А.И. ЛЕВИН, П.Е. МИНИН, А.Д. ЕГОРОВ

Научный руководитель — Ю.Ю. ШУМИЛОВ, д.т.н., профессор

*Национальный исследовательский ядерный университет «МИФИ»*

**РАСПОЗНАВАНИЕ ИНТОНАЦИИ В НЕПРЕРЫВНОЙ РЕЧИ ЧЕЛОВЕКА**

Предложен метод распознавания интонации в речи человека с использованием open source программного обеспечения Praat. В методе используется алгоритм получения тона речи человека, используемый в Praat. Точность работы данного алгоритма, а также область его применения может быть изменена посредством варьирования параметров.

По данным Всемирной организации здравоохранения, около 10% населения Земли страдают различными нарушениями слуха[1]. Помимо нарушения слуха у таких людей также имеются проблемы с речью, в связи с чем в повседневной жизни они испытывают трудности в общении. Алгоритм распознавания интонации речи, в совокупности с другими методами анализа речи (например, распознавание речи), позволяет реализовать программное обеспечение (например, для смартфона) для помощи людям с речевыми и слуховыми дефектами в общении.

Предлагаемый метод распознавания интонации основан на выделении тона из аудиозаписей голоса человека и дальнейшем поиске их в шаблонах в речи человека – интонационных конструкциях.

В данном методе выделяется три основных этапа: запись речи человека и разделение ее на законченные интонационные конструкции, выделение тона голоса в каждой из частей, построение классификатора интонационных конструкций.

На первом этапе происходит запись голоса человека. В качестве данных для исследования данного алгоритма были использованы аудиозаписи различных диалогов с разными интонационными конструкциями, во время записи которых говорящим создавались паузы между ними продолжительностью ~250 мс. После этого следует шаг разделения аудиозаписи на части по интервалам, в которых отсутствует голос.

На втором этапе происходит выделение тона голоса. Для выделения тона голоса в каждой части записи используется алгоритм выделения частоты основного тона голоса. Данный алгоритм в программе Praat реализуется следующим образом[2]:

1. Предобработка сигнала с помощью оконной функции и быстрого преобразования Фурье для компонентов сигнала.
2. Вычисление глобального абсолютного значения сигнала.
3. Выделение части сигнала; вычитание локального среднего.
4. Анализ части сигнала на наличие голоса с помощью параметров VoicingThreshold и SilenceThreshold.
5. Умножение на оконную функцию.
6. Присоединение нулевого массива (для автокорреляции значений).
7. Быстрое преобразование Фурье.
8. Расчет частей в частотной области.
9. Быстрое преобразование Фурье.
10. Деление на автокорреляционную функцию окна.
11. Поиск положения и высоты максимумов в непрерывной автокорреляционной функции.
12. Получение тона для частей с голосом и без голоса.

После второго этапа для каждой записи получены зависимости полутонов от времени. На третьем этапе данные зависимости подвергаются классификации. На этом шаге возможно использование скрытых марковских моделей или искусственных нейронных сетей.

В русском языке существует семь видов интонационных конструкций: ИК-1 – ИК-7[3]. Для различных типов интонаций характерны следующие конструкции:

* Для вопросительной интонации характерны интонационные конструкции ИК-2, ИК-3 и ИК-4.
* Восклицательная интонация в основном соответствует конструкциям ИК-5 и ИК-6.
* Повествовательная интонация встречается в конструкциях ИК-1 и ИК-7.

Каждая из этих конструкций является шаблоном тона в речи человека. В связи с этим для задачи классификации больше всего подходят скрытые марковские модели[4].

Точность работы данного алгоритма зависит от качества записей, от особенностей речи человека, от продолжительности записей, от уровня постороннего шума во время записи, и прочие факторы. Точность распознавания интонации может быть повышена за счет изменения параметров алгоритма в программе Praat, а также параметров классификатора на основе скрытых марковских моделей.

*Список литературы*

1. А.И. Осадчих, «Медицина труда и промышленная экология», №4, с. 1-4, 2002.
2. Paul Boersma, «Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound», Institute of Phonetic Sciences, University of Amsterdam, IFA Processings, vol. 17, pp. 97-110, 1993.
3. Брызгунова Е.А. Интонация и смысл предложения. – Русский язык за рубежом, 1967, №1.
4. F. Ringeval, J. Demouy, G. Szaszak, M. Chetouani, L. Robel, J. Xavier, D. Cohen, and M. Plaza, «Automatic intonation recognition for the prosodic assessment of language impaired children», IEEE Transactions on Audio, Speech & Language Processing, vol. 19, pp. 1328–1342, 2011.