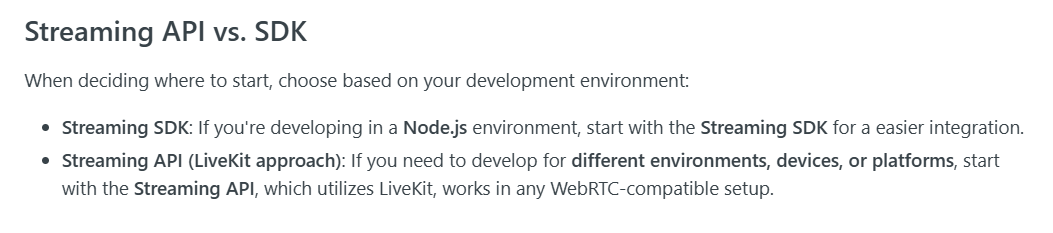
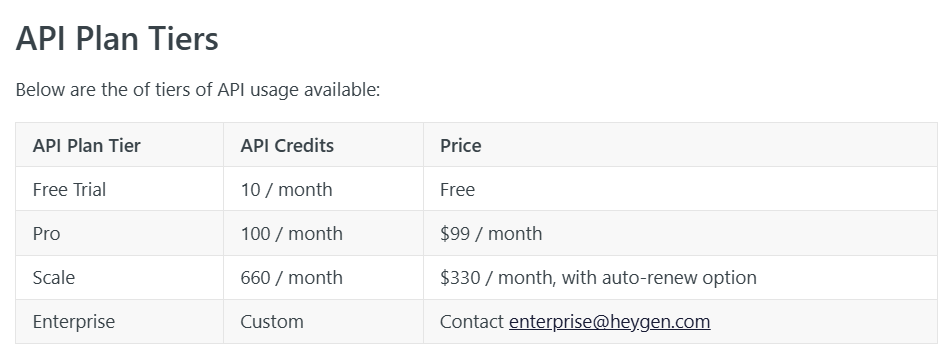
1. **HeyGen**

* Choosing API Vs SDK based on our environment setup.



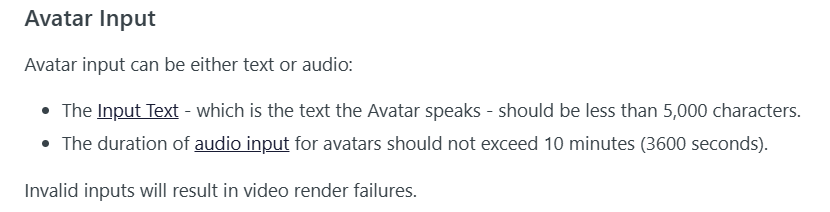
* Prices



* Credits Consumptions

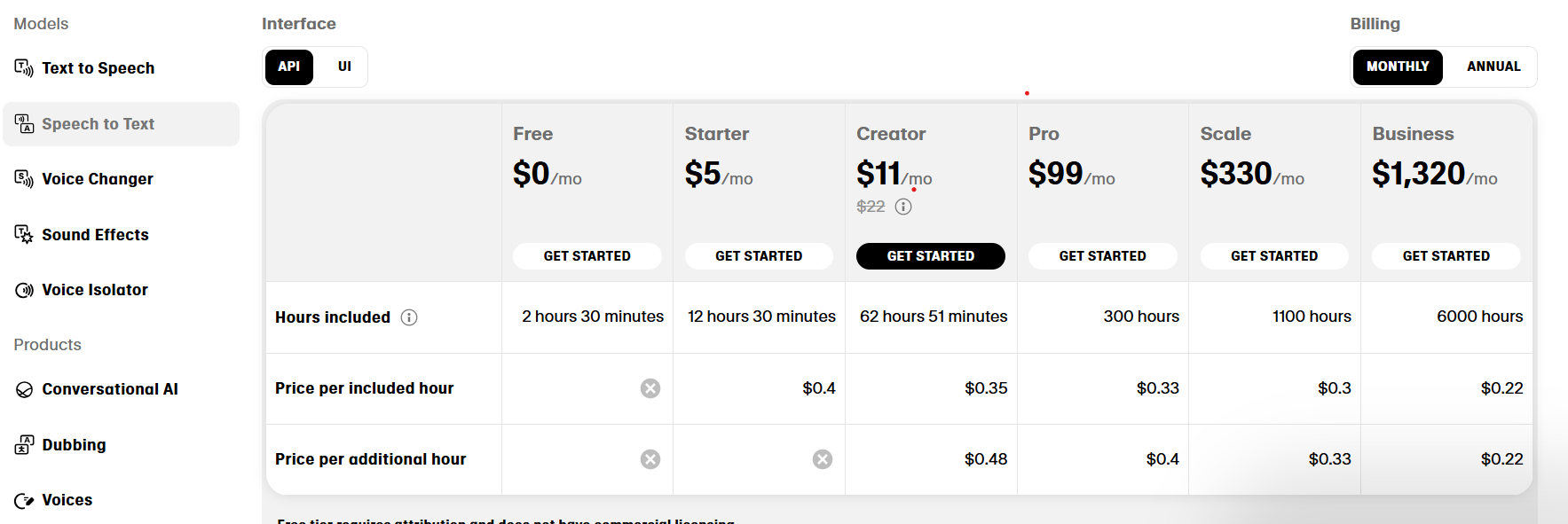


* Maximum input size.

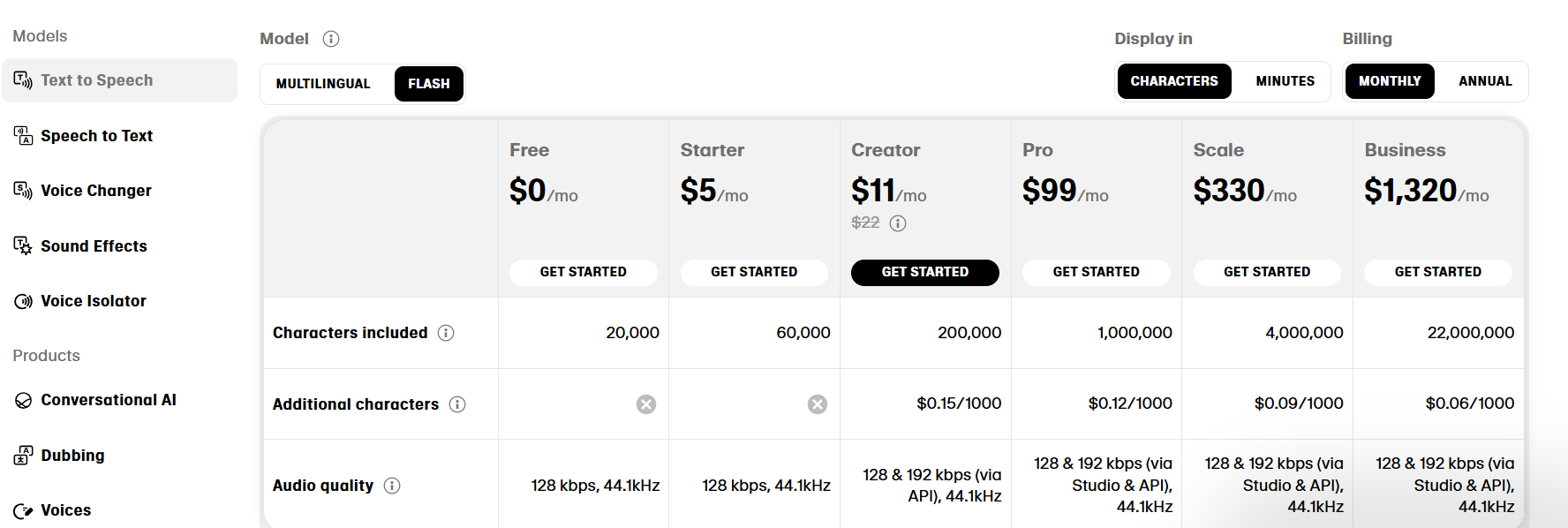


1. **ElevenLabs**

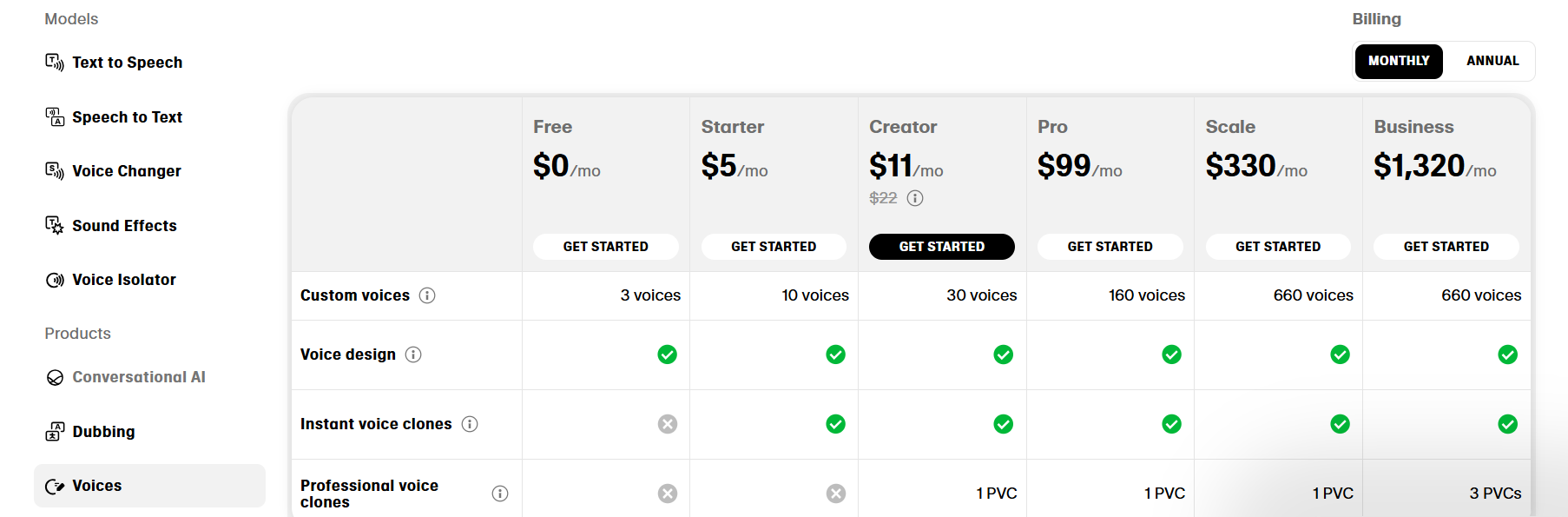
🡪 Speech To Text



* Text To Speech



* Voice Cloning



**Solution**

By reading the docs and API references of these two sites, I ended up with **three solutions.**

1. We use **HeyGen** for our all process, from creating avatar to end result with lip sync.
2. We mix up **HeyGen** and **ElevenLabs** for better performance, like we use **HeyGen** for **STT**, **LLM** and **avatar creation** and **lip sync** and **ElevenLabs** for **voice cloning** which they are best in.
3. We use **ElevenLabs** for **STT**, **TTS** and **voice cloning,** for **LLM** use **GPT-4o** and **HeyGen** for **avatar creation** and **lip sync.**

**Explanation**

* Above Screen Shots explicitly tell us that HeyGen is costly than ElevenLabs.
* We can use HeyGen for our all steps but the voice cloning feature of ElevenLabs is more powerful than HeyGen.
* But another option, in which we can mix things is also good but using more features of HeyGen may cost you high.
* Prices and Solutions are mentioned above, you can select based on also your research.
* If we use, Hegen for STT, LLM and avatar creation, its good to go, but what if some **glitches** happen on LLM response, instead of 5-6 words answer it gives us in 5-6 lines? On that cases it is better to have our own models but HeyGen don’t allow.

**Opinion**

1. We use **ElevenLabs** for **Speech to text**, **Text to speech** and **voice cloning** and finally use **HeyGen** for **avatar creation** and **lip sync** which may save your **money**. And used **GPT-4o** or why not we go with our **custom model** which may **save** more **money**.
2. Apart of these suggestions we have best open-source libraries which can be use to process our 3 steps out of 4.
3. Open-source libraries are, instead of **Whisper** we can use **Vosk** for **Speech to Text,** which is super amazing. For **Text to Speech**, we use **Coqui TTS** which is **pre-trained** model. Custom model for **LLM** and last but not least **avatar creation** with **HeyGen**.
4. Step 3 needs very **high-performance** system or you can say that required **GPU** for better **performance** to overcome the **latency** rate.