АЛГОРИТМ ФОРМИРОВАНИЯ ПАРАМЕТРИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ДЛЯ ОЦЕНКИ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ И ИНДИВИДУАЛЬНОСТИ ЧЕЛОВЕКА ПО ЕГО РЕЧИ

Карпов О.Н.,

Глушак К.Н.

В статье приведены некоторые подходы для формирования параметрического представления функционального состояния человека. А также высказывается гипотеза о дальнейшем способе их применения для решения задачи оценки индивидуальности человека.

• функциональное состояние • речь • распределение частот • частотно-временные функции • модифицированный локон Анъези

In the article are some of the approaches to generate a parametric representation of the functional state of the person. As well as it is hypothesised about the further way of their application for the solution of the problem of estimating a person's individuality.

• functional state • speech • frequency distribution • frequency-time function • modified curl Anezi

Введение

В области речевых технологий существует большое количество разнообразных задач. Но лишь в рамках некоторых из них проводится попытка анализа тембра человека. Эти задачи:

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

Большое число подходов к решению указанных задач, пытаются их решить независимо друг от друга. В частности, функциональное состояние (ФС) можно определять по динамике частоты основного тона (ОТ), но данный подход

далеко не универсален, он не будет работать для людей с дефектами речи, например, для шёпотной речи этот метод вообще неприменим. Темпоральные характеристики (ТХ), в свою очередь, требуют длительного интервала для анализа. Существуют также и другие характеристики оценки ФС, но они чаще всего являются индивидуально-зависимыми, так же как и оценки по ОТ.

Что же касается оценки индивидуальности, то здесь также можно выделить большое количество различных параметров, одним из которых может быть частота основного тона или её индивидуальная динамика на заданном словаре слов, или 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 обертоны, связанные с функционированием голосовой мышцы: продольные параметрические колебания, наложенные на спектры речевого сигнала как амплитудные, частотные и функциональные модуляции. Они присутствуют во всём частотном диапазоне спектра тонального речевого сигнала.

Таким образом, можно предположить наличие довольно большой корреляции между параметрами индивидуальности и ФС. Разделение их трудноразрешимая задача, хотя в спектральной области можно разделить компоненты, частично соответствующие сущностям смыслового описания, индивидуальности и ФС.

Вообще говоря, оценка и того и другого – это функция времени от совокупности частотных, амплитудных и временных параметров. Допустим, есть набор временных функций речевого сигнала $\{s_{rl}(\omega,t)\}$ (где ω – диапазон наблюдаемых частот: 100–11000Гц), 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, Ω – темп речи.

Тогда задачу можно сформулировать следующим образом: определить лицо l и его функциональное состояние r по заданному сигналу. В итоге образуется 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 $s_{rl}(\omega,t)$, $Z(\Omega,\omega,T_i)$ и $S_{rl}(\Omega,\omega,T_i)$. То есть необходимо реализовать некоторую операцию # (рис. 1), такую что:

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

$$z(\omega,t)= \varphi_{rl}(\omega,t)\#s_{rl}(\omega,t) \quad (1)$$

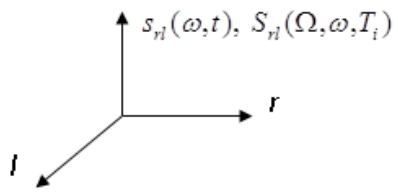

рис. 1.

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

1. Самая простая операция – это разность в частотной области.

$$Z(\Omega,\omega,T_i) - S_{rl}(\Omega,\omega,T_i) = V_{rl}(\Omega,\omega,T_i). \quad (4)$$

Во временной области операция

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

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 $S_{rl}(\Omega,\omega,T_i)$, что делает их несопоставимыми напрямую. Сопоставление длин реализуется методом динамического программирования.

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

$$Z(\Omega,\omega,T_i) = V_{rl}(\Omega,\omega,T_i)\, S_{rl}(\Omega,\omega,T_i), \quad (6)$$

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

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

$$z(\omega,t) = \int_0^{\infty} s_{rl}(\omega,\tau)\phi_{rl}(\omega,t-\tau)d\tau \quad (8)$$

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

$$z(\omega t) = \int_0^{\infty} c_{rl}(\omega\tau)\varphi_{rl}(\omega t - \tau)d\tau \quad (8)$$

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

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

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

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:

$$\omega_e = \omega_0 < \omega_1 < \ldots < \omega_m = \omega_f ,$$
$$T_c = T_0 < T_1 < \ldots < T_n = T_d .$$

Самый простой вид такого описания функции $srl(\omega,t)$ как произведение одномерных функций, представленных в виде полиномов по соответствующему аргументу и соответствующей степени

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

В итоге

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

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

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

при этом задача решается методом покоординатного спуска, при котором параметры a_j, b_p меняются последовательно по правилу $a_j\pm\Delta a_j$, $b_p\pm\Delta b_p$, минимизируя σ_g^2.

На практике же задачу поиска передаточной функции имеет смысл рассматривать в пространстве колоколообразных функций. А именно с использованием модифицированного локона Анъези.

Также в области R определены классы функций $\{W_i(\omega_k)\}$, $\{h_i(t_l)\}$ со следующими свойствами:

- форма функции должна определяться некоторыми параметрами;
- функции асимптотически приближаются к области плоскости R в произвольном направлении от максимума;
- ветви функции должны монотонно убывать, если максимум колокола находится вне области R.

Функция $S(\omega_k,t_l)$ в области R имеет произвольное количество всплесков спектральной энергии, размещенных произвольным образом в заданной области. Необходимо функцию $S(\omega_k,t_l)$ сегментировать и аппроксимировать в классе функций $\{W_i(\omega_k)\}$. Для этого необходимо определить параметры всплесков функции $S(\omega_k,t_l)$, как параметры функции $\{h_i(t_l)\}$. Для решения этой задачи в области R строится сетка MxN, где M – количество сегментов, N – количество полос по частоте. Клетки сетки определяют начальные значения параметров по частоте и времени.

Спектрально-временная составляющая ищется в классе функций модифицированного локона Анъези

$$(\omega^2+r^2)W(\omega)-a^3=0,\quad W(\dot{u}_k)=\frac{a_{(i)}^3}{c_{(i)}^2+(\dot{u}_k-\hat{a}_{(i)})^2},$$

$$h(t_l)=\frac{b_{(j)}^3}{d_{(j)}^2+(t_l-T_{(j)})^2},$$

где r, c, d определяют форму локона в виде

$$Z_{(i,j)}(k,q_{(i,j)},a_{(i,j)},c_{(i,j)},\beta_{(i,j)},b_{(i,j)},d_{(i,j)},T_{(i,j)},\omega_k,t_l) = S_{g(j)}(\omega_k)W_{(i)}(\omega_k)h_{(j)}(t_l) =$$
$$= \frac{q_{(i)}^{\ln k}}{k}\frac{a_{(i)}^3}{c_{(i)}^2+(\omega_k-\beta_{(i)})^2}\frac{b_{(j)}^3}{d_{(j)}^2+(t_l-T_{(j)})^2}.$$

По методу наименьших квадратов определяются параметры k, $q_{(ij)}$, $a_{(ij)}$, $c_{(ij)}$, $\beta_{(ij)}$, $b_{(ij)}$, $d_{(ij)}$, $T_{(ij)}$ для

На первом шаге параметры определяются для:

$$\sigma_{(i,j)}^2 = \sum_{\omega_k}\sum_{t_l}[S_{i-1,j-1}(\omega_k,t_l) - Z_{(i,j)}(k,q_{(i,j)},a_{(i,j)},c_{(i,j)},\beta_{(i,j)},b_{(i,j)},d_{(i,j)},T_{(i,j)},\omega_k,t_l)]^2$$

Параметры $q_{(ij)}$, $a_{(ij)}$, $c_{(ij)}$, $\beta_{(ij)}$, $b_{(ij)}$, $d_{(ij)}$, $T_{(ij)}$ определяются на последующих шагах.

$$\sigma_{(1,1)}^2 = \sum_{\omega_k}\sum_{t_l}[S(\omega_k,t_l) - Z_{(1,1)}(q_{(1,1)},a_{(1,1)},c_{(1,1)},\beta_{(1,1)},b_{(1,1)},d_{(1,1)},T_{(1,1)},\omega_k,t_l)]^2,$$

$$\sigma = S(\omega_k,t_l) - \frac{q_{(1)}^{\ln k}}{k}\frac{a_{(1)}^3}{c_{(1)}^2+(\omega_k-\beta_{(1)})^2}\frac{b_{(1)}^3}{d_{(1)}^2+(t_l-T_{(1)})^2}.$$

Рассмотрим задачу сопоставления двух сигналов на практике. Сравнение будем проводить на двух реализациях украинского слова «чотири».

На рис. 2 мы видим представление сигнала во временной области. Тут же можно наглядно убедиться в нереализуемости операции сопоставления этих двух сигналов в силу разности их продолжительности. После перехода к спектрально-временному представлению мы можем применить к нему модифицированный локон Анъези.

Таким образом мы можем посмотреть на восстановленное спектрально-временное представление после аппроксимации его с помощью локона Анъези.

Следующим логичным шагом является сравнение полученных величин. Для начала рассмотрим операцию разности восстановленных спектрально-временных представлений.

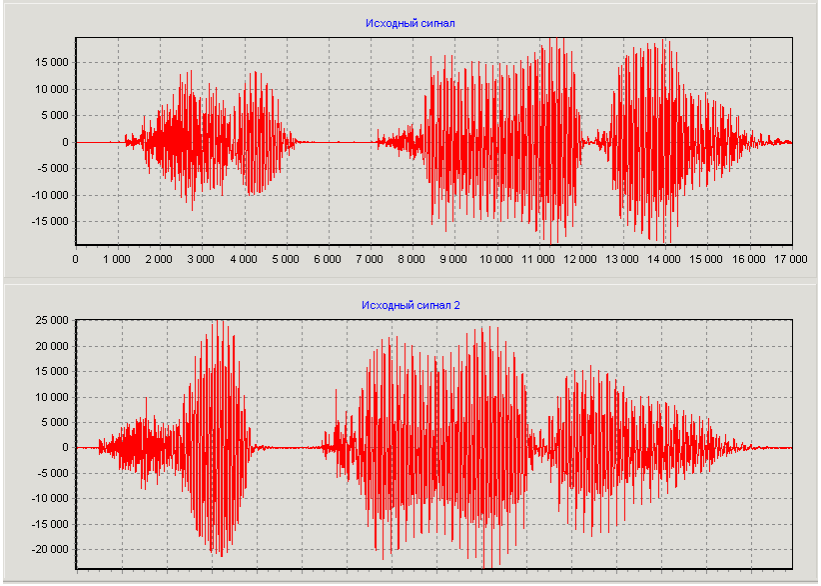

Рис. 2. Две реализации слова «чотири»

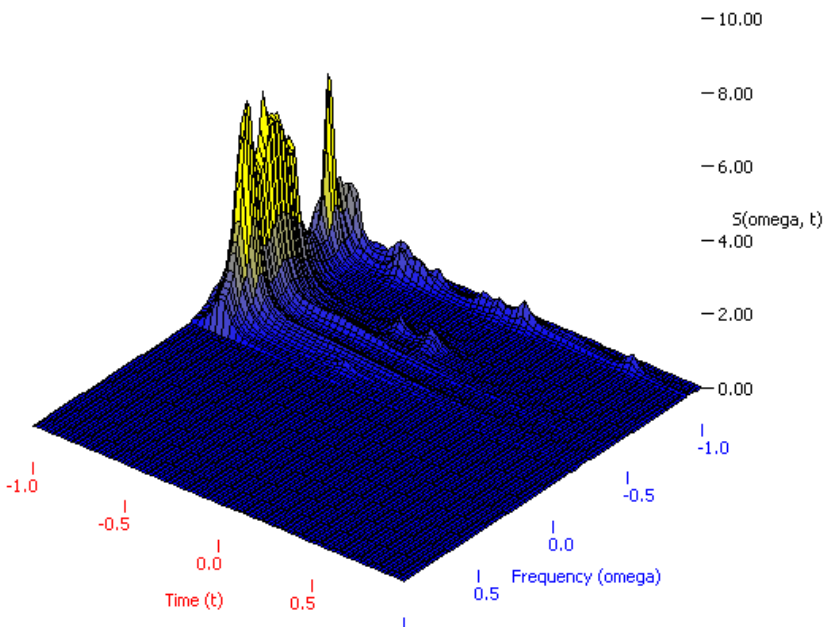

Рис. 3. Восстановленное спектрально-временное представление первой реализации слова «чотири» после аппроксимации локоном Анъези

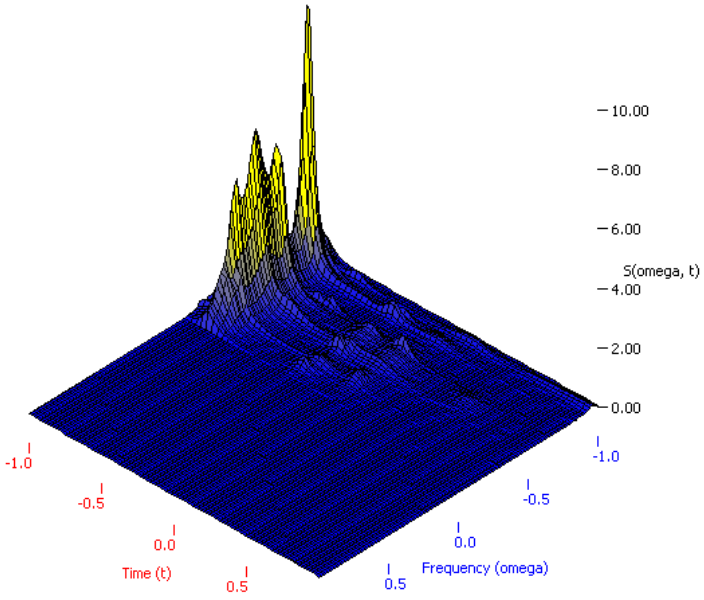

Рис. 4. Восстановленное спектрально-временное представление второй реализации слова «чотири» после аппроксимации локоном Анъези

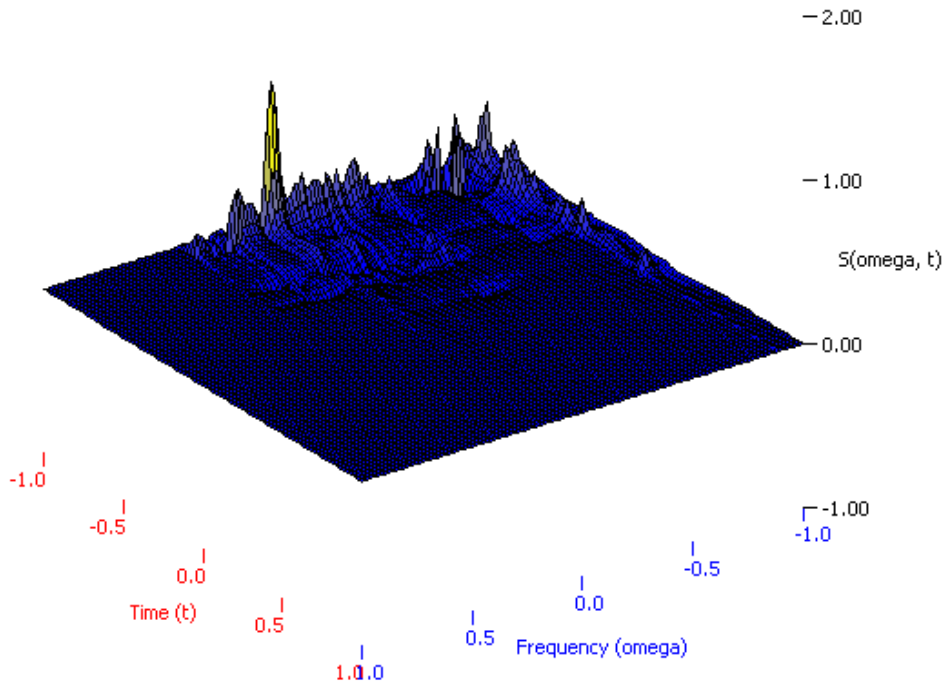

Рис. 5. Разность восстановленных спектрально-временных представлений слова «чотири» (масштаб увеличен)

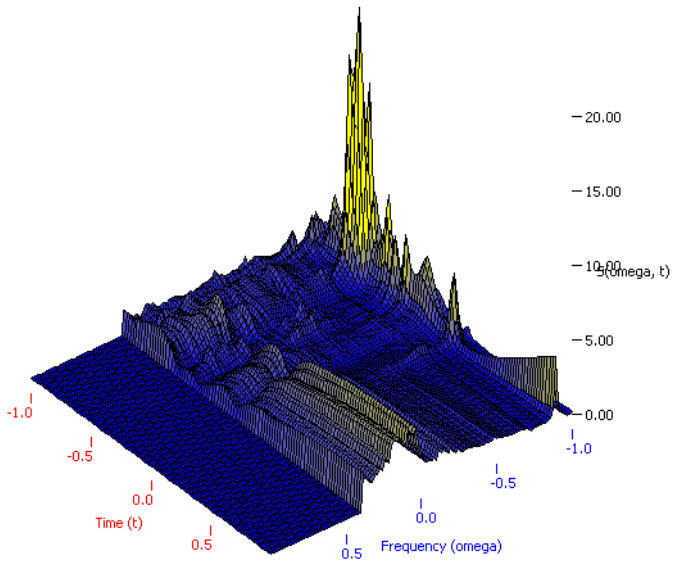

Рис. 6. Отношение восстановленных спектрально-временных представлений слова «чотири»

Как видно из рис. 5, разница между двумя реализациями сколь угодно мала. Картина выглядит иным образом для операции деления.

Как результат мы получили искомую передаточную функцию $V_{ri}(\omega,T_i)$, которая идентифицирует человека уникальным образом. То есть выделили параметрическое представление для конкретного человека, который выражает определённую эмоцию.

Для получения окончательного вывода об индивидуальности человека или его функциональном состоянии можно применить алгоритмы динамического программирования или решить задачу уменьшения размерности параметрического пространства, после чего можно использовать нейронные сети разной архитектуры.

Литература

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

Сведения об авторах:

Карпов Олег Николаевич,
доктор технических наук, профессор кафедры математического обеспечения ЭВМ факультета прикладной математики Днепропетровского национального университета им. О. Гончара, изобретатель. Награждён почётным званием «Изобретатель СССР», имеет 11 авторских свидетельств, его научные работы получили 9 серебряных и бронзовых медалей ВДНХ СССР. Область научных интересов: теория и компьютерные технологии по решению проблемы распознавания речи.

Глушак Константин,
аспирант Днепропетровского национального университета им. О. Гончара, PHP-программист компании «Арчер Софтвер». Научные интересы лежат в области искусственного интеллекта и распознавания речи.