

УДК 004.934.2

С.А. Елистратов, М.А. Косенко, Е.Ю. Костюченко, А.А. Чичерин

Сравнение параметров для выделения вокализованных сегментов и классификации гласных фонем

Исследована применимость параметров для анализа речевого сигнала для выделения вокализованных сегментов и для различия гласных фонем. Проведено сравнение лучшего параметра в случае применения в качестве решающего блока нейронной сети – двухслойного персептрона – при подаче входных данных номера канала с максимальной интенсивностью и корреляции между отсчетами в рамках окна анализа. Показаны большая информативность параметра корреляции между отсчетами по сравнению с номером канала с максимальной интенсивностью для решения задачи поиска вокализованных участков и неприменимость обоих параметров для классификации гласных фонем.

Ключевые слова: анализ речи, нейронная сеть, выделение сегментов.

Анализ речи и его приложение

Анализ речи является актуальной составляющей для решения различных проблем человеческой жизнедеятельности, таких как автоматизация ввода текста, диагностика заболеваний органов речевого тракта, идентификация говорящего или выделение его существенных признаков. Следует отметить, что решение данных задач может зависеть от языка диктора и применение решений для языков, отличных от интересующего – в данном случае русского, не всегда приводит к удовлетворительным результатам при смене языка. В данной статье рассматривается применение нейросетевого подхода применительно к задаче выделения участков речевого сигнала, обладающих определенными признаками. При этом данная задача отличается от сегментации отсутствием необходимости простановки точных границ, важным является учет нужного класса сегментов (например, при анализе гласных звуков – не пропуски их или более общих классов).

Постановка задачи

Имеется речевой сигнал от одного диктора. Требуется провести анализ сигнала с целью сравнения применимости признаков, используемых для выделения вокализованных участков сигнала и разделения гласных фонем. Необходимо оценить параметр, наиболее пригодный для выделения вокализованных участков \maxch – номер канала с наибольшей интенсивностью, применительно к задаче разделения гласных звуков, а также оценить эффективность использования коэффициента корреляции между отсчетами в рамках окна анализа corr при использовании персептрона для задач выделения признаков речевого сигнала.

Выделение параметров речевого сигнала

Рассматриваемые в данной работе параметры \maxch и corr были получены следующим образом:

1) исходный речевой сигнал (частота дискретизации 12 кГц) обрабатывается системой фильтров, учитывающих особенности слуховой системы человека [1, 2] для построения спектра. Верхняя частота анализа была выбрана 2 кГц, количество каналов анализа выбрано 128;

2) из полученного сигнала вырезалось окно для построения параметров для анализа нейронной сетью. Размер окна анализа был выбран 128 отсчетов, что соответствует продолжительности 10,7 мс, что меньше минимальной продолжительности 20 мс;

3) окна анализа нарезаются с шагов в половину окна;

4) на основе полученного окна выделяется параметр звукового сигнала. В данной работе рассматриваются параметры \maxch – нормированный по числу каналов номер канала с максимальной интенсивностью и corr – корреляция между первым и последующими временными отсчетами окна. Таким образом, параметры представляют собой векторы размерностью в 128 чисел.

Оценка информативности анализируемых параметров

Анализ информативности параметров проводился в соответствии с методикой, предложенной в [3, 4].

В качестве критерия используется

$$M'(m_1, m_2) = ((m_1)^s + (k \times m_2)^s)^{1/s} \quad (1)$$

для случая, когда более важной является ошибка второго рода, или

$$M'(m_1, m_2) = ((k \times m_1)^s + (m_2)^s)^{1/s} \quad (2)$$

для случая, когда более важной является ошибка первого рода.

$$M(m_1, m_2) = \begin{cases} M'(m_1, m_2), & M'(m_1, m_2) < 1, \\ 1, & M'(m_1, m_2) \geq 1, \end{cases} \quad (3)$$

где $k \in [1, \infty)$ – коэффициент, учитывающий взаимную значимость ошибок первого и второго рода; $m_1 \in [0, 1)$ – значение ошибки первого рода; $m_2 \in [0, 1)$ – значение ошибки второго рода; $s \in [1, \infty)$ – коэффициент, учитывающий значимость больших ошибок по сравнению с меньшими.

1. Задание параметров критерия k и s – приоритета ошибки первого и второго рода относительно идентифицируемого класса речевого сигнала (вокализованный/невокализованный и а/и), а также учета важности малых ошибок по сравнению с большими – равнозначные ошибки и в 10 раз больший приоритет для обоих классов.

2. Выбор допустимых пределов значения окна интерпретатора и шага между точками, в которых будут рассчитываться значения критерия, – выбор параметров интерпретации выходов нейронной сети.

3. Обучение нейронной сети.

4. Анализ тестовой выборки обученной нейронной сетью.

5. Интерпретация выхода нейронной сети интерпретатором, параметры которого определяются узлами сетки, заданной на шаге 2. Поиск минимального значения критерия в узлах сетки.

6. Если в точке минимума значение критерия $M = 1$, то исследуемый параметр является неинформативным, переход на этап 1.

7. Координаты точки минимума – выбранные параметры интерпретатора.

8. Значение критерия M в точке минимума – оценка информативности исследуемого параметра при помощи критерия, основывающегося на результатах обработки тестовой выборки при помощи нейронной сети.

Итоговые оценки информативности

Для оценки информативности параметров были выбраны настройки интерпретатора выходов нейронной сети. Для этого был найден минимум значения критерия информативности при различных весовых значениях критерия. Пример поверхности значений критерия в зависимости от настроек интерпретатора для параметра maxch и суммы равнозначных ошибок первого и второго рода представлены на рис. 1.

Поиск оценок информативности был проведен для задач разделения гласных [а] и [и] (была выбрана именно эта пара гласных как наиболее отличающаяся по формантной структуре [5]) и для по-

Определение критерия информативности и параметров интерпретатора для параметра maxch , равный приоритет ошибок, $k=1, s=1$

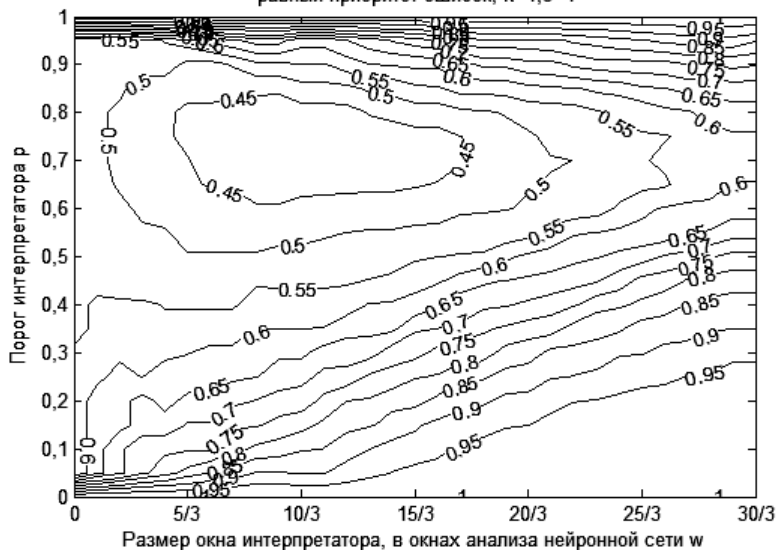


Рис. 1. Линии уровня значений критерия информативности для параметра maxch , $k=1, s=1$ для задачи поиска вокализованных участков

иска вокализованных участков. Оценки информативности для каждого из параметров при различных значениях параметров критерия представлены в таблице.

На основе полученных оценок информативности можно сделать 2 основных заключения:

1) при поиске вокализованных сегментов корреляционный параметр оказался более информативным по сравнению с наиболее информативным параметром maxch из рассмотренных

в [4]. Возможно его применение для поиска вокализованных участков с приоритетом соответствующей ошибки;

2) применительно к задаче классификации гласных оба параметра показали себя неудовлетворительно – сумма ошибок первого и второго рода более 0,5 при отсутствии возможности анализа сигнала с отсутствием ошибок одного из родов (наилучшее суммарное значение 0,9057 близко к оценке абсолютно неинформативного параметра 1).

Оценки информативности при различных значениях параметров критерия

Значения параметров	Корреляция при вокализации	Корреляция при гласных [а], [и]	Максимальный канал при вокализации	Максимальный канал при гласных [а], [и]
$k = 1, s = 1$	0,3268	0,6104	0,4054	0,5138
$k=10$, приоритет вокализованных или А, $s = 1$	0,4723	0,9057	0,6385	1
$k=10$, приоритет невокализованных или И, $s = 1$	1	1	1	1
$k = 1, s = 2$	0,2367	0,4435	0,2896	0,3907
$k = 10$, приоритет вокализованных или А, $s = 2$	0,3813	0,9057	0,6578	1
$k=10$, приоритет невокализованных или И, $s = 2$	1	1	0,7327	1

Заключение

В результате проведенного исследования показана большая информативность корреляционного параметра согг по сравнению с номером максимального по интенсивности канала. Полученные оценки информативности говорят о возможности применения корреляционного параметра для поиска вокализованных участков с применением нейросетевого подхода, что подтверждают данные, полученные в [6, 7] для ненейросетевого подхода.

Исходя из низкой информативности параметров, можно прийти к заключению о необходимости учета внутренней структуры гласных звуков (явный учет двух формант, выделяемых в разных частотных диапазонах). Учет одного максимального канала, соответствующего, как правило, первой форманте, оказывается неинформативным. Кроме того, показана неэффективность использования коэффициента корреляции для задачи классификации гласных, что говорит об их одинаковой внутренней структуре по времени – высокая степень стационарности подтверждается постоянным высоким значением коэффициента корреляции для всех гласных звуков, что не позволяет использовать его для их классификации.

Литература

1. Бондаренко В.П. Выделение особенностей структуры речевого сигнала / В.П. Бондаренко, В.П. Коцубинский, Р.В. Мещеряков // Сб. тр. XII сессии Российского акустического общества (Москва). – 2003. – Т. 3. – С. 63–66.
2. Бондаренко В.П. Модель периферии слуховой системы человека / В.П. Бондаренко, В.М. Разин // VI Всесоюзный семинар «Автоматическое распознавание слуховых образов» (АРСО-VI). – Таллин, 1972. – С. 26–29.
3. Костюченко Е.Ю. Обработка естественной информации на основе аппарата нейронных сетей // Докл. Том. гос. ун-та систем управления и радиоэлектроники. – 2009. – № 1(19), ч. 2. – С. 54–56.
4. Костюченко Е.Ю. Критерии информативности при обработке биометрических сигналов при помощи нейронных сетей / Е.Ю. Костюченко, Р.В. Мещеряков, А.Ю. Крайнов // Докл. Том. гос. ун-та систем управления и радиоэлектроники. – 2010. – № 1(21), ч. 1. – С. 118–120.
5. Бондарко Л.В. Основы общей фонетики / Л.В. Бондарко, Л.А. Вербицкая, М.В. Гордина. – 4-е изд. – СПб.: Академия, 2004. – 160 с.

6. Конев А.А. Сегментация речевого сигнала / А.А. Конев, А.А. Пономарёв // Сб. тр. XVI сессии Российского акустического общества.– М.: ГЕОС, 2005. – Т. 3. – С. 44–47.

7. Мещеряков Р.В. Система оценки качества передаваемой речи // Доклады ТУСУРа. – 2010. – № 2(22). – С. 324–329.

Елистратов Сергей Александрович

Студент 4 курса, каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) ТУСУРа

Тел.: 8 (383-2) 41-34-26

Эл. почта: mushroom3p@gmail.com

Косенко Михаил Анатольевич

Студент 4-го курса каф. КИБЭВС ТУСУРа

Тел.: 8 (383-2) 41-34-26

Эл. почта: mushroom3p@gmail.com

Костюченко Евгений Юрьевич

Канд. техн. наук, доцент каф. КИБЭВС ТУСУРа

Тел.: 8 (383-2) 41-34-26

Эл. почта: key@keva.tusur.ru

Чичерин Андрей Александрович

Студент 4-го курса каф. КИБЭВС ТУСУРа

Тел.: 8 (383-2) 41-34-26

Эл. почта: mushroom3p@gmail.com

Elistratov S.A., Kosenko M.A., Kostyuchenko E.Y., Chicherin A.A.

Comparison of parameters for the selection and classification of voiced phonemes segments

The applicability of the parameters for the speech signal analysis to extract voiced segments, and to distinguish phonemes was investigated. The best option in case of application as a crucial block of a neural network – a two-layer perceptron – feeding the input channel number with the maximum intensity and the correlation between samples within the analysis window was compared. More informative parameter of the correlation between samples in comparison with the number of channels with a maximum intensity for the solution of the problem of finding sites and voiced the inapplicability of both parameters for the classification of phonemes was shown.

Keywords: speech analysis, neural network, the selection of segments.
