

Comandos ffmpeg e transcript de audio

Comandos úteis/frequentes ffmpeg + *transcript* de audio

Extraindo audio de video

Arquivo de entrada foi AudioSemente.mov, gerou AudioSemente.wav que já é compatível com whisper.cpp

```
ffmpeg -i AudioSemente.mov -q:a 0 -map a AudioSemente.wav -y
```

Convertendo audio (mp3, ogg, etc) para wav

Arquivo gerado compatível com whisper.cpp

```
ffmpeg -i meeting-2025-03-12.ogg -ar 16000 -ac 1 -c:a pcm_s16le meeting-2025-03-12.wav
```

Mudando velocidade de vídeo

Manter o mesmo valor na opção **-vf** e **-af**, nesse caso **1.7** que acelerou o vídeo em 70%

```
ffmpeg -i Remedio-Obesidade.mov -vf "setpts=(PTS-STARTPTS)/1.7" -crf 18 -af atempo=1.7 Remedio-Obesidade-rapido.mov
```

Um exemplo de *slowdown* (metade da velocidade):

```
ffmpeg -i Remedio-Obesidade.mov -vf "setpts=(PTS-STARTPTS)/0.5" -crf 18 -af atempo=0.5 Remedio-Obesidade-rapido.mov
```

Convertendo um wav para mp3

Vendo se existe suporte a mp3 no ffmpeg:

```
ffmpeg -formats 2>/dev/null | grep mp3 # se deixar o stderr ir pra tela vai ser a lista de bibliotecas linkadas
```

O comando usando libmp3lame:

```
ffmpeg -i AudioSemente.wav -acodec libmp3lame AudioSemente.mp3
```

Outra opção de comando sem informar a lib de audio que deve ser usada (ffmpeg "escolhe" a melhor):

```
ffmpeg -i AudioSemente.wav -vn -ar 44100 -ac 2 -b:a 192k AudioSemente.mp3
```

Onde:

-vn disable video

-ar 44100 audio sampling frequency

-ac 2 number of audio channels (2 canais, não queremos stereo)

-b:a 192k converts the audio bit-rate to be exact 192 KB/s

Fazendo transcripts de audio

Instalando whisper.cpp

X Code command line tools:

```
xcode-select --install
```

Cmake, pode fazer download direto do site <https://cmake.org/download/> Mas se tiver usando brew:

```
brew install cmake
```

Baixando e compilando whisper.cpp:

```
git clone https://github.com/ggerganov/whisper.cpp.git
```

```
cd whisper.cpp
```

```
sh ./models/download-ggml-model.sh base.en
```

```
cmake -B build
```

```
cmake --build build --config Release
```

O whisper só aceita wav como entrada, o comando ffmpeg para converter áudios para ele:

```
ffmpeg -i meeting-2025-03-12.ogg -ar 16000 -ac 1 -c:a pcm_s16le meeting-2025-03-12.wav
```

Transcript de áudios em Inglês

Meu arquivo estava num diretório acima do whisper, por isso os `../`

```
./build/bin/whisper-cli -f ../AudioSemente.wav | tee -a ../AudioSemente.txt
```

Transcript de áudios em Português

Usando o modelo padrão do whisper (baixado na instalação) vai ter um output assim com áudios não-inglês:

```
[00:00:30.000 --> 00:00:33.920] (speaking in foreign language)
```

Não existe um modelo específico para português como existem os de inglês, então precisamos baixar um modelo genérico (várias línguas), dependendo do arquivo de entrada talvez precise de um modelo grande, os meus áudios todos funcionaram bem o modelo **medium**. Para baixar o modelo:

```
sh ./models/download-ggml-model.sh medium # execute o Shell sem argumentos para uma lista de modelos disponíveis
```

Comando passando o modelo e informando a língua:

```
./build/bin/whisper-cli -m ./models/ggml-medium.bin -l pt -f ../AudioSemente.wav | tee -a ../AudioSemente.txt
```

Dá para pedir para o whisper detectar a língua:

```
./build/bin/whisper-cli -m ./models/ggml-medium.bin -l auto  
-f ../AudioSemente.wav | tee -a ../AudioSemente.txt
```

Antes de começar o *transcript* ele vai imprimir qual a língua detectou:

```
whisper_full_with_state: auto-detected language: pt (p =  
0.989407)
```

Provavelmente não funciona com áudios multi-línguas, mas não testei.