

ИНТЕРАКТИВНАЯ ЛАБОРАТОРНАЯ СИСТЕМА ДЛЯ АНАЛИЗА И ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ

С.Л. Гончаров, В.Я. Чучупал

Вычислительный центр
Академия наук СССР
Москва, СССР

РЕЗЮМЕ

Доклад содержит описание программного обеспечения интерактивной системы для анализа и цифровой обработки речевых сигналов на малой ЭВМ общего назначения, оснащенной графическим дисплеем. Система представляет собой комплекс программ, в который входят как библиотека процедур, реализующих выполнение стандартных функций по обработке сигналов, так и программы, позволяющие, в диалоговом режиме, с использованием графического изображения речевого сигнала и его параметров на экране дисплея выполнять анализ речевых сигналов, выделять значения их параметров и сохранять эти значения в базе данных, а также выполнять некоторые функции по обработке речевых сигналов, например, производить анализ зашумленных фонограмм с целью улучшения их качества.

ВВЕДЕНИЕ

Исследования в области автоматического распознавания, цифровой обработки сигналов и экспериментальной фонетике зачастую носят трудоемкий и рутинный характер, так как связаны с большим объемом ручной работы, требующейся, например, при подготовке экспериментального материала, а также оценке результатов работы. С начала 80-х годов в Вычислительном центре Академии наук СССР разрабатывается специализированное программное обеспечение для цифровой обработки речевых сигналов на малых ЭВМ. Целью разработки является максимальное облегчение усилий пользователей при программировании и отладке процедур анализа и обработки речевых сигналов, в частности, создание программного обеспечения рабочего места для анализа речевых сигналов. Полученные к настоящему времени в результате проделанной работы программные средства оправдали ожидания разработчиков, как существенно повысив производительность труда специалистов, так и выполнять исследования, ранее практически невозможные.

АРХИТЕКТУРА СИСТЕМЫ

На рис. 1 изображена схематически архитектура системы, включая как программное обеспечение, так и аппаратную часть. Технической базой для рабочего места явился измерительно-вычислительный комплекс, включающий в себя мини ЭВМ, аппаратуру ввода-вывода речевых сигналов и графический дисплей. Всё разработанное программное обеспечение функционирует в среде операционной системы реального времени. Программное обеспечение построено по иерархическому принципу. В этом смысле систему можно рассматривать как совокупность четырех основных компонент - "уровней". Уровни организованы таким образом, что модули верхних уровней ссылаются при работе на модули нижних уровней и могут обмениваться информацией с ними; однако при этом передаваемые параметры стандартизированы так, что структура модулей нижнего уровня остается скрытой от модулей верхнего уровня. Модулям нижних уровней, в свою очередь, недоступна информация о существовании более высоких уровней. Внутри каждого уровня соблюдался модульный принцип построения программ, в соответствии с которым каждая программа выполняет свою, достаточно автономную функцию обработки и только ее. Подобная архитектура обеспечила большую степень независимости программного обеспечения, позволив сравнительно легко приспособлять его как к новым аппаратурным ресурсам, так и к новым областям применения.

БАЗОВОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Эта часть программного обеспечения представляет собой совокупность пяти библиотек (объектных модулей), на основе которых создано всё программное обеспечение более высокого уровня. Библиотека стандартных подпрограмм цифровой обработки речевых сигналов содержит стандартных подпрограмм численных преобразований (алгоритмы вычисления БПФ для комплексных, действительных и целочисленных данных, оценки спектральных характеристик сигнала, огибающей спектра, частоты основного тона, параметров линейного предсказания речевого сигнала, синтеза цифровых фильтров и фильтрации речевых

сигналов и т.п.). Библиотека подпрограмм матричной арифметики включает набор процедур для выполнения матричных операций на мини ЭВМ, не имеющей спецпроцессора. Использование этих подпрограмм позволяет в 4-6 раз ускорить процесс вычислений по сравнению с обычными процедурами, написанными, например, на языке Фортран-4. Библиотека подпрограмм ввода-вывода речевых сигналов предоставляет средства для обслуживания АЦП и ЦАП. При этом ввод (вывод) данных может производиться как в оперативную память ЭВМ, так и в указанный файл на носителе прямого доступа. Речевой сигнал в последнем случае хранится в формате обычного (двоичного) файла подсистемы управления файлами операционной системы.

Набор подпрограмм управления окнами включает в себя процедуры, предоставляющие наиболее удобный интерфейс между алгоритмами обработки речевых сигналов (использующих обычно пошаговую обработку данных) и самими данными, содержащимися в файлах на носителе прямого доступа. В алгоритмах обработки речевых сигналов оценка значений параметров сигнала и обработка сигнала обычно выполняются в пошаговом режиме, когда речевой сигнал рассматривается как последовательность перекрывающихся между собой сегментов ("окон") данных. Каждый такой сегмент обрабатывается алгоритмом отдельно и, в большинстве случаев, обработкой сегментов выполняется однотипным образом. С точки зрения подпрограмм управления окнами файл, содержащий речевой сигнал (или файл, состоящий из параметров речевого сигнала) имеет структуру матрицы, каждая строка которой - вектор значений параметров речевого сигнала, соответствующий определенному моменту времени. В частном случае, когда в файле содержится речевой сигнал, число элементов в каждой строке равно 1, то есть это - одномерная матрица. Прикладная программа должна содержать описание используемых ею данных и правила, по которым будет выполняться разбиение файла данных на сегменты для обработки. С этой целью в прикладной программе специальным запросом создается буфер - "окна", в которых содержатся данные. Окна данных могут быть созданы для всех входных, выходных и промежуточных потоков данных алгоритма. Поскольку данные описаны как файлы, то последовательность действий по описанию и управлению потоком данных в алгоритме включает в себя следующие этапы:

- описание окон в прикладной программе;
- установление связи между окнами и отображенными в этом окне файлами;
- инициирование операций передачи данных между окном и связанными с ним файлами в соответствии с требованиями алгоритма.

Программа сообщает размер каждого окна (в секундах), шаг, с которым оно смещается по данным, по сигналу (он может изменяться, как и длина окна в процессе обработки), тип окна (окно для чтения и записи данных или только для чтения или записи), а также ха-

рактеристики отображаемых в этом окне данных: частоту дискретизации, количество параметров и т.п. После того, как программа выдаст запрос на установление связи между созданным ею окном и файлом данных, она может полностью контролировать все функции по обмену данными, с помощью запросов типа "позиционировать окно", "сдвигать окно вверх", "сдвигать окно вниз", "получить время позиционирования окна" и т.п. Когда программа выдает запрос "позиционировать окно", она сообщает точку позиционирования (в секундах от начала записи) и в область окна выполняется передача данных из файла данных, которые соответствуют указанному программой времени. Сдвиг окна на шаг вверх подразумевает смещение данных в буфере таким образом, что начало буфера смещается по файлу данных на промежуток времени, равный шагу смещения окна в сторону возрастания времени. Сдвиг окна вниз аналогичен сдвигу вверх, но при этом окно смещается в сторону уменьшения времени, то есть к началу файла данных. Программа может получить время, которое соответствует началу расположения окна в файле данных выдать запрос "получить время".

При таких условиях работы алгоритм, который реализует пошаговую обработку сигнала от начала до конца, должен просто после очередного цикла обработки сигнала выдавать запрос на сдвиг окна вверх. Описанный выше подход позволил не только резко ускорить процесс создания программы обработки сигналов (включая их отладку), но и обусловил существенную гибкость этих программ за счет отсутствия необходимости в их модификации в случае изменения параметров анализа сигналов (например, длины анализируемого сегмента данных) или характеристик самого сигнала (например, частоты дискретизации). Файлы данных, как уже было упомянуто, могут содержать не только дискретные значения самого речевого сигнала, но также состоять из произвольных параметров речевого сигнала. Этими параметрами могут быть значения параметров модели речеобразования, измеренные на интервалах анализа, так и более сложные лингвистические характеристики, такие как точки начала или конца определенных фонем, границы участков пауз, смечек, звонких или глухих звуков и т.п. Файлы данных, содержащие значения параметров речевых сигналов, полученных в результате кратковременного спектрального анализа, удобно обрабатывать с помощью представленных выше функций управления окнами данных. Однако для организации работы с данными, характеризующими лингвистические характеристики речевого сигнала, более удобным является дополнительно к этому использование специальных функций системы управления базой данных. Основных функций несколько: это поиск какого-то входящего строки параметров (меток или маркеров) в файл параметров, начиная с указанной точки поиска. В этом случае в прикладную программу передаются времена, соответствующие точкам начала и кон-

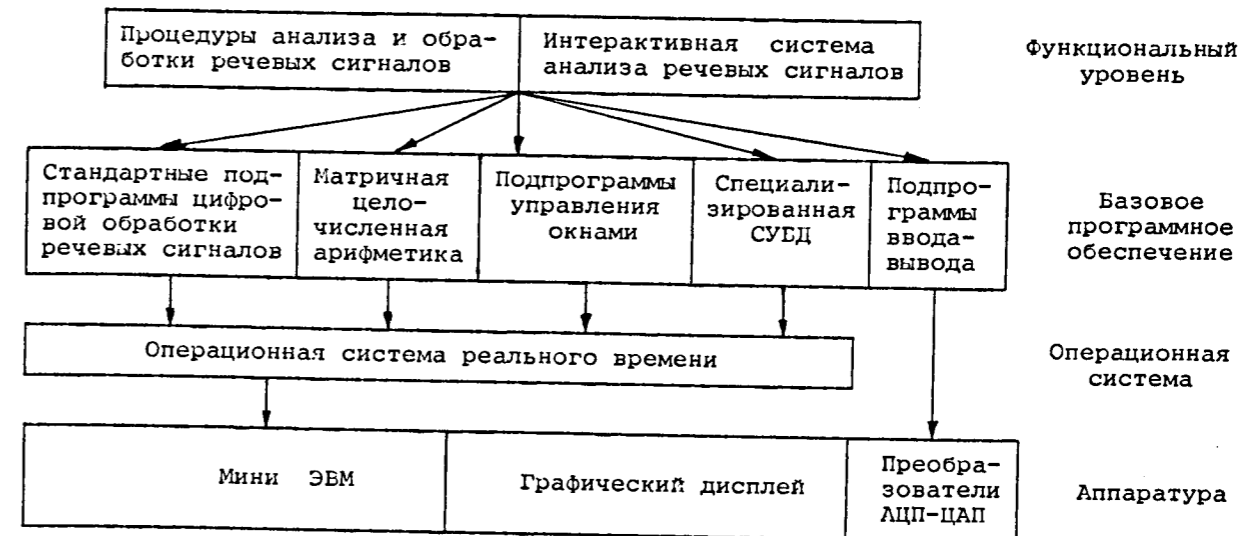


Рис. 1. Иерархическая схема организации диалоговой системы цифровой обработки речевых сигналов.

ца-го вхождения в файл параметров определенной метки или последовательности меток. Таким путем прикладная программа может, например, последовательно получить доступ ко всем сегментам речевого сигнала, содержащим квазистационарный участок ударной гласной "а" или ко всем участкам, соответствующим определенным слогам и т.п. Пользователь может также специальным вызовом занести значение определенного параметра в указанное место (по времени) файла параметров. Сервисные процедуры позволяют получить перечень значений параметров, содержащихся в указанном файле данных. Следует отметить, что базовое программное обеспечение никак не интерпретирует и не использует семантику параметров, содержащихся в файле данных: это целиком определяет сам пользователь.

В "функциональный уровень" (см. рис. 1) попали процедуры, реализующие законченные алгоритмы обработки речевых сигналов. Состав этого уровня зависит от проблемной ориентации системы. В данном случае наиболее интенсивно использовалось программное обеспечение для коррекции зашумленных сигналов. Однако при всех вариантах использования системы в этот уровень целесообразно включать процедуры, обеспечивающие ввод (вывод) речевого сигнала в файлы на дисках ЭВМ, процедуры, обеспечивающие визуализацию речевого сигнала (типа "видимая речь"), процедуры для генерации сигналов специального вида (тестовые сигналы, белый шум, розовый шум, тоновые сигналы), процедуры оценки первичных параметров речевого сигнала, аддитивного наложения сигналов и т.п. Одной из наиболее интенсивно используемых процедур функционального уровня является интерактивная графическая система для анализа речевых сигналов.

ИНТЕРАКТИВНАЯ ГРАФИЧЕСКАЯ СИСТЕМА

Интерактивная графическая система предоставляет пользователю возможность выделения, в диалоговом режиме, информативных параметров речевых сигналов, сохранение полученных таким образом значений в базе данных и использование этих значений при выполнении процедур обработки речевых сигналов. Текущая версия системы включает в себя прикладные программы обработки речевых сигналов, которые позволяют выполнять обработку зашумленных речевых сигналов (фильтрацию сигнала из смеси с аддитивным шумом), а также выполнять в диалоговом режиме разметку речевого сигнала, сохраняя в базе данных сведения об основных параметрах сигнала (как-то: значение признака тон/шум, частоты и амплитуды формант, частоты основного тона и т.п.), что может быть полезным при создании специализированных баз данных для тестирования и отладки устройств и алгоритмов обработки и распознавания речи.

Система дает возможность выполнять пользователю следующие действия:

- создавать на экране графического дисплея изображение различных характеристик речевого сигнала: кратковременного амплитудного спектра, временной формы речевого сигнала, сглаженного амплитудного спектра, контура кратковременной интенсивности;
- производить непосредственно на экране графического дисплея указание положения или значений информативных параметров речевых сигналов и запоминать найденные значения в базе данных;
- управлять просмотром сигнала, перемещая смотровые окна в нужном направлении или позиционируя их в интересующей точке;

- изменять тип отображаемых в данный момент характеристик сигнала;
- прослушивать речевой сигнал в указанных пользователем фрагментах;
- указывать границы участка обработки и обрабатывать выделенный участок с помощью заданных пользователем процедур.

Перед началом работы в системе в диалоговом режиме пользователем устанавливаются основные параметры работы с системой, к которым относятся, например, частота квантования, длина и смещение окна, тип весовой функции (прямоугольное окно или окно Хемминга), наличие или отсутствие коррекции верхних частот и т.п.

Анализ сигналов в интерактивном режиме проходит следующим образом. Изображение на экране дисплея можно рассматривать как совокупность нескольких различных представлений сигнала, которые изображаются соответственно, в различных областях экрана дисплея:

- отображение графика кратковременной интенсивности речевого сигнала (окно 3),
- отображение графика речевого сигнала (окно 2),
- отображение амплитудного спектра сигнала (окно 1).

При работе в системе пользователь должен информировать систему о том, с каким окном он работает в данное время, то есть "активизировать" окно. Остальные окна, во время активности одного из окон, пассивны. Существует набор команд работы с окнами. Выполнение команды происходит по нажатию определенной клавиши на функциональной клавиатуре графического дисплея. Система имеет в своем распоряжении следующие команды работы с окном: позиционирование окна, сдвиг окна вверх (вниз) по сигналу, переход к окну 1 (2 и 3).

После того, как окно активизировано, можно переходить к заданию параметров сигнала в интерактивном режиме. Для задания данных в распоряжении пользователя имеется (в каждом окне свой) курсор. Пользователь может указать интересующую его точку следующими способами:

1. Подогнать курсор в нужное место с помощью команд перемещения курсора.
2. Указать элемент изображения световым пером.

3. Сначала "грубо" указать интересующий нас элемент изображения, а затем "тонко" подправить положение курсора.

Кроме команд перемещения курсора (вверх, вниз, влево, вправо) в распоряжении пользователя находятся следующие команды (эти команды выполняются как нажатием клавиши на функциональной клавиатуре графического дисплея, так и указанием световым пером на команду в меню, расположенном на экране дисплея): фиксация точки, привязка к точке, удаление точки.

В описываемую систему встроены две сервисные программы общего назначения, которые требуются практически при всех видах работ.

Эти программы выполняют следующие функции: копирование - введенная точка интерпретируется как начало копируемого сигнала, а текущее положение курсора как конец участка копирования; вывод - на ЦАП с целью прослушивания или записи на магнитофон выводится сигнал на указываемом пользователем сегменте. Текущая версия системы включает в себя ряд процедур обработки зашумленных речевых сигналов. Эти процедуры реализованы в виде команд системы, задаваемых как с клавиатуры графического дисплея, так и световым пером в меню на экране дисплея. Система имеет следующие команды:

- "шум" (участок от заданной точки до текущего положения курсора используется для оценки априорных сведений о параметре шума),
- "пауза" (аналогичный участок сигнала интерпретируется как пауза. Соответственно интенсивность сигнала снижается на заданную величину),
- "вычитание" (на аналогичном участке происходит вычитание амплитудных спектров),
- "фильтр" (на аналогичном участке выполняется обработка сигналов квазиоптимальным фильтром для стационарной гауссовой помехи. При этом требуется задание еще одной точки - центра фильтрации).

При работе с речевым сигналом с помощью интерактивной системы обычно используются два типа файлов: это файл, содержащий собственно значение речевого сигнала, которые используются для последующей обработки или анализа речевого сигнала, и файл параметров, в который могут заноситься значения параметров сигнала, найденные в процессе его анализа. К таким значениям могут относиться метки акустических событий (начало и конец участков определенных фонем, пауз, смычек, признаки тон/шум и т.п.). Файл параметров используется при решении задачи накопления в базе данных значений параметров, полученных в ходе ручной или полуавтоматической разметки (анализа) речевых сигналов. В настоящее время решается задача автоматического сбора статистических сведений о накопленной таким образом информации, например, с помощью алгоритмов кластерного анализа.

К системе могут присоединяться новые прикладные программы пользователя. При этом они должны находиться на диске в виде загрузочного модуля. Основная программа системы при этом запоминает свое текущее состояние, поэтому возможен возврат из прикладной программы в систему.

ЗАКЛЮЧЕНИЕ

В докладе описано программное обеспечение интерактивной лабораторной системы для анализа и обработки речевых сигналов. Существующий вариант системы направлен на обработку зашумленных сигналов, однако при некоторых изменениях системы она станет бо-

лее универсальной. Разработчики видят следующие возможности применения системы:

- полуавтоматическая разметка сигналов. При этом один из сигналов размечается вручную, остальные такие же сигналы (полученные от разных дикторов) размечаются с использованием алгоритмов динамического программирования,
- создание банков фонем, слогов, слов языка. Эти банки данных могут быть использованы в обучающих системах для автоматического распознавания речи.

Возможны также и другие приложения. Таким образом, мы надеемся, что интерактивная система анализа и обработки речевых сигналов является универсальным инструментом в руках исследователей речи.

ЛИТЕРАТУРА

1. В.Я. Чучупал. Диалоговая система цифровой обработки зашумленных речевых сигналов. Диссертация на соискание ученой степени кандидата физико-математических наук. Москва, 1985.
2. В.Р. Женило, Р.С. Иванова, П.В. Миксов, В.Я. Чучупал. Применение измерительно-вычислительных комплексов для цифровой обработки речевых сигналов. Москва, Вычислительный Центр АН СССР, Сообщения по вычислительной технике, 1985.
3. С.Л. Гончаров, В.Я. Чучупал. Интерактивная лабораторная система для анализа и обработки речевых сигналов. Тезисы докладов и сообщений 14-го Всесоюзного семинара (АРСО-14). 26-28 августа 1986 г. Часть II, с. 70 - 71. Каунас, 1986.