УДК 004.934

ИЗВЛЕЧЕНИЕ АКУСТИЧЕСКИХ ПРИЗНАКОВ ИЗ АУДИОСИГНАЛА В СИСТЕМАХ РАСПОЗНАВАНИЯ РЕЧИ

Крейс А.В., магистрант

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Боброва Н.Л. – канд. техн. наук

Аннотация. В настоящее время системы распознавания речи получили распространение в различных сферах жизни человека. Система распознавания речи – это система, позволяющая представить человеческую речь, содержащуюся в аудиосигнале, в форме, подлежащей интерпретации компьютером. В настоящее время существуют различные варианты реализации подобных систем, в которых можно выделить общие черты, например, необходимость выделения акустических признаков из исходного аудиосигнала, которые используются алгоритмом(-ами) для распознавания слов, произнесенных диктором. В данной работе выполняется обзор процесса получения данного набора признаков.

Ключевые слова. Акустический признак, система распознавания речи, медианный фильтр, фильтрация шумов, дискретное преобразование Фурье, окно Хэмминга, Мел-частотные кепстральные коэффициенты, Перцептуальное линейное предсказание, Кодирование с использованием линейного предсказания, фрейм, критические полосы слуха, кривые равной громкости, дискретное косинусное преобразование.

Прежде, чем приступить непосредственно к извлечению акустических признаков из аудиосигнала, его необходимо очистить от присутствующих в нем шумов. Существуют различные виды шумов, которые могут присутствовать в аудиосигнале. Они могут быть классифицированы по различным признакам:

– по виду взаимодействия с сигналом (аддитивный, мультипликативный);
– по наличию стационарности (стационарный, нестационарный);
– по постоянству (постоянный, непостоянный (в т. ч. импульсный));
– по частоте и т. д.

Для устранения шумов, присутствующих в сигнале, используются различные фильтры. Приведем некоторые примеры таких фильтров.

1. Медианный фильтр. Данный фильтр представляет собой пример фильтра с конечной импульсной характеристикой. Принцип работы данного фильтра заключается в применении скользящего окна из нескольких отсчетов входного сигнала, которые сортируются по значению с последующим получением отсчета из середины списка (или среднего значения двух отсчетов в середине списка в случае его четной длины). Полученное значение подается на выход фильтра. Медианный фильтр эффективен при устранении импульсного шума.

2. Низкочастотные / высокочастотные / полосно-пропускающие / полосно-заграждающие фильтры. Они применяются для отсеивания высоко-/низкочастотных составляющих сигнала / любых частотных составляющих вне указанного диапазона / определенного диапазона частот соответственно. Данные фильтры можно использовать для устранения аддитивного шума.

Для устранения искажений в спектре сигнала, вызванных применением преобразования Фурье к ограниченному участку сигнала, может быть использовано окно Хэмминга. Применение данного окна заключается в выполнении операции свертки входного сигнала со специальной оконной функцией. Помимо окна Хэмминга для этой цели применяются окна Ханна, Кайзера и др.

Для извлечения акустических признаков из аудиосигнала используются различные методы. Среди них можно выделить следующие: Мел-частотные кепстральные коэффициенты (Mel-frequency cepstral coefficients или MFCC), Перцептуальное линейное предсказание (Perceptial linear prediction или PLP), Дискретное вейвлет-преобразование (Discrete wavelet transform или DWT), Кодирование с использованием линейного предсказания (Linear predictive coding или LPC) и др. Некоторые из данных методов рассмотрены ниже.

Различные методы извлечения акустических признаков из исходного аудиосигнала подразумевают его дробление на небольшие отрезки. Данные отрезки называются фреймами (frames). В работе [1] проводились исследования влияния размера фрейма и количества кепстральных коэффициентов (о которых речь пойдет ниже) на производительность распознавания речи. Согласно результатам исследования размеры фрейма в диапазоне 16–32 мс не оказывают существенного влияния на показатель WER (Word Error Rate).

Популярным методом получения акустических признаков является MFCC. Человеческое ухо различает разные частотные составляющие сигнала неодинаково. Мел-представление позволяют учесть значимость определенных частот из спектра для человека, а также учесть тембр его голоса

[2]. В источнике [3] отмечается, что мел-частотный анализ представляет частоты речи с позиции психоакустического параметра слуха – высоты тона.

Мел представляет собой единицу высоты звука. Данная величина может быть получена из частоты с использованием следующей формулы:

$$m = 2595log_{10}\left(1 + \frac{f}{700}\right), \quad (1)$$

где *f* – частота звука (в Гц).

Для того, чтобы найти мел-частотные кепстральные коэффициенты, необходимо выполнить несколько шагов. Описание процесса получения данных коэффициентов можно найти в источниках [2, 4]. Первый шаг – применить дискретное преобразование Фурье к исходному звуковому сигналу в рамках фрейма. В источнике [5] указано, что размер таких фреймов обычно равен 10–40 мс, при этом кадры накладываются друг на друга. К полученному спектру применяется набор мел-фильтров. Далее определяется энергия для каждого фрейма. Последний этап – применение дискретного преобразования Фурье или дискретного косинусного преобразования.

Использование мел-частотных кепстральных коэффициентов позволяет сократить количество акустических признаков по сравнению с использованием отсчетов сигнала, или его спектра, или его периодограммы [2].

В настоящее время существуют различные модификации алгоритма MFCC, среди которых можно отметить линейно-частотные кепстральные коэффициенты (Linear-frequency cepstral coefficients или LFCC), метод кепстральных коэффициентов прямоугольного набора фильтров (Rectangular-frequency cepstral coefficients или RFCC), гамматон-частотные кепстральные коэффициенты (Gammaton-frequency cepstral coefficients или GFCC) [6].

Метод LPC подразумевает использование так называемого линейного предсказателя, который позволяет получать значение очередного отсчета сигнала *s(n)*, основываясь на предыдущих значениях отсчетов сигнала. Для этого используется следующая формула [3]:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k), \quad (2)$$

где a_k – *k*-ый коэффициент линейного предсказателя;

p – порядок линейного предсказания.

Для данного предсказателя необходимо найти коэффициенты a_k, для чего используются три базовых алгоритма [3]:

– ковариационный;

– автокорреляционный;

– лестничный.

Полученные коэффициенты используются в качестве акустических признаков. Метод LPCC (Linear prediction cepstral coding), который является вариацией алгоритма LPC, подразумевает также применение к полученным коэффициентам дискретного преобразования Фурье или дискретного косинусного преобразования, в результате чего получаются коэффициенты кепстра линейного предсказателя [7]. Одним из преимуществ LPCC является снижение влияния канала передачи на параметры речевого сигнала [7].

Метод PLP, разработанный Гинеком Германски (Hynek Hermansky), исключает несущественную информацию, содержащуюся в речи и тем самым улучшает качество ее распознавания [8]. Данный метод учитывает три психоакустических фактора: критические полосы слуха с маскированием, кривую равной громкости, степенное соотношение между громкостью и интенсивностью звука [3].

Критические полосы слуха являются частотными полосами, за пределами которых субъективные ощущения звука сильно изменяются [9]. В разных условиях при различных уровнях шума слышимость звука различается. Понижение уровня слышимости в условиях шума называется маскировкой звука.

Кривые равной громкости представляют собой графическое отображение нелинейности восприятия звука человеком [10]. Другое название – кривые Флетчера-Мэнсона [10]. Эти кривые показывают, какое звуковое давление необходимо создать для одинакового восприятия громкости различных частот [10].

Далее приведена последовательность шагов метода PLP [3]:

– применение к фрейму оконной функции и дискретного преобразования Фурье;

– определение спектра мощности для данного фрейма;

– перевод полученного спектра мощности в барк-шкалу;

– перемножение спектра со спектром мощности кривой маскирования критической полосы;
– сглаживание полученного спектра функцией кривой равной громкости;
– извлечение кубического корня из амплитуды полученного спектра.

После выполнения перечисленных выше шагов осуществляется расчет коэффициентов предсказания (см. метод LPC, описанный выше), на основе которых рассчитываются кепстральные коэффициенты [3].

Источник [8] отдает предпочтение использованию методов MFCC и PLP, а не LPC, так как первые два получены из концепции набора фильтров, находящихся в логарифмическом пространстве, и концепции человеческой слуховой сенсорной системы и, следовательно, демонстрируют более хороший результат [8]. Согласно исследованиям, описанным в источнике [11] метод LPC демонстрирует более хорошие результаты, чем MFCC, в условиях невысокого уровня шума в аудиосигнале.

Таким образом, в данной работе был выполнен сбор сведений, касающихся извлечения акустических признаков из аудиосигнала в системах распознавания речи. Была рассмотрена классификация шумов, которые могут присутствовать в аудиосигнале; фильтрация аудиосигнала с использованием различных фильтров; дробление аудиосигнала на фреймы и применение к ним оконных функций. Были описаны различные методы извлечения акустических признаков из исходного аудиосигнала: MFCC, LPC(C), PLP. Собранные в работе сведения могут быть использованы при разработке систем распознавания речи.

Список использованных источников:

1. Mporas, I. Examining the Influence of Speech Frame Size and Number of Cepstral Coefficients on the Speech Recognition Performance [Electronic resource] / I. Mporas, T. Ganchev, I. Kotinas, N. Fakotakis // ResearchGate. – Режим доступа: https://www.researchgate.net/profile/Todor-Ganchev/publication/239546404_Examining_the_Influence_of_Speech_Frame_Size_and_Number_of_Cepstral_Coefficients_on_the_Speech_Recognition_Performance/links/0c9605322c413b47db000000/Examining-the-Influence-of-Speech-Frame-Size-and-Number-of-Cepstral-Coefficients-on-the-Speech-Recognition-Performance.pdf?origin=publication_detail. – Date of access: 15.03.2023.

2. Алюнов, Д. Ю. Реализация алгоритма обработки и распознавания речи / Д. Ю. Алюнов, Е. С. Сергеев, П. В. Пигачев, А. Н. Мытников // Современные наукоемкие технологии. – 2016. – № 3. – С. 225 – 230.

3. Судьенкова А. В. Обзор методов извлечения акустических признаков речи в задаче распознавания диктора / А. В. Судьенкова // Сборник научных трудов НГТУ. – 2019. – № 3–4. – С. 139–164.

4. Мел-кепстральные коэффициенты (MFCC) и распознавание речи [Электронный ресурс] // Хабр. – Режим доступа: https://habr.com/ru/post/140828/. – Дата доступа: 12.02.23.

5. Воробьева, С. А. Выделение границ фонем речевого сигнала с помощью мел-частотных спектральных коэффициентов / А. С. Воробьева // Молодой ученый. – 2017. – № 13. – С. 2–5.

6. Гуртуева, И. А. Аналитический обзор и классификация методов выделения признаков акустического сигнала в речевых системах / И. А. Гуртуева, К. Ч. Бжихатлов // Известия Кабардино-Балкарского научного центра РАН. – 2022. – Вып. 1. – С. 41–58.

7. Зо Хеин Мин. Построение системы распознавания речевых сигналов / Хеин Мин Зо, В. М. Довгаль, В. А. Кудинов // Экономика. Информатика. – 2019. – Т. 46, №2. – С. 367–374.

8. Namrata Dave. Feature extraction methods LPC, PLP and MFCC in speech recognition [Electronic resource] / Dave Namrata // International Journal For Advance Research in Engineering And Technology. – Mode of access: https://www.researchgate.net/profile/Namrata-Dave-2/publication/261914482_Feature_extraction_methods_LPC_PLP_and_MFCC_in_speech_recognition/links/562dce4908ae04c2aeb4aa1b/Feature-extraction-methods-LPC-PLP-and-MFCC-in-speech-recognition.pdf?origin=publication_detail. – Date of access: 12.03.2023.

9. Борискевич, А. А. Анализ частотных и временных свойств слухового аппарата: Метод. указания к лабораторной работе по дисциплинам «Цифровая обработка речи и изображений» и «Защита речевых сообщений и объектов связи от несанкционированного перехвата» для студентов спец. «Сети телекоммуникаций» дневной, вечерней и заочной форм обучения / Сост. А. А. Борискевич, В. К. Конопелько. – Мн.: БГУИР, 2003. – 19 с.: ил.

10. Коваленко, А. Кривые равной громкости [Электронный ресурс] / А. Коваленко // Создание электронной музыки. – Режим доступа: https://fierymusic.net/teoriya-zvuka/krivye-ravnoy-gromkosti. – Дата доступа: 17.03.2023.

11. Метод шумоочистки речевых сигналов на основе мел-частотных кепстральных коэффициентов с использованием фильтрации Калмана / С. М. Горошко, С. Н. Петров // Известия Гомельского государственного университета имени Ф. Скорины. – 2009. – № 6 (117). – С.103–107.

UDC 004.934

ACOUSTIC FEATURES EXTRACTION FROM AUDIOSIGNAL IN SPEECH RECOGNITION SYSTEMS

Kreis A.V.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Bobrova N.L. – PhD in Technical Sciences

Annotation. Nowadays speech recognition systems are widespread in different spheres of humans` life. Speech recognition system refers to a system that allows to represent humans` speech located in an audio signal in the form that can be interpreted by a computer. At the present time different variants of such systems implementation exist. They possess common traits, for instance, the necessity of the acoustic features extraction from the source audio signal that are used by the algorithm(s) for the recognition of words that were pronounced by a speaker. In this work an overview of such feature set extraction process is conducted.

Keywords. Acoustic feature, speech recognition system, median filter, noise filtration, discrete Fourier transform, Hamming window, Mel-frequency cepstral coefficients, Perceptual linear prediction, Linear predictive coding, frame, critical band, equal-loudness-level-contours, discrete cosine transform.