

УДК 616-71 + 612.78

МЕТОД СЕГМЕНТАЦИИ ГОЛОСОВОГО СИГНАЛА НА ПЕРИОДЫ ОСНОВНОГО ТОНА ДЛЯ СИСТЕМ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

Е.И. ЖОЛУД, М.И. ВАШКЕВИЧ

*Белорусский государственный университет информатики и радиоэлектроники
(Минск, Республика Беларусь)*

Аннотация. Статья посвящена исследованию методов сегментации голосового сигнала на периоды основного тона в контексте задач медицинской диагностики. В работе рассматривается метод пересечения нуля (zero-crossing) для получения грубой разметки сигнала, а также анализируется влияние длины фильтра, используемого в этом методе, на точность определения частоты основного тона. Результаты эксперимента показывают, что выбор оптимальной длины фильтра является ключевым фактором для повышения точности сегментации.

Ключевые слова: сегментация голосового сигнала, частота основного тона, автокорреляционная функция, фильтрация, zero-crossing, PP-метод, WM-метод, акустическая обработка, медицинская диагностика, адаптивный фильтр, точность определения ЧОТ, джиттер.

METHOD OF SEGMENTING VOICE SIGNALS INTO FUNDAMENTAL FREQUENCY PERIODS FOR MEDICAL DIAGNOSTIC SYSTEMS

E.I. ZHOLUD, M.I. VASHKEVICH

*Belarusian State University of Informatics and Radioelectronics
(Minsk, Republic of Belarus)*

Abstract. This article is dedicated to the study of methods for segmenting voice signals into fundamental frequency periods in the context of medical diagnostics. This paper investigates methods for segmenting speech signals into fundamental frequency periods in the context of medical diagnosis. It examines the zero-crossing method for obtaining a rough signal segmentation and analyzes the influence of the filter length used in this method on the accuracy of fundamental frequency determination. Experimental results demonstrate that choosing the optimal filter length is a key factor in enhancing segmentation accuracy.

Keywords: voice signal segmentation, fundamental frequency, autocorrelation function, filtering, zero-crossing, PP method, WM method, acoustic processing, medical diagnostics, adaptive filter, accuracy of F0 determination, jitter.

Введение

Анализ голоса – это мощный инструмент, используемый в различных областях, включая медицину, психологию, лингвистику и криминалистику [1, 2]. Он позволяет получить ценную информацию о физическом и эмоциональном состоянии человека, а также о его личности.

Одним из самых важных параметров при анализе голоса является частота основного тона (ЧОТ), которая обозначается как F0. Это самая низкая частота звуковых колебаний, генерируемых голосовыми связками во время речи. На рис. 1, а показан фрагмент голосового сигнала, а на рис. 1,б – его спектр. На рис. 1,б также обозначена ЧОТ и её две первые гармоники. ЧОТ определяет высоту голоса и является ключевым индикатором физического состояния. Изменения в ЧОТ могут указывать на патологии голосовых связок, такие как ларингит, полипы или узлы [3], а также на нарушения в работе нервно-мышечной системы [4].

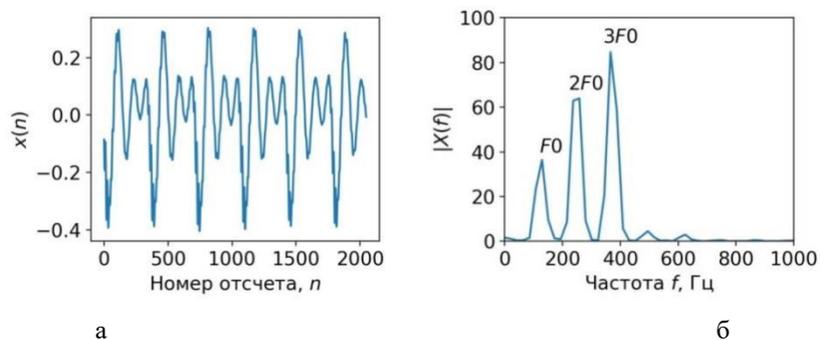


Рис. 1. Сигналы: а) голос; б) частота основного тона

Анализ ЧОТ является ценным инструментом, но для получения более глубокого анализа голоса необходимо учитывать его динамику. ЧОТ не является статичным значением, она постоянно меняется в зависимости от множества факторов, таких как нарушения в работе голосовых связок, вызванные патологиями или утомлением. Для изучения этих динамических изменений требуется сегментация голосового сигнала на периоды основного тона (ОТ). Этот процесс осложняется в тех случаях, когда голос подвержен патологическим изменениям, поскольку в этом случае в сигнале присутствуют шумы и/или нерегулярные всплески.

Методы сегментации голосового сигнала

Есть два основных метода сегментации голосового сигнала на периоды ОТ:

- подгонка формы сигнала (waveform matching) или WM-метод;
- метод отбора локальных максимумов (peak-peaking) или PP-метод.

WM-метод основывается на грубой разметке сигнала. Суть метода заключается в подборе такого смещения между двумя соседними периодами сигнала, чтобы среднеквадратичная разность между ними была минимальной [5]. PP-метод основан на нахождении отрицательных пиков на каждом периоде ОТ сигнала между точками грубой разметки [5]. Таким образом, можно заметить, что оба метода опираются на предварительно выполненную грубую разметку сигнала.

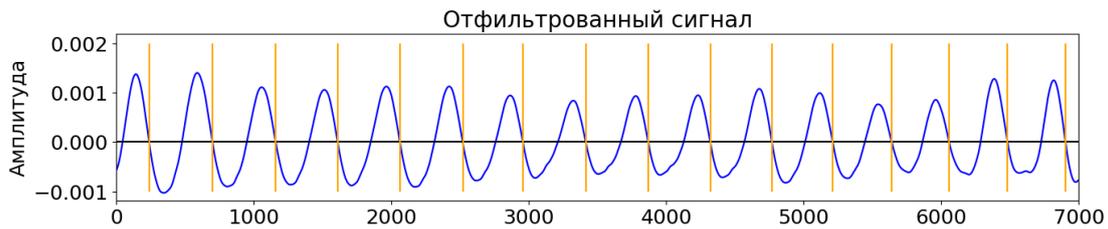
Для получения грубой разметки в [5] предложено использовать метод пересечения нуля (zero-crossing) или ZC-метод. Данный метод заключается в выполнении следующих действий.

1. Отфильтровать сигнал речи с помощью фильтра, чтобы высокочастотные компоненты, не относящиеся к ЧОТ. Рекомендуется использовать КИХ-фильтр с частотой среза равной $1,5 \cdot F_0$, при этом порядок фильтра в методе не оговаривается [5]. На рис. 2 показан голосового сигнала и его отфильтрованная версия.



Рис. 2. Входной сигнал и отфильтрованный сигнал

2. Определить моменты, когда отфильтрованный сигнал меняет знак (с плюса на минус). Обычно используют метод перехода через ноль [5]. На рис. 3 показан пример выделения моментов перехода через ноль, когда сигнал меняет полярность с положительной на отрицательную.

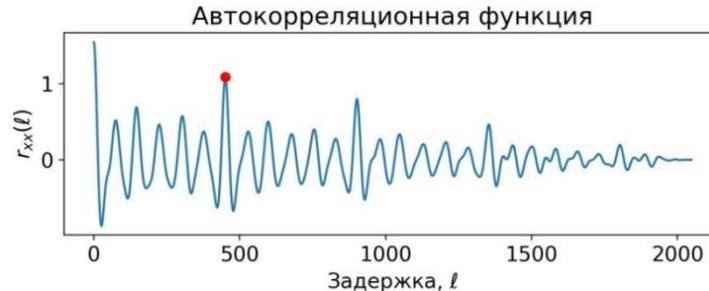


На сегментацию ZC-метода влияет:

- шум в сигнале, который может создавать дополнительные пересечения нулевой линии, что искажает оценку ЧОТ;
- нерегулярные колебания голосовых связок, которые могут привести к неравномерному

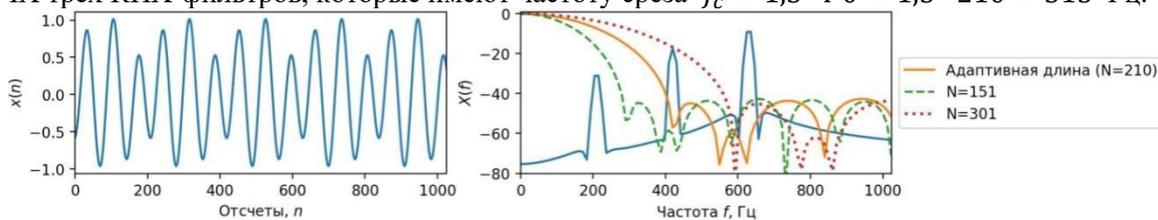
В контексте сегментации голоса, ZC- метод, как правило, используется для получения предварительной разметки сигнала, которую затем уточняют с помощью более сложных методов, таких как WM или PP [4].

Чтобы синтезировать КИХ-фильтр, необходимый в ZC-методе требуется знать приблизительное значение ЧОТ, поскольку частота среза выбирается, как $f_c = 1,5 \cdot F_0$. Для оценки ЧОТ используется автокорреляционная функция (АКФ). Она позволяет выявить периодичность сигнала, что особенно важно для определения ЧОТ [4]. Считается, что первый максимум АКФ вне нулевого задержки соответствует периоду основного тона. Пример АКФ и её первого максимума показаны на рис. 4.



Таким образом, в ZC-методе перед выполнением фильтрации на первых 4096 отсчетах голосового сигнала выполнялся расчет АКФ, по которой определялся период основного тона. На основании данного периода рассчитывалась ЧОТ после чего оконным методом синтезировался КИХ-фильтр с частотой среза $f_c = 1,5 \cdot F_0$.

Как было сказано ранее, в описании ZC-метод порядок фильтра не оговаривается [5], хотя этот фактор имеет существенное значение. На рис. 5 показан полигармонический сигнал, состоящий из трех гармоник с $F_0 = 210$ Гц. Слева на рис. 5 показаны спектр сигнала, а также АЧХ трех КИХ-фильтров, которые имеют частоту среза $f_c = 1,5 \cdot F_0 = 1,5 \cdot 210 = 315$ Гц.



Можно заметить, что случае, когда длина фильтра равна $N = 151$, то, частотная компонента ОТ хотя и выделяется, но при этом будет в выходном сигнале иметь сильное ослабление. Длина фильтра $N = 301$, позволяет выделить компоненту ОТ, но в выходной сигнал также будет проходить и вторая гармоника сигнала с небольшим ослаблением, что в результате помешает правильной работе ZС-метода. В данной работе предлагается подход, согласно которому длина КИХ-фильтра должна выбираться адаптивно. В частотности предлагается выбирать длину фильтра равной периоду ОТ, оцененному по АКФ. В этом случае, как показано на рис. 5 компонента ОТ получает небольшой ослабление, а вторая и третья гармоники попадает на ноль частотной характеристики фильтра, что позволяет эффективно подавить именно гармонические составляющие сигнала. В результате обработки таким фильтров выходной сигнал с меньшей вероятностью будет содержать компоненты, мешающие правильной работе ZС-метода.

Методика проведения эксперимента

Для проведения эксперимента с использованием базы голосов Minsk2020_ALS [6]. База содержит записи голосов людей с боковым амиотрофическим склерозом (БАС), собранные в период 2019-2020 гг. в Минске, Беларусь. База имеют 66 записей голоса здоровых людей и 62 записи пациентов с БАС. Важно отметить, что на каждого человека приходится по две записи (звуки /а/ и /и/). Таким образом, в базе представлены записи от 33 здоровых людей и 31 человека с БАС. Для исследования брались записи здоровых людей.

Целью эксперимента было определить, как длина КИХ-фильтра, используемого в ZС-методе влияет на точность определения ЧОТ голосовых сигналов. Исследовались различные фиксированные длины фильтров (201, 301, 401, 501) и адаптивная длина фильтра, равная периоду основного тона. Таким образом, КИХ-фильтр настраивался автоматически в зависимости от периода основного тона, что позволяло ему подстраиваться под характеристики конкретного сигнала. Точность определения ЧОТ оценивалась с использованием метрики джиттера – меры вариативности ЧОТ. Джиттер более 5% считался признаком ошибки в определении ЧОТ.

Результаты и их обсуждение

На рис. 6 представлены результаты в эксперимента, которые позволяет провести сравнительный анализ между длиной фильтра и количеством ошибок, возникающих в процессе обработки голосовых сигналов.

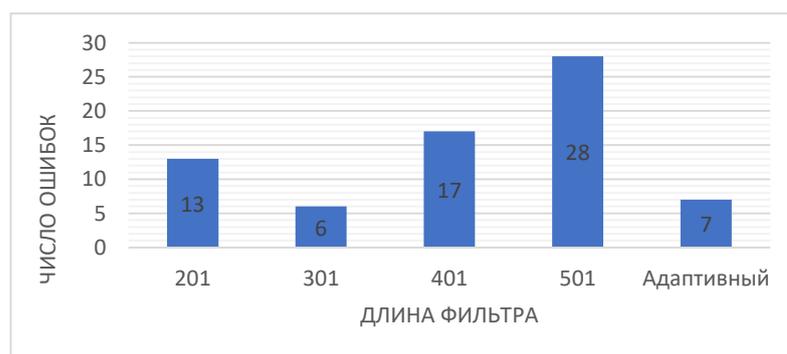


Рис.6. График количества ошибок в зависимости от длины фильтра

Согласно полученным данным, минимальное количество ошибок наблюдается при длине фильтра, равной 301, а также в условиях использования адаптивного фильтра. В этих случаях было зафиксировано всего 6 и 7 ошибок соответственно.

На основании полученных результатов эксперимента можно также сделать вывод о том, что характеристики адаптивного фильтра в значительной степени соответствуют параметрам фильтра с длиной 301 отсчет. Это свидетельствует о том, что исследуемая база данных голосовых сигналов преимущественно включает записи голосов людей, чьи частотные

характеристики наиболее эффективно обрабатываются с использованием фильтра длиной 301 отсчет.

Таким образом, результаты эксперимента не только подтверждают эффективность фильтра с адаптивной длиной для выделения периодов основного тона.

Заключение

В ходе проведенного эксперимента была исследована влияние длины КИХ-фильтра, используемого в ZС-методе, на точность сегментации голосовых сигналов, записанных у здоровых людей.

В работе было предложено выбирать длину фильтра приблизительно равной периоду основного тона. При этом основной тон предлагается оценивать на начальном сегменте голосового сигнала с использованием автокорреляционной функции.

Результаты эксперимента показали, что длина фильтра имеет значительное влияние на точность сегментации голосового сигнала. Наименьшее количество ошибок в определении ЧОТ было зафиксировано при использовании фильтра длиной 301 отсчет, а также при применении адаптивного фильтра.

Список литературы

1. Barsties B., De Bodt M. Assessment of voice quality: current state-of-the-art // *Auris Nasus Larynx*. – 2015. – Т. 42. – №. 3. – P. 183-188.
2. Hecker P. et al. Voice analysis for neurological disorder recognition—a systematic review and perspective on emerging trends // *Frontiers in Digital Health*. – 2022. – Т. 4. – P. 1-16.
3. Анализ акустических параметров голоса для выявления заболеваний гортани / Вашкевич М. И., Бурак А.А., Конойко Н.С. Долдова В.С. // *Информатика*. – 2020. –Т.17, №1. – С. 78 – 86.
4. Kul T. P., Rushkevich Y. N., Lihachev S. A. Adaptation of digital signal processing methods to the analysis of speech in neurological pathologies // *Doklady BGUIR*. – 2018. – №. 7. – P. 128-132.
5. Comparison of Extraction Methods for High-Precision Voice Perturbation Measurements / Ingo R. Titze, Haixiang Liang // *Journal of Speech and Hearing Research*, 1993. – P. 1-14.
6. https://github.com/Mak-Sim/Minsk2020_ALS_database