

**СООБЩЕНИЯ
ОБЪЕДИНЕННОГО
ИНСТИТУТА
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ
ДУБНА**

P11-86-778

В.А.Горшков, В.Л.Пахомов

РЕЧЕВОЙ ВЫВОД ДЛЯ МИКРОЭВМ

1986

Для человека речь представляет собой наиболее естественный вид связи. Поэтому идея реализации речевого общения человека с машиной привлекает все большее внимание специалистов. В данной работе рассматривается программное и аппаратное обеспечение системы речевого вывода, созданной на микроЭВМ "Электроника-60".

Авторы ставили перед собой задачу создать простую и достаточно доступную систему речевого вывода с целью использования ее в системе автоматизации проектирования и производства (САПР) печатных плат^{/1/}.

Речевой вывод позволяет разгрузить зрительный канал восприятия информации, который до сего времени является основным при работе с экранными терминалами ЭВМ, что в свою очередь снижает утомляемость оператора и уменьшает количество ошибок.

В то же время данная система позволяет вести профессиональные исследования в области распознавания и синтеза речи, а также может использоваться автономно в составе вычислительных комплексов, экспериментальных установок и роботизированных производств.

При работе с САПР экран серийного цветного телевизора используется для вывода графической информации (через телевизионный драйвер КИ15К (ЛЯР)), а его динамик - для вывода речевых сообщений.

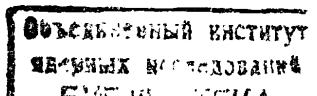
Описание программ

Все программы написаны на языке СИ^{/2/}. В том числе и программы работы с блоками КАМАК, которые обычно пишутся на Ассемблере. Система реализует следующие методы синтеза речи:

- непосредственного кодирования - восстановления,
- линейной дельта-модуляции,
- частотной модуляции^{/3/}.

В начале работы запрашивается количество байт, отводимых для запоминания речи. Затем на терминал выдается меню, по которому можно выбрать нужный вид работы. Состав меню:

- ввод с микрофона,
- вывод на динамик,
- вычисление параметров речи,
- фильтрация,
- сжатие,



- усиление,
- вывод графика на телемонитор,
- дельта-модуляция,
- частотная модуляция,
- запись файла,
- чтение файла.

После выполнения указанной работы меню вновь выдается на терминал.

При непосредственном кодировании-восстановлении входной сигнал можно отфильтровать для удаления шумов, сжать в 2 раза, усилить в 2 раза, записать на диск и т.д. Сжатие заключается в замене двух последовательных значений амплитуды сигнала на их среднее значение. Это преобразование практически не сказывается на качестве воспроизведения речи. Каждое преобразование можно применять многократно.

Вывод графика на телеэкран производится последовательно, порциями по 256 значений амплитуд. При этом можно получить копию изображения с экрана на графической печати СМ-6325.

Работа алгоритма линейной дельта-модуляции заключается в преобразовании значений амплитуды так, чтобы разность любых двух соседних значений не превышала единицы. Запоминание вместо абсолютных значений амплитуд их приращений позволяет сократить требуемый объем памяти в 4 или 8 раз.

При частотной модуляции запрашивается значение уровня обнуления, и затем каждое значение амплитуды сигнала заменяется на одно из трех фиксированных значений a , a_0 , $-a$ по следующему алгоритму. Пусть

a_0 - уровень обнуления, a - фиксированное значение, тогда если $(\text{модуль}(a) < a_0)$, то $a = a_0$, иначе если $(a > a_0)$, то $a = A$, иначе $a = -A$.

Такой клиппированный сигнал звучит вполне разборчиво, но индивидуальные особенности речи диктора уже не слышны. Большая естественная избыточность речевого сигнала позволяет сжать принятую информацию в 30-50 раз и тратить на 1 секунду речи от 200 до 400 байт памяти.

Данная система позволяет создать словарь слов или фраз на диске и затем, используя отдельную простую программу речевого вывода, работать с ним на любой машине, комбинируя предложения любой продолжительности и сложности. Это чисто программная речевая система.

Описание аппаратуры

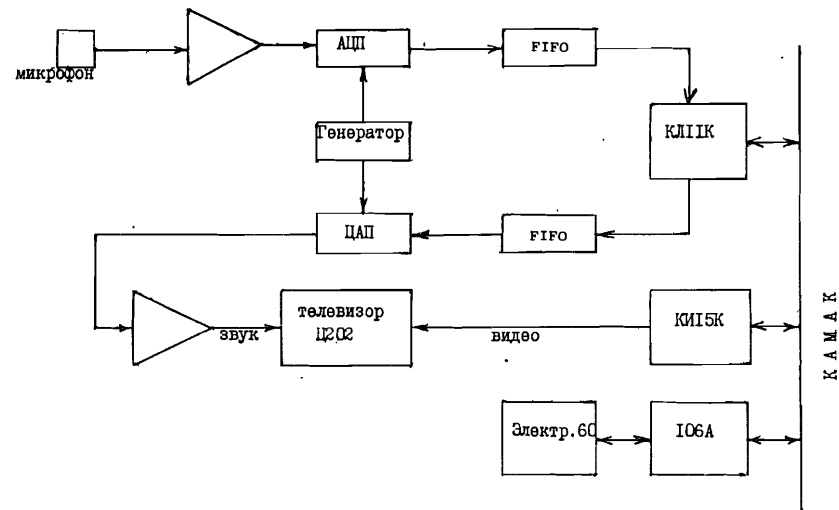
Так как основная мощность речевого сигнала сосредоточена в пределах до 5 кГц, то из теоремы Котельникова следует, что для адекват-

ной выборки речевого сигнала частота выборок должна быть не менее 10 кГц. Частота выборки 10 кГц позволяет максимально использовать возможности программной обработки на ЭВМ и снижает затраты на аппаратуру.

При такой частоте на прием и запоминание или выдачу одного байта можно тратить не более 100 мкс. Но так как оперативная память машины выполнена на микросхемах динамической МДП-памяти, она требует регенерации. Регенерация памяти производится с частотой 500 Гц, и каждый цикл регенерации занимает около 150 мкс. Это приводит к потере информации при вводе и выводе и, как результат, к заметным искажениям речи.

Для ликвидации этих потерь в данной схеме используется память FIFO. Для работы достаточно емкости этой памяти всего в несколько байт.

С микрофона сигнал поступает на корректирующий усилитель - формирователь, где он усиливается с одновременным заваливанием высоких частот (> 5 кГц). Далее сигнал поступает на 8-разрядный АЦП, где с частотой 10 кГц происходит преобразование амплитуды в цифровой код. С выхода АЦП коды поступают в память FIFO емкостью 1Кбайт и, пройдя ее, через входной регистр блока КЛП1К поступают на магистраль КАМАК и далее в ЭВМ. Блок КЛП1К - это потенциальный входной/выходной регистр^{4/}.



Для вывода речи ЭВМ через КАМАК выдает байты на выходной регистр блока КЛП1К. Далее, пройдя память FIFO, они поступают на ЦАП, который преобразует коды в аналоговый сигнал с той же частотой 10 кГц. Далее, пройдя режекторный фильтр на 10 кГц, сигнал поступает на линейный вход усилителя звукового тракта телевизора.

Данную схему можно целиком реализовать на одной плате КАМАК. Систему речевого вывода целесообразно внедрить для облегчения работы операторов больших вычислительных комплексов, операторов просмотрово-измерительных систем, передачи срочных, аварийных сообщений. Для эффективного использования ресурсов такой системы ее можно сделать многоканальной, обслуживающей большое число потребителей.

Использование возможностей речевого ввода позволяет легко реализовать систему верификации говорящего, т.е. определения личности по специфическим особенностям речи. Возможности машины в этом случае существенно превосходят возможности человека. Это позволяет, например, заменить многочисленные пароли, служащие сейчас для проверки полномочий на использование вычислительных ресурсов.

В заключение авторы благодарят Н.Н.Говоруна за поддержку этой работы.

Литература

1. Пахомов В.Л. ОИЯИ, Р11-85-910, Дубна, 1985.
2. Керниган Б. и др. Язык программирования СИ. М., Финансы и статистика, 1985.
3. Кейтер Дж. Компьютеры - синтезаторы речи. М., Мир, 1985.
4. Смирнов В.И. ОИЯИ, Р13-86-256, Дубна, 1986.

Рукопись поступила в издательский отдел
3 декабря 1986 года.

Горшков В.А., Пахомов В.Л.
Речевой вывод для микроЭВМ

Р11-86-778

На микроЭВМ "Электроника-60" создана система речевого вывода для использования в САПР печатных плат с целью разгрузки зрительного канала восприятия информации. Аппаратура выполнена в стандарте КАМАК, работает с частотой выборки речевого сигнала 10 кГц, обеспечивает накопление данных при регенерации памяти ЭВМ. Программы написаны на языке СИ, реализуют методы непосредственного кодирования-восстановления, линейной дельта-модуляции, частотной модуляции. Система позволяет создать словарь на диске /тратя около 300 байт на 1 с речи/, который можно использовать на других ЭВМ, составляя нужные речевые сообщения.

Работа выполнена в Лаборатории вычислительной техники и автоматизации и Лаборатории ядерных реакций ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1986

Gorshkov V.A., Pakhomov V.L.
Speech Output System for Micro-Computer

P11-86-778

Speech output system has been created on "Elektronika"-60 micro-computer for the use in CAD/CAM of printed circuit boards in order to discharge spectators' channel of information perception. The equipment has been made in CAMAC standard, operates with 10 kHz speech output sampling rate, provides for the data acquisition at regeneration of the computer memory. The programs are written in "C" language, realize the following methods: of direct encoding - decoding, linear delta modulation, frequency modulation. The system permits to create vocabulary on disc (about 300 bite/s of speech) which could be implemented on other computers by compiling necessary speech communications.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, and Laboratory of Nuclear Reactions, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1986