Министерство Образования и Науки Украины

**Курсовой проект**

***На тему***: «Разработка программного канального вокодера»

**по курсу** «Цифровая обработка сигналов»

2006

АННОТАЦИЯ

В данном проекте описан процесс создания канального вокодера. Описана программа на языке Matlab, модель, созданная с помощью системы Matlab Simulink, а так же программа на DSK TMS320C6711(5402), использующая возможности CODE COMPOSER STUDIO v.2, осуществляющая сжатие и восстановление речевого сигнала в реальном времени до уровня не более 4800 бит/с с удовлетворительным качеством восстановленного речевого сигнала (требуется обеспечить разборчивость речи, при этом узнаваемость диктора не обязательна).

На защиту студент представляет:

- пояснительную записку;

- электронную презентацию проекта в виде файла формата .ppt, .pdf или .ps;

- полностью отлаженное программное обеспечение.

Пояснительная записка должна выполняться в соответствии с требованиями действующих стандартов и содержать следующие разделы:

- введение;

- обзор существующих систем и методов решения задачи в соответствии с темой курсового проекта;

- обоснование метода решения задачи;

- разработка структурной схемы системы и расчет необходимых системных параметров;

- разработка программной модели системы на языке МАТЛАБ;

- результаты тестирования модели системы;

- разработка программного обеспечения системы на языке Си для реализации ее на цифровом сигнальном процессоре;

- анализ вычислительной сложности;

- анализ результатов реализации системы в реальном времени на цифровом сигнальном процессоре;

- краткое руководство пользователя;

- заключение;

- библиографический список;

- приложения:

- тексты программ на языке МАТЛАБ;

- тексты программ на языке Си.

ВВЕДЕНИЕ

В настоящее время, с развитием вычислительной техники, стоимость сигнальных процессоров довольно не велика и постоянно уменьшается, при этом увеличивается производительность вновь выпускаемых моделей процессоров. При этом стоимость каналов связи остается на достаточно высоком уровне, так как создание современных цифровых каналов связи требует значительных капиталовложений, а существующие аналоговые линии связи не выдерживают возрастающей нагрузки. Поэтому разработка систем сжатия речевых сигналов, с целью передачи их по каналам связи, является одной из актуальных задач современности.

Вокодер (от англ. voice — голос, coder — кодировщик) представляет со­бой электронное устройство, предназ­наченное для анализа и синтеза зву­ков человеческого голоса.

Впервые вокодером был назван изо­бретенный в 1936 году американским инженером Гомером Дадлеем аппарат, сужающий полосу частот, требуемую для передачи речевого сигнала по ка­налам связи. В последующие десяти­летия появилось множество разновид­ностей вокодера, применяемых в систе­мах связи. В них передается не са­ма речь, а определенные параметры речевого сигнала, по которым его за­тем можно восстановить в месте при­ема. Широко применяют вокодер в аку­стических исследованиях, при обуче­нии иностранным языкам, в речевой терапии.

Первый раздел пояснительной записки посвящен обзору существующих систем и методов построения сжатие речи с помощью канальных вокодеров обоснование метода решения данной задачи.

Во втором разделе КП разработка структурной схемы системы и расчет необходимых системных параметров для построения канального вокодера;

Разработка программной модели системы на языке Matlab и результаты тестирования модели системы приведены в третьем разделе.

В четвертом разделе описан процесс разработки программного обеспечения проектируемой системы канального вокодера на языке Си для реализации ее на цифровом сигнальном процессоре.

В пятом разделе произведен анализ вычислительной сложности, разработанной системы сжатия речи, а так же анализ результатов реализации данной системы в реальном времени на цифровом сигнальном процессоре, и краткое руководство пользователя системы сжатия.

В приложениях приводятся тексты разработанных программ на языке МАТЛАБ и Си.

ПОСТАНОВКА ЗАДАЧИ НА ПРОЕКТИРОВАНИЕ

Разработать и реализовать на ЦПОС TMS320C6711(5402) систему сжатия речи (рекомендуется реализовать канальный вокодер), осуществляющую сжатие и восстановление речевого сигнала в реальном времени до уровня не более 4800 бит/с. Считать, что исходный речевой сигнал представлен в виде последовательности 16-разрядных отсчетов c частотой дискретизации 8КГц. Необходимо обеспечить удовлетворительное качество восстановленного речевого сигнала (требуется обеспечить разборчивость речи, при этом узнаваемость диктора не обязательна).

**1.** **СИСТЕМЫ СЖАТИЯ РЕЧИ**

Голосовой тракт человека представляет собой акустическую трубу, ко­торая с одной стороны оканчивается голосовыми связками, а с другой губами. Форма голосового тракта определяется положением губ, челюстей языка и мяг­кого неба.

Звуки в этой системе образуются тремя способами. Вокализованные (звонкие) звуки - путем возбуждения голосового тракта квазипериодическими импульсами воздушного давления, создаваемыми вибрациями голосовых свя­зок. Фрикативные звуки образуются проталкиванием воздуха через сужения в определенных областях голосового тракта, в результате чего возникает турбуленция, которая является источником шума, возбуждающего голосовой тракт. Взрывные звуки образуются путем создания избыточного давления в области полного смыкания голосового тракта с последующим его быстрым размыка­нием. Все эти источники создают широкополосное возбуждение голосового тракта, который в свою очередь действует как линейный фильтр с изменяющи­мися во времени параметрами.

На рис. 14 приведена модель источника речи на основе цифрового пред­ставления речевых сигналов. Предполагается, что в этой модели дискретные отсчеты речевого сигнала формируются на выходе ЦФ с переменными параметрами, который аппроксимирует передаточные свойства голосового тракта, обусловленные формой импульсов возбуждения.

U[n]

Генератор

импульсной последовательности

Генератор

случайных

чисел

Цифровой фильтр с переменными пара-метрами

Период основного тона

Коэффициенты ЦФ

(параметры голосового

тракта)

G

Aмплитуда

X[n]

Отсчеты

речи

Рисунок 14 - Модель источника речи

На временном интервале порядка 10ms характеристики ЦФ можно считать неизменными. На каждом таком интервале ЦФ может быть охарактеризован совокупностью своих коэффициентов. В случае вокализованной речи ЦФ возбуждается генератором квазиканонической импульсной последовательности, расстояние между соседними импульсами которого соответствует периоду основного тона. На интервалах невокализованной речи ЦФ возбуждается генератором случайных чисел, который вырабатывает шумовой сигнал с равномерной спектральной плотностью. В обоих случаях сигнал, поступивший на ЦФ, управляется по амплитуде.

На рассмотренной модели базируются многочисленные способы представления речевых сигналов. По сложности реализации эти способы кодирования речи занимают широкий диапазон от простейшей периодической дискретизации до оценок параметров модели изображенной на рис.14.

Существует несколько подходов к сжатию речевых сигналов:

- кодирование формы волны речевого сигнала;

- кодирование параметров речевого тракта человека и источника возбуждения;

- кодирование символьной информации (фонем);

- кодирование лингвистической информации (слов, фраз и т.п.).

## 1.1 Непосредственное кодирование формы речевого сигнала

Исходный речевой сигнал представляет собой акустическую волну (волна давления в воздухе), которую можно преобразовать в электрический сигнал с помощью микрофона. Будем считать, что спектр речевого сигнала лежит в диапазоне от 100 до 4000 гц. Динамический диапазон изменения амплитуды, достаточный для описания речевых сигналов, составляет 12 двоичных разрядов.

Первым шагом, обеспечивающим сжатие речевого сигнала, является попытка обеспечения равномерной относительной точности измерения значения амплитуды сигнала. Для этого 14-12-ти разрядный динамический диапазон амплитуды разбивают на 8 логарифмических поддиапазонов, в каждом из которых значение амплитуды кодируют 5 разрядами и, таким образом, достигают сокращения информации до 64000 бит/с (кодирование по μ- и A- законам в соответствии со стандартом ITU -G.711). Следующим шагом является адаптивная дифференциальная импульсно-кодовая модуляция (АДИКМ), (например, в соответствии со стандартами G.721 или G.726 8-40000 бит/с), с помощью которой осуществляют кодирование (аппроксимацию) степени приращения амплитуды сигнала во времени. Таким путем удается достичь степени сжатия речевого сигнала порядка 32000-16000 бит/сек., причем приемлемое (коммерческое) качество речи (по критерию отношения: полезный\_сигнал/шум) обеспечивается до 24000 бит/сек. При более низких скоростях кодирования сохраняется разборчивость речи, но характерны сильные нелинейные и частотные искажения сигнала и ухудшение отношения сигнал/шум. Дальнейшее уменьшение информационной емкости сигнала с помощью данного подхода считается неэффективным.

***1.1.2. Параметрическое кодирование***

Низкоскоростное кодирование складывается из двух основных процессов:

- параметрическое представление речевого сигнала минимальным набором параметров, характеризующих источник возбуждения и акустический фильтр, определяющий передаточную функцию голосового тракта;

- дискретизация речевых параметров для их передачи по каналу связи при использовании минимальной емкости канала.

Для параметрического описания речи обычно используется подход, основанный на вычислении параметров, описывающих передаточную функцию речевого тракта человека и функцию возбуждения. Такими параметрами могут являться: осредненные значения энергии речевого сигнала, разбитого на ряд частотных полос, или коэффициенты линейного предсказания (или, связанные с ними, коэффициенты отражения). Обычно для кодирования речи используются 8-10 параметров (один из вышеперечисленных наборов), вычисляемых на интервалах порядка 5-30 мс (так как на таком интервале речь может считаться стационарным процессом), кроме того, вычисляется параметр, характеризующий изменение амплитуды либо мощности сигнала, период основного тона речи, а также признак типа тон/шум/пауза, характеризующий способ возбуждения речевого сигнала.

Полученный набор параметров, оптимизированный по критерию точности и минимальной разрядности представления, передается в цифровом виде по каналу связи в реальном времени, а на приемном конце осуществляется синтез речевого сигнала по перечисленным параметрам. Таким путем удается снизить информационную емкость речевого сигнала до уровня 16000 - 1200 бит/сек, причем с сохранением разборчивости и индивидуальных особенностей речи говорящего.

***1.1.3. Другие способы кодирования***

Следующим шагом в направлении дальнейшего увеличения компрессии является создание фонемного вокодера. Как известно, минимальной слогоразличительной (и словоразличительной) единицей речи является фонема. Поэтому создание устойчивого метода распознавания фонем позволит снизить скорость кодирования речевой информации до 100 бит/сек, что соответствует информационной скорости текста. Следует отметить, что на приемной стороне речь будет восстановлена синтезатором речи по фонемному тексту, при этом информация об индивидуальности диктора будет утрачена.

## 1.2. Канальные вокодеры

Канальный вокодер представляет собой совокуп­ность двух основных частей - ана­лизирующей (передающая сторона) и синтезирующей(приемная), кото­рые содержат идентичные наборы(гребенки) по­лосовых фильтров, перекрывающих оп­ределенный частотный интервал. Структура канального вокодера представлена на рисунке 9.1:

Рисунок 9.1 - Структурная схема канального вокодера

1

2

N

ВОТ

Т - Ш

КС

MUX

DEMUX

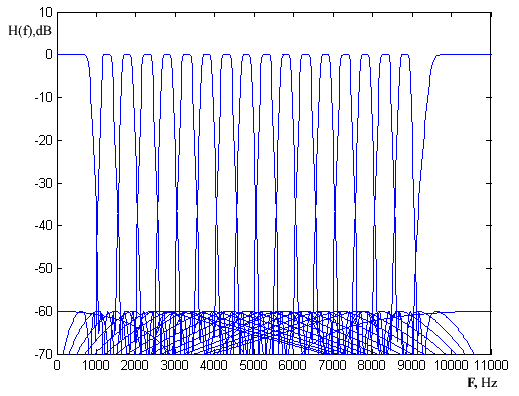
ГТ

Г Ш

Σ

Выход

Речь

Рисунок 9.2 - АЧХ гребенки фильтров

Фильтры блока анализа обеспечивают тональное раз­деление спектра сигнала. Для перекрытия всей полосы звуковых частот, наряду с полосовыми, в гребенке используют фильтры НЧ и ВЧ (в самых низкочастотном и высоко­частотном каналах). Типовая амплитудно-частотная характеристика гребенки фильтров, в случае равномерного разделения каналов может иметь следующий вид:

Детектор и фильтр НЧ в каждом канале выде­ляют огибающую сигнала данного канала, и каждая из них характеризует энергию речевого спект­ра в соответствующей полосе частот для фрагмента речи (длина фрагмента 5-30 мс). Для более компактной передачи выходы каждого из каналов могут логарифмироваться и кодироваться с помощью дельта импульсной кодовой модуляции(ДИКМ)[1].

От числа частотных полос зависит разборчивость синтезированной речи. Хорошие результаты получаются при числе каналов 15-16 (полоса 100 Гц... 4 кГц). Для улучшения качества восстановленной речи при выборе центральных частот и ширины полос пропускания канальных фильтров целесообразно использовать критические полосы слуха[1].

При выборе типа фильтров необ­ходимо иметь в виду, что спект­ральные составляющие сигнала вбли­зи центральной частоты резонансных фильтров подвергаются существенным фазовым сдвигам, а это приводит к изменению тембра, даже если ампли­тудные соотношения сохранены. Причем при увеличении порядка фильт­ров фазовый сдвиг увеличивается, явля­ясь причиной характерной для вокодера неестественности речи. С другой сторо­ны, при недостаточной крутизне спа­да АЧХ фильтров появляется “смазанность” синтезированной речи. Практика показыва­ет, что оптимальный результат соот­ветствует АЧХ фильтров при крутизне их спада при­мерно 36 дБ на октаву.

Частоту среза канального фильтра НЧ выби­рают в десять раз меньшей цент­ральной частоты полосового фильтра канала, однако возможно использование одинаковых фильтров низких частот с частотой среза около 25 Гц, что несколько снижает качество восстановленного речевого сигнала.

Кроме оценки энергии сигнала в частотных полосах, в блоке анализа также производится оценка характера речевого фрагмента. В детекторе тон/шум(Т-Ш) оценивается: вокализованным или фрикативным является данный фрагмент речевого сигнала. Такой детектор может быть реализован через определение числа переходов через ноль речевого сигнала (для вокализованных фрагментов в отличие от фрикативных число переходов через ноль невелико) или оценкой энергии сигнала в полосах до 800 Гц и более 2КГц (структурная схема такого детектора представлена на рисунке 9.3).

Признак тон/шум

Речь

Рисунок 9.3 – Структура детектора тон-шум

Также в блоке анализа происходит определение частоты основного тона для вокализованных фрагментов речи. Для этого в структуру блока анализа вводится выделитель основного тона (ВОТ). Известно, что значение ОТ для разных голосов может изменяться почти в 10 раз - от 2 до 18 мс. Это обстоятельство создает немало трудностей при оценке ОТ, так как слух очень чувствителен к его искажениям. На сегодняшний день известно большое количество алгоритмов оценки ОТ, оперирующих как непосредственно с временным представлением речевого сигнала, так и со спектром речи. Например, метод Голда-Рабинера[2,5], автокореляционный метод[1,2], SIFT(simplified inverse filter tracking) алгоритм [2].

Таким образом, для каждого фрагмента речевого сигнала на приемную сторону передается признак тон-шум(1 бит), частота основного тона, в случае вокализованного фрагмента(5-6 бит), значения энергии сигнала по каналам (при использовании ДИКМ менее 5 бит на канал).

На приемной стороне, в случае если принят вокализованный фрагмент, в качестве источника возбуждения гребенки фильтров выступает генератор периодических сигналов(ГТ). В качестве сигнала возбуждения может быть выбрана последовательность специального вида импульсов[1] с периодом равным частоте основного тона данного фрагмента (также принимаемой из канала связи). В случае невокализованого фрагмента, в качестве сигнала возбуждения используется белый шум(ГШ), который может быть получен с помощью генератора случайный чисел.

Сигнал возбуждения умножается на значения энергии для каждого из каналов, после чего поступает на фильтр соответствующего канала (гребенка фильтров блока синтеза полностью аналогична гребенке блока анализа).

Для получения выходного синтезированного речевого сигнала выходы всех каналов блока синтеза суммируются.

## 2. Синтез цифровых рекурсивных фильтров с использованием

## пакета MATLAB

Решение задачи синтеза цифровых рекурсивных фильтров(РФ) сводится к нахождению коэффициентов **ak** и **bk** разностного уравнения:

|  |  |
| --- | --- |
|  | (9.1) |

Известны прямые и косвенные методы синтеза рекурсивных цифровых фильтров[4,1]. Прямые методы основаны на непосредственном определении параметров цифровых РФ по заданным временным или частотным характеристикам. Косвенные методы синтеза цифровых РФ основаны на дискретизации аналогового фильтра-прототипа, удовлетворяющего заданным требованиям. Существуют различные способы аппроксимации аналоговых фильтров[6,1]. На практике широко применяется аппроксимация АЧХ аналоговых фильтров с помощью полиномов Баттерворта, Чебышева, Бесселя, Кауэра.

При проектировании фильтров обычно задают следующие требования к их амплитудно-частотным характеристикам(АЧХ):

- границы полос пропускания и задерживания (соответственно ω1 и ω2);

- затухание в полосе задерживания (Н2);

- коэффициент передачи в полосе пропускания(Н1- обычно он равен 1);

- допуск на отклонение реальной АЧХ от желаемой в полосе пропускания(Δ1).

В пакете МATLAB для синтеза РФ существует ряд функций:

**- butter** – аппроксимация фильтра Баттерворта;

**- cheby1** – аппроксимация фильтра Чебышева первого типа;

**- cheby2** – аппроксимация фильтра Чебышева второго типа(инверсный фильтр Чебышева);

**- ellip** – аппроксимация эллиптического фильтра(фильтр Кауэра).

Например, вызов функции **butter** в пакете MATLAB может иметь следующий синтаксис:

*[b,a] = butter(n,Wn);*

*[b,a] = butter(n,Wn,'ftype').*

Для расчета порядка фильтра Баттерворта, обеспечивающего требуемые характеристики(Н1, Н2, Δ1) в пакете MATLAB используется функция **buttord**, имеющая следующий синтаксис:

*[n,Wn] = buttord(Wp,Ws,Rp,Rs).*

Пример. С помощью пакета MATLAB найти коэффициенты и построить характеристики полосового фильтра Баттерворта с полосой пропускания от 60 до 200Гц(частота дискретизации 1кГц), с пульсациями в полосе пропускания не более 3 дБ и затуханием в полосе задерживания не менее 40 дБ.

Ниже представлен текст программы, а на рисунке 9.4 показаны характеристики полученного фильтра.

*Wp = [60 200]/500; %расчет нормированных граничных частот полосы пропускания*

*Ws = [50 250]/500; %расчет нормированных граничных частот полосы задерживания*

*Rp = 3; Rs = 40;% задаем величину пульсаций и затухание*

*[n,Wn] = buttord(Wp,Ws,Rp,Rs)% определяем порядок фильтра и частоты среза*

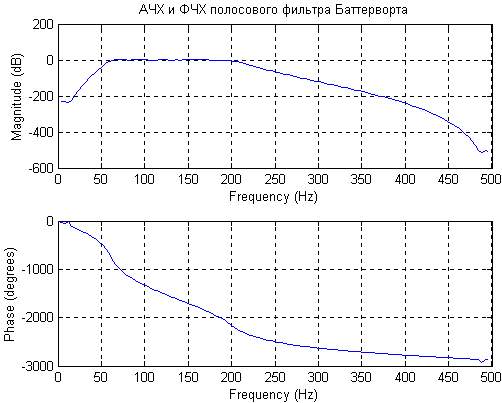
*>>n =16*

*>>Wn =[0.1198 0.4005]*

*[b,a] = butter(n,Wn);%находим коэффициенты фильтра*

*freqz(b,a,128,1000)% строим характеристики фильтра*

*title('АЧХ и ФЧХ полосового фильтра Баттерворта')*

Рисунок 9.4 - АЧХ и ФЧХ полосового фильтра Баттерворта

Синтез гребенки фильтров

От числа частотных полос зависит разборчивость синтезированной речи. Хорошие результаты получаются при числе каналов 15-16 (полоса 100 Гц... 4 кГц). По условию поставленной задачи необходимо разработать вокодер для сжатия и передачи по каналу связи речи без идентификации диктора, то для уменьшения количества передаваемой информации можно уменьшить количество каналов до 12-14. На рисунке 3.1 представлена реализация используемой при кодировании гребенки фильтров, состоящей из 12 полосовых фильтров, ФНЧ и ФВЧ.



Рисунок 3.1 – гребанка фильтров.

Нижеследующий текст программы на матлабе позволяет рассчитать необходимые характеристики фильтров и построить АЧХ гребенки фильтров.

n\_bands = 15;

fd1 = 8000

[n,Wn] = buttord(100/fd1, 300/fd1, 2, 50);

[b{1},a{1}] = butter(n, Wn);

[n,Wn] = buttord([125/fd1 300/fd1], [20/fd1 500/fd1], 4, 40);

[b{2},a{2}] = butter(n, Wn);

[n,Wn] = buttord([305/fd1 485/fd1], [105/fd1 685/fd1], 3, 40);

[b{3},a{3}] = butter(n, Wn);

[n,Wn] = buttord([500/fd1 690/fd1], [300/fd1 890/fd1], 3, 40);

[b{4},a{4}] = butter(n, Wn);

[n,Wn] = buttord([700/fd1 880/fd1], [500/fd1 1080/fd1], 2, 40);

[b{5},a{5}] = butter(n, Wn);

[n,Wn] = buttord([900/fd1 1080/fd1], [700/fd1 1280/fd1], 2, 40);

[b{6},a{6}] = butter(n, Wn);

[n,Wn] = buttord([1100/fd1 1380/fd1], [900/fd1 1580/fd1], 2, 45);

[b{7},a{7}] = butter(n, Wn);

[n,Wn] = buttord([1400/fd1 1680/fd1], [1200/fd1 1880/fd1], 2, 47);

[b{8},a{8}] = butter(n, Wn);

[n,Wn] = buttord([1700/fd1 1980/fd1], [1500/fd1 2180/fd1], 2, 50);

[b{9},a{9}] = butter(n, Wn);

[n,Wn] = buttord([2000/fd1 2330/fd1], [1800/fd1 2530/fd1], 2, 50);

[b{10},a{10}] = butter(n, Wn);

[n,Wn] = buttord([2350/fd1 2670/fd1], [2150/fd1 2870/fd1], 2, 50);

[b{11},a{11}] = butter(n, Wn);

[n,Wn] = buttord([2700/fd1 3070/fd1], [2500/fd1 3270/fd1], 2, 50);

[b{12},a{12}] = butter(n, Wn);

[n,Wn] = buttord([3100/fd1 3470/fd1], [2900/fd1 3670/fd1], 2, 50);

[b{13},a{13}] = butter(n, Wn);

[n,Wn] = buttord([3500/fd1 3880/fd1], [3300/fd1 4080/fd1], 2, 50);

[b{14},a{14}] = butter(n, Wn);

[n,Wn] = buttord(3900/fd1, 3700/fd1, 2, 50);

[b{n\_bands},a{n\_bands}] = butter(n, Wn, 'high');

hold on;

for i=1:n\_bands,

[H,W] = freqz(b{i}, a{i}, [0:0.01:pi]);

plot(W,20\*log10(abs(H)));

axis([0 pi -100 10]);

end;

hold off;

Гребенка состоит из цифровых фильтров Баттерворта, так как это наиболее распространенный БИХ-фильтров.