

ПРИМЕНЕНИЕ ВЕЙВЛЕТА В MATHCAD ДЛЯ РАСПОЗНАВАНИЯ РЕЧИ

APPLICATION OF A WAVELET IN MATHCAD FOR SPEECH RECOGNITION

**N. Leontiev
R. Kopyrin**

Summary. This article discusses the use of wavelet functions in the MathCAD package. Used window analysis of speech in yakut language to obtain a correlation pattern. The dependence of the correlation on the window width and the number of samples of the wavelet function are investigated.

Keywords: wavelet, speech recognition, yakut language.

Леонтьев Ньургун Анатольевич

К.т.н., доцент, Северо-Восточный федеральный университет им. М. К. Аммосова, Якутск
leonza@mail.ru

Копырин Рустам Русланович

Северо-Восточный федеральный университет им. М. К. Аммосова, Якутск
rustamkopyrin@mail.ru

Аннотация. В данной статье рассматривается вопрос применение вейвлет-функций в пакете MathCAD. Применяется оконный анализ устной речи на якутском языке для получения корреляционной картины. Исследуются зависимость корреляции от ширины окна и количество отсчетов вейвлет-функции.

Ключевые слова: вейвлет, распознавание речи, якутский язык.

Современные методы исследования позволяют выявить различные характеристики и параметры речевого сигнала. Накопление большой базы данных позволяет проводить преобразование речи в текст. Но такой подход требует накопление нескольких миллиардов фрагментов речи только для одного диалекта языка. В то же время совершенствуются методы распознавание речи по малому словарю. Проблема исследований по распознаванию языков народов России существует давно, так как большие корпорации пока занимаются только языками с большим количеством носителей языка.

Якутский язык или язык саха относится к тюркской группе языков, имеет более 450 тыс. носителей. Отличается от других тюркских языков наличием пласта лексики неясного происхождения. В вокализме существуют краткие и долгие гласные, дифтонги, имеет финальное ударение. Авторами исследуется возможность распознавание устной речи на якутском языке.

Большое применение имеют методы распознавания речи с применением преобразования Фурье, но и другие методы имеют свои особенности. Одним из методов является вейвлет анализ, который в данное время является весьма популярным.

«Вейвлеты, как средство многомасштабного анализа позволяют выделять одновременно как основные характеристики сигнала, так и короткоживущие высо-

кочастотные явления в речевом сигнале. Это свойство является существенным преимуществом в задачах обработки речевого сигнала по сравнению с оконным преобразованием Фурье, где, варьируя ширину окна, приходится выбирать масштаб явлений, которые необходимо выделить в сигнале» [1].

«Преимущество вейвлет-преобразования над Фурье преобразованием обуславливается большей информативностью первого из них, за счет свойства локальности вейвлетов. Что непосредственно дает исследователям только глобальные сведения о частотах (масштабах) анализируемого речевого сигнала, так как большинство вейвлетов представлено вещественными функциями, так что отсутствует необходимость привлечения для их вычисления аппарата комплексных чисел, затрудняющих вычисления» [2].

«Вейвлет-спектрограммы отчетливо выделяют такие особенности сигналов, как разрывы, изменение знаков первой и второй производных, изменение частоты нестационарного сигнала и т.д. То есть те особенности сигналов, которые плохо выделяются на спектре Фурье-сигнала. Большие уровни при вейвлет-анализе характерны для коэффициентов тех вейвлетов, которые смещены вблизи той или иной локальной особенности функции» [3].

Также исследуются возможности применения различных вейвлетов для распознавания речи. В работе 4

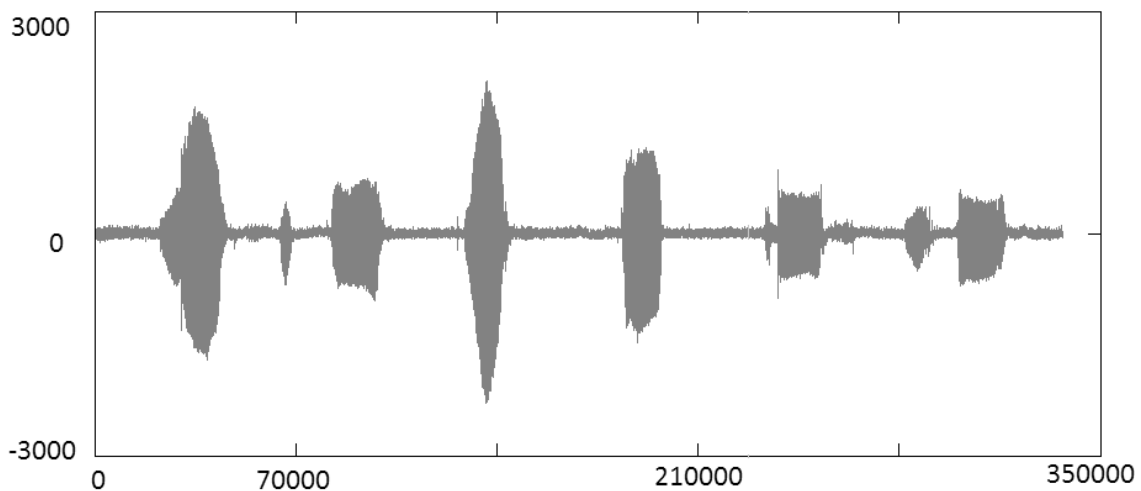


Рис.1. Звуковой сигнал из числительных.

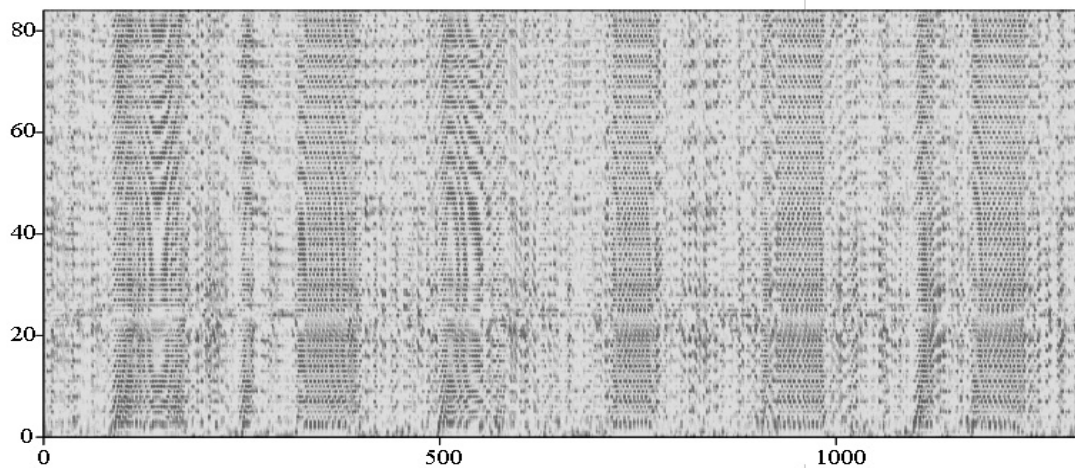


Рис. 2. Двухмерная корреляционная картина, полученная для окна размером в 512 отсчета и для вейвлета в 7 отсчетов.

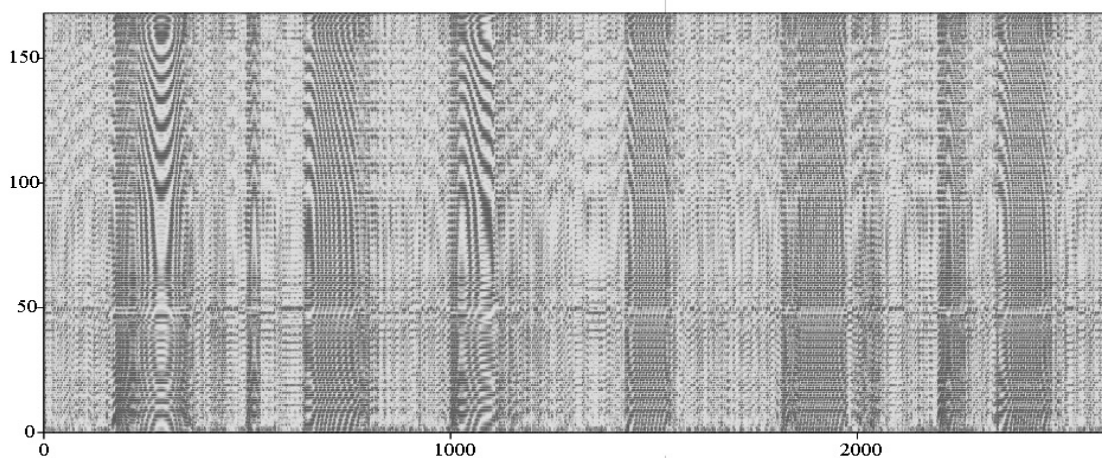


Рис. 3 Двухмерная корреляционная картина, полученная для окна размером в 256 отсчета и для вейвлета в 15 отсчетов.

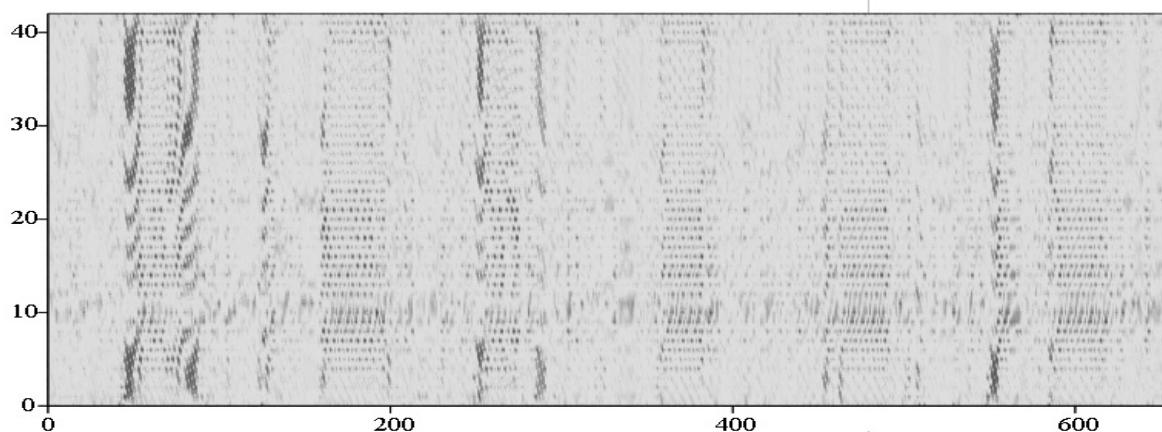


Рис. 4. Двухмерная корреляционная картина, полученная для окна размером в 1024 отсчета и для вейвлета в 15 отсчетов.

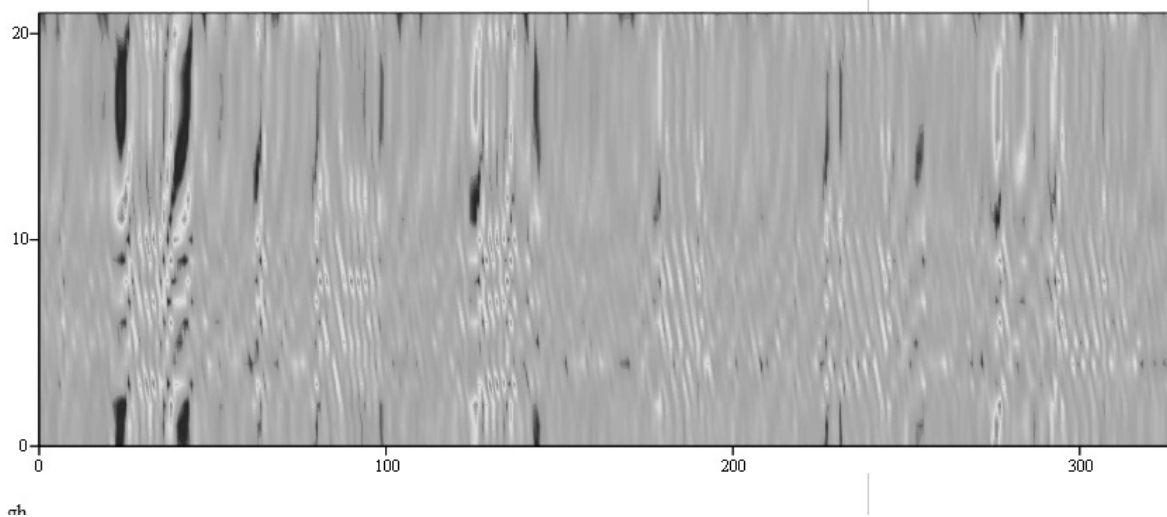


Рис. 5. Двухмерная корреляционная картина, полученная для окна размером в 2048 отсчета и для вейвлета в 15 отсчетов.

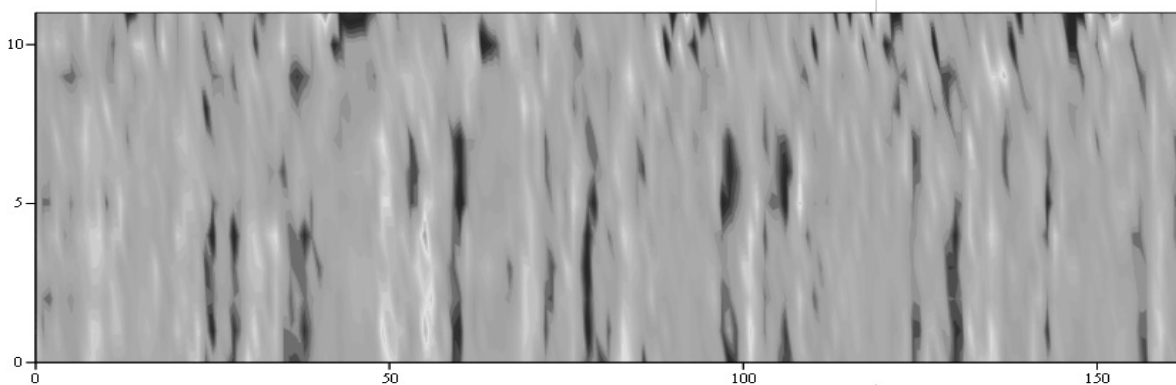


Рис. 6. Двухмерная корреляционная картина, полученная для окна размером в 4096 отсчета и для вейвлета в 15 отсчетов.

были исследованы вейвлеты Хаара, ФНАТ (французская шляпа) и МНАТ (мексиканская шляпа). «В результате проведенных исследований было выяснено, что наиболее подходящим вейвлетом для фильтрации речевого сигнала при решении задач распознавания речи является вейвлет МНАТ (мексиканская шляпа)» [4].

Автор применял вейвлет также для сегментации устной речи на якутском языке [5]. Пакет MathCAD был применен для цифровой обработки сигнала [6]. Исследование проводилось также для возможности распознавания устной речи на якутском языке с помощью вейвлетов [7].

Исследования

Для исследования была получена звукозапись из 7 слов «биир» (рус. *один* — числительное). Из данной звукозаписи были получены соответствующие окна для одного слова и преобразованы через вейвлет Добеши в последовательный массив векторов. Для функции вейвлета Добеши была использована встроенная функция *wave()* пакета MathCAD.

Для исследования был записан речевой сигнал (рис. 1), которая состоит из числительных якутского языка, от 1 до 6. Последовательность: *Биир, икки, үс, түөрт, биэс, алта*. Формат РСМ, частота дискретизации 44,1 кГц. Всего отсчетов 336373, длительность 7,62 секунд.

Для анализа был использован метод скользящих перекрывающихся окон. Окна сдвигаются на половину ширины окна, таким образом получается двойное сплошное покрытие.

Для первоначального анализа было выбрано окно размером 512 отсчетов или 11 мс. На рисунке 2 показана картина, полученная из корреляции исходного сигнала и звукового сигнала с числовой последовательностью. Двумерная корреляционная картина является неоднозначной, но там четко видно наличие слов и его границы. На картине по оси ординат приведены отсчеты эталонного сигнала, по оси абсцисс номер окна.

Увеличение числа отсчетов до 15 приводит к фильтрации сигнала, уменьшению количества совпадений.

Дальнейшее повышение количества отсчетов до 31 немного улучшает ситуацию с корреляцией. Дальнейшее увеличение отсчетов до 63 не дает большого преимущества, но увеличивает время вычислений.

Было исследовано сжатия размера окна до 256 точек для уменьшения вычислительных затрат, в результате сигнал получается зашумленным, что видно на рисунке 3.

Для окна размером 1024 отсчета или 22 мс, как видно из рисунка 4, корреляционная картина имеет более сглаженный характер, в месте нахождения первого звукового сигнала, корреляционная картина в начале и конце сигнала имеет большую корреляцию, а середина сигнала такой корреляции не имеет.

При выборе окна размером 2048 или 44 мс на корреляционной картине появились четко выделенные области при совпадении слова.

Из рисунка 6 видно, что корреляционная картина имеет большое изменение и анализ становится невозможен. Дальнейшее увеличение размера окна не дает нужной картины.

Выводы

Для анализа сигнала при окне 256 или 512 отсчетов, необходимо отслеживать характер изменения коэффициентов корреляции. На всех корреляционных рисунках четко выделяются границы слов. Распознавание слов необходимо делать по характеру изменению корреляционной картины, то есть по спаду и подъему коэффициента корреляции. Данный способ требует дополнительных исследований.

При увеличении числа отсчетов появляется корреляционные максимумы, по которым возможно сопоставить и распознать исходный сигнал. Лучший результат был достигнут при размере окна равным 2048 отсчетам и длительностью 44 мс.

Таким образом, применение пакета MathCAD дает возможность применять вейвлет-анализ для распознавания речи. Применение вейвлета Добеши позволяет распознать числительные на якутском языке.

ЛИТЕРАТУРА

1. Горшков Ю. Г. Обработка речевых сигналов на основе вейвлетов // Т-Сотт: Телекоммуникации и транспорт. 2015. Т. 9. № 2. С. 46–53.
2. Гапочкин А. В. Преимущество вейвлет-преобразования в обработке речевых сигналов // Вестник МГУП имени Ивана Федорова. 2015. № 6. С. 34–37.
3. Зубаков А. П. Фурье и вейвлет-преобразования в проблеме распознавания речи // Вестник Тамбовского университета. Серия: Естественные и технические науки. 2010. Т. 15. № 6. С. 1893–1899.

4. Балакирев Н. Е., Гуснин С. Ю., Малков М. А., Червяков Л. М. Фильтрация речевого сигнала с помощью вейвлет-преобразования при решении задач распознавания речи // Известия Юго-Западного государственного университета. 2012. № 5–2 (44). С. 044–050.
5. Леонтьев Н. А., Протодьяконова Н. А., Петухова А. А. Применение пакета MathCAD для сегментации устной речи // Международное научное издание Современные фундаментальные и прикладные исследования. 2017. № 4–1 (27). С. 74–77.
6. Леонтьев Н. А. Использование среды MathCAD для моделирования оптической обработки информации // Педагогическая информатика. 2012. № 5. С. 28–34.
7. Leontiev, N. A. The use of wavelets for the Yakut speech recognition problems // AIP Conference Proceedings 2041,050005~<2018

© Леонтьев Ньургун Анатольевич (leonza@mail.ru), Копырин Рустам Русланович (rustamkopurin@mail.ru).
Журнал «Современная наука: актуальные проблемы теории и практики»



Северо-Восточный федеральный университет им. М.К. Аммосова