NAME

cmvoiceagent − talk to cmagent with a microphone and a speaker

SYNOPSIS

cmvoiceagent −−asr *MODEL* **−−tts** *MODEL* [*OPTION*]... [*CMAGENT OPTION*]... *MODEL*

DESCRIPTION

cmvoiceagent runs **cmagent**(1) as a child over its **−−bridge** protocol. Speech from the microphone is transcribed with Parakeet (as in **cmtranscribe**(1)) and sent as a task. The answer is read aloud with Kokoro (as in **cmspeech**(1)) sentence by sentence while the model is still writing.

Options not listed below go to cmagent unchanged, including the language model *MODEL*, **−−gpu**, **−−write**, **−−run** and **−c**. cmagent also reads its own configuration files. **−−tui** and **−−once** are refused.

By default the session is half duplex: the microphone is off while the agent works and while its speech plays. An utterance ends after a pause (**−−silence**). With **−−barge−in** it stays on, and speaking stops the agent; see **INTERRUPTING**. Tool calls, the text a model writes after a call in the same turn, code blocks, markdown marks and URLs are not spoken; the whole transcript is printed.

When cmagent asks for approval, the question is spoken ("May I run cargo test?") and the first word of the answer decides: yes, yeah, sure, okay and similar allow; no, stop, cancel and similar refuse. Anything else is asked again.

Typed lines are tasks as well, and a line starting with **/** is a cmagent command, such as **/new**, **/undo** or **/quit**.

OPTIONS

−−asr *MODEL*
: Parakeet model, in any form cmtranscribe takes. Required.

−−tts *MODEL*
: Kokoro model, in any form cmspeech takes. Required.

−−voice *NAME*
: Kokoro voice; default af_heart, else the model's first voice.

−−lang *L*
: Kokoro language letter; default from the voice name or the model card.

−−speed *X*
: Speaking rate; default 1.

−−no−espeak
: Do not use espeak-ng for English words outside the lexicon.

−−record *CMD*
: Shell command that writes 16 kHz mono 32-bit float little-endian samples to standard output. Default: the first of **pw−record**, **parec**, **arecord** and sox's **rec** on PATH (**rec** first on macOS).

−−play *CMD*
: Shell command that reads 24 kHz mono 32-bit float little-endian samples from standard input. Default: the first of **pw−play**, **pacat**, **aplay** and sox's **play**.

−−silence *MS*
: Pause that ends an utterance, in milliseconds; default 700.

−−level *DB*
: Fixed speech threshold in dBFS. By default a 30 ms frame counts as speech 12 dB over the measured noise floor and above −50 dBFS. Utterances under 250 ms of speech are ignored; 60 s without a pause is taken as noise and dropped.

−−keyboard
: No microphone; type the tasks. Answers are still spoken.

−−barge−in
Listen while the agent talks or works, and stop it when the user speaks. Needs headphones or an echo-cancelled microphone; see **INTERRUPTING**.

−−agent *PATH*
The cmagent program; default the one next to cmvoiceagent, else cmagent on PATH.

−h, **−−help**
Show a summary of the options.

−−system is passed on with a sentence added: answer in short spoken sentences, without markdown or code unless asked. Without **−−system**, that sentence becomes the system prompt and replaces any **−−system** from cmagent's configuration files.

INTERRUPTING

While an utterance runs, its last 2.4 s are transcribed every 0.9 s. A snippet is taken as the agent's own voice when it is within 0.35 edits per character of a stretch of the last 20 sentences spoken, compared without its first and last word, which the window may have cut. Under 12 characters every word has to have been spoken. Anything else is an interruption. The speaker command is killed (with its process group), the queued sentences are dropped, and a running task is cancelled as with Ctrl-C in cmagent. The utterance restarts at the snippet that interrupted, and when it ends it is sent as the next task, followed by a note naming the sentence the user was hearing.

The echo check is a fallback. Loudspeakers without echo cancellation can still produce false interruptions and long utterances that mix the agent's voice with the user's. On PipeWire, load **libpipewire−module−echo−cancel** and point the commands at its nodes, for example

```
--record "pw-record --target echo-cancel-source --rate 16000 \
  --channels 1 --format f32 --raw -"
--play "pw-play --target echo-cancel-sink --rate 24000 \
  --channels 1 --format f32 --raw -"
```

(the node names are set in the module's configuration).

FILES

~/.config/cmvoiceagent/config
Options, one per line with its value after it; # starts a comment. The command line comes after them.

EXAMPLES

```
cmvoiceagent --asr Phonon-2/ --tts Kokoro-82M/ \
  Qwen3-30B-A3B-Instruct-2507.gguf --gpu --write
```

Typing instead of speaking, answers still spoken:

```
cmvoiceagent --asr Phonon-2/ --tts Kokoro-82M/ --keyboard model.gguf
```

ENVIRONMENT

CMGENERATE_THREADS
Worker threads for Parakeet and Kokoro in this process. cmagent reads it too. **−−threads** *N* overrides it here and is passed to cmagent.

SEE ALSO

cmagent(1), **cmtranscribe**(1), **cmspeech**(1)