

ПРИМЕНЕНИЕ S-ФУНКЦИИ ДЛЯ СИНТЕЗА ЭМОЦИОНАЛЬНО ОКРАШЕННОГО ЗВУЧАНИЯ ГЛАСНЫХ БУКВ

Фадеев А.С. к.т.н, доц.,
Лань Гунбо,
Томск политехнический университет
Email: langongbo2012@163.com

Введение

С ростом популярности использования компьютеров и развитием современной компьютерной науки все большее применение в различных областях человеческой деятельности находят интеллектуальные компьютерные технологии. В области работы с речью человека одной из немаловажных является технология синтеза речи. Целью данной работы является разработка метода, позволяющего анализировать эмоциональные свойства речи человека и получать аналитическое описание изменения частот отдельных формант голоса, а также синтезировать звуки фонем с эмоциональной окраской на основе полученных функций.

Частотная характеристика эмоционального звучания

Анализ оригинальной звучания

Как описано в исследованиях [Ошибка! Закладка не определена., Ошибка! Источник ссылки не найден., 3], энергия голоса распределяется во времени только в некоторых частотных диапазонах и различается по величине. В них предложены методы, позволяющие синтезировать отдельные звучания звонких букв, как сложных сигналов, состоящих из отдельных формант. Каждая форманта описывается вектором из двух параметров: амплитуда и частота звучания. Но в результате синтеза звуков с полученными параметрами в звучании отсутствует эмоциональная окраска. Интонация является одним из важных признаков для эмоциональной речи. На рис.1 показано представление динамики с изменением интонации (динамики с подъем эмоционального произношение) на буквы «А» в виде спектра.

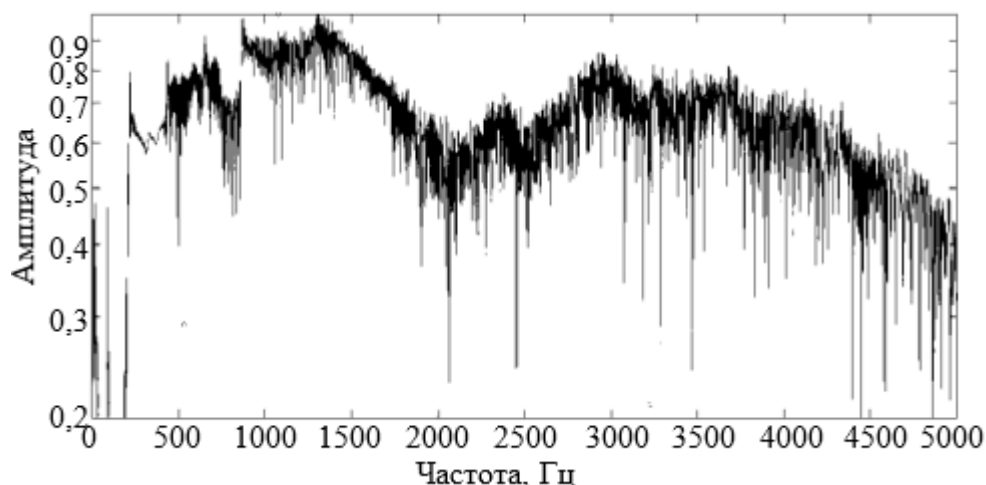


Рис. 1. Спектр буквы «А» с эмоциональным «подъемом»

Анализ изображения показывает, что звучание имеет очень сложные частотные составляющие. В спектре практически невозможно различить отдельные форманты.

В настоящей работе были изучены наборы звукозаписей речи диктора с различной эмоциональной окраской. На рисунке 2 приведена спектрограмма динамики с подъем и спадом эмоционального произношения буквы «А». Анализ изображения показывает, что значения частоты каждой форманты увеличивается почти в два раза на интервале времени с 0,4 по 0,7 секунду.

Исследование спектрограмм наборов звукозаписей показало наличие закономерностей в изменении частотных характеристик всех формант отдельной фонемы при наличии эмоциональной окраски: относительное изменение частоты во времени каждой форманты для всех формант одной фонемы подчиняется одному и тому же закону. Таким образом, для синтеза эмоционально окрашенных

звучания предложена математическая модель, описывающая в аналитическом виде изменение частот для всех формант одной фонемы.

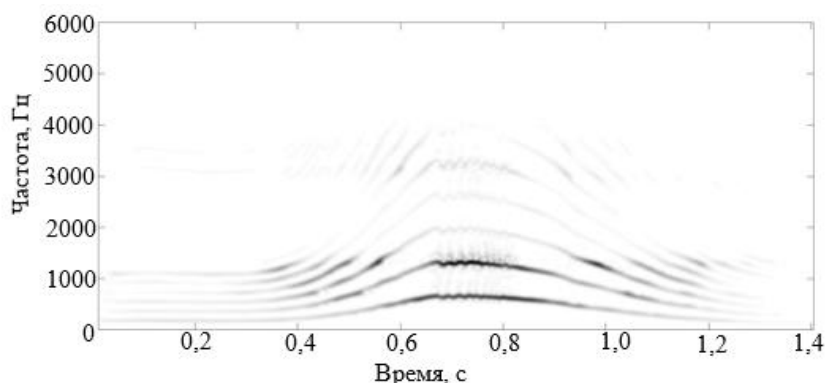


Рис. 2. Спектрограмм букв «А» с подъем и спадом

Математическая модель для генерирования изменяемой частота синтезированного звучания

Анализ звукозаписей диктора показал, что достаточно часто эмоциональная окраска выражается в изменении интонации голоса диктора: присутствует либо подъем, либо спад частоты основного тона. Поэтом в работе предложено в основу синтеза изменяющейся частоты положить функцию сигнмоида, представляющую собой гладкую монотонную возрастающую нелинейную функцию, имеющую форму буквы «S» (*s* – функция) [2], которая часто применяется для «сглаживания» значений некоторой величины. Форма сигмоида представляется в рис.5. его функция имеет следующий вид:

$$s(t) = s_1(t) \cdot s_2(t) = k \cdot \frac{1}{(k_1 + c_1 \cdot e^{a_1(t+b_1)})} \cdot \frac{1}{(k_2 + c_2 \cdot e^{-a_2(t-b_2)})} \quad (1)$$

Где k – базисное значение частоты. k_1, k_2 – коэффициент изменения частоты. a_1, a_2 – параметры, определяющие скоростью изменения частоты подъема и спада. b_1, b_2 – параметры, определяющие временные характеристики сигнала. c_1, c_2 – параметры, определяющие начальную и конечную точки изменения сигнала. t – время.

Для синтезировать фонем с изменяемыми значениями частотами будем объединить функции сигмоида и с фундаментальной частот, и добавлять их в математическую модель, получим следующую функцию:

$$f_i(t) = A_i \sin(2\pi \int s_i(t) dt). \quad (3)$$

Где A_i – значение амплитуда каждой формант.

Сигнал, имитирующий звучание отдельной фонемы или гласной буквы, получается алгебраическим сложением сигналов всех формант $f_i(t)$.

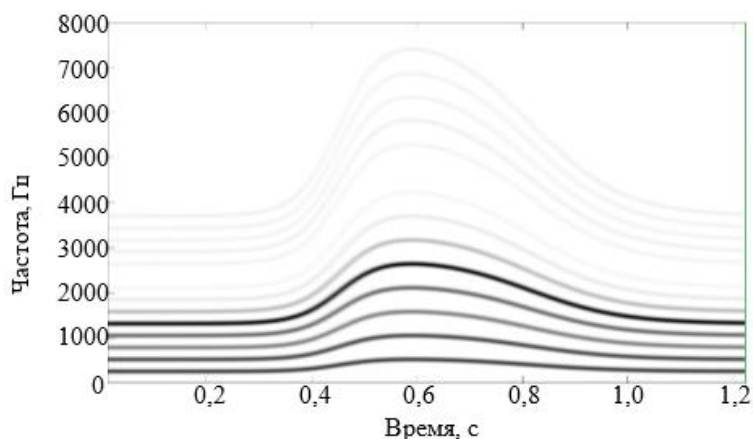


Рис. 3. Спектрограмм синтезированной эмоциональной фонем «А»

При помощи формулы 3 можно генерировать синтезированный сигнал; процесс реконструкции сигнала реализуется в MATLAB, на рис.3 показана спектрограмма синтезированной фонемы буквы «А» с эмоциональной окраской, полученной с применением данной математической модели. Для получения значений параметров амплитуды и частоты каждой форманты используются фонема «А» без эмоциональной окраски.

Анализ изображения показывает, что значение частот всех формант начинает увеличиваются в точке 0,4 с., в точке 0,5 с. значение частоты всех формант достигает заданного максимального значения и потом начинается понижение частоты всех формант. Результат анализа оригинального и синтезированного звучания фонемы показал, что интонация синтезированной эмоционально окрашенной фонемы, полученной на основе функций сигмоида, соответствует интонации оригинального звучания. Также слуховой анализ показал, что синтезированное звучание имеет электронную окраску звука, за счет чего сходство между синтезированной фонемой и оригинальной фонемой звучание было не абсолютным.

При сравнении спектрограмм синтезированной фонем (рис.3) и оригинальной фонем (рис.2) видно, что, когда значение частота формантов оригинальной фонем изменяются в спектрограмме, амплитуды форманты будут также изменяться. Цвет форманты в настоящем рисунке спектрограмм обозначает значение амплитуды, чем темнее цвет на рисунке, тем больше значение амплитуды. Таким образом, можно узнать, что изменение интонации не только связано с изменением значения частот формантов, а также зависит от изменения значение амплитуды формантов.

Заключение

В данной работе предложена методика описания закономерности изменения частоты фонем с эмоциональной окраской при анализе спектра и спектрограмм фонем голоса, записанного с различными интонациями. Предложена математическая модель с применением функции сигмоида и методика, позволяющая синтезировать сигнал эмоциональной фонем человека на основе полученных модели, а также указано недостатки данной методики и выставлено возможное решение.

Список литературы

1. Лань Г. Методика реконструкции фонем голоса человека / Г. Лань, А. Н. Моргунов // Вестник современных исследований Выпуск. – 2018 – № 10-3 (25). – С. 130–135.
2. Сигмоида, Материал из Википедии, [электронный ресурс] режим доступ: <https://ru.wikipedia.org/wiki/Сигмоида>
3. Иванов, А. В., Трушин, В. А., Маркелова, Г. В., Рева, И. Л. Исследование спектра формант форсированной речи. *Научный вестник Новосибирского государственного технического университета*, 2015, (4), 63-73.
4. Бондаренко В П, Бондарь В А. Измерение некоторых характеристик гласных звуков. *Известия Томского политехнического института*, 1974, 246: 39-41.
5. Волковец А И. Создание и обработка звука при разработке интерактивных приложений, 2017.