ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

Программа анализа интонационных конструкций русского языка

Исупова Татьяна Дмитриевна, магистр
Куликов Александр Алексеевич, студент
Шаманин Андрей Олегович, студент
Кацай Дмитрий Алексеевич, кандидат технических наук, доцент
Южно-Уральский государственный университет (г. Челябинск)

В настоящий момент существует большое множество образовательных ресурсов, у которых в наличии имеется обучение с родного языка на необходимый пользователю иностранный. Одна из сложных задач изучения русского языка как иностранного (РКИ) связана с освоением интонационных конструкций. Для решения этой задачи разработан программный комплекс, содержащий рабочее место по подготовке эталонов речевых образцов и виртуальный кабинет с web-интерфейсом для изучения интонационных конструкций русского языка.

Данная разработка является дополнением к образовательному мультимедийному тренажеру по обучению фонетике русского языка, созданному сотрудниками и студентами Южно-Уральского государственного университета [1, с. 492].

В данной работе рассматриваются структура и основные функции программы для исследования интонационных признаков речевых образцов. Программа анализа речевого образца основана на алгоритме вычисления мел-кепстральных коэффициентов, которые используются для формирования значений парной корреляции с эталоном речевого образца. Программа обеспечивает выполнение следующих функций:

- вычисление частотных и мел-кепстральных характеристик участка речевого образца;
- визуальное отображение кепстрального портрета выделенного сегмента звука;
- численное и графическое сопоставление нового речевого образца с записанным эталоном,
- поиск в речевом образце участков максимального сходства с эталоном.

В состав программы входит предварительная обработка входного сигнала, включающая в себя алгоритм обнаружения речевой активности (voice activity detection, VAD) [6, с. 102].

Человеческая речь находится в среднечастотном звуковом диапазоне, в пределах от 300 до 3400 Гц, так как, в основном, форманты, определяющие ее разборчивость, находятся именно в этой полосе частот [3, с. 20]. Опираясь на данную информацию, в программу были установлены параметры использования заданных частот, с целью избежать обработки лишнего неинформативного шума, не относящегося к речи.

Перед сравнением записанные образцы следует привести к единой метрике с эталонными образцами произношения, записанными филологами. Эталонные звуковые файлы записаны преподавателями в рекомендованных оптимальных условиях с использованием профессионального оборудования.

Для нелинейного сопоставления кадров использовался принцип алгоритма динамической трансформации временной шкалы (dynamic time warping, DTW) [4, с. 69-84], встроенный в программу. Данный алгоритм предназначен для измерения подобия между двумя временными последовательностями разной длины по заданным ограничениям и правилам. Например, это используется для установления соответствий между двумя речевыми образцами разной длины, но, предположительно, представляющими одно и то же слово, произнесённое с разной скоростью.

При разработке программного обеспечения возникла задача подбора оптимальной ширины кадра, применяемого к входным образцам звука. Здесь требуется подобрать баланс между разрешением по времени и разрешением по частоте: меньший кадр предоставляет большее разрешение по времени, но меньшее по частоте – и наоборот. Слишком большое дробление на кадры приводит к неоправданным вычислительным затратам, уменьшает информативность по частоте и, как следствие, сокращает количество доступных коэффициентов для анализа в дальнейшем. Наоборот, слишком широкие кадры могут охватывать сигнал, не являющийся условно-стационарным на выделенном промежутке времени, и не учитывать динамику сигнала. Также важным является выбор оптимального перекрытия между кадрами, позволяющего сгладить краевые эффекты от оконного преобразования.

Исходя из поставленных задач, было проведено экспериментальное тестирование по определению оптимальных заданных параметров для обработки входных звуковых сигналов. По результатам тестирования, представленного в таблице 1, было установлено, что при уменьшении ширины кадра, с определённого момента перестаёт расти информативность, начинает увеличиваться размер шума и искажений, особенно при сравнении образцов.

В тесте использовалось 3650 звуковых файлов. На основе результатов был установлен оптимальный размер ширины кадра 2048 сэмплов для частот дискретизаций 44.1 и 48 кГц. Это соответствует рекомендациям о выборе ширины кадра в районе 25-40 мс, поскольку мы можем считать сигнал на этом промежутке условно-стационарным, а именно не меняющим (или меняющим незначительно) свои характеристики от начала к концу кадра.

К каждому кадру применялась «оконная функция Хэмминга» $□(□) = 0{,}54 − 0{,}46 \cdot \cos(2□□/□ − 1)$ [5, с. 1-7] для выравнивания значений по его границам (рисунок 1).

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

Таблица 1. **Результаты программы при изменении ширины кадра без использования перекрытия**

Ширина кадра, семплов	Максимальное сходство с эталоном, %	Минимальное сходство с эталоном, %	Среднее сходство с эталоном, %	Правильное определение интонации, %
8192	93,24	78,32	85,67	62,34
4096	90,12	75,43	85,12	67,12
2048	87,51	68,24	85,33	78,32
1024	89,42	73,14	84,56	74,21
512	79,54	54,21	77,18	73,82
256	75,74	52,18	73,24	71,94

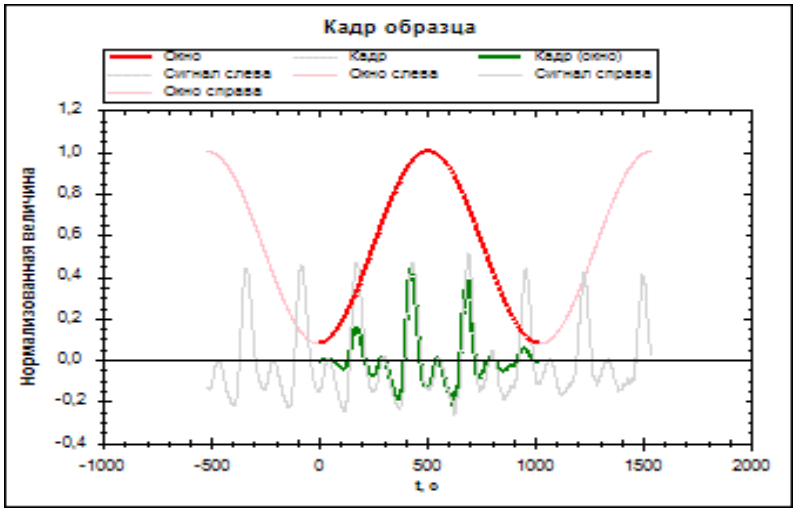

Рис. 1. **Окно Хэмминга, применённое к звуковому сигналу**

В поставленном эксперименте в записи звуковых образцов участвовало 18 дикторов в возрасте от 20 до 35 лет, из них восемь носителей русского языка (4 мужского и 4 женского пола) и десять носителей китайского языка (5 мужского и 5 женского пола). По результатам были отобраны наиболее качественные записи для участия в эксперименте в равных количествах. Результаты представлены в таблице 2.

Таблица 2. **Выборки речевых образцов, отобранных для участия в эксперименте, распределенные по интонации и типу носителя языка**

Носители языка	Интонация	Количество речевых образцов	
		По интонационным характеристикам	По типу носителей
Русского	Вопросительная	600	1200
	Повествовательная	600	
Китайского	Вопросительная	1225	2450
	Повествовательная	1225	

Для выявления мел-частотных коэффициентов применялись треугольные оконные фильтры [7, с. 117-121]. Разрешение по частоте является вдвое меньшим, чем размер кадра. При установленном размере кадра в 2048 семплов частотный спектр будет составлять 1024 значений, и, опираясь на ранее установленную методологию расчетов [2, с. 1-4], по заданным параметрам максимальным количеством получаемых характеристик является 50 мел-частотных коэффициентов.

Следующим этапом идёт экспериментальное определение процентного отношения перекрытия кадров (таблица 3) для уменьшения искажения на краях кадров с помощью сглаживания [8, с 267-278].

Таблица 3. **Результаты программы при параметризации размера перекрытия**

Ширина перекрытия, %	Максимальное сходство с эталоном, %	Минимальное сходство с эталоном, %	Среднее сходство с эталоном, %	Правильное определение интонации, %	Время сравнения, с
90	92,64	78,98	90,23	86,96	16,4
75	92,32	78,87	89,70	86,65	4,1
50	92,51	79,18	90,43	87,20	2,6
25	90,21	78,40	88,16	85,45	2,3
12,5	88,15	76,87	83,66	84,32	1,9

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

По результатам сравнительного анализа оптимальным оказалось перекрытие в половину кадра.

Реализация алгоритма идентификации была выполнена в среде разработки Visual Studio 2019, языком программирования был выбран C#.

Результаты работы программы можно увидеть на рисунках 2-5.

Разработанная программа даёт приемлемый результат в распознавании интонационных признаков в произношении слов. В будущем алгоритм программы можно связать с обучением нейронной сети по выборкам из звуковых фалов для повышения результата тестирования. Данное ПО можно модульно подключить к мультимедийному тренажеру по обучению фонетике русского языка на базе Южно-Уральского государственного университета. Подобный сервис может служить инструментом как для преподавателей, занимающихся обучением иностранных студентов, так и для иностранных студентов, изучающих русский как неродной.

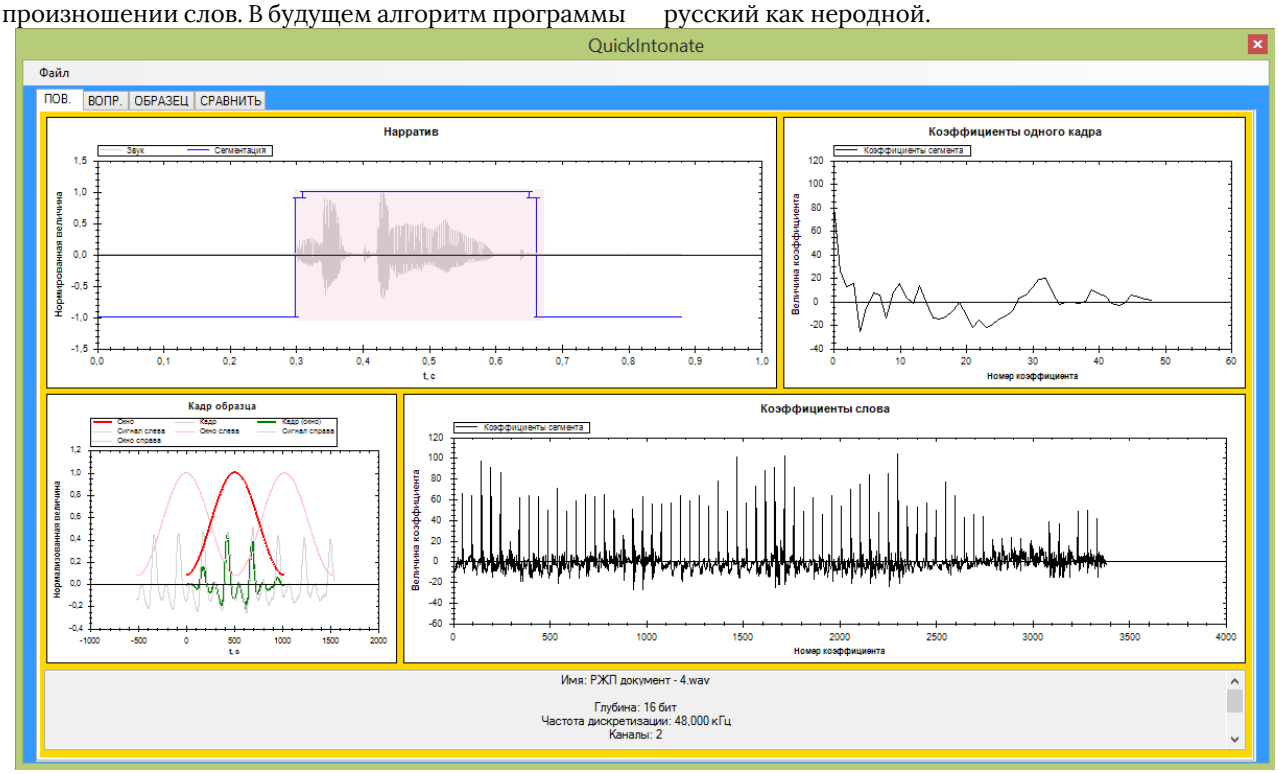

Рис. 2. **Загрузка образца русскоязычного диктора женского пола повествовательной интонации (РЖП) слова «документ» в программу**

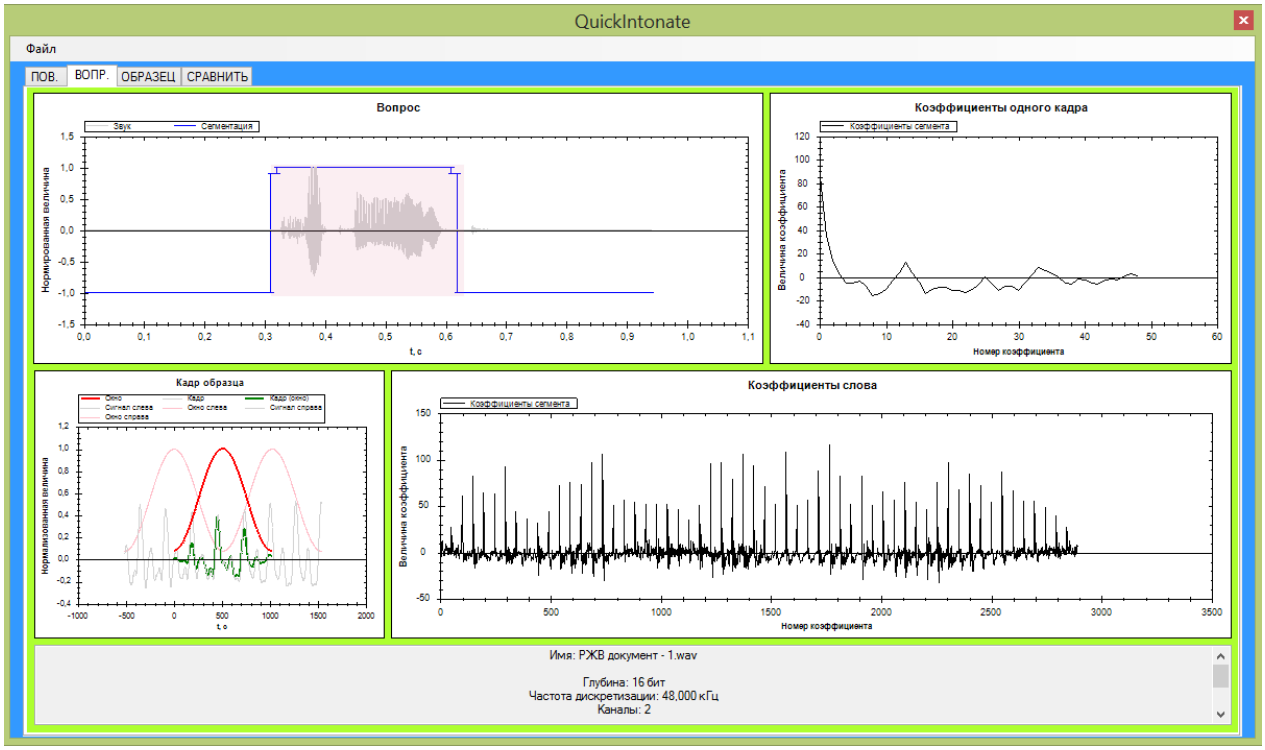

Рис. 3. **Загрузка образца русскоязычного диктора женского пола вопросительной интонации (РЖВ) слова «документ» в программу**

ЕВРАЗИЙСКОЕ НАУЧНОЕ ОБЪЕДИНЕНИЕ
EURASIAN SCIENTIFIC ASSOCIATION
www.esa-conference.ru

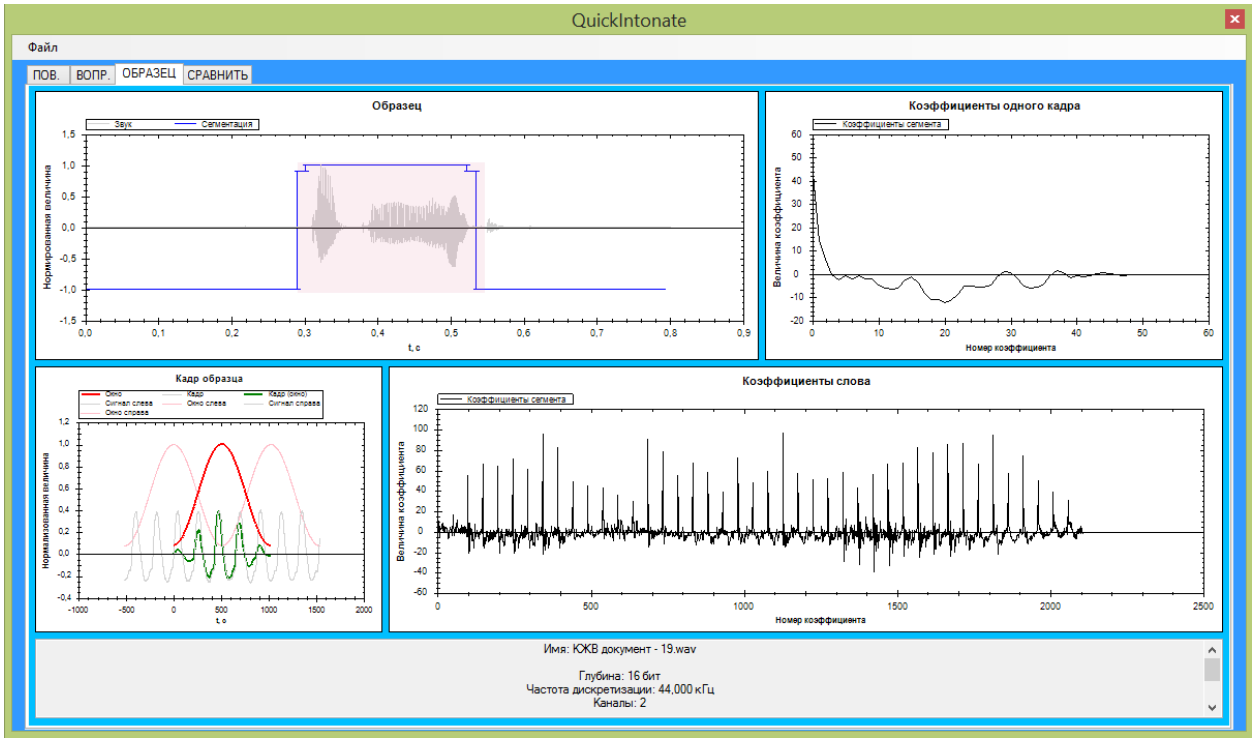

Рис. 4. **Загрузка образца китайского диктора женского пола вопросительной интонации (РЖВ) слова «документ» в программу**

Рис. 5. **Результат программы по оценке распределения коэффициентов корреляции с помощью использования мел-кепстральных коэффициентов**

Литература:

1. Кацай, Д.А. Структура тренажера для совершенствования речи для иностранных студентов, изучающих иностранный язык / Д.А. Кацай, Т.Д. Исупова, Е.В. Харченко, А.Д. Бирюков, Д.С. Филиппов // Сборник материалов международной научно-практической интернет-конференции (Москва, 27 ноября – 1 декабря 2017 г.) «Актуальные вопросы описания и преподавания русского языка как иностранного/неродного». – М.: Государственный институт русского языка им. А.С. Пушкина, 2018. – С. 492–496.

2. Bhadragiri J.M. Speech recognition using MFCC and DTW / J.M. Bhadragiri, B.N. Ramesh // 2014 International Conference on Advances in Electrical Engineering (ICAEE). – 2014. – P. 1–4.

3. ITU-T P.310–2000. International telecommunication union. Telecommunication standardization sector of ITU. – Geneva: ITU, 2001 – P. 20.

4. Muller M. Dynamic time warping. / M. Muller // In Information Retrieval for Music and Motion. – 2007. – Chap. 4. – P. 69–84.

5. Podder P. Comparative Performance Analysis of Hamming, Hanning and Blackman Window. / P. Podder, T. Zaman, M. Haque // International Journal of Computer Applications. – 2014. – V. 65, №18. – P. 1–7.

6. Ramachandran, R. Modern Methods of Speech Processing / R. Ramachandran, R. Mammone. – Springer Science & Business Media, 2012. – 102 p.

7. Sirko M. Computing mel-frequency cepstral coefficients on the power spectrum / M. Sirco, M. Pitz // Lehrstuhl fur Informatik VI, Computer Science Department – 2002. – V. 65, №2. – P. 117–121.

8. Trethewey M.W. Window and overlap processing effects on power estimates from spectra / M.W. Trethewey // Mechanical Systems and Signal Processing. – 2000. – V. 12, № 2. – P. 267–278.