УДК 004.9

АНАЛИЗ ОТКЛИКА СИСТЕМЫ ОСЦИЛЛЯТОРОВ ДЛЯ РАСПОЗНАВАНИЯ ЗВУКОВ РУССКОЙ РЕЧИ

Д. В. Болотнов, С. А. Запрягаев

Воронежский государственный университет

Поступила в редакцию 01.02.2012 г.

Аннотация. В статье описывается метод выделения вектора акустических признаков на основе анализа отклика системы связанных осцилляторов. Производится исследование зависимости результатов распознавания от выбора коэффициентов осцилляторной системы. В качестве коэффициентов исследуются также формантные области спектра сигнала. Использование данного подхода рассматривается на проблеме распознавания речи.

Ключевые слова: распознавание речи, система осцилляторов, вектор отличительных признаков.

Annotation. The article describes the method of allocation vector of acoustic features based on analysis of the oscillator's system response. Performed to research the dependence of recognition results on the choice of the coefficients of the oscillator system. As the factors it's also used formants of a signal's spectrum. Using this approach, it can be used to improve a speech recognition system.

Keywords: speech recognition, oscillator system, vector of acoustic features.

Звук представляет собой колебания плотности среды, в которой они распространяются. Колебания среды, дошедшие до барабанной перепонки, трансформируются в слуховом тракте в систему сигналов поступающих в головной мозг человека, интерпретирующий эти сигналы в виде осмысленной речи или шума.

В целом, слуховая система человека состоит из наружного уха, среднего уха и улитки внутреннего уха [1]. Наружное ухо выполняет функцию волновода, проводящего колебания среды к барабанной перепонке, которая разграничивает наружное и среднее ухо. Колебания барабанной перепонки передаются через систему слуховых косточек вестибулярной перилимфе[1]. Колебания перилимфы вызывают колебания базилярной мембраны. В результате базилярная и текториальная мембраны испытывают относительное друг друга смещение, что приводит к изгибанию стереоцилий[2] волосовых клеток и изменению мембранного потенциала последних. Слуховой тракт начинается от первичных чувствительных нейронов вблизи улитки. Аксоны нейронов следуют к ядрам мозга. В целом, в тракте реализуется правило в соответствии с которым, чем выше расположен нейрон по тракту, тем более сложные характеристики колебаний требуются для его активации. Так, например, в так называемых, кохлеарных ядрах большинство нейронов возбуждается звуками строго определённой частоты. Соответственно в оливарном комплексе имеются нейроны, которые реагируют на частотно-модулированные тоны. В четверохолмии, как правило, нейроны реагируют на амплитудно-модулированные тоны (звуки переменной громкости). Среди нейронов слуховой коры есть клетки, реагирующие только на начало или окончания возбуждения.

Таким образом, в слуховом тракте колебания непрерывно анализируются, причём на каждом последующем шаге анализа распознаются всё более тонкие характеристики звукового колебания. На основании анализа физиологического распознавания звуковых колебаний в первом приближении можно представить звуковой тракт системой связанных осцилляторов с различными характеристиками «жёсткости» пружин и масс, к которой приложена вынуждающая сила S(t) моделирующая исходное звуковое колебание:

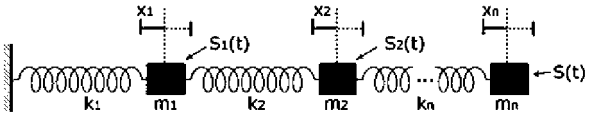

Рис. 1. Гармонический осциллятор

[1] Перилимфа – вязкая жидкость, заполняющая полости улитки

[2] Стереоцилии волосковых клеток – реснички

Идея распознавания речевых образов с использованием системы связанных осцилляторов, состоит в выделении дополнительных отличительных признаков речевого сигнала (помимо стандартных признаков), анализируя отклик системы осцилляторов при подаче на неё сигнала.

Общая теория колебаний в цепочках однородных и периодически чередующихся элементов хорошо известна [2]. Обозначая смещение i-ой массы от положения равновесия через x_i, система уравнений движения цепочки для n тождественных масс m с одинаковыми коэффициентами жесткости пружин k имеет вид [2]:

$$\ddot{x}_i + k(2x_i - x_{i+1} - x_{i-1}) = 0.$$

При этом n собственных частот удовлетворяют соотношению:

$$\omega_m = 2\sqrt{\frac{k}{m}} \sin \frac{m \to k\pi}{2(n+1)}; k = 1,2,...n$$

В случае вынужденных колебаний при наличии трения система уравнений имеет вид [2]

$$m\ddot{x}_i + \alpha\dot{x}_i + k(2x_i - x_{i+1} - x_{i-1}) = 0.$$

С граничным условием, определяющим движение крайнего элемента цепочки под действием вынуждающего колебания. В результате решение имеет структуру бегущей волны. Приведенные выше системы уравнений обобщаются и на случай линейных цепочек периодически чередующихся элементов.

В настоящей работе слуховой тракт моделируется системой связанных осцилляторов, находящихся под воздействием звукового сигнала и распознавание сигнала осуществляется на основе совокупных данных стандартного Фурье анализа исходного сигнала и Фурье анализа отклика системы осцилляторов на звуковой сигнал.

АНАЛИЗ И РАСПОЗНАВАНИЕ ЗВУКОВОГО СИГНАЛА

Процедура обработки звукового сигнала состоит в предварительном очищении сигнала от шума, путём простого отсечения амплитуды сигнала по заданному порогу. На следующем этапе к полученному сигналу применяется оконное преобразование Хэмминга [3],

$$a_n = 0.53836 - 0.46164\cos\left(\frac{2\pi n}{N-1}\right).$$

которое позволяет избежать резких всплесков сигнала по границам окна и повысить чёткость сигнала в середине. Пример использования окна Хэмминга представлен на рисунке 2:

В процессе распознавания, вектор отличительных признаков формируется в несколько этапов. Звуковой сигнал предварительно обрабатывается и подаётся на систему, связанных осцилляторов. Поданное возмущение вызывает движение связанных осцилляторов. Координаты частиц передаются на модуль преобразования Фурье. Учитывая особенности слухового восприятия человека анализируются только частоты в диапазоне от 0Hz до 3kHz. Исходный сигнал поступает в формате Wave 16-bit, 11025Hz. Весь диапазон анализируемых частот разбивается на 20 равных областей и среди каждой находится максимальное значение. Полученные максимальные значения образуют вектор отличительных признаков для каждого элемента системы связанных осцилляторов.

Для распознавания звуков используется нейронная сеть с одним скрытым слоем, на котором располагается 21 нейрон. Обучение осуществлялось методом обратного распространения ошибки [4, 5].

Общая схема работы программы представлена в виде последовательной работы спроектированных модулей (рис.3):

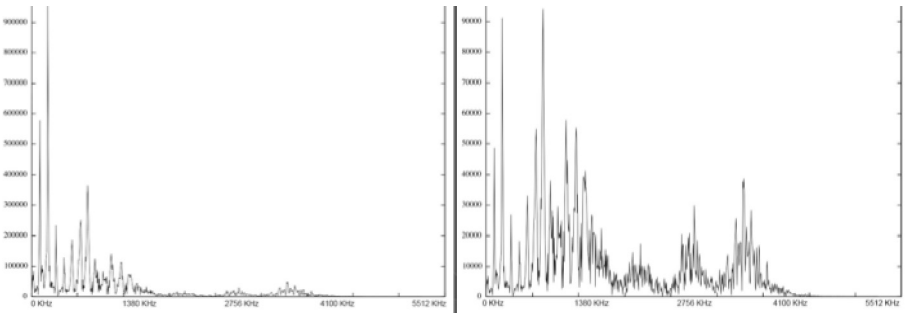

Рис. 2. Преобразование Хэмминга. Слева – исходный сигнал, справа – преобразованный

Так как цепочка осциллирующих элементов может содержать элементы с различными параметрами, то очевидным является вопрос об оптимизации параметров системы по результату распознавания звуковых образов. Оптимизация параметров системы может быть направлена как на подбор оптимального количества осциллирующих элементов, так и на определение оптимальных коэффициентов «упругости пружин» и оптимизацию масс осциллирующих элементов.

Для определения оптимального числа осциллирующих элементов в системе, были исследованы три системы с различным количеством осциллирующих элементов: 2, 3 и 4. Во всех случаях сигнал подаётся на последний элемент, при этом его масса и коэффициент упругости соответствующей ему пружины в два раза больше остальных. Так для случая с двумя элементами, параметры системы выбирались следующими:

$m_1 = 1000$; $m_2 = 2000$; $k_1 = 0.001$; $k_2 = 0.002$;

Для случая с тремя элементами:

$m_1 = 1000$; $m_2 = 1000$; $m_3 = 2000$;

$k_1 = 0.001$; $k_2 = 0.001$; $k_3 = 0.002$;

Наконец для случая с четырьмя элементами коэффициенты выбирались в виде:

$m_1 = 1000$; $m_2 = 1000$; $m_3 = 1000$; $m_4 = 2000$;

$k_1 = 0.001$; $k_2 = 0.001$; $k_3 = 0.001$; $k_4 = 0.002$;

Во всех случаях коэффициенты и массы подбирались из условия принадлежности главных резонансов системы выбранному диапазону частот.

Для первого эксперимента были выбраны пять русских гласных звуков: [а, и, у, э, о] [6]. Общее количество образцов: по 20 на каждый звук. Из них: 5 образцов на обучение, 15 образцов для оценки качества распознавания. Нейронная сеть содержала один скрытый слой, на котором размещены 5 нейронов. После успешного обучения, были получены следующие результаты распознавания данной выборки (табл. 1).

Как видно из таблицы 1, все три системы распознали пять гласных звуков и дальнейших экспериментов по увеличению числа осцилляторных элементов не производилось.

Для большей информативности была взята выборка из 21 звука русского языка. Нейронная сеть также состояла из одного скрытого слоя, на котором расположен 21 нейрон. Допустимая ошибка составляла 10^{-5}. Проверка распознавания производилась изначально на обучающей выборке на каждой из осциллирующих систем. Результаты распознавания в процентах представлены в таблице 2:

Из полученного результата видно, что увеличение осциллирующих элементов даёт значительного увеличения в эффективности распознавания. Но, одновременно приводит к существенному снижению вычислительной производительности. В качестве примера на рис. 4 приведены данные по времени распознавания 10 образцов на каждый звук для данного эксперимента:

Все эксперименты выполнялись на двухъядерном процессоре Intel Core2Duo 2.26GHz с 8Gb оперативной памяти. Оценивая результаты данных экспериментов, выбор был осуществляем на системе с тремя осциллирующими элементами, как на оптимальном соотношении скорости и эффективности. Четыре осцилляторных элемента не дают существенного прироста эф-

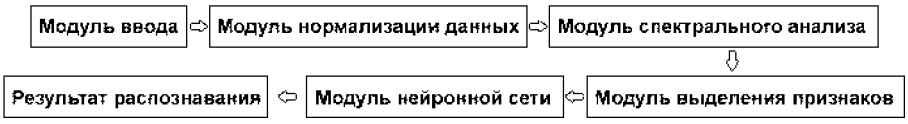

Рис. 3. Общая схема работы программы

Таблица 1

Зависимость результатов распознавания от количества осциллирующих элементов

	Количество элементов		
	2	3	4
а	100%	100%	100%
и	100%	100%	100%
у	100%	100%	100%
э	100%	100%	100%
о	100%	100%	100%

Таблица 2

Распознавание выборки из 21 звука русских букв

	Распознавание по обученным (в %)			Распознавание по необученным (в %)		
	Число осциллирующих элементов			Число осциллирующих элементов		
звук	2	3	4	2	3	4
а	100%	100%	100%	87%	87%	87%
и	100%	100%	100%	100%	100%	100%
у	80%	100%	60%	100%	100%	100%
э	100%	100%	100%	100%	100%	100%
о	100%	100%	100%	100%	100%	100%
б	40%	40%	40%	73%	80%	73%
в	80%	40%	100%	100%	100%	87%
г	100%	60%	100%	100%	100%	100%
д	80%	80%	40%	93%	93%	93%
ж	20%	100%	100%	100%	100%	100%
к	100%	100%	60%	100%	100%	93%
л	100%	100%	80%	100%	100%	93%
м	100%	100%	80%	100%	100%	100%
н	100%	100%	100%	100%	100%	100%
п	100%	20%	100%	93%	80%	100%
р	100%	100%	100%	93%	87%	93%
с	60%	100%	100%	93%	100%	100%
т	60%	40%	60%	87%	93%	87%
ф	40%	60%	40%	100%	93%	100%
х	40%	100%	100%	87%	87%	100%
ш	100%	100%	100%	100%	100%	100%
Среднее:	80.95%	82.86%	83.81%	95.55%	95.23%	95.52%

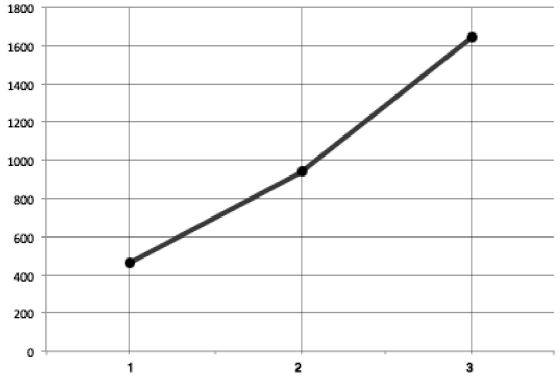

Рис. 4. График зависимости количества осциллирующих элементов (ось OX) от времени работы системы (ось OY выраженная в секундах)

фективности и значительно увеличивают время, затрачиваемое на распознавание. В то время как система с двумя осцилляторными элементами даёт ошибки чаще и в большем количестве звуков, нежели система осцилляторов с тремя элементами.

Низкая скорость вычислений в реальных условиях не является существенной проблемой, поскольку производительность может быть увеличена за счёт запуска исполняемого файла на высокопроизводительном кластере. Также, распараллеливание может быть реализовано в виде отдельной (самостоятельной) надстройки над системой распознавания. Если алгоритм не предусматривает параллельных вычислений, можно использовать процессор видеокарты, который последнее время часто используется для сложных вычислений.

Второй задачей при исследовании откликов системы осцилляторов является исследование зависимости отдельных параметров системы на результат обучения нейронной сети и распознавания. В качестве параметров выступают – коэффициенты упругости пружин и массы осциллирующих элементов.

Эксперимент состоял в изменении третьего коэффициента упругости пружины, поскольку в нашем случае на третий осциллирующий элемент подаётся звуковой сигнал. В данном эксперименте выполнен анализ работу системы распознавания со следующими коэффициентами:

$$k_1 = 10;\ k_2 = 10;\ k_3 = 0...4000;$$

$$m_1 = 0.001;\ m_2 = 0.001;\ m_3 = 0.001;$$

В данном случае, третий коэффициент жёсткости пружины будет варьировался в диапазоне от 0 до 4000. Каждый такой проход занимает длительное время, поэтому шаг выбирался переменным. После работы данного алгоритмы был получен график зависимости распознавания как функция параметров третьего осциллятора. Вектор отличительных признаков в данных экспериментах формировался только из коэффициентов, полученных по анализу работы системы осцилляторов.

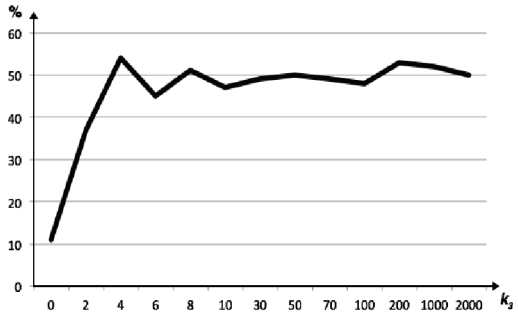

Рис. 5. График зависимости результатов распознавания от параметров осцилляторной системы при $k_1 = 10;\ k_2 = 10;\ k_3 = 0...4000$

На графике (рис. 5) на оси ординат указан результат распознавания в процентах, как функция от значения k_3. Как видно результат распознавания следует признать неудовлетворительным. Результат распознавания в аналогичной конфигурации системы, но при значениях жесткости пары пружин увеличенной в десять раз ($k_1 = 100;\ k_2 = 100$) представлен на рис. 6. ($m_1 = 0.001;\ m_2 = 0.001;\ m_3 = 0.001$)

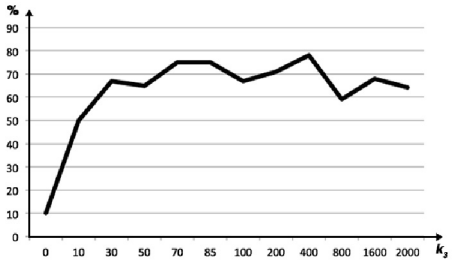

Рис. 6. График зависимости результатов распознавания от параметров осцилляторной системы при $k_1 = 100;\ k_2 = 100;\ k_3 = 0...3000$

На следующем этапе исследовались коэффициенты вида $k_1 = 1000;\ k_2 = 1000;\ k_3 = 0..20000;$ $m_1 = 0.001;\ m_2 = 0.001;\ m_3 = 0.001$. Результат представлен на рис 7.

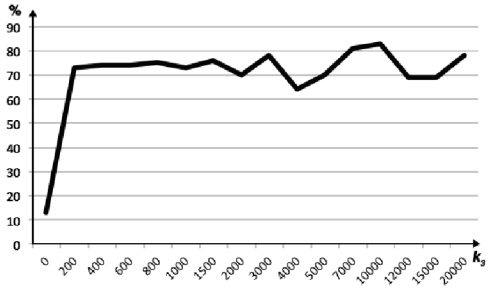

Рис. 7. График зависимости результатов распознавания от параметров осцилляторной системы при $k_1 = 1000;\ k_2 = 1000;\ k_3 = 0..20000$

Данные результаты отличаются в лучшую сторону от результатов, полученных от предыдущих двух исследований. Это происходит потому, что осцилляторная система более приближена по собственным частотам к реальным частотам сигнала. Очевидно, что при некоторых параметрах, характер отклика осцилляторной системы на поданный звуковой сигнал, позволяет эффективнее разделять звуковые образы.

Результат исследований показывают, что наблюдается зависимость степени распознавания как от порядка, так и от значений самих коэффициентов. Для лучшего варианта распознавания звуков с использованием системы связных осцилляторов, необходим поиск и установка подходящих коэффициентов. Целью данных экспериментов было выявление таких коэффициентов, при которых наилучшим образом можно было бы классифицировать звуковые образы.

С целью дальнейшего улучшения результатов распознавания был проверен отклик осцилляторной системы настроенной на чувствительность к формантам говорящего. Форманты –

часть тонового звукового спектра (частотная область), описываемая усреднённой частотной величиной. В данной частотной области, под воздействием резонанса усиливается некоторое число гармоник, которые производит голосовой аппарат человека. В следствии этого, на спектре появляется достаточно отчётливо заметная область усиленных частот, определяемые по усреднённой частотной величине. Форманты являются важнейшим информационным параметром в спектре сигнала, которые характеризуют распределение и концентрацию энергии в ограниченной частотной области. Форманта может быть характеризована частотой, шириной и амплитудой. Под частотой форманты подразумевается частота максимальной амплитуды в пределах форманты. Таким образом, форманта это – определённый амплитудный всплеск на графике звукового спектра и его частота при этом – частота пика всплеска. Количество формант определяется количеством резонансных полостей в речевом тракте человека. Обозначаются форманты латинским символом: F с указанием порядкового индекса форманты. В речевом спектре выделяют несколько формант: F1 – 500Hz, F2 – 1500Hz, F3 – 2500Hz, F4 – 3500Hz. Среднее расстояние между формантами у мужских голосов составляет порядка одной тысячи герц, у женских и детских – несколько больше. Считается, что для характеристики звуковой речи достаточно четырёх формант. Хотя в большинстве случаев для различения гласных звуков оказывается достаточным анализ, всего лишь, двух первых формант. Но практически всегда, количество формант, присутствующее в звуковом спектре, более, чем две, что может говорить о наличии более сложной зависимости артикуляции и акустическими характеристиками звука.

При анализе гласных звуков, используемых в экспериментах, для диктора отчётливо были заметны формантные области на следующих диапазонах частот: 700–800 Гц, 1700–1800 Гц и 2700–2900 Гц. При настройке системы осцилляторов так, чтобы его собственные частоты совпадали с форматными областями частот. Необходимо выбрать следующие коэффициенты упругости пружин в виде $k_1 = 600$, $k_2 = 3000$, $k_3 = 8000$.

При этом собственные частоты системы попадают в частоты форматных диапазонов говорящего. После того, как данные параметры установлены в системе, полученные результаты распознавания оказались равными 89%, что является лучшим результатом из всех экспериментов с параметрами системы.

Для улучшения уровня распознавания требуется более тщательная оптимизация параметров и числа элементов системы. Однако как установлено из данных экспериментов, система осцилляторов состоящая из небольшого числа осциллирующих элементов, в первом приближении вполне пригодна для решения задачи распознавания элементов русской речи.

СПИСОК ЛИТЕРАТУРЫ

1. *Смирнов В.* Физиология человека. – М.: Медицина, 2002. – 608 с.

2. *Ланда П.С.* Нелинейные колебания и волны. Книжный дом «Либроком». Москва. 2010. – 552 с.

3. *Хемминг Р. В.* Цифровые фильтры. – М.: Недра, 1987. – 221 с.

4. *Хайкин С.* Нейронные сети. – М.: Вильямс, 2005. – 1104с.

5. *Терехов В.А., Ефимов Д. В., Тюкин И. Ю.* Нейросетевые системы управления. – 1-е. – Высшая школа, 2002. – 184 с.

6. *Болотнов Д.В., Запрягаев С.А.* Распознавание гласных звуков русского алфавита. – Информатика: проблемы, методология, технологии : материалы XI междунар. науч.-метод. Конф., 2011

7. *Сергиенко А. Б.* Цифровая обработка сигналов. – 2-е. – СПб.: Питер, 2007. – 751 с.

8. *Рабинер Л., Гоулд Б.* Теория и применение цифровой обработки сигналов. – М.: Мир, 1978. – 848 с.

Болотнов Денис Владимирович – аспирант цифровых технологий Воронежского государственного университета. Тел.: 8-903-030-08-42. E-mail: denis1503@gmail.com

Запрягаев Сергей Александрович – доктор физико-математических наук, профессор кафедры цифровых технологии Воронежского государственного университета. Тел.: (473) 2208-257. E-mail: zsa@main.vsu.ru

Bolotnov D. V. – Post-graduate student of the dept. of digital technologies Voronezh State University. Tel.: 8-903-030-08-42. E-mail: denis1503@gmail.com

Zapryagaev S. A. – Doctor of Physics-math. Sciences, Professor of the dept. of digital technologies Voronezh State University. Tel.: (473) 2208-257. E-mail: zsa@main.vsu.ru