ИДЕНТИФИКАЦИЯ ДИКТОРОВ НА ОСНОВЕ СРАВНЕНИЯ ПАРАМЕТРОВ РЕАЛИЗАЦИИ МЕЛОДИЧЕСКИХ КОНТУРОВ ВЫСКАЗЫВАНИЙ

SPEAKER IDENTIFICATION BASED ON THE COMPARISON OF UTTERANCE PITCH CONTOUR PARAMETERS

Смирнова Н.С*. (nsmirnova@speechpro.com), Центр речевых технологий (http://speechpro.ru)*

В докладе представлен формализованный подход к использованию параметров мелодического контура высказывания для идентификации говорящего. Приводятся результаты исследований, даётся предварительная оценка эффективности предлагаемого метода, обсуждаются перспективы его дальнейшей разработки.

Введение

В современных условиях, когда большое внимание уделяется обоснованности выводов экспертного заключения, отмечается повышенный интерес к автоматическим методам идентификации, проверенным на представительных речевых базах. В то же время автоматические методы, по крайней мере на данном этапе их разработки, не способны к учёту целого ряда произносительных особенностей, информативность которых подтверждена многолетней экспертной практикой. К таким особенностям относится, в частности, специфика реализации ключевых элементов в структуре контура, обусловленная как физическими характеристиками говорящего, так и усвоенными стереотипами интонирования. Однако отсутствие в данной области устоявшегося общепринятого понятийно-терминологического аппарата, преимущественная опора на описательные (качественные) категории, а также недостаточность экспериментальных данных, отражающих степень идентификационной надёжности мелодических признаков, затрудняют их полноценное использование. С целью формализовать и объективировать процесс экспертного сравнения особенностей мелодического оформления высказываний был разработан метод анализа мелодического контура, позволяющий представлять и сравнивать основные описательные характеристики в численном виде. В настоящем докладе рассматриваются теоретические предпосылки метода и некоторые экспериментальные данные, позволяющие оценить его надежность.

Теоретическое обоснование

Способ анализа структуры интонационного контура, принятый за основу при разработке метода, включает элементы наиболее влиятельных подходов [1,2,3]. В качестве ключевого элемента мелодического контура рассматривается ядро (центр). Именно ядро является единственным обязательным элементом контура, формирует его «лицо» и играет решающую роль при восприятии типа высказывания. В большинстве языков, за исключением случаев эмфатического выделения, ядро чаще всего реализуется в конце высказывания и, таким образом, совпадает с мелодическим завершением.

В качестве основных параметров, различающих типы ядерной мелодики, используются:

- направление тона (восходящий, нисходящий, нисходяще-восходящий и т.д.);
- уровень или регистр (высокий-низкий)
- интервал (узкий-широкий)
- полнота (полный-неполный относительно границ дикторского диапазона)
- крутизна или скорость изменения тона (крутой-пологий)
- тайминг[1] (ранний-поздний)
- форма (вогнутый-выпуклый)

[1] Тайминг - временная точка достижения «целевого» значения ЧОТ относительно границ опорного элемента сегментной цепи. Термин введен в обращение лингвистами т.н. Голландской школы [3]. Тайминг определяет характер соотнесённости мелодического изменения с акцентным выделением. Как правило, в качестве целевой рассматривается либо предельная (например, минимальная для падений или максимальная для подъемов) частота, либо точка начала мелодического изменения. В качестве опорного элемента, относительно которого определяется тайминг, может выбираться как ударный (ядерный) слог или ударный гласный, так и вся слоговая последовательность, на которой реализуется изменение тона.

Предъядерный участок (в структуре которого различаются предшкала и шкала) обладает некоторой автономностью и, в отличие от заядерного участка, допускает определенную степень варьирования параметров, не связанную напрямую с типом реализуемой ядерной мелодики. Его роль при восприятии коммуникативного значения высказывания по сравнению с ядерным участком незначительна, однако этот элемент контура связывается иногда с передачей ряда нюансов, характеризующих речевую манеру говорящего. В частности, одним из полезных признаков в экспертной практике является оценка степени изрезанности шкалы. Другие используемые параметры описания традиционно включают уровень, направление и характер изменения тона (шкалы или предшкалы), а акцентно выделенные слоги шкалы могут описываться с использованием тех же параметров, что и ядерные акценты. Однако следует отметить, что предъядерный участок как отдельный элемент контура исследован очень слабо, и его описания гораздо хуже формализованы.

Изложенный подход к анализу мелодического контура пригоден для большинства нетональных языков, т.е. является относительно языконезависимым.

Как свидетельствуют данные литературы, приведённые выше параметры реализации ядерного тона не только дифференцируют различные функционально нагруженные мелодические типы, но и могут использоваться для отражения региональной специфики реализации однотипной мелодики. Например, неполнота конечных нисходящих тонов является одной из характерных черт русской речи жителей Дальневосточного региона, а тайминг мелодического пика, как показывают исследования на материале ряда языков [4,5,6], различается в зависимости от диалектной принадлежности говорящего.

Логично предположить, что индивидуальные интонационные особенности могут быть описаны с использованием тех же параметров, только междикторские различия в этом случае окажутся более тонкими, а границы междикторской вариативности, вероятнее всего, менее четкими.

На основании упомянутых выше интонационных описаний, а также с учетом достаточно скупых сведений, содержащихся в литературе относительно стабильности проявления отдельных мелодических особенностей [7-12] и их различительном потенциале [13,14], был разработан метод сравнения мелодических характеристик, в основе которого лежит набор элементов мелодического контура с относящимися к ним реализационными параметрами.

Описание метода

Цель предлагаемого метода анализа – по возможности представить основные мелодические параметры локального уровня, используемые в экспертной практике, в численном виде, автоматизировать их вычисление и статистическую обработку, а также предложить способ интерпретации результатов в форме экспертного решения с оценкой его надёжности.

На данном этапе обсуждается только анализ ядерной мелодики двух видов – простое падение и простой подъем. Эти два типа ядерной мелодики являются наиболее частотными и легко выделяются в потоке речи.

В составе параметров, используемых для характеристики ядерной мелодики обоих типов, можно выделить как «физические» (связанные преимущественно с анатомо-физиологическими особенностями говорящего), так и собственно лингвистические:

1. Начальная частота – значение первого отсчета (в Гц) в начальной точке ядерного фрагмента контура;

2. Конечная частота – значение последнего отсчета (в Гц) в конечной точке ядерного фрагмента контура;

3. Максимальная частота – максимальное значение частоты ОТ (в Гц) в пределах ядерного фрагмента контура;

4. Минимальная частота – минимальное значение частоты ОТ (в Гц) в пределах ядерного фрагмента контура;

5. Средняя частота – среднее значение ЧОТ (в Гц) в пределах ядерного фрагмента контура;

6. Время максимума – координата максимального значения в процентах от общей длительности ядерного фрагмента, соответствует параметру «тайминг»;

7. Время минимума – координата минимального значения в процентах от общей длительности ядерного фрагмента, соответствует параметру «тайминг»;

8. Время половинной частоты – координата значения половинной частоты (от интервала между максимумом и минимумом) в процентах от общей длительности ядерного фрагмента; частично соответствует описательной категории выпуклость-вогнутость. Относительное время половинной частоты как полезный индивидуализирующий параметр предложено Ф. Ноланом в [13].

9. Интервал – разница между максимальным и минимальным значением частоты ОТ (в Гц и в полутонах);

10. Скорость изменения тона – средняя скорость убывания или возрастания тона на выделенном участке контура в Гц/мсек., соответствует описательной категории «крутизна».

Эффективность сравнения во многом определяется сопоставимостью стилистической и коммуникативной направленности реплик, а также сегментных основ, на которых реализуются мелодические контуры. Поэтому

при работе по данной методике от эксперта требуется корректное выделение фрагментов контура для получения статистических данных для сравнения. В общем случае для получения достоверной статистики нежелательно использовать фрагменты с различным количеством слогов, с начальным и/или конечным глухим согласным, а также фрагменты, взятые из высказываний различной коммуникативной и/или эмоционально-стилистической направленности.

Численные представления мелодических категорий дают возможность осуществлять дальнейшую статистическую обработку данных – вычислять дисперсию, среднее квадратическое отклонение, коэффициент вариативности, доверительный интервал по каждому параметру. Однако корректная интерпретация результатов не может опираться только на сравнение статистики двух образцов – необходимы полученные на представительном материале пороги внутридикторской вариативности по каждому из используемых параметров, а также степень их идентификационной надёжности. Некоторые экспериментальные данные, позволяющие дать предварительную оценку надёжности метода, представлены в следующем разделе.

Данные экспериментов

С целью оценки различительного потенциала параметров из приведённого выше набора было проведено исследование характеристик ядерных восходящих и нисходящих тонов, реализованных на одном слоге с сегментной структурой «звонкий согласный+гласный» и «звонкий согласный+гласный+звонкий согласный» в речи на таджикском языке. Речевой материал составили записи 10 дикторов-мужчин. Краткая характеристика дикторов приведена в таблице 1.

Диктор	**Возраст**	**Регион**
Диктор 1	25	Душанбе
Диктор 2	21	Душанбе
Диктор 3	20	Душанбе
Диктор 4	35	Душанбе
Диктор 5	18	Душанбе
Диктор 6	35	Юг
Диктор 7	51	Юг
Диктор 8	35	Север
Диктор 9	38	Юго-восток
Диктор 10	35	Юго-восток

Таблица 1. Характеристика дикторов

Звукозапись производилась в цифровом формате непосредственно на портативный компьютер с помощью программы Wave Assistant (ЦРТ); частота дискретизации 22050 Гц, разрядность 16 бит. Для каждого диктора анализировались 2 записи (подхода), сделанные с интервалом не менее недели.

Из материала для исследования мелодических параметров были отобраны фрагменты (квази-) спонтанной речи – рассказ о себе, о своем городе и т.п. На основе анализа неконечных и конечных синтагм повествовательных высказываний в речи каждого диктора было выделено от 10 до 30 ядерных слогов с тонами обоих типов.

Потенциальная различительная способность параметров традиционно оценивается при помощи т.н. F-критерия (критерия Фишера), который вычисляется как отношение вариативности средних значений параметра у исследуемых дикторов к средней вариативности параметра в речи каждого из дикторов по формуле:

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

, где σ_1^2 - дисперсия средних значений параметра у разных дикторов, а σ_2^2 - среднее значение дисперсии параметра в речи одного диктора.

Перспективными параметрами для использования считаются те, для которых значение F-критерия превышает единицу. Высокие значения F-критерия не гарантируют высокую эффективность идентификации – это лишь означает, что по данному параметру дикторы могут быть четко разделены как минимум на две идентификационные группы.

На основе обработки речевого материала с помощью разработанного в ЦРТ специализированного модуля

мелодического анализа были получены значения реализационных параметров для нисходящего и восходящего ядерных тонов в речи исследуемых дикторов. Затем для реализационных параметров был рассчитан F-критерий. Внутридикторская вариативность рассчитывалась на основе объединённых данных по обоим подходам каждого диктора, междикторская – как средняя дисперсия средних значений параметров первого и второго подхода. Поскольку алгоритмы расчёта структурно-временных параметров находятся в стадии разработки, в настоящем докладе не приводятся данные для параметров «время максимума» и «время половинной частоты».

Гистограмма значений F-критерия для подъёмов и падений по исследованным параметрам приведена на рис.1.

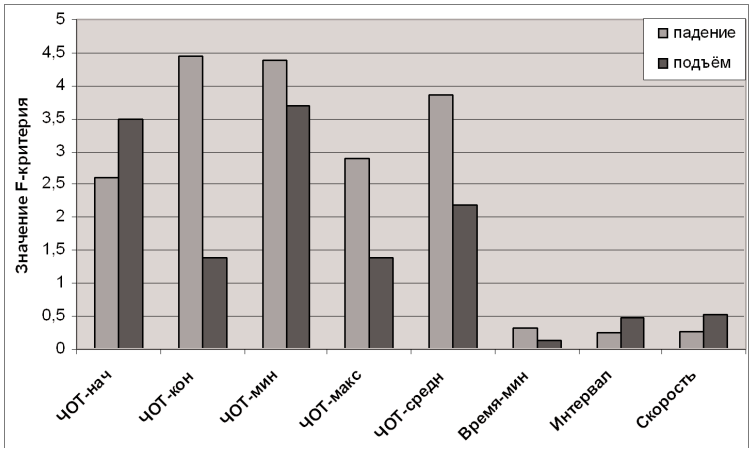

Рис. 1. Значения F-критерия для параметров ядерных восходящих и нисходящих тонов

Из всех анализируемых параметров обоих типов мелодики значение F-критерия превысило единицу только для т.н. «физических». Причём наилучшие результаты получены для конечного и минимального значения нисходящего тона – междикторская вариативность по этим параметрам превышает внутридикторскую более чем в 4 раза. Данный результат был в целом ожидаемым, поскольку стабильность значений частоты нисходящего завершения подтверждена многими исследованиями [7:64]. Примечательно, что лучший результат имеет не собственно минимальное, а именно конечное значение, которое не всегда совпадает с минимальным. Для восходящих тонов наиболее информативными параметрами оказались значения минимальной и начальной частоты – междикторская вариативность для них превысила внутридикторскую более чем в 3 раза.

Значения F-критерия для остальных параметров даже не приблизились к единичной отметке, что свидетельствует в целом об их низкой различительной способности. Из этого не следует, что данные параметры можно совершенно исключить из рассмотрения как неинформативные, однако их использование ограничится, по-видимому, теми случаями, когда в речи диктора отмечаются стабильно проявляющиеся ярко выраженные отличия от типичного характера реализации того или иного признака. Такие случаи имеют место и в исследованной базе. Один из них проиллюстрирован гистограммой, приведённой на рис. 2.

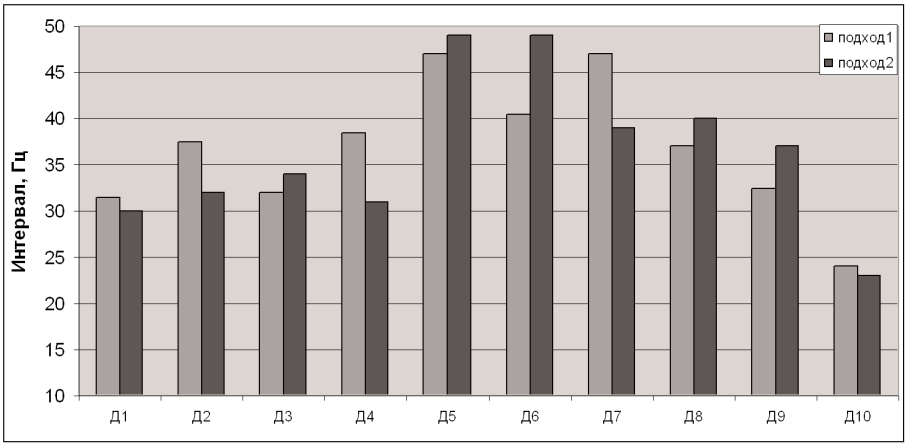

Рис. 2. Среднее значение интервала восходящего тона в двух подходах 10 дикторов

На гистограмме показаны значения интервала ядерного подъёма в Гц для двух подходов 10 дикторов. Как видно, большинство дикторов (8 из 10) реализуют подъём тона с интервалом от 30 до 40 Гц хотя бы в одном подходе (6 дикторов – в обоих подходах), три диктора имеют значения выше 45Гц хотя бы в одном подходе (один диктор – в обоих), и только один диктор (Д15) в обоих подходах имеет специфически узкий интервал реализации ядерного подъёма – менее 25 Гц. Учитывая, что вариативность по данному параметру у данного диктора существенно ниже среднедикторской (значение F-критерия составляет 1,4), данную особенность можно отнести к идентификационно значимым.

Аналогичный подход применим и к другим параметрам с низкими значениями F-критерия.

Реальная эффективность параметров для идентификации (в отличие от потенциальной, выражаемой через F-критерий) проверяется с помощью величины т.н. «равной ошибки» (EER) - для каждого из параметров определяется порог варьирования, при котором ошибка ложного принятия равна ошибке ложного отбрасывания. Поскольку исследованная база включала только 10 дикторов, а для сравнения использовались средние значения, полученные для 2 подходов одного диктора, ошибка ложного отбрасывания всегда составляла целое число, кратное 10 (1 диктор – 10%, 2 диктора – 20% и т. д.). В то же время ошибка ложного принятия была получена для 180 сравнений (10 дикторов*2подхода*9 сравнений), поэтому в большинстве случаев не могла совпасть с ошибкой первого рода. Для приблизительной оценки значения равной ошибки было решено определять порог, при котором оба типа ошибки имели максимально близкое значение (например, 20% и 23%, 40% и 37% и т.п.), после чего вычислять их среднее, которое и принималось за значение равной ошибки. Полученные таким образом значения ошибки для исследованных параметров обоих типов ядерного тона приведены на рис.3.

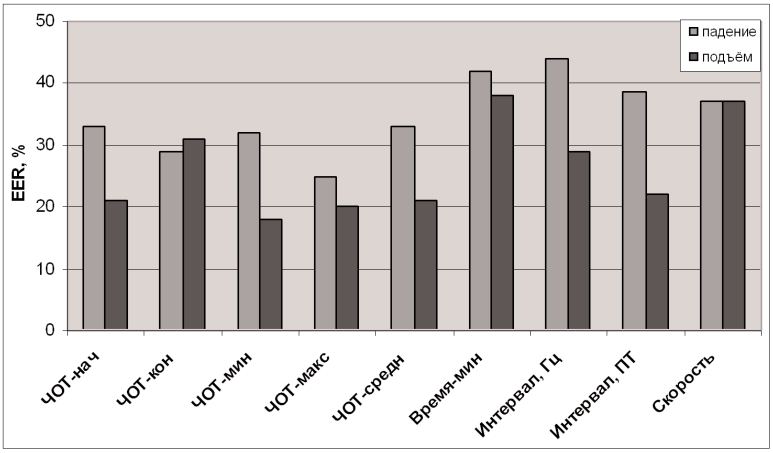

Рис. 3. Величина равной ошибки (EER) для параметров нисходящего и восходящего тона

Вопреки ожиданиям, ошибка распознавания на основе параметров нисходящих тонов оказалась в целом выше, чем для аналогичных параметров восходящих тонов. Исключение составила конечная частота восходящего тона, которая характеризуется значительным варьированием. Для самых перспективных по F-критерию параметров – конечной и минимальной частоты нисходящего тона – ошибка превысила 25%. Наименьшее же значение ошибки распознавания зафиксировано для минимальной частоты восходящего ядерного тона – 18%.

В целом в рамках исследуемой речевой базы из 10 дикторов надежность идентификации не менее 70% по двум подходам обеспечивают следующие параметры:

- Минимальная частота восходящего тона – 82%
- Максимальная частота восходящего тона – 80%
- Начальная частота восходящего тона – 79%
- Средняя частота восходящего тона – 79%
- Интервал восходящего тона в полутонах – 78%
- Максимальная частота нисходящего тона – 75%
- Конечная частота нисходящего тона – 71%
- Интервал восходящего тона в Гц – 71%

Ошибка распознавания по шкале полутонов была дополнительно рассчитана для интервалов тона. Следует отметить, что среди исследователей нет единого мнения о том, какая именно шкала измерения относительной высоты тона является более точной – у каждого из используемых сегодня типов шкал (герцы,

полутона, мелы, барки, эрбы) есть свои преимущества и недостатки [15]. Однако в рамках данного исследования переход на шкалу полутонов позволил снизить ошибку распознавания для нисходящих тонов на 5,5%, а для восходящих – на 7%.

Заключение

В докладе предложен метод идентификационного сравнения структурных элементов мелодического контура высказывания. Предварительные результаты применения метода, полученные на речевой базе парных записей 10 дикторов, свидетельствуют о возможности его использования в задачах идентификации диктора, по крайней мере при сравнении коммуникативно и сегментно сопоставимых участков естественно произносимой стилистически нейтральной речи. Лучший из исследованных параметров – минимальная частота восходящего ядерного тона – обеспечивает надежность идентификации 82%.

Дальнейшая разработка метода предполагает, в частности, расширение числа единиц анализа (элементов контура), введение в качестве самостоятельных единиц заполненных пауз хезитации, уточнение критериев сопоставимости фрагментов в зависимости от конкретных параметров сравнения, выявление степени взаимозависимости используемых параметров, введение дополнительных параметров, в том числе производных и отношений величин, изучение влияния речевого стиля и эмоционального состояния говорящего на стабильность проявления мелодических характеристик. Полученные данные позволят установить степень надёжности каждого из используемых параметров в процессе идентификации и определить эффективность метода по совокупности всех параметров. Предполагается, что завершением данной исследовательской работы станет разработка процедуры формирования частной оценки близости мелодического оформления сравниваемых образцов речи, которая будет учитываться при принятии общего вероятностного идентификационного решения на основе различных видов анализа речевого сигнала.

Список литературы

1. Брызгунова Е.А. Интонация // Русская грамматика, М.: 1980. Т.1, С.96-123.
2. O‘Connor J., Arnold G. Intonation of colloquial English // London: Longman, 1973.
3. ’t Hart J., Collier R., Cohen A. A Perceptual Study of Intonation: An experimental-phonetic approach to speech melody // Cambridge: Cambridge University Press, 1990.
4. Bruce G., Frid J., Thelander I. Swedish Accent Navigation // International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages. Beijing: 2004.
5. Nolan F., Farrar K. Timing of F0 peaks and peak lag // Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco: 1999. V.2, PP.961–4.
6. Peters J. The timing of nuclear high accents in German dialects // Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco: 1999. V.3, PP.1877-1880.
7. Ladd D. R. Intonational Phonology // Cambridge: Cambridge University Press, 1996.
8. Menn L. and Boyce S. Fundamental frequency and discourse structure // Language and Speech: 1982, № 25, PP.341-383.
9. Ashby M. A study of two English nuclear tones. Language and Speech: 1978, № 21, PP.326-336.
10. Grundstrom A. L’intonation des questions en Français Standard // A. Grundstrom and P. L´eon (eds) Interrogation et Intonation. Paris: Didier, 1973. № 8, PP.19–51.
11. Smirnova N. On the phonological status of the HL*H vs. H*LH timing-related tonal opposition in Dutch // Speech Prosody-2002, PP.651-654.
12. Jessen M., Köster O., Gfroerer S. Influence of vocal effort on average and variability of fundamental frequency // The International Journal of Speech, Language and the Law: 2005. V.12, №2, PP.174-203.
13. Nolan F. Intonation in speaker identification: an experiment on pitch alignment features // Forensic Linguistics: 2002. V.9(1), PP.1-21.
14. Kraayeveld J. Idiosyncrasy in prosody: speaker and speaker group identification in Dutch using melodic and temporal information // PhD thesis. Catholic University Nijmegen: 1997.
15. Nolan F. Intonational equivalence: an experimental evaluation of pitch scales // Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona: 2003. PP.771-774.