Правительство Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования
«Санкт-Петербургский Государственный Университет»

Кафедра Теоретической Кибернетики

Лебедев Владимир Вячеславович

Синтез речевого сигнала

Дипломная работа

Допущен к защите.
Зав. кафедрой:
д. т. н., профессор Фрадков А. Л.

Научный руководитель:
д. ф.-м. н., профессор Барабанов А. Е.

Рецензент:
к. ф.-м. н., доцент Бондарко В. А.

Санкт-Петербург
2016

SAINT-PETERSBURG STATE UNIVERSITY

Department of Theoretical Cybernetics

Vladimir Lebedev

Synthesis of speech signal

Graduation Thesis

Admitted for defense.
Head of department:
Professor Alexander Fradkov

Scientific supervisor:
Professor Andrey Barabanov

Reviewer:
Docent Vladimir Bondarko

Saint-Petersburg
2016

Оглавление

Введение

В данной работе будет решаться задача улучшения качества синтезированного сигнала. Всегда есть смысл делать это, например, для фонетиков - людей изучающих речь, важно работать с сигналом наиболее приближенным к идеалу. Иногда, из-за недостаточно достоверной модели возникают так называемые «хлопки». На вход программы подается речевой сигнал. В таких согласных как «ж», «ш», «ц» происходят быстроменяющиеся оценки частоты основого тона. Все программы были написаны в математическом пакете MATLAB. Имеется программа VOCODER с помощью которой происходит высчитывание параметров модели по данному сигналу, а затем воспроизведение сигнала по построенной модели.[2]

1. Постановка задачи

На вход программы подается речевой сигнал. Фрагмент записи будет рассматриваться наполовину перекрывающимися частями, которые называют фреймами. Звуки речи делятся на вокализованные(тоны) и невокализованные(шумы). В данной работе будут рассматриваться только вокализованные. На вокализованном участке голосовые связки совершают периодические колебания с частотой основного тона. Далее представлен фрагмент аудиозаписи «жу»:

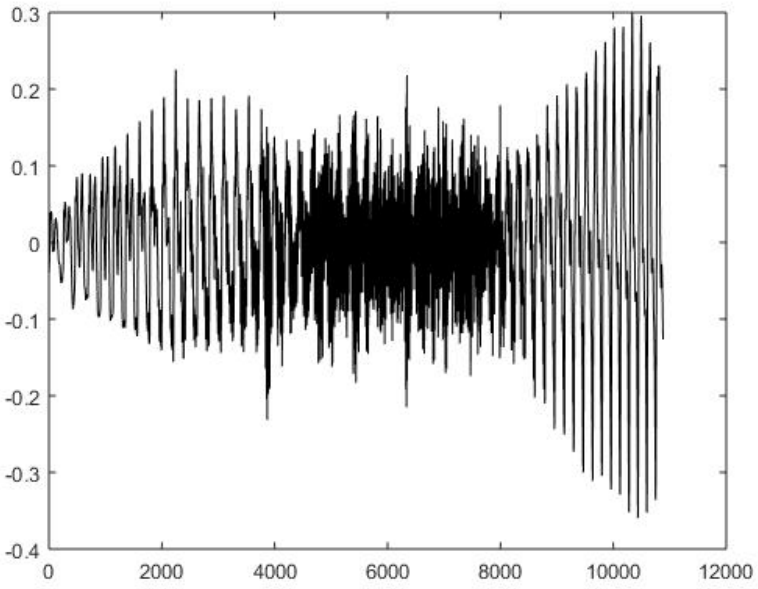

Рис. 1: Речевой сигнал

1.1. Цель работы

Целью работы является разработка методов синтеза речевого сигнала по параметрическим моделям его коротких сегментов, а также оптимизация параметров моделей речевого сигнала по методу ”анализ через синтез”. Предполагается, что фиксирован способ оценки амплитуд и фаз всех гармоник полигармонической модели речевого сигнала в коротком сегменте при фиксированной частоте основного тона. Для этого применяется метод наименьших квадратов в комплексном расши-

рении задачи. Требуется оценить быстро меняющуюся частоту основного тона и принять решение о вокализованности звука.

2. Математическая часть

Рассмотрим задачу о стыковке соседних голосовых фреймов[1]. Заданы амплитуды, фазы и период основного тона для левого и правого фреймов. Для левого фрейма это $(A_{k,l})_{k=0}^{K_l}, (\phi_{k,l})_{k=0}^{K_l}$ и P. Для правого фрейма соответственно, $(A_{k,r})_{k=0}^{K_r}, (\phi_{k,r})_{k=0}^{K_r}$ и P. Фазы относятся к середине промежутка времени соответствующего фрейма. Задана дробная часть τ_l середины левого фрейма.

Требуется синтезировать сглаженный сигнал, который в середине левого и правого фрейма имеет частоты, амплитуды и фазы, соответствующие своей модели, а на границе фреймов сохраняет гладкость перехода гармоник.

2.1. Соединение двух голосовых фреймов в непрерывном времени

Введем шкалу времени с нулевым значением на границе фреймов. Модели сигналов для левого и правого отрезков:

$$s_l(t) = \sum_{k=0}^{K_l} A_{k_l} \cos(2\pi k F_l * (t - t_l) + \phi_{k,l}), \quad t \in [-P, 0]$$

$$s_r(t) = \sum_{k=0}^{K_r} A_{k_r} \cos(2\pi k F_r * (t - t_r) + \phi_{k,r}), \quad t \in [0, P]$$

Предполагается, что предыдущий фрейм уже построен до точки $t = -P/2$. Для одной гармоники из левого фрейма ставится в соответствие гармоника из правого. Каждой паре гармоник сопоставляется гармоника в сумме $s(t)$, которая определяет декодированный сигнал на промежутке $[-P/2, P/2]$. Она задается по формуле:

$$s(t) = \sum_{k=0}^{K} A_k(t) \cos(2\pi F_k(t) t + \phi_k), \quad t \in [-P/2, P/2]$$

Предполагается, что гармоники сигналов s_l и s_r распадаются на па-

ры стандартным образом по сеткам частот, пропорциональным числам

$$F_l = \frac{1}{P}, \qquad F_r = \frac{1}{P}$$

Каждой паре гармоник ставится в соответствие индекс k и гармоника в сумме $s(t)$ Рассмотри некоторую пару гармоник с левым индексом k_l и правым индексом k_r. Введем обозначение для частот слева и справа

$$f_l = k_l F_l, \qquad f_r = k_r F_r,$$

Функция амплитуды $A_k(t)$ рассчитывается следующим образом: на концах промежутка $[-P/2, P/2]$ устанавливается значение $A_{k_1,1} A_{k_2,2}$ и линейно интерполируется внутрь.

Функция амплитуды $A_k(t)$ рассчитывается следующим образом. Пара амплитуд $(A_{k_l,l}, A_{k_r,r})$ устанавливается на краях промежутка времени $[-P/2, P/2]$ и линейно интерполируются внутрь этого промежутка.

Функция для фазы $\phi_k(t)$ и частоты $F_k(t)$ рассчитывается следующим образом: на границах должны быть точные значения фаз и частот:

$$F_k(-P/2) = k_l F_l, \quad F'_k(-P/2) = 0, \quad \phi_k(-P/2) = \phi_{k_l,l}, \quad \phi'_k(-P/2) = 0,$$

$$F_k(P/2) = k_r F_r, \quad F'_k(P/2) = 0, \quad \phi_k(P/2) = \phi_{k_r,r}, \quad \phi'_k(P/2) = 0$$

Далее вводится сглаживающая функция $\gamma(t)$ соединяющая гладко две гармоники, которая удовлетворяет следующим критериям:

$$\gamma(-P/2) = -1/2, \quad \gamma(-P/2) = 0, \quad \gamma(P/2) = 1/2, \quad \gamma(-P/2) = 0$$

Следующие значения помогут для вычисления фазы и частоты:

1. Среднее значение и приращение частоты

$$f^0 = \frac{1}{2}(f_1 + f_2)$$

$$\Delta f = f_2 - f_1$$

2. В средней точке $t = 0$ определяются фазы стационарных моделей

слева и справа:

$$\phi^- = (\phi_{l,k_l} + \pi k_1) \mod 2\pi$$

$$\phi^+ = (\phi_{r,k_r} - \pi k_2) \mod 2\pi$$

3. Среднее арифметическое и невязка фаз в центральной точке

$$\phi^0 = \frac{1}{2}(\phi^+ + \phi^-)$$

$$\Delta\phi^+ = [(\phi^+ + \phi^- + \pi) \mod 2\pi] - \pi$$

4. Искомые функции определяются как

$$F_k(t) = f^0 + \gamma(t)\Delta f$$

$$\phi_k(t) = \phi^0 + \gamma(t)\Delta\phi$$

2.2. Расчет сигнала в дискретном времени

В полученной модели сглаженного сигнала нужно выбрать отсчеты в равноотстоящие моменты времени

$$t_n = -P/2 - \tau_l + n, \qquad 1 \le n \le N,$$

где число N выбирается максимальным из условия $t_N \le P/2$. А именно

$$N = [P + \tau_l]$$

Дробная часть середины правого фрейма по отношению к целой сетке отсчетов равна

$$t_r = P + \tau_l$$

Отсчеты синтезированного сигнала рассчитываются по формуле

$$s_n = \sum_{k=0}^{K} A_k(t_n) cos(2\pi F_k(t_n) t_n + \phi_k(t_n)), \qquad 1 \le n \le N.$$

3. Методы выделения пар гармоник

При быстроменяющихся оценках частоты основого тона количество гармоник в соседних фреймах неодинаково, тогда возникает проблема: как соединить по парам гармоники от левого и правого фреймов так, чтобы получившийся сигнал оказался лучшим. Будем рассматривать фреймы, в которых число гармоник различно.

3.1. Объединение по номеру гармоники

Допустим $P_1 > P_2$, где P_1 и P_2 период основного тона левого и правого фрейма соответственно. Так как $P_1 > P_2$, то количество гармоник в левом фрейме будет больше. Соединим первую гармонику из левого и первую из второго фрейма, вторую со второй и так далее. В конечном результате, в первом массиве останутся несовмещенные. Добавим во второй гармонику с той же частотой, но с амплитудой равной нулю. Соединим их непрерывно, тем самым происходит избегание возникновения щелчков.

3.2. Равномерное соединение гармоник

Начальные данные те же: $P_1 > P_2$, где P_1 и P_2 период основного тона левого и правого фрейма соответственно. Соединим i-ю гармонику справа с $i * P1/P2$ гармоникой слева. Несовмещенные таким же способом, совместим непрерывно с гармоникой, у которой амплитуда равна нулю. Приведены графики построенного сигнала на одном выбранном фрейме(Рис: 2 и Рис: 3). Синим обозначен спектр исходного сигнала, красным спектр синтезированного сигнал по заданному методу, желтым разность между ними.

Затем были произведены опыты по улучшению данной модели. Например, рассмотрим соседние гармоники у выбранной гармоники: $i * P1/P2+1$ и $i*P1/P2-1$. Естественно было бы предположить, что если у одной из них разность амплитуд с i-ой во много раз меньше чем у $i*P1/P2$, то возьмем ее, а предыдущую непрерывно соединить с "нулевой" гармоникой. В ходе экспериментов, данное предположение только ухудшило качество сигнала.

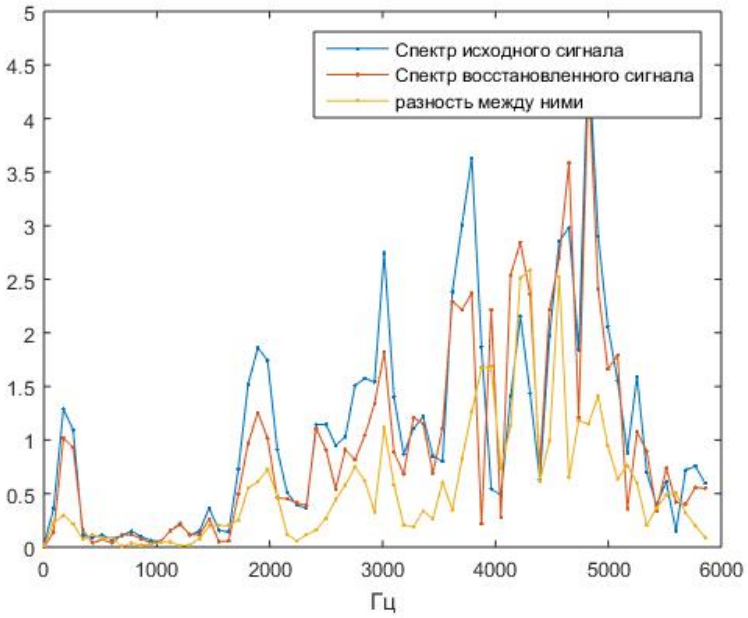

Рис. 2: Объединение по номеру гармоники

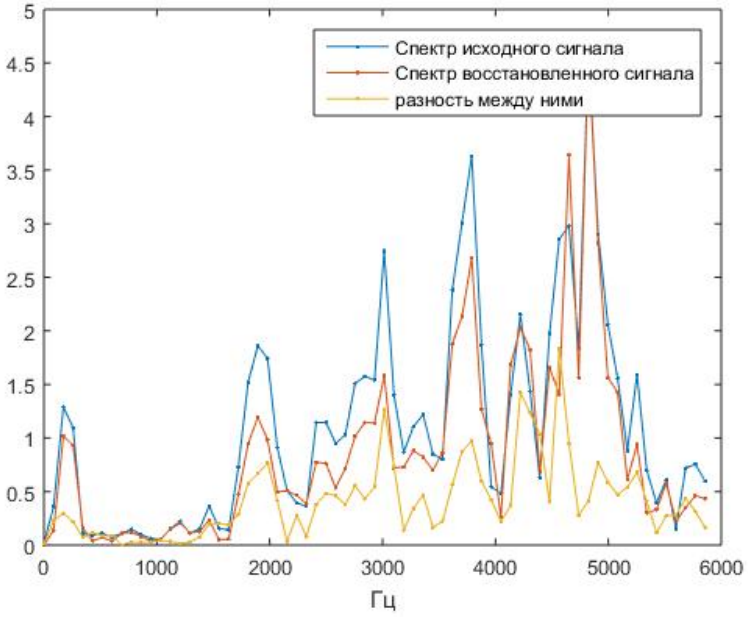

Рис. 3: Равномерное соединение гармоник

3.3. Анализ через синтез

Предыдущие два метода работали с построенной моделью сигнала. Теперь рассмотрим как происходит нахождение параметров модели на каждом фрейме. Строится график полигармонической модели, который наилучшим образом приближает сигнал на промежутке. Поэтому необходимо минимизировать функцию от конечного числа параметров(период основного тона, амплитуда и фаза). Эта функция по методу наименьших квадратов сводится к минимизации по одной переменной - периоду основного тона. В предыдущих методах искался минимум данной функции, и его значение присваивалось периоду. Найдем локальные минимумы, которые по значению отстоят от глобального минимума на некоторое значение.

После выполнение предыдущей операции, в каждом фрейме имеются кандидаты на период. Теперь, из них требуется выбрать один. Для этого воспользуемся динамическим программированием. Возьмем два последовательных фрейма, по периоду, фазе и амплитуда можно однозначно построить сигнал на этих двух фреймах. Пусть $bel(i, j)$ стоимость прихода в j-ый фрейм и в i-ый период в правом фрейме. Получем следующую формулу:

$$bel(i, j) = \min_{\forall l \in K_{j-1}} (bel(l, j-1) + F(l, i))$$

Где K_{j-1} множество кандидатов на период в $j-1$ фрейме, $F(l, i)$ ошибка между исходным сигналом на данном промежутке и построенным сигналом, где периодом в левом фрейме является l-ый член в K_{j-1}, а в правом i-ый член в K_j множестве. Период для первого фрейма определяется однозначно.

Во время работы предыдущей операции происходит запоминание номера периода из левого множества, тем самым осуществляется запоминание оптимального пути, благодаря этому можно восстановить последовательность взятых периодов. В результате найдены наилучшие периоды основного тона для каждого фрейма. Подставим их в исходную модель. Так как периоды взяты оптимальным образом то

качествео сигнала соответственно улучшилось. Далее, для определенно взятого фрейма, приведены графики до использования этого анализа и после (Рис: 4 и Рис: 5). Обозначения такие же: синим обозначен спектр исходного сигнала, красным спектр синтезированного сигнал по заданному методу, желтым разность между ними. На графике 6 изображен исходный и восстановленный сигнал по методу «Анализ через синтез».

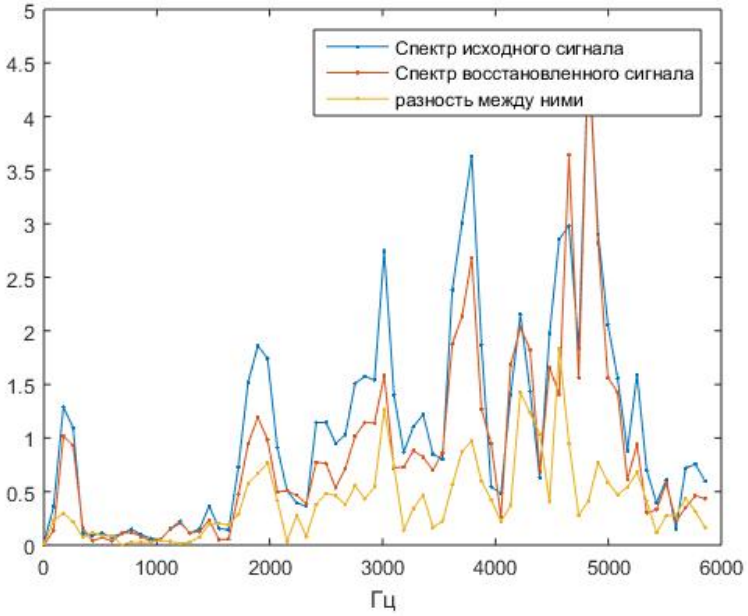

Рис. 4: До оптимизации

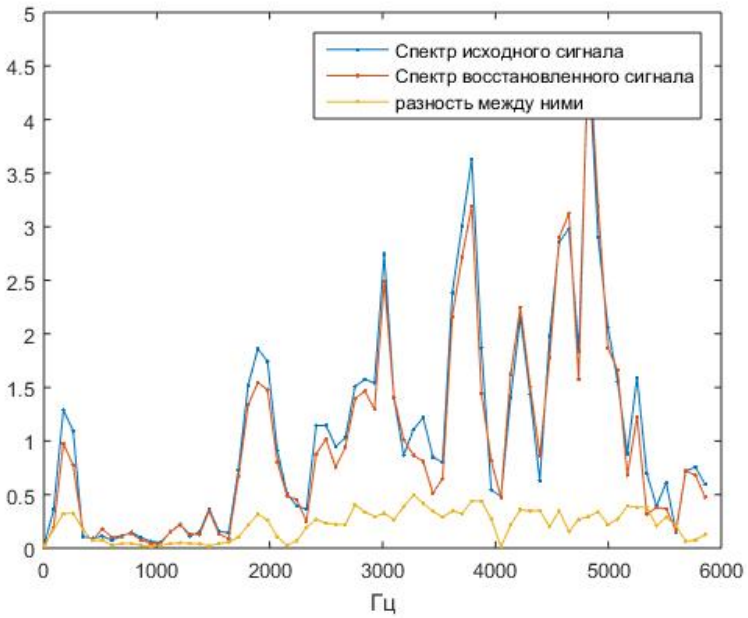

Рис. 5: После оптимизации

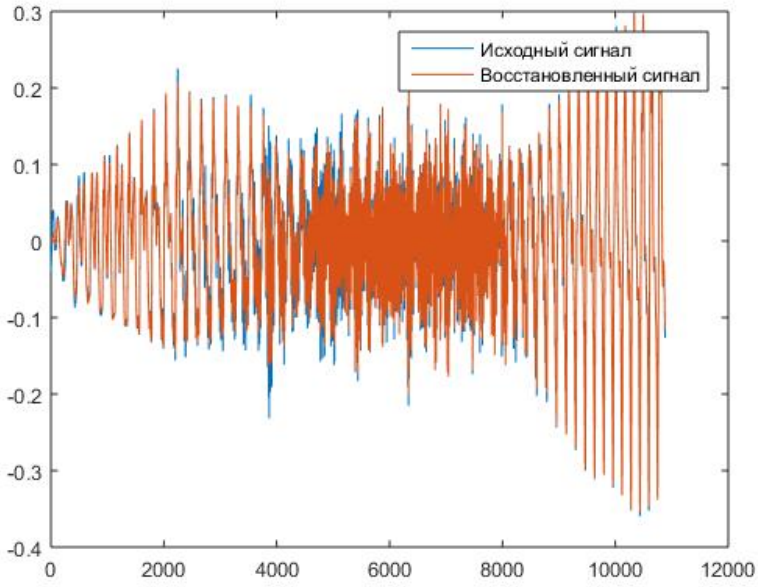

Рис. 6: График исходного и восстановленного сигнала

Заключение

В ходе проделанной работы был разработан метод по улучшению существующей модели. Осуществлена программная реализация придуманного алгоритма. Тем самым, расширена программа оценивания параметров VOCODER. Показано, что новая модель лучше приближает сигнал при возникновении проблемы быстроменяющихся оценках частоты основого тона.

Список литературы

[1] Benesty Jacob, Sondhi M. Mohan, Huang Yiteng (Arden). Springer Handbook of Speech Processing. — Secaucus, NJ, USA : Springer-Verlag New York, Inc., 2007. — ISBN: 3540491252.

[2] Daniel W. Griffin, Jae S. Lim. Multiband Excitation Vocoder. - IEEE Trans. on Acoustic, Speech and Signal Processing, v. 36, no. 8, August 1988, pp. 1223-1235.