

**ВЫДЕЛЕНИЕ НАИБОЛЕЕ ВАЖНЫХ ФОНЕМ В  
ЗАДАЧАХ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ  
РЕЧИ**

***Карпухин Иван Александрович***

*Аспирант*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: karpuhini@yandex.ru*

Системы автоматического распознавания речи давно вышли из категории узкоспециализированных инструментов и стали массовым продуктом. Качество распознавания для некоторых языков в офисных условиях приблизилось к возможностям человека. Однако все еще остается ряд нерешенных задач, стоящих на пути создания полноценных систем, способных грамотно функционировать в повседневной жизни. К числу таких задач относятся борьба с шумами звукового сигнала и оптимизация производительности распознающих программ.

Интересно отметить, что большинство современных систем, как коммерческих, так и бесплатных, основаны на принципах 20-и летней давности. Это относится и к алгоритмам, как то использование признаков MFCC и скрытых марковских моделей, и к структуре обучающих данных, содержащих фонетически репрезентативную либо фонетически сбалансированную речь.

В данной работе исследуется альтернативный подход к построению обучающих данных и подбору признаков входного сигнала. Среди всевозможных фонем русского языка ищется относительно небольшое их подмножество, успешное распознавание которого теоретически позволит добиться максимального качества распознавания речи. При построения такого подмножества оптимизируются две метрики:

1. Максимально достижимое качество распознавания речи при использовании только заданной группы фонем и игнорировании прочих
2. Различимость признаков описаний данной группы фонем

Например, в число наиболее полезных фонем попадают звуки «Э», «Л», «И», «Т», «А», «С».

В работе приводится формальная постановка соответствующей оптимизационной задачи и доказывается её принадлежность классу

сложности *NPC*, что оправдывает использование жадных алгоритмов её решения.

Полученные результаты предлагается использовать как для планирования структуры обучающих данных, например, за счет увеличения в них числа наиболее полезных фонем, так и для создания новых признаков, способных достаточно просто и точно описать необходимые звуки. В работе приведены примеры признаков, способных описать некоторые из предложенных фонем. Результаты применимы также к области мультимодального (аудио–видео) распознавания речи.

Все результаты показаны на примере русского языка. Насколько известно автору, данный подход к выявлению наиболее полезных фонем для задачи распознавания речи исследуется впервые.

Работа поддержана грантом по программе «УМНИК».

### **Литература**

1. Baker J. Developments and directions in speech recognition and understanding, part 1 // Signal processing magazine, IEEE, 2009, Т. 26, №3. С. 75–80.
2. Deng L. Machine learning paradigms for speech recognition: an overview // IEEE Transactions on audio, speech and language processing, 2013, Т. 21, №5. С. 1060–1089.
3. Кривнова О. Ф. Речевые корпуса на новом технологическом витке // Речевые технологии. 2008. № 2. С. 13–23.