

**Тезисы доклада**

Начало формы

1. **НАЗВАНИЕ ДОКЛАДА:**

Некоторые особенности решения задачи разделения дикторов на фонограмме.

Some features of solving the speaker segmentation problem on a phonogram.

1. **АВТОРЫ:**

Е. А. Петров

1. **ОРГАНИЗАЦИЯ (полное наименование, без аббревиатур):**

Петрозаводский государственный университет

Petrozavodsk State University

1. **ГОРОД:**

Петрозаводск

Petrozavodsk

1. **ТЕЛЕФОН:** (+78142) 71-10-69
2. **ФАКС:**
3. **E-MAIL**: johnp@petrsu.ru
4. **АННОТАЦИЯ:**

В статье представлены текущие результаты и проблемы, возникшие в ходе решения задачи разделения дикторов на фонограммах защиты кандидатских диссертаций. Обсуждены возможные варианты решения возникших проблем.

The article presents the current results and problems which have been found during the process of solving the "speaker segmentation" problem on phonograms. Possible solutions of the problems were presented.

1. **КЛЮЧЕВЫЕ СЛОВА:**

фонограмма, речь, сегмент, кластеризация, LIUM

phonogram, speech, segment, clustering, LIUM

1. **ТЕКСТ ТЕЗИСОВ ДОКЛАДА:**

 Задача разделения дикторов на фонограмме является одной из еще нерешенных задач в области обработки речи. Перед нами поставлена задача разработать систему, которая будет осуществлять выделение речевых сегментов из фонограмм защиты кандидатских диссертаций и производить последующую кластеризацию выделенных сегментов по принадлежности к дикторам. Существующие алгоритмы и свободно распространяемые программные библиотеки для решения задачи разделения дикторов не позволяют ее решить с достаточной точностью.

 В статье [1] представлен обзор свободно распространяемых систем разделения дикторов и результаты тестирования рассмотренных систем на 6 файлах из корпуса NIST2008-ENG. В качестве критерия оценки эффективности работы систем разделения дикторов используется критерий DER (Diarization Error Rate), описание которого представлено в статье [1]. Результаты тестирования свободно распространяемых систем представленные в статье [1] показали, что наименьшую ошибку дала система LIUM[2]. Было принято решение вести дальнейшую разработку системы разделения дикторов на основе системы LIUM, так как она показала наилучшие результата, а так же имеет открытый исходный код, большое количество уже готовых модулей и функций для создания систем разделения дикторов.

Система LIUM по умолчанию в качестве акустических признаков использует 13 мел-частотных кепстральных коэффициентов (MFCC, Mel-frequency cepstral coefficients). Результаты проведенных экспериментов, с использованием различного количества акустических признаков, от 13 до 19 MFCC, показали, что для различных аудио записей наименьшее значение DER было получено при разном количестве используемых признаков MFCC.

Одним из шагов решения задачи разделения дикторов является решение задачи нахождения точек смены дикторов, т. е. разбиение фонограммы на непересекающиеся сегменты, содержащие речь только одного диктора. В системе LIUM для поиска точек смены дикторов на фонограмме возможно использовать четыре различные алгоритма: GLR, BIC, KL2, H2. Было выдвинуто предположение, что при объединении результатов работы системы с использованием нескольких алгоритмов нахождения точек смены диктора, получится снизить значения DER. Была разработана система, которая учитывала результаты работы четырех алгоритмов. Окончательная точка смены дикторов определяется только в случае, если она встречалась в определенном количестве алгоритмов. Проведены две группы экспериментов. В первой группе точка смены диктора ставилась в случае, если она встречалась в 2-х и более алгоритмах. Во второй группе точка смены дикторов ставилась в случае, если она встречалась в 3-х и более алгоритмах. В каждой группе экспериментов проводились эксперименты с различной длинной окна от 1,5 до 2,5 секунд. Результаты экспериментов показали, что для различных аудио записей наименьшие значения DER получаются в разных группах экспериментов и для разных значений длинны окна.

Были проведены эксперименты в которых из исходных аудиозаписей с помощью фильтров высоких и низких частот, были удалены звуки не входящие в звуковой диапазон среднестатистического диапазона речи человека. Далее на полученных аудио записях, была протестирована работа системы разделения дикторов. Результаты также показали, что для одних аудиозаписей фильтры снижают значение DER а для других увеличивают.

Все описанные выше эксперименты проводились на аудио записях их корпуса NIST2008-ENG. Особенность этих аудио-записей заключается в том, что они содержат запись телефонного разговора двух людей, в них отсутствуют сегменты речи содержащие одновременную речь двух человек, и на записях практически отсутствую помехи.

Для использования в экспериментах аудио записей защит кандидатских диссертаций необходимо в соответствии с рекомендациями NIST [3] создать специальный файл ключевой разметки дикторов для каждой аудио записи.

В ходе создания файлов эталонной разметки выяснились следующие особенности фонограмм защит, которые негативно скажутся на дальнейшем решении задачи разделения дикторов:

* фонограммы содержат большое количество посторонних шумов;
* на фонограммах содержатся участки речи дикторов длительностью менее 3 секунд;
* на фонограммах имеются участки содержащие одновременную речь нескольких дикторов;
* фонограммы содержат речь более 10 человек.

Так как описанные выше попытки улучшить показатели DER не принесли ожидаемого стабильного улучшения, в будущем предполагается опробовать следующие подходы для улучшения результатов работы системы разделения дикторов:

1. Применить обучение системы на небольшом участке фонограммы для подбора наиболее подходящих конфигурационных параметров системы разделения дикторов для данной фонограммы.
2. Применить в работе системы не только алгоритмы, осуществляющие разделение дикторов без наличия какой-либо априорной информации о дикторах, но и добавить в работу системы алгоритмы, которые будут выделять дикторов на основе заранее загруженных в систему голосовых моделей.
3. Использовать подходы и методы обработки участков фонограмм, содержащих голос сразу нескольких дикторов.

Работа выполнена при финансовой поддержке Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности.

Литература

1. Рогов А.А., Петров Е.А. Анализ существующих свободно распространяемых систем разделения дикторов на фонограмме// Фундаментальные исследования. – 2015. – № 6–1. – С. 67-72;
URL: [www.rae.ru/fs/?section=content&op=show\_article&article\_id=10007663](http://www.rae.ru/fs/?section=content&op=show_article&article_id=10007663) (дата обращения: 29.08.2015 NIST, "The 2009 (RT-09) Rich Transcription Meeting Recognition Evaluation Plan" [Электронный ресурс]. - Режим доступа: http://www.itl.nist.gov/iad/mig/tests/rt/2009/docs/rt09-meeting-eval-plan-v2.pdf (дата обращения:05.04.2015))
2. LIUM Speaker Diarization Wiki. [Электронный ресурс].- Режим доступа: http://www- lium.univ-lemans.fr/diarization/doku.php. (дата обращения:29.08.2015)
3. NIST, "The 2009 (RT-09) Rich Transcription Meeting Recognition Evaluation Plan" [Электронный ресурс]. - Режим доступа: http://www.itl.nist.gov/iad/mig/tests/rt/2009/docs/rt09-meeting-eval-plan-v2.pdf (дата обращения:29.08.2015)