

ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ ДЛЯ РАСПОЗНАВАНИЯ ИСКАЖЕННЫХ ФОНЕМ

Пылькин А.Н., Вавилова О.С.

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Рязанский государственный радиотехнический университет» (ФГБОУ ВПО «РГРТУ»), 390005, г. Рязань, ул. Гагарина, д.59/1.

В статье описан метод распознавания фонем с помощью вейвлет-преобразования, приведены экспериментальные исследования в системе Matlab. Также показано, что вейвлет-преобразование является лучшим средством обработки речевых сигналов. Дано описание устройству человеческого уха, показано, что в улитке располагается мембрана, которая является своеобразным набором биомеханических фильтров, каждый участок которой при обработке сигнала пропускает колебания только одной частоты. Также показано, что обрабатываемое и передаваемое в мозг преобразование речевого сигнала с точностью до константы совпадает с вейвлет-разложением этого сигнала. Далее представлено вейвлет-преобразование в применении к распознаванию искаженной речи. Также представлены экспериментальные исследования, проведенные в системе Matlab.

Ключевые слова: вейвлет-преобразования, кратномасштабный вейвлет-анализ, определение степени искаженности звуков речи, цифровая обработка речи, акустические свойства звуков речи, распознавание речи, распознавание фонем, обучение глухих произношению звуков речи.

VEVLET TRANSFORMATION FOR RECOGNITION OF THE DISTORTED PHONEMES

Pilkin A.N., Vavilova O.S.

Federal public budgetary educational institution of the higher professional education "Ryazan state radio engineering university" ("RGRU" FGBOU VPO), 390005, Ryazan, Gagarin St., 59/1.

In article the method of recognition of phonemes by means of vevlet-transformation is described, pilot studies are given in Matlab system. It is also shown that vevvlet-transformation is the best means of processing of speech signals. The description is given to the device of a human ear, shown that in a snail the membrane which is a peculiar set of the biomechanical filters which each site when processing a signal passes fluctuations of only one frequency settles down. It is also shown that transformation of a speech signal processed and transferred to a brain to within a constant coincides with vevlet-decomposition of this signal. Further it is presented vevlet-transformation in application to recognition of the distorted speech. The pilot researches conducted in Matlab system are also presented.

Keywords: vevlet-transformations, vevlet-analysis, definition of a degree of sounds of speech, digital processing of speech, acoustic properties of sounds of speech, recognition of speech, recognition phonem training of deaf persons to a pronunciation of sounds of speech.

Введение. Использование вейвлетов в задачах обработки и распознавания речи продиктовано особенностями речевого акустического сигнала. Вейвлеты как средство многомасштабного анализа позволяют выделять одновременно как основные характеристики сигнала, так и короткоживущие высокочастотные составляющие в речевом сигнале, которые существенно влияют на качество произношения.

Люди с врожденной глухотой, но с не нарушенным речевым аппаратом имеют возможность научиться говорить и не являются истинно глухонемыми. Их немота обусловлена тем, что они не слышат звуки, которые произносят, что и является основной трудностью при обучении их речи. Представленная система дает глухонемым возможность обучения правильному произношению фонем [5].

Целью написания данной статьи является описание метода распознавания искаженных фонем с помощью вейвлет-преобразования. Сейчас 3 % людей в обществе являются глухонемыми, их немота обусловлена тем, что они не слышат произносимые ими звуки. Эти люди с ограниченными возможностями произношения сталкиваются с серьезными проблемами в общении с другими людьми. С помощью рассмотренного в статье метода можно создать систему, которая поможет глухонемым общаться, понимать и быть понятыми окружающими людьми, без знания сложных аспектов дактильной и жестовой речи [4].

Таким образом, появляется возможность самостоятельного обучения глухонемых правильному произношению звуков речи в домашней обстановке или в школе для детей при наличии компьютера.

Устройство человеческого уха. Внутреннее ухо человека представляет собой сложный лабиринт соединенных друг с другом каналов, которые расположены в височной кости и заполнены специальной жидкостью. Часть этого лабиринта образует вестибулярную систему, а одна из костных структур, имеющая форму спирали, относится к слуховой системе и называется улиткой.

В улитке происходит преобразование механических процессов в нервные сигналы, которые затем передаются в мозг и анализируются.

При передаче звуковых колебаний основание стремечка в такт с изменением звукового давления вдавливают мембрану овального окна в жидкость вестибулярного канала, создавая в нем избыточное давление, которое практически мгновенно распространяется по всей длине улитки. При этом с информационной точки зрения со звуком происходят определенные преобразования, которые сводятся в основном к усилению. Таким образом, большинство элементов в улитке имеют определенный коэффициент усиления k . Увеличение давления приводит в движение податливые структуры среднего канала – Рейснерову мембрану, жидкость среднего канала и базилярную мембрану, что, в свою очередь, вызывает изменение давления жидкости в тимпанальном канале и смещение мембраны круглого окна. Реакция базилярной мембраны на отдельный импульс увеличения давления в жидкости улитки не является мгновенной и может быть описана следующим образом. Сначала мембрана «выбухает» в сторону тимпанального канала у своего основания, затем в силу эластичности возвращается в положение равновесия. Возникшее выбухание перемещается вдоль длины мембраны к вершине за время, равное нескольким мс. Таким образом, на мембране возникает бегущая волна смещения ее локальных участков относительно положения, которые они занимают вне звука.

Реакция базилярной мембраны на частоту звуковых колебаний является следствием ее упругих свойств и анатомического строения. Разные участки мембраны отличаются шириной и жесткостью. Ширина мембраны увеличивается по направлению от основания к вершине

примерно в десять раз, а упругость постепенно уменьшается: у основания мембрана более чем в 100 раз жестче, чем у вершины. Приблизительно половина мембраны, считая от ее вершины, реагирует на колебания с частотами от 0 до 2000Гц, в то время как оставшаяся часть (у основания) реагирует на остальные частоты звукового диапазона. Такое распределение означает, что различия между звуками в области низких частот обнаруживаются слуховой системой гораздо лучше, чем различия в высоких частотах. Можно сказать, что реакция мембраны на частоту сигнала приблизительно следует логарифмической шкале.

Интенсивность звука также влияет на колебательный процесс, возникающий в мембране: чем сильнее звук, тем больше размах колебаний всех участков мембраны, в том числе и участка с максимальным смещением. Размер зоны возбуждения вблизи точки максимального смещения определяется как частотой звука, так и его амплитудой. Связь местонахождения максимального возбуждения базилярной мембраны с частотой звука свидетельствует о частотно-избирательной или фильтрующей функции этого слухового органа, в физиологии слуха связь между местом максимального смещения базилярной мембраны и частотой возбуждающего сигнала называют тонотопической организацией. Этот же термин распространяется и на более высокие уровни слуховой системы для обозначения пространственной упорядоченности нервных элементов в соответствии с их частотной избирательностью [7].

Таким образом, во внутреннем ухе человека осуществляется перекодирование частоты и интенсивности звукового сигнала в активность, локализованную на определенном участке базилярной мембраны. Различные участки мембраны являются своеобразными биомеханическими фильтрами, а мембрана в целом может рассматриваться как набор фильтров, упорядоченных по частоте и покрывающих в совокупности всю область частот, доступных восприятию человека. В техническом спектральном анализе фильтр является устройством, которое, обрабатывая сигнал, пропускает колебания только определенных частот. Результирующее преобразование сигнала во внутреннем ухе человека с точностью до константы совпадает с вейвлет-преобразованием сигнала. Следовательно, восприятие человеческим ухом звуковых колебаний, и в том числе человеческой речи, с математической точки зрения соответствует вейвлет-разложению. Поэтому естественно в системе распознавания речи использовать вейвлет-разложение [6].

Вейвлет-преобразование речевого сигнала

Вейвлет-преобразования одномерного сигнала (формула 1) – это его представление в виде обобщенного ряда или интеграла Фурье по системе базисных функций, сконструированных из материнского вейвлета $\psi(t)$, обладающего определенными свойствами за счет операций сдвига во времени b и изменения временного масштаба a . Множитель $1/\sqrt{a}$ обеспечивает независимость нормы этих функций от масштабирующего числа a .

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (1)$$

Для заданных значений параметров a и b функция $\psi_{ab}(t)$ и есть вейвлет, порожаемый материнским вейвлетом [8].

В частотной области вейвлеты похожи на всплески с пиком на частоте w_0 , то есть имеют вид полосового фильтра, при этом w_0 и Δw уменьшаются с ростом параметра a .

Следовательно, вейвлеты локализованы как во временной, так и частотной областях.

Вейвлет-преобразование обеспечивает двумерное представление исследуемого сигнала в частотной области в плоскости частота-положение. Аналогом частоты при этом является масштаб аргумента базисной функции – времени, а положение характеризуется ее сдвигом. Это позволяет разделять крупные и мелкие особенности сигналов, одновременно локализуя их на временной шкале. Иными словами, вейвлет-анализ можно охарактеризовать как спектральный анализ локальных возмущений.

Спектральное представление вейвлетов аналогично заданию окна в оконном преобразовании Фурье. Но отличие состоит в том, что свойства окна, его ширина и перемещение по частоте, присущи самим вейвлетам. Это служит предпосылкой их адаптации к сигналам, представляемым совокупностью вейвлетов. Поэтому с помощью вейвлетов можно осуществить анализ и синтез локальной особенности речевого сигнала.

Вейвлет-преобразование не просто «режет» исследуемый объект на куски, а выделяет из него компоненты разных масштабов, и каждая компонента анализируется с той степенью детальности, которая соответствует его масштабу. Следовательно, вейвлет-преобразование можно использовать для обработки нестационарных сигналов, которым и является речевой сигнал.

Представим речевой сигнал как функцию f из бесконечно мерного Гильбертова пространства $L^2(R)$. Эту функцию f разложим на некотором заданном уровне разрешения j_n , (где $n=1,2,3\dots$) в ряд вида, представленного формулой 2.

$$f(t) = \sum_k s_{j_n,k} \varphi_{j_n,k} + \sum_{j \geq j_n,k} d_{j,k} \psi_{j,k} \quad (2)$$

где $\varphi_{j_n,k}$ и $\psi_{j_n,k}$ – масштабированные и смещенные версии скейлинг-функции (масштабной функции) φ и «материнского вейвлета» ψ ; $s_{j,k}$ – коэффициенты аппроксимации; $d_{j,k}$ – детализирующие коэффициенты.

Вейвлет-коэффициенты аппроксимации соответствуют передаточной характеристике фильтра низких частот, а коэффициенты детализации соответствуют высокочастотному фильтру [2].

Вейвлет-коэффициенты $s_{j,k}$ и $d_{j,k}$ можно вычислить по формулам 3 и 4.

$$s_{j,k} = \int f(x)\varphi_{j,k}(x)dx \quad (3)$$

$$d_{j,k} = \int f(x)\psi_{j,k}(x)dx \quad (4)$$

Первая сумма в (6) со скейлинг-функциями $\varphi_{j,k}$ содержит средние значения f по диадным интервалам $[k2^j, (k+1)2^j)$. Усреднение проводится с весовыми функциями $\varphi_{j,k}(x)$. Второй член содержит все флуктуации f на данном интервале. Эти флуктуации проистекают из всех меньших интервалов, заключенных внутри данного и соответствующих большим значениям параметра масштабирования f . Можно сказать, что этот член как бы фокусирует наше внимание на все более тонких деталях изучаемого сигнала. Это последовательное изучение свойств сигнала происходит через равные промежутки, как следует из постоянства масштабирующего множителя, причем детали размером 2^j располагаются в точках $k2^j$. На низшем уровне разрешения j_0 , наиболее широких интервалах, первая сумма содержит всего один член, дающий общее усредненное взвешенное значение сигнала $\langle f \rangle = s_{j_0} k_0$, где k_0 обозначает центр гистограммы. Вторая сумма в (6) показывает флуктуации на всех без исключения уровнях. На следующем уровне с более мелким разбиением $j_1 > j_0$ в первой сумме содержатся два члена, отвечающих за средние значения f в полуинтервалах с центрами, расположенными в k_1, k_2 . Число членов во второй сумме уменьшается на единицу – исчезает член, который раньше показывал величину флуктуаций на половинном масштабе. Общее число членов в разложении остается неизменным. Число членов в каждой сумме зависит от выбранного уровня разрешения. Меняя на единицу этот уровень, мы передвигаем некоторые члены в другую сумму, причем каждое из этих представлений является истинным представлением исходной гистограммы на разных уровнях разрешения [3].

Экспериментальные исследования. В процессе исследования фоном были проведены эксперименты в системе Matlab, в результате которых были выделены характерные признаки каждой из фонем, представленных с помощью вейвлет-преобразования [1]. В следующих таблицах приведены вектора из восьми коэффициентов, которые представляют собой средние значения энергий детализирующих коэффициентов, рассчитанные на каждом уровне вейвлет-разложения. Жирным шрифтом выделены характерные особенности каждой из фонем.

Таблица. Вектора средних значений энергий детализирующих коэффициентов, рассчитанные на каждом уровне вейвлет-разложения для фонем ‘а’, ‘и’, ‘о’

a1	и2	o1
1.0e+003	1.0e+003	1.0e+003
0.0020	0.0142	0.0001
0.0117	0.0575	0.0006
0.1657	0.0099	0.0041
5.1521	0.0231	0.2986

0.7928	0.1897	7.1588
0.3462	2.8531	0.2790
0.0895	0.0623	0.0159
0.0388	0.0081	0.0133

Из таблицы видно, что вейвлет-преобразование

методом многомасштабного вейвлет-анализа позволяет выявить характерные признаки каждой фонемы, также как это происходит на базилярной мембране внутреннего уха человека. Затем вектора подаются на вход нейросети и распознаются, так же как в человеческом мозге. Следовательно, предложенная система работает по алгоритму человеческого уха, идеальной системы распознавания фонем созданной самой природой.

На рисунке 1 показаны результаты распознавания фонем системой с помощью оконного преобразования Фурье.

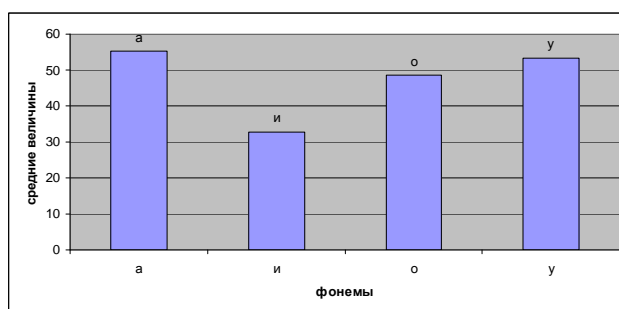


Рисунок 1. График средних арифметических не взвешенных величин результатов распознавания фонем ‘а’, ‘и’, ‘о’, ‘у’ нескольких людей одного пола системой, использующей оконное преобразование Фурье, обученной на фонемах одного из них

Из рисунка 3 следует, что предполагаемая модель и ее исполнение показали хорошие результаты правильности распознавания фонем, но все средние величины результатов распознавания меньше шестидесяти. Это обусловлено тем, что метод Фурье преобразования не решает проблему дикторозависимости системы. Следовательно, все системы, построенные на преобразовании Фурье, являются дикторозависимыми, то есть предназначены для работы только с одним пользователем, на фонемах которого обучалась данная система. Однако в нашем случае это невозможно, так как человек, чьи фонемы должна распознавать система, не может правильно произнести их.

На следующем рисунке 2 приведены результаты распознавания фонем системой с помощью вейвлет-преобразований.

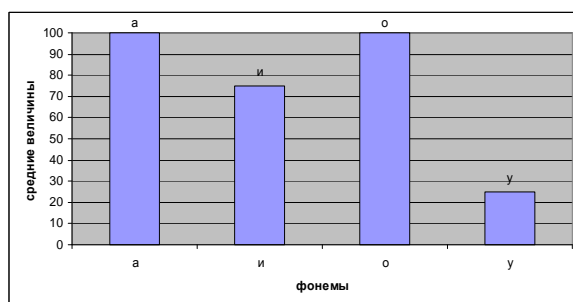


Рисунок 2. График средних арифметических не взвешенных величин результатов распознавания фонем ‘а’, ‘и’, ‘о’, ‘у’ нескольких людей одного пола системой, использующей вейвлет-преобразование, обученной на фонемах одного из них

Из рисунка 4 видно, что система, основанная на методе вейвлет-преобразования, близка к дикторонезависимым, что для решения нашей задачи является необходимым. Система обучается на векторах фонем правильного произношения, а затем распознает искаженное произношение этой фонемы.

Заключение. В результате проделанной работы можно сделать вывод, что вейвлет-анализ как средство многомасштабного анализа может быть успешно применен для исследования особенностей речевого сигнала. Исходя из результатов эксперимента, можно сказать, что при использовании вейвлет-преобразования для решения задач распознавания речи процент распознавания фонем гораздо выше, чем при использовании преобразования Фурье. Это объясняется тем, что вейвлет-преобразование гораздо ближе к естественному восприятию звука человеческим ухом.

Список литературы

1. Вавилова О.С., Пылькин А.Н., Дондик Е.М., Автоматизированный анализ искажений произносимых фонем // Оптимизация процесса образования и воспитания детей с ограниченными возможностями. Монография. – Пенза, 2013. – С.101-111.
2. Витязев В.В. Вейвлет-анализ временных рядов: Учеб.пособие. – СПб.: Изд-во С.-Петербур. ун-та, 2001. – 58 с.
3. Дондик Е.М., Пылькин Н.Е. Скоробогатова А.Н. Информационная система распознавания образов переменной структуры // Вестник Рязанского государственного радиотехнического университета. – 2011. – № 3 (вып. 37). – С. 17-23.
4. Дондик Е.М., Пылькин А.М., Скоробогатова Н.Е. Математическое представление селекции распознаваемых знаков дактильной речи. – Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – 2012. – № 1 (140). – С.55-60.
5. Кроцилин А. В., Крошила С.В. Проектирование систем поддержки принятия решений для оценки состояния здоровья пациентов в условиях неопределенности // Информатика и системы управления. – 2010. – № 4 (26). – С. 82-94.
6. Кроцилин А.В., Крошила С.В., Пылькин А.Н. Некоторые аспекты построения систем поддержки принятия решений для эффективного управления товарными запасами // Радиотехника. – 2012. – № 3. – С.103-108.

7. Фланаган Д.Л. Анализ, синтез и восприятие речи: Пер. с англ. под ред. Пироговой А.А. – М.: СВЯЗЬ, 1968. – 392 с.

8. Яковлев А.Н. Введение в вейвлет-преобразования: Учеб. пособие. – Новосибирск: Изд-во НГТУ, 2003. – 104 с.

Рецензенты:

Кузнецов А.Е., д.т.н., профессор, заместитель директора научно-исследовательского института обработки аэрокосмических изображений (НИИ «Фотон»), г. Рязань.

Мусолин А.К., д.т.н., профессор, зав. кафедрой АИТП Рязанского государственного радиотехнического университета, г. Рязань.