

From Voice to Text

```
Discord (Legcord)
  ↓
PipeWire graph
  ↓
discord_sink (virtual null sink)
  ↓
discord_sink.monitor (loopback source)
  ↓
whisper_mic (remap-source, mono, 16kHz)
  ↓
ffmpeg
  ↓
audio.wav (growing file)
  ↓
whisper-stream
```

On the device used for transcription, a Discord client (Legcord) is running and joins the session's Discord voice channel. Legcord's audio output is moved to a virtual null sink (named `discord_sink`) via PipeWire. The sink's loopback source (`discord_sink.monitor`) is remapped by `whisper_mic` and fed into `ffmpeg`.

Whisper-stream uses the wav file created by `ffmpeg` to create the transcript.

Using a minimal language model (`tiny.en`), the voice input is transcribed with a delay of 5 to 10 seconds.