Beitrag ID: **48** Beitragskennung: **121** Typ: **Poster**

In-the-wild Speech Emotion Conversion Using Disentangled Self-Supervised Representations and Neural Vocoder-based Resynthesis

Speech emotion conversion aims to convert the expressed emotion of a spoken utterance to a target emotion while preserving the lexical information and the speaker's identity. In the context of human-machine interaction systems (e.g., social robots), to improve the naturalness of machine communication, the generation of emotionally expressive speech is required. In this work, we introduce a methodology that uses self-supervised networks to disentangle the lexical, speaker, and emotional content of the utterance, and subsequently uses a HiFiGAN vocoder to resynthesise the disentangled representations to a speech signal of the targeted emotion. Results confirm that the proposed approach is aptly conditioned on the emotional content of input speech and is capable of synthesising natural-sounding speech for a target emotion.

Find me @ my poster

1, 2

Keywords

Speech Emotion Conversion
HiFiGAN
Self-supervised representations
Speech Synthesis
Human-machine interaction

TentID

Autor: RAJ PRABHU, Navin (Signal Processing)

Co-Autoren: Dr. LEHMANN-WILLENBROCK, Nale (Industrial and organizational psychology); GERKMANN, Timo