ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1976 год Выпуск 67

УДК 621.391.19

О ВЫВОДЕ ИНФОРМАЦИИ В ВИДЕ СЛИТНОЙ РЕЧИ МЕТОДОМ КОМПИЛЯЦИИ ИЗ СЛОВ

В.А.Данилов, И.И.Мякишева

Одной из актуальных задач при создании систем автоматического управления является организация диалога с ЭВМ. Системы речевого общения с машиной могут быть полезны в различных областях современного производства. В частности, речевой ответ ЭВМ может использоваться весьма эффективно в оперативных АСУ и в ряде случаев иметь преимущество по сравнению с визуальной обратной связью.

Для вывода информации в форме устной речи в настоящее время известен ряд методов с применением синтеза речи [1].

В статье описывается один из способов дискретного синтеза слитной речи с ограниченным словарем, в основу которого положена компиляция из слов фраз. Были поставлены следующие задачи:

а) поиск способа формирования фраз из изолированно произнесенных слов;

б) поиск способа интонирования компилированных фраз;

в) оценка возможности компиляции словоформ из морфем с целью сокращения объема словаря, а следовательно, памяти ЭВМ.

Для решения этих задач был использован макет тракта полосного вокодера, сочлененного с ЭЦВМ. Анализатор и синтезатор тракта имеют 12-полосную гребёнку фильтров в диапазоне частот 150 – 12000 гц, каналы выделения и управления частотой ОТ и канал управления тональными и шумовыми видами возбуждения. Рабочий словарь для синтеза был заранее записан на магнитную ленту с голоса одного диктора и введен в ЭВМ через анализатор спектра в виде 14-ти параметров, проквантованных с частотой 50 гц.

Введенные в ЭВМ параметры слов по специальной программе переписывались в долговременную память ЭВМ. Одновременно был предусмотрен вывод полученных данных на печатающее устройство с последующим редактированием (выбрасыванием длительных пауз между словами и акустических помех), определением начала и конца каждого слова с нумерацией проб и зоны его расположения в долговременной памяти ЭВМ. По имеющейся программе вывода информация может непосредственно, без каких-либо изменений поступать с выхода ЭВМ на вход синтезатора речи, либо подвергаться следующим преобразованиям, позволяющим получить на выходе слитную речь (фразы):

а) вставке между словами пауз заданной длительности;

б) выбрасыванию заданного числа проб, добавлению проб или замене значений параметров в пробах другой информацией, введенной с перфокарт;

в) линейной интерполяции значений граничных проб слов, оставшихся после преобразований по п."б" на заданном временном интервале;

г) замене значений частоты ОТ в любой части слова и заданию мелодии речи методом линейной интерполяции полученных значений.

В качестве исходного речевого материала (рабочего словаря) было использовано 150 слов, произнесенных одним диктором изолированно с повествовательной и вопросительной интонациями. У ряда слов были записаны отдельные словоформы (косвенные падежи у существительных и прилагательных).

Для опытов по компиляции из записанных слов были составлены следующие фразы:

1. Вы меня слышите?
2. Вы слышите меня?
3. Слушаю Вас.
4. Нажмите вторую кнопку справа!
5. Я вас понял.
6. Что делать дальше?
7. Поднять груз над палубой!
8. Повторите, пожалуйста, что я сказал.

Для контроля в ЭВМ были введены и сами фразы, произнесенные тем же диктором.

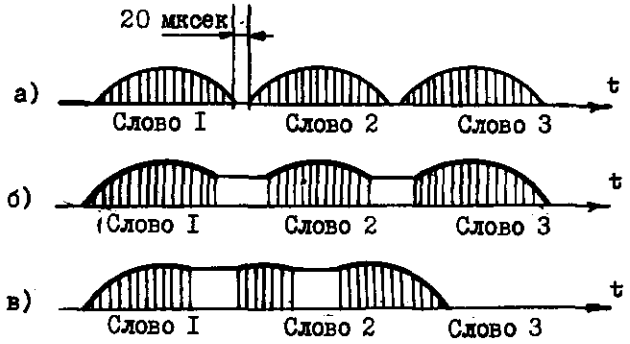

Рис.I. Компиляция фраз из слов: а) без интерполяции переходов; б) с выбрасыванием граничных проб и интерполяцией переходов; в) со сближением слов во времени.

В качестве первого опыта по стыковке было опробовано последовательное воспроизведение слов с паузой между ними, равной 20 мсек, без каких-либо преобразований параметров слов (рис.I,а). Собственно, только этот способ и можно назвать компиляцией в чистом виде, в других способах, описанных ниже, будут применены определенные правила для синтеза речи на уровне слов.

Для определения влияния длительности паузы на восприятие предложений был поставлен предварительный опыт по стыковке слов с паузами разной длины: 20, 40, 60 мсек. Оказалось, что даже с минимальной паузой речь воспринимается как диктовка отдельных слов, что и следовало ожидать, так как слушатель ориентируется на характерное оформление начала – конца слов. Такая речь вполне разборчива и может быть применена в случаях вывода информации в виде многословных команд.

Во втором опыте слова, сочленяемые по фразе, подвергались определенным преобразованиям. Для создания плавного перехода между ними у последнего звука предыдущего слова и первого звука последующего слова выделялись переходные участки, соседние с паузой. Затем была осуществлена операция выбрасывания спектральной информации из проб, входящих в переход к паузе у последнего звука и в переход от паузы у первого звука каждого слова вплоть до квазистационарного участка. После этого слова сближались по оси времени на n проб и между крайними пробами производилась линейная интерполяция значений параметров (рис.I,б,в). При формировании переходов учитывались место и способ образования начальных – конечных звуков в слове, для чего все звуки разбивались на группы. В зависимости от этих групп выбрасывание производилось на участках от 3 до I8 проб (60–360 мсек) с последующим сближением до 50% длительности переходных участков для нединамичных звуков и до 30% для динамичных (взрывных, дрожащих). Можно предположить, что будет достаточно 5–8 групп для получения плавных переходов, близких к естественным.

В работе по определению правил сочленения слов использовался опыт синтеза по правилам на уровне фонем [2], где решалась аналогичная задача формирования переходных процессов между звуками речи с учетом коартикуляции.

Прослушивание фраз, компилированных описанным выше способом, показало, что достигается заметная слитность по сравнению с результатами первого опыта. Бригада аудиторов в составе I0 человек оценила при 3-балльной системе второй режим на 0,6 балла выше, чем первый. В первых двух опытах мелодическая кривая слов оставалась неизменной, благодаря чему на стыках слов при компиляции фраз происходили интонационные разрывы, что не могло не сказаться на слитности и естественности речи. Поэтому в третьем опыте была предпринята попытка искусственного интонирования синтезированных вторым способом фраз с имитацией трех основных типов интонации: повествования, вопроса, побуждения. Сначала была поставлена задача повышения слитности и естественности за счет линейной интерполяции параметра основного тона.

Во фразе, компилированной из слов, задавались точки отсчета естественной частоты ОТ, расположенные в начале и в середине первого слова, в середине последующих слов, в середине и конце последнего слова. Между этими значениями производилась интерполяция, и полученной аппроксимированной кривой мелодии управлялся генератор импульсов ОТ синтезатора (рис.2,а). Было отмечено, что таким образом удается повысить естественность звучания фраз.

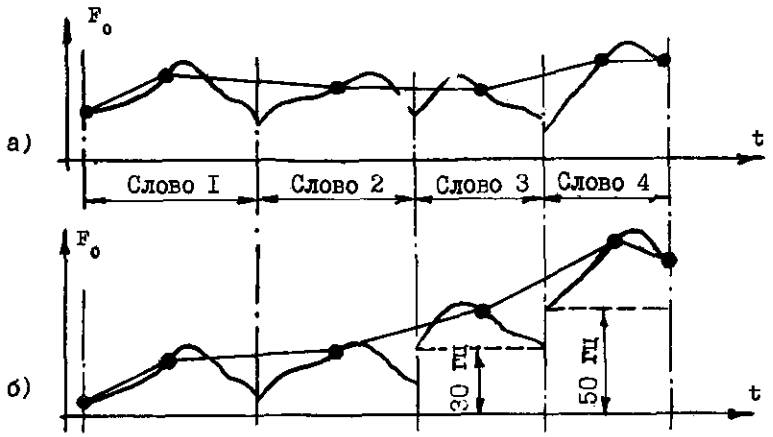

Рис.2. Способы интонирования фразы при компиляции: а) интерполяция значений ОТ по заданным точкам отсчёта; б) интерполяция транспонированных значений ОТ.

Далее с применением этого способа пытались имитировать виды интонации, перечисленные выше. Для этого были использованы слова, заранее произнесенные с требуемой интонацией вопроса и повествования.При составлении фраз из таких слов выяснилось, что тип интонации зависит в основном от интонации последнего слова: воспринимают как вопросительные фразы, в которых последнее слово было произнесено как вопрос.

В следующем опыте интонации побуждения и вопроса имитировались при помощи искусственного подъема частоты ОТ на определенных участках компилированной фразы. Оказалось,что достаточно поднять частоту ОТ в середине предпоследнего слова на 30 гц, а в середине последнего – на 50 гц, как начинает восприниматься вопросительная интонация,хотя слова были произнесены диктором в повествовательной форме (рис.2,б).

Для оценки результатов интонирования этим способом фразы в случайном порядке были предъявлены 8 аудиторам.Оказалось,что вопрос и повествование воспринимаются достаточно уверенно: повествование в I00% случаев, вопрос в 70% (в I5% – повествование и в I5% – побуждение). Значительно хуже опознается побуждение – в 3I% случаев (а в 69% – повествование).

Задание кривой мелодии речи методом линейного интерполирования весьма перспективно, так как помимо простоты реализации позволит в дальнейшем полностью формализовать процесс интонирования в системах синтеза речи. Для оценки возможности сокращения объема словаря, хранящегося в памяти ЭВМ, был поставлен небольшой опыт по образованию словоформ в процессе синтеза. В этом случае в рабочий словарь включаются только исходные формы слов, а словоформы синтезируются методом компиляции.

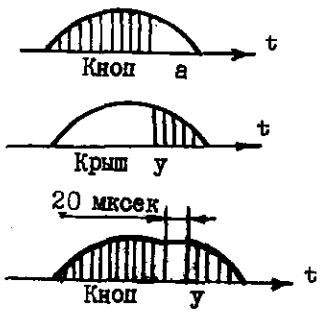

Рис. 3

Принцип образования словоформ иллюстрируется рис.3. В исследуемой фразе "Нажмите вторую кнопку справа" словоформы "вторую" и "кнопку" были скомпилированы из слов "вторая", "кнопка" путем удаления спектральной информации частей слов, относящихся к флексии, и заменой информации о соответствующих морфемах -у, -ую, взятой из других слов.Прослушивание синтезированной таким образом фразы обнаружило ее слитное и вполне естественное звучание. Кроме этой фразы, было образовано и просинтезировано таким способом еще несколько слов:

дом/а = дом + (кнопк)а
сказа/ть = сказа(л) + (нажа)ть
метр/ов = метр + (метр)ов
слуш/ать = слуш(аю) + (ех)ать
на/втор/ом = на(д палубой) + втор(ая) + (д)ом

Полученные слова были предъявлены бригаде аудиторов вместе со словами, просинтезированными непосредственно с голоса диктора (без компиляции). Слова предъявлялись в случайном порядке, и аудиторам предлагалось оценить их с точки зрения естественности. Оказалось, что в пользу слов, компилированных на морфемном уровне, было подано 47% всех голосов. Можно считать, что аудиторы практически не различали естественных и компилированных слов.

Проведенный небольшой эксперимент позволяет надеяться,что разработка этого метода перспективна для использования его в системах, где требуется сокращение памяти машины, правда, за

счет определенного усложнения алгоритма синтеза слитной речи методом компиляции.

Все результаты, полученные в описанных опытах, носят предварительный характер и требуют проверки на большом расчетном материале.

Литература

1. ГОЛУНОВ С.В. Методы распознавания и синтеза речи с ограниченным словарем. -Настоящий сборник, с.106-134.

2. ГОЛУНОВ С.В. Синтез речи. В кн.: Труды Всесоюз.семинара АРСО-IУ, Киев-Канев, 1968.

Поступила в ред.-изд.отд.
7 мая 1975 года