

УДК 534.4: 621.391

О.Н. Карпов

Днепропетровский национальный университет имени Олеся Гончара

РАСПРЕДЕЛЁННЫЙ АЛГОРИТМ ОЦЕНКИ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ И ИНДИВИДУАЛЬНОСТИ ЧЕЛОВЕКА ПО ЕГО РЕЧИ

Розглянуто проблему визначення функціонального говорячого за набором часових функцій мовного сигналу за допомогою розподіленого алгоритму визначення відхилень у частотно-часовій області або відхилень у часовій області. Наведено алгоритм оцінки функціонального стану та індивідуальності людини за її мовою.

Ключові слова: *функціональний стан, індивідуальність людини, мова, мовні сигнали, темпоральні характеристики, частота основного тону, тембр, розподілений алгоритм*

Рассматривается проблема определения функционального состояния говорящего по набору временных функций речевого сигнала при помощи распределённого алгоритма определения отклонений в частотно-временной области или отклонений во временной области. Приведен алгоритм оценки функционального состояния и индивидуальности человека по его речи.

Ключевые слова: *функциональное состояние, индивидуальность человека, речь, речевые сигналы, темпоральные характеристики, частота основного тона, тембр, распределенный алгоритм.*

We consider the problem of a speaker functional state determining on a set of time functions of the speech signal using a distributed algorithm for determining deviations in frequency-time domain, or deviations in the time domain. An algorithm for assessment of functional status and individuality in his speech.

Key words: *functional state, human individuality, speech, speech signals, temporal characteristics, main pitch frequency, timbre, distributed algorithm*

Вступление. Существуют разные методы исследования речевых сигналов, отдельные для каждой из проблем:

- распознавание индивидуальности говорящего;
- распознавание функционального человека по его речи.

Каждый метод исходит из независимости и уникальности проблемы, которую он решает. В частности, функциональное состояние (ФС) можно определять по динамике частоты основного

тона (ОТ), но для людей с дефектами речи, например шепотная речь, этот метод вообще не применим. Темпоральные характеристики (ТХ) предполагают длительный интервал анализа. Имеются и другие характеристики оценки ФС, но они чаще всего являются индивидуально-зависимыми, также как и оценки по ОТ.

Что же касается оценки индивидуальности, то один из параметров – это средняя частота основного тона или её индивидуальная динамика на заданном словаре слов, также средний показатель ТХ, например, средний темп речи, усреднённый показатель длительности пауз раздумья. Эти характеристики зависят от ФС и предметной области, в которой ведётся разговор. Для оценки индивидуальности есть другие характеристики, а именно то, что называют «тембр». Это, с одной стороны, обертоны, связанные с функционированием голосовой мышцы: продольные параметрические колебания и функциональные модуляции и присутствуют во всём частотном диапазоне спектра тонального речевого сигнала, а также индивидуальных резонансных областей, связанных с состоянием зубов, миндалин и других особенностей (шепелявость), которые наложены на речевой сигнал независимо от генератора речеобразования.

Таким образом, существует достаточно большая зависимость между параметрами индивидуальности и ФС. Разделение их трудно разрешимая задача, хотя в спектральной области можно разделить компоненты, частично соответствующие сущностям смыслового описания, индивидуальности и ФС. При всём том речеобразующий тракт в процессе функционирования находится в динамике и средние показатели отражают некоторые условные понятия ФС и индивидуальности.

В целом, оценка и того и другого – это функция времени от совокупности частотных, амплитудных и временных параметров. Допустим, есть набор временных функций речевого сигнала $\{s_{rl}(\omega t)\}$, для них могут быть вычислены спектрально-временные функции вида $\{S_{rl}(\Omega, \omega, T_i)\}$, где T_i – интервал анализа, r – номер реализации ФС, l – номер лица.

Общая задача: определить лицо l и его функциональное состояние r . В итоге образуется распределённый алгоритм определения $V_{rl}(\Omega, \omega, T_i)$ – отклонение в частотно-временной области или $\varphi_{rl}(\omega, t)$ – отклонения во временной области.

Задача сопоставления речевого высказывания некоторого лица, находящегося в каком-то состоянии $z(\omega, t)$, $Z(\Omega, \omega, T_i)$ это определение меры различия $z(\omega, t)$ и $s_{r_l}(\omega, t)$, $Z(\Omega, \omega, T_i)$ и $S_{r_l}(\Omega, \omega, T_i)$. Т. е. необходимо реализовать некоторую операцию (рис.1)

$$Z(\Omega, \omega, T_i) = V_{r_l}(\Omega, \omega, T_i) \# S_{r_l}(\Omega, \omega, T_i). \quad (1)$$

$$z(\omega, t) = \varphi_{r_l}(\omega, t) \# s_{r_l}(\omega, t). \quad (2)$$

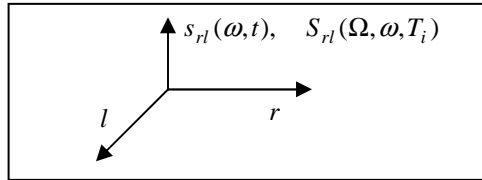


Рис.1. Сопоставление речевого высказывания

Суть операции «#» – это метод решения задачи определения $V_{r_l}(\Omega, \omega, T_i)$ или $\varphi_{r_l}(\omega, t)$.

1. Самая простая операция – это разность в частотной области

$$Z(\Omega, \omega, T_i) - S_{r_l}(\Omega, \omega, T_i) = V_{r_l}(\Omega, \omega, T_i). \quad (3)$$

Во временной области операция

$$z(\omega, t) - s_{r_l}(\omega, t) = \varphi_{r_l}(\omega, t) \quad (4)$$

не реализуема из-за разных фаз гармоник речевых сигналов, но реализуема (3) в частотной области. Для данной операции проблема заключается в разных длительностях реализаций $Z(\Omega, \omega, T_i)$ и $S_{r_l}(\Omega, \omega, T_i)$, что делает их несопоставимыми напрямую. Выравнивание длин реализуется методом динамического программирования.

2. Другая операция – это построение фильтра в предположении, что $Z(\Omega, \omega, T_i)$ может быть получено из $S_{r_l}(\Omega, \omega, T_i)$ путём умножения вида

$$Z(\Omega, \omega, T_i) = V_{r_l}(\Omega, \omega, T_i) S_{r_l}(\Omega, \omega, T_i) \quad (5)$$

$$z(\omega, t) = \varphi_{r_l}(\omega, t) \# s_{r_l}(\omega, t). \quad (6)$$

Операция (5) в частотной области в общем случае реализуема, но во временной области- это свёртка

$$z(\omega, t) = \int_0^{\infty} s_{r_l}(\omega, \tau) \varphi_{r_l}(\omega, t - \tau) d\tau. \quad (7)$$

Искомая задача – это определение $\varphi_{r_l}(\omega, t)$, например, как решение уравнения Винера-Хопфа.

$$z(\omega, t) = \int_0^{\infty} c_{r_l}(\omega, \tau) \varphi_{r_l}(\omega, t - \tau) d\tau,$$

где $c_{r_l}(\omega, t)$ – взаимокорреляционная функция между $z(\omega, t)$ и $s_{r_l}(\omega, t)$, которая может быть вычислена.

Выражения (3), (5) являются мерами близости между частотно-временными функциями $Z(\Omega, \omega, T_i)$ и $S_{r_l}(\Omega, \omega, T_i)$.

В частотной области решение задачи выглядит как (5), где $V_{r_l}(\Omega, \omega, T_i)$ некоторая передаточная функция системы, преобразующей $S_{r_l}(\Omega, \omega, T_i)$ в $Z(\Omega, \omega, T_i)$. Таким образом, задача определения характеристик индивидуальности и эмоционального состояния можно свести к задаче наилучшего приближения $Z(\Omega, \omega, T_i)$ к $S_{r_l}(\Omega, \omega, T_i)$, подбирая $V_{r_l}(\Omega, \omega, T_i)$.

Есть решение задачи описания спектрально-временного представления в классах колоколообразных [1] и колебательных функций [2; 3] Задача построения функции $V_{r_l}(\Omega, \omega, T_i)$ имеет как общее, так и частные решения.

1. Общее решение – определить личность, для найденной личности определить функциональное состояние (ФС).
2. Определить ФС для известной личности.
3. Определить личность.
4. Определить n личностей с заданным ФС.
5. Определить ФС для m личностей.

Возможны и другие задачи и вид их решения может быть представлен в соответствии с теоремой Колмогорова о представлении функций многих переменных как суперпозиция функций меньшего числа переменных или функций одной переменной

Исходное спектрально-временное представление рассматривается в прямоугольной области $R = [\omega_e, \omega_f] \times [T_c, T_d]$, в области R задана таблично спектрально-временная функция $S(\omega_k, T_i)$, где ω_k – дискретно заданная частота, t_l – дискретно заданное время:

$$\omega_e = \omega_0 < \omega_1 < \dots < \omega_m = \omega_f$$

$$T_c = T_0 < T_1 < \dots < T_n = T_d.$$

Самый простой вид такого описания функции $V_{r_l}(\omega, T_i)$ как произведение одномерных функций, представленных в виде

полиномов по соответствующему аргументу и соответствующей степени

$$V_{rl}(\omega, T_i) = \sum_{j=0}^m a_j \omega^j \sum_{p=0}^n b_p T_i^p.$$

Выводы. В итоге

$$Z(\Omega, \omega, T_i) = S(\Omega, \omega, T_i) \sum_{j=0}^m a_j \omega^j \sum_{p=0}^n b_p T_i^p$$

В функции $V_{rl}(\omega, T_i)$ присутствуют частоты Ω , как низкочастотные колебания поверхности, которую описывает $V_{rl}(\omega, T_i)$. Решение задачи определения параметров a_j, b_p решается методом наименьших квадратов в виде

$$\sigma^2 = [Z(\Omega, \omega, T_i) - S(\Omega, \omega, T_i) \sum_{j=0}^m a_j \omega^j \sum_{p=0}^n b_p T_i^p]^2$$

при этом задача решается методом покоординатного спуска, при котором параметры a_j, b_p меняются последовательно по правилу

$$a_i \pm \Delta a_i, \quad b_p \pm \Delta b_p$$

минимизируя σ_g^2 . Степени полиномов последовательно увеличиваются, уменьшая минимумы σ_g^2 , где g - номер итерации повышения степени.

Библиографические ссылки

1. Карпов О.Н. Компьютерные технологии распознавания речевых сигналов. Монография. / О.Н. Карпов, А.Г. Габович, Б.Г. Марченко, В.А. Хорошко, Л.Н. Щербак. –К., 2005. – 138 с.
2. Карпов О.Н. Описание спектрально-временного представления речевых сигналов в классе производных функций Гаусса второго порядка. / О.Н. Карпов, Г.В. Зирнеева // Питання прикладної математики та мат. моделювання. – 2004.–С.88–97
3. Карпов О.Н. Сравнение свойств колебательных функций в задаче анализа спектров речевых сигналов. / О.Н. Карпов, Г.В. Зирнеева // Актуальні проблеми автоматизації та інформаційних технологій. – Т.8.–2004. С. 14–19.

Надійшла до редколегії 20.08.10

УДК 534.4: 621.391

О.І. Лучинкіна, П.Ю. Аксьоненко, О.М. Карпов

Дніпропетровський національний університет імені Олеся Гончара

ТРАНСКРИПТОР ЯК ОДИН З ВУЗЛІВ ЗАГАЛЬНОЇ ЕКСПЕРТНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ МОВИ

Проводиться аналіз структури транскриптору та пропонується схема побудови транскриптору як експертної системи.

Ключові слова: експертна система, надійність, розпізнавання мови, транскриптор

Проводится анализ структуры транскриптора и предлагается схема построения транскриптора как экспертной системы.

Ключевые слова: экспертная система, надежность, распознавание речи, транскриптор

The analysis of the structure of transcripтор holds and the scheme of transcripтор as an expert system is proposed.

Keywords: expert system, reliability, speech recognition, transcripтор

Постановка проблеми. На сьогоднішній день, під поняттям «розпізнавання мови» приховується ціла сфера наукової та інженерної діяльності. Загалом, кожна задача розпізнавання мовлення зводиться до того, щоб виділити, класифікувати, і, відповідним чином, відреагувати на людську мову з вхідного звукового потоку. Це може бути і виконання певної дії на команду людини, і виділення певного слова-маркера з великого масиву телефонних переговорів, і системи для голосового вводу тексту.

Аналіз досліджень та постановка задачі. Згідно з опублікованими на сьогоднішній день даними надійність розпізнавання мови для систем, які працюють з однослівними командами, сягає 99,5 %, командами, що складаються з двох слів – 97,5 %, з трьох слів – 92,5 %, і чотирьох слів – 91,9 %. Але при проектуванні системи розпізнавання злитного мовлення перед розробниками постає проблема розпізнавання не лише мовного сигналу, але й лінгвістичного змісту. Тобто необхідно розробити методи підвищення надійності розпізнавання ланцюжка з великої кількості слів.

Системи розпізнавання мовлення класифікують за такими ознаками [1; 2; 3; 4]:

- тип речі (злитий або ізольований);

© О.І. Лучинкіна, П.Ю. Аксьоненко, О.М. Карпов, 2010