

требуется выбрать последовательность переменных, по которой ведётся многоуровневое разложение Шеннона. На следующем этапе строится диаграмма двоичного выбора по заданной перестановке переменных. Далее происходит поиск одинаковых поддеревьев в дереве (графе представления системы функций). Одинаковые поддеревья удаляются и заменяются уже найденным аналогичным поддеревом, тем самым удаляются лишние узлы, и сложность представления функций уменьшается.

Ключевым этапом в оптимизации является поиск одинаковых поддеревьев. Он начинается с листьев и продолжается до тех пор, пока не обнаружится, что одинаковых поддеревьев в дереве больше нет. Также важным этапом оптимизации булевых функций с помощью диаграмм двоичного выбора является выбор порядка переменных, по которым ведётся разложение Шеннона. От этого выбора зависит количество одинаковых поддеревьев, а следовательно, и степень оптимизации. Хорошей эвристикой при выборе перестановки переменных является порядок, в котором сначала идут наиболее часто встречающиеся переменные, а в конце - наименее встречающиеся в исходной системе дизъюнктивных нормальных форм (ДНФ), задающей исходную систему либо уже полученные подфункции. Этот подход прост в реализации и даёт хорошие результаты, так как существует большая вероятность обнаружения «больших» одинаковых поддеревьев, и это позволяет упростить граф.

Зачастую встречаются ситуации, когда используются неполностью определённые булевы функции. Неполностью определённые функции, или частичные функции, - это функции, значения которых не определены на некоторых наборах значений переменных, на данных наборах значения функций при оптимизации заменяются определёнными (0,1). При логической оптимизации неполностью определённых функций при помощи диаграмм двоичного выбора алгоритмы выбора перестановки и построения диаграммы аналогичны. Однако при поиске одинаковых поддеревьев неопределённые листовые значения следует дополнять до значений 0 или 1, чтобы получилось наибольшее число одинаковых поддеревьев. Таким образом, общий алгоритм логической оптимизации неполностью определённых функций начинается с этапа выбора порядка переменных разложения. Далее идёт этап разложения по выбранной перестановке переменных. На третьем этапе производится анализ всего дерева и дополнение неопределённых листьев до значения 0 или 1. На заключительном этапе производится поиск одинаковых поддеревьев и удаление избыточностей.

Список использованных источников:

1. Карпов, Ю. Верификация параллельных и распределённых программных систем /Ю. Карпов. - СПб: «БХВ-Петербург», 2010 – 560 с.
2. Бибило, П. Н. Применение диаграмм двоичного выбора при синтезе логических схем /П. Н. Бибило. – Минск: «Беларуская навука», 2014 – 231с.
3. Бибило, П. Н. Логическое проектирование дискретных устройств с использованием продукционно-фреймовой модели представления знаний /П. Н. Бибило, В. Н. Романов – Минск.: «Беларуская Навука», 2011. - 279 с.
4. Бибило, П. Н. Минимизации многоуровневых представлений систем булевых функций на основе разложения Шеннона / П. Н. Бибило, Ю. Ю. Ланкевич // Информатика, 2017 – № 2. – С. 45- 57.
5. Бибило П. Н. Алгоритм построения диаграммы двоичного выбора для системы полностью определённых булевых функций / П. Н. Бибило, П. В. Леончик - // Управляющие системы и машины. - 2009. - № 6. – С. 42 – 49.

ИССЛЕДОВАНИЕ КРИТЕРИЕВ ОЦЕНКИ КОРРЕКТНОСТИ ПРОИЗНОШЕНИЯ ФРАЗ ДЛЯ СИСТЕМЫ ИСПРАВЛЕНИЯ РЕЧЕВЫХ ДЕФЕКТОВ

Демидович В.С.

*Белорусский государственный университет информатики и радиоэлектроники
Г. Минск, Республика Беларусь*

Лихачев Д.С. – к.т.н., доцент

Речь человека имеет огромное количество характеристик таких как темп речи, громкость, высота голоса, интонация, тембр и др. Эти характеристики делают речь каждого человека уникальной и неповторимой, что в свою очередь усложняет ее анализ.

В данной статье описан метод, позволяющий определить правильность произношения фразы или слова, основываясь на *DTW-алгоритме* с использованием мел-кепстральных коэффициентов. Данные коэффициенты, расположенные на мел-шкале, позволяют выделить наиболее значимые для восприятия человеком частоты.

Для определения десяти мел-кепстральных коэффициентов используется частотный диапазон от 75 Гц до 5000 Гц. Для преобразования частоты в мел используется формула (1):

$$M = 1127 * \log \log \left(1 + \frac{F}{700} \right) \quad (1)$$

Обратное преобразование производится по формуле (2):

$$F = 700 * \left(e^{\frac{M}{1127}} - 1 \right) \quad (2)$$

Условно разделим человеческую голос на мужской (высокий, средний и низкий) и женский (высокий, средний и низкий). В качестве эксперимента был использован мужской средний голос. Важно отметить, что эталонные образцы голоса и образцы голоса тестируемого человека имеют одинаковую длину. Для каждой фразы имеется по три записи: эталонная запись (эталон), запись человека с нормальным произношением (норма) и запись человека с нарушением речи (дефект).

Для анализа речевого сигнала весь аудиосигнал разбивается на фреймы длительностью 50 мс. Далее рассчитывается спектр сигнала с помощью быстрого преобразования Фурье.

Для расположения полученного ранее спектра на мел-шкале необходимо создать «ребенку» фильтров. Каждый фильтр – это треугольная функция, позволяющая получить суммарное количество энергии на определенном диапазоне частот.

Диапазон частот [75; 5000] на мел-шкале согласно формуле (1) равен [114.7; 2363.5]. Для построения десяти треугольных фильтров необходимо двенадцать опорных точек.

Получим: $m = [114.7; 319.1; 523.5; 729.9; 934.3; 1138.7; 1343.1; 1547.5; 1751.9; 1956.3; 2160.7; 2363.5]$.

Используя формулу (2) переведем мел-шкалу в герцы.

Получим: $h = [75; 229; 414; 638; 904; 1223; 1605; 2063; 2613; 3272; 4061; 5000]$.

Далее необходимо наложить шкалу на спектр фрейма по формуле (3):

$$f(i) = (frameSize + 1) * \frac{h(i)}{sampleRate}, \quad (3)$$

где $frameSize$ – длина фрейма; $h(i)$ – значение опорной точки в Гц; $sampleRate$ – частота звука в Гц, равная половине частоты дискретизации; $f(i)$ – соответствующие опорные точки на оси спектра.

Зная опорные точки на оси спектра, строим фильтры, используя формулу (4):

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, & f(m) \leq k \leq f(m+1) \\ 0, & k > f(m+1) \end{cases} \quad (4)$$

где k – номер отсчета фрейма; $f(m)$ – опорные точки на оси спектра.

Для получения мел-коэффициентов необходимо полученные значения попарно перемножить с энергией спектра, после чего прологарифмировать (5). Таким образом понижается чувствительность коэффициентов к шумам.

$$S[m] = \log \left(\sum_{k=0}^{N-1} |X[k]|^2 * H_m[k] \right), 0 \leq m \leq 10 \quad (5)$$

Для получения мел-кепстральных коэффициентов необходимо применить дискретное косинусное преобразование (ДКП) к мел-коэффициентам. ДКП используется для повышения значимости первых коэффициентов и уменьшения значимости последних.

В результате для каждого аудиосигнала получена матрица $N(m,k)$, где m – количество кепстральных коэффициентов (в нашем случае десять), k – количество фреймов аудиосигнала.

Таким образом, для каждой фразы имеется три матрицы (эталон, норма, дефект). С помощью dtw-алгоритма определяется минимальная сумма евклидовых расстояний между матрицами эталона и нормы; матрицами эталона и дефекта.

Проведя десяток экспериментов было установлено, что среднее арифметическое разности сумм евклидовых расстояний равно 7,73, а минимальная разность – 6,91.

Кепстральный коэффициент 3 для слова «двери» в зависимости от времени показан на рисунке 1а, на рисунке 1б изображены кепстральные коэффициенты четвертого фрейма для данного слова.

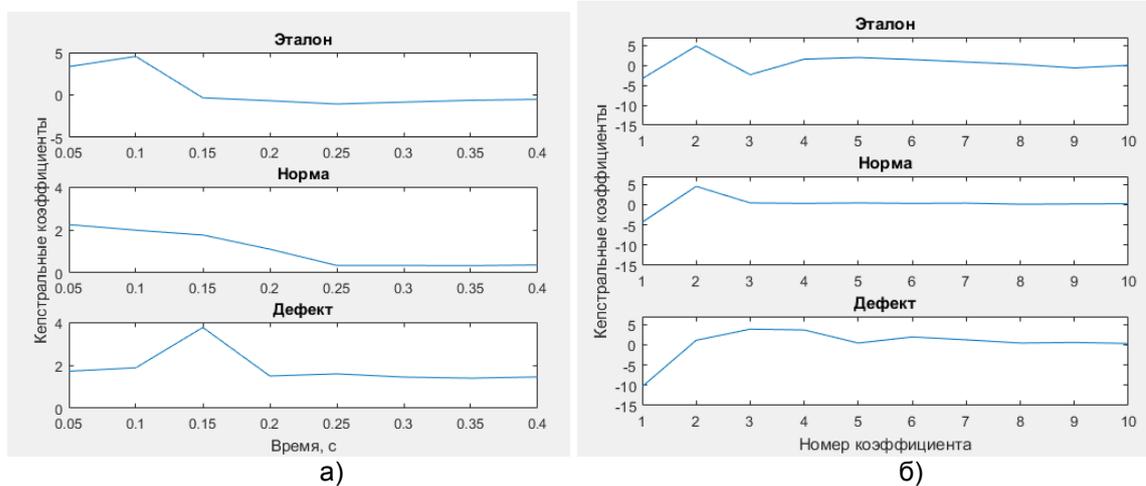


Рисунок 1 - Кепстральный коэффициент 3 (а), кепстральные коэффициенты четвертого фрейма (б).

Можно сделать вывод о возможности использования кепстральных коэффициентов для определения правильности произношения фразы или слова. Граница разности сумм евклидовых расстояний при которой можно считать, что человек верно произнес фразу равна 6,8.

Список использованных источников:

1. Davis, S. Mermelstein, P. (1980) *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences*. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp. 357-366.
2. Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, Prentice Hall, 2001, ISBN:0130226165.

ИНТЕЛЛЕКТУАЛИЗИРОВАННАЯ СИСТЕМА УПРАВЛЕНИЯ СВЕТОВЫМИ ЭФФЕКТАМИ

Добровольский Д.А.

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

Шемаров А.И. – д-р. техн. наук, профессор

В современных реалиях большинство областей развлекательной индустрии подразумевают использование автоматизированных систем для генерации контента. Однако существуют области, где всё ещё необходим исключительно человеческий труд, и одна из этих областей - работа со сценическим оборудованием. Поскольку конструкция освещения сцен варьируется от случая к случаю, время, затрачиваемое на подготовку и программирование освещения велико, в основном из-за полностью ручного программирования каждого осветительного устройства.

Данная система направлена на удешевление и ускорение процесса подготовки сцены к выступлению путем применения системы предварительного программирования управляемых прожекторов и создание базовой программы освещения сцены. Используя сгенерированную программу освещения, режиссеру освещения остается лишь внести небольшие изменения, соответствующие желанию заказчика - такие как смена цветовой гаммы и добавление уникальных сцен освещения.

Сценическое освещение - это набор управляемого светового оборудования, способное менять свойства направления, яркости, цвета, форму светового луча или сразу несколько из них. Устройства освещения подключаются последовательно к общей сети передачи данных и работают независимо друг от друга. Для передачи команд управления на данные устройства часто используется порт RS-485 с использованием протокола DMX512. Порт представляет собой разъем с тремя или пятью пинами, отведенные под передачу, прием данных и заземление соответственно. Протокол DMX (Digital Multiplex) является простой реализацией передачи данных, состоящих из 513 байт, где главный байт отводится на инициализацию и информирование о формате сессии, а последующие 512 байт несут информацию для устройств. Если взять за основу монохромный статичный прожектор с единственным элементом контроля яркостью, то на одну шину возможно подключить до 512 таких устройств. Если же добавить к функционалу прожектора добавить возможность изменять цвета, то количество информации к каждому прожектору потребует передавать 3 байта информации в RGB-