

ТЕХНОЛОГИЯ КЛОНИРОВАНИЯ ИНТОНАЦИИ

«Единственная известная мне роскошь — это роскошь человеческого общения», писал некогда великий мыслитель своего времени Антуан Сент-Экзюпри. Любому общению предшествует мысль. Мысли представляют собой энерголучи. Все мысли выглядят одинаково и формируются в мыслительном аппарате, которые затем ретранслируются через кору головного мозга и являются лингвистически оформленными, таким образом, приобретают форму того языка, на котором происходит процесс речевого взаимодействия между людьми, т.е. сам процесс общения.

Голос является самым выразительным инструментом общения. Устная речь имеет двойственную природу: семантическую и акустическую. Звуковая волна передает от говорящего к обращенному в слух оппонента целый комплекс смысловой информации экстралингвистического характера. Вариативность речи проявляется в ее информационной, кодовой и модуляционной структуре. Речь зависит от персональных акустических характеристик голоса человека, прежде всего от анатомической особенности строения и функционирования элементов речевого аппарата человека (гортань, голосовые связки, глотка, полость рта, зубы и т.д.), а также от динамических особенностей взаимодействия колебаний голосовых связок и резонаторов речевого аппарата.

Попытки имитации персональных характеристик голоса человека в системе преобразования текста в речь на основе моделирования физиологических и акустических процессов речеобразования приводят к тому, что создаются близкие копии компьютерного существа. При этом не менее важной остается задача сохранения личностных акустических и фонетических особенностей голоса человека, его произношения и акцента, учитывающих такие характеристики как мелодика, ритмика, динамика. Сходным с генетикой образом, в компьютерном «клонировании» персонального голоса и речи происходит клонирование акустических характеристик голоса человека. В основе технологии клонирования акустических особенностей человека закладываются акустические особенности голоса одного человека, фонетические особенности произношения другого, а интонация или совокупность просодических характеристик — третьего. Синтезированная речь уступает по своему звучанию человеческой речи, а также аффлектирует восприятие, поэтому разумно использование отрезков натуральной речевой волны в качестве минимального исходного материала для клонирования голоса. Выбор выпадает на позиционно комбинаторные варианты фонем — аллофоны или мультиаллофоны.

Методика эксперимента

Для реализации эксперимента технологии клонирования интонации в качестве основы были взяты интонационные конструкции Е.А.Брызгуновой,

которые используются в качестве методического материала при обучении русскому как иностранному. Особенностью классификации интонационных конструкций (ИК) является характер движения тона на ударном и прилегающих к нему слогах. Е.А. Брызгунова выделяет *семь* [2] основных *интонационных конструкций* русского языка, различающих смысл звучащих предложений. Б.М. Лобанов более 20 лет назад предложил модель портретов акцентных единиц, которая до сих пор актуальна [3]. Основу портретов мелодических единиц составляет минимальный просодический компонент, из которого составляется интонация синтагмы, Акцентная Единица (АЕ). В своем составе 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». Модель портретов акцентных единиц предполагает, что для определенного типа интонации топологические свойства мелодического контура АЕ не зависят от количественного и качественного содержания предъядра, ядра и заядра.

Семь основных интонационных конструкций, согласно классификации Е.А. Брызгуновой, можно реализовать в виде мелодических портретов в контексте таких нормированных координат как частота и время.

Проведение эксперимента осуществляется в несколько этапов.

1. На начальном этапе проводится расшифровка и коррекция текста и фонограммы. При необходимости убираются ошибочно произнесенные слова и звуки, помехи, а также происходит корректировка амплитудно-частотных характеристик звуков.

2. Путем последовательного прослушивания происходит разбиение фонограмм и текста на синтагмы.

3. После происходит просодическая маркировка синтагм. Каждый ударный слог маркируется с помощью следующих маркеров: сильное (+) и слабое (=) или ударение отсутствует. Слова, в которых ударение отсутствует, объединяются в единое фонетическое слово. Затем слова со слабым ударением объединяются в единую Акцентную единицу (АЕ) с одним из слов с сильным ударением и в тексте указываются границы АЕ.

Система **PhonoClonator** [5] реализует для каждого из аудио сигналов эталонную ИК, при которой осуществляется автоматическая сегментация и разметка аудиосигнала на последовательность аллофонов фонем и питчей (периодов основного тона) в соответствии с текстом анализируемой фразы. Пример пользовательского интерфейса ФоноКлонатора при обработке этих фраз приведен на рисунке 1.

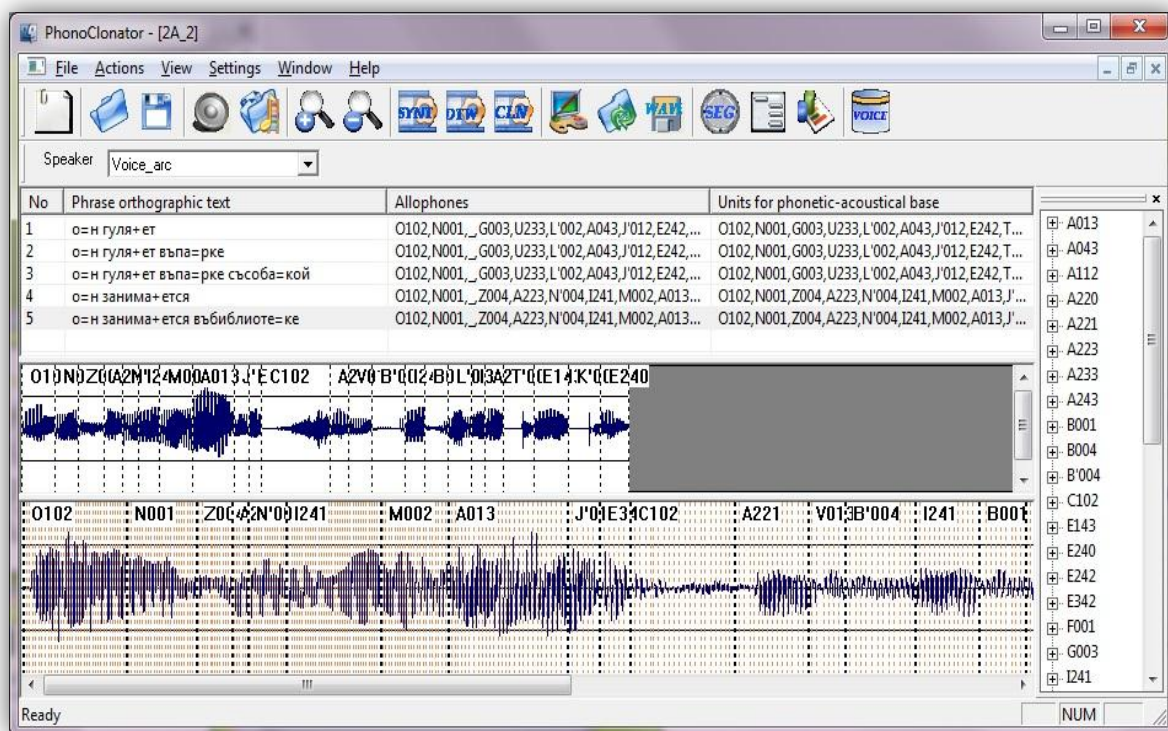


Рис. 1. PhonoClonator.

Далее размеченные ФоноКлонатором аудио сигналы подаются на вход системы **IntoClonator** [5], с помощью которого отображаются границы ядра, предъядра и зядра, мелодическая кривая и кривая интенсивности сигнала (см. рисунок 2).

Следует отметить, что предварительно устанавливаются 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. В приведенном на рис. 2 примере для фразы с ИК1 «Он заним¹ается в библиотеке», произнесённой диктором женщиной выбраны: $F_{0\min} = 115$ Гц, а $F_{0\max} = 420$ Гц [5].

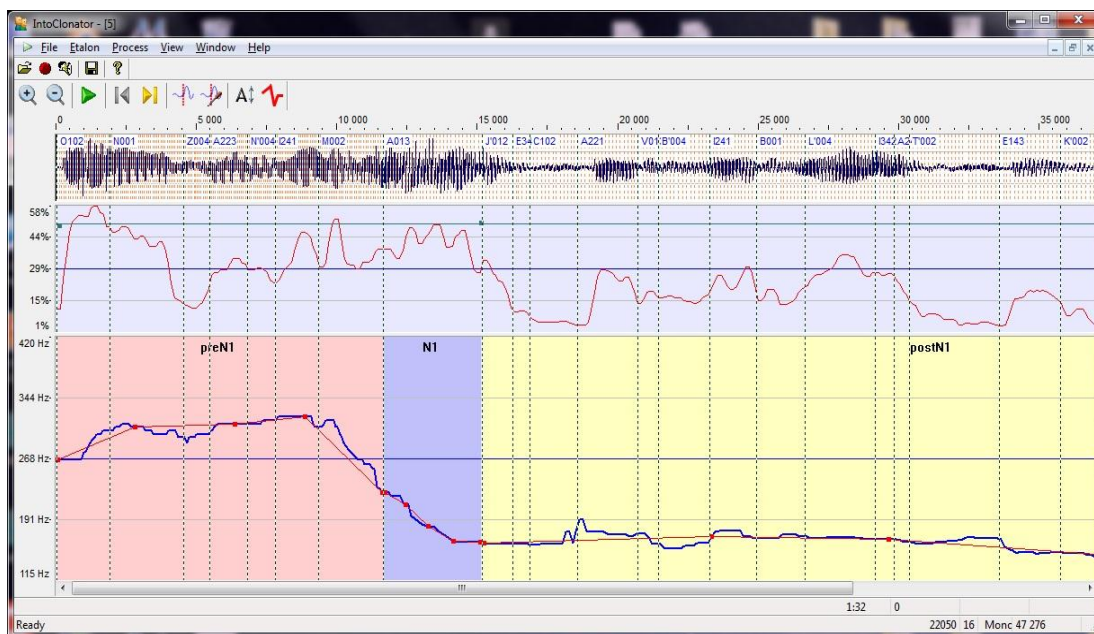


Рис. 2. IntoClonator

На заключительном этапе с помощью системы **ShapeEditor** (см. Рис. 3) обработанная ИнтоКлонатором информация используется для построения мелодических портретов ИК1 для фразы «Он занима+ется в библиотеке» в описанном выше нормированном виде.

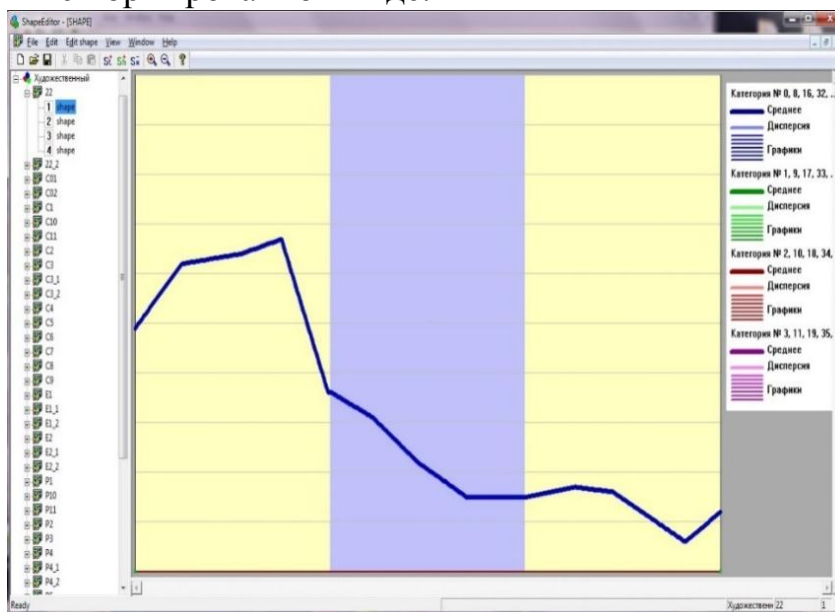


Рис. 3. ShapeEditor

Результат эксперимента

Интонационная конструкция № 2 согласно классификации Е.А.Брызгуновой имеет следующую характеристику особенностей типичного употребления и описания характера движения тона:

- [ИК2](#) реализуется в вопросе с вопросительными словами:
Кто пьет сок? Как поет Наташа?

При ИК2 ударная часть произносится с некоторым повышением тона.

При анализе фраз типичной интонационной конструкции №2 (ИК2), представленной в качестве наглядного примера, в соответствии

нормированной частотной шкалой (ИК1 значение $F_{0\min} = 75$ и 115 Гц, ИК5 значение $F_{0\max} = 210$ и 420 Гц), были получены универсальные мелодические портреты (УМП) [5] для женского и мужского голосов, на которых жирная кривая описывает усреднённые данные для ИК.

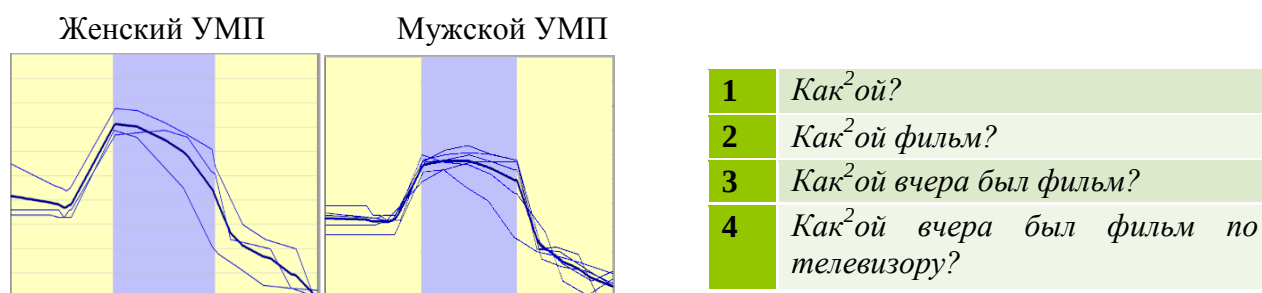


Рис. 4. ИК2 — УМП мужского и женского голосов + тексты фраз

Материал, полученный в ходе эксперимента при анализе всех семи ИК Е.А. Брызгуновой, позволяет сделать вывод: универсальные мелодические портреты являются эффективным средством представления тональной структуры интонационных конструкций Е.А. Брызгуновой, которые обеспечивают инвариантность относительно количественного и качественного состава произносимых фраз, а также высоты и диапазона голоса диктора.

Вывод

Сосредоточив внимание лишь на представлении акустических компонентов интонации, был создан эталонный набор УМП интонационных конструкций, который может быть размечен способом просодической маркировки синтагм, для того, чтобы в последующем озвучить этот материал желаемыми интонационными характеристиками с использованием **MultiPhone** [5], в котором созданные УМП ИК применяются в качестве эталонов. Используя данный материал в составе обучающих программ РКИ, да и любого другого языка, преподаватель или учащийся сможет самостоятельно формировать тренировочные упражнения для освоения звучания различных ИК русской речи, в различных контекстах и с различными голосами.

Эталонный набор УМП ИК может быть применен в анализаторе интонации произносимых фраз, таким образом создается специализированный программный комплекс, с помощью которого произносимая учащимся в микрофон фраза анализируется и на основе анализа спектральных, мелодических и энергетических характеристик строится её УМП. Затем полученный УМП сравнивается с эталонным и по результатам сравнения даётся оценка правильности интонирования произнесённой фразы, а также рекомендации по исправлению интонирования предъядра, ядра или заядра фразы.

Данный программный комплекс УМП ИК эффективен при создании компьютерных программ обучения русскому языку как иностранному с

использованием синтезаторов и анализаторов речи. УМП ИК могут быть использованы в составе русскоязычных синтезаторов речи для освоения русской интонации на слух и в составе анализаторов речи — для тренировки правильности интонирования в процессе говорения.

ЛИТЕРАТУРА

1. Брызгунова, Е.А. Звуки и интонация русской речи. — М., 1968.
2. Брызгунова, Е.А. Интонация // Русская грамматика. — М., 1980.
3. Lobanov, B. The phonemophon text-to-speech system // International Congress of Phonetic Sciences: proc. of the 11-th seccion ICPHS'87. — Tallin, 1987. — V.1. — P. 120–124.
4. Lobanov, B., Tsirulnik, L., Zhadinets, D., Karnevskaya, E. Language- and Speaker Specific Implementation of Intonation Contours in Multilingual TTS Synthesis // Speech Prosody: proceedings of the 3rd International conference. — Dresden, Germany, 2006. — Vol. 2. — P. 553–556.
5. Лобанов, Б.М., Цирульник, Л.И. Компьютерный синтез и клонирование речи. — Минск, 2008. (См. также: http://www.pselab.ru/Books/Lobanov_Cirulnik_2008.pdf)
6. Муханов, И.Л. Интонация в практике русской диалогической речи. — Ойкумена. — М., 1995.
7. Одинцова, И.В. Звуки. Ритмика. Интонация. — М., 2011.
8. Янко, Т.Е. Интонационные стратегии русской речи в сопоставительном аспекте. — М., 2008.