

Средства анализа и визуализации речевых сигналов при подготовке студентов инженерных специальностей

Ю.Г. Горшков¹

Московский государственный технический университет имени Н.Э. Баумана,
Москва, Россия

¹ ORCID: 0000-0003-0483-4603, y.gorshkov@np0-echeleon.ru

Аннотация

В работе рассмотрен зарубежный опыт обучения студентов инженерных специальностей анализу речевых сигналов инструментальными методами. Приводятся примеры получения спектрограмм, а также возможности программных средств анализа речи используемые на инженерных курсах бакалавриата и магистратуры Калифорнийского университета (Лос-Анджелес, США). Показаны недостатки анализа и визуализации речи на основе Фурье-преобразования. Предложены новые решения обработки и визуализации речевых сигналов на основе многоуровневого вейвлет-анализа. Рассмотрены основные характеристики разработанных программ WaveView и WaveView-MWA, обеспечивающих повышенное частотно-временное разрешение гласных звуков. Впервые представлены результаты высокоточного анализа и визуализации согласных звуков - нестационарных сигналов, недоступных спектральному анализу с использованием преобразования Фурье. Выполнен сравнительный анализ частотно-временного разрешения спектрограмм и вейвлет-сонограмм при визуализации английской речи. Разработанная технология высокоточного анализа и визуализации речевых сигналов используется при подготовке специалистов кафедры «Информационная безопасность» факультета «Информатика и Управление» МГТУ имени Н.Э. Баумана в ходе выполнения лабораторных работ по курсу «Криминалистическое исследование фонограмм».

Ключевые слова: визуализация речевых сигналов, спектральный анализ, многоуровневый вейвлет-анализ, сонограмма.

1. Введение

Анализу речевого сигнала при подготовке студентов инженерного профиля ведущими отечественными и зарубежными университетами придается важное значение. Так, например, в работе [1] представлены результаты внедрения методов обучения «чтению» спектрограмм на инженерных курсах бакалавриата и магистратуры в Калифорнийском университете, Лос-Анджелес (по версии журнала Forbes в 2019 году вошел в число лучших колледжей мира). Анализ спектрограмм - важный навык для изучения акустики речи, необходим для визуализации причинно-следственных связей между движениями речевого артикулятора и производимым в результате звуком. Чтение спектрограмм часто является сложной задачей для студентов, не имеющих предварительного образования в области акустической фонетики.

Математический аппарат построения спектрограмм был разработан в 1960-е годы. Определение кратковременного преобразования Фурье (логарифмическая величина которого отображается на спектрограмме) сформулировано в 1967, а с 1970 года стало возможным его вычисление с помощью компьютера [2]. Некоторые из оригинальных спектрографических параметров и терминов сохранились до наших дней, включая ис-

пользование «широкополосных» и «узкополосных» для описания спектрограмм, вычисленных, соответственно, с использованием «коротких» и «длинных» окон анализа [3].

Для получения спектрограмм речевого сигнала или изображений «видимая речь» в Калифорнийском, а также ряде зарубежных университетов используются такие программы, как Praat [4] и Audacity Team [5].

Цель данного исследования - выполнить оценку частотно-временного разрешения спектрограмм речевого сигнала, полученных на основе преобразования Фурье, а также рассмотреть возможности нового решения обработки и визуализации гласных и согласных звуков речи на основе многоуровневого вейвлет-анализа.

2. Программные средства получения спектрограмм речевого сигнала

2.1. Программа Praat

Компьютерная программа Praat (doing phonetics by computer), разработана институтом фонетических исследований Амстердамского университета, Нидерланды, 2009 год. Назначение - Фурье-анализ и визуализация формантных характеристик речевого сигнала [6, 7].

Под формантой понимают полосы передаточной функции речевого тракта, характеризующиеся частотой F_i , амплитудой A_i , и полосой пропускания B_i [8, 9]. На амплитудно-частотном спектре форманты проявляются в виде заметных максимумов на гласных звуках (рис. 1).

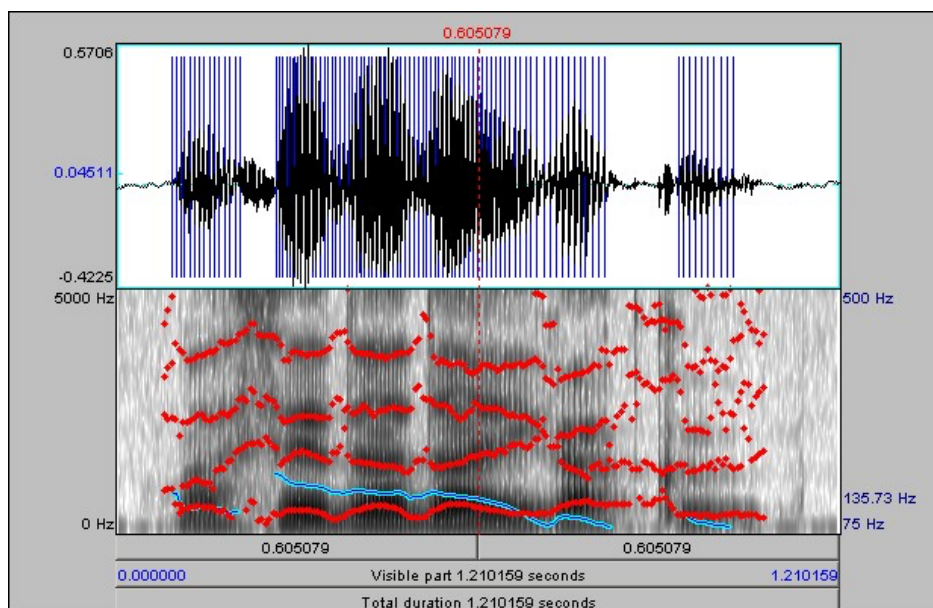


Рис. 1. Praat. Окно «Sound». Красным цветом обозначены формантные траектории гласных звуков

2.2. Программа Audacity

Audacity - аудиоредактор с открытым исходным кодом и прикладное программное обеспечение для записи, доступное для Windows, macOS, Linux и других Unix-подобных операционных систем; разработан университетом Карнеги-Меллона, США, 2000 год. Назначение - инструментальный Фурье-анализ и визуализация звуковых сигналов. На рис. 2 представлен внешний вид интерфейса программы Audacity, режим Multi-view.

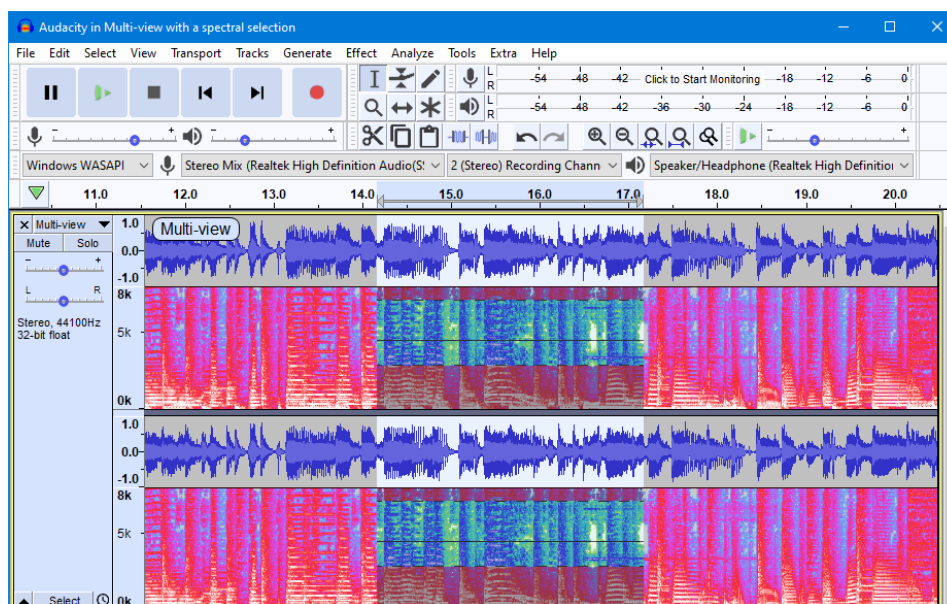


Рис. 2. Внешний вид интерфейса программы Audacity, режим Multi-view

2.3. Примеры спектрограмм речевого сигнала

Рассмотрим спектрограммы двух фраз: «We saw the goal to win boons» и «We sue the bowl to bin beans», полученных с помощью программы Praat и рассмотренных ученым Bruce Hayes, UCLA [6]. (Bruce Hayes – заслуженный профессор лингвистики, Калифорнийский университет, Лос-Анджелес).

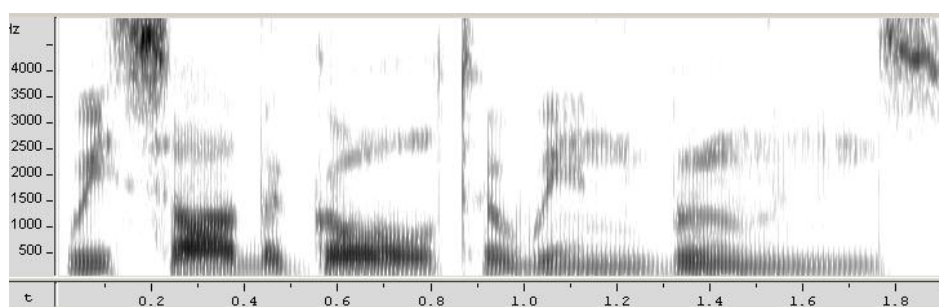


Рис. 3. Спектрограмма фразы «We saw the goal to win boons»

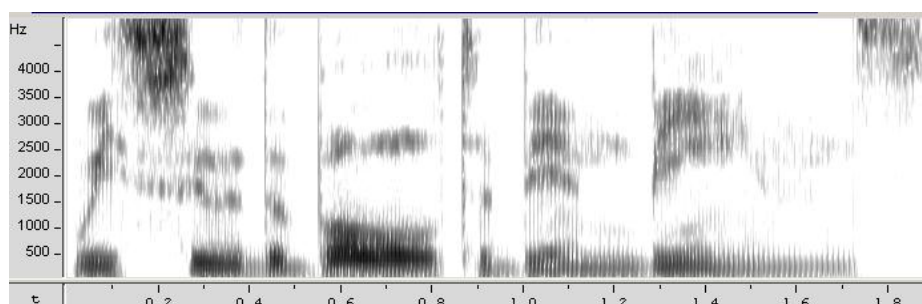


Рис. 4. Спектрограмма фразы «We sue the bowl to bin beans»

В работе [6] также представлены комментарии – результаты анализа спектрограмм рис. 3, 4:

saw/sue, vowels at about 800 msec. in both spectrograms: note high F1 for the low vowel (IPA open o) in «saw» (about 600 Hz.) vs. low F1 for [u] in «sue» (about 300 Hz). High vowels have low F1, low vowels have high F2.

goal/bowl: initial stop at about 1100 msec. in both spectrograms. First spectrogram: [g] has a velar pinch in F2/F3. Second spectrogram: formants rise out of the stop closure for [b].

win/bin: First spectrogram: [w] at about 1500 msec. has a gentle decrease/increase in amplitude; it is a sonorant consonant. Second spectrogram: [b], a stop, has a sudden increase in amplitude, with a small burst, at the moment of release.

boons/beans: First spectrogram: [u] at about 1800 msec. has F2 around 1300. Second spectrogram: [i] at around 1800 msec. has F2 around 2200 Hz.; back vowels, and round vowels, have lower F2 than front/unrounded vowels.

Из полученных данных следует, что рассмотренная методика оценки формант по спектрограммам, полученным на основе преобразования Фурье, позволяет выполнить первичный анализ речевого сигнала, причем только гласных звуков (в английском языке 6 гласных, 21 согласных). В то же время, для решения таких задач, как автоматическое распознавание речи, аутентификация диктора, определения диалекта языка и акцента носителя иностранного языка необходимы новые высокоточные решения анализа сигналов [10, 11].

3. Анализ и визуализация речевого сигнала на основе вейвлет-преобразования

3.1. Программа вейвлет-анализа WaveView

Речевые сигналы относятся к сложным нестационарным сигналам, в связи с этим применение вейвлет-преобразования при решении задач анализа речи позволит получить более точное его представление в частотно-временной области [12]. Первая версия программы WaveView разработана в 2003 году [13].

Программа вейвлет-анализа WaveView предоставляет следующие возможности [14]: отображение осциллограммы сигнала; анализ участка сигнала с возможностью выбора полосы частот и частотно-временного разрешения; визуализация - отображение результатов анализа в виде вейвлет-сонограммы (используется вейвлет Морле); получение частотного сечения в заданный момент времени; поддержка большого числа форматов звуковых файлов. Требования к ПК: операционная система Windows.

Построение вейвлет-сонограмм можно осуществлять также в дистанционном режиме с использованием портала «Акустокардиограф» (<http://acustocard.ru>) [15]. На рис. 5. представлен пример визуализации гласных звуков А, Э, И, О, У, Ы, а также точек монтажа фонограммы, получаемые студентами 6 курса кафедры «Информационная безопасность» МГТУ им. Н.Э. Баумана при выполнении лабораторной работы «Определение подлинности фонограмм с использованием технологий БПФ и вейвлет-анализа» при освоении курса «Криминалистическое исследование фонограмм» [16].

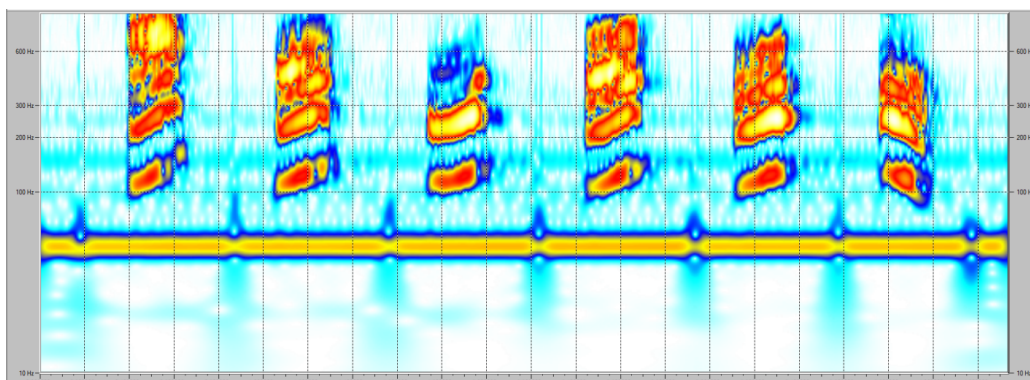


Рис. 5. Вейвлет-сонограмма гласных звуков А, Э, И, О, У, Ы. В нижней части изображения видны точки монтажа – разрывы фазы сигнала фона сети питания 50 Гц

3.2. Программное обеспечение WaveView-MWA

В программном обеспечении (ПО) WaveView-MWA [17] реализовано несколько алгоритмов построения вейвлет-сонограмм, выполнения вейвлет-фильтрации, вычисления фазы речевого сигнала. Вейвлет-сонограмма это диаграмма, на которой по оси абсцисс откладывается время, по оси ординат – частота, а амплитуда соответствующей частотной составляющей отмечается интенсивностью цвета в данной точке графика. При построении сонограммы для каждого момента времени вычисляются значения спектра сигнала по заданным параметрам вейвлет-преобразования. Полученные данные амплитуд - значения одного столбца графика. Вейвлет-сонограмма обеспечивает высокое частотно-временное разрешение исследуемого сигнала.

3.2.1. Описание главного меню

На рис. 6 представлен вид главного меню настроек построения вейвлет-сонограммы ПО WaveView-MWA.

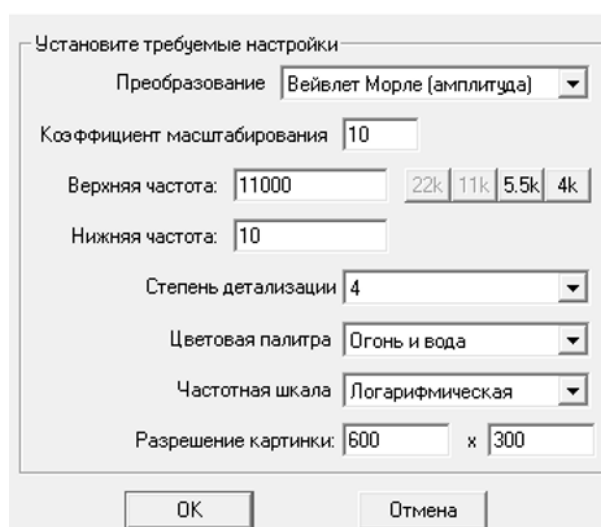


Рис. 6. Вид главного меню ПО WaveView-MWA

Предусмотрены следующие режимы обработки и задания параметров:

1. «Преобразование» – выбор материнского вейвлета и отображаемой на экране компоненты вейвлет-преобразования (амплитуда, фаза).

2. «Коэффициент масштабирования» – параметр, указывающий степень локализации вейвлет-сонограммы по частоте и времени. Малые значения дают высокую локализацию по времени, но низкую по частоте. Большие значения – наоборот.

3. «Верхняя частота»/«нижняя частота» - диапазон частот, в котором будет построена вейвлет-сонограмма. Для установки значения верхней частоты, обозначения: «22k», «11k», «5.5k» и «4k» соответствуют: 22050 Гц, 11025 Гц, 5500 Гц и 4000 Гц.

4. «Степень детализации» – показатель уровня частотно-временного разрешения вейвлет-сонограммы.

5. «Цветовая палитра» – выбор цветового представления вейвлет-сонограммы.

6. «Частотная шкала» – может быть представлена в логарифмическом либо линейном масштабе. «Логарифмическая» дает детальное представление низкочастотной области сигнала; «Линейная» - высокочастотной.

7. «Разрешение картинки» – задается размер изображения сонограммы в пикселях.

На рис. 7 представлена структура изображения вейвлет-сонограммы речевого сигнала, слово «крюк».

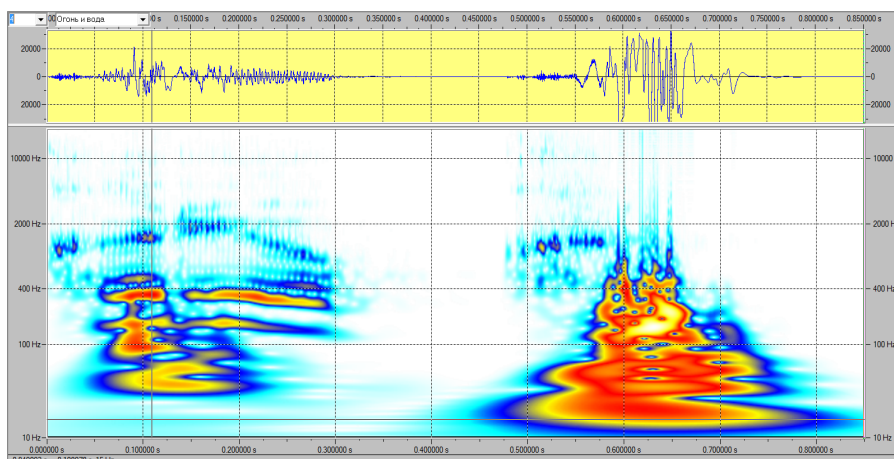


Рис. 7. Структура изображения вейвлет-сонограммы речевого сигнала, слово «кряк»

Проведенное тестирование ПО WaveView-MWA показало возможность выделения и визуализации нестационарных сигналов малого уровня (до -60 дБ). Использование режима «Звуковой микроскоп» позволяет получать характеристики гласных звуков речи с повышенным частотно-временным разрешением. Достигнуто получение параметров согласных звуков, недоступных программным средствам анализа с использованием преобразования Фурье.

3.3. Примеры вейвлет-сонограмм речевого сигнала

На рис. 8 представлена вейвлет-сонограмма фразы: «We saw the goal to win boons», полученная с помощью ПО WaveView-MWA. (Спектрограмма для сравнения – рис. 3).

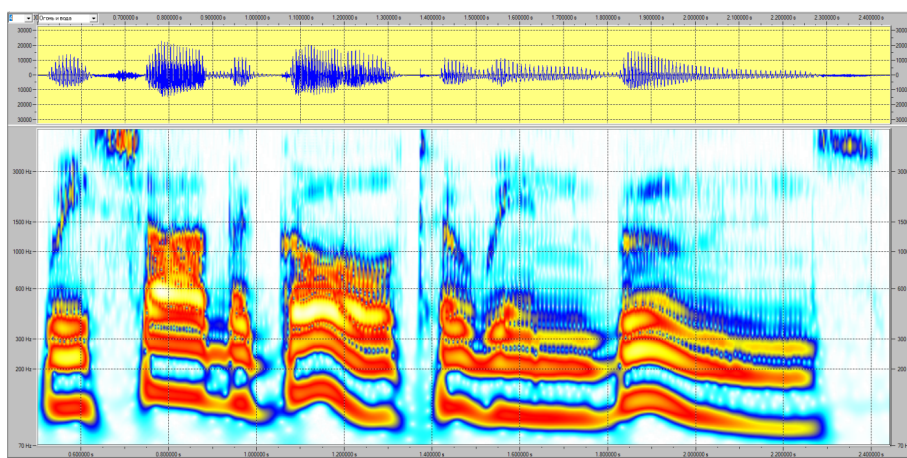


Рис. 8. Вейвлет-сонограмма фразы «We saw the goal to win boons»

В отличие от спектрограммы рис. 3 видим повышенное частотно-временное разрешение тональных участков речевого сигнала - гласных звуков. Основные характеристики - цифровые значения формант, периода основного тона и его динамика отображаются с использованием курсора в реальном масштабе времени.

Рассмотрим возможность выделения и визуализации согласных звуков речи. На рис. 9 – вейвлет-сонограмма сочетания согласный-гласный «go» (goal).

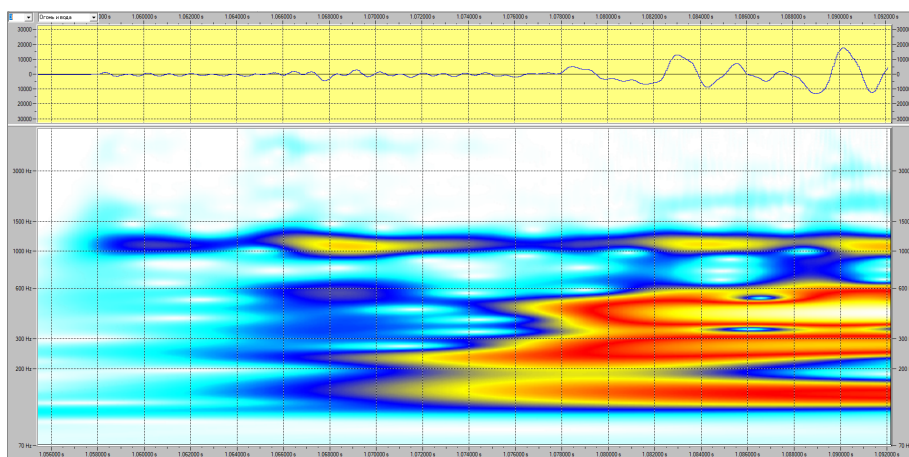


Рис. 9. Вейвлет-сонограмма «go» (goal), 1.058-1.092 сек. Полоса частот «g»: 1000-1300 Гц; «o»: 100-600 Гц

На рис. 10 представлена вейвлет-сонограмма сочетания согласный-гласный «to» (to win).

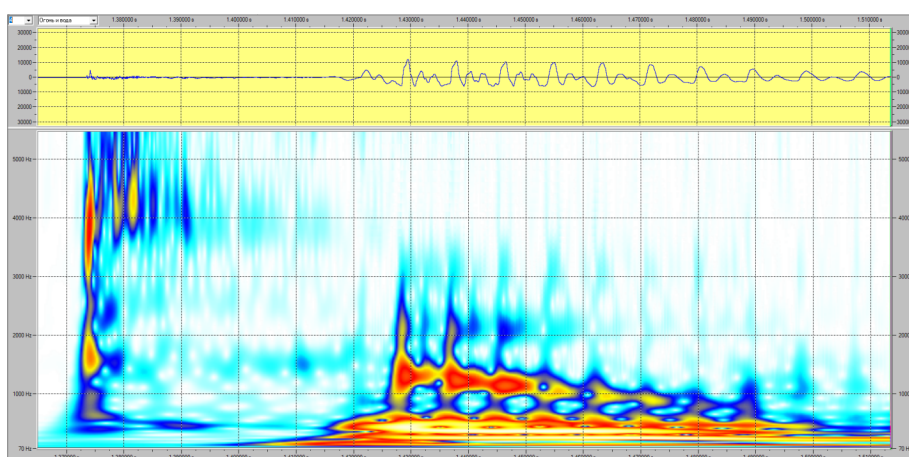


Рис. 10. Вейвлет-сонограмма «to» (to win), 1.37-1.51 сек. Полоса частот «t»: 295-5405 Гц; «o»: 88-3070 Гц. Длительность «t»: 0.025 сек; «o»: 0.115 сек

На рис. 11 представлена вейвлет-сонограмма второй из анализируемых фраз: «We sue the bowl to bin beans», также полученная с помощью ПО WaveView-MWA. (Спектрограмма для сравнения – рис. 4).

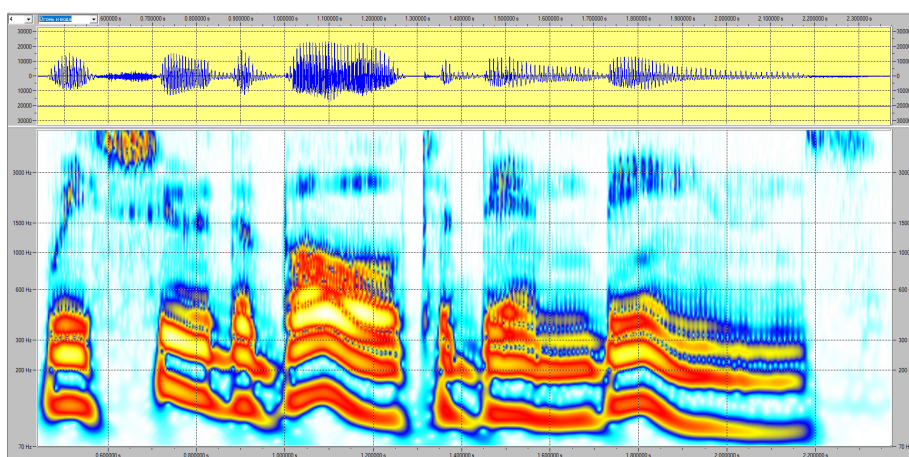


Рис. 11. Вейвлет-сонограмма фразы «We sue the bowl to bin beans»

По сравнению со спектрограммой рис. 4 также заметно повышенное частотно-временное разрешение гласных звуков.

На рис. 12 представлена вейвлет-сонограмма сочетания согласный-гласный «to» (**to bin**).

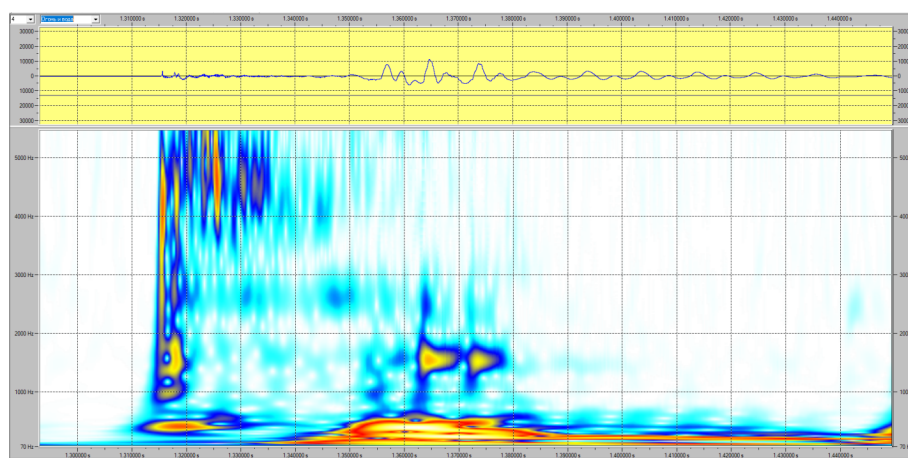


Рис. 12. Вейвлет-сонограмма «to» (**to bin**), 1.31-1.335 сек. Полоса частот «t»: 327-5613 Гц; «o»: 88-2885 Гц. Длительность «t»: 0.025 сек; «o»: 0.06 сек

Из анализа спектрограмм (рис. 3, 4) и сонограмм (рис. 8-12) следует, что вейвлет-сонограммы обладают повышенным частотно-временным разрешением по сравнению с Фурье-спектрограммами. Визуализация данных многоуровневого вейвлет-анализа нестационарных участков речевого сигнала малого уровня позволяет получать частотно-временные характеристики согласных звуков.

4. Заключение

Разработанные программные средства многоуровневого вейвлет-анализа и визуализации WaveView и WaveView-MWA позволяют получать частотно-временное описание речевого сигнала с разрешением, превосходящим возможности программ Фурье-анализа, таких как Praat и Audacity Team. Вейвлет-сонограммы представляют наглядную объективную и полную информацию параметров гласных звуков. Впервые показана принципиальная возможность получения характеристик согласных звуков речи.

Предложенная технология визуализации речевого сигнала подтвердила ее высокую эффективность при выполнении студентами 6 курса кафедры «Информационная безопасность» факультета «Информатика и Управление» МГТУ имени Н.Э. Баумана лабораторных работ по курсу «Криминалистическое исследование фонограмм» [18].

Программные средства WaveView-MWA нашли также применение при создании систем телемедицины нового поколения [19, 20], высокоточной визуализации звуков сердца [21], легочных звуков [22],

эмоциональной напряженности человека по речевому сигналу [23].

Список литературы

1. Alexander Johnson. An integrated approach for teaching speech spectrogram analysis to engineering students. The Journal of the Acoustical Society of America. 152. 2022. pp. 1962-1969. (<https://doi.org/10.1121/10.0014172>)
2. Flanagan, J.L. Speech Analysis, Synthesis and Perception. Springer, New York, 1972. 446 p.
3. Sean A. Fulop. The beginning of time-frequency analysis. The Journal of the Acoustical Society of America. 152. 2022. R9-R10. (<https://doi.org/10.1121/10.0014987>)
4. P. Boersma and D. Weenink. Praat: Doing phonetics by computer (version 6.2.14) [computer program], 2022. (<http://www.praat.org/>)

5. Audacity Team Audacity(r): Free audio editor and recorder (version 3.0.0) [computer application], 2021. (<https://audacityteam.org/>)
6. B. Hayes. Spectrogram reading practice, 2021. (<https://linguistics.ucla.edu/people/hayes/103/SpectrogramReading/index.htm>)
7. Столбов М.Б. Основы анализа и обработки речевых сигналов. СПб.: НИУ ИТМО, 2021. 101 с.
8. Лобанов Б.М. Речевой интерфейс интеллектуальных систем: учебное пособие / Б.М. Лобанов, О.Е. Елисеева; под науч. ред. проф. В.В. Голенкова. Минск: БГУИР, 2006. 152 с.
9. Елисеева О.Е. Речевой интерфейс. Лабораторный практикум: учебно-методическое пособие для студентов учреждений, обеспечивающих получение высшего образования по специальности «Искусственный интеллект»: в 2 ч. Ч. 1 / под ред. проф. В.В. Голенкова. Минск: БГУИР, 2008. 44 с.
10. Тампель И.Б., Хитров М.В. Автоматическое распознавание речи. Учебное пособие по дисциплине «Распознавание речи». СПб: СПбНИУ ИТМО. 2014. 119 с.
11. Леонов А.С., Сорокин В.Н. Формантный анализ речевого сигнала в фазовой области. Информационные процессы. 2021. Т. 21. № 2. С. 125–134. (<http://www.jip.ru>)
12. Горшков Ю.Г. Обработка речевых и акустических биомедицинских сигналов на основе вейвлетов. Научное издание. М.: Радиотехника, 2017. 240 с.
13. Горшков Ю.Г., Кузин А.Ю. Применение Wavelet-преобразования при решении задач анализа речевого сигнала. Проблемы информационной безопасности в системе высшей школы. X Всероссийская научная конференция. Москва, 2003. С. 51.
14. Горшков Ю.Г., Каиндин А.М., Марков А.С., Цирлов В.Л. WaveView. Вейвлет-анализ речевых и акустических биомедицинских сигналов. Свидетельство о регистрации программы для ЭВМ RU 2017662425, 07.11.2017. Заявка № 2017619325 от 14.09.2017.
15. Портал «Акустокардиограф» (<http://acustocard.ru>)
16. Горшков Ю.Г. Криминалистическое исследование фонограмм: Методические указания к выполнению лабораторных работ. Учебно-методическое пособие. М.: Изд-во МГТУ им. Н.Э. Баумана, 2017. 32 с.
17. Горшков Ю.Г., Каиндин А.М., Марков А.С., Цирлов В.Л. WaveView-MWA. Многоуровневый вейвлет-анализ речевых и акустических биомедицинских сигналов. Свидетельство о регистрации программы для ЭВМ RU 2017662094, 27.10.2017. Заявка № 2017619124 от 08.09.2017.
18. Горшков Ю.Г. Визуализация многоуровневого вейвлет-анализа фонограмм. Научная визуализация. 2015. Т. 7. № 2. С. 96-111.
19. Горшков Ю.Г. Новые решения визуализации биомедицинских сигналов в системах телемедицины. Научная визуализация. 2019. Т. 11. № 2. С. 56-72. DOI: 10.26583/sv.11.2.05.
20. Горшков Ю.Г. Визуализация помех сети питания в телемедицинских системах мобильной электрокардиографии. Научная визуализация. 2021. Т. 13. № 1. С. 44-53. DOI: 10.26583/sv.13.1.04.
21. Горшков Ю.Г. Визуализация звуков сердца. Научная визуализация. 2017. Т. 9. № 1. С. 97-111.
22. Gorshkov Yu.G. Visualization of Lung Sounds Based on Multilevel Wavelet Analysis. Scientific Visualization. 2022. Vol. 14. No. 2. pp. 18-26. DOI: 10.26583/sv.14.2.02.
23. Горшков Ю.Г. Визуализация эмоциональной напряженности человека по речевому сигналу. Научная визуализация. 2023. Т. 15. № 2. С. 102-112. DOI: 10.26583/sv.15.2.09.

Means of Analysis and Visualization of Speech Signals in the Preparation of Engineering Students

Yu.G. Gorshkov¹

Bauman Moscow State Technical University, Russian Federation

¹ ORCID: 0000-0003-0483-4603, y.gorshkov@npo-echelon.ru

Abstract

The paper considers the foreign experience of teaching engineering students to analyze speech signals using instrumental methods. Examples of obtaining spectrograms are given, as well as the capabilities of speech analysis software used in undergraduate and graduate engineering courses at the University of California (Los Angeles, USA). The disadvantages of speech analysis and visualization based on Fourier transform are shown. New solutions for processing and visualizing speech signals based on multilevel wavelet analysis are proposed. The main characteristics of the developed WaveView and WaveView-MWA programs that provide increased time-frequency resolution of vowel sounds are considered. For the first time, the results of high-precision analysis and visualization of consonant sounds - non-stationary signals inaccessible to spectral analysis using the Fourier transform are presented. A comparative analysis of the time-frequency resolution of spectrograms and wavelet sonograms in the visualization of English speech is performed. The developed technology of high-precision analysis and visualization of speech signals is used in the training of specialists of the Department of «Information Security» of the Faculty of «Informatics and Management» of the Bauman Moscow State Technical University during laboratory work on the course «Forensic study of phonograms».

Keywords: visualization of speech signals, spectral analysis, multilevel wavelet analysis, sonogram.

References

1. Alexander Johnson. An integrated approach for teaching speech spectrogram analysis to engineering students. The Journal of the Acoustical Society of America. 152. 2022. pp. 1962-1969. (<https://doi.org/10.1121/10.0014172>)
2. Flanagan, J.L. Speech Analysis, Synthesis and Perception. Springer, New York, 1972. 446 p.
3. Sean A. Fulop. The beginning of time-frequency analysis. The Journal of the Acoustical Society of America. 152. 2022. R9-R10. (<https://doi.org/10.1121/10.0014987>)
4. P. Boersma and D. Weenink. Praat: Doing phonetics by computer (version 6.2.14) [computer program], 2022. (<http://www.praat.org/>)
5. Audacity Team Audacity(r): Free audio editor and recorder (version 3.0.0) [computer application], 2021. (<https://audacityteam.org/>)
6. B. Hayes. Spectrogram reading practice, 2021. (<https://linguistics.ucla.edu/people/hayes/103/SpectrogramReading/index.htm>)
7. Stolbov M.B. Fundamentals of speech signal analysis and processing. St. Petersburg: ITMO Research Institute. 2021. 101 p.
8. Lobanov B.M. Speech interface of intelligent systems: a textbook / B.M. Lobanov, O.E. Eliseeva; under the scientific editorship of prof. V.V. Golenkov. Minsk: BGUIR, 2006. 152 p.

9. Eliseeva O.E. Speech interface. Laboratory workshop: an educational and methodological guide for students of institutions providing higher education in the specialty «Artificial Intelligence»: in 2 hours, Part 1 / edited by prof. V.V. Golenkov. Minsk: BGUIR, 2008. 44 p.
10. Tampil I.B., Khitrov M.V. Automatic speech recognition. Textbook on the discipline «Speech recognition». St. Petersburg: SPbNIU ITMO. 2014. 119 p.
11. Leonov A.S., Sorokin V.N. Formant analysis of a speech signal in the phase domain. Information processes. 2021. Vol. 21. No. 2. pp. 125-134. (<http://www.jip.ru>)
12. Gorshkov Yu.G. Processing speech and acoustic biomedical signals based on wavelets. Scientific edition. Moscow: Radio Engineering, 2017. 240 p.
13. Gorshkov Yu.G., Kuzin A.Y. Application of the Wavelet transform in solving speech signal analysis problems. Problems of information security in the higher school system. X All-Russian Scientific Conference. Moscow, 2003. p. 51.
14. Gorshkov Yu.G., Kaindin A.M., Markov A.S., Cirlov V.L. WaveView. Wavelet analysis of speech and acoustic biomedical signals. Certificate of registration of the computer program RU 2017662425, 07.11.2017. Application No. 2017619325 dated 09/14/2017.
15. The portal «Acoustocardiograph» (<http://acustocard.ru>)
16. Gorshkov Yu.G. Forensic examination of phonograms: Guidelines for laboratory work. Educational and methodical manual. Moscow: Publishing House of Bauman Moscow State Technical University, 2017. 32 p.
17. Gorshkov Yu.G., Kaindin A.M., Markov A.S., Cirlov V.L. WaveView-MWA. Multilevel wavelet analysis of speech and acoustic biomedical signals. Certificate of registration of the computer program RU 2017662094, 27.10.2017. Application No. 2017619124 dated 08.09.2017.
18. Gorshkov Yu.G. Visualization of multilevel wavelet analysis of phonograms. Scientific visualization. 2015. Vol. 7. No. 2. pp. 96-111.
19. Gorshkov Yu.G. New visualization solutions for biomedical signals in telemedicine systems. Scientific visualization. 2019. Vol. 11. No. 2. pp. 56-72. DOI: 10.26583/sv.11.2.05.
20. Gorshkov Yu.G. Visualization of power supply network interference in telemedicine mobile electrocardiography systems. Scientific visualization. 2021. Vol. 13. No. 1. pp. 44-53. DOI: 10.26583/sv.13.1.04.
21. Gorshkov Yu.G. Visualization of heart sounds. Scientific visualization. 2017. Vol. 9. No. 1. pp. 97-111.
22. Gorshkov Yu.G. Visualization of Lung Sounds Based on Multilevel Wavelet Analysis. Scientific Visualization. 2022. Vol. 14. No. 2. pp. 18-26. DOI: 10.26583/sv.14.2.02.
23. Gorshkov Yu.G. Visualization of human emotional tension by speech signal. Scientific visualization. 2023. Vol. 15. No. 2. pp. 102-112. DOI: 10.26583/sv.15.2.09.