стороны, можно встретить утверждения, что возникновение религии обусловлено чисто земными причинами [3], а с другой стороны, даже слонам приписываются религиозные чувства и молитвы [4]; если это так, то, вероятно, религия — всеобщее явление. Между прочим, исследование религиозной истории для общей теории темпов развития любых цивилизаций интересно и потому, что религия (во всяком случае на поздних этапах ее эволюции) служит примером аспекта, тормозящего развитие цивилизации.

Любопытны сопоставления темпов развития искусства и науки. Дискуссия «физиков и лириков» на страницах нашей печати в конце 50-х годов скорее пыталась поставить вопросы, чем отвечала на них. Е. Л. Фейнберг [5] отмечал, что темпы развития науки уже давно превышают темпы роста искусства и существующую ситуацию можно проэкстраполировать на будущее. Менее ясно, однако, как меняется значимость науки и искусства в жизни общества и его членов. Наше понимание закономерностей современной эволюции искусства еще слишком далеко от возможности применить к нему количественные методы.

Важное значение для характеристики эволюции цивилизации имеет также вопрос о развитии языка и вообще способов коммуникаций между членами общества.

Язык и коммуникации. Общение между членами популяции, между различными особями стада задолго до возникновения человечества привело к стихийной выработке специальных систем условных знаков, в частности, звуковых систем. Основным отличием языка человека от «языка» животных является возможность вводить новые знаки, объясняя их смысл средствами только самого языка [6]. Следует заметить, что, несмотря на ускорение развития весьма многих аспектов цивилизации, языки в последние тысячелетия эволюционируют примерно постоянным темпом, сохраняя за 1000 лет около 85% лексики [7].

Несколько быстрее в последние века и десятилетия стал расти словарный состав в результате обогащения самой действительности и развития науки. Но заметнее всего бурное развитие цивилизации проявилось в создании массы новых знаковых систем: дорожных и речных знаков, химических формул, алгебры и анализа, теории множеств и математической логики, библиотечных шифров и шифров стандартов. Медленному изменению языка противостоит быстрое изменение пропорции устной и письменной речи, места формул и чертежей в письменной речи, роли чертежей, диапозитивов и кинофильмов в устной. Обычный язык и не пытается конкурировать с языком формул: сравним значение слова «твердость» в двух следующих предположениях: «Твердость гипса равна 2» и «Сплавы этого типа отличаются твердостью».

Возможность различения весьма непохожих логически значений одного и того же слова обеспечивается контекстом, семантической избыточностью речи. Подобные же расхождения обнаруживаются у разных значений слов «качественный», «высота», «емкость» и т. д. Лишь в редких случаях говорящий прибегает к поясняющим словам «степень», «величина», «мера», или к составлению сложного слова, присоединяя, скажем, морфему «-grade» (нем.). Следует полагать, что «слова» для языков, на которых будут общаться удаленные друг от друга цивилизации, могут не совпадать со словами современных языков.

В каких же количественных характеристиках коммуникации проявляются уровень цивилизации и темпы ее развития? Скорее всего, не в структуре каждого отдельного естественного языка, а в количестве различных знаковых систем, но систем, действительно необходимых, остающихся после унификаций и стандартизаций. Это статистические характеристики используемой лексики [8] и в первую очередь запаса используемых терминов. Еще более ярко отражается научно-технический прогресс в статистических характеристиках чисел, встречающихся в текстах. В ходе истории повышается разброс чисел по порядкам (однозначные, двузначные и т. д.), все чаще

появляются записи со степенями десятки. Округление приближенных чисел становится более мотивированным, все реже значащие цифры обрываются на разряде именно единиц. Вслед за ростом порядков мы встречаемся с округлением в показателе степени: чаще встретишь 10^{60} , чем 10^{59} . Подобные тенденции достаточно просты, поддаются исследованию без больших материальных затрат и объясняются весьма общими и фундаментальными причинами. Можно предполагать правомерным обобщение этих тенденций и на внеземные цивилизации.

Более сложно эволюционирует соотношение доли точных (дискретных) и приближенных («непрерывных») чисел в текстах. Средневековые тексты отличаются преобладанием дискретных величин: число людей, объектов, операций; формы небесных тел и их траектории рассматриваются как идеальные сферы и окружности. Числа 7 и 3 наблюдатели видят чаще, чем они встречаются в реальности. Промышленная революция и развитие капитализма ведут к отказу от некоторых предрассудков, к измерению вместо подсчета и рассуждения. Внимание инженеров сосредоточивается на физических процессах, прежде всего энергетических. Дискретные величины сдают позиции. Но в конце XIX в. физика возвращается к ним, обнаружив элементарные заряды, квантование величин в микроявлениях. Вскоре в изучении как живой природы, так и техники, внимание переключается с энергетических процессов на информационные. В описаниях многих явлений внешнее описание дополняется структурным, что способствует бурному развитию «дискретной» математики. Такой путь и для внеземных цивилизаций представляется достаточно правдоподобным.

Наоборот, многообразие живых языков на нашей планете и отсутствие видимого родства языковых семей можно приписать особенностям земного рельефа, трудностям морских путешествий и т. п. На планете с более компактной сушей, меньшим числом гор и пустынь, но с большим числом судоходных рек еще на заре цивилизации мог бы сложиться единый ми-

ровой язык. Другим достаточным условием возникновения такого языка могла бы явиться лучшая техника транспорта или связи.

Нет также оснований считать, что цивилизация не может искусственно создать единый международный или мировой язык. Почему этого не произошло на Земле? Международный язык ученых и теологов средневековья, латынь, был труден для изучения и достаточно далек от любого из живых языков. В противном случае он, вероятно, остался бы международным или даже стал мировым языком. В период промышленной революции наука, техника и искусство развивались в ряде ведущих стран, среди которых не было хотя бы двух с одинаковым языком или с близкими языками. Не потому ли ни один из этих языков не сделался мировым? Во второй половине XIX в. появились проекты искусственных языков. Многие ожидали широкого их распространения (в первую очередь эсперанто*). Но первая мировая война усилила националистические барьеры, а богатая событиями последующая история отвлекла внимание от этой проблемы. В настоящее время вопрос в значительной степени потерял свою остроту из-за перспективы машинного перевода, совершенствования методики изучения иностранных языков и более широкого распространения образования.

С другой стороны, нет никаких оснований считать, что цивилизация не сможет развиваться и даже выйти в космос при гораздо большем разнообразии жаргонов, диалектов, языков или языковых семей, чем у земного человечества.

Демографические характеристики цивилизации. Имеющиеся в распоряжении статистики данные о росте численности населения являются более или менее точными лишь за последние несколько сотен лет. Данные о распределении населения между городом и

*) Имеется свидетельство, что в некоторых сельских местностях Британской Гвинеи он действительно сохраняется как живой язык [9].

деревней, распределении по возрастам и т. п. для большинства стран имеются только за полтора-два последних столетия.

Весьма богатая демографическая литература или концентрирует внимание на изучении и прогнозе роста и изменения состава населения отдельных стран или ограничивается сравнительно краткими периодами времени. В работах, рассматривавших рост населения Земли за несколько прошлых столетий, вопрос о подборе единой аналитической зависимости, в которую укладывались бы данные, либо не рассматривался, либо в качестве такой зависимости применялась экспонента и степень точности оказывалась весьма низкой. Между тем данные о населении Земли хорошо укладываются в гиперболическую зависимость от времени, как это показал И. С. Шкловский [10]. Впрочем, в той же работе отмечается, что эта зависимость должна в скором времени нарушиться за счет совершенствования управления обществом. Рост процента ежегодного прироста может замедлиться также вследствие того, что с распространением комплексной механизации и автоматизации численность населения страны уже не определяет столь жестко, как прежде, потолок ее экономических возможностей. Для быстрого экономического роста нужен передовой социально-экономический строй, нужна хорошая организация производства, планирования, образования, но уже в меньшей степени необходим прирост народонаселения.

При интерпретации того или иного прогноза численности населения Земли на будущее могут встать трудности, связанные, во-первых, с возможным открытием существ, стоящих на грани разума [11], а, во-вторых, с созданием систем, близких к разумным [2], в результате обучения дельфинов [12] или в результате хирургического или селекционного получения более крупного мозга у обычных животных, или наконец, в результате конструирования имитирующих человека машинных программ. Другого рода трудности подсчета могут быть связаны с возможным сохранением людей в состоянии анабиоза. Разумеется, не

все из названных факторов могут дать значительную ошибку при определении численности населения Земли.

В применении к другим цивилизациям понятие численности населения может оказаться еще более неопределенным, как за счет указанных причин, так и за счет неясности самого понятия «организм». Например, в применении к земным пчелам и муравьям некоторые авторы предлагают рассматривать в качестве организма не генетическую особь, а улей или муравейник [13].

Социальные и культурные изменения в обществе могут привести к нарушению количественных закономерностей роста как населения в целом, так и отдельных демографических показателей. Так, процент населения земного шара, жившего в городах с населением свыше 10^5 человек, составлял в разные годы t [14]:

t	1800	1850	1900	1950	1960
p	1,7	2,3	5,5	13,1	20,1

Легко проверить, что рост этой доли p населения, а тем более отношения городского населения к сельскому $\frac{p}{100-p}$, происходит не в геометрической прогрессии от времени, а с ускорением относительных темпов роста.

Данные этой таблицы можно аппроксимировать уравнением

$$p\% = \frac{110}{110 + (2000 - t)^{1/3}} \cdot 100\%.$$

Эту модель можно интерпретировать как превращение всей Земли (или каждого материка) к концу XX в. в один большой город. Между тем в настоящее время проявляется тенденция дезурбанизации, сохранения малых поселков и стремления части жителей в эти поселки. Можно предположить, что распространение телевидения, единой телефонной и видеофонной сети, дистанционное пользование библиотечными фондами задержат ускоренное развитие городов. ...

Развитие индивидуальных способностей. Для оценки достигнутого уровня и темпов развития цивилизации представляется интересным оценить средний (или рекордный) уровень способностей составляющих ее индивидуумов, а также меру полноты использования этих особенностей. При этом возникают многие трудности, прежде всего трудности построения объективного и воспроизводимого способа измерения отдельной способности, полноты охвата этими измерениями всего человечества и выбора из построенных таким образом показателей наиболее значимых (и независимых друг от друга).

Удобный для измерения показатель может оказаться мало эффективным для оценки динамики развития, если мы не в состоянии оценить его величину для прошлых эпох. До сих пор не выработано эффективных способов измерения общих интеллектуальных способностей. Распространенные на Западе тесты на одаренность дают числовой скалярный показатель, за которым скрываются не только творческие способности, но и запас знаний и навыков и культурный уровень проходящего тест. В последние годы сами западные ученые признают это.

Стремясь отобрать не только знающих, но и способных абитуриентов, преподаватели некоторых вузов и факультетов приобретают опыт (скорее интуитивный, чем научно обоснованный) диагностирования именно творческих способностей, примером чего могут служить математические олимпиады, проводимые во многих городах СССР.

Но этот интуитивный опыт нелегко применить к оценке способностей ученых или изобретателей прошлого, когда запас знаний, общая картина мира и способы обучения были совсем иными, была иной и трудность решавшихся задач. В настоящее время население Земли во много раз больше, чем в античную эпоху или в средние века, образование доступно в тысячи раз большему числу людей, чем, скажем, в античной Греции. Не будет ли естественным предположить, что среди современных ученых есть десятки умов, превосходящих Архимеда, Лейбница,

Ньютона и Ломоносова? Быть может, на месте Декарта гораздо больше сделал бы любой из таких ученых, как Бор, Винер или Ландау? Подобные вопросы можно было бы задать о разных областях проявления таланта: о поэзии (Пушкин или Блок), о предвидении техники будущего (Роджер Бэкон или Уэллс), о шахматах (Морфи или Алехин). Вряд ли удастся однозначно ответить на них без серьезных успехов как в психологии творчества, так и в психологии развития способностей, без анализа задач, которые ставила каждая эпоха перед культурой, без анализа возможностей развития, которые она давала людям. Трудно найти задачи, действительно одинаковые по трудности для разных эпох. Число изученных иностранных языков зависит от особенностей не только ума, но и памяти; к тому же в разные эпохи языки занимали разное место в начальном обучении.

Существует очень мало творческих задач, независимые решения которых появлялись бы на протяжении многих эпох. Утерянное доказательство великой теоремы Ферма не было повторено за три столетия, но где гарантия, что в первом доказательстве не содержалось ошибки?

Но разве сами достигнутые индивидами результаты, полнота использования возможностей мозга (и всего тела) не служит мерой успехов цивилизации? И здесь далеко не все поддается измерению. Для некоторых профессий производительность труда хорошо прослеживается на протяжении многих эпох, но уже не так легко сопоставлять успехи полководца, учителя, популяризатора. Можно сопоставлять шахматистов по глубине и сложности найденных (или пропущенных) ими комбинаций, но не видно способов оценки того, насколько недостаток комбинационного зрения компенсируется исторически накопленными знаниями.

В лучшем положении оказываются измерения достижений собственно физической культуры. Однако и здесь сопоставление взаимно удаленных эпох не обходится без затруднений. Только с XIX в. спортивные достижения регистрируются регулярно, точно и в

точно заданных условиях. Эти условия также не вполне стабильны: изменилась обувь и условия старта спринтера, материал шеста у прыгуна и др. Бурно растет число новых видов спорта, поэтому на каждый отдельный вид приходится все меньшая доля спортсменов. С другой стороны, несколько возрастают специализация и профессионализация.

Выше мы кратко рассмотрели некоторые аспекты развития цивилизации, которые наглядно показывают, сколь трудно судить об их темпах, об изменении этих темпов, а тем более дать количественные выражения этим закономерностям. Гораздо лучше обстоит дело с оценкой темпов развития таких аспектов цивилизации как экономика, техника и наука. Эти аспекты представляют больший интерес и с точки зрения выяснения возможностей межзвездной связи.

В следующих двух параграфах мы несколько подробнее рассмотрим технический и научный аспект развития цивилизации, но по-прежнему ограничимся больше качественным описанием — для нас сейчас важнее именно общая картина ускорения темпов развития, условия смены показателей, характеризующих это развитие, и подобные вопросы. Дело в том, что конкретные количественные характеристики пока что относятся к очень небольшим промежуткам времени и распространять их на условия внеземных цивилизаций очень рискованно.

§ 3. Показатели технического прогресса

Вероятно, важнейшей характеристикой уровня цивилизации является объем технических знаний, показатели отдельных технических средств, количество и качество выпускаемой продукции, количество используемой энергии и т. д. Важно и то, насколько системным, организованным является управление всей совокупностью технических средств, использование техники как целого.

Наука, медицина и образование также существенно зависят от уровня технических знаний. Примеры этого общеизвестны. Но техника играет в на-

стоящее время большую роль не только как средство научных исследований, но и как причина развития науки, ставя перед ней все новые задачи.

В настоящее время имеется богатая литература по экономической, технической и научной истории отдельных стран и всего человечества. Много работ посвящено исследованию роста отдельных показателей развития (мощности, скорости и т. п. [15, 16]). Однако эти сведения в основном относятся к последним десятилетиям. Значительно менее полны данные за последние несколько столетий.

Сопоставление числовых данных, взятых из разных эпох технического развития, не для всех показателей одинаково правомерно: скажем, рост производства энергии является более корректным показателем, чем рост парка металлообрабатывающих станков, так как производительность и точность станка резко изменяются в процессе технического прогресса, а перевозки в тонно-километрах являются более точным показателем, чем число судов. Вообще более обоснованные выводы можно делать на основе роста таких показателей, единица которых сохраняет одинаковое значение для разных эпох.

Среди показателей роста продукции, прослеживаемых на большом промежутке времени и сопоставимых в течение всего этого периода, имеются такие величины, как мировая добыча серебра или угля, темпы которой все время возрастали, начиная с XVI—XVII вв. (когда стали регистрироваться более или менее точные количественные данные) до конца XIX в., а затем эти темпы несколько замедлились. Зато вместо угля стала быстро возрастать добыча нефти, причем темпы роста этой добычи намного выше настоящих (и прошлых) темпов роста добычи угля. Подобным же образом серебро уступило место другим химически стойким материалам (в том числе пластмассам), производство которых растет сейчас невиданными для серебра темпами.

В развитии транспорта рано или поздно рост скоростей и мощности одного вида транспорта сменяется переходом к другому виду и т. п.

Все это означает, что темпы роста цивилизации характеризуются сменой ведущих показателей.

О преемственности показателей. Приведенные выше примеры показывают, что на каждом этапе развития цивилизации существуют определенные основные показатели, которые, однако, на следующих этапах уступают место другим показателям. Более того, можно даже говорить об определенной преемственности показателей между эволюцией животного мира и развитием человеческого общества. В развитии общества продолжаются ускорения темпов (также со сменой показателей), наметившиеся еще в ходе биологической эволюции [17]. Можно даже составить шкалу эволюции от $5 \cdot 10^9$ лет до 50 лет, с каждым последующим уменьшением периода в десять раз, и сопоставить этой шкале важные этапы в развитии жизни, общества и техники [18].

Выбирая существенный показатель в ряду сменяющих друг друга технических средств, нельзя быть уверенным, что этот показатель остается столь же существенным и для будущего.

Человеку средних веков был, например, важен прогресс в термонизоляционных материалах, в прочности вообще, а позже большее значение приобрел прогресс в электронизоляционных материалах или в прочности на единицу веса (с переходом от строительства крепостей к высотным зданиям).

В ходе развития техники меняется не только относительная роль роста различных показателей в каждом из направлений технического прогресса, но и относительная роль самих этих направлений, примером чего служит появление управляющих и вычислительных устройств, которые перехватывают название «машины» у энергетических. Вместе с тем в литературе отмечается исключительная роль энергетического направления на всем протяжении истории не только техники, но и всей жизни на планете.

Можно было бы продолжать и дальше этот ряд укрупнений, рассматривая смену ведущих по темпам развития (или по значимости) аспектов цивилизации,

или даже, возможно, смену периодов, в которых темпы значимы для выживания цивилизации (или просто сравнительно высоки), и периодов, когда они незначимы (или сравнительно низки). Но в данный момент наука вряд ли располагает достаточно определенными подходами к измерению уровня основных аспектов цивилизации, темпов и значимости их развития. Не представляется достаточно точно выясненной даже функция каждого аспекта на каждом этапе развития цивилизации. Не вполне ясно, скажем, имела ли первобытная религия какую-либо позитивную функцию, хотя бы мнемоническую, конструирование ложных, сказочных связей для того, чтобы богатый набор эмпирических фактов приобрел осмысленность и тем самым легче удерживался в памяти.

Если динамику развития каждого отдельного технического средства можно считать близкой к экспоненциальной, т. е. процент ежегодного прироста соответствующего показателя постоянным, то при охвате этим показателем ряда сменяющих друг друга средств следует учесть и смену величины ежегодного прироста. Без специального «шкалирования» времени по общим темпам развития техники, накопления технической информации, не представляются обозримыми многообразные числовые данные по развитию различных отраслей в различные эпохи.

При известном сходстве медицины и образования с техникой их прогресс, рост уровня их достижений поддается измерению со значительно большим трудом. Вот некоторые причины.

Успехи медицины могут оцениваться по показателям смертности, по состоянию здоровья населения, его работоспособности. Но состояние здоровья и работоспособность не так легко выразить в виде объективного, воспроизводимого численного показателя; кроме того, они зависят не только от уровня медицины, но и от социально-экономических условий, от условий труда и быта, от наследственных предрасположений людей, доживших до определенного возраста. Нелегко, например, учесть то обстоятельство, что в настоящее время вероятность дожить, скажем, до 60 лет

различается у людей, имеющих патологические отклонения в наследственности, и у людей, не имеющих таковых, в меньшей степени, чем прежде.

Трудности измерения прогресса в эффективности и качества образования связаны с тем, что результаты образования разнородны. Они включают знания, конкретные и абстрактные, различные «умения», в том числе умение использовать приобретенные знания и приобретать новые. Они включают развитие способностей, а также результаты, связанные с воспитанием. Социальные, экономические и технические изменения жизни общества заметно сказываются как на развитии приходящих в школу детей, так и на задачах, которые ставятся перед специалистом по окончании его обучения.

Математические функции, описывающие темпы роста. Мы не приводим конкретных числовых данных главным образом потому, что с точки зрения поиска сигналов внеземных цивилизаций вопрос о росте добычи серебра на Земле или о прогрессе в изучении океанских глубин не представляет особого интереса. Темпы роста потребления энергии уже обсуждались в гл. I и III. О закономерностях роста показателей, характеризующих развитие радиотехники, по мнению некоторых авторов, еще рано судить, так как слишком мал соответствующий период. В еще большей мере это относится к технике космических полетов. Главной целью этого параграфа было указать на смену и преемственность ведущих показателей развития техники. Поэтому сейчас практически невозможно сказать, чем и как можно охарактеризовать технику сверхцивилизаций, с которыми мы надеемся установить связь, ее уровень и темпы развития.

Тем не менее, хотя здесь и не приводятся количественные характеристики темпов развития цивилизаций, мы сделаем несколько замечаний по поводу математического выражения этих темпов.

Постоянная скорость роста дает линейную зависимость соответствующего показателя от времени. Вероятно, для основных показателей подобная зависи-

мость вряд ли когда-либо имеет место. У большинства показателей, характеризующих развитие цивилизаций, темпы роста быстро увеличиваются. Если скорость роста пропорциональна самой величине показателя (т. е. относительная скорость роста постоянна), то соответствующий показатель растет со временем по экспоненте.

Наконец, если имеет место увеличение также и относительной скорости роста, т. е. если с увеличением самого показателя относительная скорость его изменения достаточно быстро растет, то сам показатель увеличивается по гиперболическому закону. Характерная особенность этого закона состоит в том, что здесь показатель достигает бесконечного значения за конечный промежуток времени. Надо заметить, что многие показатели роста и развития земной цивилизации лучше всего укладываются именно на гиперболическую кривую (например, демографический показатель). Однако из того, что было сказано выше о смене показателей, следует, что гиперболический закон для каждого показателя должен в конце концов нарушаться.

В ряде случаев величины, характеризующие развитие цивилизации, по самой своей природе имеют определенные ограничения (например, показатель отношения числа научных работников к численности всего населения). Рост подобного показателя уменьшается с приближением к соответствующему пределу. Обозначим такой показатель через n и предположим, что на ранних стадиях имеет место обычная экспоненциальная зависимость

$$\frac{dn}{dt} = \text{const } n. \quad (5.1)$$

На поздних стадиях, при приближении к пределу, рост показателя должен замедляться. Можно предположить, что здесь $\frac{1}{n} \frac{dn}{dt}$ пропорционально разности между максимальным значением (например, 100%) и значением в данный момент времени, т. е.

$$\frac{dn}{dt} = \text{const } (n_{\text{max}} - n)n. \quad (5.2)$$

Или, в иных обозначениях,

$$\frac{dn}{dt} = \frac{n}{t_0} \left(1 - \frac{n}{n_{\max}} \right), \quad (5.3)$$

где t_0 — масштаб времени. Решение этого уравнения дает так называемую «логистическую» кривую

$$\frac{n}{n_0} = \frac{n_{\max}}{n_0 + (n_{\max} - n_0) e^{-t/t_0}}, \quad (5.4)$$

где n_0 — начальное значение данного показателя (при $t=0$). Особенностью графика логистической кривой является симметричность относительно точки перегиба. При несимметричности говорить о логистичности [20] неправомерно.

Наряду с перечисленными кривыми используются и другие зависимости. Например, вместо гиперболы можно использовать формулу Земана [19]

$$n = \text{const} \lg \frac{\text{const}}{t_0 - t}, \quad (5.5)$$

также дающую ускорение относительных темпов, но зато оставляющую конечным интеграл от этой кривой (т. е. полное количество «событий» осталось ограниченным).

При изменении по экспоненте остаются постоянными промежутки удвоения значений показателя. Именно поэтому часто темпы прогресса характеризуются временем удвоения соответствующего показателя [20, 21] (см. также следующий параграф). Однако для других кривых время удвоения не является инвариантной величиной.

§ 4. Темпы развития науки

Один из важнейших аспектов цивилизации, роль которого все время и притом очень быстро возрастает — это наука. Обычная деятельность токаря (или врача) дает результаты непосредственно там, где он работает. Плодами деятельности ученого может пользоваться все человеческое общество.

С другой стороны, десять одинаковых деталей, сделанных десятью токарями, имеют в общей сложности десятикратную ценность по сравнению с одной деталью. Десять же одинаковых работ, выполненных десятью учеными, не имеют много большей ценности, чем одно такое исследование. Наука отличается от материального производства, медицины, обучения и т. д., в частности, тем, что, с одной стороны, каждый ее продукт должен быть новым, а с другой, — раз полученный продукт не расходуется при его широком и многократном использовании, он не требует ремонта и может репродуцироваться в любых количествах. Научный факт может использоваться одновременно многими предприятиями мира, в то время как станок — только одним предприятием. Продукт науки является не просто слагаемым в общей сумме богатств, он увеличивает коэффициент воспроизводства всего общества, играя мультипликативную роль. Другое отличие науки от техники состоит в том, что научный труд становится коллективным, механизированным и автоматизированным на более поздних этапах развития общества, чем промышленный. Требование новизны каждого научного продукта ведет к более высокой роли случайности в процессе получения научного знания, чем в процессе материального производства или лечения.

Все эти отличия заставляют рассматривать науку отдельно от техники, хотя наука и становится в настоящее время непосредственной производительной силой, хотя научные исследования и разработки непосредственно служат конкретным задачам производства. Вопросы количественной характеристики темпов развития науки также будут рассмотрены в отдельном разделе.

Проблема измерения темпов роста значения науки — это прежде всего задача выбора существенных показателей. Число занятых сотрудников и величина финансирования не вполне отражают степень проникновения науки в жизнь общества, хотя с развитием цивилизации эти параметры быстро увеличиваются.

Главным продуктом научного исследования являются научные знания. Их рост и должен служить основным показателем прогресса науки. Однако объективно судить о количестве знаний очень трудно, и поэтому в качестве основных параметров часто все же выбирают именно количество научных работников, количество научных публикаций и отчетов.

За последние триста лет эти показатели растут в среднем по экспоненциальной кривой [20], но для разных областей науки периоды удвоения различны. Если в физике период удвоения числа публикаций составляет примерно 10—15 лет [20], то в некоторых разделах математической статистики он равен (за последние годы) всего двум годам [22]. Рост затрат и относительного числа научных сотрудников также следует экспоненциальному закону, хотя в дальнейшем этот параметр должен перейти на логистическую кривую.

Между прочим, с течением времени меняется и доля творчески работающих ученых, а также распределение времени между творческим и нетворческим трудом научных работников [28]. По оценке Прайса число подлинно творчески работающих ученых примерно равно квадратному корню из общего числа научных работников, причем первым принадлежит примерно половина всех публикаций и 70—80% существенных результатов.

В настоящее время мы можем сопоставлять количество ценных, значимых открытий в разные эпохи только на основе интуиции специалиста. Обратимся к этим данным, установленным весьма субъективно, но относящимся именно к количеству значимых шагов в науке. Е. Л. Фейнберг [5] считает, что научные открытия нашего столетия по своей относительной значимости существенно не отличаются от научных открытий каждого из трех предшествующих столетий. На этом основании он отстаивает экспоненциальное представление о развитии науки, указывая, что возрастает только абсолютный прирост научного знания, научного понимания мира, что увеличение этого прироста, а также тот факт, что именно в нашем столе-

тии наука догнала и перегнала искусство по своему значению в жизни общества, создают ложное впечатление большей значимости каждого последующего десятилетия, каждого последующего века для изменения состояния научного знания, изменения научной картины мира.

Однако Е. Л. Фейнберг приводит примеры прежде всего из области физики, а эта область знания начала развиваться раньше других; многие ее проблемы применительно к макромиру были в первом приближении твердо решены уже в прошлые столетия. Физика, вероятно, недостаточно репрезентативный материал для суждений о темпах развития науки в целом. Разные соотношения значимости различных пятидесятилетий выглядят по-разному в истории каждой из таких наук как астрономия и геофизика, химия и биохимия, физиология и учение о наследственности.

Впрочем, подобным образом можно оспаривать и любое другое субъективное мнение. Попробуем рассмотреть конкретный рост числа научных открытий в некоторых областях науки. Сам выбор значимого открытия еще остается субъективным, но обработка данных уже будет проводиться достаточно воспроизводимыми методами. Проследим, насколько часто приводятся даты из разных эпох в обзорных монографиях по истории некоторых отраслей науки.

Вот распределение дат по векам в одной из достаточно популярных книг по истории математики [25]. Не регистрировались даты жизни, а также даты, относящиеся к развитию истории математики, а не самой математики, даты в примечаниях и даты, записанные буквами, а не цифрами.

Период (века)	до н. э.	I—XIV	XV	XVI	XVII	XVIII	XIX	XX
Число дат	30	50	9	19	79	119	255	9
Возможная модель			8	20	50	125	312,5	

С XV по XIX в. рост достаточно близок к геометрической прогрессии, кроме замечательного для истории математики XVII века. Значение математических открытий последних десятилетий, естественно, еще не

осмыслено всесторонне, и трудно для популярного изложения. Это проявляется уже в том, что на первую половину XIX в. приходится 139 ссылок, а на вторую — 116. Фактически экспоненциальный рост количества дат имеет место лишь до первой половины XIX в.

Рассмотрим теперь распределение дат в указателе источников курса истории психологии [26]. Даты начинаются с IV в. до н. э. и число их растет вплоть до начала 30-х гг. нашего столетия. Находя годы t_y , к которым накопилось количество дат, равные целым степеням двойки 2^y , получаем следующую таблицу (округление до десятилетий):

y	4	5	6	7	8	9
t_y	1580	1670	1780	1850	1890	1930
Период уд- воения	90	110	70	40	40	

В хронологической таблице «важных открытий и изобретений в области химии» [27] количества накопленных событий, равные целым степеням двойки, приходятся на следующие периоды (двойные даты, записанные через тире, заменялись средним арифметическим):

y	2	3	4	5	6	7	8	9	10
	50 в.	22 в.	1 в.	15 в.	1630 г.	1730 г.	1820 г.	1860 г.	1915 г.
	до н. э.	до н. э.	до н. э.						
Период уд- воения	3000	2000	1500	200	150	40	40	55	

Из последних двух таблиц видно, что число накопленных ссылок растет с более или менее постоянным периодом удвоения только последние полтора-два столетия. В целом даже за последние три-четыре столетия единой экспоненциальной кривой нет. В некоторых книгах число накопленных ссылок замедляет свой относительный рост уже начиная с XVIII в. Весьма четко это проявляется в одном из курсов биологии [28]. В ряде курсов по истории науки число ссылок по

векам или более мелким единицам времени растет неравномерно: например, в истории лингвистики [29] наблюдается резкий всплеск числа ссылок на XVI в., заметны всплески также на XIII в. и на последнюю четверть XIX в.

Таким образом, пытаясь оценить объем знаний и результативность научных исследований на каждом этапе развития науки в целом или ее отдельных отраслей, мы не находим достаточно объективных и вместе с тем значимых показателей. Число журнальных публикаций — показатель слишком поверхностный, число открытий — слишком субъективный.

Возможно, что в каждой области науки можно будет найти более уверенные и объективные данные — такие, как мера полноты, точности и надежности знаний. Мерой полноты может служить отношение числа изученных к числу еще не изученных объектов того же класса, мерой точности могло бы служить количество значащих цифр в измерении тех или иных параметров, мерой надежности — период, в течение которого не обнаруживается недостаточная точность тех или иных величин. В настоящее время исследования в направлении поиска объективных параметров развития науки еще только начинаются.

Приведенное в последних параграфах подразделение аспектов развития человечества является традиционным и далеко не самым лучшим для задач настоящего сборника; их удобнее решать в более общих терминах стационарности, детерминированности или регулярности в потоках и запасах информации, энергии, вещества и т. д. Однако скороспелый переход к новой системе понятий вряд ли был бы целесообразен, равно как и попытка сформулировать окончательные выводы в традиционных терминах, которые способны хорошо служить только для решения традиционных задач, более узких и частных, чем задачи настоящего сборника и настоящей главы. На данном этапе разработки проблематики темпов и прогнозирования с точки зрения связи с внеземными цивилизациями необходимо накопление широкого круга разнообразных фактов, их проверка и систематизация на

новой основе. Всякие попытки категорических оценок интерпретации фактов и даже их подбора с точки зрения частной концепции или частной отрасли знания, будь то термодинамика или семиотика, могут только помешать рождению эффективных подходов в столь новой и необычной области.

§ 5. Прогнозирование

Управляющие системы в природе, технике и человеческом обществе снабжаются устройствами, способными принимать информацию, т. е. классифицировать воздействия извне с точки зрения тех решений, тех выборов, реализация которых и составляет функцию управляющей системы, способствующую сохранению или развитию этой или включающей ее более крупной кибернетической системы [30—32]. С некоторыми оговорками сюда можно отнести также информацию, получаемую изнутри. Более простые системы используют эти воздействия только для учета состояния внутренней и внешней среды в момент воздействия, более сложные — функционируют в соответствии с теми изменениями среды, которых следует ожидать в будущем исходя из ее состояния в настоящий момент. Такая экстраполяционная или прогнозирующая функция управляющих систем в последние годы подвергалась пристальному изучению в биологии [33] и в технической кибернетике [34].

С усложнением жизни человеческого общества, с развитием цивилизации возрастает и роль прогнозов. У вождя первобытного племени или шамана прогноз или лжепрогноз будущего был одной из ряда его функций. В рабовладельческом обществе появляются профессиональные оракулы и нередко специализирующиеся на прогнозах жрецы, причем несомненный в отдельных случаях успех предсказаний (например, затмений) объясняется чисто эмпирическим знанием. Общественное признание, которым длительное время пользовались и пользуются еще сейчас за рубежом оракулы, астрологи, гадалки и др. объясняется их относительными успехами за счет тонкого знания прак-

тической психологии людей. Интуитивные способы правильного (а чаще — просто понятного для заказчика) предсказания обычно неправильно мотивируются линиями руки, положением планет или игральных карт и т. д. Первые же успехи науки нового времени решительно подорвали авторитет лжепророков в культурных слоях общества. Подобная «стихийная» общественная оценка была закреплена теоретическими работами французских материалистов XVIII в. К сожалению, при этом не были изучены ни степень правильности предсказаний различных аспектов судьбы людей (что затруднялось, помимо всего прочего, воздействием предсказания на судьбу за счет внушения), ни психологические механизмы решения прогностических задач. Последние два-три столетия характерны тем, что наиболее интересные, значимые и правильные прогнозы исходили от лучших специалистов данной области человеческой деятельности. Правильных прогнозов исхода военной операции ждали не от оракула, а от известного полководца или политического деятеля, а правильных прогнозов исхода шахматного матча — от гроссмейстеров. В последние десятилетия положение несколько изменилось за счет более тесного взаимодействия разных видов деятельности и разных областей человеческого знания. Помимо процесса нарушения отраслевой специализации прогнозов намечается процесс их профессиональной специализации, появляются научные коллективы, ставящие своей прямой целью прогнозирование будущего. С другой стороны, все большую роль в духовной жизни общества начинают играть писатели-фантасты. Они становятся чистыми фантастами: Жюль Верн написал немало романов просто о путешествиях и приключениях, а сейчас эти жанры не так часто чередуются с фантастикой у одного и того же автора. Фантастами становятся ученые: Г. Уэллс был дипломированным биологом, А. Азимов — крупный исследователь в области биохимии, А. Кларк — астроном, И. Ефремов — палеонтолог. Сами фантасты, с другой стороны, переходят от сюжетного способа подачи своих взглядов на будущее к форме научного трактата или монографии

[35, 36]. Таким образом, в нашей цивилизации снова формируются и начинают играть важную функциональную роль подсистемы, специализирующиеся на прогнозе будущего.

О способах классификации прогнозов. Прогнозирование можно подразделить на прогнозы массовых событий (которые наблюдались много раз без существенной вариации условий их протекания) и редких или уникальных событий (которые, если и наблюдались, то очень малое число раз). Естественно, что первые прогнозировать легче.

Помимо приведенной классификации следует указать ряд других классификаций прогнозов.

Классификация по широте задачи позволяет различать: а) частные прогнозы («эту пешку я могу потерять»), б) прогнозы, относящиеся к существенным сторонам судьбы человека, коллектива, научного эксперимента и т. д., в) прогнозы развития отдельной отрасли промышленности или науки, г) прогнозы развития цивилизации в целом, д) прогнозы о том, как будут реагировать иные цивилизации на принятый ими сигнал и на обнаруженный ими аппарат, отправленный прогнозирующей это событие цивилизацией.

В литературе широко известно подразделение прогнозов по дальности [21]. При этом, к сожалению, в качестве единицы измерения длительности выбирается год или столетие, а не эпоха, не промежуток удвоения существенного показателя. Выбор календарных единиц длительности препятствует сопоставлению прогнозов, выполнявшихся цивилизацией в периоды с разными темпами развития или разными цивилизациями.

По способу получения прогнозы можно разделить на логически обоснованные и интуитивные, а также на прогнозы, использующие только качественные данные или также и количественные. Это подразделение можно распространить и на способ отбора экспертов для выполнения прогнозов: прогноз о пригодности тех или иных лиц к прогнозу также может быть логическим или интуитивным.

С точки зрения формы результата прогноз может быть детерминистическим и вероятностным; в последнем случае он может содержать дискретное распределение вероятностей, непрерывное распределение числовой величины, распределение в пространстве функций и т. д.

Не углубляясь в систематизацию этих способов классификации прогнозов, в их уточнение и выяснение их взаимосвязей, попытаемся рассмотреть пути получения прогноза и повышения его точности.

О точности прогноза. Разрабатывая и совершенствуя методы выработки прогноза, нужно прежде всего уметь сопоставлять правильность различных прогнозов или серий прогнозов.

Для вероятностных прогнозов измерителем правильности (или точности) может служить количество информации, недополученное в данном прогнозе или серии. Пусть возможным дискретным исходам K прогнозируемого события были приписаны вероятности $P_K > 0$. Предположим, что истинным оказался исход K_1 . Тогда величина $-\lg P_{K_1}$ и является недополученным количеством информации, обращающимся в нуль при $P_{K_1} = 1$ и неограниченно, хотя и медленно, возрастающим с приближением P_{K_1} к нулю.

Сопоставляя ряд способов (или систем) прогнозирования $1, 2, \dots, j, \dots$, примененных к событиям $1, 2, \dots, i, \dots, n$, составим для каждого способа сумму

недополученной информации: $S_j = - \sum_{i=1}^n \lg P_{jiK_1}$. Тогда

S_j служит измерителем точности серии прогнозов при способе j : чем меньше значения, приписанные истинным исходам, тем хуже прогноз и тем S_j больше. Этот измеритель аддитивен и прост в вычислении. Его нетрудно обобщить на случай непрерывного распределения вероятностей, если результатом проверки прогноза служит не точное, а приближенное значение неизвестной величины.

Приведенный критерий может быть полезен для сопоставления различных разновидностей способа

прогнозирования, прогностической квалификации различных специалистов, различных научных школ и т. д. Критерий может применяться и в случае словесно выраженных степеней уверенности в том или ином исходе дискретного распределения: по частотам ошибок оцениваются вероятности, скрывающиеся за утверждениями «возможно», «маловероятно», «невозможно», «абсолютно невозможно».

Ознакомление эксперта с этой статистикой ошибок позволяет в дальнейшем вместо словесного выражения степени уверенности получать от него устойчивые численные оценки вероятностей, в то время как человек, не прошедший подобной практики, легко меняет называемые вероятности одного и того же исхода с 10% до 0,01% и обратно.

Если общество заинтересовано в изучении состояния своей науки, в выяснении того, насколько наука способна оценивать правдоподобие гипотез на основе косвенной информации, то оно должно чаще практиковать анкетирование ученых своих научных коллективов, чаще проводить испытания «разумных» машин различных типов, начиная с программ для распознавания образов.

Особенно интересно проводить анкеты перед выяснением результатов критических экспериментов, наблюдений принципиально новых явлений, путешествий в новые районы своей планеты, своей планетной системы и на новые планетные системы. Лучше познаёт себя та культура, которая будет достаточно полно регистрировать свой процесс изучения внешнего мира.

Поскольку в нашей земной науке подобные исследования еще не проведены, да и психологические механизмы прогнозирования изучены весьма слабо, трудно дать четкое и завершенное сравнение разных методов прогноза. В прогнозе направлений научно-технического прогресса бросается в глаза беспомощность и косность мысли многих ученых и фантастов, когда речь заходит о сколько-нибудь далеком будущем [36]. Практические неудачи здесь соседствуют с теоретическими неясностями: как в психологии, так

и в кибернетике весьма слабо разработанной является проблема выдвижения новой гипотезы, в противоположность проблеме выбора гипотезы из заданной совокупности [37]. Между тем подлинно творческую деятельность человека не удастся эффективно уложить в представление об акте выбора, чем, кстати, пытаются объяснить относительные неудачи работы по решению задач машиной, «эвристическому программированию» и т. д. [38]. Только при охвате очень широкой области, при работе с весьма общими понятиями вроде «материалы» и «энергия», «управление» и «информация» можно было бы на основе перебора заблаговременно предвидеть логическую возможность, скажем, изобретения вычислительных машин или квантовых генераторов.

Прогноз темпов научно-технического прогресса.

Этот частный вид прогнозов исследован лучше, чем прогнозы направления. Примеры перспективного прогноза темпов и сроков имеются во многих исторических эпохах.

Среди них встречаются и очень резкие промахи, и отдельные совпадения. Основными причинами промахов являются, с одной стороны, недооценка трудностей исследования или изобретения, а с другой, — недооценка ускорения развития науки и техники. Например, десяток-полтора лет назад ученые плохо представляли трудности машинного перевода или теоретико-информационного объяснения истории языка, недооценивали болезненные последствия пересадки органов или химических мер борьбы с вредителями леса. В 1900 г. Герберт Уэллс предсказывал использование атомной энергии уже в начале века [39]. Зато появления самолетов он ожидал лишь в середине века. Лет семьдесят — сто назад признавались невозможными (или возможными лишь через миллионы лет) многие технические достижения, ставшие реальностью в наши дни [36, 40].

Простейшим способом прогноза роста какого-либо показателя является экстраполяция на будущее теоретической функции, достаточно близкой к эмпи-

рическим данным роста показателя в прошлом. Степень близости может определяться, например, методом наименьших квадратов. Этот метод способен давать весьма различные результаты при различных преобразованиях шкал по осям абсцисс и ординат, причем наилучшими для задач прогноза следует считать такие преобразования, которые делают колебания функции на всем диапазоне времени статистически однородными.

Слабо разработанным является вопрос о предпочтении одних типов функций другим при использовании метода наименьших квадратов и об определении числа изменяемых параметров при заданном типе функций.

Одни авторы работают только с экспонентой и логистической кривой [20, 21], другие прямо отмечают [10, 16] или подразумевают, судя по шкалам на их графиках [17, 41] применимость к отдельным периодам развития гиперболы, экспоненты от экспоненты, дробно-линейной функции и т. д.

Формальная экстраполяция динамического ряда неизбежно приведет к ошибкам или к чрезмерной неопределенности прогноза, если мы не будем учитывать содержательную сторону дела: социально-экономические изменения, возможные открытия и изобретения и т. п. Указанное обстоятельство особенно сильно сокращает сроки правомерности конкретных прогнозов (развитие отдельных технических средств, узких отраслей науки и т. д.; ср. § 2), в то время как общие закономерности динамики изменения темпов развития сохраняются дольше [20].

Психологически многие ошибки прогнозов срока можно связать с тем, что не срок определяется темпом, а темп сроком. Ученый или писатель-фантаст оценивал прогресс предстоящего столетия по предыдущему столетию и предстоящего тысячелетия также по прошедшему тысячелетию и т. д. Но последние лет 300—400 как раз характерны более высоким темпом прогресса, чем предшествующее тысячелетие. По ряду существенных показателей темп этот непрерывно рос и на протяжении новой истории. Таким образом, темп

прогресса будущего по этому прогнозу оказывался замедляющимся, ход развития в будущем — зеркальным отображением развития в прошлом, а сроки — сильно запаздывающими по сравнению с ускоряющимся ходом реальной истории техники. Если, например, задана последовательность дат 1500, 1800, 1900, 1950, то ее симметричным, зеркальным продолжением в будущее служит последовательность 1950, 2000, 2100, 2400, в то время как сохранение темпов ускорения предполагало бы экстраполяцию вроде 1950, 1980, 2000, 2015.

Нет основания категорически настаивать на экстраполяции последнего типа, но предыдущая экстраполяция является предельным субъективизмом, признанием своей эпохи за центр симметрии кривой развития во времени.

Этот субъективизм проявляется даже на первой странице обложки книги Артура Кларка [36], где друг под другом поставлены числа 1800, 1900, 1950, 2000 и 2100. Критика консервативных прогнозов ограничивается в этой книге психологическим уровнем и уровнем конкретной истории техники. Будущее развитие техники по собственному прогнозу Кларка не является, правда, замедленным, но и особого ускорения в его оценках также усмотреть нельзя. Впрочем, следует оговорить, что признание концепции сокращения промежутков удвоения заметно проявляется в прогнозе темпов развития. Между тем сколько-нибудь полные прогнозы далекого будущего (см., например, [40]) являются именно прогнозами сроков, как и прогноз Кларка.

Первоочередность рассмотрения темпов, как условие правильности прогноза, хорошо показал Стайн [16], чрезмерно, однако, увлекшийся формальной стороной экстраполяции без содержательного анализа. За шесть лет, прошедшие после опубликования его статьи, многие из рассмотренных им показателей задержались в своем ускорении (например, скорости транспорта). Их гиперболический рост сменился экспоненциальным или еще более медленным. Указанная поправка не зачеркивает основной мысли Стайна о

резком отставании предсказанных фантастами темпов от действительных, потому что график, соответствующий представлению фантастов о будущем, скорее является выпуклой вверх кривой, чем прямой, как считал Стайн.

При прогнозировании темпов развития цивилизации, вышедшей в космос, заселившей целый район галактики и продолжающей быстро развиваться, возникнут специфические трудности, связанные с тем, что никакая информация не может распространяться быстрее скорости света и одна часть цивилизации будет с большим запозданием узнавать о прогрессе другой. В этом случае общение разных частей одной цивилизации будет в некоторых отношениях напоминать общение разных цивилизаций между собой.

О прогнозе темпов развития земной цивилизации.

Рассмотрим некоторые из факторов, упоминаемых в нашу эпоху в связи с прогнозами будущего развития человечества. Прежде всего отметим, что прогноз темпов развития (если он дается в форме средней кривой, а не целого множества кривых с вероятностной мерой на этом множестве) обычно строится для случая нормальных условий: не разразится ядерная война, не будут занесены смертоносные микробы с других планет и т. д. Вероятность подобной глобальной катастрофы многим авторам представляется более высокой для нашего времени, чем для прошлых веков (когда имели место случаи если не гибели, то значительной задержки развития цивилизации в целом, что произошло, например, при падении Древнего Рима). Но если мы даже согласимся с тем, что вероятность подобной катастрофы стала выше в расчете на один год или на длительность жизни одного человека, то из этого не следует, что эта вероятность стала выше в расчете на длительность одной эпохи, так как длительности эпох заметно сократились.

С другой стороны, резко упала вероятность гибели человечества за счет какого-либо стихийного бедствия, вероятность гибели, вызванная внешними причинами, а не собственной активностью цивилизации. Вряд ли

человечество не сумеет себя сохранить при резком понижении солнечной радиации или при близкой вспышке сверхновой звезды. Поскольку вероятность подобных событий ничтожна, мы можем считать гибель цивилизации от внешних стихийных причин практически исключенной для нашего уровня техники и не следует ставить в один ряд причины стихийные и внутренние, как это делают некоторые авторы [42].

В литературе наиболее часто можно встретить прогнозы развития, основанные на представлении об экспоненциальном росте того или иного показателя, например, расхода энергии [10] или роста науки и техники в целом [30]. Следует ли основываться на подобных представлениях в нашу эпоху, в конце 60-х годов XX в.? Против этого следует выдвинуть ряд возражений.

Во-первых, признавая ту или иную количественную закономерность скорости развития, мы не в праве считать, что выразителем ее все время будет оставаться один и тот же показатель, а не будет происходить смены ведущих, важнейших характерных показателей для развития всей цивилизации, ее отдельного аспекта, ее отдельной отрасли.

Во-вторых, как это уже указывалось выше, многие показатели до последнего времени ускоряли темп своего развития. Нет оснований думать, что это ускорение прекратится именно сейчас. Наоборот, более логично предположить, что именно в ближайшие годы или десятилетия произойдет ряд скачков, которые еще более ускорят темп развития земной культуры. После второй мировой войны ведущие страны мира расходовали значительные средства на такие отрасли, как конструирование вычислительных машин, программированное обучение (и другие новые учебные методы), машинный перевод, исследования элементарных частиц, космические исследования. Все эти затраты дали пока лишь небольшую часть того эффекта, которого от них ждали. Приходится слышать сомнения в том, что мы стоим на пороге скачка в некоторых из этих отраслей. Правомерны ли такие сомнения? Ясно, что

космические исследования быстро движутся вперед и уже приносят весомые научные плоды.

Прекрасные результаты обнаруживаются и у новых методов обучения, но пока в экспериментальном масштабе. Более далекой от разрешения представляется проблема машинного перевода, но работы в этой области позволили четко увидеть и сформулировать многие вопросы, связанные с восприятием письменной речи, решение которых сулит крупные продвижения в понимании психической деятельности человека, а следовательно, в решении многих задач обучения, организации творческого труда, в конструировании «умных машин».

Важным фактором ускорения темпов развития цивилизации, несомненно, послужит и социальное преобразование мира, переход новых стран на социалистический путь развития.

Существенному ускорению темпов прогресса может содействовать и расшифровка сообщений внеземных разумных систем. Основаниями для надежды на прием или перехват таких сообщений в самые ближайшие годы могут служить бурный расцвет радиоастрономии и установка оптических телескопов на небесных телах без атмосферы, например, на Луне, а также обнаружение возможных следов техники на астероидах и спутниках, не подверженных выветриванию.

Третьим возражением против экспоненциальных представлений служит ряд обстоятельств, которые в более далеком будущем (XXI в. или конец XX в.), вероятно, замедлят развитие земной цивилизации. Среди этих причин нередко отмечают чисто земные [21], связанные с трудностями ориентировки в растущем потоке научной информации, с отрицательными сторонами специализации ученых и т. п. Весьма фундаментальной причиной также служит большое различие (десятки тысяч раз) между расстояниями до последних планет Солнечной системы и до ближайших звезд. Это различие, по мнению автора главы, означает, что после изучения и освоения Солнечной системы некоторое время изучать и осваивать будет нечего.

Нечто сходное имело место в прошлом, когда к концу XIX в. на карте мира исчезли белые пятна, а для серьезного исследования дна океанов еще не было технической базы. Это послужило одной из тех причин, которые задержали быстро ускорявшийся ранее рост темпов научно-технического прогресса и способствовали популярности экспоненциальной модели. Говоря о быстром ускорении темпов до XX в., автор имеет в виду быстроту ускорения, разумеется, не по годам, а по промежуткам удвоения ведущих показателей (см. § 4).

Трудно судить и о том, как скажется на темпах нашего развития ожидаемый в еще более далеком будущем прямой контакт с представителями иных цивилизаций или с созданными там автоматами. В научной фантастике можно встретить весьма различные представления о возможностях воздействия на темп научного и социального прогресса отстающего общества со стороны более передового. Например, в повести братьев Стругацких «Трудно быть богом» [43] последовательно обосновывается идея невозможности изменения темпов за счет такого космического вмешательства, при котором отстающее общество не теряет своей индивидуальности. Нет никаких гарантий, что понятие вражды и дружбы, обучения и обмена информацией, наблюдения и эксперимента не окажутся антропоморфными, что они смогут отразить все разнообразие и сложность возможных отношений между двумя цивилизациями. А если когда-нибудь наша цивилизация займет свое место в бесконечной системе связанных между собой действительно дружественных цивилизаций, то не окажется ли антропоморфным и само понятие темпов развития?

Если это случится, то когда? Неизвестно, произойдет ли это через месяц или через миллионы лет. От идеи уникальности Земли человечество шло к представлению об обитаемости всех небесных тел [44], а сейчас оно пришло к осознанию своей неосведомленности в этом вопросе. Одни предполагают вероятным существование в Галактике цивилизованных систем. Другие авторы подчеркивают, что время

освоения всей Галактики цивилизацией вряд ли составит много миллионов лет и что оно во всяком случае гораздо короче всей истории Галактики, так что в случае возникновения иной цивилизации она бы с большей вероятностью появилась раньше, а не точно в то время, что и на Земле, и успела бы дать о себе знать. Слабым местом этой точки зрения является принятие как очевидных следующих факторов: скорость развития ни у какой цивилизации не может быть во многие тысячи раз меньше, чем у нашей *); всякая цивилизация сможет выйти в космическое пространство; всякая цивилизация попытается это сделать. Молчаливое допущение этих посылок — не более, чем дань антропоморфизму, и у нас пока нет оснований считать очень маловероятным существование иных разумных существ в нашей Галактике или в соседних галактиках.

ЛИТЕРАТУРА

1. Macgovan R., Ordway F., *Intellegence in the Universe*, Prentice Hall, N. Y., 1966.
2. Schmeck H., *Semi-artificial Man*, London, G. G. Harrap and Co, 1965.
3. Поршнев Б. Ф., Рецензия, «Вестник древней истории», № 1, 1963.
4. Sanderson I., *The dynasty of Abu*, N. Y., Knopf, 1962.
5. Фейнберг Е. Л., Обыкновенное и необычное. «Новый мир», № 8, 1965.
6. Berrill N., *Worlds apart*, London, 1965.
7. Кондратов А. М., Звуки и знаки. «Знание», 1966.
8. Фрумкина Р. М., Статистические методы изучения лексики, «Наука», 1964.
9. Norwood V. G. Ch., *Man alone*, London, Boardman, 1958.
10. Шкловский И. С., *Вселенная, жизнь, разум*, изд. 2-е, «Наука», 1965.

*) Мнение о сходстве темпов развития всех цивилизаций хорошо сформулировано Стругацкими [43]: «Мы не физики, мы историки. У нас единица времени не секунда, а век...». В действительности история не имеет постоянных единиц времени даже в рамках одной и той же цивилизации.

11. Поршнев Б. Ф., Возможна ли сейчас научная революция в приматологии? «Вопросы философии», № 1, 1966.
12. Белькович О. М., Клейненберг С. Е., Яблоков А. В., Загадка океана, «Молодая Гвардия», 1965.
13. Шовен Р., От пчелы до гориллы, «Мир», 1965.
14. Hauser P., (ed.), The study of urbanisation, N. Y., 1965.
15. Rousseau P., Histoire de la vitesse, Paris, Press Universitaire de France, 1946.
16. Stine G. H., Science fiction is too conservative, «Analog Science», N. Y. 67, № 3, 1961.
17. Meyer F., L'acceleration evolutive, Paris, 1946.
18. Bruner J., Toward a theory of instruction, Belknap, 1966.
19. Zeman J., Poznání a informace, Praha, 1962.
20. Price O., Little science, big science, N. Y., Columbia, 1963.
21. Добров Г. М., Наука о науке, Киев, «Наукова думка», 1966.
22. Налимов В. В., Чернова Н. А., Статистические методы планирования экстремальных экспериментов, «Наука», 1965.
23. Лаврентьев М. А., Беречь время ученого! «Организация и эффективность научных исследований», Новосибирск, «Наука», 1965.
24. Price D., Regular patterns in the organization of Science, «Organon», Warszawa, № 2, 1965.
25. Стройк Д. Я., Краткий курс истории математики, «Наука», 1964.
26. Spearman C., Psychology down the ages, Vol. 2, London, Macmillan, 1937.
27. Walden P., Chronologische Übersichtstabellen, Berlin, Springer, 1952.
28. Sirks M. J., Conway Z., The evolution of biology, N. Y., 1964.
29. Haus A., Sprachwissenschaft — der Gang ihrer Entwicklung von der Antike bis zur Gegenwart, Freiburg, 1955.
30. Ляпунов А. А., Тезисы доклада на конференции по философским вопросам кибернетики, М., 1962.
31. Эшби Р. У., Конструкция мозга, ИЛ, 1959.
32. Бир Ст., Кибернетика и управление производством, изд. 2-е, «Наука», 1965.
33. Крушинский Л. В., Экстраполяционные рефлексy как элементарная основа рассудочной деятельности у животных, ДАН СССР 121, № 4, 1958, 762—765.
34. Рубинштейн С. Л., О мышлении и путях его исследования, Изд. АН СССР, 1958.
35. Лем Ст., Две эволюции — сходства и различия, «Наука и техника», Рига, № 8, 1965.
36. Кларк А., Черты будущего, «Мир», 1966.
37. Пушкин В. Н., Оперативное мышление в больших системах, «Энергия» 1965.

38. Пушкин В. Н., Эвристика и кибернетика, «Знание», 1965.
39. Уэллс Г. Дж., Предсказания о судьбах, ожидающих человечество в XX столетии, Спб., тип. Монтвида, 1903.
40. Рюриков Ю., Через 100 и 1000 лет, «Искусство», 1961.
41. Перельман Р. А., Цели и пути освоения космоса, «Наука», 1967.
42. Зигель Ф. Ю., Жизнь во Вселенной, Минск, «Наука и техника», 1966.
43. Библиотека современной фантастики, т. 7, «Молодая гвардия», 1966.
44. Фламмарion К., Популярная астрономия, Детиздат, 1939.

Некоторые общие вопросы проблемы внеземных цивилизаций

§ 1. Введение

В настоящее время довольно четко обрисовывается возможность нового, научного подхода к изучению проблемы существования и развития разумной жизни во Вселенной. Нет надобности доказывать научную и мировоззренческую важность исследований в этой области. Вопрос о существовании внеземных цивилизаций в той или иной форме ставился на всех этапах истории развития науки. Задача эта фундаментальна и сложна, затрагивает множество отраслей и областей науки, поэтому на разных уровнях знания ее можно рассматривать в различных аспектах и разрешать в той или иной степени.

На первом этапе развития идеи множественности обитаемых миров ее объективной материалистической основой являлось стремление осознать закономерность появления человека, человеческого общества, отбросить теологические тезисы об уникальности живого и разумного в природе и о существовании принципиального разрыва между «духом» и материей.

В дальнейшем гипотезы о существовании других разумных миров появлялись в связи с определенными успехами в научно-техническом прогрессе. Раздвижение границ наблюдаемой Вселенной благодаря бурному развитию оптической астрономии давало обильную пищу многочисленным попыткам обнаружить «экологические ниши», подходящие для существования жизни, подобной земной, на различных космических объектах.

Возможно, с точки зрения современных знаний гипотезы эти представляются весьма наивными и мало-доказательными, но здравая методологическая идея, лежащая в основе таких поисков, привела в дальнейшем к созданию астробиологии и современной научной дисциплины — экзобиологии, в задачу которой, в частности, входит исследование возможности существования в космосе условий, подходящих для белковой формы жизни.

С другой стороны, проблема внеземных цивилизаций неоднократно ставилась в связи с возможным прогнозом дальнейшего развития человеческого общества. Эта тема волновала многих ученых, философов, социологов и писателей. В частности, необходимо отметить высказывания К. Э. Циолковского, который одним из первых указал на возможность «энергетического кризиса» при дальнейшем росте производства, науки и техники, обратил внимание на практическую необходимость исследования и овладения космическим пространством. Эти идеи, как известно, являются основополагающими в подавляющем числе современных работ, посвященных проблеме космических цивилизаций.

На первый взгляд представляется естественным и даже единственно возможным рассматривать и решать проблему внеземных цивилизаций, используя упомянутые выше две методологические идеи, которые можно условно назвать «экзобиологической» и «прогнозной». В самом деле, наши суждения будут опираться тогда на хорошо известные знания о белковой форме жизни (единственно наблюдаемой) и о закономерностях развития человеческого общества (также единственно известной нам форме цивилизации).

Однако существует возможность более широкого подхода к проблеме внеземных цивилизаций. При таком подходе вопрос об их существовании рассматривается как часть еще более общей и сложной проблемы, включающей в себя исследование генеральных принципов строения, функционирования и эволюции сложных больших систем, частным случаем которых являются биологическая эволюция и человеческая ци-

вилизация в целом. Подобное рассмотрение несет в себе принципиально новые методологические черты и непосредственно связано с возникновением теории сложных систем, которая успешно развивается в последние годы на основе идеи теоретической и технической кибернетики. Этот взгляд не исключает «экзобиологических» и «прогнозных» инструментов исследования, но позволяет оценить объективные возможности их вклада в решение проблемы космических цивилизаций.

Утверждение о необходимости системного подхода к изучению возможностей существования внеземных цивилизаций отчетливо проводится в работе С. Лема [1], к которой мы будем неоднократно в дальнейшем обращаться.

Эту постановку задачи можно сформулировать и несколько иначе: нельзя ли попытаться уже сейчас рассмотреть основные принципы проблемы внеземных цивилизаций с тем, чтобы по возможности построить некую «общую теорию цивилизации», которая опиралась бы на данные современной науки. С нашей точки зрения, такая возможность представляется реальной. Эта возможность позволяет четко очертить тот круг вопросов, которые принадлежат проблеме внеземных цивилизаций и определить место ее и в «общей теории цивилизации» и место самой «общей теории» в классификации современных научных дисциплин.

Значение идеи системного изучения отчетливо проявляется на примерах тех принципиальных трудностей, которые возникают в частных «астрономическом» и «радиоастрономическом» аспектах проблемы внеземных цивилизаций.

§ 2. Методология «радиоастрономического» аспекта проблемы. «Энергетическая» гипотеза

Успехи радиоастрономии привели к возможности постановки вопроса об обнаружении в радиодиапазоне космических сигналов искусственного происхождения. На данном этапе изложения нет надобности в

уточнении понятий «сигнал», «искусственность происхождения». Во всяком случае, из дальнейшего будет ясно, что в каждой конкретной работе подразумевается под этим термином, а возможные уточнения будут сделаны в третьем разделе главы. Идея, положенные в основу большей части исследований, посвященных «радиоастрономической части» проблемы внеземных цивилизаций, весьма близки друг к другу. Для их обзора достаточно поэтому рассмотреть те работы, которые появились за последние несколько лет. Это прежде всего книга И. С. Шкловского [2], статья Н. С. Кардашева [3], а также статьи в сборниках трудов «Внеземные цивилизации» [4] и «Межзвездная связь» [5].

Возможность поиска сигналов от внеземных цивилизаций основывается в этих работах на следующих соображениях:

1. Радиодиапазон является оптимальным для передачи информативного сигнала на сверхдальние расстояния.

2. Космические цивилизации в процессе развития увеличивают свою энергоспособность.

3. На определенном этапе своего развития цивилизации с необходимостью посылают информативные сигналы в космическое пространство.

4. Принятые сигналы могут быть расшифрованы.

Справедливость этих предпосылок представляется неравноценной. Первое положение кажется наименее дискуссионным. Оно полностью соответствует уровню современных знаний. Мы не знаем пока, или по крайней мере не видим путей к практическому использованию более оптимальных способов связи на космических расстояниях. Все другие предположения носили бы схоластический или спекулятивный характер [2].

Второе положение очень сложно по своей структуре. Во-первых, дается определенный прогноз развития человеческой цивилизации. Считается, что дальнейший научно-технический прогресс будет неуклонно идти по пути «энергетического развития» и освоения все больших областей космоса. Во-вторых, утверждается универсальность этого способа развития, по

меньшей мере для широкого класса «антропоморфных» цивилизаций во Вселенной *).

Проблема возможности прогнозирования дальнейшего хода научно-технологического прогресса широко обсуждается в своевременной научной печати [6, 7, 8]. Все более утверждается мнение, что необходим целостный, систематический подход к проблемам научно-технического прогресса, и специальный научный аппарат для рассмотрения таких проблем. Исследования должны вестись не только в плане социально-экономического анализа факторов научно-технического прогресса, но и с точки зрения общего хода развития, общих внутренних тенденций эволюции науки и техники. Все эти требования привели к заложению фундамента новой научной дисциплины — «науки о науке» или «науковедении» (см., например, [8]). В этой отрасли знания широко используются достижения теории сложных систем.

До сих пор наиболее яркими примерами составления «долгосрочных прогнозов» дальнейшего развития человеческой цивилизации были научно-фантастические и социально-фантастические произведения. Анализ методов, используемых в наиболее логических и продуманных работах такого рода, показывает весьма интересные проявления своеобразной «психологии прогнозов». Основной чертой этих методов является стремление к линейной экстраполяции на будущее тех возможностей, которые реализуются или потенциально зреют на современном уровне. Этот метод С. Лем называет «ортоэволюционным прогнозом». Как правило, проверка временем показывает, что прогнозируемые картины даже не очень далекого будущего весьма отличаются от действительности. Диалектика развития, как известно, включает количественные изменения как один (и не самый важный) этап в процессе

*) Последняя фраза требует некоторого разъяснения. Как будет видно из дальнейшего, соответствующие утверждения [3] можно понимать двояко. С одной стороны, можно подумать, что все космические цивилизации считаются антропоморфными. С другой стороны, возможна интерпретация, что рассмотрение ограничивается только цивилизациями антропоморфного типа.

эволюции. На определенных периодах развития человеческого общества открытие новых возможностей (например, новых видов энергии, материалов) скачком качественно меняло способы производства, тем самым влияя на образ жизни данного поколения и на дальнейшее развитие общества. Недостатком «ортоэволюционного» метода является и недопустимая альтернативность предсказания. При этом абсолютизируется одна из сторон наблюдаемого явления, которая антагонистически противопоставляется всем другим возможностям. Развитию же свойственно взаимодействие полярностей, являющейся одной из основ динамики движения вперед.

Слабой стороной «энергетической» гипотезы развития земной цивилизации является именно ее «ортоэволюционный» характер. «Молчаливо подразумевается, что увеличение темпа технического прогресса, который мы наблюдаем на Земле, по крайней мере в течение последних 200 лет является динамическим непрерывным процессом, который задержать могут только деструктивные причины («вырождение», «самоубийство» цивилизации)» ([1], стр. 85).

Положения «энергетической» гипотезы, безусловно, опираются на известные факты в истории развития науки и техники за последнее время. Основным мотивом, стимулирующим эту идею, является стремление нащупать пути предотвращения угрозы намечающихся «энергетического» и «демографического» кризисов, реальное наличие которых отмечается во многих источниках [6, 7]. Следует отметить, что идея о необходимости непрерывного овладения космическим пространством прямо следует из основных идейных посылок «энергетической» гипотезы и не является дополнительным предположением в ее рамках. Однако возможность преодоления указанных выше кризисов представляется в «энергетической» гипотезе в виде чисто количественного роста энергонакопления. Поиски энергии и «жизненного пространства» возводятся в ранг вечных проблем. Постулируется неизменность характеристик динамического процесса развития человечества на неопределенно долгий промежуток вре-

мени (фактически все гипотезы этого класса утверждают такой путь на все время существования цивилизации) *). Но это ведет в рамках энергетической гипотезы к выводу, что поиск дополнительных источников энергии и свободного пространства ставит человечество на путь постоянного балансирования «на острие ножа», так как малейшая задержка в освоении дополнительных ресурсов может привести к катастрофическим последствиям. Особое значение поэтому приобретает для этой гипотезы обсуждение проектов «космоинженерии» (типа проекта Дайсона), которые позволили бы доказать реальность построения моделей, обеспечивающих разрешение энергетического и демографического кризисов. Но беспредельная «космическая диффузия» и энергонакопление увеличивают возможность проявления других кризисных явлений, которые намечаются уже на современном этапе развития. Мы имеем в виду «информационный» и «организационный» кризисы, рассмотрение которых в энергетической гипотезе игнорируется. В критике проектов «космоинженерии» указывается на возникновение труднопреодолимых препятствий такого рода. С другой стороны, подробное рассмотрение возможностей преодоления в развитии цивилизации таких эффектов, как, например, информационный кризис, приводит к гипотезам, качественно отличным от «энергетической» [1].

*) Следует отметить, что, как и основоположник идеи проникновения человечества в космос К. Э. Циолковский, советские авторы связывают эту мысль прежде всего с оптимистическим этически-философским утверждением безграничных возможностей человеческого разума. В то же время в ряде работ западных ученых заметна тенденция истолковывать продвижение в космос как «исход» от трудностей жизни и противоречий современного общества. Показательна в этом смысле статья Р. Симона [9]. Автор в «американском» стиле рекламирует космическое пространство как великолепное место для развития частнособственнического индивидуализма, которому «слишком тесно» на Земле. Наибольшим достоинством космоса, по мысли Симона, является его «емкость за счет трехмерности». Поэтому «излишки человечества» могут уходить в межпланетную жизнь, дабы избежать тревожащих неустroенного мира и тем самым способствовать «гармонизации» человеческого общества.

Резюмируя выше сказанное, можно отметить, что «энергетическая» гипотеза является по крайней мере одной из возможных, опирающихся на исследование некоторых тенденций современного развития человечества.

Рассмотрим теперь вопрос об универсальности «энергетического» способа развития для других космических цивилизаций. Единственно наблюдаемая нами форма существования земной цивилизации развивается «технологическим» путем *). Теоретически можно предполагать, однако, существование «нетехнологического» пути развития сложной живой системы [1]. Примером может служить биологическая эволюция, идущая по пути пластического приспособления к среде. Система, развивающаяся подобным образом, может достигать очень высокой степени организованности. Но с нашей точки зрения она не является «разумной». Это правильно применительно к наблюдаемой нами биологической эволюции. Но можно представить себе направленную деятельность типа спланированной автоэволюции, изменения биологического вида для лучшего приспособления и деятельности в окружающем пространстве. Такая цивилизация представлялась бы весьма странной с точки зрения земной. Однако не исключено, что это следствие антропоцентрического образа мышления, отрицающего возможность существования других форм разума.

Во всяком случае, определение «разумной» системы и степени ее принадлежности к «цивилизациям» требует дополнительного обсуждения, к которому мы еще вернемся.

Если же говорить о «технологических», «антропоморфных» цивилизациях во Вселенной, то «энергети-

*) Человечество пошло по пути активного преобразования природы, создания подходящих условий для поддержания параметров, нужных для существования вида. Деятельность эта далека от «гармонической». Вольно или невольно в своем активном вмешательстве в природу человек нарушает естественный баланс условий, создавая искусственную «сверхнизкоэнтропийную» среду.

ческая» гипотеза утверждает для них универсальность общего пути развития.

Перейдем к рассмотрению третьей предпосылки гипотез о возможности «радиоастрономического» обнаружения сигналов внеземной цивилизации. С одной стороны, третье положение является необходимым для логической замкнутости постановки задачи. Если цивилизации не передают сигналов в радиодиапазоне, то их нельзя и обнаружить с помощью радионаблюдений *). С другой стороны, возможно попытаться логически обосновать третье положение.

Естественно рассмотреть случай, когда цивилизация намеренно передает информацию в окружающее космическое пространство. Для обоснования подобного рода деятельности космических цивилизаций приводятся обычно две группы аргументов:

а) считается, что посылка информативных сигналов связана с экспериментальной, исследовательской деятельностью развитой «технологической» цивилизации на поиск других подобных себе цивилизаций [2];

б) предполагается, что стремление к передаче информативных сведений непременно вытекает из психологически-этических тенденций развития высокоорганизованной цивилизации [3].

Аргументы первой группы, пожалуй, естественно вытекают из предположения об «энергетическом» развитии «технологической» цивилизации. Цивилизации «технологического» типа нуждаются во всемерном изучении всех явлений в окружающем их мире. Наиболее логичным инструментом исследования подобных себе цивилизаций будет установление канала информативной связи между цивилизациями. «Задержка» двусторонней связи из-за огромных расстояний в космическом пространстве в такой трактовке не представляется принципиальной.

*) Это заключение полностью справедливо в «антропоцентрической» постановке задачи. Вообще же говоря, следует рассмотреть пути исследования цивилизаций, не посылающих специальных сигналов о своем существовании, но «проявляющихся» в виде специфических форм «поведения» (см. § 3).

Вторая группа аргументов также приводится при доказательстве существования «бескорыстной» посылки цивилизацией информации неопределенному абоненту. Это положение существенно при расчетах о вероятности обнаружения искусственных сигналов от сверхцивилизации [3].

Представляется, что любые функциональные действия любой сложной системы можно обосновать только определенной необходимостью для целенаправленного существования этой системы. В этом смысле этические и психологические моменты являются следствием более глубоких «принципов действия» цивилизации. Поэтому обоснования гуманизма или, наоборот, «агрессивности» цивилизации в той форме, в которой они часто делаются, являются определенным новым добавлением к ряду уже перечисленных допущений, базирующихся на экстраполяции современных понятий и представлений.

Рассмотрим теперь вопрос о возможности расшифровки принятого сигнала. Логика построения предыдущих попыток приводит к выводу, что «антропоморфная», «технологическая» цивилизация должна посылать информацию в виде семантической, языковой системы, переданной в определенной кодовой форме. Многочисленны попытки построить по возможности формализованную, языковую систему, с помощью которой можно было бы передать систему антропоморфных понятий. Таковой является система Lincos Фройдентала [10]. Возможность расшифровки связывается при этом подходе с идеей общности понятий и сходства систем знаний у цивилизаций. Это предположение не является логически вытекающим из предыдущих. На теоретическую возможность существования у различных цивилизаций систем знания, со значительно отличающимися даже элементарными понятиями, указал А. В. Гладкий [11]. В таких случаях расшифровка информации должна наталкиваться на серьезные трудности.

Для целей нашего изложения следует подчеркнуть, что характерным для «радиоастрономических» гипотез поиска сигналов внеземных цивилизаций является

стремление ориентироваться и в проблеме расшифровки сигнала на антропоморфные признаки.

Таким образом, из анализа предположений, лежащих в основе [3] и сходной группы гипотез следует, что предметом поиска являются сигналы от космических цивилизаций, которые во всех своих основных проявлениях сходны с земной цивилизацией (вплоть до деталей прогнозируемого развития).

Ясно, однако, что придерживаться во всей полноте «чистых» принципов типа предложенных в [3] было бы весьма затруднительно. В конце концов, рассмотренные гипотезы являются рабочими и сами их создатели непрерывно ищут новые возможности для развития идей и методов поиска сигналов от внеземных цивилизаций. В этом смысле предлагаемое критическое рассмотрение не ставит своей целью «ниспровержение» высказываемых положений, а претендует лишь на методологическое выяснение принципов, заложенных в этих гипотезах.

Переходной ступенью к новым конструктивным возможностям исследования служит обсуждение вопроса о критериях искусственности происхождения сигналов. И. С. Шкловский предложил концепцию «космического чуда» [2]. Под «космическим чудом» понимается наблюдаемое с помощью астрономических методов проявление в космических масштабах деятельности разумных существ. В плане радиоастрономических поисков сигналов от космических цивилизаций вопрос об обнаружении «космического чуда» звучит как проблема структуры «позывных» — сигнала, который несет информацию лишь о том, что он искусственного происхождения (см. гл. III настоящего сборника). Естественно встает вопрос об однозначности интерпретации подобного явления. Так, например, расшифровка некоторой семантической системы сигналов, принятых в радиодиапазоне от космического объекта, несомненно, свидетельствовала бы о наличии передающей цивилизации. В этом случае критерием искусственности служил бы сам факт возможности расшифровки информации, заключенной в сигнале. Однако ясно, что это был бы исключительно удачный,

редкий случай. Как же быть в других ситуациях? И. С. Шкловский пишет: «...мы чисто объективно, путем наблюдений, можем заметить весьма удаленные от нас сверхцивилизации потому, что связанные с ними объекты не подчиняются закономерностям неживой материи или приводят к характеристикам удивительным и даже неестественным» [2].

Слова «не подчиняются закономерностям неживой материи», «удивительные характеристики» носят неопределенный характер. Логика развития естественных наук и применение их к изучению астрономических объектов накладывает определенные жесткие ограничения на возможность истолкования даже весьма «странных» явлений в космосе, как проявлений «разумной» деятельности. Отсутствие четких критериев, позволяющих отличить продукт деятельности цивилизации от естественных космических объектов, открывает дорогу ненаучным спекуляциям на тему об искусственном происхождении некоторых явлений (как, например, печально известная история с Тунгусским падением), а, с другой стороны, накладывает ненужный «запрет» на открытие и исследование в астрономии принципиально новых «естественных» явлений, обогащающих наши земные знания в области физики и других областей науки.

Как можно решать проблему критериев искусственности с точки зрения системного подхода, будет изложено в § 3. Пока же рассмотрим еще один вопрос, оживленно обсуждающийся в научной литературе в связи с проблемой внеземных цивилизаций.

Важным экспериментальным фактом является отсутствие наблюдаемых «космических чудес». Оставим пока в стороне соображение о том, что причиной является ненадежность критериев «искусственности» и встанем на точку зрения «антропоморфных» гипотез. В соответствии с постулатами «энергетической» гипотезы все цивилизации должны проходить «технологический» этап в своем развитии, связанный с освоением и преобразованием окружающего космического пространства, причем, как это отмечается многими авторами (см. гл. I), темп технологического развития дол-

жен быть очень высоким. Практически это означает, что проявления космической деятельности цивилизаций будут заметны в течение сроков, весьма малых по сравнению с космической шкалой времени. Тогда отсутствие космических чудес в рамках предлагаемых гипотез может объясняться: а) чрезвычайной редкостью цивилизаций, б) тем, что все цивилизации находятся примерно на одинаковом «раннем» этапе развития, в) кратковременностью жизни цивилизаций. Подробное критическое обсуждение этих возможностей было проведено С. Лемом.

Итак, что можно сказать в защиту тезиса о чрезвычайной редкости космических цивилизаций? А. И. Баумштейн [12] доказывал уникальность возникновения жизни на Земле в связи с вероятностно-комбинаторным подсчетом правдоподобия наличия различных дополнительных условий, необходимых для био- и антропогенеза (определенной силы тяжести спутника-Луны, «нужной» смены климатических условий и т. д.). С. Лем [1] подверг справедливой критике возможность применения комбинаторных подсчетов для вероятности реализации существования сложной динамической системы эволюции. Система эволюции отличается весьма высокой пластичностью и наличие или отсутствие определенных условий не ставит ее перед выбором «жизнь или смерть», а накладывает лишь определенные рамки на пути дальнейшего развития. Подобный расчет доказывает только весьма малую вероятность возникновения в других мирах дословного повторения земного антропогенеза и тем самым ослабляет позиции «антропоморфных» гипотез.

Далее, под словами «чрезвычайная редкость» (если считать, что остальные возможности не реализуются) надо понимать такую разреженность космических цивилизаций, что одна наша земная цивилизация приходится чуть ли не на всю наблюдаемую часть Вселенной (иначе мы бы наблюдали проявления деятельности сверхцивилизаций типа постулируемых Н. С. Кардашевым). Такое предположение требует согласования с принятыми космогоническими представлениями о типичности условий в Солнечной

системе, типичности самой Солнечной системы для Галактики и типичности нашей Галактики.

Одной «редкостью» цивилизаций объяснить факт отсутствия наблюдений их деятельности можно, только приписывая земной цивилизации уникальность.

Предположение б) также представляется противоречащим тем же космологическим взглядам на различный возраст объектов и эволюцию Галактики и Метагалактики. Вряд ли приемлемо положение о том, что условия для возникновения жизни возникли сравнительно недавно, по космической шкале времени, для всей наблюдаемой области Вселенной.

Для обоснования тезиса о кратковременности жизни цивилизаций приводятся обычно высказывания о возможных катастрофических кризисах, которые возникают по мере поступательного развития [13]. Однако логика таких высказываний, применительно к объяснению отсутствия «космических чудес», требует, чтобы неизбежная гибель на ранней стадии «технологического» развития сопровождала всем цивилизациям [1]. Только при такой фаталистической неизбежности мы не будем наблюдать проявлений деятельности хотя бы небольшого числа «оставшихся», бурно прогрессирующих цивилизаций. Постулировать такой чудовищный детерминизм, однако, весьма затруднительно *).

*) Отметим особо, что речь идет о непримлемости тезиса о неизбежности уничтожения или распада цивилизационной системы в «подходящий» момент, т. е. на грани активного выхода в космос или до этого. В принципе неизбежность «умирания» отдельной цивилизации не только не противоречит, а соответствует общим положениям диалектической философии, утверждающей, что только материя в целом «неуничтожима», все другие явления возникают, развиваются и гибнут, сменяясь новыми формами существования [12]. Противоположный взгляд приводит к неправомерным утверждениям типа, что «Вселенная проникнута сознанием» или «сознание — неуничтожимый атрибут материи». Став на такие позиции, мы были бы вынуждены (опять-таки из-за эффекта отсутствия «космических чудес») вернуться к идее об уникальных свойствах наблюдаемой Вселенной, где существует одна и только одна земная цивилизация.

Поэтому, видимо, правильно считать, что высокоорганизованная форма существования регулярно возникает в различных

Базируясь на факте отсутствия признаков деятельности космических цивилизаций при учете слабости позиций «антропоморфных» гипотез, С. Лем выдвинул интересную гипотезу о типичности «нетехнологического» развития для большинства существующих космических цивилизаций. Для «технологической» цивилизации С. Лем считает современную «энергетическую» фазу, с выходом в космос, весьма кратковременной, которая заменяется (отчасти под давлением ряда кризисов типа «информационного», «организационного» и др.) качественно новой формой развития. Гипотеза Лема непринужденно объясняет факт отсутствия «космических чудес» и не требует категорического ответа на вопрос о длительности существования цивилизаций. Правда, отметим, что с точки зрения традиционных представлений, нам трудно согласиться с мыслью о «запрете», накладываемом на космическую деятельность в развитии цивилизации. Но во всяком случае предположение Лема имеет меньше внутренних противоречий, чем любая «антропоморфная» гипотеза.

Попробуем теперь обобщить высказывания о слабых сторонах теорий существования внеземных цивилизаций и проблем связи с ними, опирающихся на «экзобиологический» и «прогнозный» принципы.

Существенным методологическим недостатком построения гипотез такого плана является обилие дополнительных предположений. Эти априорные предположения связаны с «ортоэволюционным» методом рассуждения (прогноз будущего развития, основанный на линейной экстраполяции) и «антропоморфизмом», утверждающим универсальность этого способа развития для всех (или большинства) космических цивилизаций. При этом методе абсолютизируются отдельные стороны явления, которым придается

уголках Вселенной. Искры возникшей жизни ярко разгораются, а затем угасают и снова возникают там, где существуют подходящие условия. Сколько времени существуют такие системы, сказать трудно. Современные знания, на наш взгляд, не позволяют дать даже примерной оценки типичной длительности «психозойской» эры.

решающий смысл, в то время как другие возможности игнорируются (например, последствия «информационного» кризиса).

В итоге оказывается, что построенные подобным образом теории не разрешают трудностей, связанных даже с учитываемыми эффектами. Так, например, проблема «энергетического» кризиса «выносится» в космическое пространство и решение ее откладывается на «лучшие времена».

Антропоморфизм наших представлений не позволяет удовлетворительно развить важную для этого класса гипотез концепцию «космического чуда». Критерии искусственности могут черпаться только из внутренней системы антропоморфных понятий. Поэтому единственным четким критерием может являться в этих рамках только дословное повторение внешних проявлений типа человеческой деятельности на других космических объектах (например, прием «антропоморфной» семантической системы сигналов).

В то же время основой, на которой строятся экстраполяции антропоморфных гипотез, является наше знание о белковой форме жизни и структуре земной цивилизации на современном этапе. Встает вопрос, возможно ли найти другой путь, чтобы заложить принципы менее противоречивой, более доказательной «теории» внеземных цивилизаций?

Безусловно, любая неспекулятивная гипотеза должна опираться на свод научных знаний. В этом смысле можно утверждать, что решение проблемы космических цивилизаций надо (по крайней мере в настоящее время) искать на Земле! Достаточно ли учитываются все современные научные представления в построении такой сложной теории? И. С. Шкловский [2] указал на принципиальную важность для теории космических цивилизаций использования таких разрабатываемых современной наукой проблем, как функциональное определение жизни, «разума». Однако в высказываниях И. С. Шкловского звучит тенденция, что эти исследования находятся на весьма раннем этапе и не дали еще принципиальных конкретных выводов, которые уже сейчас могли бы быть примене-

ны при изучении вопроса о существовании и формах разумной жизни во Вселенной.

Конечно, для теории внеземных цивилизаций было бы весьма важным установление четких определений жизни, мышления, не связанных с узостью описания конкретной формы и особенностей био- и антропогенеза. Безусловно и то, что до разработки таких исчерпывающе четких определений еще далеко. Но, с другой стороны, современные научные дисциплины, связанные с кибернетическими методами исследования, предлагают новый подход к изучению явлений окружающей нас действительности. Именно принципиальность этой новизны открывает широкие перспективы перед соответствующими областями знания. Новый метод «кибернетического мышления» весьма эффективен именно в таких проблемах, которые связаны с изучением сложных больших систем.

«Кибернетические методы» позволяют навести определенный порядок в проблеме внеземных цивилизаций, уточнить терминологические понятия, оценить объективность высказываемых предположений, наконец, правильно сформулировать этапные задачи. Ниже мы попытаемся последовательно изложить некоторые конструктивные принципы использования этого метода в приложении к проблеме существования цивилизаций, по возможности стремясь дать переоценку отдельных положений существовавших ранее теорий. Но прежде остановимся на тех работах, в которых используются методы системного подхода.

§ 3. Другая точка зрения. С. Лем и его «Сумма технологий»

В своей книге «Сумма технологий» [1] С. Лем подробно осветил ряд вопросов, связанных с проблемой внеземных цивилизаций. Собственно, «астрономической» части проблемы в этой книге уделено скромное место. Но, как мы пытались показать, в действительности вопрос о существовании космических цивилизаций является частью более широкой

задачи изучения свойств и закономерностей развития очень сложных систем.

С. Лем обсуждает эти «смежные» проблемы, отводя основную роль изучению возможностей «прогнозирования» дальнейшего развития цивилизации. Подробно рассматриваются в свете современных данных основные свойства биологической эволюции. Обсуждаются потенциальные возможности естественного биогенеза в рациональном устройстве и управлении «живой» системой. Эти возможности сравниваются с теми требованиями, которые выдвигает наука при создании сложных искусственных систем.

В [1] указывается, что во всех теоретических построениях, связанных с «прогнозом» пути развития человечества, несколько идеализируется «доскональность» биогенеза. Связано это с тем, что сейчас человечество еще не способно достигнуть такого совершенства в технологии синтеза органических веществ, как естественная биологическая эволюция. Поэтому нам присуще стремление абсолютизировать «данные» нам природой возможности и игнорировать рассмотрение путей к «улучшению» человеческой природы путем искусственной «автоэволюции». «Когда химический синтез, теория информации, общая теория систем уйдут далеко вперед, человеческое тело будет выглядеть в свете этих достижений наиболее несовершенным элементом». Тогда перед человечеством с необходимостью встанет задача улучшения самого «несовершенного» элемента в системе цивилизации. Следует отметить, что это, видимо, должен быть именно путь «автоэволюции», а не путь улучшения «бытовых условий», так как само несовершенство природного устройства человеческого организма ставит сравнительно ограниченный предел человеческому существованию, не более 100 лет. Часто встречающийся взгляд, что пределы человеческой жизни можно довести легко до 150—200 лет соответствующей системой медико-гигиенических методов, весьма сомнителен. Основой для подобного мнения является, прежде всего, вера в «огромные потенциальные возможности», якобы заложенные в организме человека. В процессе эволюции

организмы достигают высокой пластичности, приспособленности и имеют известные «резервы надежности». Однако биологическая эволюция «не планирует» на будущее свои достижения. Выживают именно те организмы, которые лучше всего приспособлены к существующим условиям.

В биологической эволюции отбираются только те черты, которые имеют значение для «цели» генезиса, т. е. для выживаемости вида в целом. Что же касается долговечности, то природу «не интересует», что будет с индивидуумом, сколько он проживает после выполнения им жизненных функций, существенных для системы (размножение, опека над потомством). Превышение некоторого «необходимого предела» длительности жизни определяется в природе случайными причинами. «Антропоморфизм» в подходе к вопросу о биологической долговечности человеческого организма заключается в приписывании биоэволюции несвойственного ей наличия потенциальных резервов в реализуемых возможностях, которые не нужны для «немедленной реализации». В конечном счете это ведет к телеологическим представлениям о биогенезе. Кроме того, «методы» эволюции носят вероятностный характер. Из того, что отдельные особи доживают до некоторого предельного возраста, не следует принципиальная долговечность вида. Тот факт, что отдельные люди живут 150—170 лет, аналогичен факту, что отдельные люди гениальны, а вывод о возможности «дотягивания до гениальности» улучшением условий (обучение, быт) не может быть сделан всерьез.

Важное место в своей книге С. Лем посвящает анализу проблемы «информационного» кризиса. Возможность этого кризиса обсуждается в связи с характером развития науки на современном этапе. Поступательное развитие производительных сил (в частности, и поиски дополнительных источников энергии) прежде всего требует все большего количества научной информации. Проследить это можно на историческом примере, изучая количество научных исследований, которые надо было производить при переходе от одного вида энергии к другому. Эта величина

непрерывно возрастала. Без увеличивающегося темпа накопления информации развитие современного общества не может продолжаться. С другой стороны, мы не можем планировать «нужные» научные открытия. «Стратегия» науки носит «вероятностный» характер. В ходе научного прогресса земная цивилизация обеспечивает научными силами все участки возможных исследований, ибо неизвестно заранее, где именно будут произведены фундаментальные открытия. Ясно, что при таком положении, с одной стороны, будет возрастать лавина получаемой информации, с другой, будет увеличиваться число людей, занимающихся научными исследованиями. Высокий темп развития процесса должен вскоре привести (при условии количественного решения проблемы) к кризисным явлениям (например, к исчерпанию людских резервов для научных исследований). Вероятностный характер «производства» важных открытий не позволяет «разумно ограничить» области ведущихся исследований. При современных условиях это грозило бы еще большими бедами. Лем замечает, что и так на современном этапе чрезмерная гиперболизация отдельных отраслей знания, связанных с ракетной техникой и космическими исследованиями (тут, разумеется, важны и социально-политические причины), пагубно отражается на фундаментальных исследованиях в других областях. Разные же области знания не являются изолированными. Поэтому искусственное торможение развития исследований в одних областях за счет роста в других, в конце концов, не дает возможности продвигаться и в привилегированных участках науки. Гипертрофированное развитие «модных» отраслей приводит часто к стремлению решать все проблемы «в лоб», за счет количественного возрастания мощностей эксперимента (увеличение мощности ускорителей частиц, площади радиотелескопов и др.) [14].

Как избежать «информационного» кризиса? Из характера развития науки в человеческом обществе вытекает, что особую роль играет увеличение «канала информации» между природой и цивилизацией. До сих пор расширение этого «канала» шло за счет уве-

личения числа областей знания и количества ученых. Успехи кибернетики позволяют надеяться, что удастся создать сложные машины, которые помогут человечеству в сборе и переработке увеличивающегося количества информации («усилители разума»). Это не означает изменения характера научных исследований, а лишь «автоматизацию» процесса.

Однако ясно, что путь чисто количественного увеличения «синтетических ученых» только отдалил бы кризис, но не устранил причины, его порождающие. Единственной возможностью (при сохранении прочих условий), по мысли Лема, является создание искусственной системы, которая непосредственно выделяла бы нужную информацию из окружающей среды, т. е. такой машины, которая являлась бы не помощником ученого в обработке информации, а мощной аналитической системой, возможности которой превышали бы возможности мозга. Лем развивает концепцию некоторой синтетической эволюционирующей системы, которая была бы способна увеличивать «выходную» информацию в процессе своего развития наподобие накопления генетической информации в биогенезе, но только с «направленным и улучшенным» действием, — это своего рода система, «выращивающая» научные выводы. Лем показывает, что возможность реализации подобной системы не противоречит выводам современной науки и анализирует детали создания такой системы с точки зрения механизма научного познания и способов передачи информации в биогенезе.

Однако и создание автономной системы переработки информации может разрешить лишь часть возникающих проблем. «Информационный» кризис существует не сам по себе, а связан с целым рядом других важных сопутствующих процессов. В «Сумме технологий» обсуждаются возможности появления и других кризисных явлений. Если считать, что процесс добывания информации будет как-то разрешен, то отсюда не обязательно следует вывод о «спокойном» развитии в виде «космической колонизации» с расширением сферы «рабочего» жизненного пространства при неограниченном росте населения и

энергетических ресурсов. Одним из основных принципов существования и деятельности сложных систем является их регулируемость. Мы не знаем пока, существует ли некоторый предел сложности структурного состава сложной системы (количества элементов, ее составляющих). Не исключена возможность, что превышение определенного числа составляющих элементов приводит к тому, что система становится неуправляемой и дезинтегрируется. Применительно к будущему «чрезмерно разросшихся» цивилизаций вопрос управления не сводится к банальному: как найти занятие всем членам общества? Речь идет о таких фундаментальных проблемах, как, например, сохранение единства культуры такой гигантской системы. Важным связующим звеном в развитии человечества является именно преемственность различных этапов, передача «информации» от поколения к поколению при условии интенсивного обмена ею между членами общества. При этом непрерывно расширяются возможности личности. Затрудненность информационного обмена в гигантской сверхцивилизации грозит опасностью «обезличения» общества, каждый член которого будет превращаться в узкоспециализированную «ячейку», выполняющую определенные служебные функции (иначе трудно себе мыслить «целесообразную управляемость» подобной системы).

Лем замечает, что цивилизация вовсе не означает роста всех возможных индивидуальных свобод. Наоборот, развитие общества сопровождается постоянным введением «запретов», связанных с познанной необходимостью. Встает интересный вопрос: если научный анализ подтвердит, например, невозможность регулируемости систем, состоящих из слишком большого числа элементов, или выяснится, что управление такой системой можно обеспечить лишь ценой «нивелировки» членов общества, — не будет ли это слишком дорогой ценой за «свободу» безграничного продвижения в космос?

Теоретически логично предполагать, что общество, встретившееся с принципиальными трудностями на пути «космической экспансии», «разумно» откажется

от такого плана развития. Это, кстати, вовсе не будет означать «деградации» цивилизации. Никто не доказал, что путь «стихийного» развития человеческой цивилизации является «естественно правильным» и не может привести к отрицательному результату или к огромным иррациональным затратам в будущем. Достаточно хотя бы вспомнить такое нежелательное последствие использования атомной энергии, как опасность генетического вырождения, или последствия неразумного, «бесхозяйственного» расходования природных ресурсов Земли.

Осознанная необходимость легко усваивается цивилизацией и не воспринимается как нечто «противоестественное». Например, вполне серьезным кажется суждение, что «демографический» кризис в общем легко разрешается в этической сфере разумным ограничением деторождения (которое, кстати, и проявляется по мере роста материальной культуры).

В противовес «ортоэволюционному», «энергетическому» прогнозу развития Лем выдвигает гипотетическую идею иного возможного пути развития цивилизации. Об «автоэволюции» мы уже говорили. Другой возможностью, сопутствующей «автоэволюционному пути», Лем считает создание «мира в мире», т. е. конгломерата искусственных условий в достаточно большом объеме пространства, который подчинялся бы системе запрограммированных искусственных законов этого «мира».

Законы движения, распространения сигналов, структурные элементы всех материальных объектов такого «заповедника» были бы выбраны такими, как это представлялось бы наиболее рациональным с точки зрения «управления» этими объектами и помещенной в такой мир, «замкнутой» в нем цивилизации. Такая система представляла бы собой искусственную машину, состоящую из двух основных частей: «среды» и «цивилизации», причем «действие» машины заключалось бы во взаимодействии этих двух частей. В «Сумме технологий» разбираются принципиальные возможности создания подобной системы.

Безусловно, что эта гипотеза Лема весьма дискуссионна. Однако важны методологические выводы, которые могут быть сделаны из рассмотрения подобной концепции: наряду с «энергетическим» путем возможны другие варианты развития цивилизации, которые теоретически по меньшей мере так же вероятны, как и энергетический. Самый главный вывод состоит в том, что если принципы действия сложных систем познаваемы, то рано или поздно прогресс науки позволит определить, какие из всех возможных путей развития оказываются оптимальными. Лем подводит нас к постановке принципиального вопроса: какова должна быть «цель» цивилизации и каким в соответствии с этой «целью» представляется путь ее развития.

Под «целью» сложной системы можно понимать внутренне осознанный принцип ее действия. Можно говорить об объективности категорий «разум», «сознание», «логика систем», подчеркивая, что они являются функциональными свойствами сложных систем. Появление определенной «метафизики», «догм» неразрывно связано с практической деятельностью такой системы, как цивилизация. На каждом этапе система не обладает «полным знанием» об окружающей действительности, но действует так, как если бы это знание было полным. Определенность действия сопряжена с уверенностью в его «правильности», иначе система не действовала бы. В этом и заключается естественная «метафизика», «догматичность». В своем развитии цивилизация движется через целый ряд этапных «догматических», «рабочих гипотез», определяющих ее практическую деятельность и являющихся определенным приближением к объективной действительности. Процесс превращается в «чистую догму» в том случае, если он не проверяется постоянно экспериментом, сравнением с реальностью. В таком случае все объективно происходящие события извращаются в сознании, подгоняются под одну из каких-либо догматизированных «рабочих гипотез», и фактически поступление дополнительной информации прекращается (примером является религиозная догматика). Сила «рабочих догматов» заключается в их

постоянном изменении, приведении в соответствие с научным знанием. Решать вопрос о «цели» цивилизации можно, только обладая знаниями о принципиальных свойствах сложных эволюционирующих систем. Пока достаточных данных еще нет. Но бурное развитие науки позволяет надеяться, что в ближайшее время теория систем будет давать определенные «рекомендации» для путей развития всей человеческой цивилизации. Безусловно, принимать и воплощать в жизнь эти рекомендации может только единое, гармонично развивающееся и наиболее «саморегулируемое» коммунистическое общество.

Вопрос о том, каков будет этот путь — распространение в космос или создание искусственного автоэволюционирующего «мира Лема», не может, как представляется, быть решен в наши дни. Пока мы, скорее, видим трудности различных вариантов предстоящего пути, а не конструктивные возможности его осуществления.

Заканчивая на этом краткий обзор книги Лема, отметим еще раз ценность методологического подхода, выдвигаемого в ней для решения круга проблем, связанных с поиском внеземных цивилизаций.

Методы, обсуждавшиеся в предыдущей части статьи, применимы не только к вопросам «общей теории цивилизаций» — анализу принципиальных свойств сложных систем. Подобная методика плодотворна и при изучении «частных» вопросов. Такой проблемой является возможность «естественного» образования сложных систем во Вселенной. Характерным примером «кибернетического» подхода к решению задачи появления жизни во Вселенной является работа М. Таубе [15]. Исходя из функционального определения жизни, данного А. А. Ляпуновым [16], Таубе рассматривает те «естественные» процессы и материальные объекты в космосе, которые могли бы служить материалом для построения живого организма. При этом учитываются такие необходимые условия, как достаточная распространенность определенных элементов, способность различных соединений образовывать структуры и выполнять определенные

служебные функции (перенос высоко- и низкоэнтропийной энергии, передача и хранение информации). Таубе приходит к выводу, что материальным носителем жизни могут быть в естественных условиях только спонтанно синтезирующиеся в неживой среде молекулы веществ, представляющие соединения с водородом. «Использование» в качестве «строительного материала» для форм жизни соединений без участия водорода, кислорода и углерода представляется автору принципиально невозможным. Представим дальнейшему анализу вопрос о корректности категорических выводов Таубе и остановимся на методе его рассуждений. Во-первых, четко сформулирована задача с точки зрения изучения принципов действия систем. Затем исследуется широкий класс явлений, которые могли бы выполнять функции «живой системы». Кроме того, рассматривается (хотя далеко не полно) вопрос о том, когда и как могут реализоваться условия естественного создания сложных «живых» систем.

Такой подход лишен «антропоцентризма» поисков только белковой формы жизни. Во всяком случае, ищутся основания к доказательству «универсальности» белковой формы жизни, как одной из немногих возможных в типичных условиях космической среды.

Это ведет, безусловно, к большей объективности выводов о возможности существования «жизни» в космосе. Успехи подобных разработок позволяют определить возможное «внешнее» конструктивное оформление систем, претендующих на звание «живых» и тем самым уточнить критерии различия «живого» и «неживого» во Вселенной.

В ходе обсуждения проблемы внеземных цивилизаций на современном этапе часто ставится вопрос о возможных последствиях контактов с инопланетным разумом. Как правило, ответ на этот вопрос ищется также в рамках «антропоморфных» представлений. Космические цивилизации рассматриваются с точки зрения их «гуманности» (или, наоборот, «агрессивности») [4]. Разница в цивилизациях усматривается в чисто количественном «ортоэволюционном» плане. Справедливости ради, отметим, что в ставшей библи-

ографической редкостью книге Степлдона [17] впервые подробно рассматривались проблемы контактов с «иначе устроенными» цивилизациями.

Следствием «антропоморфизма» представлений является и обсуждение последствий «космических контактов» в виде «взаимопомощи» цивилизаций в ходе их существования.

В свете изложенных ранее соображений вопрос о последствиях контактов с «цивилизациями» может рассматриваться как проблема «конкуренции» различных видов «разума». Более «рационально устроенные» системы обладают лучшей приспособленностью и этот факт будет иметь решающее влияние на другие «цивилизации». Ситуация будет напоминать ту, которая обсуждалась на стр. 408 в связи с обоснованием необходимости «автоэволюции». «Узнав» о своем «нерациональном» устройстве, цивилизация, несомненно, будет использовать полученное знание для своего преобразования. Обоснование «бездействия» по причине «неприемлемости» других путей развития для человеческого общества было бы равносильно признанию теологического тезиса об уникальности человечества и «предначертанности» его пути. Поэтому следствием контактов с внеземными цивилизациями будет получение «чисто научной информации» о сравнительных характеристиках принципов действия других систем. Несомненно, с методологической точки зрения полезными будут выводы, ведущие к избавлению от антропоцентризма в научном мышлении. Детализацию спекулятивных рассуждений о возможных «недоразумениях» при космических контактах можно предоставить авторам научно-фантастических произведений.

§ 4. Проблемы космических цивилизаций с точки зрения общей теории систем

Новые научные дисциплины, объединяющиеся под флагом кибернетических методов, возникли как обобщение принципов, которыми наука и до сих пор пользуется в изучении окружающей нас действитель-

ности. При изучении очень сложных по структуре явлений появилась необходимость уточнить сами принципы построения научных теорий и методов анализа. Поэтому круг кибернетических идей не приносит в науку «нового способа мышления», но уточнение смысла фундаментальных понятий позволяет осуществить новый эффективный подход к исследованию самых запутанных проблем естествознания. Естественно, что ценность корректных методов научного исследования, даже если они ведутся в сугубо «традиционной» классической манере, нисколько не умаляется открытием новых обобщенных принципов. Эту мысль о преемственности научных методов проиллюстрируем на следующем примере. В предыдущей части обзора подвергалась критике некоторая концепция проблемы внеземных цивилизаций. Однако в этой критике не использовалась и не противопоставлялась «классическому» взгляду некоторая «кибернетическая идейность», хотя по ходу изложения и высказывались замечания о необходимости системного подхода к изучению проблемы. В общем, сомнению были подвергнуты принципы, положенные в основу «энергетической» гипотезы. С другой стороны, решение конкретных частных задач проблемы существования цивилизаций ставится, как правило, правильно, и совпадает, по сути дела, и с решением в духе «кибернетических» принципов.

Так, при обсуждении вопроса об исследовании радионизлучения от подозреваемых в «искусственности» космических объектов В. И. Сифоров [18] обратил внимание на анализ статистической структуры радиосигналов. «В частности, — замечает Сифоров, — если предположить, что приходящие сигналы являются узкополосными, то целесообразно при их анализе экспериментально определить двумерную плотность распределения вероятностей расположения на плоскости конца вектора, изображающего амплитуду и фазу принимаемых колебаний. По характеру поверхностей, изображающих эти двумерные плотности, можно судить о том, использована или не использована для генерации колебаний обратная связь. Нам предста-

вляется, что изучение статистической структуры принимаемых сигналов должно оказаться полезным при решении вопроса о том, являются ли эти сигналы «искусственными» или они обусловлены естественными процессами, не связанными с деятельностью разумных существ.

Мы хотим обратить внимание на соответствие такого подхода методу «черного ящика» (стр. 428), являющемуся одним из основных в современной кибернетике. Последовательное применение такого метода при переходе от узких задач (свойства генератора радиосигналов — по статистике сигнала) к более широким (критерии искусственности) представляется — мы будем стараться в дальнейшем развить эту идею — наиболее плодотворным *).

Для последующего необходимо изложение определенного круга идей «новых» научных дисциплин, таких, как кибернетика, теория информации и другие.

Под термином «система» в кибернетике понимается взаимосвязь различных элементов, представляемых в виде совокупности переменных величин, которые должны приниматься в расчет. Обнаружение систем согласно этому определению связывается с анализом взаимосвязанных явлений в мире. Основной упор делается на изучении принципов связи, а не конкретных проявлений систем в виде определенных материальных конструкций. В этом смысле системой будет и атом, и агрегатное состояние твердого тела, язык, игра в шахматы, разговор двух собеседников и т. д.

Важным обстоятельством является возможность классификации систем. Обычно такая классификация производится по степени сложности системы. Кроме того, классификацию можно произвести по другим принципам, например, по характеру связи: детерминированные и вероятностные системы.

*) Подобный методологический подход к проблеме космических цивилизаций был, по-видимому, впервые выдвинут Голеем [19].

Развитие понятия системы и ее свойств приводит к определению термина «машина». Под «машинной» подразумевается система, изменяющая свое состояние так, что переменные величины, характеризующие ее, связаны определенным законом преобразования. В соответствии с характером системы могут существовать детерминированные и вероятностные машины с простыми и сложными законами преобразования их текущих параметров. Машину можно понимать как «целесообразную» систему, т. е. систему, организация которой с определенной точки зрения подчинена осуществлению поставленных перед нею задач [20]. Слово «задача» не следует понимать как «предусмотренная и поставленная другой системой (например, «человеком») цель», а только как принцип действия*). Машинной с этой точки зрения будет широкий класс явлений от атомов отдельных элементов до планетных систем, клетки и ткани живого существа, сами живые существа, «биосфера», биологическая эволюция в целом.

Как систему или «машину» можно рассматривать и проблематично существующую космическую цивилизацию.

Важной характеристикой машин является характер динамических связей между различными частями системы, например, наличие положительной и отрицательной обратной связи.

В своей книге [1] С. Лем рассматривает чрезвычайно интересную классификацию машин, предложенную де Латилем. Де Латиль различает три главные группы машин по способу действия. Свойства представителей каждого следующего класса включают свойства, характерные для предыдущих. К первой группе (детерминированные системы) относятся по

*) Вопрос о том, как избежать в кибернетических описаниях приписывания рассматриваемым системам осознанной целенаправленности, очень важен не только с точки зрения построения правильного «метаязыка» для описания некоторых свойств сложных систем, но и для выяснения объективной сущности таких явлений, как «самосознание», «психология», «цель существования» сложной системы (см. [21]).

классификации Латиля простые и сложные орудия труда (неавтоматические устройства) и те системы, которые не связаны с окружением обратной связью. Ко второму классу относятся организованные регулирующие и саморегулирующиеся системы с обратной связью. В этот широкий класс попадают и механические автоматические регуляторы с обратной связью, программированные машины и самопрограммирующиеся устройства (в том числе человек и животные, взятые как отдельные представители). Третья группа включает системы, которые могут менять свое устройство и «принципы действия», используя данный этой системе материал. Примером является биологическая эволюция. Де Латиль предполагает существование четвертой группы систем, которые имеют вдобавок свободу и возможность выбирать из окружения подходящий материал, из которого «сами себя делают». Такой системой можно считать научно-техническую деятельность человечества в целом. По мнению Лема, классификацию Латиля можно расширить, включив в нее еще одну группу систем, которые материал для «самоорганизации» выбирают не из имеющегося в наличии во Вселенной «естественно существующего» вещества или используя физико-химическую технологию для производства синтетических материалов, а создавая «синтетические» условия, которые не могут образовываться при естественно-физических процессах. Речь идет об искусственно создаваемых новых формах существования материи, что, как считает Лем, в будущем способно будет делать человечество.

Классификации, подобные приведенной, имеют важное значение. Они позволяют определить место интересующей нас системы в ряду прочих по определенным принципиальным признакам, что позволяет научно сформулировать постановку задачи анализа данной системы в нужном плане. Четкость же постановки задачи особенно важна в рассматриваемой проблеме поисков космических цивилизаций.

Следующим важным шагом в систематике кибернетических понятий является обобщение понятия устойчивости и стабильности систем. Существенным в

этом смысле представляется принцип гомеостазиса [20, 21]. Гомеостазис — это такой принцип действия целесообразной системы, при котором значения некоторых существенных внутренних переменных поддерживаются в пределах некоторых границ, несмотря на разнообразные (регулярные или нерегулярные) внешние воздействия.

Гомеостат в соответствии с этим представляет собой машину, обладающую необходимым механизмом регулирования, поддерживающим на определенном уровне всю систему «жизненно важных» внутренних параметров этого устройства. Существенно, что явления живой природы носят гомеостатический характер. Само понятие гомеостазиса введено в науку путем обобщения результатов биологических наблюдений. Обнаружение «гомеостатического характера поведения» системы играет поэтому большую роль в выяснении сущности данного явления. Можно утверждать, что и космические цивилизации представляют собой очень сложные вероятностные системы, имеющие гомеостатическую природу.

Фундаментальность принципа гомеостазиса вытекает также из того, что это, пожалуй, один из немногих четко сформулированных тезисов, определяющий возможную «цель» или «задачу» для развития эволюции сложных самоорганизующихся систем. Такие свойства живой системы как приспособляемость, выживаемость, сознание в общем подчинены выполнению принципа гомеостатичности. Поэтому эти свойства можно рассматривать как подсистему эффективных регуляторов, выполняющих роль поддержания общего гомеостазиса системы.

Понятие «разумности» весьма существенно, как мы видели, для попыток определить эффективные критерии искусственности в проблеме космических цивилизаций. Мы уже упоминали о том, что методы кибернетики дают возможность построения функционального понятия «разумности». Путь к этому определению прокладывают возможности теории сложных самопрограммирующихся и самоорганизующихся уст-

ройств, к созданию которых вплотную подходит современная наука.

Необходимым признаком «разумности» является, во-первых, способность системы хранить и перерабатывать информацию. В самом общем смысле информацию можно определить как меру упорядочения, уменьшения неопределенности состояния системы. В этом плане любая машина обладает информацией, так как определенность ее закона преобразования ограничивает разнообразие других возможностей (состояний), которые не реализуются. Поэтому любую машину можно рассматривать как машину для переработки информации [22]. Это один из основных принципов кибернетики. Современная теория информации успешно разрабатывает вопросы измерения количества информации и способов оптимальной передачи ее. Нас же сейчас интересует другой вопрос, а именно: какие способы хранения и переработки информации свойственны системам «высшего порядка», т. е. каковы достаточные признаки «разумности»?

Одна из важных особенностей передачи информации состоит в том, что информация передается в кодированной форме [20]. Информация в кодированной форме может храниться в системе, являя собой накопитель информации, «память» системы. Примером могут служить способы генетической информации в биологической эволюции. В соответствии с представлениями о формах хранения и передачи информации можно подойти к более сложному представлению — «логике системы». Под наличием логики в сложной высокоорганизованной системе можно понимать наличие способности отражать внешние закономерности окружения *) с помощью некоторого множества внутренних реакций, представимых в кодированной форме

*) Сама структура внутренних параметров может выступать для логического аппарата такой системы, как проявление «внешней» ситуации, т. е. как предмет для логического анализа (например, исследование человеком своего собственного тела). Такое расширение понятия «окружение» необходимо, чтобы избежать наложения ограничений на возможности аналитического аппарата подобного устройства.

и использования системы таких множеств для анализа и предвидения внешних ситуаций с целью поддержания и «улучшения» гомеостазиса своего существования (гомеостазиса всей системы). Следовательно, предполагается наличие специального «логического» приспособления, оперирующего с множеством кодовых знаков («понятий»). Таким устройством является, например, человеческий мозг. В настоящее время успешно разрабатываются вопросы о «формах логик», присущих различным сложным автоматам, что дает надежду получить уверенную классификацию особенностей различных «логик», в зависимости от структуры и принципа действия систем *)).

Изложенный круг «кибернетических» взглядов позволяет подойти к определению той системы, которая подразумевается под понятием «разумная». Можно определить условно космическую цивилизацию как существующую в космосе очень сложную вероятно-

*) Приведенные рассуждения показывают, что трудно найти резкую грань, разделяющую «разумные» и «неразумные» системы. Это подтверждается и некоторыми данными современной биологии, которые свидетельствуют о наличии определенных элементов «самосознания», «логики» у некоторых видов животных. Это лишний раз свидетельствует о функциональном характере такого понятия, как «разум», определяющегося материальной основой — структурой и наличием определенных механизмов. Такие понятия, как «сознание», «эмоциональное восприятие» и др. связаны с определенными свойствами сложных систем [21].

Традиционные интуитивные взгляды утверждали, что именно наличие «сознания», «логичности», «эмоций» является принципиальным и определяющим «разумность» системы. Современная наука открывает дорогу к истолкованию «служебного» значения таких явлений, как религия, искусство. Искусство можно в свете представлений о сложных системах трактовать не только как средство для получения дополнительной информации, но более как необходимый «тренировочно-обучающий» процесс, регулирующий и упорядочивающий «эмоциональные» и «эстетические» свойства сложных высокоорганизованных систем [1]. Во всяком случае, «надстроечные» явления в сложной системе человеческой цивилизации могут и должны интерпретироваться как явления, объективно связанные с «принципом действия» системы, свойства же их должны раскрываться через изучение функционального значения для системы [1]. На этой точке зрения и стоит исторический материализм.

стную машину гомеостатического характера, обладающую необходимыми механизмами в виде «логических способностей» для хранения и переработки информации, способную к анализу ситуации и применению результатов этого анализа для направленной эволюции системы в соответствии с определенными принципами направленного действия.

Как уже неоднократно подчеркивалось, существенное значение имеет именно функциональный характер кибернетического определения. Возможность (хотя бы принципиальная, теоретическая) функционального определения заставляет четко формулировать задачу о том, что мы хотим обнаружить при поиске космических цивилизаций.

Функциональное определение «цивилизационной системы» закрывает возможность «антропоморфизма» высказываний *). В класс космических цивилизаций попадают не только «человекоподобные цивилизации», но любые другие формы «разумного» существования, лишь бы они обладали достаточным набором параметров, необходимых для осуществления планируемой целесообразной деятельности. В связи с этим отпадает нужда в ограничениях, накладываемых на поиски возможных проявлений «разумной деятельности» определенным классом типологии и неопределенность утверждений о степени достоверности и однозначности интерпретации критических «труднообъяснимых» явлений. Разумеется, то же относится к критериям искусственности сигнала в радиодиапазоне. Открывается в принципе возможность «расшифровки» не только семантической информации, но и возможность

*) Во всяком случае, «антропоморфизм» высказываний «понижается» до такого уровня, на каком мы способны «неантропоморфно» рассуждать о закономерностях окружающего нас мира. Человеческому сознанию свойственно анализировать мир с помощью определенного логического аппарата (включающего, например, «математизацию» методов исследования). Говорить об «антропоморфизме» на этом уровне, значит, утверждать, что внешний мир устроен «хаотично», что в нем нет «причинных связей» и т. д. Такие утверждения противоречат материалистической теории познания.

структурного анализа происхождения сигнала постепенным рафинированием методов исследования сигнала.

Возможность функционального определения «космической цивилизации» дает право наметить характер подхода к развитию некоторых общих принципов «теории внеземных цивилизаций». Задача из «астрономической» превращается в «земную». Успехи теории сложных систем в развитии представлений об основных свойствах высокоорганизованной формы существования должны явиться фундаментом для построения корректных основ «частной» теории высокоорганизованных космических систем. В этом смысле значение теоретической кибернетики предстает для теории внеземных цивилизаций в том же плане, как, скажем, значение теоретической физики для современной астрофизики. Разумеется, теория внеземных цивилизаций может вносить свой вклад и в развитие кибернетических идей, например, за счет специфики тех возможностей, которые могут проявляться на различных космических объектах.

Многое может дать изучение кибернетических способов анализа. В кибернетике очень сложные системы исследуются методом «черного ящика».

«Черный ящик» представляет собой модель некоторой системы, к изучению которой мы приступаем, не зная ничего о внутреннем ее устройстве. Можно представить эту систему в виде машины со «входом» и «выходом». Под «входом» понимается возможность оказывать различные воздействия на систему, под «выходом» — ответные реакции, которыми система отвечает на определенные внешние воздействия. В принципе реальная система может иметь бесчисленное множество возможных «входов» и «выходов». Количество «входов» и «выходов», в некотором смысле, определяется самим методом и возможностями экспериментов, которые производятся над исследуемым объектом.

Прежде чем перейти к изложению метода «черного ящика», надо сказать несколько слов о весьма плодотворном понятии модели в науке. Ф. Энгельс еще бо-

лее ста лет назад провел блестящий анализ генезиса научного познания в связи с практической деятельностью человечества и обратил внимание на то, что непознанные явления мира, «вещи в себе», превращаются в окончательно познанную «вещь для нас» в том случае, если мы можем воссоздать такое явление искусственно. Превращение «вещи в себе» в «вещь для нас» представляет длительный поэтапный процесс, в ходе которого мы изучаем все новые и новые черты явления, все ближе подходим к знанию обо всех основных свойствах объективно существующей «вещи в себе». В этих положениях заложено ядро современного научного понятия «модели». Методы современной науки (например, физики) почти всегда используют для математического описания процессов модели тех явлений, которые нужно изучать. В сущности, все теории современной физики опираются на определенные физические или математические модельные представления. В том случае, если модель полностью соответствует оригиналу и на ней можно исчерпывающе изучать все свойства объекта, мы называем модель изоморфной оригиналу. В этом смысле можно, например, категорически утверждать, что представители одного класса объектов полностью изоморфны друг другу. Например, один атом водорода полностью изоморфен другому атому водорода. Но с позиций кибернетики изоморфной будет модель и «не похожая» на оригинал, но выполняющая всю совокупность операций, которые способен совершать оригинал.

Как правило, построение моделей сложных систем не приводит к полному изоморфизму. Поэтому появляется понятие гомоморфных моделей. Гомоморфная модель соответствует другому явлению только в определенной степени, на некотором уровне, и дает правильную интерпретацию лишь определенных свойств явления. Например, электронная счетная машина гомоморфна на определенном уровне человеческому мозгу, так как она производит ряд определенных логических операций счета, хотя ход процессов в машине сильно отличается от хода и характера аналогичных

операций в мозге. Существенно отметить, что если бы мы построили машину, способную совершать все операции, которые может производить мозг, и «области возможностей» такой машины и мозга совпадали бы, то подобное искусственное устройство с точки зрения кибернетики было бы изоморфной моделью мозга. Это замечание лишний раз подчеркивает, что кибернетику интересуют принципы происходящих процессов, а не конкретное материальное выражение в виде «вещи» со всей совокупностью ее внешних признаков.

Теперь можно изложить принципы исследования методом «черного ящика». Метод «черного ящика» приложим к изучению очень сложных систем, устройство которых для нас недоступно.

Совершая определенные воздействия на вход системы (что равносильно наблюдению определенных воздействий на вход со стороны среды или других систем, не обязательно являющихся искусственно созданными «экспериментальными приборами»), мы изучаем их функциональную связь с выходными реакциями «черного ящика». На каждом этапе создается соответствующая гомоморфная явлению модель (например, в виде рабочей гипотезы). Целью является установление закона преобразования «входа» в «выход», т. е. принципа действия машины. Накопление данных позволяет строить гомоморфные модели все более высоких уровней. Как предел, мы получаем изоморфную модель явления. Из сути метода вытекает парадоксальный на первый взгляд вывод: в принципе мы можем получить исчерпывающее представление о «черном ящике», не зная, как он устроен в действительности. Но этот вывод следует автоматически, если вспомнить определение изоморфной модели. Изоморфная модель эквивалентна реальной системе по своему действию. Такая модель взаимозаменяема с реальным объектом, так как может выполнять все аналогичные функции. Построение изоморфной модели соответствует искусственному воспроизведению явления. Вспомним, что Энгельс положил в основу критерия превращения «вещи в себе» в «вещь для нас» именно тезис о возможности искусственного воспроизведения!

Для того чтобы можно было говорить о применении метода «черного ящика» к проблеме космических цивилизаций, необходимо обсудить, как протекает информационный процесс при исследованиях, которые ведутся этим методом.

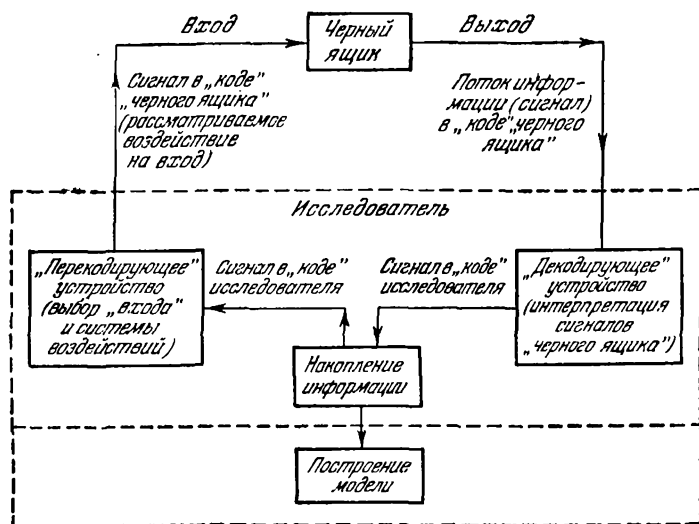


Рис. 72. Исследования методом «черного ящика с входом», представленные в виде информационной машины.

К сказанному выше об информации следует добавить, что передачу потока информации можно осуществить только в определенных системах, различные части которых имеют между собой каналы связи.

Процесс осуществления исследований методом «черного ящика» можно представить в виде некоторой информационной машины, где исследователь и сам объект исследования представляют собой единую систему, связанную обратной связью (рис. 72). Существенно, что информация передается в кодированной форме. Поэтому если у различных частей системы разные коды, то для движения потока информации

необходимы соответствующие перекодирующие устройства для перевода сигналов от одной части системы в «понятную» другой части системы форму. В соответствии с этим на рис. 72 изображен цикл процесса переработки информации в методе «черного ящика». Воздействие на вход «черного ящика» определяет реакцию выхода в виде некоторого сигнала*), в той форме, в какой эта система может производить. Исследователь «декодирует» эти сигналы, давая им соответствующее истолкование. Цепочку обратной связи создает изучение исследователем возможностей оказания воздействия на вход системы. Сама операция определения «входа» и выбора для изучения подходящей группы воздействий на вход «черного ящика» представляет в информационном плане операцию перекодирования «сигнала», идущего от исследователя, в сигнал, «понятный» «черному ящику». Каждый цикл обращения информации дает некоторые сведения исследователю, позволяет, накопив достаточное количество информации, построить модель «черного ящика» на соответствующем уровне.

Приведенная схема представляет собой фактически рассмотрение «модели познания» неизвестного явления, которое проводится обычно научными методами.

Исследования методом «черного ящика» особенно хороши при изучении явлений, происходящих на удаленных от нас космических объектах. Проблемы астрономии сводятся фактически к проблеме изучения недоступных для нас космических «черных ящиков». Возможно, что астрономы всех времен, не зная кибер-

*) Понятие «сигнал» довольно сложно и иногда допускает двусмысленность истолкования. В самом общем смысле под «сигналом» можно подразумевать способ переноса информации. При этом определении мы не связаны необходимостью обязательно иметь в виду определенный материальный носитель, «материальное» воздействие одной части информационной машины на другую. Так, например, отсутствие воздействия может также служить в «машине познания» сигналом, так как оно свидетельствует (несет информацию) о том, что в данных условиях какие-то конкретные явления не реализуются (Например, отсутствие «космических чудес» является определенным информативным сведением для теории внеземных цивилизаций.)

нетики, поступали методически правильно, «говорили прозой, сами того не замечая».

Специфика астрономических исследований заключается в том, что проблема «черного ящика» в ней усложняется. Астрономические объекты исследования представляют собой в некотором смысле «черные ящики» без «входа». Понимать это надо так: удаленность

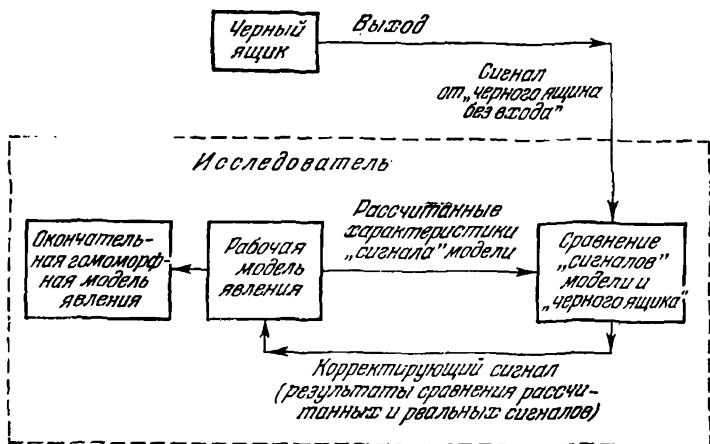


Рис. 73. Исследования методом «черного ящика без входа», представленные в виде информационной машины.

космических объектов позволяет судить о них только по тем сигналам, которые они нам присылают. Единственной используемой формой сигнала является электромагнитное излучение. Астроном лишен возможности «экспериментировать» с объектом, менять условия существования явления. Поэтому при наблюдениях астрономических явлений обратная связь в системе «исследователь» — «объект исследования» оказывается нарушенной (рис. 73).

В такой информационной машине циклическое движение информации происходит в подсистеме «исследователь». Поступающие от выхода «черного ящика» сигналы сравниваются с теми, которые «шли бы» от

некоторой модели явления. Фактически в астрономии мы имеем дело с методом аналогий. Многозначность выбора модели пытаются ограничить рассмотрением класса аналогичных явлений. Большое значение имеет возможность «предсказуемости» новых сторон явления и т. д. Однако с точки зрения рассматриваемых идей существующие методы исследования астрономических объектов, по существу, представляют собой способы «отбрасывания» некорректных гомоморфных моделей. Построение же изоморфной модели наталкивается на принципиальные трудности. Из-за отсутствия обратной связи «исследователь» — «объект исследования» возможным оказывается только построение гомоморфизма определенного уровня. Все это и заставляет астрофизику привлекать и обсуждать новые возможности получения сигналов от космических объектов и детализировать старые методы. Поэтому особое значение в последние годы приобретает учет и исследование «тонких» эффектов.

Однако забвение того факта, что в астрономии отсутствует изоморфное представление объектов, зачастую влечет за собой известную спекулятивность астрофизических теорий. С одной стороны, нередки случаи неохотного отказа от привычных концепций; с другой, чрезмерное увлечение поисками «модных» эффектов и попытками универсального объяснения явлений с точки зрения наиболее привлекательных на данный момент гипотез.

Современная кибернетика еще не разработала «теории изучения черных ящиков без входа». Но введение некоторых методологических идей этой науки в астрономию, видимо, бесполезно. Оно может дать дополнительный критерий надежности многочисленных теоретических положений науки о Вселенной. В последнее время появились некоторые работы, в которых астрономические явления анализируются с помощью такого «квазикибернетического» подхода.

Так, Л. И. Гудзенко и В. Е. Чертопруд [22, 23] предприняли попытку исследования солнечной активности методом, основанные на изучении статистических свойств «сигнала» — изменяющихся со временем

числовых параметров, характеризующих солнечную активность. Подобный анализ дает возможность выяснить, каков принцип работы механизмов, ответственных за проявление активности (являются ли они, например, автоколебательной системой). Такой подход имеет, на наш взгляд, преимущество перед традиционным, когда изучение солнечной активности связывается в первую очередь с поисками конкретных носителей этой активности, а затем уже подбирается модель «их возможного действия», объясняющая наблюдаемую картину. Сами же характеристики принципа действия «машины солнечной активности» оказываются недоуясненными.

Что же касается проблемы космических цивилизаций, то в ней учет методов исследований в духе кибернетики играет особую роль. В принципе, изучая все более и более детально структуру сигналов от «черного ящика без входа» по схеме рис. 73, мы можем получать гомоморфные модели все более высоких уровней. При этом будут возрастать наши знания о принципах действия сложной системы. Теоретически возможно построить классификационную типологию возрастающих по сложности функциональных действий систем; такой исследовательский метод дает возможность определить принадлежность того или иного объекта к определенному классу типологии.

Например, очевидно, очень широкий класс объектов будет представлять системы с наличием обратной связи всех видов и разной степени структурной сложности. Такие данные можно получать уже сейчас, изучая, скажем, статистическую структуру принимаемого радионизлучения. Далее, может быть, удастся выделить из объектов такого класса более узкий подкласс, в котором проявляется гомеостатичность. Наконец, из класса гомеостатических объектов выделить группу с наличием еще более сложных функциональных свойств. В конце концов, в принципе, можно обнаружить класс объектов, которые по своим проявлениям гомоморфны на высоком уровне земной цивилизации. Разумеется, задача эта невероятно сложна и ее

разрешение затянется на долгие годы и даже десятилетия. Но привлекательной стороной подобного метода является возможность четкой и последовательной постановки цели исследования. При таком подходе исключается неопределенность интерпретации явлений с той точки зрения, что они являются «искусственными» или «естественными». О каждом явлении можно будет сказать, что оно свидетельствует об определенных конкретных особенностях порождающего его механизма, не больше и не меньше. Если мы сможем доказать, что «особенности» эти характеризуют «логическую систему», то тогда и только тогда можно будет утверждать, что мы обнаружили космическую «цивилизацию» *).

Конечно, в класс «цивилизаций», в принципе, могут попасть системы, весьма отличающиеся по внешним конструктивным признакам (например, теоретически мыслимые «нетехнологические» системы с другими элементарными понятиями и т. п.).

При этом важно, разумеется, учитывать все дополнительные данные и соображения и о свойствах белковой формы жизни, и о специфике космических условий в различных областях Вселенной, и многие другие сведения, которые сейчас в той или иной мере используются при постановке проблемы внеземных цивилизаций (в том числе и организация поисков «антропоморфных» космических цивилизаций! **) .

*) Рассматриваемый метод предполагает, по своей сути, «минимальную» оценку. Так, если мы установим, что объект по признакам изучаемого сигнала попадает в нижний класс типологии, то это означает, что объект, безусловно, принадлежит этому классу, но не исключена возможность, что он принадлежит и к более высокому классу типологии. Видимо, такая форма утверждений наиболее адекватна для результатов астрономических исследований из-за отсутствия возможности доказать изоморфизм полученного результата реально существующему явлению.

**) Не исключена возможность успеха поиска «антропоморфной» цивилизации в виде обнаружения семантической системы сигналов сообщения. Но такая возможность несколько не подкрепляет принципиальности приводимых рассуждений. Речь идет не о невозможности обнаружения такого типа цивилизаций, так как подкласс «антропоморфных» систем вкладывается в соответствующую кибернетическую типологию. Если бы «человеко-

В принципе, методика поисков космических цивилизаций представляется следующей:

Необходимо разрабатывать вопросы о принципах действия и основных закономерностях очень сложных систем (включая не только структурный анализ этого явления, но и проблемы законов эволюции, прогноза дальнейших форм существования системы). Целью должно быть создание надежной классификации систем по существенным признакам. Это вопросы, подлежащие компетенции «земных наук», — теории сложных систем. Очень трудной задачей является определение «критерия искусственности», т. е. такой системы сигналов, которая бы однозначно свидетельствовала о том, что она посылается весьма сложно организованной системой («цивилизацией») *). Подобные сигналы должны были иметь какие-то специфические признаки, по которым их можно было бы отличить от других, сколь угодно сложных по строению сигналов, производящихся «нецивилизационными» астрономическими объектами. С кибернетической точки зрения это — проблема определения уровня организации «черного ящика без входа» по его выходному сигналу [24] или синтез производящих сигнал систем по характерным особенностям наблюдаемого сигнала.

Оказывается, возможно теоретически построить некоторые последовательности сигналов (структуры) такие, что производящая их система должна обладать

подобный» разум был обнаружен в том виде, как это предсказывает «антропоморфно-энергетическая» гипотеза, то это свидетельствовало бы о том, что справедливо утверждение об универсальности «технологического» и энергетического пути развития, но только в силу соответствия этого пути развития некоторым глубоким принципам развития сложных систем.

*) Здесь мы, естественно, оставляем в стороне вопрос о существовании «антропоморфной связи» в виде возможностей обнаружения, дешифруемой в понятиях информации. Безусловно, такой особый случай вполне доказателен как «критерий искусственности». Однако надо иметь в виду, что «семантическая связь» требует достаточно «длинных посланий» для последовательного объяснения значения отдельных символов, так что и в этом случае встает проблема «позывных», т. е. достаточно коротких и эффективных для целей опознания посылок сигналов (см. гл. II).

набором весьма сложных особенностей (типа наличия «памяти», умения «распознавать образы», создавать «абстракции» и т. п.). Такие последовательности сигналов представляют собой особым образом организованные по определенным теоретико-множественным принципам закономерные структуры, в которых как бы «демонстрируются» определенные функциональные возможности системы к производству сложных операций [25]. Разработка таких проблем также не является астрономической задачей.

В соответствии с успехом указанных выше исследований необходимо последовательно изучать «принципы действия», функциональные свойства астрономических объектов, с четким представлением о том, какие именно свойства мы хотим обнаружить. Это, безусловно, астрономическая проблема. Она включает и обобщающую часть: выяснение того, как специфика различных космических условий может влиять на изменение принципов действия сложных систем.

При этом не закрывается возможность ведения частных разработок по конкретным вопросам, связанным, например, со спецификой радиоастрономических наблюдений. Отдельную часть комплексных исследований может составить изучение возможных «помех» — шумов, затрудняющих прием сигналов и искажающих их значение.

Совокупность имеющихся научных данных позволяет утверждать, что проблема внеземных цивилизаций вполне может быть поставлена для решения на современном уровне в четком методологическом плане. Сложность проблемы заключается в том, что она неразрывно связана с еще более фундаментальными задачами. Поэтому только успехи в развитии методов кибернетического анализа, общей теории систем, биологии и других дисциплин позволят последовательно продвигаться и в решении задачи космических цивилизаций. Нельзя надеяться на «случайную удачу», счастливое «угадывание» ответа на вопросы, не всегда четко сформулированные. То же самое относится и к экспериментам, направленным на обнаружение астрономических сигналов, обладающих признаками

«искусственного» происхождения. Здесь первоочередной задачей является четкая формулировка критериев — «признаков искусственности» и детальный анализ очень широкого класса астрономических явлений. В общем это не является принципиально новой задачей. Так, С. Э. Хайкин отмечает, что задача планомерных поисков радиосигналов искусственного происхождения в целом совпадает с основной задачей радиоастрономии: накопления информативных данных по радионизлучению космических источников [26]. Специфика же проявится только на уровне применения «критериев искусственности» к конкретным классам объектов.

Проблема «общей теории цивилизаций» будет, безусловно, одной из «больших» тем современной и будущей науки. Успехи в ее разрешении будут определяться продвижением общего фронта научного прогресса. В том, что эта область исследований займет важное место в системе научных дисциплин, сомнения нет.

ЛИТЕРАТУРА

1. Lem S., *Summa technologiae*, Wyd. Lit. Kraków, 1964. Русский перевод: «Мир», 1968.
2. Шкловский И. С., *Вселенная, жизнь, разум*, изд. 2-е, «Наука», 1965.
3. Кардашев Н. С., *Астрон. ж.* **41**, 282, 1964.
4. «Внеземные цивилизации», Тр. совещания, Бюракан, 20—23 мая 1964 г., изд. АН Арм. ССР, 1965.
5. «Межзвездная связь», сб. статей, под ред. А. Камерона, пер. с англ., «Мир», 1965.
6. Lilley S., *Sociology of Science*, N. Y., 1962.
7. Price D., *Discovery* **6**, 240, 1956.
8. Добров Г. М., *Наука о науке*, «Наукова думка», Киев, 1966.
9. Simon R., *Astronaut and Aeronaut.* **3**, 59, 1965.
10. Freudental H., Linkos, Amsterdam, 1960.
11. Гладкий А. В., в [4], стр. 145.
12. Баумштейн А. И., *Природа*, № 12, 1961.
13. Брейсуэлл Р., в [5], стр. 271.
14. Hoyle F., *Of Men and Galaxies*.
15. Taube M., *Hydrogen the carrier of Life*, Nucl. Energ. Inform. center, Warsaw, 1965.
16. Ляпунов А. А., в сб. «Кибернетика, мышление, жизнь», «Мысль», 1964, стр. 127.

17. Stapledon O., Last and first man, Peng. B., London, 1939.
18. Сифоров В. И., в [4], стр. 121.
19. Голей, см. [5].
20. Эшби Р. У., Введение в кибернетику, ИЛ, 1959.
21. Бир Ст., Кибернетика и управление производством, Физматгиз, 1963.
22. Гудзенко Л. И., Чертопруд В. Е., Астрон. ж. 41, 597, 1964.
23. Гудзенко Л. И., Чертопруд В. Е., Астрон. ж., 43, 113, 1966.
24. Пановкин Б. Н., Докл. на сессии НТОРиЭ им. А. С. Попова, М., май 1967 г.
25. Пановкин Б. Н., Докл. на 1-й конф. по космической радиосвязи, М., 1968 г.
26. Хайкин С. Э., в [4], стр. 83.

**Внеземные цивилизации
Проблемы межзвездной связи**

М., 1969 г., 440 стр. с илл.

Редактор *В. Н. Комаров*

Техн. редактор *К. Ф. Брудно*

Корректор *Э. Ш. Сукачева*

Сдано в набор 24/X 1968 г. Подписано к печати 19/V 1969 г. Бумага $84 \times 108/_{32}$. Физ. печ. л. 13,75. Условн. печ. л. 23,10. Уч.-изд. л. 21,88.
Тираж 13 800 экз. Т-06037.
Цена книги 86 к. Заказ № 1631.

Издательство «Наука»
Главная редакция
физико-математической литературы
Москва, В-71, Ленинский проспект, 15.

Ленинградская типография № 2 имени
Евгении Соколовой Главполиграфпрома
Комитета по печати при Совете
Министров СССР
Измайловский проспект, 29.

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
Москва, В-71, Ленинский проспект, 15

ИМЕЮТСЯ В ПРОДАЖЕ КНИГИ ПО АСТРОНОМИИ:

- Астрономия невидимого. Перев. с англ. Под ред.
С. И. Сыроватского, 1967, 248 стр., 67 коп.
- Гребеников Е. А., Рябов Ю. А., Что такое небесная
механика, 1966, 111 стр., 21 коп.
- Зигель Ф. Ю., Звезды ведут в бесконечность. Картины ми-
роздания. Изд. 2-е, исправл. и дополн., 1966, 248 стр.,
46 коп.
- Комаров В. Н., Увлекательная астрономия, 1968, 432 стр.,
89 коп.
- Мензел Д. Г., Наше Солнце. Перев. с англ., 1963, 328 стр.,
79 коп.
- Паршин И. А., Луна, 1960, 53 стр., 10 коп.
- Элисон М. А., Солнце и его влияние на Землю. Введение
в исследование проблемы Земля — Солнце. Перев. с англ.
Под ред. И. С. Шкловского, 1959, 216 стр., 71 коп.

Перечисленные выше книги требуют в магазинах Книго-
торга. Письменный заказ можно направить также в ближай-
ший отдел «Книга — почтой» республиканского, областного,
краевого книготорга. Литература будет выслана наложенным
платежом. При отсутствии этих книг на месте следует обра-
щаться по адресу: Москва, К-50, ул. Медведева, 1, Магазины
№ 8 Москниги, отдел «Книга — почтой».