Voice Synthesis, Transformations, and Sound Effects: A portfolio

Erkut, Cumhur

Publication date:
2025

Voice Synthesis, Transformations, and Sound Effects: A portfolio

Cumhur Erkut,
with help of Christie Laurent, Juan Alonso, David Südholt, and Anders R Bargum

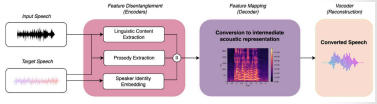
Intro

- Started physically on aging formants
- Benefited in DL disentanglements in DDSP, RAVE -> SRAVE, back to music

Methods

1. Tunable Voice Aging Model
2. MFCC playback using Z in DDSP
3. DDSP cross-synthesis, plugins
4. Full embeddings in S-RAVE

Results



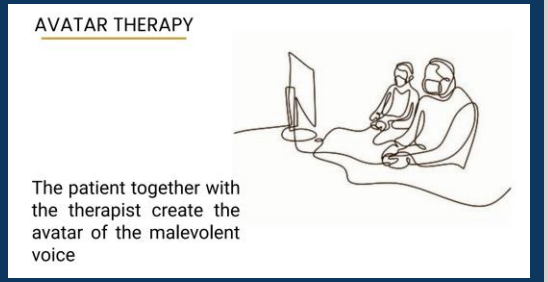
- Significant performance: RT from 1 to 15, transformation data from 20 minutes to 30 seconds
- Concerns: Voice spoofing

Discussion

- Diverse portfolio (Hearing Aids, XR, Robotics, Embedded AI)
- PINNS <.-> Transformers
- Active Inference (Bayesian)
- Evaluation: Audio Aesthetics?

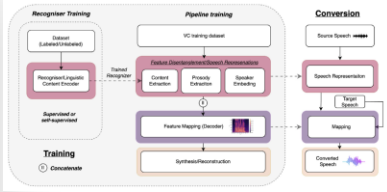
AALBORG UNIVERSITET

Interplay between sound, music, and speech. Transformers for embeddings, diffusion for real-time synthesis.



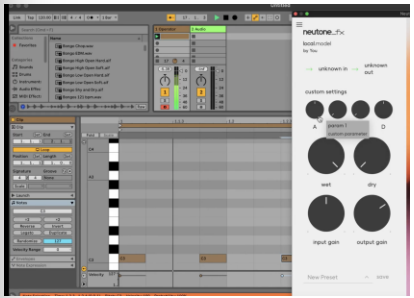
Scan to hear more at cerkut.github.io, **click on photos to read more about talented Christie Laurent, Juan Alonso, David Südholt, and Anders R Bargum**

Recent Work



General Voice Conversion pipeline: S-RAVE

Commercial deployment possible: BabyAudio Humanoid



Back to musical instruments