

NAME

cmspeech - text to speech with a local Kokoro-82M model

SYNOPSIS

cmspeech [*OPTION*]... *MODEL* [*TEXT*]...

DESCRIPTION

cmspeech speaks *TEXT* (the arguments joined by spaces), the file given with **--file**, or standard input, and writes a WAV file: 24 kHz, 16-bit, mono. It runs hexgrad's Kokoro-82M (StyleTTS 2 with an iSTFTNet decoder) on the CPU.

MODEL is the **hexgrad/Kokoro-82M** snapshot directory, holding **config.json**, **kokoro-v1_0.pth** and **voices/**, or the directory the download created (**models--hexgrad--Kokoro-82M**), in which case the newest snapshot is used. It can also be a model of the same architecture with other weights, as a directory or its **.pth** file: a repository without **config.json** gets Kokoro-82M's, which is built in, and weights saved with PyTorch's newer weight norm (**parametrizations.weight**) are read as well.

Kokoro-82M was not trained on German. For German, use a German fine-tune such as **kikiri-tts/kikiri-german-victoria** or **kikiri-tts/kikiri-german-martin** (one voice each, Apache-2.0); German text goes through **espeak-ng**, as for their training.

English text is turned into phonemes with misaki's US or GB lexicon, built in, including its rules for numbers, currency, stress and the endings -s, -ed and -ing. A word the lexicon lacks is read as two lexicon words where it splits into them ("bookshelves"), else by **espeak-ng** where it is installed, else by letter-to-sound rules learned from the lexicon (about two thirds of such words right; names fare worse). German, Spanish, French, Hindi, Italian and Brazilian Portuguese go through **espeak-ng** entirely. Japanese and Mandarin voices need **--phonemes**.

Text is split at line breaks, and longer paragraphs before 510 phonemes at the last sentence, clause or comma boundary; the pieces are spoken one after another into one file.

OPTIONS

-o, --output *PATH*

The WAV file to write, or - for standard output. Default: **speech.wav**.

--voice *NAME*

A voice in *MODEL*/voices (**af_heart**, **bm_george**, **ef_dora**, **victoria**, ...), a path to a **.pt** voice file, or several separated by commas, which are averaged. A name like **af_heart** sets the language by its first letter (see **--lang**). Default: **af_heart** where the model has it, else its only or first voice.

-s, --speed *X*

Speaking rate; 2 is twice as fast. Default: 1.

-f, --file *PATH*

Read the text from *PATH*.

--lang *L*

a American English, **b** British English, **d** German, **e** Spanish, **f** French, **h** Hindi, **i** Italian, **p** Brazilian Portuguese. Default: the voice name's first letter, else the language in the model card's **README.md** front matter, else American English.

--phonemes

TEXT is already in Kokoro's phoneme alphabet (as **--print-phonemes** shows it); it is cut into pieces at spaces.

--no-espeak

Do not run **espeak-ng**.

-p, --print-phonemes

Print the phonemes of each piece on standard error.

--seed *N*

Seed of the noise in the vocoder's excitation. Default: 0, so the same input gives the same file.

-v, --verbose

Report the load time, whether **espeak-ng** is used, and the audio length, time taken, speed as a multiple of real time and number of pieces, on standard error.

-h, --help

Show the options.

ENVIRONMENT

CMGENERATE_THREADS

Worker threads.

CMSPEECH_ESPEAK

The **espeak-ng** program. Default: **espeak-ng** on the path.

EXAMPLES

`cmspeech ~/.cache/huggingface/hub/models--hexgrad--Kokoro-82M "Hello, world." -o hello.wav`

```
cmspeech --voice bf_emma -f chapter.txt -o chapter.wav kokoro/  
cmspeech --voice af_bella,af_sarah -o - kokoro/ "Mixed voices." | aplay  
cmspeech ~/.cache/huggingface/hub/models--kikiri-tts--kikiri-german-victoria "Guten Tag." -o tag.wav
```

SEE ALSO**cmtranscribe(1), espeak-ng(1)**