

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC

**«АҚПАРАТТЫҚ ЖӘНЕ ТЕЛЕКОММУНИКАЦИЯЛЫҚ
ТЕХНОЛОГИЯЛАР: БІЛІМ, ҒЫЛЫМ, ТӘЖІРИБЕ»
атты II Халықаралық ғылыми-тәжірибелік конференциясының
ЕҢБЕКТЕРІ**

Алматы, Қазақстан, 3-4 желтоқсан, 2015 жыл

I том

ТРУДЫ

II Международной научно-практической конференции
**«ИНФОРМАЦИОННЫЕ И ТЕЛЕКОММУНИКАЦИОННЫЕ
ТЕХНОЛОГИИ: ОБРАЗОВАНИЕ, НАУКА, ПРАКТИКА»,**

Алматы, Казакстан, 3-4 декабря, 2015 года

I том

THE PROCEEDINGS

Of the II International scientific - practical conference
**«INFORMATION AND TELECOMMUNICATION
TECHNOLOGIES: EDUCATION, SCIENCE AND PRACTICE»,**

Almaty, Kazakhstan, December 3-4, 2015

I volume

Алматы 2015

10.	Бисаринова А.Т., Мамырова А.К., Тусупова Б.Б., Иргалеев А.Б. Обоснование для применения ВЕБ-ГИС технологий разработки ГИС МЭОМ.....	137
11.	Билалов Б.О., Волобуева О.П. Разработка информационно-управляющей системы для АПК.....	139
12.	Божанов Е.Т., Тулешева Г.А., Мурзасанмова К.Д., Хойлан К. Об одной модели Б-6 относительного прогиба трубчатой конструкции из композита.....	143
13.	Боскебеев К. Д., Акматалиева Ж.З. Использование теории систем и CASE – технологий для моделирования информационных систем.....	147
14.	Большакова Н.А., Гусарова Н.М., Жубат К.Ж., Калжанов З.Б. Геостатистический анализ экологических данных с применением среды «R».....	152
15.	Букенова И.Н. Теоретические основы и методы снятия неопределенности при исследовании погрешностей исходных данных с применением имитационного моделирования.....	155
16.	Дадаева А.Н., Божанов Е.Т., Акимжанова Ш.А., Курмангали А. Об одной математической модели технологии усадки порошкового материала при неподвижной матрице.....	158
17.	Дауренбаева Н.А., Алданғарқызы А., Сыдыков Р.А. Принципы построения нейросистем на базе нейрочипа.....	163
18.	Dzalilov Z., Irzhanova A., Khadzhabergenova N., Amanzholova N.I., Kalizhanova A.U. Information and Imaging technologies for learning brain function.....	166
19.	Егоров С.В. Исследование параметров функций распределений, которые используются для оценки вероятности безотказной работы.....	171
20.	Еримбетова А.С., Мурзин Ф.А., Сагнаева С.К. Табиғи тілде сөйлемдердің байланысын талдау.....	175
21.	Есимова Ж.А., Сейдалиева Г.О., Сейдалиева Г.О. Математические модели и методы в логистике.....	179
22.	Ескендирова Д.М., Сыдыбаева М.А., Адилбекова А.А.. Роль искусственных нейронных сетей в распознавании образов.....	181
23.	Зейнуллина А., Мустафин С. Задача оптимизации и анализ данных.....	184
24.	Иманбаев К.С., Джанузаков С.Д. Некоторые особенности искусственных нейронных сетей.....	187
25.	Карабеков А.Б., Нуркаманова М.А., Санатбек С., Шекербекова Ш.Т. О применении онтологий в задачах построения интеллектуальных транспортных систем..	190
26.	Касымова Д.Т., Утепбергенов И.Т., Ескендирова Д.М. Деректердің үлкен көлемін өндеуге заманауи көзқарастар.....	194
27.	Киселева О.В., Балгабаева Л.Ш. Реализация информационной системы микро ГЭС с микропроцессорным управлением....	197
28.	Косников В.А., Сейдахметова К. Применение информационных и телекоммуникационных технологий в исследовании проблем и аспектов социального развития личности в Республике Казахстан.....	200
29.	Косынбай Е.Б., Тусупова Б.Б., Мамырова А.К. О мобильном приложении «Электронно-терминологический словарь «BilimDerekkor».....	202
30.	Курманкожаева А.А., Есенбеков К. Схема концепции математико-логического моделирования и принятия решений.....	206
31.	Курманкожаева А.А., Купарова А.Т. Математические методы в формальных языках.....	210
32.	Леменкова П.А. Пространственный анализ и геостатистическая интерполяция для моделирования распределения данных на поверхности.....	214
33.	Мамырбаев О., Калимолдаев М.Н., Турдалыгулы М, BagherBabaAli Методы построения системы многомодального распознавания речи.....	217
34.	Маслий Р.В. Классификация аэрофотоизображений с использованием случайных лесов.....	222
35.	Назирова А.Б. Сравнительная характеристика существующих геоинформационных технологий в геофизике.....	225

специальном окне построения картографических SQL. В ответ на каждый запрос получена новая грид-тема, показывающая удовлетворяющие запросу ячейки с указанными районами.

Построение гидрологических моделей требуют логически безошибочных данных, для проверки которых использовалась встроенная в состав модуля функция очистки данных от ошибок и пробелов. Ошибочными могут быть, например, отдельные провалы или выступы в ячейках грида, содержащих величины концентрации веществ. Их сравнение с преобладающим фоном, соседними значениями позволяет провести агрегирование данных или восполнение пробелов. Для оценки достоверности данных используется также модуль Geostatistical Analyst, который позволяет использовать геостатистический метод интерполяции. В отличие от детерминированного в Spatial Analyst, он опирается как на статистические, так и на математические методы, которые используются для оценки ошибки интерполяции, указывающей погрешность вычислений. Данные без указания погрешностей мало пригодны для экологического мониторинга, поэтому геостатистический метод используется для оценки точности измерений [3]. С использованием перечисленных групп функций модуля Spatial Analyst проводился анализ грид-тем пространственного распространения загрязнителей в пределах акваторий и их совместного использования с векторными данными. Результаты пространственного анализа и геостатистических методов выявляют основные источники загрязнения, участки акваторий с разным типом гидрологических обстановок и гидрохимическим загрязнением, значительно превышающим норму и используются для построения карт итогового экологического районирования.

На основе работы с модулем Spatial Analyst создаются информационно-справочные инвентаризационные карты загрязнения отдельных компонентов природной среды по типам загрязнителей, отражающие пространственное распределение загрязняющих веществ (тяжелых металлов Cd, Cr, Cu, Hg, Ni, Pb, Zn, радиоактивных элементов ^{137}Cs , ^{60}Co , $^{239,240}\text{Pu}$, ^{90}Sr и хлорорганических соединений ДДТ, ПХБ, ГХЦГ) по акватории Арктического бассейна в целом и по отдельным его акваториям (например, Баренцево и Печорское моря). Для серии карт на акваторию Арктики в целом оптимальное решение принять за базовую азимутальную стереографическую проекцию. Карты отдельных морей можно выполнить в конической равнопромежуточной проекции как имеющей оптимальное соотношение искажений площадей и углов объектов карты. Характер допустимых искажений обуславливается масштабом карт и не превышает допустимые нормы. В результате работы над проектом моделировано состояние экосистем, разработана методика ГИС картографирования с использованием системного подхода в процессе интерполяции геопространственных данных, создан обширная база тематических данных на основе слоев пространственной информации.

ЛИТЕРАТУРА

1. Зейлер М. Моделирование нашего мира. Руководство ESRI по проектированию базы геоданных – Нью-Йорк, ESRI Press, 1999. – 254 с.
2. Трофимов А.М., Редюков А.В. ГИС как инструмент пространственного анализа и моделирования. //ГИС для устойчивого развития территорий. Материалы международной конференции. ИнтерКарто - 8. СПб., 2002. С.142-146.
3. Савельев А.А., Мухарамова С.С., Пилюгин А.Г., Чижикова Н.А. Геостатистический анализ данных в экологии и природопользовании (с применением пакета R): Учебное пособие. – Казань: Казанский университет, 2012. - 120 с.

УДК 519.7, 519.97

**Оркен Мамырбаев¹, Максат Калимолдаев¹,
Муса Турдалыулы, BagherBabaAli²**

¹Институт информационных и вычислительных технологий
Казахстан, Алматы,

²College of Science, University of Tehran,
morkenj@mail.ru

МЕТОДЫ ПОСТРОЕНИЯ СИСТЕМЫ МНОГОМОДАЛЬНОГО РАСПОЗНАВАНИЯ РЕЧИ

Аннотация. В настоящее время совокупность данных поясняет, что хорошо спроектированная мультимодальная система, сочетающую в себе два или более источников информации может быть эффективным средством для увеличения распознавания. В данной статье эксплуатационные преимущества были продемонстрированы в различных комбинациях модальности (речи и движения губ), для различных задач

и в различных средах. В работе предложен двухмодальный метод распознавания казахской речи, использующий каналы передачи сигналов, как естественная речь и движение губы. Проведен сравнительный анализ с другими известными системами распознавания речи.

Ключевые слова Распознавание речи, Voiceactivitydetection, Сегментация речи, Многомодальные системы.

Введение. Разработка системы эффективного взаимодействия на естественном языке человека с компьютером сегодня является одним из приоритетных направлений развития области распознавания речи. Это связано с тем, что ресурсы вычислительной техники не используются в полной мере из-за отсутствия полноценного, привычного человеку, интерфейса для взаимодействия пользователя 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. Речевая форма диалога обладает рядом преимуществ, таких, как естественность, оперативность, смысловая точность ввода, освобождение рук и зрения пользователя, возможность управления и обработки в экстремальных условиях. Считая, что речь является наиболее естественной формой взаимодействия, многие разработчики пытались заменить ряд интерфейсов многих систем речевым интерфейсами и заставить компьютеров и пользователей выполнять все функции с помощью голоса.

Исследования, посвященные распознаванию речи, лица, положения человека в окружающем пространстве ведутся уже более полувека. Однако, системы объединяющие различные способы ввода информации в единой форме стали разрабатываться совсем недавно. Такие распознающие системы используют многомодальных (мультимодальных) методов распознавания речи. Многомодальные методы распознавания речи обрабатывают данные полученных из двух и более каналов ввода информации например, как речь и движения губы. Китлер (Kittler) и другие ученые [1, 2] объединили вероятностные классификаторы Байеса в одну систему для распознавания речи. В такой системе учитываются данные полученные из трех каналов - речь, фронтальная и профильная части лица. Чибулуши (Chibelushi) и другие [3, 4,5] разработали метод комбинированного распознавания речи, состоящие из речи и изображений лица.

Созданная нами система рассматривает двухмодальный (речь и губы) метод распознавания. По первому каналу поступает речевой сигнал от микрофона, а из второго канала поступает сигнал от видео камеры Kinect, описывающий движение губы. Разработанная многомодальная система, получая информацию из двух каналов совместно обрабатывая их, позволяет более качественно распознавать речь. В результате, установлено, что качество распознавания речи на основе двухмодального подхода выше, чем качество, полученное от отдельно взятой речи, учитываемое в системах.

Освоенности казахского языка для распознавания речи. Звуковой строй казахского языка стал предметом изучения еще с XIX века. Казахский язык, как один из тюркских языков, относится к кыпчакской группе. Тюркология в последнее время сделала заметный поворот к изучению закономерностей, лежащих в сфере речи. Обращение к фактам разговорной речи обнаруживает ряд синтаксических примеров, которые не только интересны по организации и функции, но и расширяют наши представления о соотношении между языком и речью, о возможностях речевой реализации общезыковых моделей, схем.

Фонетическую систему языка образуют не только сегментные средства, но и суперсегментные, которые накладываются на фонемную (линейную) структуру речи. Минимальными единицами звуковой системы языка являются фонемы (гласные, согласные), являющиеся элементами звуковой оболочки слов и морфем. Фонемы составляют сегментный уровень речевого звучания и определяют линейности речи, то вторую линию речевого звучания составляет суперсегментный уровень, т. е. интонация. Интонационный (просодический) уровень имеет достаточно сложное строение и выполняет весьма многообразные функции. Просодические средства в каждом языке функционируют в виде тональных, динамических и темпоральных модификаций и изменений одновременно с сегментами речевого потока.

Звуки казахского языка по их артикуляционной близости разделены на две категории: восемь звуков отнесены к группе гласных (а -э, е, ы, і, о, ө, ұ, ү), девятнадцать звуков – к согласным (п, б, м, у, н, д, т, ж, з, ш, с, р, л, ј, к, г, қ, ғ, ң). Переднеязычный широкий э не выделяется как самостоятельная фонема. В составе согласных отсутствуют аффрикаты дж, ч и спирант h.

Приведем отдельные сведения о некоторых артикуляционных особенностях казахских звуков, комбинаторных чередованиях их (б~м, н~б~т, н~ң, к~г и т. п.) и ударении выделяются твердые и мягкие варианты согласных.

В составе казахского вокализма отмечает девять гласных (а, ә, е, о, ө, ұ, ү, ы, і), в консонантизме – 20 согласных звуков. Согласные классифицируются им по четырем группам: глухие (қ, к, т, п, с, ш), звонкие (ғ, г, д, б, з, ж), сонорные (н, ң, м, р, л, л') и промежуточные (у, й). По мнению ученых, не совсем обоснованно выделены: у, й из состава других сонорных. Представляется неверным также выделение мягкой и твердой разновидности сонорного л в качестве самостоятельных фонем[6].

Состояния губ, соответствующие фонемам устной речи, называются виземами. Предварительный анализ возможности автоматической классификации образцового алфавита показал необходимость их существенного сокращения в направлении использования базовых или опорных визем. Поэтому, для дальнейших исследований в направлении разработки системы автоматического чтения с губ можно принять следующий рабочий алфавит визем, за основу которого приняты опорные виземы, показанные в рисунке 1.

1	2	3	4	5	6	7
						
а, ә /a/, /æ/	ж, ш /jo/, /ʃ/	ы, ң /w/, /ŋ/	д, т, н, р, л /d/, /t/, /n/, /r/, /l/	е /e/	с, з /s/, /z/	і /i/
8	9	10	11	12	13	
						
о, ө /o/, /ø/	у, ұ, ү /w/, /ʊ/, /y/	б, п, м /b/, /p/, /m/	й /j/	в, ф /v/, /f/	г, ғ, қ, к /g/, /ɣ/, /q/, /k/	

Рис. 1. Состав визем казахского языка

Предварительная обработка речевого сигнала. Voiceactivitydetector (VAD) – является технология сжатия речевого сигнала, за счет поиска речи и пауз и их кодирования. Алгоритм VAD работает в процессе кодирования речевого сигнала перед распознаванием речи. Наличие пауз в речевом сигнале определяется на основе анализа и синтеза речевых данных. Предполагаем что, речь содержит паузу, которую необходимо предсказывать определения паузы является самым сложным аспект в алгоритме VAD (Рисунок 2). Алгоритм VAD работает на основе вычисления значений

кратковременной энергии $E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2$ (или кратковременное значение модуля энергии).

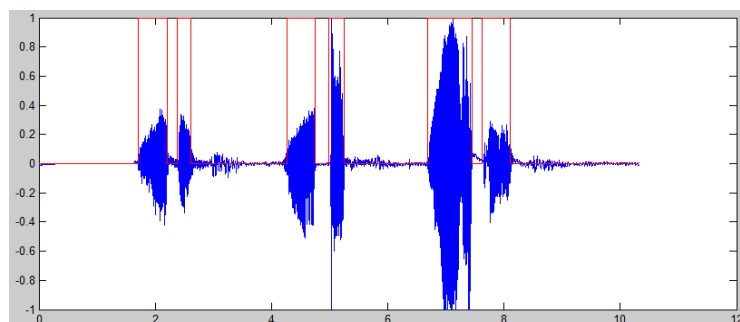


Рис. 2. Результаты определения паузы в речевом сигнале с помощью VAD

В наиболее простых реализациях наличие паузы в наборе цифровых отсчетов определяется на основе сравнения суммарной энергии пакета речевых данных с некоторым пороговым значением. В таком случае необходимо подобрать порог так, чтобы не допустить слишком часто устранение ошибочных пауз, т.к. это может привести к ухудшению качества и потере полезных данных, что может послужить снижению эффективности алгоритма VAD. Обычно, для определения пауз, применяется сложный алгоритм, учитывающий не только энергию пакета, но и энергию спектральных составляющих отрезка сигнала.

1. Проверим числа переходов через нуль слева от точки с минимальной энергией, чтобы точно определить точку границы между слогами это будет точное разделение двух слогов, соответствующих двум словам (рисунок 3).

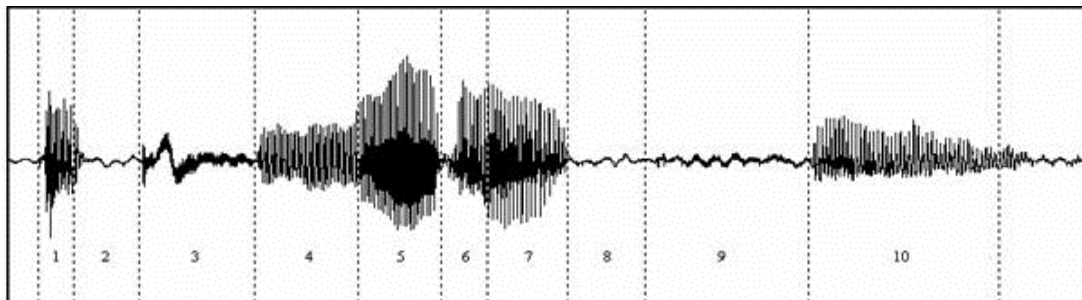


Рис. 3. Сегментация речевого сигнала

Распознавание движения губы. Для распознавания движения губы мы используем активные модели внешнего вида (Activeappearancemodel, AAM) [7]. В первую очередь распознаются вся область лица диктора, а потом выделяет контур губы для распознавания движения губы, что бы улучшить распознавания речи с чтением движения губы (Рисунок 4).

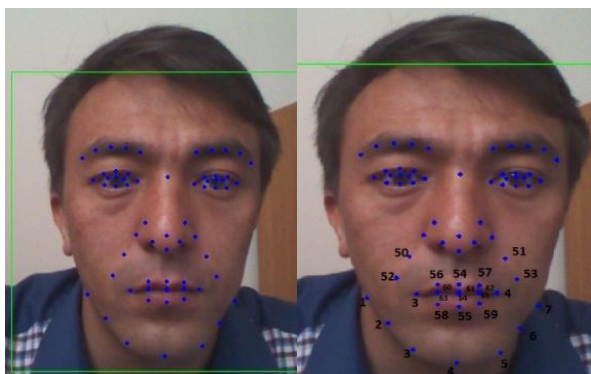


Рис. 4. Распознавание лица и губы

При выделении контура губы и их составляющие точки по оси x и y формируют матрицу $F = [f_1, f_2, \dots, f_n]$, где $f_n = [x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n]$ и форма губы имеет следующий вид:

$f = f_0 + M_n p_n$, где f_0 - нормальная форма губы, M_n - матрица главных векторов, p_n - параметры формы. Для быстрого и надежного отслеживания губы, мы предлагаем метод слежения Lukas-Kanede feature tracker [15]. Изменяя вектор параметров p_n можно получить разного рода деформации формы для сопоставления ее с виземами. После сопоставления мы получаем множество Q_{lip} .

Распознавание лица и чтение движения губы осуществляются с помощью алгоритма FaceSDK и для видеозаписи мы использовали камеру Kinect model 1517.

Заключение. Для анализа результатов системы, мы рассмотрим 10 дикторов, показанных в таблице 1, среди которых 5 дикторов мужчины и 5 дикторов женщины. Каждый диктор говорит по 200 слов с разными сложностями. По результатам данного анализа показателей системы распознавания речи мы получили следующий график (Рисунок 6):

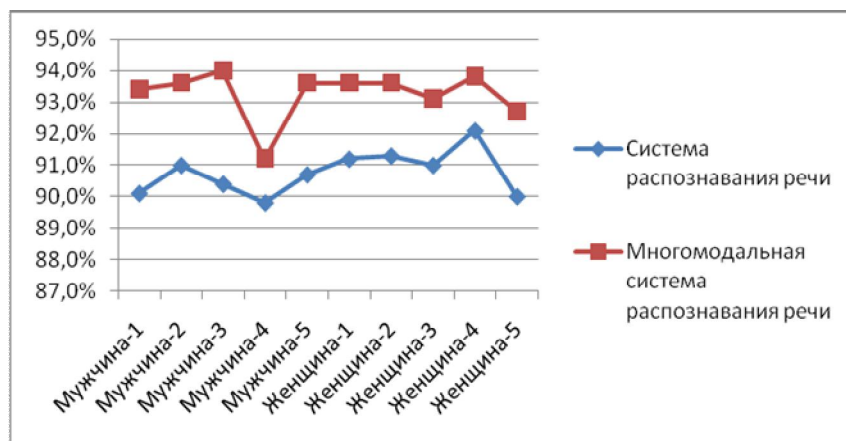


Рис. 6. Результаты сравнительного анализа системы распознавания речи

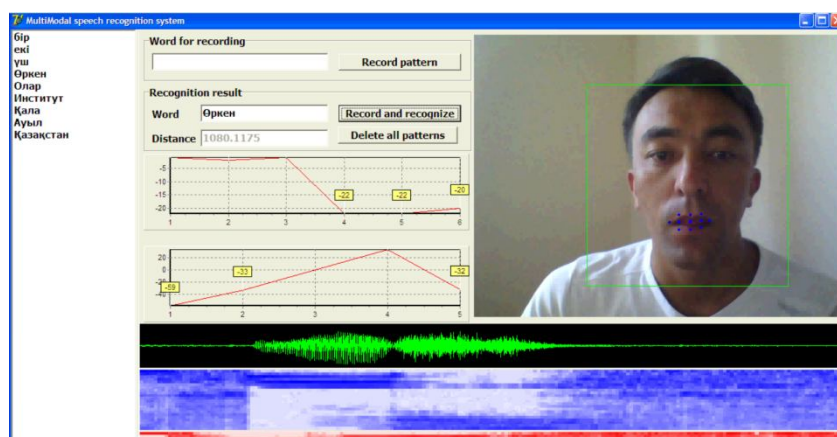


Рис. 7. Интерфейс многомодальной системы распознавания речи

Созданную систему можно предложить в автомобильную навигационную систему и для удобного использования людьми с ограниченными возможностями. До сих пор методы чтения губы пока не были эффективно использованы в режиме реального времени из-за трудности отслеживания губы. В этой работе предлагаем, что системы чтения губ может эффективно применяться в качестве опоры для существующей системы распознавания речи. Мы предлагаем многомодальную систему распознавания речи в качестве эффективного инструмента для распознавания казахской речи.

ЛИТЕРАТУРА

1. Jain A.K., Ross A., Prabhakar S., An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14 (2004) P. 4–20.
2. Raghavendra R. (2010). Robust Algorithms for Person Verification using Multimodal Biometrics, Ph.D Thesis, Feb.
3. Raghavendra R., Ashok Rao, Hemantha Kumar (2010). Multisensor Biometric Evidence Fusion of Face and Palmprint for Person Authentication using Particle Swarm Optimization (PSO), International Journal of Biometrics, Vol.2, No.1, P. 19–33.
4. Sanderson C., & Paliwal K. (2003). Noise compensation in a person verification system using face and multiple speech features, Pattern Recognition 36 (2), P. 293–302.
5. Lanitis A., Taylor C., & Cootes T. (1997) Automatic interpretation and coding of face images using flexible models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), P.743–756.
6. Aralbayeva J.A. (1970). Vowels Kazakh language. The Science, P.178–181.
7. Kalimoldayev M. N., Amirgaliyev Ye., Mussabayev R. R., Mamyrbayev O. J. (2014). Methods of forming a dictionary vizem for multimodal speech recognition. X- International Conference, Vol.35, No. 1, P. 60–65.