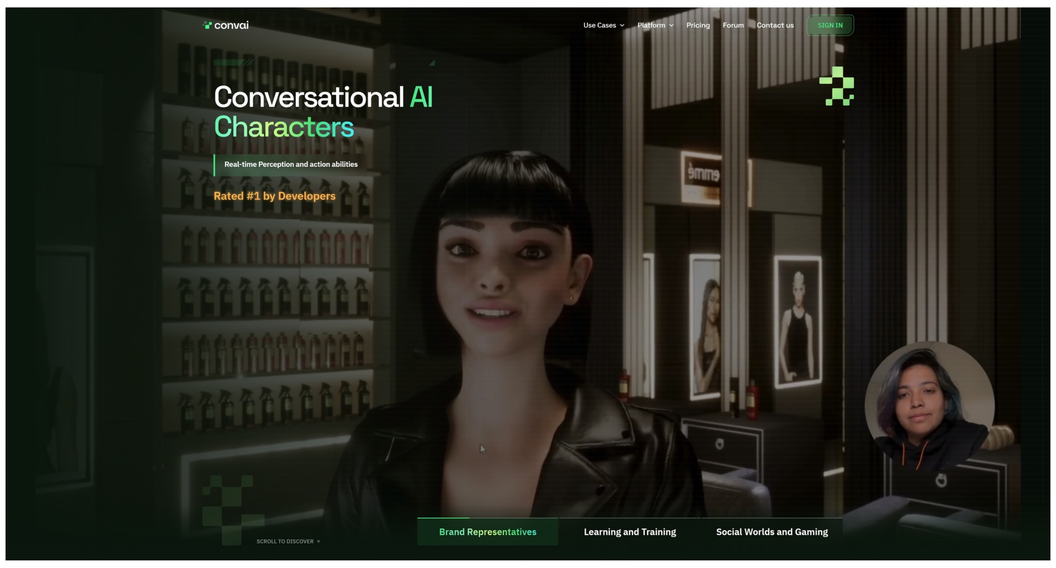
Towards Human–Avatar Communication

Training Hearing Impaired via Emotionally Adaptive Voice Synthesis

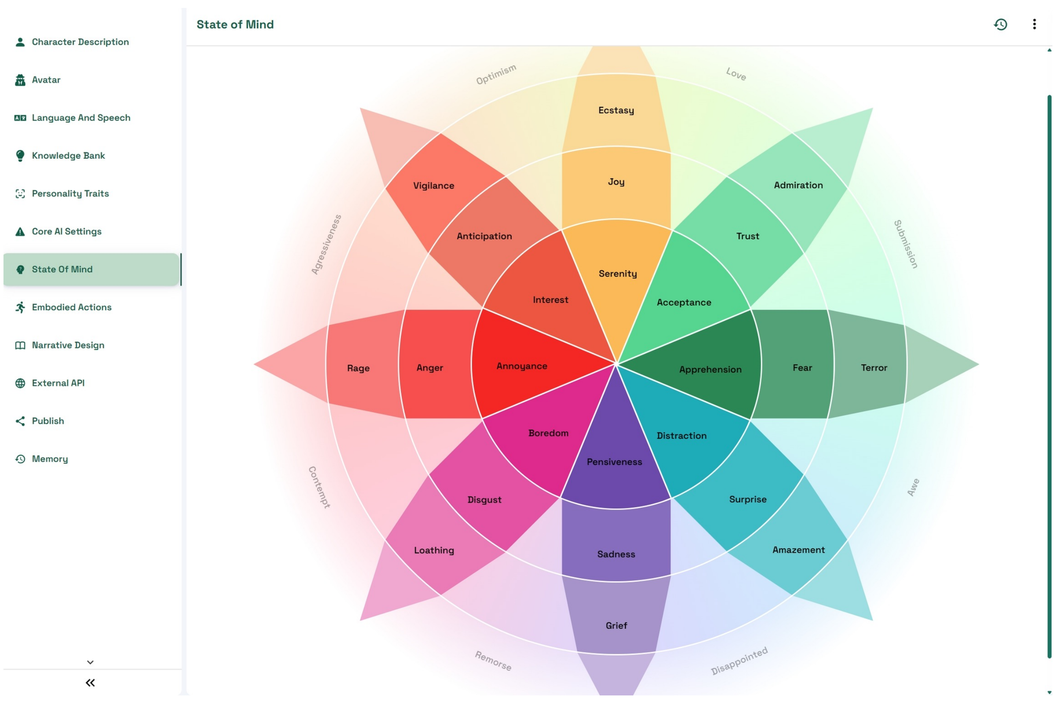
Background: LLM-driven avatars are increasingly employed in VR for training. How about Hearing-impaired?

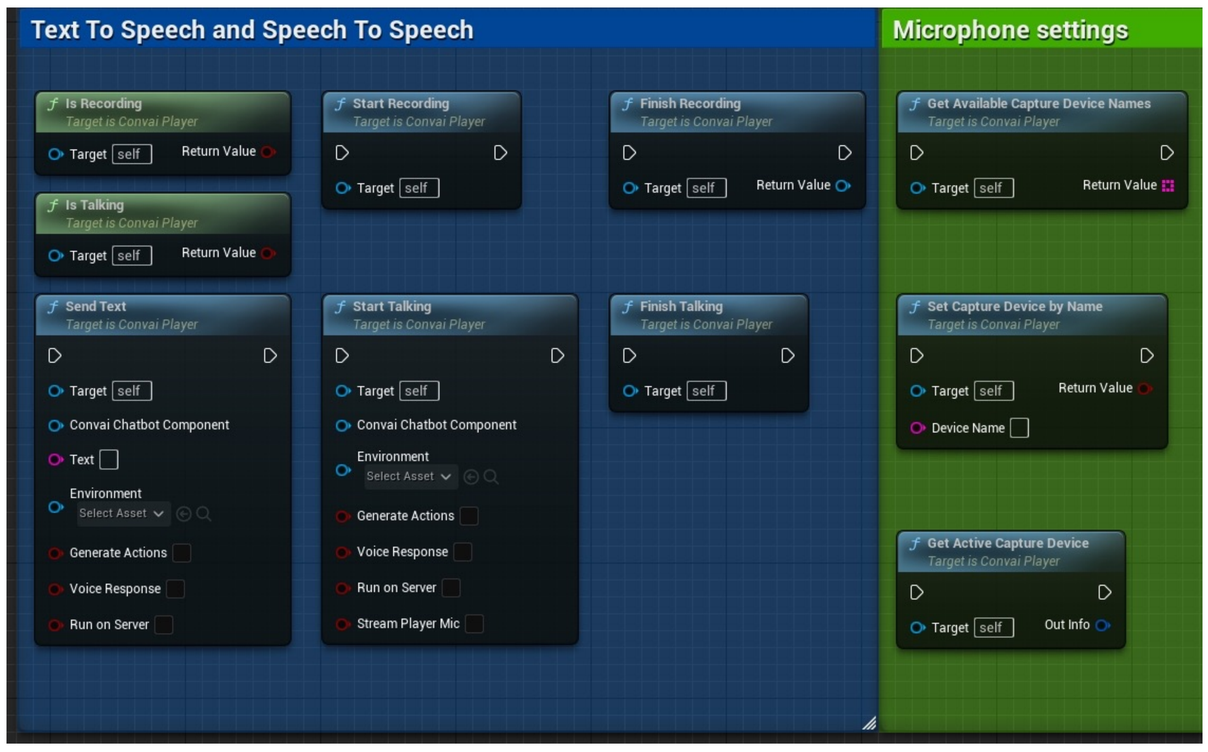
Bottleneck: Text-to-speech (TTS) : high latency, limited expressiveness, and lack of distinct speaker identities.

Can our real-time voice conversion plug-in independently control emotional sentiment and vocal timbre?



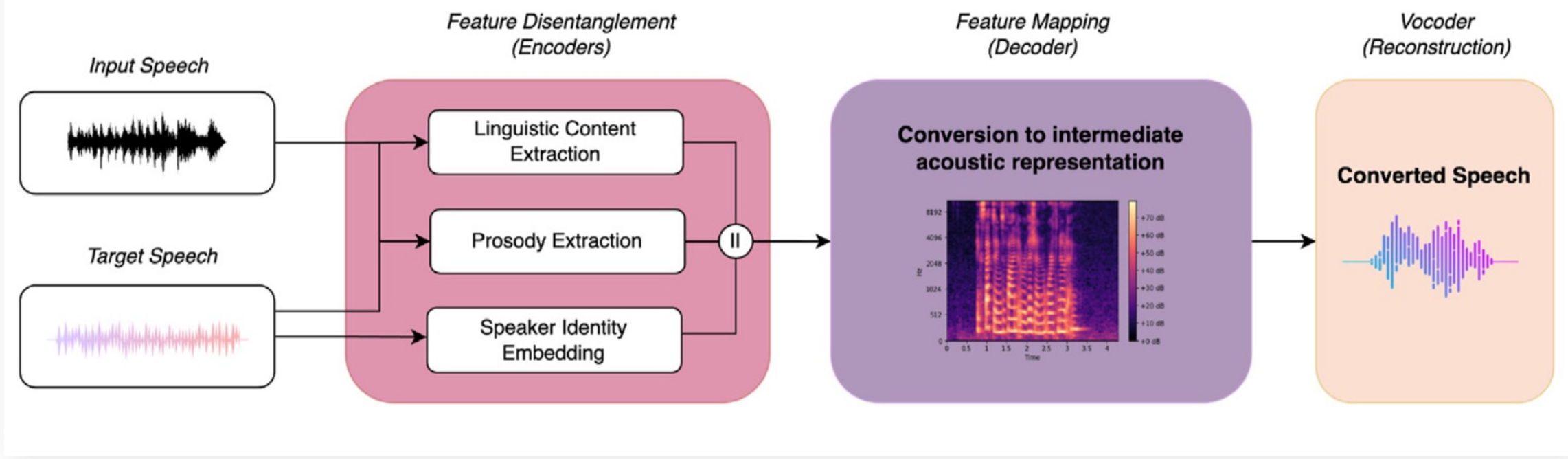
SoTA: convai





Our work

Plugin | Post TTS inference | Extensive Disentanglement | Voice Cloning (20 sec)



Aim: Test the recognition of emotional states and speaker variations in VR.

AALBORG UNIVERSITET

Anders R Bargum, Stefania Serafin, Cumhur Erkut
cuerlab
Copenhagen, Denmark