УДК 004.93

Особенности распознавания тональности в речевом потоке

Балакирев Н.Е.,[*] Нгуен Х.З.[**]

Московский авиационный институт (национальный исследовательский университет), МАИ, Волоколамское шоссе, 4, Москва, А-80, ГСП-3, 125993, Россия

[*]*e-mail: balakirev1949 @yandex.ru*

[**]*e-mail: nguyenhoangzuy@gmail.com*

Аннотация

В работе рассматривается один из возможных подходов к решению задачи распознавания особых аспектов речи, связанных с тональностью, которая занимает важное место в человеческой коммуникации. Градус особого состояния напряжения, прежде всего, отражается не в содержании слов, а способе их произношения, который несет иногда другую смысловую нагрузку относительно содержания слов. Решение такой задачи может использоваться в авиационной технике, в частности, для автоматического распознавания эмоционального состояния на борту, для выделения эмоциональных отрезков в записях бортовых самописцев с речью пассажиров местных и международных рейсов. Особое ключевое значение тональность имеет в тональных языках Юго-Восточной Азии, хотя тональность и для европейских языков имеет немаловажное значение, отражая характер произносимой фразы и привнося дополнительный смысл в содержание распознаваемых слов и предложений. Так или иначе, сама тональность проявляется идентично, но имеет свои особенности по отношению информационного содержания фонемы или слова. И это прежде всего касается рассмотрения самого

объекта, несущего информацию о тоне. В отличие от решения задачи распознавания последовательности слов, где ориентиром является совокупность частот, задача распознавания тональности не может опираться на общепринятые математические методы обработки и распознавания волн. Рассмотрение вопросов распознавания тональности обычно выходит за рамки широкого обсуждения в этих методах, а также весьма ограничен круг предложений алгоритмического разрешения этой задачи. Поэтому рассматривается на примерах тональная составляющая, прежде всего, фонемы, которая может быть получена специальными методами, отличными от традиционных методов. Предлагаются методы, опирающиеся на установление отношений между характерными точками и представление конфигурации этих отношений в виде матричной модели. Фактически, такая модель является качественной характеристикой тональности, не зависящей от значения амплитуд, что позволяет сравнивать разные проявления тональности, выраженные в громкости произношения и в особенностях артикуляционного аппарата конкретного человека. Само сравнение предполагает наличие качественной меры, которая позволяет отражать степень различия рассматриваемых фонем в речевом потоке. Все эти вопросы обсуждаются в данной статье.

Ключевые слова: распознавание речи, тональность, тон, ударение, интонация, фонема, структурная матрица.

Введение

Создание различного вида распознающих систем [1] и, в особенности, систем распознавание речи, несмотря на уже существующие практические примеры применения, остается актуальной сферой интересов исследователей и разработчиков, решающих задачу «человеческого» общения с информационным пространством компьютерного мира. Основной тренд исследований и разработок ведется, прежде всего, для английского языка и языков европейского типа. Значительные наработки в распознавании речи – например, последние работы Яндекса, имеются и для русского языка. Языки же Юго-Восточной Азии слабо представлены в исследованиях и разработках [2-4], что может быть связано с особым характером фонемной базы, характеризуемой тональностью в гласных фонемах [5]. Следует отметить большое разнообразие как языков, обладающих этой особенностью, так и многообразие тональных рядов и уровней гласных фонем, представленных в этих языках. К сожалению, невозможно отказаться от специального рассмотрения вопроса тональности в распознавании речи при распознавании языков этого региона, и механистически перенести методы и алгоритмы, применяемые для европейских языков на языки Юго-Восточной Азии. Слова, одинаково слышимые европейцами, при изменении тона имеют абсолютно разные смысловые значения.

В рассмотрении данного вопроса нас будут интересовать не смысловые или чисто лингвистические аспекты тональности в частности фонетики и фонологии[5], а так называемый акустический аспект фонетики – т.е. изучение звуков речи с точки зрения их физических характеристик и математических методов их распознавания[6-8]. В ходе исследований было установлено, что с технической

точки зрения информационной основой тональности являются не частотные или амплитудные характеристики звукового потока сигналов, а их особая совокупность или сочетание, которое локально неповторимо и представлено в единственном экземпляре такой информационной характеристики как тон.

Особый характер тональности в речевом потоке

В данной статье **тональность,** как явление, не имеющее строгих рамок в лингвистике и, уж тем более, не формализованная точными науками и не фиксируемая в настоящее время при распознавании речи, рассматривается, прежде всего, с технических позиций. Сопредельные понятия: **тон, интонация** и другие, так или иначе, являются сторонами этого языкового явления. В данном рассмотрении нас будут интересовать не лингвистические аспекты, важность или значимость тональности в коммуникации людей, а вполне практичный аспект – как формализовать и технически распознать «множественный алфавит» тонов речи.

Тональность (мелодика), вполне осязаемая при распознавании речи человеком и соответственно используемая в конструкции построения слов, фраз и предложений того или иного языка, не вызывает затруднений в её фиксации носителями языка. Правда, следует заметить, что богатство речевого использования тональности подкрепляется весьма бедным аппаратом грамматического отображения в письменных источниках. В связи с этим в сценариях к спектаклям и кинофильмам присутствует контекст, предписывающий способ произношения той или иной фразы. То есть, это подтверждает важность тональности в речи человека, которая добавляет дополнительный смысл в текст повествования. Но это и говорит о трудностях строгого описания формализации этого явления в речи людей и

фиксации тональности в письменной форме. Несколько другая ситуация возникает, когда рассматривается тональность, проявляемая в фонемах.

Тональность, если подходить не очень углубленно, имеет три формы проявления: тональность на уровне фонемы, тональность на уровне слова и межсловная тональность. В первом случае уместно использовать понятие **тон**. Во втором случае это будет именоваться **ударением**. И в третьем случае будем использовать понятие **интонация**. Для европейских языков смысловое очертание или проявление мелодики в речи имеет, прежде всего, как между словная тональность, с помощью которой можно выразить различные эмоциональные окраски и характер произносимого предложения. А вот в языках Юго-Восточной Азии тональность характерна для фонем, а точнее гласных фонем. К разряду явлений тональности, как промежуточному между двумя этими проявлениями, следует отнести и явление ударных и безударных гласных, с помощью которых мы придаем иногда разный смысл одному и тому же фонетически звучащему слову. Итак, в случае наличия **тона**, слова, произнесенные представителем Юго-Восточной Азии, которые для европейца будут звучать абсолютно одинаково, для жителей этого региона могут означать абсолютно разные понятия. Приведем пример тональности одной из фонем вьетнамского языка и? используя собственные инструментальные средства и логико-лингвистический подход, применяемый в исследованиях [8-10], рассмотрим осциллограмму этой фонемы в различных тонах. Шесть тонов фонемы «а» вьетнамского языка (шесть фонем типа «а») определяются специфической динамикой основного тона на звонком участке слога. А следуя техническому языку, они определяются характерной изменчивостью значения

амплитуд (энергии), которые однозначно определяются в процессе формирования фонемы артикуляционным аппаратом человека.

1-й тон (по-вьетнамски **ngang**) характеризуется относительно постоянными значениями на слоге. В письменности этот тон не обозначается никакими специальными знаками сверху или снизу буквы, соответствующей гласному. Например: ki – «килограмм»

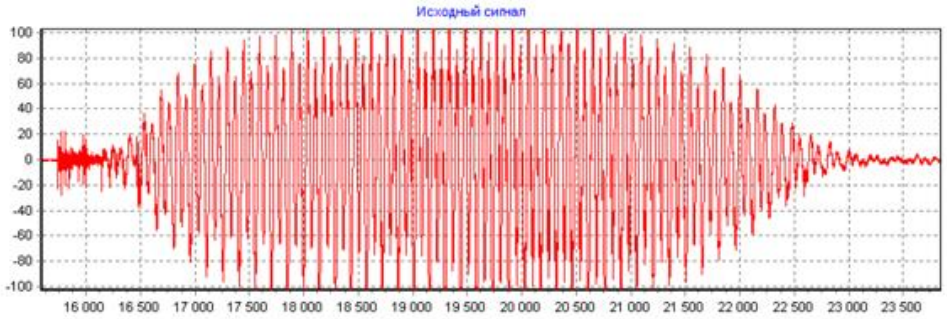

Рис. 1 Осциллограмма – 1-й тона в слове ki

2-й тон (по-вьетнамски **huyền)** характеризуется постепенным понижением от начала к концу слога. В письменности этот тон отмечается знаком " \ " над гласной. Например: kì – «флаг», «период».

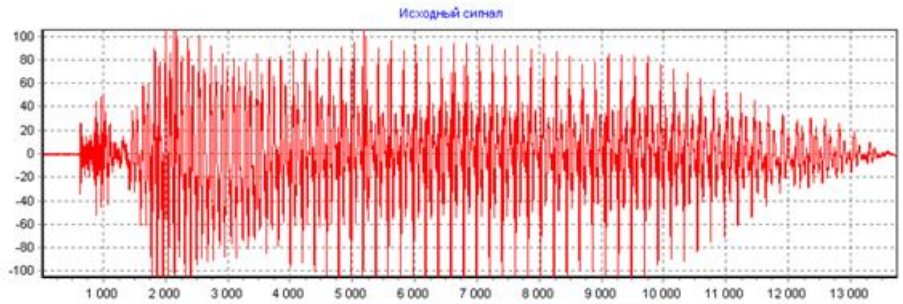

Рис. 1 – Осциллограмма – 2-го-й тона в слове kì.

3-й тон (по-вьетнамски **ngã**) характеризуется последовательными участками нарастания, спада и повторного нарастания. В письменности он отмечается знаком " ~ ". Например: k**ĩ** – «тщательно», «умело».

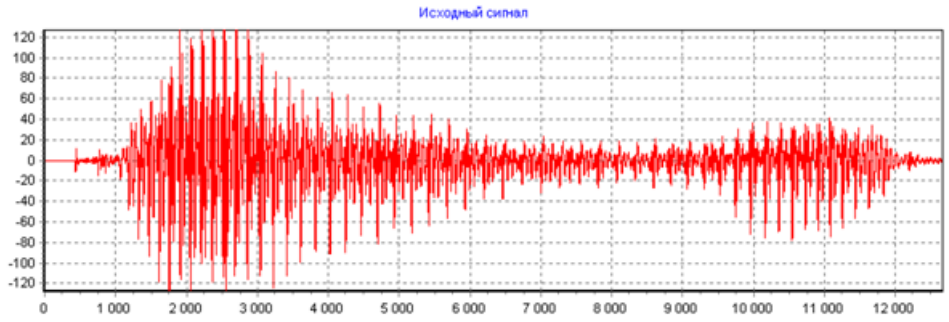

Рис.3 – Осциллограмма – 3-го-й тона в слове kĩ.

4-й тон (по-вьетнамски ***hỏi***) характеризуется нарастанием на первой половине звонкого участка спада и спадом на второй. В письменности он отмечается знаком " ? ". Например: k**ỉ** – «век».

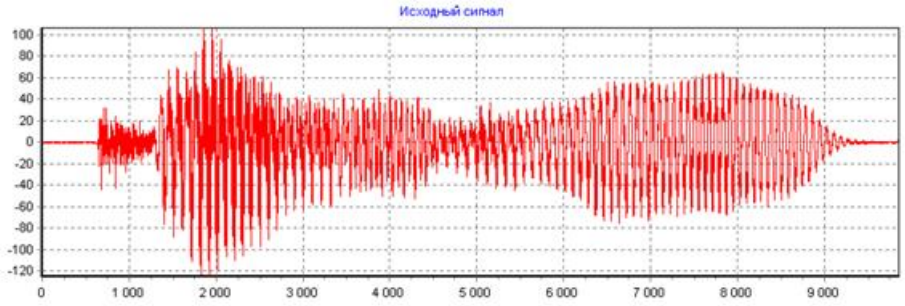

Рис. 4 – Осциллограмма – 4-го-й тона в слове kỉ.

5-й тон (по-вьетнамски **sác**) характеризуется нарастающей частотой на звонком участке. Отмечается знаком " / ". Например: kí – «подпись».

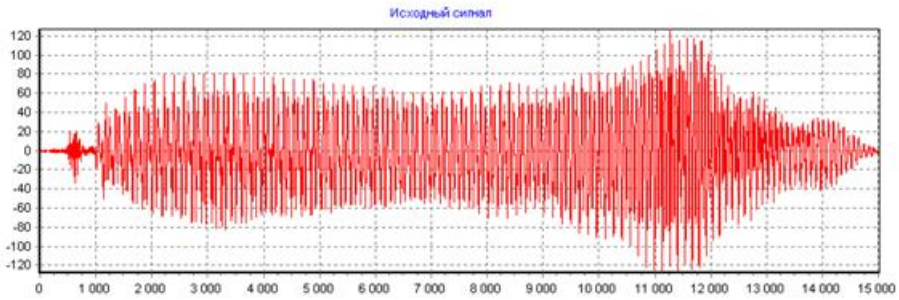

Рис. 5 – Осциллограмма – 5-го-й тона в слове kí

6-й тон (по-вьетнамски **nặng**) чаще всего характеризуется спадом мелодики и последующим подъемом на прежний уровень. Такая мелодика дает основания определять этот тон как нижний ломаный. Например: k**ị** – «неприязнь».

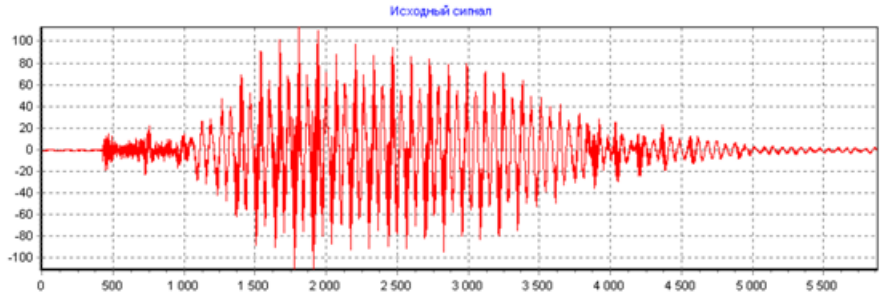

*Рис. 6 – Осциллограмма – 6-го-й тона в слове k**ị**.*

Таким образом, мы видим абсолютно разные по характеру рисунки и различную длительность их звучания, хотя при прослушивании для «русского уха» во всех случаях было слышно и распознано одно слово – «ма». Аналогичные

примеры можно привести и продемонстрировать, показывая межсловную тональность или проявление ударение на ту или иную гласную. Картина будет подобной, но более сложной и будет охватывать существенно более длинный участок звуковой последовательности. Понимая проявление мелодии речи, нарушение которой для определенного языка называется акцентом, как определенная, устоявшаяся осмысленность разных видов тональности, решено было сосредоточить усилия на фонемной тональности гласных фонем языков Юго-Восточной Азии, которые несущественно различаются между собой [2,3], но в качестве основы рассматривается вьетнамский язык [3,4].

В соответствии с вышеприведенными примерами тональные фонемы мы можем различать визуально, но с безусловным привлечением человека – носителя языка, который воспроизводил и распознавал указанные тональные фонемы на слух.

Как мы видим, тональность на уровне фонемы проявляется на более коротком участке звукового потока и известно фиксированное множество разновидностей тонов, которое относительно невелико и маркировано особыми пометами при их письменной записи, как например, для вьетнамского языка [5]. Известно, что сама фонема (в данном случае «и», без учета её тонального содержания) в большинстве исследований по распознаванию речи фиксируется на основе частотных характеристик, имеющих отличие относительно численного значения частоты для разных фонем [11-15]. Условиями для получения таких характеристик с технической точки зрения является повторяемость частот на достаточно большом промежутке звукового потока. Но, такого повторения не наблюдается при проявлении тональности. Она проглядывается в характерном обрамлении рисунка

следования одних и тех же частот для различных вариантов тональных фонем. Итак, тональность не повторяется в процессе звучания, а присутствует в «единственном экземпляре».

Постановка задачи

Так образом, для дальнейшего вскрытия сущности тональности необходимо решить следующие вопросы:

1. Определить, что такое тональность в техническом плане.
2. Выделить все варианты проявления тональности для гласных фонем, используя в качестве примера вьетнамский язык.
3. Найти способы (модели) фиксации тональности и их вариантов.

В случае успеха, полученные результаты можно будет распространить на другие виды тональности.

Первый вопрос, как видно из примеров, на сознательном уровне частично решен и тональность в техническом плане – это рисунок обрамляющей линии, наблюдаемый поверх сменяемости максимумов и минимумов колебаний звуковой волны. Вручную возможно выделить такую линию, но как её описать математически и, соответственно, иметь возможность её фиксировать?

Второй вопрос теоретически также частично решен, множество тональностей определено на лингвистическом уровне для гласных фонем, и они вполне могут привязаны к осциллограмме этих тональных фонем. Но опять возникает вопрос, как и какими математическими моделями описывать это множество значений, характеризующих тональность?

Фактически, так или иначе, всё сводится к нахождению ответа на третий вопрос, имеющий принципиальное значение для решения задачи распознавания тональности в любом варианте её проявления.

Итак, переходя к технической стороне вопроса, дадим кратное описание представления звука в современных системах записи[16]. Не трогая длинную цепочку преобразований аналогового сигнала в цифровой, затронем только верхний уровень. Фактически, с точки зрения обработки звукового потока у нас в распоряжении имеются только значения амплитуд (отклонение от нулевой отметки в положительном или отрицательном направлении), которые отображают реальное изменение давления воздуха через фиксированные и одинаковые промежутки времени. С учетом математического взгляда и постановки задачи, концептуально идеальное решение было бы, если бы мы имели или могли предложить способ получения всего спектра волн входящих в звуковой поток, слышимый нами. Но это «звуковое блюдо», замешанное на множестве составляющих элементов, невозможно разложить на исходные элементы взаимно однозначно [17-20]. Это тоже самое, что по значению суммы, даже известного количества слагаемых попытаться определить, какие это были слагаемые. Имеется всем известный способ приближения волновых функций с помощью ряда Фурье [6], который прекрасно работает в моделировании волновых явлений и даже используется на практике, когда генерируются относительно стабильные волны. Но этот аппарат требует выполнения жестких начальных условий, чаще всего невыполнимых на практике, особенно для звуков речи. Существует и другие математические методы, которые в любом случае

стремятся к точному образу поведения волны в виде некоторый функции (ряд Фурье) или алгоритмов, основанных на численном приближении.

В рамках исследований по распознаванию речи был предложен новый подход или, можно сказать, взгляд на объект распознавания – звуковой поток [8,9]. В рамках логико-лингвистического подхода нас должна интересовать, прежде всего, качественная картина «поведения» звукового потока и при необходимости привлекаться количественная оценка.

Отношения упорядоченных трех точек любое множество чисел разбивает на 13 классов подмножеств и такой набор базовых отношений мы назвали **примитивами**. Подобная ограниченность количества классов подмножеств наблюдается при дальнейшем увеличении точек. Фактически такие отношения отношений определяют **конфигурацию,** или **фигуру** отношений. Интуитивно понятно, что сравнивать имеет смысл близкие по структуре фигуры отношений при одинаковом количестве выбранных характерных точек. Но и число таких точек должно быть в «разумных» пределах, чтобы обеспечить и узреть повторяемость конфигурации. Так как здесь при фиксированном потоке действует закон, чем больше точек, тем меньшая вероятность узреть повторяющиеся по конфигурации участки. В теоретических рассуждениях всё вроде бы логично, но как сделать шаг в сторону практического применения и «запечатления» конфигураций? Была предложена[10] широко используемая форма в виде матрицы, которую мы назвали **структурной матрицей**. Совокупность и иерархия таких структур определяет общую конфигурацию звукового потока. Она оказалась той **качественной мерой,** которая позволяет разбивать звуковой поток на иерархию структур, сравнивать

такие структуры, и также, привлекая числовую меру, дополнять и углублять далее процесс распознавания. В подобной раскладке отсутствует проблема совмещения характерных участков, так как построение матриц производится с учетом всей иерархии построения потока и наряду с характерными точками числовой меры, учитываются точки с качественной мерой, фактически, отражающей спектральное содержание характерной точки.

В структурной матрице картина отношений будет характеризоваться некоторыми последовательностями характеристик $\chi_{i,j}$, которые фактически характеризует **конфигурацию отношений.** Без потери общности, возьмем общеизвестное множество отношений, состоящего из трех характеристик и пустого значения ⊗:

$$\chi \in \{\otimes,\ <,\ >,\ =\}$$

Обозначим $a_{i,j}$ элемент матрицы, который может принимать одну из символьных характеристик отношений элементов **i,j** и которую будем представлять в виде числового идентификатора от 0-3 соответственно.

В силу упорядоченности и отсутствия необходимости рассмотрения отношений объектов в обратном порядке, будем пользоваться только **верхней «бездиагональной» треугольной матрицей. «**Бездиагональная» верхнетреугольная матрица или просто верхнетреугольная матрица **(ВТМ)** – это квадратная матрица, у которой всегда все элементы ниже главной диагонали и по диагонали (0 диагональ) являются пустыми ($\mathbf{a_{ij}} = \otimes$**,** при $i \geq j$). Все остальные элементы могут иметь любое из указанного выше списка значений. При этом нулевое значение выше диагонали в

этом случае будет означать, что пока это отношение **не выведено,** или **не установлено.** Процесс **установления отношений** разбивается на три этапа:

- **структуризация** – получение отношений в виде примитивов:

- **выведение** некоторых отношений на основе полученных примитивов по правилам выведения отношений;

- явное **установление** отношений для отношений, не выводимых неявно, предполагающее процедуру явного сравнения двух количественных характеристик в структурированном потоке с последующим повторением **этапа выведения**.

Если в качестве идентификатора объекта отношений использовать их номера, а отношения между ними фиксировать, как некоторое значение на пересечении строк и столбцов, то можно представить всю совокупность отношений в виде **BTM** матрицы, на рис. 1.

i/j	**0**	**1**	**2**	**3**	**4**	**5**	**6**	…	**n-1**	**n**
0	⊗	a_{01}	a_{02}	$?_1$	$?_3$	$?_6$	$?_{10}$	…	•	•
1		⊗	a_{12}	a_{13}	$?_2$	$?_5$	$?_9$	…	•	•
2			⊗	a_{23}	a_{24}	$?_4$	$?_8$	…	•	•
3				⊗	a_{34}	a_{35}`	$?_7$	…	•	•
4					⊗	a_{45}	a_{46}	…	•	•
5						⊗	a_{56}	…	•	•
6							⊗	…	•	•
•						…			•	•
n-1									⊗	$a_{n-1,n}$
n										⊗

Рис.1. Индексация матрица отношений при её заполнении.

В данной матрице отношений имеется два вида нумерации: как элемент стандартной квадратной матрицы $a_{i,j}$ – матричная индексация, так и последовательную внутреннюю индексацию, которая указана в индексе вопроса

«**?**$_i$». Две диагонали вверх заполняются на основе примитивов, получаемых в процессе первичной структуризации. Таким образом, 1 и 2 диагонали будут заполнены изначально.

Необычный вариант внутренней индексации снизу вверх, а затем вправо и опять снизу вверх объясняется последовательностью продвижения по неизвестным значениям отношений, которое удобно использовать для **выведения** некоторых первоначально неизвестных значений элементов матрицы, исходя из предыдущих известных значений отношений.

Полностью заполненная матрица отражает качественную характеристику упорядоченной последовательности значений объектов или конфигурацию последовательности объектов безотносительно численного значения. Можно сказать, что это матрица полной связанности объектов (каждый с каждым) с установленными значениями таких отношений. Это весьма важная характеристика для определения подобия объектов без привязки к их размерности, что вообще то и соответствует распознаванию объектов! Содержательно, первая диагональ от центральной диагонали устанавливает отношения между соседними элементами. Вторая диагональ отношения между элементами, идущими через 1, далее третья через 2 и т.д. Строка матрицы отражает отношения со всеми далее идущими элементами характерных точек.

Фундаментальное значение такого представления состоит в том, что в практическом плане появляется **качественная мера**, которая дает возможность качественного сравнения различных объектов через сравнение таких матриц. Более того, при такой связанности объектов отношениями, число возможных

представлений матриц и соответственно отношений, как оказывается, весьма ограничено по отношению к возможным комбинациям заполнения матрицы. Содержательно, всё это отражает факт ограниченности качественных представлений связанных событий, и безграничность их численного представления. Именно эта ограниченность и распознавание через подобие защищает наше мышление от бескрайности фактического потока информации из окружающего нас мира.

В основе подхода при распознавании тона лежит установление и фиксация значимых изменений в совокупности значений амплитуд и установление отношений между этими характерными точками [9, 10]. Обрамляющие линии по локально экстремальным характерным точкам реализуется итерацией последовательностей отношений при схлопывании характерных точек в случае наличия промежуточного максимума или минимума между ними. Достаточное и необходимое количество итераций для выделения обрамляющей тона будет определяться на экспериментальном уровне. Предварительные эксперименты показали возможность такого получения обрамляющих линий и это открывают дорогу, как на распознавание фонем, так и на распознавание тональности в речи в разных её проявлениях.

Заключение

В работе предложен один из возможных подходов к решению задачи распознаванию тональности, прежде всего фонем языков Юго-Восточной Азии. В основе средства решения задачи предлагается использовать модель структурной матрицу и качественную меру при распознавании тональных фонем. Данный подход

проверен на практике, а основные алгоритмы реализованы и используется в экспериментальных исследованиях.

Библиографический список

1. Гусейнов А.Б., Маховых А.В. Структурно-параметрический синтез рационального бортового распознающего устройства в составе беспилотного летательного аппарата // Труды МАИ. 2016. № 90. URL: http://trudymai.ru/published.php?ID=74833

2. Аунг Вин, Балакирев Н.Е., Мью Ту Наинг, Щербаков А.И. Вопросы создания фонемной базы Мьянманского языка // Всесоюзная научно-техническая конференция «Новые материалы и технологии НМТ-2008». Сборник докдадов. (Москва, 11-12 ноября 2008) - М.: МАТИ, 2008. Т. 2. С. 146 - 148.

3. Nguyen V.L., Edmondson J.A. Tones and voice quality in modern northern Vietnamese: Instrumental case studies // Mon-khmer Studies Journal, 1998, vol. 28, pp. 1 - 18.

4. Нгуен Ван Хунг. Исследование и разработка алгоритмов и программ автоматического распознавания ограниченного набора команд вьетнамской речи. Автореферат диссертации на соискание ученой степени кандидата технических наук. - М.: МЭИ, 2010. - 20 с.

5. Сандакова Л.Л., Тюменева Е.И. Вьетнамский язык. Пособие по переводу. – М.: Восток-Запад, 2004. - 211 с.

6. Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением // Речевые технологии. 2008. № 3. С. 3 - 15.

7. Рабинер Л.Р. Скрытые марковские модели и их применение в избранных приложениях при распознавании речи // Труды института инженеров по электротехнике и радиоэлектронике. 1989. Т. 77. № 2. URL: https://b-ok.org/book/3079373/ed4973

8. Балакирев Н.Е. Логико-лингвистический подход по распознаванию содержания физических волн // Материалы XV Международной конференции «Информатика: проблемы, методология, технологии» (Воронеж, 12-13 февраля 2015). – Воронеж: ВГУ, Т. 1. С. 31 - 36.

9. Балакирев Н.Е. Количественная и качественная оценка исследуемых объектов на примере простейших отношений // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2016. № 2. С. 65 - 72.

10. Балакирев Н.Е., Нгуен Х.З., Малков М.А., Фадеев М.М. Структуризация и качественное рассмотрение звукового потока в системе синтеза и анализа речи // Программные продукты и системы. 2018. Т. 31. № 4. С. 768 – 776.

11. Grenander U. A Calculus of Ideas: A Mathematical Study of Human Thought, World Scientific, 2012, 219 p.

12. Grenander Ulf, Miller Michael. Pattern Theory: From Representation to Inference, Oxford University Press, 2007, 608 p.

13. Куприянов А.И., Шевцов В.В. Потенциальная чувствительность и дальность действия лазерного микрофона // Труды МАИ. 2012. № 55. URL: http://trudymai.ru/published.php?ID=30112

14. Самохин В.Ф., Мошков П.А. Экспериментальное исследование акустических характеристик силовой установки самолета «Ан-2» в статических условиях // Труды МАИ. 2015. № 82. URL: http://trudymai.ru/published.php?ID=58711

15. Балакирев Н.Е. Малков М.А. Метод идентификации голосового сообщения // Информационные технологии. 2008. № 12. С. 66 - 68.

16. Galunov V.I., Galunov G.V. Science perspectives of speech technology, SpeeCom, 2001, 302 p.

17. Разумихин Д.В. Использование нейронных сетей на уровне семантики в системе распознавания речи // IV Всероссийская конференция "Нейрокомпьютеры и их применение». Тезисы докладов. - М.: Радиотехника, 2001. - 288 с.

18. Soloviev A.N., Victorova K.O., Razumikhin D.V. About using non-informational functions in models of speech communication, International workshop "Speech and Computer" Proceedings SPb, Russian, 2002, pp. 27 – 31.

19. Пуртов И.С., Синча Д.П. Исследование методов и разработка алгоритмов обработки видеоинформации в задачах локализации положения беспилотного летательного аппарата на основе распознавания изображений при помехах и искажениях // Труды МАИ. 2012. № 52. URL: http://trudymai.ru/published.php?ID=29444

20. Королев В.О., Гудаев Р.А., Куликов С.В., Алдохина В.Н. Решение задачи распознавания типа объекта на основании использования диаграммы

направленности антенны в качестве признака // Труды МАИ. 2017. № 94. URL: http://trudymai.ru/published.php?ID=81109

Статья поступила 01.03.2019