

# Системы распознавания речи

*М.С. Кухтинова, Н.А. Позолотина, В.Г. Трубин  
ФГБОУ ВПО НГТУ, Новосибирск, Россия*

**Аннотация:** Простой способ реализации системы распознавания речи на языке программирования Python.

**Ключевые слова:** Голосовое управление, система распознавания речи, Python.

## ВВЕДЕНИЕ

Речь является одним из основных способов передачи информации человеком. Речевой аппарат и естественные языки совершенствовались и усложнялись в течение тысяч лет, привычка передавать сигналы речью объясняет желание создания голосового устройства вода для взаимодействия с техникой.

## ОБЗОР СУЩЕСТВУЮЩИХ ПРОГРАММНЫХ СРЕДСТВ

В связи с разнообразием фонем, а также тембров голосов, задача распознавания требует весьма сложных алгоритмов. Решения этой задачи представляют некоторые компании. Приведем несколько примеров программных продуктов, преобразующих устную речь в письменную:

1. Google speech – облачный сервис, имеет открытый API, поддерживает ряд языков, в т. ч. русский.
2. Yandex.SpeechKit – тоже облачный сервис, имеет API для мобильных платформ, поддерживает русский язык.
3. Стандартное средство распознавания речи Windows – поставляется с операционной системой, не поддерживает русский язык.
4. NUANCE Dragon – коммерческий продукт (стоимость варьируется от \$74 до \$799), поддерживаются платформы Windows и MacOS, доступен только английский язык.
5. Real speaker – коммерческий продукт (стоимость \$36.7), поддерживаемая платформа Windows, предназначена для русского языка.

Основные две группы систем распознавания речи – это облачные системы и ПО для PC и MAC. Некоторые облачные системы имеют открытые API, что позволяет использовать их для своих приложений, но при этом требуется устойчивое интернет-соединение. Так, Yandex.SpeechKit имеет открытые API для android, iOS и WindowsPhone, что позволяет использовать его для мобильных платформ. Google speech имеет открытый JavaScript API и поддерживает ряд языков, в т.ч. русский и английский. Значительным минусом облачных

систем является время отклика, которое может составлять несколько секунд в зависимости от скорости интернет-соединения.

Специальное ПО, например NUANCE Dragon, платформозависимо и требует определенных ресурсов для работы, а кроме того может иметь высокую цену. Также самые популярные коммерческие системы рассчитаны на английский язык, а качество продуктов для русского языка хуже. Некоторые компании (например, ООО «Центр речевых технологий»), предоставляют готовые ASR-средства (automatic speech recognition – автоматическое распознавание речи) для бизнеса, то есть такие сервисы, как речевая аналитика для систем записи, стенографирование, голосовое самообслуживание. Это подходит для использования в крупных компаниях, но абсолютно неуместно в небольших проектах.

## ПРИМЕР ПРОГРАММНОГО РАСПОЗНАВАНИЯ РЕЧИ

Рассмотрим пример преобразования речевой команды в текстовую, в котором будем использовать русскоязычную систему, чтобы увеличить шанс корректного распознавания, поскольку качество дикции на родном языке значительно выше, а акцента либо нет, либо он незначителен. Для демонстрации возможностей достаточно минимального количества текста, поэтому целесообразно использовать бесплатные системы. Из наиболее популярных систем приведенным выше требованиям удовлетворяют Google speech и Yandex.SpeechKit, но Yandex.SpeechKit имеет API только для мобильных платформ, поэтому будем использовать Google speech, который имеет web API.

Для реализации распознавания через Google speech необходимо сделать следующее:

1. Записать голосовую команду.
2. Конвертировать запись в формат FLAC со скоростью звукового потока 8 или 16 кГц.
3. Отправить этот файл POST-запросом на URL системы распознавания [1].
4. Получить ответ, который приходит в формате JSON (JavaScript Object Notation – текстовый формат обмена данными, основанный на JavaScript).

Записать аудиофайл, содержащий речь, и конвертировать его во FLAC можно любым доступным способом, в т. ч. и программно. Ниже приведен скрипт на языке Python, записывающий сигнал с микрофона в течение пяти секунд в wav-

файл. Для его исполнения, помимо самого Python [2], необходимо установить библиотеку Pyaudio [3]. Там же можно найти и пример скрипта, записывающего звук (здесь приведен модифицированный вариант):

```
# -*- coding: cp1251 -*-
import pyaudio, wave
CHUNK = 512; FORMAT = pyaudio.paInt16
CHANNELS = 1; RATE = 16000;
RECORD_SECONDS = 5
# Открытие потока
p = pyaudio.PyAudio()
stream = p.open(format = FORMAT, channels =
CHANNELS,
rate = RATE, input = True, output = True,
frames_per_buffer = CHUNK)

frames = []; print "Recording"
# Считывание информации из потока
for i in range(0, RATE / CHUNK *
RECORD_SECONDS):
    data = stream.read(CHUNK);
frames.append(data)

# Запись файла
wf =
wave.open("C:\\Python26\\Work\\audiofile.wav",
"wb");
wf.setnchannels(CHANNELS)
wf.setsampwidth(p.get_sample_size(FORMAT))
wf.setframerate(RATE);
wf.writeframes(b''.join(frames))
wf.close(); print "Done"

# Завершение потока
stream.stop_stream(); stream.close();
p.terminate()
```

Конвертация в формат FLAC (FLAC - Free Lossless Audio Codec, бесплатный аудиокодек, не имеющий потерь) не представляет сложностей, если у вас есть конвертер. Так, например, при наличии консольного конвертера [4], можно использовать следующий скрипт, написанный на Python:

```
import os
# Преобразование файла wav -> flac,
ПРОВЕРЬТЕ ВСЕ пути к файлам !
os.system('C:\\Program Files
(x86)\\FLAC\\flac.exe -8 -f
C:\\Python26\\Work\\audiofile.wav')
```

Следующий скрипт отправляет POST-запрос с содержимым аудиофайла и выводит результат.

```
# -*- coding: cp1251 -*-
import urllib2
# Передаем flac файл в Google и получаем
ответ в виде текста
GOOGLE_SPEECH_URL="https://www.google
.com/speech-api/v2/recognize?output=json&lang=ru-
```

```
RU&key=XXXXX&client=chromium&maxresults=
6&filter=2"
f = open('C:\\python26\\Work\\audiofile.flac','rb')
# Проверьте путь к файлу !!!
flac_content = f.read(); f.close()
hdrs = {"User-Agent": "Mozilla/5.0 (X11; Linux
i686) AppleWebKit/535.7 (KHTML, like Gecko)
Chrome/16.0.912.63 Safari/535.7", 'Content-type':
'audio/x-flac; rate=16000'}
req =
urllib2.Request(GOOGLE_SPEECH_URL,
flac_content, hdrs)
print "Sending request to Google"
p = urllib2.urlopen(req); response = p.read();
print response.decode('utf8')
```

В выше приведённом тексте программы XXXXX следует заменить на Ваш пароль. За время написания статьи API Google успел измениться с v1 на v2, в частности стал другим адрес обращения, поэтому при поиске информации по данному вопросу следует обращать внимание на дату публикации информации.

В результате работы этих скриптов в консоль Python будет выведен ответ в формате JSON, содержащий одну или несколько гипотез о том, что было сказано во время записи файла. Ждать ответа приходится около двух секунд, поэтому такой метод управления не подходит для быстродействующих систем.

## ВЫВОДЫ

Системы автоматического распознавания речи можно использовать в качестве устройства ввода информации. Такой способ получения текстовой информации от пользователя значительно быстрее и удобнее, чем ввод с клавиатуры, хотя имеет более низкую точность и причиняет ряд неудобств (необходимость наличия специального ПО, интернет-соединения и др.). Хотя системы автоматического распознавания речи требуют дальнейших разработок и улучшений, для некоторых случаев использовать их можно уже сейчас.

## ЛИТЕРАТУРА

- [1] Сервис распознавания речи Google Speech <https://www.google.com/speech-api/v2/recognize?output=json&lang=ru-RU&key=XXXXX&client=chromium&maxresults=6&filter=2> (дата обращения .05.14).
- [2] Сайт Python <https://www.python.org/> (дата обращения 16.05.14).
- [3] Сайт библиотеки Pyaudio <http://people.csail.mit.edu/hubert/pyaudio/> (дата обращения 16.05.14).
- [4] Сайт flac: <https://xiph.org/flac/index.html> (дата обращения 16.05.14).



**Мария Степановна Кухтинова** -  
Студент 4-го курса НГТУ кафедра  
Автоматики.  
E-mail: [mariyakukhtinova@gmail.com](mailto:mariyakukhtinova@gmail.com)



**Виталий Геннадьевич Трубин** -  
зав. лаб. кафедры Автоматики НГТУ,  
директор ООО «КБ Автоматика».  
Автор 18 научных статей. Область  
интересов – разработка  
специализированной электроники.  
E-mail: [trubin@ngs.ru](mailto:trubin@ngs.ru)



**Наталья Андреевна Позолотина** -  
Студент 4-го курса НГТУ кафедра  
Автоматики.  
E-mail: [sms-ka-93@mail.ru](mailto:sms-ka-93@mail.ru)

## **Speech Recognition Systems**

**M.S. KUKHTINOVA, N.A. POZOLOTINA,  
V.G. TRUBIN**

*Abstract:* The paper gives simple method for the realization of speech recognition system on computer language Python.

*Ключевые слова:* voice-activated control, speech recognition system, Python