УДК 004.934.2.

А.А. ШАРИПБАЕВ, А.К. БУРИБАЕВА, Г.Т. БЕКМАНОВА, А.К. КАЛИЕВ

ГЛАСНЫЕ ЗВУКИ КАЗАХСКОГО ЯЗЫКА И МЕТОДЫ ИХ ВЫДЕЛЕНИЯ В НАЧАЛЕ СЛОВА

Евразийский национальный университет имени Л.Н. Гумилева, Астана

В статье описан метод распознавания гласных звуков казахского языка в начале слова при помощи алгоритма DTW (Dynamic time warping). Это может быть использовано для ускорения распознавания, так как определение первого звука слова может существенно уменьшить список кандидатов-слов при распознавании. Также приведен акустический анализ казахских гласных звуков и показана их транскрипция при распознавании речи.

Ключевые слова: *распознавание речи, гласные звуки казахского языка, транскриптор, алгоритм DTW*

Автоматическое распознавание устной речи естественного языка является одним из актуальных направлений развития искусственного интеллекта и информатики в целом, так как результаты в этом направлении позволят решить проблему создания средств эффективного речевого взаимодействия человека с компьютером. С развитием современных речевых технологий появилась принципиальная возможность перехода от формальных языков-посредников между человеком и машиной к естественному языку в устной форме как универсальному средству выражения целей и желаний человека.

Исследованием проблемы распознавания речи уже более 50 лет занимаются специалисты нескольких научных областей. Достаточно эффективные методы выделения и сегментации гласных предлагают для английского, русского, арабского, сербского языков с использованием формантного анализа [1, 2, 3, 4].

Проведенный анализ литературы показал, что в настоящее время нет описания фонетического строя казахского языка, содержащего акустические характеристики звуков. Это необходимо для построения автоматического транскриптора, являющегося неотъемлемой частью системы распознавания речи. В связи с этим авторы ведут работу в этом направлении. Так, например, полный формантный анализ гласных казахского языка приведен в труде одного из авторов данной работы [5]. Также был разработан упрощенный транскриптор для распознавания речи [6] и реализованы алгоритмы пофонемного сегментирования казахской речи [7]. Однако в ходе работы выяснилось, что для надежного распознавания нужна более подробная транскрипция, которая приводится в этой статье. Выделение первого звука слова способствует ускорению процесса всего распознавания, уменьшая список кандидатов-слов для распознавания.

Акустический анализ гласных звуков

В казахском языке имеются 9 гласных и 19 согласных звуков. В алфавите на основе кириллицы они обозначаются следующим образом:

Гласные звуки: А, Ә, Е, О, Ө, Ұ, Ү, Ы, І, из них А, О, Ұ, Ы и Е являются фонемами, а Ә, Ө, Ү, І – аллофонами фонем А, О, Ұ, Ы;

Акустический анализ гласных звуков казахского языка основан на следующие сингармонические тембры:

А, Ы – твердые негубные сингармонические гласные звуки;

Ә, І – мягкие негубные сингармонические гласные звуки;

Ұ, О – твердые губные сингармонические гласные звуки;

Ү, Ө – мягкие губные сингармонические гласные звуки;

Е – мягкий негубной сингармонический гласный звук.

Система сингармонических признаков гласных звуков казахского языка приведена в табл. 1.

Таблица 1. Система сингармонических признаков гласных звуков.

Гласные звуки	Палатальный тембр		Лабиальный тембр	
	твердый	мягкий	негубной	губной
А	+	-	+	-
Ә	-	+	+	-
Ы	+	-	+	-
І	-	+	+	-
Ұ	+	-	-	+
Ү	-	+	-	+
О	+	-	-	+
Ө	-	+	-	+
Е		+	+	

Транскрипция гласных звуков казахского языка

Транскриптор реализован как программа, заменяющая одни символы другими в соответствии со следующими правилами подстановок, которые содержатся в управляющем файле:

1) #е=йе, #о=уо, #ө=уө;

2) #л=ыл, #р=ыр, #l=іl, #r=іr;

3) о^ы=о^ұ, ұ^ы=ұ^ұ, ө^і=ө^ү, ү^і=ү^ү,

4) ө^е=ө^ө, ү^е=ү^ө;

5) і^а=і^ә, ү^а=ү^ә, ө^а=ө^ә, ү^е=ү^ө;

6) а^у=а^ұу, о^у=о^ұу, ұ^у=ұ^ұу, ы^у=ы^ұу, ә^у=ә^үу, ө^у=ө^үу, ү^у=ү^үу, і^у=і^үу, у^а=ұу^а, у^о=ұу^о, у^ұ=ұу^ұ, у^ы=ұу^ы, у^ә=үу^ә, у^ө=үу^ө, у^ү=үу^ү, у^і=үу^і;

7) а^и=а^ый, о^и=о^ый, ұ^и=ұ^ый, ы^и=ы^ый, ә^и=ә^ій, ө^и=ө^ій, ү^и=ү^ій, і^и=і^ій, и^а=ый^а, и^о=ый^о, и^ұ=ый^ұ, и^ы=ый^ы, и^ә=ій^ә, и^ө=ій^ө, и^ү=ій^ү, и^і=ій^і; қи=қый, ғи=ғый, иқ=ыйқ, иғ=ыйғ.

Каждое правило подстановки состоит из двух частей, разделенных между собой знаком «=». Слева от этого знака стоят исходные символы буквенной записи слова, справа – символы, которыми они должны замениться в транскрипции.

Для транскрибирования заданного слова последовательно ищется в нем вхождение левой части очередного правила, и если таковое обнаруживается, то вместе неё подставляется правая часть этого правила.

В качестве транскрипционных знаков для гласных звуков использованы в основном соответствующие казахские буквы. Твердые казахские согласные транскрибируются также казахскими буквами, а соответствующие мягкие согласные – аналогичными латинскими буквами.

Знак «#» означает начало или конец слова в зависимости от местоположения: если «#» стоит перед символами, то это начало слова; если «#» стоит после символов, то это конец.

Знак «^» означает любые символы в любом количестве между двумя звуками.

Для удобства читателя в данном тексте правила разбиты на группы, которые занумерованы. Рекомендуется внести в управляющий файл эти группы в порядке номеров, не меняя порядка правил в группах, поскольку порядок замен, очевидно, важен.

Поясним правила транскрипции по каждому пункту:

1) в казахском языке если слово начинается на гласное «е», то при произношении перед ней слышится «й», если слово начинается на гласные «о», «ө», то при произношении перед ними образуется краткая вставка «у», например, «ет» – «йет», «он» – «уон», «өнер» – «уөнер».

2) если слово начинается на согласные «р» или «л», то при произношении перед этими звуками слышится гласные «ы», «і», в зависимости от твердости или мягкости согласных, здесь «r», «l» означает мягкие аналоги «р» и «л», например, «рас» – «ырас», «рет» – «ірет», «лас» – «ылас», «лезде» – «ілезде».

3) гласные звуки «ұ», «ү», «о», «ө» в начале или в первом слоге слова при произношении изменяют в следующих слогах гласные звуки «ы», «і» на гласные звуки «ұ», «ү» соответственно. Например, «қолтық» – «қолтұқ», «құлын» – «құлұн», «күлкі» – «күлкү», «көлік» – «көлүк»;

4) гласные звуки «ү», «ө» в начале или в первом слоге слова при произношении изменяют в следующих слогах гласный звук «е» на гласный «ө», например, «үлкен» – «үлкөн», «өнер» – «өнөр».

5) гласные звуки «ә», «ү», «і» в начале или в первом слоге слова при произношении изменяют в следующих слогах гласный звук «а» на его аллофон «ә», например, «ләззат» – «ләззәт», «діндар» – «діндәр».

6) при произношении дифтонга «у» в составе слова слышится «ұу», «үу», в зависимости от твердости или мягкости гласных в остальных слогах. Например, «туыс» – «тұуыс», «күту» – «күтүу».

7) при произношении дифтонга «и» в составе слова слышится «ый», «ій», в зависимости от твердости или мягкости гласных в остальных слогах. Например, «ине» – «ійне», «жина» – «жыйна». Если перед или после «и» идут согласные «қ», «ғ», то при произношении звука «и» всегда слышится «ый». Например, «қиын» – «қыйын», «қиғаш» – «қыйғаш».

Данный транскриптор можно использовать так и для синтеза речи.

Распознавание гласных звуков в начале слова

Определение начала речи

Для начала нужно надежно определить начало речи. Мы воспользовались алгоритмом В.Ю. Шелепова [8]. Опишем вкрадце этот алгоритм.

Используется 8-битная запись с частотой 22050 Гц. По нажатии кнопки записи записываются последовательные отрезки звука по 300 отсчетов (окна). Для каждого из них вычисляется отношение V/C, где

$$V=\sum_{i=0}^{298}\left|x_{i+1}-x_i\right|$$

- численный аналог полной вариации, C – количество точек постоянства, то есть таких моментов времени, что в следующий момент величина сигнала остается той же самой. Берется среднее этого отношения по первым 10 окнам. Назовем эту величину «текущий StartPorog». Она характеризует верхний порог «молчания». Ждем момента, когда этот порог будет превышен не менее 5 раз подряд. Возвращаемся на 20 окон назад (начальный запас) и, начиная с этого момента, заносим записываемые отсчеты в буфер 1. Тем самым начинается запись того, что мы предполагаем речью. Определим «текущий EndPorog» как пятикратный текущий StartPorog. Заполнение буфера 1 продолжается до момента, после которого величины V/C на протяжении 10 тысяч отсчетов будут меньше, чем текущий EndPorog. В него заносятся также упомянутые 10 тысяч отсчетов (запас в конце). Таким образом, запись предполагаемого речевого отрезка останавливается. Отметим, что при каждой записи вычисляются новые значения величин «текущий StartPorog» и «текущий EndPorog».

Записанное проверяется на наличие речи с использованием квазипериодичности [9]. Если наличие речи обнаруживается, содержимое буфера 1 передается в буфер 2.

Выбор признаков для распознавание по эталону

Остановимся на используемой нами при DTW-распознавании системе признаков [9]. В соответствующий буфер заносится 10 тысяч чисел:

$$y_1, y_2, \ldots, y_{10000} \tag{1}$$

значения напряжения на выходе микрофона в последовательные моменты времени (Эти моменты времени будем называть отсчетами). Сам ряд чисел (1) и соответствующую функцию

$$y(i) = y_i \tag{2}$$

будем называть сигналом. Таким образом, числа (1), в конечном счете, отражают изменение давления на мембрану микрофона как функцию времени. На экран монитора может быть выведен график сигнала, как функция времени (визуализация сигнала).

Сглаживанием сигнала мы называем обработку его 3-точечным скользящим фильтром

$$y_i = \frac{y_{i-1}+y_i+y_{i+1}}{3},\ i = 2,3,\ldots,9999 \tag{3}$$

Дальнейшая работа происходит с поточечной разностью исходного и десятикратно сглаженного сигнала. Это позволяет в некоторой степени "очистить" его от индивидуального тембра говорящего и тем самым сделать шаг в направлении дикторонезависимости системы распознавания. Далее, если не оговорено противное, под сигналом будем понимать указанную разность и, чтобы не усложнять обозначений, считать, что (1) и (2) соответствуют именно ей.

Пусть l – число отсчетов между двумя соседними локальными максимумами функции (2) (назовем сужение функции на соответствующий интервал полным колебанием). Если максимумы – не строгие, то под l будем понимать число отсчетов от начала первого максимума до начала второго. Определим величину z:

$$z=l \text{ при } 2 \le l < 20;\ z = 20 + \frac{l-20}{6} \text{ при } 20 \le l < 50;$$

$$z=25 + \frac{l-50}{10} \text{ при } 50 \le l < 90;\ z = 29 \text{ при } l \ge 90.$$

Ближайшее целое число, не превосходящее z, назовем длиной соответствующего полного колебания. Таким образом, длина полного колебания учитывается тем более точно, чем оно короче. Выделим участок сигнала и обозначим через n общее число полных колебаний на этом участке, через n_1 – число полных колебаний длины 2,...,через n_{28} – число полных колебаний длины 29.

Поставим в соответствие выделенному участку вектор

$$(x_1,\ldots, x_{28}, \varepsilon) \tag{4}$$

где $x_k = n_k/n$, $k = 1,2,\ldots 28$, ε – отношение амплитуды (разность наибольшего и наименьшего значений) рассматриваемого участка сигнала к амплитуде всего сигнала. Величина ε вводится для того, чтобы надежно отделить паузу от значащей части сигнала, а нормировка ее делается, чтобы отвлечься от громкости произносимого.

Разобьем записанный сигнал в 10 тысяч отсчетов на отрезки по 368 отсчетов в каждом (удвоенный квазипериод основного тона для мужского голоса средней высоты). Для каждого из 27-ми полных отрезков вычислим вектор (4). Последний неполный отрезок просто отбросим. В результате мы представляем сигнал в виде траектории, то есть последовательности 27-ми точек в 29-мерном пространстве:

$$A = (a_1, a_2,\ldots, a_{27}).$$

Выше было описано представление целого слова, а для определения гласного звука в начале слова достаточно трех первых векторов, то есть:

$$A = (a_1, a_2,\ldots, a_3).$$

Далее применяем для распознавания ставший уже классическим алгоритм Т.К. Винцюка, известный под названием алгоритма DTW (Dynamic time warping) [10].

Результаты тестирования

Для тестирования были отобраны по 17 слов каждому звуку с разными сочетаниями звуков. Так, например, для звука «ө» были выбраны слова: өкше, өнер, өгей, өшіргіш, өшпес, өзен, өзбек, өрік, өркеш, өлке, өжет, өсек, өсиет, өмір, өтірік, өтем, өбектеу. Во всех вариантах, кроме слова «өгей» программа надежно распознавала звук «ө».

В рисунках 2,3 представлены визуализация слова «өкше» и распознавание звука «ө» в ее начале.

Рис. 1. Визуализация слова «өкше»

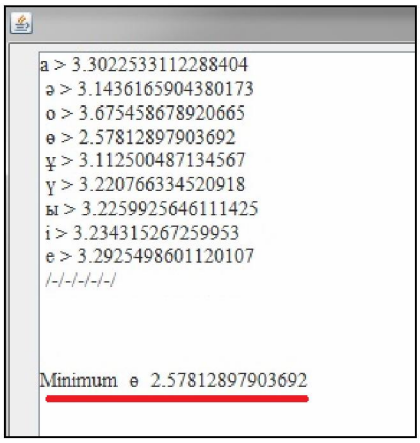

Рис. 2. Распознавание звука «ө» в начале слова

Результаты эксперимента представлены в таблице 2.

Таблица 2. Результаты эксперимента.

Звук	Распознавание
А	100%
Ә	88,23%
Ы	88,23%
І	88,23%
Ұ	88,23%
Ү	94,11%
О (уо)	94,11%
Ө (уө)	94,11%
Е (йе)	100%

Как видно из таблицы, звуки «а» и «е» распознаются без ошибок, при распознавании звуков «ұ», «о», «ө» программа ошиблась по одному разу. Со звуками «ә», «ү», «ұ» дела обстоят немного хуже, так как программа путает звук «ә» со звуком «а», а звуки «ұ», «ы», «і» между собой. Более надежного распознавания можно добиться при помощи дополнительного обучения, так как при каждом обучении эталоны звуков усредняются.

Выводы

Авторами данной работы были получены следующие результаты:

- сделан акустический анализ гласных звуков казахского языка;
- разработан и программно реализован транскриптор казахских гласных для распознавания речи;
- разработан и программно реализован алгоритм распознавания гласных звуков в начале слова.

Далее планируется:

- расширение и адаптирование транскриптора для слитной речи;
- разработка алгоритмов и программная реализация разпознавания согласных звуков казахского языка в начале слова;
- разработка алгоритмов и программная реализация распознавания слитной казахской речи на основе межфонемных переходов.

ЛИТЕРАТУРА

1. Iverson P., Smith Ch. A., Evans B.G. *Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration.* J. Acoust. Soc. Am., Vol. 120, No. 6., **2006.**
2. Сорокин В.Н., Цыплихин А.И. *Сегментация и распознавание гласных.* Информационные процессы, Том 4 , № 2, **2004.** стр. 202-220.
3. Alotaibi A., Hussain A. *Comparative Analysis of Arabic Vowels using Formants and an Automatic Speech Recognition System International.* Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 3, No. 2, **2010**.
4. Prica B., Ili´c S. *Recognition of Vowels in Continuous Speech by Using Formants.* Facta universitatis (ni s). Vol. 23, No. 3, **2010.** pages 379-393.

5. Yessenbayev Zh., Karabalayeva M. , Sharipbayev A. *Formant Analysis and Mathematical Model of Kazakh Vowels*, UKSim 14th International Conference on Computer Modelling and Simulation, London, **2012.**
6. Бекманова Г. Т. *Транскриптор казахских слов для распознавания речи*, Вестник НАН РК.– № 6. Алматы, **2009.** с. 12-17.
7. Карабалаева М.Х., Шарипбаев А.А, *Алгоритмы пофонемного распознавания казахской речи в амплитудно-временном пространстве*, II Всероссийская конференция «Знания – Онтологии – Теории», **2009.**
8. Шелепов В.Ю., Ниценко А.В. *Новый подход к определению границ речевого сигнала.* Проблемы конца сигнала. Речевые технологии, Москва, **2012**.
9. Шелепов В.Ю. *Лекции о распознавании речи.* Донецк, ПШШ, Наука і освіта, **2009.**
10. Винцюк Т.К. *Анализ, распознавание и интерпретация речевых сигналов.* Киев, Наук. думка, **1987.**

REFERENCE

1. Iverson P., Smith Ch. A., Evans B.G. *Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration.* J. Acoust. Soc. Am., Vol. 120, No. 6., **2006.**
2. Сорокин В.Н., Цыплихин А.И. *Сегментация и распознавание гласных.* Информационные процессы, Том 4 , № 2, **2004.** стр. 202-220.
3. Alotaibi A., Hussain A. *Comparative Analysis of Arabic Vowels using Formants and an Automatic Speech Recognition System International.* Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 3, No. 2, **2010**.
4. Prica B., Ili´c S. *Recognition of Vowels in Continuous Speech by Using Formants.* Facta universitatis (ni s). Vol. 23, No. 3, **2010.** pages 379-393.
5. Yessenbayev Zh., Karabalayeva M. , Sharipbayev A. *Formant Analysis and Mathematical Model of Kazakh Vowels*, UKSim 14th International Conference on Computer Modelling and Simulation, London, **2012.**
6. Бекманова Г. Т. *Транскриптор казахских слов для распознавания речи*, Вестник НАН РК.– № 6. Алматы, **2009.** с. 12-17.
7. Карабалаева М.Х., Шарипбаев А.А, *Алгоритмы пофонемного распознавания казахской речи в амплитудно-временном пространстве*, II Всероссийская конференция «Знания – Онтологии – Теории», **2009.**
8. Шелепов В.Ю., Ниценко А.В, *Новый подход к определению границ речевого сигнала.* Проблемы конца сигнала. Речевые технологии, Москва, **2012**.
9. Шелепов В.Ю. *Лекции о распознавании речи.* Донецк, ПШШ, Наука і освіта, **2009.**
10. Винцюк Т.К. *Анализ, распознавание и интерпретация речевых сигналов.* Киев, Наук. думка, **1987.**

Шәріпбаев А.Ә., Бөрібаева Ә.К., Бекманова Г.Т., Қалиев А.Қ.

ҚАЗАҚ ТІЛІНІҢ ДАУЫСТЫ ДЫБЫСТАРЫ ЖӘНЕ ОЛАРДЫ СӨЗДІҢ БАСЫНДА ТАНУ ӘДІСТЕРІ

Мақалада қазақ тілінің дауысты дыбыстарын DTW (Dynamic time warping) алгоритмінің көмегімен сөздің басында тану әдісі сипатталған. Бұл тануды жылдамдату үшін қолданылады, себебі сөздің алғашқы дыбысын анықтау сөйлеуді тану кезінде сөз кандидаттарының тізімін елеулі түрде азайтады. Сонымен қатар қазақ тілінің дауысты дыбыстарының акустикалық талдауы мен оларды транскрипциялау ережесі келтірілген.

Sharipbayev A.A., Buribayeva A.K., Bekmanova G.T., Kaliyev A.K.

KAZAKH VOWELS AND METHODS OF THEY RECOGNITION AT WORD BEGINNING

This paper presents the description of the method of Kazakh vowels recognition at word beginning using DTW (Dynamic Time Warping) algorithm. This can be used for acceleration of recognition since word's first sound identification can significantly decrease the list of words-candidates during recognition. Also the acoustic analysis of Kazakh vowels and their transcription during speech recognition are presented.