

УДК 004.934.2

**А.А. Каргин**

Донецкий национальный университет, г. Донецк  
kargin@dongu.donetsk.ua

**Т.В. Шарий**

Донецкий национальный университет, г. Донецк  
tsphere@mail.ru

## Применение нечеткой логики в системах фонологической классификации звуков речи

Статья посвящена вопросу фонологической классификации звуков речи по признакам. Предложен нечеткий подход классификации, в соответствии с которым выходные значения классификационных блоков представляют собой нечеткие величины. Предложены акустические характеристики, на основе которых производятся процедуры фазификации. Приведены результаты экспериментов разделения фонем на классы.

### Введение

Проблема автоматического распознавания речи не теряет своей актуальности на протяжении нескольких десятилетий. За последнее время удалось достичь заметных успехов, однако ключевая задача описания речевых единиц не решена однозначно и в полной мере. Все современные автоматические системы распознавания речи (АСРР) включают модули предобработки и постобработки речевого сигнала [1-3]. В первом модуле осуществляется параметризация сигнала – представление сигнала в виде последовательности компактных векторов признаков. Для этого применяются различные методы цифровой обработки сигналов (кепстральная обработка, вейвлеты, линейное предсказание); при этом, до сих пор не обнаружено пространство их истинно-инвариантных признаков [2]. В модуле постобработки сигнала традиционно используются статистические методы, основанные на скрытых марковских моделях (СММ) [3]. Но, несмотря на относительно хорошие результаты, они обладают недостатками. Главная проблема заключается в том, что СММ не адекватны природе речевого сигнала (речевой сигнал нельзя рассматривать как стационарный случайный процесс, удовлетворяющий предположению о марковости 1-го порядка).

С другой стороны, в 50-60-х годах XX века были проведены обширные исследования в области фонологии и восприятия речевых сигналов человеком. Были предложены системы классификации звуков речи по т.н. различительным признакам (которые впоследствии не нашли применения в АСРР) [4-7]. Данные признаки связаны с конкретными особенностями спектра звука речи и его изменения во времени. **Основная идея и цель статьи** заключается в отказе от исключительно статистического моделирования речевых сигналов и дополнительном нечетком моделировании речевых сегментов на основе фонологических различительных

признаков. Для этого в статье предлагаются алгоритмы вычисления соответствующих признаковых характеристик и их фаззификации.

## Фонологическая признаковая классификация фонем

Ограниченные возможности спектрально-временных методов распознавания звуков речи побудили к поискам методов распознавания с использованием фонологических различительных признаков. Наиболее известна классификация Якобсона, которая была предложена первой [5]. В соответствии с ней, выделяются 12 бинарных признаковых классов звуков. Каждый звук речи (фонема) может иметь данный признак (принадлежать соответствующему признаковому классу, +) или не иметь (не принадлежать, –). В этой классификации выделены основные акустические признаки, на основе которых формируются важнейшие звуковые противопоставления. Каждый из 12 признаков представляет собой противопоставление между двумя относительными проявлениями одного и того же акустического свойства [6]. Одной из особенностей данной системы признаков является то, что для описания гласных и согласных звуков используется один и тот же набор признаков. Этот набор включает следующие признаки [6]:

- 1) Гласный / негласный (гласные звуки акустически характеризуются четко выраженной формантной структурой);
- 2) Согласный / несогласный (согласные звуки имеют более низкий общий уровень интенсивности);
- 3) Прерванный / непрерывный (у прерванных согласных акустически выражен интервал отсутствия или сильного ослабления звуковой энергии в полосе частот выше основного тона, после которого следует «взрыв» или резкое изменение формантной картины);
- 4) Глоттализированный / неглоттализированный (глоттализированные согласные характеризуются резким включением интенсивного источника шума);
- 5) Резкий / нерезкий (в основе признака лежат относительные различия в интенсивности, длительности и степени упорядоченности фрикативного шума, т.е. своего рода шкала шумности. Резкие согласные имеют интенсивный и длительный шум);
- 6) Звонкий / глухой (звонкие звуки произносятся с участием голосового источника);
- 7) Компактный / диффузный (компактные звуки характеризуются большей концентрацией энергии в относительно узкой серединной (у диффузных – краевой) части частотного диапазона спектра, большей интенсивностью и длительностью);
- 8) Низкий / высокий (к низким (низкотональным) относятся звуки, у которых энергия сосредоточена в более низких частотах, чем у высоких (высокотональных));
- 9) Бемольный / простой (у бемольных звуков частоты всех формант понижаются и энергия высокочастотной области спектра ослабляется);
- 10) Диезный / простой (у диезных вокальных звуков увеличиваются частоты нижних формант и возрастает интенсивность по сравнению с простыми (недиезными) звуками);

11) Напряженный / ненапряженный (напряженные звуки характеризуются большей длительностью, большей интенсивностью, более отчетливым и богатым спектром);

12) Носовой / ртовый (у носовых звуков, образуемых с участием носового резонатора, в спектре появляются устойчивые, мало изменяющиеся форманты назализации (200..300 Гц)).

Однако наиболее исчерпывающая работа в области фонологической классификации была проведена Вайреном и Штубсом, которые предложили электронную схему бинарной селекции звуков (рис.1). Данная схема уточняет схему классификации Якобсона и в большей степени соответствует реальным фонологическим данным [4].

Согласно этой схеме, звуки сначала делятся на глухие и звонкие. Это разделение производится по признаку наличия (для звонких звуков) или отсутствия (для глухих) основного тона. Далее звонкие делятся на шумовые и нешумовые. Это разделение производится по интенсивности спектра в области первой форманты: шумовые имеют большую интенсивность в высокочастотной области. Глухие, в свою очередь, делятся на взрывные и шипящие. Это разделение производится по уровню энергии сигнала на начальном участке звука. Из остальных блоков классификации были разработаны только блок разделения сонорных и блоки разделения гласных. Работа указанных блоков основана на принципе сравнения уровней спектра в различных областях частотного спектра [7].

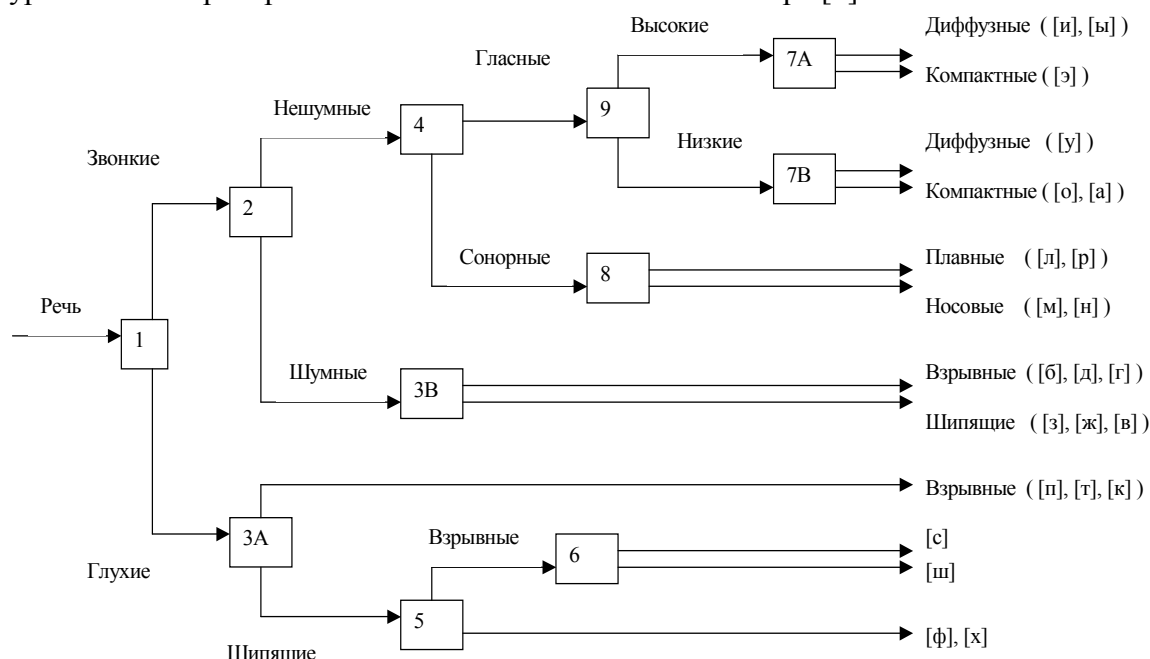


Рисунок 1 – Схема Вайрена-Штубса бинарного разделения звуков на признаковые классы.

## Переход от бинарной классификации звуков к нечеткой

Бинарный подход к признаковой классификации звуков речи обладает главным недостатком: при ошибочном срабатывании одного из блоков, вся дальнейшая

цепочка вычислений оказывается неверной и бессмысленной. И если вероятность правильного разделения звуков на «звонкие» и «глухие» (блок 1) довольно высока (95%, согласно [4]), то удовлетворительное разделение звуков на «гласные» и «сонорные» (блок 4) практически оказывается проблематичным. По мнению авторов статьи, для определения признакового класса фонемы целесообразно рассчитывать степени соответствия текущего речевого образа *каждому* из признаковых классов, на всех уровнях, и принимать окончательное решение на основе всей информации.

Для нечеткой классификации фонем необходимо ввести числовые величины, характеризующие акустические свойства выраженности того или иного различительного фонологического признака, а также разработать процедуры фазсификации для каждого признакового класса на основе данных величин. Для этого были изменены и уточнены процедуры анализа спектров, предложенные Вайреном и Штубсом. Таким образом, в нашей классификационной системе признак звонкости (блок 1), к примеру, представлен не в виде «Фрейм X: звонкий – да, глухой – нет», а в виде 2 нечетких переменных:

$$\langle \text{Звонкий}, \{x, \mu_{\text{Звонкий}}(x)\} \rangle, \quad \langle \text{Глухой}, \{x, \mu_{\text{Глухой}}(x)\} \rangle \quad (1)$$

Далее приводятся формулы для расчета акустических характеристик признаковых классов и получения на основе их степеней принадлежности  $\mu$  текущей фонемы конкретному классу.

**Мера звонкости фонемы.** В блоке 1 на рис.2 происходит разделение звуков на звонкие и глухие. Для построения нечеткой модели введем меру звонкости. Эта величина должна отражать степень выраженности частоты основного тона (ЧОТ) и ее гармоник в спектре звука. ЧОТ представляет собой частоту колебаний голосовых связок. При образовании вокализованных звуков речи воздушный поток проходит через колеблющиеся голосовые связки, поэтому в их спектре четко видна частота основного тона и ее гармоники. На невокализованных участках речи основной тон отсутствует.

В работе ЧОТ выделяется на основе спектра фрейма с помощью следующего алгоритма. Находятся локальные пики в спектре, в интервале частот 100-450 Гц (диапазон голосов от низких мужских до высоких женских). Рассматриваются только те пики, амплитуда которых превышает порог  $\alpha$ :

$$N = \{v_0 \mid v_0 \in [100; 450] \wedge f(v_0) \geq \alpha\}, \quad (2)$$

где  $v_0$  – частота, на которой наблюдается пик в спектре;  $f(v)$  – значение амплитуды на частоте  $v$  в спектре. Для каждой пиковой частоты  $v_0$  из множества  $N$  вычисляется значение:

$$S(v_0) = \sum_{k=1}^{BN} \ln(f(k \cdot v_0)), \quad \forall v_0 \in N, \quad (3)$$

где  $BN$  – число анализируемых гармоник основного тона;  $k$  – номер предполагаемой гармоники основного тона. Сумма (3) характеризует степень

выраженности гармоник конкретной ЧОТ. Частота  $\nu_0$ , дающая максимальную сумму (3), полагается частотой основного тона:

$$F0 = \begin{cases} \arg \max_{\nu_0} S(\nu_0) & , |N| > 0 \\ 0 & , |N| = 0 \end{cases} \quad (4)$$

Мера звонкости (Vocality Measure, VM) выражается формулой:

$$VM = \begin{cases} 1 - \frac{\sum_{k=1}^{BN} f(k \cdot F0_2) \cdot w(k \cdot F0_2)}{\sum_{k=1}^{BN} f(k \cdot F0) \cdot w(k \cdot F0)} & , F0 > 0 \\ 0 & , F0 = 0 \end{cases} \quad (5)$$

где  $F0$  – ЧОТ;  $F0_2 = F0 + \frac{F0}{2}$  – частота, смещенная относительно ЧОТ на половину (в спектре вокализованных звуков на этой и кратной ей частотам наблюдаются провалы);  $BN$  – число анализируемых гармоник основного тона;  $w(\nu)$  – весовая функция (значения весов для разных частотных диапазонов спектра приведены в табл.1). Чем сильнее проявляется основной тон и его гармоники, тем ближе величина VM к 1.

Таблица 1 – весовые коэффициенты для расчета акустических характеристик фонем

Частотные диапазоны	100-700 Гц	700-1500 Гц	1500-2500 Гц	2500-4000 Гц	4000-7000 Гц	> 7000 Гц
Весовые коэффициенты	1	1.4	1.8	2.0	2.3	2.5

Определим на основе меры звонкости нечеткое множество «Звонкий звук» с S-образной функцией принадлежности (график функции принадлежности показан на рис.2а):

$$\tilde{VM} = \{ VM, \mu_{\tilde{VM}}(VM) \} \quad , \quad \mu_{\tilde{VM}}(x) = \begin{cases} 0 & , x \leq 0 \\ x & , 0 < x \leq 1 \\ 1 & , x > 1 \end{cases} \quad (6)$$

**Мера согласности фонемы.** В схеме Вайрена-Штубса, в отличие от схемы Якобсона, отсутствует признаковый класс «Согласные», однако в нашей классификационной схеме он используется как вспомогательный признак. Мера согласности (CM, Consonantness Measure) должна отражать относительную интенсивность участка речи [4,7], поэтому предложено для ее вычисления использовать формулу:

$$CM = 1 - \frac{E_{cur}}{E_{max}} \quad , \quad (7)$$

где  $E_{cur}$  – энергия текущего речевого сегмента;  $E_{max}$  – максимальная энергия сигнала в интервале речи между паузами, содержащем текущий сегмент. Энергия сегмента сигнала вычисляется по формуле:

$$E(k, L) = \frac{1}{L} \sum_{i=k}^{k+L} |x_i|, \quad (8)$$

где  $k$  – позиция начального отсчета сегмента сигнала;  $L$  – число отсчетов в сегменте (длительность);  $x_i$  – значение  $i$ -го отсчета сигнала.

Введем на основе меры согласности нечеткое множество «Согласный звук» с S-образной функцией принадлежности (график функции принадлежности приведен на рис.2б):

$$\tilde{CM} = \{ CM, \mu_{\tilde{CM}}(CM) \}, \quad \mu_{\tilde{CM}}(x) = \begin{cases} 0 & , x \leq 0.2 \\ \frac{x-0.2}{0.6} & , 0.2 < x \leq 0.8 \\ 1 & , x > 0.8 \end{cases} \quad (9)$$

Константы 0.2 и 0.8 были подобраны экспериментальным путем.

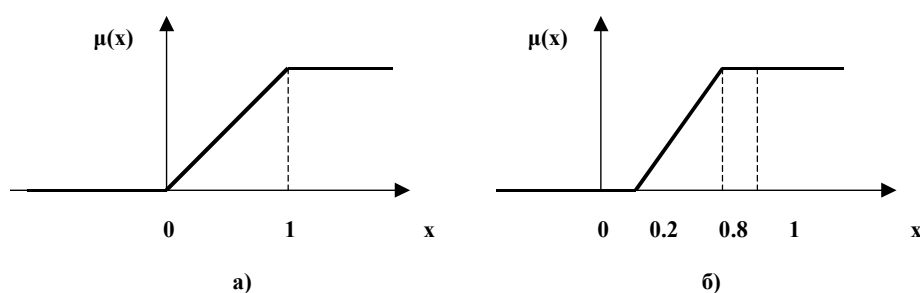


Рисунок 2 – Графики функций принадлежности нечетким множествам фонологических признаков

**Мера гласности фонемы.** В блоке 4 на рис.1 производится разделение нешумных звонких звуков на гласные и сонорные. Согласно фонологическим исследованиям, гласные звуки характеризуются четко выраженной формантной структурой (формантами называют специфические спектральные «подъемы» гласных). Данное определение довольно размыто; кроме того необходимо дополнительно производить формантный анализ фонем, т.к. разные фонемы имеют разные форманты. В статье предлагается мера «выраженности» формантной картины спектра речевого фрагмента (FM, “Formantness” Measure):

$$FM = 1 - \frac{\min_{i \in [200; 1000]} \left\{ \sum_{j=i}^{i+240} f(j) \cdot w(j) \right\}}{\max_{i \in [200; 1000]} \left\{ \sum_{j=i}^{i+240} f(j) \cdot w(j) \right\}}, \quad (10)$$

где  $f(j)$  – значение амплитуды на частоте  $j$  в спектре;  $w(j)$  – весовая функция (значения весов для разных частотных диапазонов спектра берутся из табл.1). Смысл формулы (10) заключается в том, что в частотном диапазоне [200Гц; 1000Гц] находится минимальный и максимальный по амплитуде диапазоны шириной 240Гц (среднестатистическая ширина форманты), и чем меньше их отношение, тем ярче проявлены “провал” и “подъем” в спектре. На основе характеристики  $FM$  строится нечеткое множество «Формантный»:

$$\tilde{FM} = \{ \tilde{FM}, \mu_{\tilde{FM}}(\tilde{FM}) \}, \quad \mu_{\tilde{FM}}(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & x > 1 \end{cases} \quad (11)$$

Нечеткое множество «Гласный» определяется как:

$$\tilde{VCM} = \overline{\tilde{CM}} \cap \tilde{FM} \quad (12)$$

**Мера шумности звонкого.** В блоке 2 на рис.1 производится разделение звонких звуков речи на шумные и нешумные. В [7] шумные выделяются по признаку большей интенсивности в высокочастотной области спектра, по сравнению с областью первой форманты. Однако проведенные авторами статьи эксперименты показали ненадежность указанного способа, т.к. «шумные» согласные необязательно имеют большую энергию в высокочастотной области спектра, в то время как «нешумные» гласные, наоборот, могут иметь некоторый шум. В связи с этим было предложено использовать информацию о звонкости и гласности фонемы (формулы (5) и (9)). Нечеткое множество «Шумный» определяется как:

$$\tilde{NM} = \overline{\tilde{VCM}} \cup \tilde{VM} \quad (13)$$

**Мера сибилантности.** В блоках 3А и 3Б на рис.1 происходит разделение звуков на взрывные и шипящие. В [7] это разделение производится по уровню энергии сигнала на начальном участке звука. Однако выделить автоматическим образом начальный участок произвольного звука довольно проблематично, а чаще всего невозможно (перед началом анализа система не имеет никакой информации о том, какая именно фонема звучит). С другой стороны, шипящие (сибилантные) звуки имеют легко узнаваемый вид спектра – наличие шума в довольно широком диапазоне высокочастотной области. В связи с этим, в работе предлагается следующая мера сибилантности звука (SBM, Sibillance Measure):

$$SBM = 1 - \frac{\sum_{i=400}^{1000} f(i) \cdot w(i)}{\sum_{j=1000}^{8000} f(j) \cdot w(j)}, \quad (14)$$

где  $f(j)$  – значение амплитуды на частоте  $j$  в спектре;  $w(j)$  – весовая функция (значения весов для разных частотных диапазонов спектра берутся из табл.1).

На основе характеристики  $SBM$  строится нечеткое множество «Сибилантный»:

$$\tilde{SBM} = \{ \tilde{SBM}, \mu_{\tilde{SBM}}(\tilde{SBM}) \}, \quad \mu_{\tilde{SBM}}(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & x > 1 \end{cases} \quad (15)$$

**Меры диффузности, высотности, назальности.** Работа блоков 7А, 7Б, 8 и 9 на рис.1 основана на принципе сравнения амплитуд в различных областях частотного спектра. В качестве числовых характеристик, отражающих акустические особенности каждого из этих признаков, в статье вводятся мера диффузности (DM, Diffuseness Measure), мера высотности (ACM, Acuteness Measure) и мера назальности (NSM, Nasality Measure). Все предлагаемые меры имеют общий вид:

$$M(FNL, FNL, FDL, FDH) = \frac{\sum_{i=FNL}^{FNH} f(i) \cdot w(i)}{\sum_{j=FDL}^{FDH} f(j) \cdot w(j)}, \quad (16)$$

где  $f(j)$  – значение амплитуды на частоте  $j$  в спектре;  $w(i)$  –  $i$ -ый вес (значения весов для разных частотных диапазонов спектра берутся из табл.1);  $FNL, FNH, FDL, FDH$  – параметры, которыми отличаются три введенные меры, и представляющие собой границы анализируемых частотных диапазонов спектра.

Значения  $FNL, FNH, FDL, FDH$  приведены в табл.2.

Таблица 2 – Границы частотных диапазонов для мер DM, ACM, NSM

Границы диапазонов	FNL (Гц)	FNH (Гц)	FDL (Гц)	FDH (Гц)
<b>DM</b>	220	380	340	1100
<b>ACM</b>	2400	3700	800	1200
<b>NSM</b>	220	340	450	550

На основе мер DM, ACM и NSM вводятся нечеткие множества «Диффузный», «Высокий» и «Носовой», соответственно. Каждое из них имеет S-образную функцию принадлежности (рис.2а).

## Аккумуляция выходных нечетких значений блоков

Заметим, что выше были определены нечеткие множества, связанные с *одним* фонологическим признаком каждого из блоков 1-9 в схеме Вайрена-Штубса. Но блоки имеют два выхода, поэтому необходимо также определить нечеткое множество для альтернативного признака. В работе второе нечеткое множество строится путем нечеткого отрицания первого. Например, нечеткое множество «Нешумный» (блок 2) является отрицанием (нечетким дополнением) нечеткого множества «Шумный». Отметим также, что в общем случае два выходных нечетких множества блока могут быть не связаны между собой и строиться на основе *разных* акустических характеристик.

В соответствии с нашей схемой нечеткой классификации звуков речи по признакам, выходом каждого блока на рис.1 являются значения функций



принадлежности нечетким множествам двух выходных признаков классов блока. Эти значения вычисляются автономно для каждого блока и могут быть использованы по отдельности, однако необходим также комплексный анализ работы блоков, входящих в одну «цепочку». Для расчета функций принадлежности нечетким множествам финальных блоков (блоки 3,5,6,7,8 на рис.1) используются следующие альтернативные операции:

- 1) алгебраическое пересечение (произведение)  $N$  нечетких множеств в степени  $1/N$ :

$$\mu_{\tilde{A}}(x) = \left[ \prod_{i=1}^N \mu_{\tilde{A}_i}(x_i) \right]^{1/N} \quad (17)$$

- 2) модифицированный вариант алгебраического объединения  $N$  нечетких множеств:

$$\mu_{\tilde{A}}(x) = \frac{1}{N} \sum_{i=1}^N \mu_{\tilde{A}_i}(x_i) \quad , \quad (18)$$

где  $\tilde{A}$  – нечеткое множество выходного признака финального блока;  $\tilde{A}_1, \dots, \tilde{A}_N$  – нечеткие множества выходных признаков  $N$  блоков, входящих в цепочку с финальным блоком  $\tilde{A}$ ;  $x_i$  – значение акустической меры, используемая для вычисления степени принадлежности множеству  $\tilde{A}_i$ . Нормировка на  $N$  в обоих случаях вводится в связи с тем, что разные цепочки включают разное количество блоков.

## Результаты экспериментов

Для экспериментов использовались файлы речевой базы VoxForge [8] и инструментальная среда анализа речевых файлов, разработанная авторами [9]. Всего обработано 25 файлов, содержащих предложения, сказанные 3 дикторами (голоса с ЧОТ в диапазоне 100-200Гц). Общее количество проанализированных фонем – 412.

Для оценки качества нечеткой классификации и предлагаемых в статье акустических мер использовалась величина «разброс выходных нечетких значений блока»:

$$\Delta \mu(\tilde{A}, \tilde{B}, x) = \mu_{\tilde{A}}(x) - \mu_{\tilde{B}}(x) \quad , \quad (19)$$

где  $\tilde{A}$  – нечеткое множество первого выходного признакового класса блока;  $\tilde{B}$  – нечеткое множество второго выходного класса;  $x$  – значение акустической меры, используемой в блоке для разделения признаков. Стоит отметить, что гипотезой в данном случае является утверждение «текущий фрагмент  $x$  принадлежит признаковому множеству  $\tilde{A}$ ». Таким образом, величина  $\Delta \mu$  может находиться в пределах от 1 (в случае, когда степень принадлежности величины  $x$  множеству  $\tilde{A}$  равна 1, а множеству  $\tilde{B}$  – 0) до  $-1$  (в противоположном случае). Статистика разностей (19) для отдельных признаковых классов позволяет оценить качество разделения звуков речи на данные классы. В табл.3 приведены значения математического ожидания, дисперсии, медиан показателя разброса выходных нечетких значений блоков 1-9 (рис.1).

Таблица 3 – Статистические показатели (среднее, дисперсия, медиана) величины  $\Delta\mu$ 

Блок	$M\Delta\mu$	$D\Delta\mu$	$med \Delta\mu$	Блок	$M\Delta\mu$	$D\Delta\mu$	$med \Delta\mu$
<b>1</b>	0.59	0.21	0.72	<b>6</b>	0.57	0.1	0.58
<b>2</b>	0.53	0.12	0.6	<b>7</b>	0.49	0.08	0.48
<b>3</b>	0.55	0.16	0.68	<b>8</b>	0.54	0.13	0.51
<b>5</b>	0.52	0.08	0.54	<b>9</b>	0.74	0.1	0.84

Из табл.3 видно, что автономная работа блоков дает вполне удовлетворительные результаты. Лучше всего разделяются звуки речи (большие значения  $\Delta\mu$ ) в блоке 9 («Высокий» / «Низкий»). При этом дисперсия показателя разброса выходных нечетких значений блока 9 невелика (равна 0.1), и половина значений находится выше величины 0.84 (медиана). Кроме того, в остальных блоках средние значения  $\Delta\mu$  также вполне удовлетворительны; они находятся в диапазоне [0.49, 0.59]. При этом значения дисперсий тоже относительно невелики. В таблице не приведены результаты работы блока 4 разделения звонких звуков на гласные и сонорные, т.к. в ходе экспериментов и теоретических исследований так и не было найдено адекватной акустико-фонетической характеристики, на основе которой можно было бы с высокой степенью точности разграничивать данные классы. Очевидно, для разделения звуков на эти классы необходим учет фонемного контекста и информации более высокого, лингвистического, уровня абстракции, и это является предметом дальнейших исследований.

Как уже было отмечено, в цепочку классификации входит несколько блоков и для принятия окончательного решения о принадлежности текущего звука речи признаковому классу необходим анализ степеней соответствия (17)-(18). Для оценки качества аккумуляции нечетких выходных значений блоков цепочки использовался модифицированный вариант величины (19):

$$\Delta \mu(A, B, x) = \mu_{\tilde{A}}(x) - \mu_{\tilde{B}}(x) \quad (19)$$

где  $A$  – нечеткое множество выходного признакового класса финального блока в цепочке (одного из блоков 3,5-8);  $B_i$  – нечеткие множества выходных классов других финальных блоков. Значение  $\mu$  вычисляется по одной из формул (17)-(18). Таким образом, показатель (20) позволяет оценить, насколько степень соответствия текущего речевого сегмента верному признаковому классу отличается от максимальной из степеней соответствия «чужим» признаковым классам. В табл.4 приведены значения математического ожидания, дисперсии, медиан и интервала показателя (20), где  $\Delta\mu$  вычисляется по формулам (17) и (18).

Таблица 4 – Статистические показатели величины  $\Delta\mu$  выходных блоков 3,5,6,7,8

Блок	$M\Delta\mu_1$	$D\Delta\mu_1$	$med\Delta\mu_1$	интервал	Блок	$M\Delta\mu_2$	$D\Delta\mu_2$	$med\Delta\mu_2$	интервал
<b>3</b>	0.01	0.08	0.07	1.12	<b>3</b>	0.11	0.1	0.1	1.31
<b>5</b>	0.01	0.02	0.01	0.32	<b>5</b>	0.17	0.05	0.18	1.09
<b>6</b>	0.04	0.03	0.04	0.29	<b>6</b>	0.2	0.06	0.21	1.03
<b>7</b>	0.05	0.01	0.02	0.25	<b>7</b>	0.19	0.03	0.14	0.62
<b>8</b>	0.07	0.01	0.05	0.49	<b>8</b>	0.06	0.08	0.04	0.98

В табл.4  $\Delta\mu_1$  – показатель разброса степеней соответствия, вычисляемых по формуле (17);  $\Delta\mu_2$  – показатель разброса степеней соответствия, вычисляемых по формуле (18). Из табл.4 видно, что в среднем величина  $\Delta\mu_2$  дает больший разброс, однако в обоих случаях значения  $\Delta\mu$  относительно невелики (в среднем – от 0.01 до 0.19) с относительно большим интервалом (не менее 0.25). Данная ситуация была ожидаема, т.к. она подтверждает проблемы, связанные с попытками точной (бинарной) классификации звуков речи. Однако в нашем случае положительным фактом можно считать, что средние значения  $\Delta\mu$  и медианы  $\Delta\mu$  больше 0.0. Кроме того, отдельные проблемы вызваны несоответствиями теоретических рекомендаций практическим наблюдениям. Например, в случае с блоком 3, фонема [в] часто не имела спектр, характерный для шипящих звуков. Таким образом, перераспределение отдельных фонем по признаковым классам для большего соответствия акустическим мерам, введенным в работе, может в дальнейшем улучшить показатели разделения.

## Выводы

Проблема параметризации речевого сигнала в контексте создания АСРР актуальна и нуждается в решении. В статье показано, что современные методы параметризации речевых сигналов можно дополнить анализом различительных фонологических признаков звуков речи. Фонологические признаки, такие как «звонкий», «гласный», «шипящий» и т.д., менее чувствительны к шуму и особенностям диктора, по сравнению с традиционно используемыми кепстральными коэффициентами. Однако предложенные фонологами бинарные схемы широкой классификации фонем обладают недостатком категоричности: в ситуации плохой различимости звука он может быть отнесен к противоположному классу. Поэтому нечеткий подход к фонологической классификации, предлагаемый в работе, обоснован теоретически и практически. Как показали эксперименты, средние значения показателя разброса выходных нечетких значений блоков фонологической классификации вполне удовлетворительны; они находятся в диапазоне [0.49, 0.74] («плохими» значениями являются значения, меньшие либо равные 0.0). Значения медиан в большинстве случаев близки к значениям средних. При этом значения дисперсий тоже относительно невелики. Все перечисленное можно считать «благоприятной» статистикой для применения разработанных алгоритмов фазификации на практике.

Дальнейшая работа связана с разработкой и исследованием нечетких моделей, позволяющих определять степень соответствия текущего речевого сегмента конкретной фонеме, на основе нечеткой информации как «признакового» уровня, так и «кепстрального».

## Литература

1. Research Developments and Directions in Speech Recognition and Understanding, Part 1 [electronic resource] / URL: <http://research.microsoft.com/pubs/80528/SPM-MINDS-I.pdf> / 30.05.2009.
2. X.Huang. Spoken Language Processing: A guide to theory, algorithm, and system development [text] / X.Huang, A.Acero, H.Hon. – Prentice Hall. – 2001. – 980p.
3. J.K.Baker. Spoken language digital libraries: The million hour speech, project [text] / J.K.Baker // *Proceeding of Int. Conference Universal Digital Libraries, Alexandria, Egypt.* – pp.1-13. – 2006.
4. Сапожков М.А. Речевой сигнал в кибернетике и связи [текст] / М.А.Сапожков. – М.: Связьиздат, 1963. – 453с.
5. Jakobson R. Preliminaries to Speech Analysis. The Distinctive Features and Their Correlates [text] / Jakobson R., Fant C.G.M., Halle M. – MIT, Acoustics Lab. Tech. Rep, №13. – 1952.

6. М.Ф.Бондаренко. Проблемы распознавания речевых образов спектральными методами в области прикладной лингвистики [текст] / М.Ф.Бондаренко, А.В.Работягов, С.В.Щепковский // *Бионика интеллекта: Научн.-техн. журнал.* – 2006. – № 1(64). – С.78–85.
7. Wiren J. Electronic Binary Selection System for Phoneme Classification [text] / Wiren J., Stubbs H. // *JASA.* – Vol.28(6). – 1956. – pp.1082-1091.
8. VoxForge home page [электронный ресурс] / URL: <http://www.voxforge.org/home> / 30.10.2009.
9. Каргин А.А. Разработка инструментальной среды интеллектуального анализа аудиальных данных [текст] / А.А.Каргин, Т.В.Шарий // *Труды VIII международной конференции «Интеллектуальный анализ информации ИАИ-2008», г.Киев.* – 2008. – С.558–564.

**А.А. Каргин, Т.В. Шарий**

**Застосування нечіткої логіки в системах фонологічної класифікації звуків мовлення**

Стаття присвячена питанню фонологічної класифікації звуків мовлення за ознаками. Запропоновано нечіткий підхід класифікації, згідно з яким вихідні значення класифікаційних блоків представляють собою нечіткі величини. Запропоновано акустичні характеристики, на основі яких виконуються процедури фазифікації. Наведені результати експериментів розділення фонем на класи.

**A.A. Kargin, T.V. Shariy**

**Applying the Fuzzy Logic in the Systems of Phonological Classification of Speech Sounds**

The article is devoted to the task of phonological classification of the speech sounds by features. The novel fuzzy approach to the classification is offered, according to which the output values of the classification blocks are fuzzy values. The acoustic characteristics for the fuzzification procedures are offered. The results of experiments of the class discrimination of phonemes are given.

*Статья поступила в редакцию 27.05.2010.*