

Watson Text to Speech API

IBM API 개발 가이드를 먼저 참고해 주세요.

목차

1. 음성 리스트 확인
2. `text`를 합성해 음성 파일 생성
3. `watson`에서 제공하는 음성 `api`가 특정 단어를 어떻게 읽는지 ‘발음’ 확인
4. 커스텀 모델

1. 사용 할 수 있는 음성 리스트 확인

```
import json
from ibm_watson import TextToSpeechV1
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator

authenticator = IAMAuthenticator('발급받은 API키 입력')
text_to_speech = TextToSpeechV1(
    authenticator=authenticator
)
text_to_speech.set_service_url('발급받은 URL 입력')

voices = text_to_speech.list_voices().get_result()  #Watson에 존재하는 voice 리스트를 받아옴
print(json.dumps(voices, indent=2))              코드
```

실행 결과

```
{
  "voices": [
    {
      "gender": "female",
      "supported_features": {
        "custom_pronunciation": true,  #커스텀 발음설정 가능
        "voice_transformation": false  #보이스 변경은 불가
      },
      "name": "en-US_LisaV3Voice",
      "customizable": true,
      "description": "Lisa: American English female voice. Dnn technology.", #설명
      "language": "en-US",             #해당 목소리의 국가 - 합성 가능 언어
      "url": "https://api.kr-seo.text-to-speech.watson.cloud.ibm.com/instances/2441732c-8626-4970-aa5e-a8a359865261/v1/voices/en-US_LisaV3Voice"
    },
    .....
  ]
}
```

2. 음성 합성하여 오디오 파일로 추출

```
import json
from ibm_watson import TextToSpeechV1
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator

authenticator = IAMAuthenticator('발급받은 API키 입력')
text_to_speech = TextToSpeechV1(
    authenticator=authenticator
)
text_to_speech.set_service_url('발급받은 URL입력')

# hello_world_ko : 합성 음성을 저장할 파일 이름
with open('hello_world_ko.wav', 'wb') as audio_file:
    audio_file.write(
        text_to_speech.synthesize(
            '안녕하세요',                                # 합성할 단어
            voice='ko-KR_YunaVoice',                    # 원하는 음성 선택
            accept='audio/wav'                           # 음성파일 설정
        ).get_result().content)                          #음성파일 설정은 다음 슬라이드 참고
```

코드 실행 결과는 아래와 같습니다.



2. 음성 합성하여 오디오 파일로 추출

물론 영어도 가능합니다.

.....

```
# hello_world_ko : 합성 음성을 저장할 파일 이름
with open('hello_world_ko.wav', 'wb') as audio_file:
    audio_file.write(
        text_to_speech.synthesize(
            '안녕하세요',                                # 합성할 단어
            voice='ko-KR_YunaVoice',                    # 원하는 음성 선택
            accept='audio/wav'                           # 음성파일 설정
        )
    )
#음성파일 설정은 다음 슬라이드 참고
```

코드 실행 결과는 아래와 같습니다.



2. 음성 합성하여 오디오 파일로 추출 - 설정

Audio formats (accept types)

The service can return audio in the following formats (MIME types).

- Where indicated, you can optionally specify the sampling rate (`rate`) of the audio. You must specify a sampling rate for the `audio/l16` and `audio/mulaw` formats. A specified sampling rate must lie in the range of 8 kHz to 192 kHz. Some formats restrict the sampling rate to certain values, as noted.
- For the `audio/l16` format, you can optionally specify the endianness (`endianness`) of the audio: `endianness=big-endian` or `endianness=little-endian`.

Use the `Accept` header or the `accept` parameter to specify the requested format of the response audio. If you omit an audio format altogether, the service returns the audio in Ogg format with the Opus codec (`audio/ogg; codecs=opus`). The service always returns single-channel audio.

- `audio/basic` - The service returns audio with a sampling rate of 8000 Hz.
- `audio/flac` - You can optionally specify the `rate` of the audio. The default sampling rate is 22,050 Hz.
- `audio/l16` - You must specify the `rate` of the audio. You can optionally specify the `endianness` of the audio. The default endianness is `little-endian`.
- `audio/mp3` - You can optionally specify the `rate` of the audio. The default sampling rate is 22,050 Hz.
- `audio/mpeg` - You can optionally specify the `rate` of the audio. The default sampling rate is 22,050 Hz.
- `audio/mulaw` - You must specify the `rate` of the audio.
- `audio/ogg` - The service returns the audio in the `vorbis` codec. You can optionally specify the `rate` of the audio. The default sampling rate is 22,050 Hz.
- `audio/ogg; codecs=opus` - You can optionally specify the `rate` of the audio. Only the following values are valid sampling rates: `48000`, `24000`, `16000`, `12000`, or `8000`. If you specify a value other than one of these, the service returns an error. The default sampling rate is 48,000 Hz.
- `audio/ogg; codecs=vorbis` - You can optionally specify the `rate` of the audio. The default sampling rate is 22,050 Hz.
- `audio/wav` - You can optionally specify the `rate` of the audio. The default sampling rate is 22,050 Hz.
- `audio/webm` - The service returns the audio in the `opus` codec. The service returns audio with a sampling rate of 48,000 Hz.
- `audio/webm; codecs=opus` - The service returns audio with a sampling rate of 48,000 Hz.
- `audio/webm; codecs=vorbis` - You can optionally specify the `rate` of the audio. The default sampling rate is 22,050 Hz.

For more information about specifying an audio format, including additional details about some of the formats, see [Audio formats](#).

3. 제공하는 음성이 특정 문장을 읽는 발음 확인

```
import json
from ibm_watson import TextToSpeechV1
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator
```

```
authenticator = IAMAuthenticator("발급받은 API키 입력")
text_to_speech = TextToSpeechV1(
    authenticator=authenticator
)
```

```
text_to_speech.set_service_url("발급받은 URL 입력")
```

```
pronunciation = text_to_speech.get_pronunciation(
    text='IEEE',
    voice='en-US_LisaVoice',
    format='ibm' #ipa도 있습니다.
    #customization_id = '{ }' 나중에 할거
).get_result()
print(json.dumps(pronunciation, indent=2))
```

ibm 방식의 결과는

```
{
  "pronunciation": "[.1Y] `[.1i] `[.1i] `[.1i]"
}
```

ipa의 경우 국제 발음기호로 나오는 것인데

```
{
  "pronunciation": ".\u02c8a\u026a .\u02c8i .\u02c8i .\u02c8i"
}
```

깨져서 확인이 힘들다..

4. Custom model

IEEE를 아이 이 이 이로 읽지 말고 아이 튜플 이 이런 식으로 읽도록 만들고 싶다.!
이때 커스텀 모델을 만들어서 사용함. 단... 무료 요금제는 사용이 불가능합니다.

This feature is not available for the Bluemix Lite plan. Please upgrade to a paid plan to activate this feature:

<https://console.bluemix.net/catalog/services/text-to-speech>, Code: 400 ,
X-global-transaction-id: d46cec56d7f806c0fe029e061c919539

따라서 해당 내용은

<https://cloud.ibm.com/apidocs/text-to-speech?code=python&locale=ko#create-a-custom-model>

로 참고해 주십시오.