

Д. Р. Дорошин, Д. А. Леднов, А. В. Хацкевич (Компания «Стел-КС»). **Акустическая и фонотактическая системы текстонезависимой идентификации дикторов.**

Система текстонезависимой речевой идентификации диктора представляет собой средство определения принадлежности речевых сообщений определенному диктору. Рассматриваемая система представляет собой комбинацию двух методов идентификации — акустического и фонотактического.

В качестве предварительной обработки аудиоданных используются известные алгоритмы PLP (Perceptual linear prediction) и MFCC (Mel-frequency cepstral coefficients), которые преобразуют последовательность амплитуд сигнала в последовательность векторов признаков [1].

Алгоритм акустической идентификации основан на вероятностной модели речи диктора в виде гауссовой смеси (GMM) [1]. Параметры модели оцениваются при помощи EM-алгоритма (Expectation-Maximization), максимизирующего функцию правдоподобия [1].

Первый шаг алгоритма акустической идентификации — сегментация аудиоданных на участки речи и пауз. Данный этап производится автоматически на основе моделей речи и паузы. Далее для выделенных речевых участков вычисляются вероятности принадлежности каждого из речевых участков интересующим нас дикторам. Диктор, которому соответствует максимальная вероятность по речевым фрагментам, считается идентифицированным.

В качестве второго алгоритма текстонезависимой идентификации используется фонотактический алгоритм. Основная идея алгоритма заключается в использовании различий в произношении русских фонем для различных дикторов.

На первом шаге строятся математические модели фонем. Для их моделирования используется подход, основанный на скрытых марковских моделях (СММ) [2]. На этом этапе обучения фонетической модели диктора, аудиозаписи диктора сегментируются на фонемы. Сегментирование выполняется при помощи фонемной модели русского языка. Модель фонемы — это СММ с тремя состояниями, которые описываются плотностями распределений в виде гауссовских смесей. Сегментирование записей производится автоматически, с помощью алгоритма Витерби [3].

Сегментированные на фонемы аудиоданные с помощью алгоритма Баума–Уелча [4] используются для обучения фонемной модели данного диктора.

При выполнении процедуры идентификации для каждого диктора, который участвовал в обучении, на основе данных из записи вычисляется правдоподобие. Диктор, которому соответствует максимальное правдоподобие, считается идентифицированным.

Акустический и фонотактический алгоритмы выполняются последовательно. Фонотактический алгоритм используется в том случае, если на основе акустического алгоритма нельзя принять достоверного решения.

СПИСОК ЛИТЕРАТУРЫ

1. *Аграновский А. В., Леднов Д. А.* Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. М.: Радио и связь, 2004.
2. *Young S. J.* The HTK hidden markov model toolkit: design and philosophy. Cambridge university engineering dept. Technical report. 1993.
3. *Vaseghi S. V.* Advanced digital signal processing and noise reduction. John Wiley & Sons Ltd, 2006.
4. *Baum L. E.* An inequality and associated maximization technique in statistical estimation for probabilistic functions of a Markov process. Inequalities, 1972.