

# СИСТЕМЫ РАСПОЗНАВАНИЯ РЕЧИ

Крейс А. В., Боброва Н. Л.

Факультет компьютерных систем и сетей, Кафедра информатики, Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: kreiu0@bk.ru, bobrova@bsuir.by

*В данной статье делается обзор современных систем распознавания речи, рассматривается их применение в различных сферах жизнедеятельности человека, обосновывается польза от использования подобных систем, изучаются принципы их реализации, приводятся сведения из истории развития данных систем.*

## ВВЕДЕНИЕ

В настоящее время системы распознавания речи распространены очень широко в различных сферах жизнедеятельности человека. Во многих случаях они позволяют упростить взаимодействие человека с компьютером, автоматизировать различные операции. Целью данной статьи является обзор сведений о системах распознавания речи. К задачам данной статьи следует отнести определение и классификацию систем распознавания речи, обзор истории их развития, изучение принципов их работы, сбор сведений о современных реализациях и сферах применения систем распознавания речи.

### I. ОПРЕДЕЛЕНИЕ И КЛАССИФИКАЦИЯ

Распознавание речи – это технология, с помощью которой речь человека возможно трансформировать в текст [1].

Системы распознавания речи можно классифицировать по различным критериям:

1. по зависимости от диктора:
  - системы, зависимые от диктора. Для работы с другим диктором такие системы требуют полной перенастройки;
  - системы, не зависящие от диктора [2];
2. по размеру словаря (по количеству слов, которое может распознать система);
3. по слитности речи:
  - слитная речь (пользователь может не делать дополнительных пауз между отдельными словами и фразами);
  - раздельная речь;
4. по типу структурной единицы (алгоритмы, использующие в качестве эталонов слова целиком или лишь части слов) [3];
5. по используемым алгоритмам:
  - скрытые Марковские модели (СММ);
  - динамическое программирование;
  - нейронные сети;
  - методы дискриминатного анализа [4];
6. по назначению.

### II. ИСТОРИЯ РАЗВИТИЯ

Первое устройство, способное распознавать речь, появилось в 1952 году. Это устройство по-

лучило название Audrey и было создано в компании Bell Labs. Оно было способно распознавать произнесенные человеком цифры [5].

Еще один этап развития технологии распознавания речи датируется 1962 годом, когда на рынке была представлена IBM Shoebox, которая могла распознавать не только цифры, но и некоторые команды для выполнения арифметических операций.

В 1976 году в рамках проекта DARPA, профинансированным министерством обороны США, была представлена машина Harry, которая могла распознавать 1011 слов. Вместо того, чтобы слышать целиком услышанные слова с образцами, Harry разделяла их на аллофоны (образец звучания фонемы (наименьшая акустическая единица языка [6]) в зависимости от окружающих ее букв) [7].

В 1986 году компания IBM выпустила пишущую машину с голосовым управлением, которая получила название Tangora. Она могла распознавать 20000 слов английского языка и использовала СММ. Это устройство могло приспосабливаться к голосу говорящего [5].

В 1997 году был выпущен первый в мире универсальный распознаватель речи, который не требовал делать паузу перед каждым словом, в виде программного обеспечения Dragon's NaturallySpeaking. Он был способен распознавать 100 слов в минуту [5,7].

### III. ПРИНЦИП РАБОТЫ

Принцип работы системы распознавания речи зависит от используемого алгоритма(-ов) (перечислены выше).

Подаваемый звуковой сигнал необходимо предварительно очистить от шумовых помех. Существуют различные методы очистки сигнала от помех, такие как методы, связанные с выделением признаков, которые являются инвариантными относительно шума; алгоритмы, основанные на коэффициентах линейного предсказания, мел-кепстральных коэффициентах; алгоритмы, связанные с цифровой обработкой сигналов и др [3].

Важным этапом предварительной обработки входного сигнала является нормализация

уровня сигнала. Это позволяет уменьшить погрешности распознавания, связанные с тем, что диктор может произносить слова с разным уровнем громкости [8].

Как правило, система распознавания речи состоит из двух моделей: акустической и лингвистической.

Акустическая модель отвечает за преобразование речевого сигнала в набор признаков, в которых отображена информация о содержании речевого сообщения. Программа выполняет сложный анализ речи, сравнивая аудиофрагменты с записанными в память речевыми образцами.

Лингвистическая модель анализирует информацию, получаемую от акустической модели, и формирует окончательный результат распознавания. На основе вероятностного расчета компьютер определяет, что именно мог произнести пользователь [6].

Если используются нейронные сети, они извлекают из потока речи фонемы и сопоставляют их с буквой, слогом или словом. Далее образуется порядок из известных программе слов, а неизвестные слова вставляются по контексту [1].

Если используются СММ, то производится анализ признаков распознаваемого слова и формирование соответствующей последовательности наблюдений [9]. Данные последовательности подаются моделям СММ, которые хранятся в базе, вычисляется вероятность того, что данная последовательность подходит к очередной СММ, слово, которому принадлежит СММ с наибольшей вероятностью, считается искомым [10].

При использовании алгоритма DWT (динамическое программирование) определение слова может осуществляться путем сравнения числовых форм сигналов или путем сравнения спектрограммы сигналов [11].

#### IV. СОВРЕМЕННЫЕ РЕАЛИЗАЦИИ И СФЕРЫ ПРИМЕНЕНИЯ

Системы распознавания речи используются в различных сферах жизнедеятельности человека. Они используются в системах голосового обслуживания и интерактивных автоответчиках [12], в медицине [1] и т.д.

Распознавание речи используется в различных голосовых помощниках. Примером такого помощника является Siri от компании Apple.

Кроме того, системы распознавания речи используются при расшифровке аудио- и видеозаписей, например, для подготовки документов по результатам собеседований [13].

Ряд компаний предоставляют разработчикам API для использования систем распознавания речи в своих приложениях. Примером может служить Google Cloud Speech-To-Text API.

Использование распознавания речи в различных системах делает возможным их использование людьми с нарушениями здоровья, например, слабовидящими людьми.

#### ЗАКЛЮЧЕНИЕ

В данной статье был выполнен обзор сведений о системах распознавания речи. Их высокая актуальность объясняется не только тем, что они автоматизируют различные процессы и делают взаимодействие человека с различными системами более комфортным, но и потому, что значительно облегчают использование таких систем людьми с нарушениями здоровья.

1. Mango Office [Электронный ресурс]. – Режим доступа: <https://www.mango-office.ru/newsletter/kak-rabotaet-raspoznavanie-rechi/>. – Дата доступа: 11.10.2022.
2. Речевые Технологии [Электронный ресурс]. – Режим доступа: <https://speetech.by/index.php?q=technologies/gaspoznavanie>. – Дата доступа: 11.10.2022.
3. Электронный научно-практический журнал «Современные научные исследования и инновации» [Электронный ресурс]. – Режим доступа: <https://web.snauka.ru/issues/2016/05/67547>. – Дата доступа: 11.10.2022.
4. Хабр [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/135087/>. – Дата доступа: 11.10.2022.
5. itnan.ru [Электронный ресурс]. – Режим доступа: <https://itnan.ru/post.php?c=1I&p=346928>. – Дата доступа: 11.10.2022.
6. Казачкин, А. Е. Методы распознавания речи, современные речевые технологии / А. Е. Казачкин // Молодой ученый. – 2019. – № 39 (277). – С. 6–8.
7. Хабр [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/company/toshibarus/blog/490732/>. – Дата доступа: 11.10.2022.
8. Электронная библиотека книг братьев Фроловых [Электронный ресурс]. – Режим доступа: <https://www.frolov-lib.ru/books/hi/ch05.html>. – Дата доступа: 11.10.2022.
9. Савин, А. Н. Разработка системы распознавания речи на основе скрытых марковских моделей отдельных слов / А. Н. Савин, Н. Е. Тимофеева, А. С. Гераськин, Ю. А. Мавлютова // Изв. Сарат. ун-та. Нов. сер. Сер. Математика. Механика. Информатика. – 2017. – Т. 17, № 4. – С. 452–464.
10. International Journal «Software Journal: Theory and Applications» (SJTA) [Электронный ресурс]. – Режим доступа: <http://swsys-web.ru/speech-recognition-system-based-on-hidden-markov-models.html>. – Дата доступа: 11.10.2022.
11. Хабр [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/135087/>. – Дата доступа: 11.10.2022.
12. Polygant [Электронный ресурс]. – Режим доступа: <https://polygant.net/ru/ai/sistema-raspoznavaniya-rechi/>. – Дата доступа: 11.10.2022.
13. VK Cloud [Электронный ресурс]. – Режим доступа: <https://mcs.mail.ru/blog/slushaet-i-ponimaet-kak-rabotaet-tehnologija-avtomaticheskogo-raspoznavaniya-rechi>. – Дата доступа: 11.10.2022.