А. Ю. Петрович

АЛГОРИТМ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ТЕМПА РЕЧИ

В работе рассматривается возможность решения задачи определения темпа речи посредством сегментации сигнала и использования частотного анализа сигнала.

Введение

В работе под характеристикой темпа речи понимается количество слогов внутри просодической единицы на единицу времени. В качестве такого анализируемого фрагмента, как правило, выступает фонетическое слово. Данным термином называется отрезок речевой цепи, объединяемый одним (словесным) ударением, который произносится как единое целое («на улице», «в университете»). Ограничение просодической единицей необходимо для уменьшения влияния пауз в речи.

Задачу идентификации слогов можно свести к задаче определения гласных в речевом потоке. Главное отличие гласных от согласных фонем – наличие резонансной составляющей спектра. Таким образом определение гласных можно свести к анализу резонансных частот спектра гласных, называемых формантами. Форманты гласных русского языка были определены эксперементально во множестве исследований [1] и их можно использовать для создания математической апроксимации человеческой речи.

I. Построение синтезатора речи

Для моделирования сигнала, близкого к гласным воспользуемся методами линейного предсказания речи [2]. Для симуляции гласных используется комбинация синтезатора частот с рекурсивным фильтром. Для симуляции других звуков – белый шум пропущенный через этот же фильтр. Передаточная функция фильтра будет иметь вид:

$$T(z) = \frac{1}{1 - H(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}, \qquad (1)$$

где a_k – коэфициенты фильтра, получаемые методом минимизации квадрата ошибки сравнения полученного сигнала с реальными образцами речи.

Схема фильта преставлена на рисунке 1.

II. Распознавание гласных

Для анализа спектра полученного сигнала воспользуемся библиотекой SFS [3]. В ней реализован метод получения Формантных частот методом линейного предсказания речи и использует методы автокореляции и весовые функции для анализа. На выходе получаем матрицу содержащую частоты первых трех формант и аплитуды, что позволяет составить гистограмму и отметить на ней чатсотное распределение для различных формант и гласных. Из них также можно сделать вывод о ошибочном опредении фильтра гласных, анализируя перекрытия гистограмм.

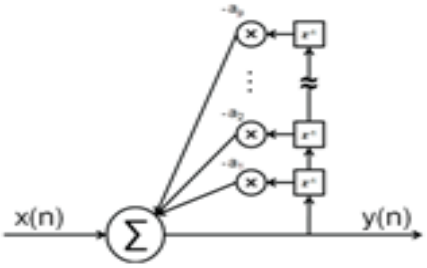

Рис. 1 – Схема БИХ-фильтра для моделирования гласных звуков речи

III. Выводы

Процесс определения темпа речи может рассматриваться, как состоящий из серии статистически независимых событий детектирования гласных, что дает возможность складывать, а не умножать ошибки от факта неверного определения. В результате экспериментов была достигнута точность определения гласных около 80%. Таким образом возможно определение темпа речи путем частотного анализа на основе линейного предиктивного програмирования.

1. Информационные процессы, Том 4 №2 / Сорокин В. Н., Цыплихин А.И. – Институт проблем передачи информации, Российская академия наук, Москва, Россия, 2004.
2. Recognition of Vowels in Continuous Speech by Rising Formants / FACTA UNIVERSITATIS (NIS)// SER.: ELEC. ENERG. vol. 23, no. 3, December 2010, 379-393.
3. Официальная документация SFS [Электронный ресурс]. Режим доступа: https://phon.ucl.ac.ak/resource/sfs/help/intex.php Дата доступа: 22.03.2020.

Петрович Андрей Юрьевич, магистрант кафедры систем управления Белорусского государственного университета информатики и радиоэлектроники, petrovichandrey17@gmail.com.

Научный руководитель: Захарьев Вадим Анатольевич, доцент кафедры систем управления Белорусского государственного университета, кандидат технических наук, zahariev@bsuir.by.