

УДК 004.934.8'1

ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ РЕЧЕВОГО СИГНАЛА В МОДЕЛИ ТЕКСТОНЕЗАВИСИМОЙ СИСТЕМЫ ИДЕНТИФИКАЦИИ ГОВОРЯЩЕГО

Минаева И.А., Белов Ю.С.

*Калужский филиал ФГБОУ ВО «Московский государственный технический университет имени
Н.Э. Баумана (национальный исследовательский университет)»*

В статье рассматриваются этапы извлечения признаков речевого сигнала при текстонезависимой идентификации говорящего на основе использования динамических мел-частотных кепстральных коэффициентов (DMFCC) в комбинации с мел-частотными кепстральными коэффициентами (MFCC) для надежной текстонезависимой идентификации говорящего. Рассматривается применение окна Хэмминга, особенности извлечения речи, быстрое преобразование Фурье и предварительная коррекция полученного звука.

Ключевые слова: идентификация по голосу, мел-частотные кепстральные коэффициенты, динамические мел-частотные кепстральные коэффициенты, окно Хэмминга, обработка речи, быстрое преобразование Фурье.

FEATURE EXTRACTION OF VOICE SIGNAL IN THE MODEL OF TEXT-DEPENDENT SYSTEM IDENTIFICATION OF THE SPEAKER

Minaeva I.A., Belov Yu.S.

This article discusses the steps of extracting features of a speech signal with text independent speaker identification based on the use of dynamic chalk-frequency cepstral coefficients (DMFCC) in combination with chalk-frequency cepstral coefficients (MFCC) for reliable speaker-independent identification. The application of the Hamming window, features of speech extraction, fast Fourier transform and preliminary correction of the received sound are considered.

Keywords: speaker identification, chalk-frequency cepstral coefficients, dynamic chalk-frequency cepstral coefficients, Hamming window, speech processing, fast Fourier transform.

Введение

Речевой сигнал содержит информацию о языке, а также о личности говорящего. Распознавание говорящего делится на две задачи: идентификация говорящего и верификация говорящего. Цель идентификации говорящего - определить человека по его голосу. Существует два типа систем идентификации говорящих: текстозависимые, текстонезависимые [1]. Говорящий должен произносить одну и ту же фразу во время обучения и тестирования. При текстонезависимой идентификации фраза во время тестирования может отличаться от обучающей фразы.

Среди различных особенностей речи MFCC считаются важными параметрами для исследователей в области распознавания говорящего из-за предпочтительной симуляции воспринимающей способности слуховой системы человека. MFCC основаны на спектральной информации, полученной из сегмента речи с коротким временным окном. Они отличаются главным образом деталями представления спектра мощности.

Результаты данной статьи в дальнейшем будут использованы для тектонезависимой идентификации при помощи модели Гауссовых смесей.

Этапы извлечения мел-частотных кепстральных коэффициентов. На рисунке ниже (рис.1) приведены все этапы извлечения кепстральных коэффициентов. Предполагается, что речевой сигнал уже прошел предварительную обработку, к звуку был применен фрейминг, а также была произведена очистка от шума.

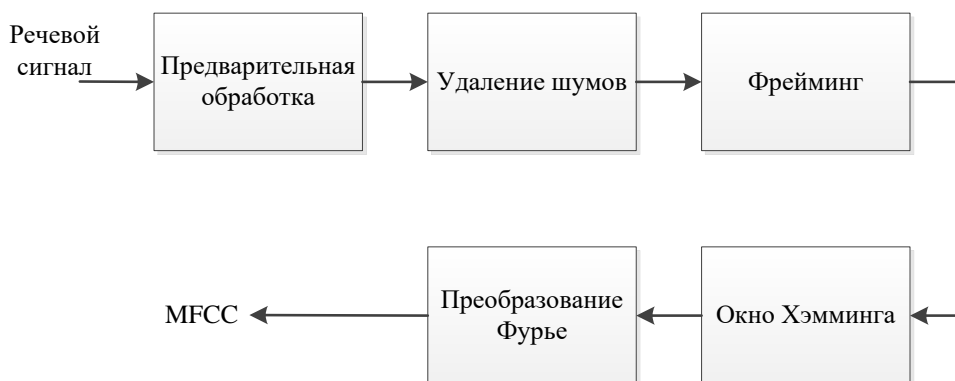


Рисунок 1 - Этапы извлечения мел-частотных кепстральных коэффициентов

Окно Хэмминга. Окно применяется для обеспечения спектрального сглаживания [4]. Это делается для каждого отдельного кадра, чтобы сузить сигнал до нуля в начале и в конце кадра. Оконное управление также важно для захвата динамических характеристик системы голосового тракта в механизме речевого производства. При использовании окна Хэмминга достигается подавление боковых лепестков при одновременном расширении главного лепестка, что делает его фильтром нижних частот с меньшими утечками [2]. Окно Хэмминга $w(n)$ имеет вид

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (1)$$

где N представляет ширину в выборках оконной функции с дискретным временем. Как правило, это целое число 2, такое как $2^{10} = 1024$. n является целым числом со значениями $0 \leq n \leq N-1$.

Особенности извлечения. Признаки извлекаются, затем создается банк фильтров с использованием гауссовых фильтров. При обработке речи MFCC является представлением кратковременного спектра мощности речевого звука, основанного на линейном косинусном преобразовании логарифмического спектра мощности по нелинейной шкале мел-частот, что получается путем дискретного косинусного преобразования в логарифмические энергии [5]. В итоге 20 коэффициентов MFCC получаются из каждого кадра.

Быстрое преобразование Фурье. Признаками являются репрезентативные векторы шаблонов для речевых сигналов. После оконного речевого сигнала быстрое преобразование Фурье (БПФ) применяется к каждому кадру, и вычисляется его квадратная величина. Для выборочных векторных данных анализ Фурье выполняется с использованием дискретного преобразования Фурье (ДПФ).

Предварительная коррекция. Высокочастотная составляющая речевого сигнала, как правило, слабая, поэтому высокочастотная энергия может отсутствовать для выделения признаков в верхнем частотном диапазоне. Во многих приложениях обработки речи необходимы высокочастотные компоненты [3]. Предварительная коррекция используется для повышения энергии высокочастотных сигналов. Таким образом, предварительное выделение помогает выровнять спектральный наклон в речи, и сигнал спектрально выравнивается. Выход предварительного выделения $\hat{s}(n)$ связан с входом $s(n)$ как

$$\hat{s}(n) = s(n) - \alpha s(n - 1) \quad (2)$$

где α - коэффициент вытеснения, значение которого варьируется от 0,9 до 1.

Практический пример. При предварительной обработке сигнала важно выделить вокализованную область от тишины. Пример сигнала представлен на рисунке 2, где явно наблюдаются области тишины, а также вокализованные области.

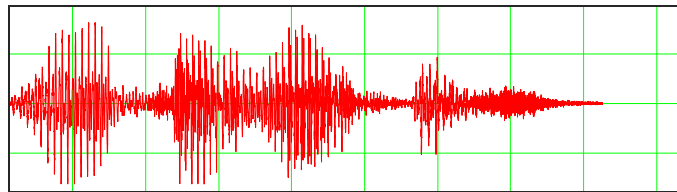


Рисунок 2 – Пример сигнала речи

После предварительной обработки, в сигнале необходимо убрать посторонние шумы. На рисунке 3 представлен пример сигнала речи с шумом льющейся воды (а), а также отфильтрованный сигнал методом Вейвлет-фильтрации (б).

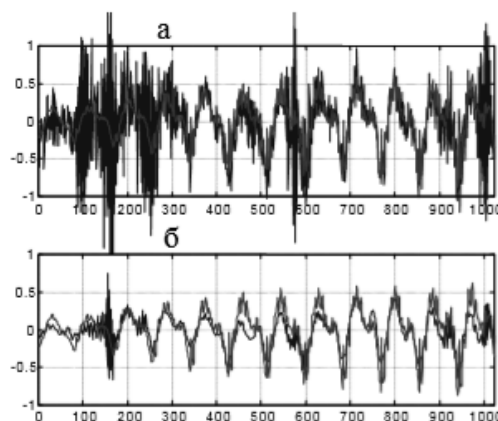


Рисунок 3 - Удаление окрашенного шума: а — зашумленный сигнал; б — отфильтрованный сигнал

После удаления шума, сигнал разбивается на фреймы (сегменты). Сигнал, разбитый на фреймы можно увидеть на рисунке 4.

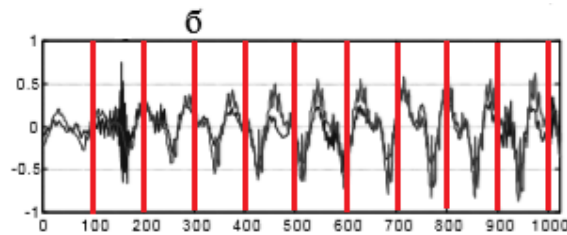


Рисунок 4 – Сигнал, разбитый на фреймы

После разбиения сигнала на сегменты, для каждого фрейма применяется окно Хэмминга, результатом его применения станет выделение центральной части кадра и плавное затухание амплитуд на его краях. Затем, применив преобразование Фурье, в результате чего осуществляется переход к мел-шкале, которая содержит мел-частотные кепстральные коэффициенты.

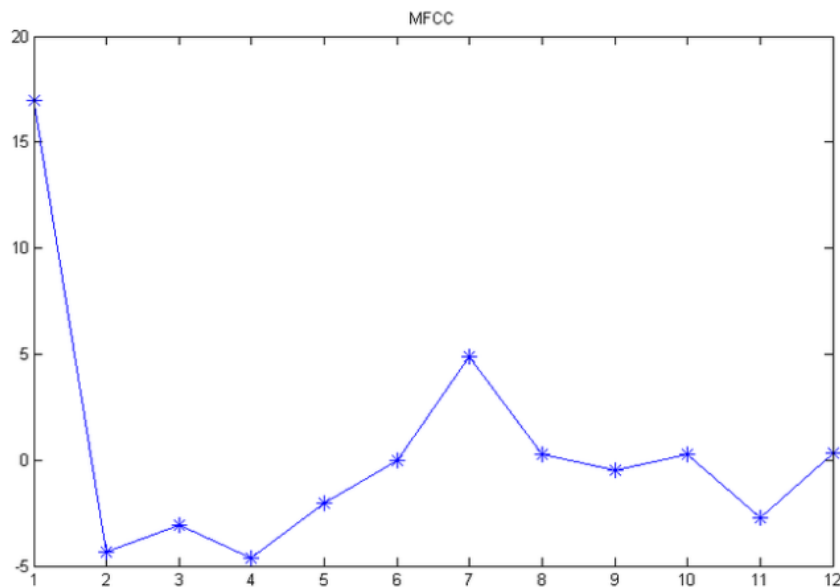


Рисунок 5 – Последовательность мел-частотных кепстральных коэффициентов

Заключение

В статье были рассмотрены этапы, необходимые для извлечения мел-частотных кепстральных коэффициентов. Этапы включают: применение окна Хэмминга, быстрое преобразование Фурье и предварительная коррекция полученного звука. В статье также был приведен практический пример извлечения признаков.

Список литературы

1. Гришунов С.С., Белов Ю.С. Основные математические методы выделения речевых особенностей в системах распознавания диктора // электронный журнал: наука, техника и образование. 2015. № 3 (3). с. 53-58.

2. Гришунов С.С., Бурмистров А.В., Молчанов А.Н. Математические методы классификации дикторов // вопросы радиоэлектроники. 2016. № 10. с. 13-17.
 3. Спажакин Ю.Г. современные средства проектирования систем голосовой биометрии // Методы и устройства передачи и обработки информации, 2016. С. 69-79
 4. Козырев М.О., Орлов М.Ю. Оконные функции и преобразование Фурье // Инновационные научные исследования: теория, методология, практика. Сборник статей IX Международной научно-практической конференции, 2017. С. 21-25
 5. Сорокин В.Н., Вьюгин В.В., Тананыкин А.А. Распознавание личности по голосу: аналитический обзор // Информационные процессы, 2012. С. 1-30
-