МЕТОДИКА РЕКОНСТРУКЦИИ ФОНЕМ ГОЛОСА ЧЕЛОВЕКА

Лань Г.
Научный руководитель – Фадеев А.С.
Томский политехнический университет
lanber@tpu.ru

1. Введение

С повсеместным использованием цифровой техники высокую актуальность приобрела задача распознавания речи человека с целью управления устройствами с помощью голосовых команд. Традиционно [1, 2, 3] задача распознавания голоса декомпозируется на три подзадачи: идентификация отдельных звуков букв и слов, идентификация отдельно голоса человека на фоне звучания других голосов и общего шума, идентификация эмоций и интонационной окраски речи.

Целью настоящей работы является формирование аналитического описания фонем на основе связных параметров, которое позволит проводить исследование динамических свойств речи человека, осуществлять моделирование и синтез звуков, букв, слов и предложений человеческой речи.

2. Анализ формант человеческой речи

Слово состоит из одного или нескольких слогов, которые в свою очередь состоят из одной или нескольких фонем [4]. В настоящей работе анализируется отрезок времени, в течение которого фонема становится квазистационарной. На протяжении этого времени фонема почти неизменна, как и ее спектральные составляющие.

Для синтеза квазистационарной части фонем было предложено декомпозировать звук выдержки фонемы на отдельные простые компоненты частотного спектра – форманты [4, 5], получить их аналитическое описание и исчерпывающий перечень параметров, позволяющих создать математическую модель для синтеза процесса выдержки каждой форманты и звука фонемы в целом.

Векторы параметров формант были представлены в виде параметрической матрицы:

$$M_F = \left(F_1 \quad F_2 \quad F_3 \quad \dots \quad F_N\right), \qquad F_i = \left(P_1 \quad P_2 \quad P_3 \quad \dots \quad P_M\right), \qquad (1)$$

где F_i– вектор параметров i-ой форманты, M_F– матрица параметров фонемы.

Для анализа фонемы и ее формант использовались амплитудно-частотные характеристики (АЧХ), представляющие собой спектр сигнала. АЧХ были получены применением быстрого преобразования Фурье (БПФ) к исходной последовательности заранее дискретизированного сигнала.

На рисунке 1 представлена АЧХ фонемы «А». На рисунке показано, что основная энергия звука сконцентрирована в диапазоне 100–1500 Гц, а наибольшей амплитудой обладает шестая форманта, имеющая частоту 1251 Гц. Более высокочастотные гармоники, частоты выше 1500 Гц, имеют меньшую амплитуду, фактически они едва различимы на фоне шума. В виду малой амплитуды, такими частотами можно в дальнейшем пренебречь.

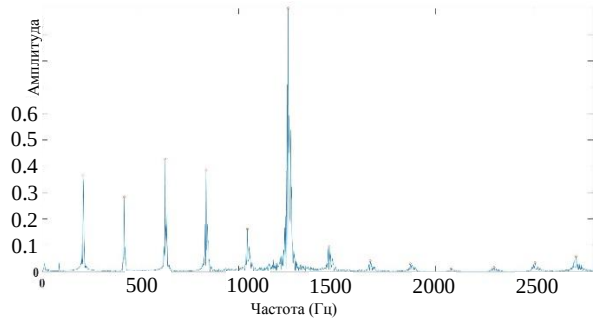

Рис. 1. Спектр фонемы буквы «А».

Для реконструкции отдельных формант был предложен метод параметрического описания формант речи человека на основе полученных АЧХ. Определено два параметра, однозначно характеризующих статические свойства формант в процессе выдержки: амплитуда и частота.

$$F_i = \left(A_i \quad v_i\right), \qquad (2)$$

Где A_i– значение амплитуды i-ой форманты; v_i – значение частоты i-ой форманты.

В частотном распределении (рисунок 1) имеется несколько пиков – один такой пик соответствует одной форманте. Соответственно точка экстремума пика характеризуется двумя величинами: амплитудой A и частотой v.

Анализ АЧХ и полученных параметров позволил выделить отличительные особенности формант гласных букв.

3. Синтез фонем

Для реализации возможности оценки качества описанного метода, была предложена модель реконструкции фонем речи человека по полученным параметрическим матрицам:

Где $A_i \cdot \sin\left(2\pi v_i t\right)$ – формантная компонента (форманта) сигнала; A_i – значение амплитуды; v_i – значение частоты форманты; t – время.

В таблице 1 приведены матрицы параметров

$$f\left(t\right) = \sum_{i=1}^{N} A_i \cdot \sin\left(2\pi v_i t\right), \qquad (3)$$

двух гласных букв алфавита «О» и «А». Форманты буквы «О» обладают следующими характеристиками: номера основных формант – 2, 3, 4; диапазон частот распределения энергии – 80—1200 Гц.

Таблица 1. Значения параметров формант гласных букв

	«О»		«А»	
	A_i	v_i	A_i	v_i
F_1	0,0474	87	0,3645	208
F_2	0,9109	224	0,2842	417
F_3	0,6099	447	0,4272	625
F_4	1	671	0,3857	833
F_5	0,0765	905	0,1626	1043
F_6	0,0209	1118	1	1251
F_7	-	-	0,0959	1459
F_8	-	-	0,0417	1668
F_9	-	-	0,0304	1871

Оценка параметрической матрицы, значения которой определены эмпирически, происходит путем сравнения синтезированного по данным параметрам сигнала с оригинальным. На рисунке 2 представлены АЧХ оригинальной и синтезированной фонем.

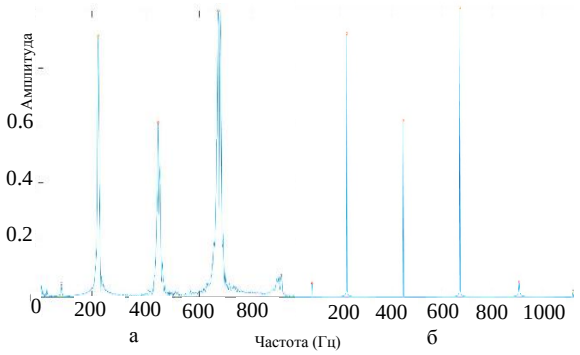

Рис. 2. Спектры оригинальной (а) и синтезированной (б) фонем буквы «О».

Набор параметров значений матриц зависит от свойств фонемы. Так, для реалистичной реконструкции звукозаписи гласной буквы «О» применена матрица, состоящая из восемнадцати числовых параметров, описывающих девять значимых формант. Для построения более точной модели, необходимо учитывать все значимые форманты фонемы. Другим условием точности сравнения оригинального и синтезированного сигнала является равная продолжительность звучания сигналов.

Заключение

В настоящей статье предлагается метод получения связанных параметров, описывающих амплитудно-частотные свойства отдельных формант гласных букв. В процессе получения значений параметрических векторов формант показана возможность декомпозиции сигнала на отдельные гармоники для относительного изолированного описания динамики их АЧХ, дальнейшего модерирования сигнала на их основе и оценки качества созданной модели.

Разработан метод получения параметрической матрицы для описания звука голоса человека, состоящей из параметров всех векторов всех значимых формант. Данные параметры отражают относительную громкость каждой форманты в общем потоке, а также они характеризуют общие закономерности фонемы. Метод позволяет выполнять моделирование и реконструкцию звука отдельной фонемы на основе полученных параметрических матриц. Состоятельность данного метода подтверждена созданием банка параметрических матриц на основе записей речи нескольких человек и проведенной оценкой синтезированных для данного банка сигналов.

Список использованных источников

1. Zhang, X. Digital Speech Processing and MATLAB Simulation (Second Edition) / Электронная промышленность // X. Zhang – ЦНИИ «Электроника», 2016.
2. Rabiner, L. Theory and Applications of Digital Speech Processing / Journal of Applied Mechanics // L. Rabiner, R. Schafer – 2010. – №30. – vol.1. – Pp. 445—447.
3. Фролов, А.В. Синтез и распознавание речи. Современные решения [Электронный ресурс] / А.В. Фролов, Г.В. Фролов – 2013. – Режим доступа: http://frolov-lib.ru/books/hi/index.html (дата обращения: 17.06.2018).
4. Лань Г., Моргунов А.Н., Моделирование и синтез фонем гласных букв, «Вестник современных исследований» Выпуск № 10-3 (25) 2018, ISSN 2541-8300. с.130 – 135.
5. Фадеев А. С., Кочегурова Е. А. Метод преобразования форматов музыкальной информации /А. С. Фадеев, Е. А. Кочегурова//Цифровая обработка сигналов. - 2007. -№ 3. -С. 49.