

4. Устройство отображения с шаговой разверткой, управляемое с диска. Оно оказалось почти незаменимым при визуальном исследовании колебаний, спектров и т. д.

5. Устройство ввода с перфокарт и выдачи на перфокарты (для исправления программ вне ЦВМ) с быстродействием 300 карт/мин.

6. Печатающее устройство для распечатки результатов и отладочного материала.

7. Накопитель на магнитной ленте для хранения больших программ и массивов чисел.

8. Устройство считывания с перфолент для ввода программ, поставляемых изготовителем машины.

Операционная система — дисковая; она хранится на диске в защищенной от записи зоне объемом 48 К слов. Резидентная программа-загрузчик объемом в 15 слов, прошитая на магнитных сердечниках, обеспечивает удобный доступ к программам, хранящимся на диске. Весьма сложная система графического отображения позволяет использовать кинескоп с шаговой разверткой для отображения текста, колебаний и различной графической информации. Графические изображения результатов на бумаге могут быть получены через центральный процессор после предварительного вывода данных из оперативной памяти на магнитную ленту.

Как уже отмечалось, для эффективного использования быстродействующего процессора (подобного описанному в данной главе) необходимо иметь хорошую систему ввода—вывода, обеспечивающую отладку и работу с программами. В этом отношении разнообразные периферийные устройства, перечисленные выше, оказались весьма полезными.

ЛИТЕРАТУРА

1. Gold B., Lebow I. L., McHugh P. G., Rader C. M., The FDP, A Fast Programmable Signal Processor, *IEEE Trans. on Computers*, C-20, 33—38 (Jan. 1971).
2. Hornbuckle G. D., Ancona E. I., The LX-1 Microprocessor and its Application to Real-Time Signal Processing, *IEEE Trans. on Computers* (Aug. 1970).
3. Gschwind H. W., Design of Digital Computers, Springer-Verlag, Austria, 1967, pp. 235—243.
4. Blankenship P., Gold B., McHugh P., Weinstein C. J., Design Study of the Advanced Signal Processor, Lincoln Lab. Tech. Note, 1972-17, 1972.

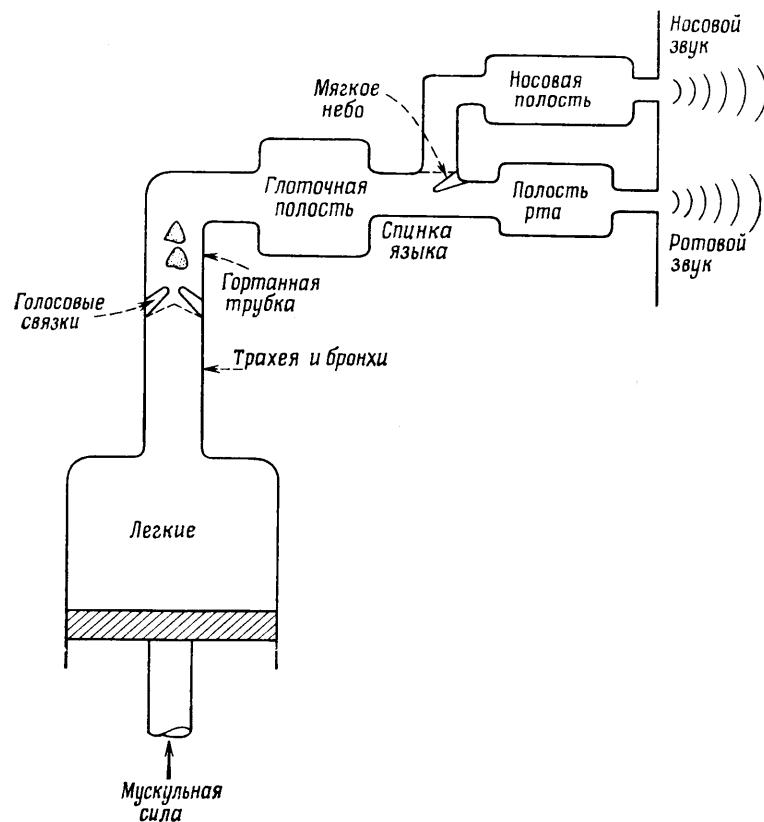
ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ

12.1. Введение

Одной из наиболее важных областей применения цифровых методов является обработка речевых сигналов. Фактически значительная часть теоретических результатов, составляющих основу цифровых методов обработки сигналов, была получена исследователями, изучавшими речевые сигналы. Ниже мы увидим, что цифровая обработка использовалась для решения широкого круга вопросов, включая спектральный анализ, полосные вокодеры, гомоморфные системы обработки, синтезаторы речи, системы линейного прогнозирования и системы голосового управления вычислительными машинами. В данной главе достаточно подробно рассмотрено несколько типичных речевых систем, при создании которых важную роль сыграла цифровая обработка. В главу включены примеры как аппаратурной, так и программной реализации систем обработки речи. Прежде чем перейти к конкретным примерам, будет дан обстоятельный обзор моделей образования речи.

12.2. Модель образования речи

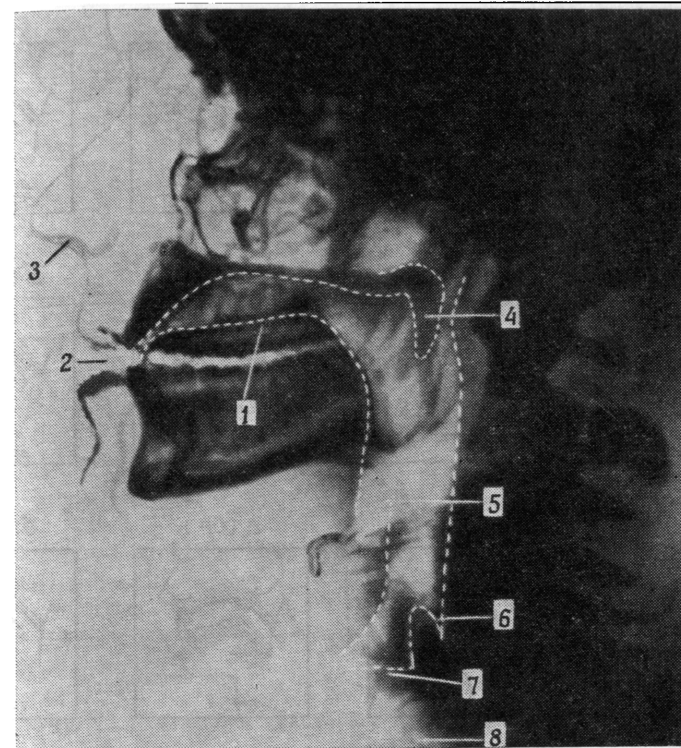
На фиг. 12.1. изображена схема, описывающая механизм образования речи в человеческом организме. При разговоре грудная клетка расширяется и сжимается, прокачивая поток воздуха из легких по трахее через голосовую щель. Если голосовые связки напряжены, как при образовании звонких звуков типа гласных, то они вибрируют подобно релаксационному генератору и модулируют поток воздуха, превращая его в короткие импульсы (порции). Если голосовые связки расслаблены, воздух свободно проходит через голосовую щель, не подвергаясь модуляции. Воздушный поток проходит через глоточную полость мимо основания языка и в зависимости от положения мягкого нёба — через ротовую и (или) носовую полости. Поток воздуха выходит наружу через рот или нос (или обоими путями) и воспринимается как речь. В случае глухих звуков, таких, как *s* в слове *snow* или *r* в слове *pit*, голосовые связки расслаблены. При этом возможны два



Фиг. 12.1. Схема механизма образования речи (по Фланагану).

режима: либо образуется турбулентный поток, когда воздух проходит через сужение в голосовом тракте (как при образовании звука *s*), либо возникает короткий взрывной процесс, вызванный повышенным давлением воздуха за точкой перекрытия голосового тракта (как в звуке *p*). При изменении положения артикуляторов (губ, языка, челюсти, мягкого неба) во время произнесения непрерывной речи форма отдельных полостей голосового тракта существенно меняется. На рентгеновском снимке (фиг. 12.2) голосового тракта мужчины показано положение некоторых артикуляторов и пунктиром обведены контуры отдельных полостей.

Голосовой тракт человека представляет собой неоднородную акустическую трубку, простирающуюся от голосовой щели до губ. У взрослого мужчины она имеет длину около 17 см и частота



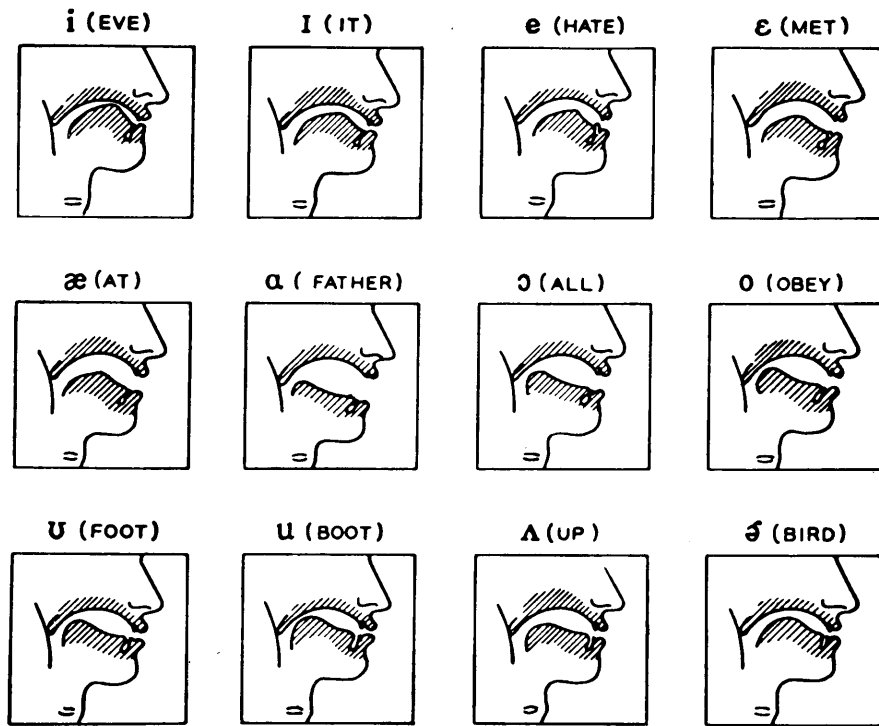
Фиг. 12.2. Рентгеновский снимок голосового тракта мужчины (по Фланагану).

1 — язык; 2 — рот; 3 — ноздри; 4 — мягкое небо; 5 — надгортанный хрящ; 6 — кольцеобразный хрящ; 7 — голосовая щель; 8 — трахея.

ее первого (четвертьволнового) резонанса равна

$$F_1 = \frac{1}{4} \frac{c}{l} = \frac{1}{4} \frac{34000 \text{ см/с}}{17 \text{ см}} = 500 \text{ Гц.} \quad (12.1)$$

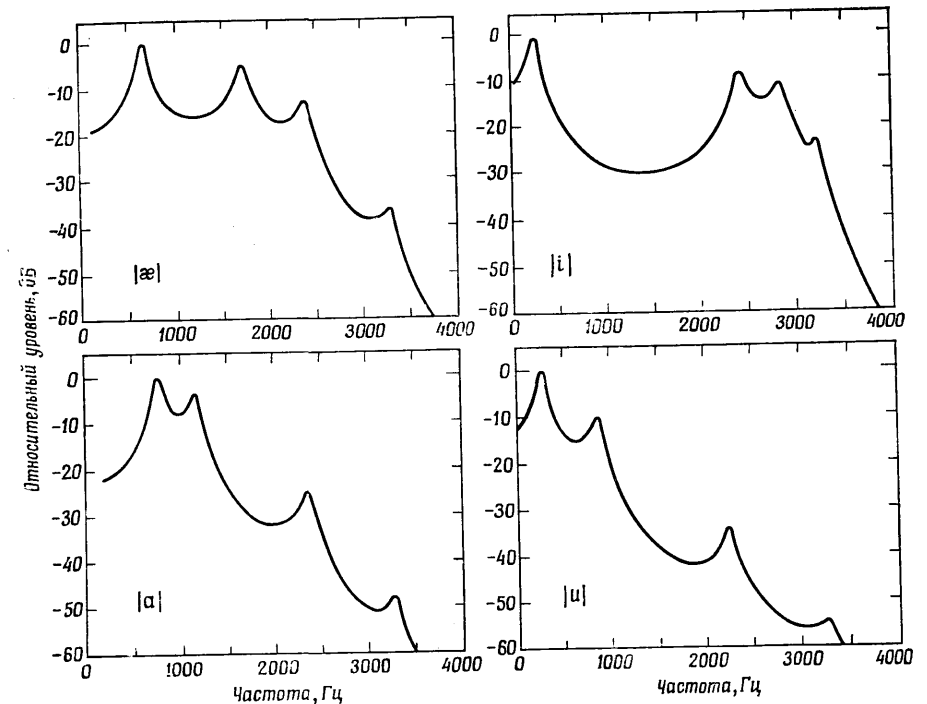
Площадь поперечного сечения акустической трубки неодинакова и зависит от положения артикуляторов, изменяясь от 0 до 20 см². Голосовой тракт имеет некоторые устойчивые виды резонансных колебаний, называемые *формантами*, которые существенно зависят от расположения артикуляторов. Профили голосового тракта для некоторых гласных и типичные для них значения частот первых трех формант (в Гц) схематически представлены на фиг. 12.3. Спектры этих гласных показаны на фиг. 12.4. Они содержат отчетливые резонансы. Полезно отметить, что при восприятии звуков на слух основную роль играют только первые три форманты,



Форматные частоты

Машинный символ гласного звука	Символ по системе IPA	Типичное слово	F_1	F_2	F_3
IY	i	(beet)	270	2290	3010
I	I	(bit)	390	1990	2550
E	ε	(bet)	530	1840	2480
AE	æ	(bat)	660	1720	2410
UH	Λ	(but)	520	1190	2390
A	a	(hot)	730	1090	2440
OW	ɔ	(bought)	570	840	2410
U	u	(food)	440	1020	2240
OO	μ	(boot)	300	870	2240
ER	ε̄	(bird)	490	1350	1690

Фиг. 12.3. Схематические профили голосового тракта и таблица формантных частот для некоторых гласных (по Фланагану).

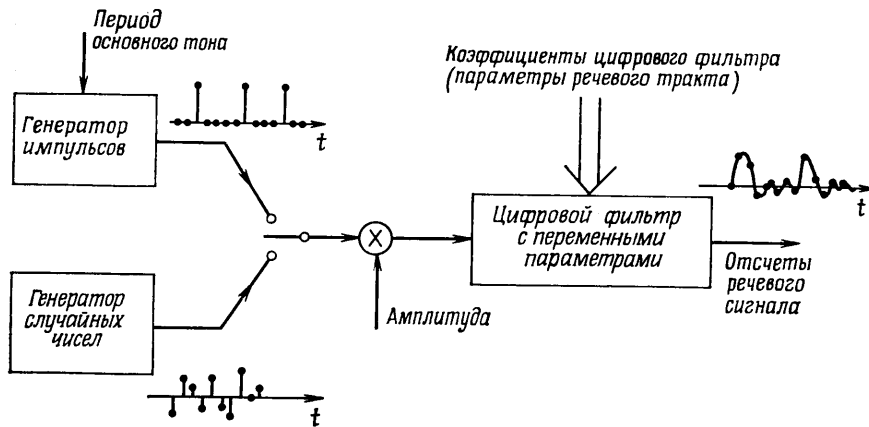


Фиг. 12.4. Спектры некоторых гласных (по Фланагану).

хотя высшие форманты и необходимы для обеспечения качества звучания. На этом факте основана работа нескольких систем сжатия полосы речевого сигнала, которые будут рассмотрены ниже.

Как уже упоминалось, существуют три основных механизма возбуждения голосового тракта. Для звонких звуков источник находится в голосовой щели и состоит из широкополосных квазипериодических порций воздуха, формируемых колеблющимися голосовыми связками. Для глухих звуков типа *s* источник находится в точке сужения голосового тракта и является турбулентным квазислучайным воздушным потоком. И наконец, для глухих звуков типа *p* (как в слове *pop*) источник расположен в точке перекрытия голосового тракта и образуется под действием скачка давления за точкой перекрытия тракта.

Предположение о независимости вида источника возбуждения и характеристик голосового тракта является основным почти для всех систем обработки речи. Именно эта независимость источника и тракта и позволяет ввести передаточную функцию голосового тракта и рассматривать его возбуждение любым из трех источников.



Фиг. 12.5. Цифровая модель образования речи (по Шафэру).

В большинстве случаев это предположение вполне допустимо. Однако в некоторых случаях (например, для глухих взрывных звуков, таких, как p в слове *pot*) оно неверно, и основная модель образования речи становится непригодной. В большей части данной главы будем считать, что предположение о независимости источника и тракта справедливо. В этом случае можно построить простую цифровую модель образования речи (фиг. 12.5). Источниками возбуждения служат генератор импульсов с внешней синхронизацией с периодом основного тона, а также генератор случайных чисел. Генератор импульсов через каждые N_0 отсчетов вырабатывает импульс, соответствующий очередной порции воздуха. Интервал между импульсами называется периодом основного тона. Он равен величине, обратной частоте следования порций воздуха или частоте колебания голосовых связок. Выходная последовательность генератора случайных чисел имитирует и квазислучайный турбулентный поток, и спад давления при образовании глухих звуков.

Каждый из источников (или оба) может быть соединен со входом линейного цифрового фильтра с переменными параметрами, моделирующего голосовой тракт. При этом коэффициенты фильтра отражают свойства голосового тракта в зависимости от времени при непрерывной речи. В среднем через каждые 10 мс коэффициенты фильтра изменяются, отражая тем самым изменение состояния голосового тракта.

Регулировка усиления, введенная между источниками и фильтром, позволяет управлять громкостью выходного сигнала. Последовательность на выходе фильтра эквивалентна речевому сигналу, дискретизованному с соответствующей частотой.

Для управления такой моделью необходимо знать зависимость соответствующих параметров (частоты основного тона, положения переключателя, громкости и коэффициентов фильтра) от времени. Основной задачей почти всех систем анализа речи является оценка параметров модели по реальной речи. Задача большинства систем синтеза речи состоит в том, чтобы, используя эти параметры, полученные некоторым способом, образовать искусственный речевой сигнал, неотличимый на слух от настоящей речи. В системах анализа—синтеза эти две задачи решаются совместно с общей целью увеличения *эффективности* (т. е. понижения частоты дискретизации в системе синтеза до величины, меньшей, чем при обычном представлении речевых сигналов) и *гибкости* (т. е. возможности изменять речь некоторым желаемым образом путем управления параметрами модели). В последующих разделах этой главы обсуждаются различные аспекты нескольких систем, разработанных с учетом этих соображений.

12.3. Кратковременный спектральный анализ

Преобразование Фурье последовательности $x(nT)$, $-\infty < n < \infty$, определяется как

$$X(e^{j\omega T}) = \sum_{n=-\infty}^{\infty} x(nT) e^{-j\omega nT}. \quad (12.2)$$

Как было показано в гл. 6, для нестационарных сигналов типа речевых сигналов преобразование Фурье не имеет смысла, так как спектр речи изменяется во времени. Более полезной характеристикой распределения энергии речевого сигнала является преобразование Фурье на коротком интервале, определяемое как

$$X(\omega, nT) = \sum_{r=-\infty}^n x(rT) h(nT - rT) e^{-j\omega rT}. \quad (12.3)$$

Равенство (12.3) можно рассматривать как фурье-преобразование речевого сигнала на бесконечном интервале, если выделить вблизи момента времени nT участок конечной длины с помощью весовой функции («временного окна») вида $h(nT)$ (фиг. 12.6). Используя свертку, равенство (12.3) можно записать иначе:

$$X(\omega, nT) = [x(nT) e^{-j\omega nT}] * h(nT). \quad (12.4)$$

Левую часть равенства (12.3) можно представить в виде

$$X(\omega, nT) = a(\omega, nT) - jb(\omega, nT), \quad (12.5)$$