

*О.Н. Карпов, д.т.н., проф., К.М. Глушак, аспирант,
(Днепропетровский национальный университет, Украина)*

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ФОРМИРОВАНИЯ ПАРАМЕТРОВ И ОЦЕНКИ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА ПО ЕГО РЕЧИ

В работе приведен вариант технологии определения функционального состояния человека по его речи. Также делаются предположения относительно определения специфических признаков речи диктора

Задача оценки функционального состояния человека по его речи является сложной комплексной задачей, которая содержит физически разнородные параметры. Решение данной задачи связано с оценкой взаимодействия психической реакции человека и моторной реакцией всех частей речеобразующего тракта: резонансной системы, созданной положением языка, губами, носовой полостью, голосовой мышцей и дыхательной системой. Соответственно, параметры речеобразования формируются как на психическом, так и на психоакустическом уровнях.

Функции индивидуальных или функциональных отклонений от эталона диктора в частотной области могут быть представлены как некоторые множества $W(\omega)$ для спектральных огибающих $S_s(\omega)$:

$$S(\omega) = W(\omega)S_s(\omega) \quad (1)$$

или как спектры сигналов разностей $S_p(\omega)$, которые вычитаются или добавляются к эталонным спектрам:

$$S_p(\omega) = S(\omega) - S_s(\omega) \quad (2)$$

(функция амплитудной индивидуальности голоса АИГ также может входить в сигнал голосового источника как произведение или сумма);

Акустический сигнал в речеобразующем тракте модулируется сигналом динамики психофизического состояния, которые представляют собой паузы раздумья или дыхательные паузы. Речевой сигнал может быть на фоне шума, который имеет сложный состав.

Таким образом, речевой сигнал имеет вид сложной многомерной функции $f(\Omega_1, \dots, \Omega_2, \omega, t)$, где $\Omega_1, \dots, \Omega_2$ – частоты компонентов речевого сигнала с неизвестными операциями их взаимодействия, ω – частота спектральных параметров, t – текущее время.

Компоненты нестационарности речевого сигнала. Изменчивость речевого сигнала обусловлена множеством факторов. Из них основными являются следующие [1]:

а) речевой сигнал - случайный нестационарный временной процесс; б) вариативность речи, обусловлена трансформацией (редукцией), транскрипцией; в) индивидуальность; г) диалект; д) состояние диктора и изменчивость состояния.

Каждый из этих факторов вызывает своё отклонение параметров реальной речи от некоторого идеала параметров правильной речи $S_r(\omega, t)$ (right – правильный).

А. Нестационарность речевого сигнала вызванная несколькими факторами, имеющими разную физическую природу: амплитудная нестационарность $S_a(\omega, t)$; частотная нестационарность $S_f(\omega, t)$; временная нестационарность $S_t(\omega, t)$.

Амплитудная нестационарность – это изменчивость громкости (интенсивности) речевого высказывания и динамика интенсивности сигнала как функции времени, обусловленной дикцией, выражением, ударением.

Частотная нестационарность $S_f(\omega, t)$:

- 1) это перераспределение энергии спектра вдоль оси частот в данном высказывании;
- 2) изменение перераспределения энергии для разных вариантов данного высказывания;
- 3) изменение плотности энергии для конкретной полосы частот (ширина и высота форманты).

Временная нестационарность $S_t(\omega, t)$ – это изменение общей длительности высказывания и составляющих его фонем.

Б. Вариативность речи обусловлена как нестационарностью, так и изменчивостью структуры речи $S_v(\omega, t)$:

– переход одного звука в другой (редукция) в соединении фонем (например, “глухой - звонкий”, “звонкий - глухой”);

– разной транскрипцией произнесения слова (например, “50 - пидисят, пядесят”, “60 - шиздисят, шиисят”);

– трансформация гласных фонем “ударных - неударных”.

В. Индивидуальность $S_p(\omega, t)$ – это многофакторное понятие. Оно характеризуется как структурой речевого сигнала, обусловленной индивидуальным состоянием речеобразующего тракта, как-то: частота основного тона (ОТ), структура резонансных областей, артикуляция губ; так и движением мышц индивидуума дефектами той или другой части системы речеобразования, предназначенной для формирования и излучением акустического сигнала.

Г. Диалект $S_d(\omega, t)$ – это вариативность структуры, ударения, фонемного состава и длительности речевого сигнала.

Д. Изменчивость состояния диктора $S_s(\omega, t)$ – это динамика ОТ (сдвиг и \pm изменение диапазона частоты), изменение структуры резонансных областей речеобразующего тракта, которое меняет временную, частотную и амплитудную структуру речевого сигнала.

При построении алгоритмов решения задач оценки функционального состояния и идентификации личности приведенные факторы требуют разделения общей задачи на частные, учитывающие специфику каждого фактора и способ его взаимодействия с другими факторами

Один из алгоритмов решения задачи. Для организации решения входная функция $s(t)$ дискретизируется с частотой f_d , формируя сигнал $s(t_j)$, который затем разбивается на интервалы анализа T_{lN} с количеством отсчетов N на интервале при $l = \overline{0, n-1}$, n – количество интервалов T_{lN} , представляя входной сигнал как двухмерную функцию времени и частоты вида $S(\omega_k, T_l)$, при $j = \overline{lN, (l+1)N}$, $S(\omega_k, T_{lN})$ – спектральный срез.

Решение задачи оценки функционального состояния диктора по его речи может быть реализовано в виде некоторой иерархии уровней, которые отличаются привлекаемыми для решения параметрами и целями.

Нулевой уровень - распознавание слов речи заданного словаря (изолированных слов или фраз), для данного варианта оценки параметров функционального состояния не ставится задача распознавания слитной речи произвольной длительности

Первый уровень - сопоставление параметров динамики основного тона и темпоральных характеристик речи.

Второй уровень – сопоставление спектрально-временных характеристик речевых высказываний слов или фраз).

Для решения задачи оценки функционального состояния говорящего первого уровня формируются две группы параметров [2, 3]:

- сигнал с частотой основного тона для тонального сигнала;

– темпоральные характеристики как количество пауз (КП) в речи, длительность речи (ДР) на интервале анализа, длительность пауз на интервале анализа, темп артикулирования (ТА) – отношение длительность пауз до длительность речи.

В данной работе приведены результаты обработки сигнала основного тона.

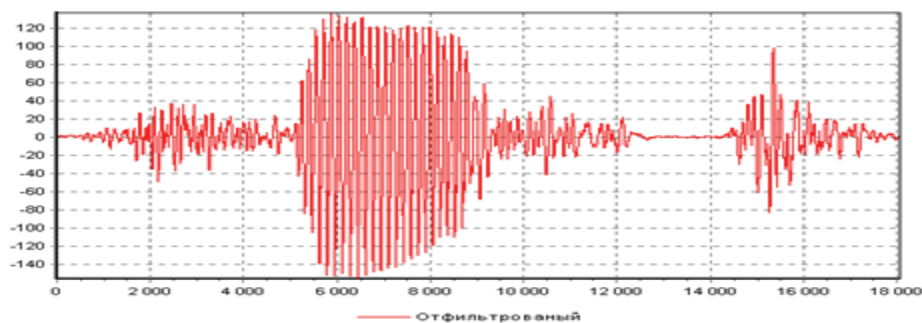


Рис. 1 Сигнал частоты основного тона слова “шесть”.

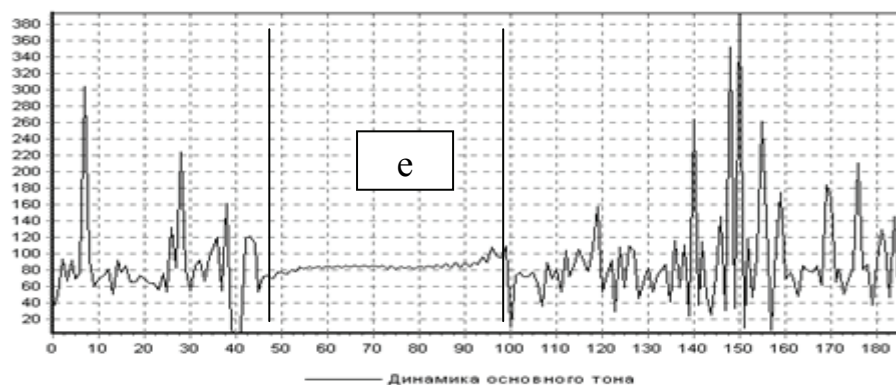


Рис. 2 Динамика частоты основного тона.

На рис.2 вертикальными линиями ограничена область динамики длительности периодов основного тон для фонемы “йэ”. За пределами этой области показана динамика низкочастотного компонента шумных звуков “ш, с”, глухой паузы и глухого взрывного звука “ть”

Решение задачи выявления чередования тональных и шумных звуков строится по параметрам тональности, шумности и интенсивности на интервалах анализа.

Графики, которые иллюстрируют процесс формирования функции частоты перехода через нуль $\rho(T_i)$, представлены на рис. 3.

Как видно из графика рис. 3 уровень частоты переходов через нуль $\rho(T_i)$ для тональных звуков не превышает 15, в то же время как для шумных звуков и глухой паузы в зависимости от фонемы может достигать 60 для интервала анализа $T_{iN} = 220$ отсчётов сигнала $s(t_j)$.

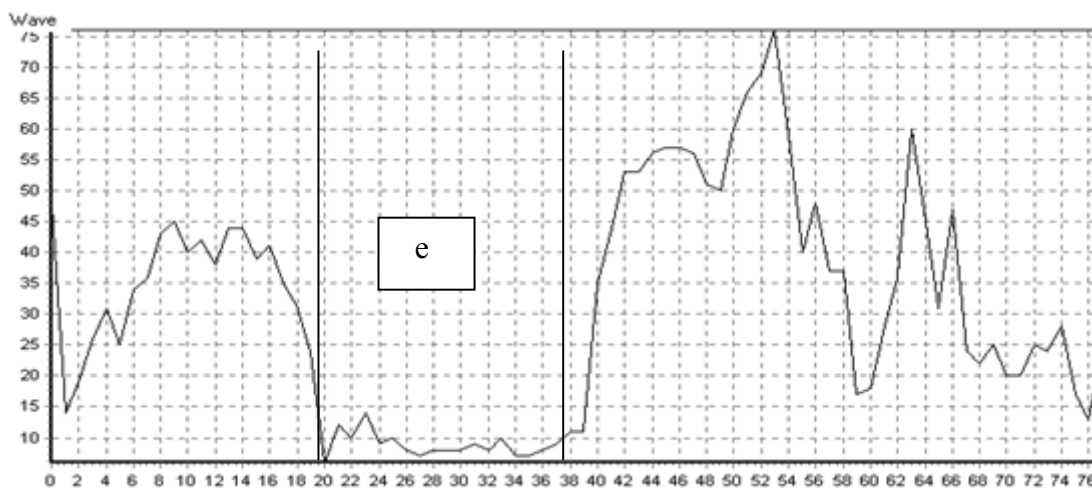


Рис. 3 Функция частоты переходов через нуль $\rho(T)$ сигнала слова «шесть»

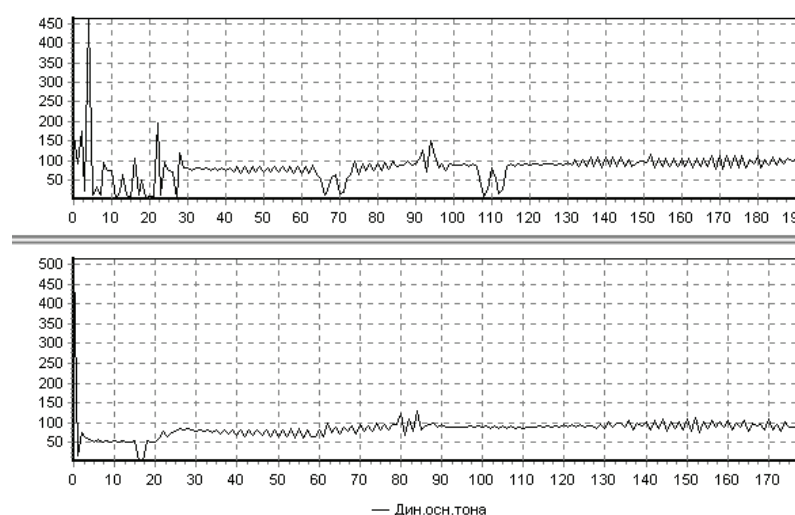


Рис. 4. Динамика частоты сигнала основного тону слова “один” двух реализаций

Задача сопоставления решается для функции динамики частоты сигнала основного тона решается методом динамического программирования только для тональной части сигнала, полученной во временном окне типа рис. 3 по граничному уровню параметра $\rho(T)$.

Выводы. Речевой сигнал имеет две основные компоненты: информационную и индивидуальную. Информационная часть – это слова, которые произносит диктор. Индивидуальная часть – это оттенки речи, связанные с дикцией, функциональным состоянием диктора, индивидуальностью его голоса. В устройствах распознавания речи определяют информационную часть.

Список литературы

1. Карпов О.М., Чугай О.А., Зірнєєва Г.В., Асадулін В.А. Методи та алгоритми оцінки ситуативних відхилень параметрів мови людини: Навч. пос. – Д.: РВВ ДНУ, 2007. – 64 с.
2. Носенко Э.Л., Карпов О.Н., Чугай А.А. Система автоматического определения эмоционального состояния человека по акустическим и темпоральным характеристикам речи./ Матер. симпоз. “Речь, эмоции и личность”. Ленинград, 1975. С. 108-113
3. Карпов О.Н., Носенко Э.Л., Чугай А.А. Устройство для выделения частоты основного тона речи. /А.С. № 773689, “Официальный бюллетень изобретений и открытий”, № 39, 1980.