

ОБ ОБМЕНЕ УСТНОЙ ИНФОРМАЦИЕЙ МЕЖДУ ЧЕЛОВЕКОМ
И ВЫЧИСЛИТЕЛЬНЫМИ СИСТЕМАМИ

Н.Г. Загоруйко

При разработке вычислительных систем (ВС) производительностью свыше миллиарда операций в секунду и объемом памяти в 10^{10} двоичных знаков [1] возникает проблема коренного усовершенствования средств связи вычислительных устройств с внешними источниками и потребителями информации, в первую очередь с человеком. Использование существующих устройств ввода и вывода информации, обладающих ограниченными возможностями и малой производительностью, может значительно снизить эффективность ВС и сузить круг решаемых задач. В связи с этим появляется необходимость в создании таких вводных и выводных устройств, которые позволяли бы ВС обмениваться информацией в удобной для человека форме, в частности, в виде зрительных и звуковых образов.

Сравнение сигналов, несущих световую и звуковую информацию, и строения зрительных и слуховых анализаторов человека указывает на то, что звуковая информация, при всей своей сложности, имеет значительно более простую структуру, чем зрительная. Поэтому кажется более вероятным, что решение проблемы автоматического опознавания звуковых образов (в том числе человеческой речи) будет найдено раньше, чем зрительных (если под последними понимать не только сравнительно простые статические изображения). Более того, широкий интерес к проблеме

опознавания звуковых образов позволяет надеяться, что практически важные результаты в этой области будут получены уже в ближайшие годы. В расчете на это построены требования к вводным и выводным устройствам не только ВС, но и других проектируемых в настоящее время высокопроизводительных ЭВМ, в частности ЭВМ Стэнфордского исследовательского института (США), параметры которой близки к параметрам ВС [2].

Большая мощность ВС и автоматизированная технология ее изготовления из микроминиатюрных элементов, которая снижает стоимость единицы оборудования, позволят возложить на ВС всю работу по анализу и синтезу звуковых образов, не прибегая к использованию дополнительных аналоговых устройств (за исключением, естественно, электроакустических преобразователей).

Поэтому, конечная цель работы над проблемой обмена звуковой информацией между человеком и ВС должна заключаться в разработке главным образом не аппаратуры, а методов (алгоритмов и программ), которые позволили бы вести свободный обмен звуковой информацией в реальном масштабе времени, без использования дополнительных аналоговых устройств и при незначительных затратах мощности ВС.

Решение проблемы автоматического опознавания информации, содержащейся в речевом сигнале, будет играть большую роль не только при построении ВС, но и в дальнейшем развитии различных областей современной науки и техники. Практически теми же методами смогут быть решены такие задачи, как диагностика заболеваний по фонокардиограммам, опознавание неисправностей механизмов по характеру их шумов, ряд задач гидроакустики, экспериментальной фонетики и т.п. В настоящее время большие усилия направлены на решение этой проблемы для компрессии речевых сигналов, передаваемых по линиям связи, для построения автоматов, управляемых устными командами, для ввода устных команд в современные ЭВМ и т.д. *)

В отличие от ВС многие области современной техники могли бы быть удовлетворены решением частных задач этой проблемы, т.е. разработкой методов и устройств опознавания сравнительно небольшого набора слов (от нескольких десятков до нескольких сотен), при ограничении стилем произношения, числа дикторов и надежности опознавания. Поэтому работу над поставленной про-

*) Обзор работ по автоматическому опознаванию речевых сигналов содержится в [3], [4], [5].

блемой целесообразно построить так, чтобы в качестве промежуточных результатов могли быть получены решения перечисленных выше частных задач.

Методика исследования

Знакомство с работами по автоматическому опознаванию речевых сигналов показывает, что, несмотря на значительные усилия, предпринимаемые в последние годы, результаты работ пока незначительны. Объясняется это главным образом тем, что во многих случаях использовались методы и средства, недостаточно эффективные для решения столь сложной проблемы.

В большинстве работ делались попытки ограничиться малым числом "ключевых" параметров для опознавания речевых сигналов, причем в качестве объекта опознавания выбирался какой-нибудь один элемент речи (чаще всего фонема или слово), без учета взаимосвязи между ним и другими элементами речи. К такой методике многие исследователи вынуждены были прибегать из-за того, что для экспериментов использовалась узкоспециализированная аналоговая аппаратура.

Построение аппаратуры, предназначенной для комплексного исследования речевых сигналов с учетом сложной взаимосвязи между элементами речи и большого числа их параметров, связано со значительными техническими трудностями. Кроме того, использование специализированной аппаратуры чрезвычайно затрудняет всякую перестройку программы исследования. Электронные вычислительные машины, мощность и универсальность которых снимает эти трудности, многими исследователями либо совсем не использовались, либо использовались лишь на отдельных этапах работы.

Большое количество элементов устной речи, широкое разнообразие их форм, тесная взаимосвязь и иерархическая подчиненность их друг другу, а также сложный комплекс их физических характеристик дают основание отнести систему устного языка к разряду сложных систем. Поэтому, по нашему мнению, работу над автоматическим опознаванием речевых сигналов необходимо основывать на методах разработки сложных систем [6].

В свете этого основные черты методики исследования речевых сигналов сводятся к следующему:

1. Основным орудием эксперимента на всех этапах разработки является универсальная электронная вычислительная машина.

2. Объектом опознавания являются сочетания всех элементов устного языка (синтагм, слов, слогов, фонем и сегментов).

3. Характеристикой, по которой производится опознавание речевых элементов различных иерархических уровней, является система параметров сигнала, выбранная с помощью оценки её информационной эффективности.

Рассмотрим содержание каждого из этих пунктов.

1. Электронные вычислительные машины как основное орудие эксперимента на всех этапах работы

Необходимость применения ЭВМ объясняется, в первую очередь, тем, что такие сложные и решаемые в едином комплексе задачи, как:

накопление и статистическая обработка большого количества информации;

получение различных характеристик сигнала с последующим анализом динамики их изменения;

выбор оптимальных гипотез опознавания путем проверки большого их количества и другие

практически не поддаются решению без использования ЭВМ, обладающих высоким быстродействием и достаточно большим объемом памяти.

В отличие от использования специализированных устройств, применение универсальных ЭВМ позволяет вести исследование широким фронтом, значительно облегчая переход от одной системы параметров к другой и от одного набора объектов опознавания к другому.

Ориентация на ЭВМ потребовала разработки алгоритмов и программ, с помощью которых можно было бы моделировать любой известный метод обработки речевых сигналов, выполняемый в настоящее время специальной аналоговой аппаратурой или вручную.

Эта задача к настоящему времени, в основном, решена. Под руководством автора разработаны программы вычисления следующих параметров речевого сигнала:

- энергии сигнала за время T ;
- огибающей сигнала за время T ;
- частоты перехода сигнала через нуль на промежутке времени T (для исходного $\dot{x}$ проинтегрированного сигнала);
- распределения промежутков между нулями клипированной речи на отрезке времени T (для исходного $\dot{x}$ проинтегрированного сигнала);

- функции автокорреляции и взаимной корреляции;
- спектрального состава речевого сигнала.

Кроме того, для проведения экспериментов потребовались программы, которые реализуют алгоритмы, ранее не использовавшиеся для обработки сигналов речи. К настоящему времени разработаны программы получения и обработки инфразвуковой части спектра речи, проведения экспериментов по членению речевого сигнала на квазистационарные участки, определения информационной эффективности независимых параметров.

При машинных методах выделения параметров речевых сигналов появляется возможность оценивать точность получаемого результата, без чего нельзя судить о достоверности получаемых статистических закономерностей и достоверности опознавания элементов речи. Исследованию погрешностей, возникающих при выделении некоторых параметров речевых сигналов на ЭВМ, посвящены работы [8] и [9].

Проведение исследований речевых сигналов с помощью ЭВМ предполагает разработку технических средств для ввода звуковой информации в вычислительную машину. С этой целью разработан преобразователь аналог-цифра с частотой отсчетов сигнала от 0 до 27 тысяч в секунду при числе двоичных разрядов в отсчете равном 8 [10].

2. Сочетание элементов устного языка различных иерархических уровней как объект опознавания

Надежное опознавание в реальных условиях, т.е. при наличии помех и искажений, вносимых аппаратурой и дикторами, возможно только за счет избыточности, содержащейся в достаточно крупных элементах речи (например, фразах или синтагмах).

Однако использование в качестве объекта опознавания только крупных элементов речи привело бы к очень большому "словарю", так как число возможных фраз и синтагм чрезвычайно велико.

Использование только мелких элементов речи (фонем и сегментов) представляет собой другую крайность. "Словарь" образов в этом случае будет небольшим, но надежность выделения и опознавания этих элементов в слитной речи высокой быть не может из-за их слабой помехоустойчивости.

Слова и слоги более устойчивы, чем фонемы. Но так как в беглой речи части слов могут сильно искажаться и даже полностью пропадать, то нельзя надеяться на полное решение пробле-

ни опознавания речи, если в качестве объектов будут выбраны только элементы этого иерархического уровня.

Наиболее целесообразным является выбор в качестве объекта опознавания определенного сочетания всех упомянутых элементов речи, а именно: "словаря" фраз и синтагм для опознавания часто встречающихся словосочетаний; словаря широко распространенных слов; "словаря" слогов для опознавания редко встречающихся слов и "словаря" фонем или даже сегментов для опознавания редких звуко сочетаний.

При решении частных задач (например, при малом объеме словаря), вероятно, будет целесообразно ограничиться не полным набором элементов, а некоторой их частью. Переход к общей проблеме будет сопровождаться расширением набора элементов.

Вопрос о том, из какого количества и каких конкретно фраз, синтагм, слов, слогов, фонем и сегментов будет составлен "словарь" опознаваемых элементов, должен быть решен на основании изучения статистических законов их распределения. Необходимо стремиться к минимуму общего числа объектов опознавания. Но при этом следует учитывать возможность автоматического членения сигнала слитной речи на выбранные элементы и возможность их надежного опознавания. Конкретный состав набора элементов будет зависеть от области применения ВС. В памяти ВС, работающих, например, в определенной области науки, целесообразно иметь, в первую очередь, "словарь" речевых элементов, характерных для этой области.

В настоящее время в Институте математики СО АН СССР ведется статистическое исследование различных элементов речи.

Результаты анализа слогового состава русского языка приведены в работе [7].

3. Система параметров как характеристика для опознавания элементов речевого сигнала

Одновременное рассмотрение элементов различных иерархических уровней по-новому ставит задачу выбора параметров речевых сигналов и существенно меняет требования к величине информации, извлекаемой с помощью каждого параметра.

Действительно, нет необходимости стремиться к предельной надежности опознавания элементов какого-нибудь одного уровня, так как определенная доля информации об этих элементах может

быть извлечена из элементов другого уровня при меньших затратах.

Критерием правильности выбора системы параметров и правильности использования каждого параметра может служить величина их информационной эффективности.

Под информационной эффективностью системы N параметров подразумевается величина

$$R = \frac{I_{\Sigma}}{n_{11} + n_{12} + \dots + n_{ij} + \dots + n_{MN}},$$

где n_{ij} - количество операций, затрачиваемое на измерение i -го параметра элементов j -го уровня, а I_{Σ} - суммарное количество информации об элементах всех M уровней, извлекаемое с помощью данной системы параметров.

Система параметров будет достаточно сложной, так как однотипные параметры, несмотря на очевидную целесообразность их использования, не могут быть одинаково эффективными для опознавания элементов разных уровней.

Кроме физических параметров сигнала, вероятно, окажется целесообразным привлекать и лингвистические характеристики языка (априорную вероятность появления элементов и их сочетаний и т.д.).

Эффективность системы параметров зависит как от эффективности каждого параметра в отдельности, так и от степени их взаимной зависимости.

Оценка эффективности системы зависимых параметров может быть получена путем изучения их условных пространственных распределений для всех объектов опознавания.

Трудность этой задачи заключается в том, что для её решения требуется произвести обработку большого объема исходных данных. Еще большую трудность представляет практическое получение достоверных условных распределений сложной системы параметров. К сожалению, большой статистический материал, накопленный к настоящему времени различными организациями, не может быть использован для этих целей, так как многие параметры исследовались в предположении, что они независимы, и часто на различном исходном материале.

В ряде случаев важно иметь оценку эффективности отдельного параметра, рассматриваемого в качестве независимого. Это оправдано в случае использования действительно независимых параметров (например, длительности слова и его спектрального состава) или когда заведомо зависимые параметры используются

как независимые из-за ограниченных возможностей опознающей аппаратуры.

Методика оценки информационной эффективности независимо-го параметра речевых сигналов изложена в [II].

4. Этапы исследования

На наш взгляд, исследование проблемы целесообразно вести в последовательности, которую можно расчленить на три этапа.

На первом этапе:

- разрабатываются алгоритмы и программы, предназначенные для обработки речевых сигналов на ЭВМ, и оцениваются погрешности, вносимые машинными методами обработки сигналов;

- проводятся статистические исследования элементов устной речи и поиск машинных методов членения слитной речи на различные элементы;

- разрабатывается методика оценки информационной эффективности параметров речевых сигналов и выбора оптимальной точности измерения параметров;

- разрабатывается (в качестве частной задачи) практически приемлемая методика опознавания небольшого количества устных команд (20 + 30 слов) с некоторыми ограничениями на дикцию и стиль произношения (последний должен быть полным). Объектом опознавания выбирается отдельное слово, усредненные параметры которого служат характеристиками для его опознавания.

На втором этапе:

- завершается обработка статистики элементов русского устного языка и поиск машинных методов членения слитной речи на элементы различного типа;

- заканчивается разработка методики оценки информационной эффективности системы зависимых параметров, заданных их условными распределениями, и методики выбора оптимальной точности измерения этих параметров;

- проводятся работы по синтезу речевых сигналов, главным образом, для проверки гипотез анализа ("анализ через синтез");

- разрабатывается методика анализа устной речи с объемом словаря 200-500 слов без существенных ограничений на дикцию и стиль произношения. Соответствующим образом выбирается набор опознаваемых элементов речи и наиболее эффективная система их параметров.

На третьем этапе исследования должна быть разработана ме-

тодика обмена устной информацией между ВС и человеком в реальном масштабе времени без использования дополнительных устройств, при незначительных затратах мощности ВС и без существенных ограничений на объем словаря, дикцию и стиль произношения.

Заметим, что на втором и особенно на третьем этапе формальный подход к решению проблемы анализа и синтеза речевых сигналов потребует сочетать с бионическим. Подготовку к этому, в частности создание экспериментальной базы и систематизацию литературных материалов по исследованию слухового анализатора и центров речеобразования, следует начинать уже на первом этапе.

На завершающей стадии работа над анализом речевых сигналов должна проводиться в тесном сотрудничестве со специалистами в области математической лингвистики, машинного перевода и других смежных научных направлений.

В ы в о д ы

1. В связи с разработкой ВС возникает необходимость в создании вводных и выводных устройств, способных обеспечивать обмен устной информацией между ВС и человеком.

2. Мощность ВС позволяет выполнять на ней всю работу по опознаванию звуковых образов без использования дополнительных устройств, поэтому основные усилия должны быть направлены на разработку машинных методов анализа и синтеза речевых сигналов.

3. Система устной речи относится к категории сложных систем. Методика её исследования, как сложной системы, должна основываться на использовании:

- а) ЭВМ как главного орудия эксперимента на всех этапах работы;

- б) сочетания речевых элементов различного иерархического уровня в качестве объекта опознавания;

- в) системы параметров с максимальной информационной эффективностью в качестве характеристики для опознавания речевых элементов всех выбранных иерархических уровней.

1. Евреинов Э.В. и Косарев Ю.Г. О возможности построения вычислительных систем высокой производительности. Изд-во СО АН СССР, 1962 год, г. Новосибирск.
2. Advances in Computers, vol. 2, p. 135. Academic Press, N.Y.L., 1961.
3. Сапожков М.А. Речевой сигнал в кибернетике и связи. Москва, Связьиздат, 1963.
4. Цеммель Г.И. Автоматическое (объективное) распознавание звуков речи (обзор). Зарубежная радиоэлектроника, 1961, № 4, стр. 52-73.
5. Advances in Computers, vol. 1, p. 83. Academic Press, N.Y.L., 1961.
6. Евреинов Э.В. и Косарев Ю.Г. О методике разработки сложных систем. Сб. "Вычислительные системы", вып. 6, Новосибирск, 1963.
7. Ёлкина В.Н. и Юдина Л.С. Статистика слогов русской речи (данный сборник).
8. Загоруйко Н.Г. Об ошибках вычисления энергии и огибающей речевых сигналов с помощью ЭВМ (данный сборник).
9. Волошин Г.Я. Спектральный анализ речевых сигналов на ЭВМ (данный сборник).
10. Волошин Г.Я. Преобразователь аналог-цифра для ввода речевых сигналов в ЭВМ (данный сборник).
11. Загоруйко Н.Г. Методика оценки информационной эффективности независимых параметров речевого сигнала (данный сборник).