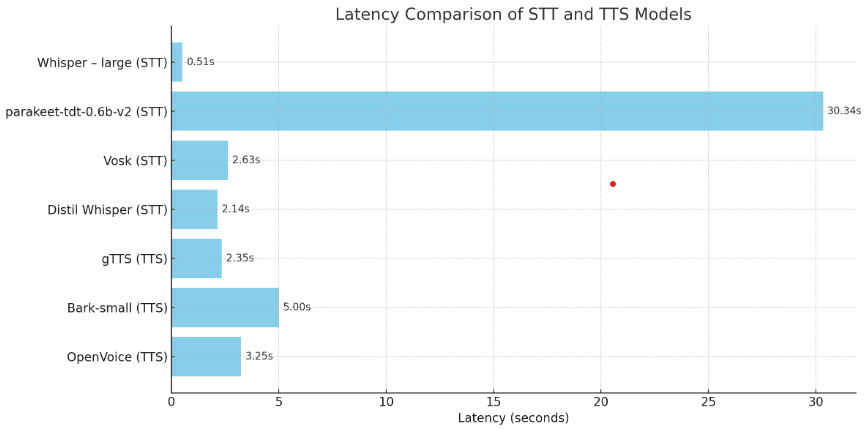
***Model Comparison Speech-to-Text,Text-to-Speech***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Model Name*** | ***Latency*** | ***Audio-duration*** | ***Link*** | ***Code Link*** |
| Whisper – large-STT | Latency: 0.51 seconds | 15.24 seconds | [GroqDocs - Build Fast](https://console.groq.com/docs/overview) | [Model-Comparison-Speech-to-Text/Whisper\_Latency\_Score.ipynb at main · Warishayat/Model-Comparison-Speech-to-Text](https://github.com/Warishayat/Model-Comparison-Speech-to-Text/blob/main/Whisper_Latency_Score.ipynb) |
| parakeet-tdt-0.6b-v2- STT | Latency: 30.34 seconds | 15.24 seconds  Gpu-required | [nvidia/parakeet-tdt-0.6b-v2 · Hugging Face](https://huggingface.co/nvidia/parakeet-tdt-0.6b-v2) | [Model-Comparison-Speech-to-Text/nvidia\_parakeet.ipynb at main · Warishayat/Model-Comparison-Speech-to-Text](https://github.com/Warishayat/Model-Comparison-Speech-to-Text/blob/main/nvidia_parakeet.ipynb) |
| Vosk- STT | Latency: 2.63 seconds | 15.24 seconds | [VOSK Offline Speech Recognition API](https://alphacephei.com/vosk/) | [Model-Comparison-Speech-to-Text/Vosk (1).ipynb at main · Warishayat/Model-Comparison-Speech-to-Text](https://github.com/Warishayat/Model-Comparison-Speech-to-Text/blob/main/Vosk%20(1).ipynb) |
| Destil Whisper- STT | 2.14 seconds | 15.24 seconds | [distil-whisper/distil-small.en · Hugging Face](https://huggingface.co/distil-whisper/distil-small.en) | [Model-Comparison-Speech-to-Text/Distil\_Whisper.ipynb at main · Warishayat/Model-Comparison-Speech-to-Text](https://github.com/Warishayat/Model-Comparison-Speech-to-Text/blob/main/Distil_Whisper.ipynb) |
| gTTS | 2.35 seconds | 28.68 seconds | [gTTS·PyPI](https://pypi.org/project/gTTS/) | [Model-Comparison-TTS-STT/gTTs.ipynb at main · Warishayat/Model-Comparison-TTS-STT](https://github.com/Warishayat/Model-Comparison-TTS-STT/blob/main/gTTs.ipynb) |
| Bark-small-TTS | ~4–6 seconds | Depends on text | [suno/bark-small · Hugging Face](https://huggingface.co/suno/bark-small) | ***-----*** |
| OpenVoice by MyShell-TTS | ~2.5–4.0s (varies) | Depends on text | [myshell-ai/OpenVoice: Instant voice cloning by MIT and MyShell. Audio foundation model.](https://github.com/myshell-ai/OpenVoice) | ***----*** |

***Graph:***

******