

Ю. А. ЗДОРОНОК

ЛИНГВИСТИЧЕСКИЕ И АКУСТИЧЕСКИЕ РЕСУРСЫ КОМПЬЮТЕРНОЙ СИСТЕМЫ АНАЛИЗА И ИНТЕРПРЕТАЦИИ РЕЧЕВОЙ ИНТОНАЦИИ

Белорусский национальный технический университет

В данной статье предлагается новый подход к распознаванию высказываний на родном и иностранном языке на основе супraseгментарных явлений, которые составляют интонацию синтагмы. В каждом языке речь представляет собой разнообразное количество супraseгментных явлений, которые работая вместе, создают определенную окраску звучащей речи, интонационно выдавая при этом эмоциональную вовлеченность или отстраненность диктора. Любому языку, присущ разнообразный вариант интонационных контуров, которые впервые были определены как таковые Е. А. Брызгуновой, и используются в учебном пособии И. В. Одицовой при обучении русскому языку как иностранному. Согласно данным семи базовым интонационным конструкциям Е. А. Брызгуновой, Б. М. Лобановым были созданы универсальные мелодические портреты (УМП), на базе которых возможна нормализация выражения во времени в соответствии с положением его предъядерной, ядерной и заядерной части. С помощью интонации передается информация на лингвистическом и паралингвистическом уровне. От правильного интонирования зависит многое: восприятие, понимание и смысл. Мысль, облаченная в слова с правильной звуковой волной, рождает в сознании человека правильное представление о сути вещей. Таким образом, исключается так называемая интерференция при изучении иностранного языка. Сегодня существует большая потребность в новых и инновационных компьютерных программах, которые обеспечивают изучение иностранного языка с учетом просодической структуры целевого языка. Представленная работа описывает лингвистические и акустические ресурсы, которые лежат в основе обучения интонации с помощью компьютера, а также алгоритм, согласно которому возможно проектирование компьютерной системы для анализа и интерпретации речевой интонации.

Ключевые слова: компьютер, система, обучение языкам, лингвистические и акустические ресурсы.

Введение

Практически владеть иностранным языком означает сформировать навыки и развить умения мыслить на иностранном языке и понимать мысли других людей. Для того, чтобы сообщение было воспринято и понято слушателем, слова и словосочетания, предложения должны быть донесены в правильном интонационном оформлении. Это значит, что внутреннее или внешнее проявление речи должно быть произнесено с присущей ей динамической звуковой окраской в соответствии с правилами данного языка. С помощью интонации передается информация на лингвистическом и паралингвистическом уровне. От правильного интонирования зависит многое: восприятие, понимание и смысл [1].

Так или иначе, в обучении просодии иностранного языка необходима палитра интонационных контуров, чтобы передать все разноо-

бразие мыслей в речи. Поэтому интонацию следует изучать в контексте хорошо-структурированных диалогов или дискурса/речи.

Речь – это универсальное средство общения. Речь включает процессы порождения и восприятия (приема и анализа) сообщений для целей общения на всех языках мира, где ведущей является мысль или мыслительный образ, который реализуется в речи посредством акустических инструментов.

Предложение является совокупностью грамматически и фонетически оформленной реализацией человеческой мысли в речи. Как известно, предложение обладает определенными фонетическими характеристиками: мелодика речи, фразовое ударение, темп, ритм, паузация и тембр. Фраза – это самая крупная фонетическая единица с законченной интонацией. Минимальным элементом членения фразы является акцентная единица: совокупность слов

с главной выделенной вершиной – «ядро» вокруг которого концентрируются сопутствующие ему элементы «предъядро» и «заядро» [2].

Основная часть

1. Универсальные мелодические портреты акцентных единиц

ПАЕ-модель (модель портретов акцентных единиц) была предложена более 20 лет назад [3] и с тех пор успешно использовалась в нескольких системах синтеза русской речи по тексту. В соответствии с ПАЕ-моделью, минимальной просодическим компонентом, из которого составляется интонация синтагмы, является Акцентная Единица (АЕ). АЕ может состоять из одного или более фонетических слов, но должна иметь в своём составе только один полноударный слог. Каждая АЕ, в свою очередь, состоит из ядра (полноударная гласная фонема), предъядра (все фонемы, предшествующие полноударной гласной) и заядра (все фонемы за полноударной гласной). Или иначе – согласно терминологии Брызгуновой – «центр», «предцентр» и «постцентр». ПАЕ-модель предполагает, что для определенного типа интонации топологические свойства мелодического контура АЕ не зависят от количественного и качественного содержания предъядра, ядра и заядра.

ПАЕ-модель обеспечивает возможность представления семи интонационных конструкций Брызгуновой – $\{ИК_i\}$ – в виде набора их Универсальных Мелодических Портретов (ЭМП) в нормированных координатах «Частота – Время» – $\{УМП ИК_i\}$.

Нормализация по времени осуществляется путём приведения к стандартной длине элементов АЕ: предъядерных, ядерных и заядерных участков. Этот вид нормализации устраняет различия мелодической кривой, связанные с количественным составом предъядерных и заядерных участков АЕ.

Для нормализации по частоте определяются минимальное ($F_0 \min$) и максимальное ($F_0 \max$) значения частоты основного тона – F_0 для всего ансамбля мелодических кривых $\{ИК_i\}$ в произнесении данного диктора. Нормализация осуществляется в соответствии с формулой: $F_{0N} = (F_0 - F_{0\min}) / (F_{0\max} - F_{0\min})$. Этот вид нормализации устраняет различия мелодической кривой, связанные с индивидуальными дикторскими различиями в высоте голоса.

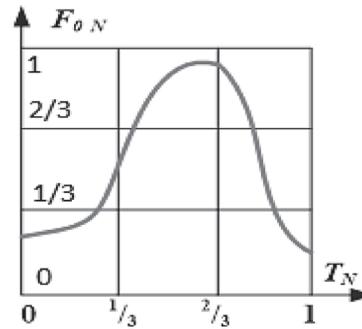


Рис. 1. Общий вид УМП ИК

Таким образом, нормированное пространство для отображения УМП $ИК_i$ может быть представлено в виде прямоугольника с координатными осями (T_N, F_{0N}) , как это представлено на схематическом рис. 1. При этом интервалам на оси абсцисс соответствуют: $[0-1/3]$ – предъядро, $[1/3-2/3]$ – ядро, $[2/3-1]$ – заядро. Интервалам на оси ординат соответствуют: $[0-1/3]$ – низкий уровень тона, $[1/3-2/3]$ – средний, $[2/3-1]$ – высокий.

Целью настоящей работы является экспериментальная проверка эффективности представления $ИК_1 - ИК_7$ в виде набора Универсальных Мелодических Портретов – $\{УМП ИК_i\}$ – в условиях варьирования количественного и качественного состава произносимых фраз, а также в условиях произнесения этих фраз женским и мужским голосами.

2. Структура лингвистических и акустических ресурсов

В настоящем исследовании нам предоставляется возможным провести анализ на промежуточном лингвистическом уровне для того, чтобы выявить супraseгментную вариативность, характеризующую иностранную речь на примере ситуационных диалогов «Situational Dialogues» Michael Ockenden [4], которые включают:

1. 44 повседневных ситуаций, каждая из которых содержит четыре диалога на аутентичном английском языке;

2. Все диалоги состоят из 1051 предложений, в том числе 704 утвердительных предложений, 325 вопросительных предложений и 22 восклицательных предложения;

3. Ситуации подходят тем, кто учится или путешествует в Англии. Они включают в себя тематические разделы: еда, развлечения и путешествия, а также фразы более общего характера, как приветствие, жалобы и извинения.

Каждый диалог структурировался по:

- 1) исполнению (диктор: мужчина и/или женщина),
- 2) количеству участников диалога (мужчина-женщина, женщина-женщина, мужчина-мужчина),
- 3) типу предложений: вопросительные, утвердительные и восклицательные,
- 4) количеству фраз в предложении,
- 5) количеству акцентных единиц во фразе.
- 6) определению предъядра, ядра, заядра.

В своей работе «Советы иностранным ученикам» А. С. Гимсон подчеркивает необходимость изучения «английского языка с использованием нисходящего и восходящего тона, чтобы обозначить настроение говорящего, так как чрезмерное использование восходящего тона создает непреднамеренное впечатление неуверенности или жалобы, а слишком частое использование нисходящего тона создает невольный эффект пренебрежительного отношения к оппоненту или ситуации» [5].

По этой причине обработка акустических материалов проводилась в соответствии со следующими критериями интонации:

- А Нисходящий тон;
- Б. Восходящий тон;
- В Нисходящий восходящий тон.

(А) Нисходящий тон. Здесь высота голоса падает с высокого до низкого тона на один ударный слог. Нисходящий тон используется в кратких утверждениях, вопросах, начинающихся с вопросительного слова, а также в вопросительных предложениях, когда оратор более чем уверен в правильности высказывания. Нисходящий тон используется в повелительных или восклицательных высказываниях.

(Б) Восходящий тон. Здесь высота голоса поднимается на последнем ключевом слове во фразе или на безударных слогах после ударного слога в слове. Восходящий тон используется для заявлений, предназначенных для того, чтобы поощрять ответы на вопросы, начинающиеся с вопросительных слов, когда оратор хочет проявить какой-то особый интерес и то, что он говорит, является правильным. Восходящий тон используется в предложениях, заканчивающихся на «пожалуйста», «извините», «спасибо» или в предложениях, чтобы выразить благодарность за просьбу (передать соль и т. д.).

(В) Нисходящий восходящий тон. Здесь высота голоса падает на самую значимую часть предложения и снова поднимается. Нисходящий восходящий тон используется для того, чтобы высказать извинение или некое суждение.

Согласно грамматическим правилам английского языка существуют два типа часто используемых вопросительных предложений: общие и специальные вопросы. В приведенных ниже статистических данных (табл. 1 и 2) показано, как часто используются вопросительные предложения в задействованных в исследовании диалогах.

Таблица 1. Статистика: Общие вопросы

Предложение начинается с ...	Количество предложений
Is	16
Are	12
Am	1
Was	2
Were	1
Will	6
Do	26
Does	4
Did	2
Have	21
Need	1
Has	4
Can	23
Could	17
Shall	2
May	2
Must	1
Would	16
ВСЕГО	140

Таблица 2. Статистика: Специальный вопрос

Вопросительное слово	Количество вопросов
What	66
When	13
Why	2
Where	14
Who	2
How	52
Which	8
Whom	0
Whose	0
Wherefore	0
Whatever	0
Wherewith	0
Whither	0
Whence	0
However	0
ВСЕГО	157

Табл. 3 (английский язык) и табл. 4 (русский язык) дает нам четкое представление того, как варьируется минимальное (F_{0min}) и максимальное (F_{0max}) значения частоты основного тона – F_0 для всего ансамбля мелодических кривых $\{ИК_i\}$ в произнесении диктора, носителя английского языка и для диктора, носителя русского языка. Это делает очевидным тот факт, что высота движения частоты основного тона в английском языке выше высоты движения частоты основного тона в русском языке.

Таблица 3. показывает минимальное (F_{0min}) и максимальное (F_{0max}) изменение значения частоты основного тона в английском языке

Тип интонации	Утверждение		Специальный вопрос		Общий вопрос	
	min	max	min	max	min	max
F_0 [Hz]						
Образец 1	92	184	100	330	109	280
Образец 2	90	180	98	280	98	286
Образец 3	100	230	60	235	101	252
Образец 4	105	230	65	232	99	211
Среднее значение	96.75	206	80.75	268.25	101.75	257.25
Диапазон	2.13		3.34		2.53	

Таблица 4. показывает минимальное (F_{0min}) и максимальное (F_{0max}) изменение значения частоты основного тона в русском языке

Тип интонации	Утверждение		Специальный вопрос		Общий вопрос	
	min	max	min	max	min	max
F_0 [Hz]						
Образец 1	80	147	85	154	85	170
Образец 2	78	150	85	155	91	196
Образец 3	81	144	80	155	84	185
Образец 4	82	146	83	157	84	185
Среднее значение	80.25	146.75	83.25	157.5	86.25	182.25
Диапазон	1.83		1.89		2.11	

Это делает очевидным тот факт, что высота движения частоты основного тона в английском языке выше высоты движения частоты основного тона в русском языке.

Результаты исследовательской работы над лингвистическими и акустическими ресурсами выявили следующее: алгоритм, согласно которому происходит оценивание интонационных конструкций слушателем, является сходным при овладении готовыми речевыми образцами. При подкреплении звуковой волной готовых речевых оборотов будет создаваться матричный

алгоритм, благодаря которому оценивается и ритм, и структура предложенных фразовых конструкций без погружения в более сложные языковые конструкции. Возможно, данный подход может выявить отсутствие многовариантности, но, в данном случае, ограниченность необходима, так как в ограниченных условиях у изучающего иностранный язык сохраняется правильно озвученный мыслительный образ изучаемых фраз, которые беспрепятственно могут быть идентифицированы в процессе общения.

3. Алгоритм компьютерной системы для обучения интонации речи

Компьютеры использовались для изучения языка с 1960 года прошлого века. Однако в настоящее время важность использования интонационных аспектов речи при преподавании иностранного языка и при создании компьютерных систем для анализа и синтеза речи имеет такое же значение, как 50 лет назад.

Для разработки интеллектуальной системы обработки речевых сигналов необходим строгий алгоритм анализа интонации, который является составляющей компонентой компьютерной обучающей системы [8].

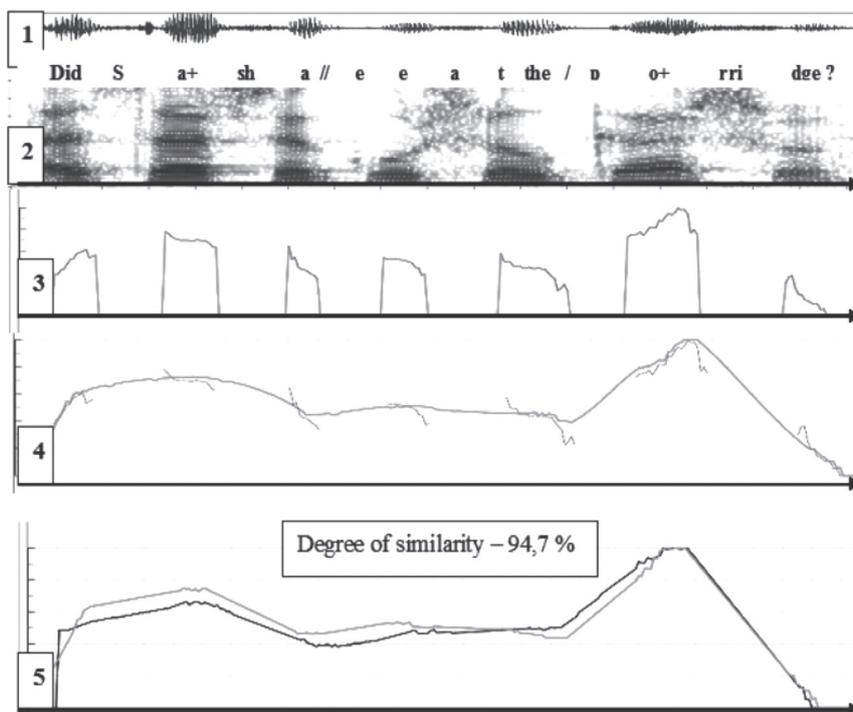
Основное назначение системы – предоставить учащемуся возможность компактного и наглядного отображения результатов анализа мелодического и энергетического контуров фраз с интонацией различного типа, а также получить визуальную, слуховую и численную оценку качества обучения интонации иностранной речи (см. рис. 2).

В БД эталонных сигналов хранятся образцы фраз различных интонационных конструкций. Для каждой из эталонных фраз проводится предварительная просодическая разметка, включающая указание фразовых границ и положение ядерных гласных. Учащийся, в соответствии с предлагаемой методикой обучения интонационным навыкам, выбирает нужные фразы, прослушивает и произносит их. При этом осуществляется их запись в буфер сигнала и производится описываемая ниже обработка эталонного и произнесённого сигналов.

Эталонный и произнесённый сигналы подвергаются спектральному анализу, затем методом динамического программирования (ДП)



Рис. 2. Структурная схема системы анализа и интерпретации интонации речи

Рис. 3. Иллюстрация обработки сигналов речи: 1 – осциллограмма, 2 – спектр, 3 – F_0 (до обработки), 4 – F_0 (после интерполяции и медианного сглаживания), 5 – сравнение мелодических кривых $F_0(t)$ эталонной и произнесённой фразы

осуществляется их сопоставление, перенос просодических меток и разметка произносимых фраз. Далее производится расчёт просодических параметров фраз – частоты основно-

го тона F_0 и энергии сигнала A_0 . Над этими параметрами осуществляется операции интерполяции на неголосовых участках, медианное сглаживание и их нормировка.

На рис. 3 представлена иллюстрация произнесённой 2-х акцентной вопросительной фразы: «А Саша кушал кашу?»: её осциллограмма, спектр и пример обработки частоты основного тона – F_0 , в результате которой получается искомая непрерывная мелодическая кривая фразы $F_0(t)$. Аналогичная обработка осуществляется для получения энергетической кривой $A_0(t)$.

Заключение

Компьютерное обучение интонации предназначено специально для оценки и улучшения произношения на изучаемом языке. Благодаря компьютеризированной системе обучения интонации конкретные ошибки произношения будут определяться на уровне слова или подслова, что дает возможность улучшить произношение в соответствии с визуальной обратной связью.

Литература

1. **Language** and speaker specific implementation of intonation contours in multilingual TTS synthesis / B. Lobanov [et al.] // *Speech Prosody: proceedings of the 3-rd International conference, Dresden, Germany, May 2–5, 2006*. Dresden, 2006. Vol. 2. P. 553–556.
2. **Лобанов, Б. М.** Опыт создания мелодических портретов сложных повествовательных предложений русской речи / Б. М. Лобанов // *Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 1–4 июня 2016 г.)*. Вып. 15 (22) / гл. ред. В. П. Селегей. – М.: РГГУ, 2016. – С. 382–392.
3. **Лобанов, Б. М.** Универсальные мелодические портреты интонационных конструкций русской речи / Б. М. Лобанов, Т. И. Окрут // *Компьютерная лингвистика и интеллектуальные технологии: Материалы ежегодной Международной конференции «Диалог»*. Вып. 13 (20). – М.: Изд-во РГГУ, 2014. – С. 330–339.
4. **Окенден М.**, (2005) *Ситуационные диалоги* // Английский центр, Истборн / П. – Лонгман – С. 98.
5. **Гимсон А. С.**, *Введение в произношение английского языка*, Лондон, 1966, С. 261.
6. **Брызгунова Е. А.** *Интонация* // *Русская грамматика* / – Наука, М., 1982. 785 с. (С. 96–122).
7. **Лобанов, Б. М.** Прототип компьютерной системы обучения речевой интонации / Б. М. Лобанов, В. А. Житко // *Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems: материалы междунар. науч.-техн. конф. Вып.1 (Минск, 16–18 февраля 2017 г.)*. / редкол.: В. В. Голенков (отв. ред.) [и др.]. – Минск: БГУИР, 2017. – С. 163–166.
8. **Лобанов, Б. М.** Компьютерная система анализа и интерпретации интонации речи / Б. М. Лобанов, В. А. Житко, Ю. А. Здоронко // *Международный конгресс по информатике: информационные системы и технологии = International Congress on Computer Science: Information Systems and Technologies / БГУ; под ред. С. В. Абламейко*. – Минск: БГУ, 2016. – P. 589–594.

References

1. **Language** and speaker specific implementation of intonation contours in multilingual TTS synthesis / B. Lobanov [et al.] // *Speech Prosody: proceedings of the 3-rd International conference, Dresden, Germany, May 2–5, 2006*. Dresden, 2006. Vol. 2. P. 553–556.
2. **Lobanov B.** Comparison of Melodic Portraits of English and Russian Dialogic Phrases // *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialogue»*. Iss. 15 (22). M.: RSHU, 2016. P. 382–392.
3. **Lobanov B., Okut T.** Universal Melodic Portraits of Intonation Patterns of Russian Speech // *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialogue»*. Iss. 13 (20). M.: RSHU, 2014. P. 330–339.
4. **Ockenden M.** *Situational Dialogues* // The English Centre, Eastbourne; Revised Edition. Longman, 2005.
5. **Gimson A. C.** *Introduction to the Pronunciation of English*, London, 1966, p. 261.
6. **Bryzgunova E.** *Intonation* // *Russian grammar* // Science. M., 1982. P. 96–122.
7. **Lobanov B.** A Prototype of the Computer System for Speech Intonation Training / B. M. Lobanov, V. A. Zhitko // *Open Semantic Technologies for Intelligent Systems: Papers from the Annual International Conference. Vol.1 (Minsk, 16–18 February 2017)*. / edited by.: V. V. Golenkov. – Minsk: BSUIR, 2017. – P. 163–166.
8. **Lobanov, B. M.** Computer-based System of Analysis and Interpretation of Speech Intonation / B. M. Lobanov, V. A. Zhitko, Y. A. Zdaranok // *International Congress on Computer Science: Information Systems and Technologies / BSU; edited by. S. V. Ablameiko*. – Minsk: BSU, 2016. – P. 589–594.

Поступила
20.09.2017

После доработки
25.10.2017

Принята к печати
15.12.2017

Zdaranok Y. A.

LINGUISTIC AND ACOUSTIC RESOURCES OF THE COMPUTER-BASED SYSTEM FOR ANALYSIS AND INTERPRETATION OF SPEECH INTONATION

Belarussian National Technical University

This article describes a novel approach to discriminating native and nonnative utterances based on suprasegmental features that constitute the intonation of the syntagma. In each language speech presents a diverse number of suprasegmental phenomena, which working together, create a certain coloring of spoken speech, intonationally giving out the emotional involvement or detachment of the speaker. To any language, there is a variety of intonational patterns, which were first identified as such by E. A. Bryzgunova, and are used for teaching Russian as a foreign language by I. V. Odintsova. According to the seven basic intonational patterns of E. A. Bryzgunova, BM Lobanov created the universal melodic portraits (UMP), which can be normalized in time in accordance with the position of its pre-nuclear, nuclear and bender parts. With the help of intonation, information is transmitted at the linguistic and paralinguistic level. A lot depends on correct intoning: perception, understanding and meaning. A thought covered in words with the right acoustic pronunciation gives a correct impression of things in human mind. Thus, the so-called interference while studying a foreign language is excluded. Linguistic and acoustic resources provide the basis for computer-aided intonation training in and outside the classroom. There is a great need for new and innovative computer-aided programs. But there is no necessary software available for such teaching system and, therefore, such system appears to be of great relevance. Presented work describes as well the algorithm of designing of the computer-based system for analysis and interpretation of speech intonation.

Keywords: *suprasegmental Parameters, Intonation, Pronunciation, Linguistic and Acoustic Resources, Computer Pronunciation Training.*



Здоронок Юлия Александровна, аспирант кафедры прикладной лингвистики Белорусского государственного университета (специальность 10.02.21 – Прикладная и математическая лингвистика). Окончила Минский государственный лингвистический университет (2012), магистратуру МГЛУ (2013). С 2014 года являюсь преподавателем кафедры «Межкультурная профессиональная коммуникация», ФММП, БНТУ. Научные интересы: прикладная лингвистика, программирование.

Zdaranok Y. A., postgraduate student of Belarussian State University, department of Applied linguistics (specialization 10.02.21 – Applied and mathematical linguistics). Graduated from Minsk State Linguistic University (2012), Master of Education (2013). Since 2014 faculty member of the Department of Intercultural professional communication of FMMP, BNTU. Scientific interests: Applied linguistics, software engineering. E-mail: yuliyazdaranok@gmail.com.

Работа выполнена в рамках диссертационного исследования на соискание степени кандидата наук по специальности «Прикладная и математическая лингвистика».