

Research on the method of transforming speaker emotion in speech using non-negative matrix factorization technique

Nguyen Van Dan

Abstract: This study proposes a novel method for speaker emotion transformation in speech using Non-negative Matrix Factorization (NMF). The approach leverages NMF's ability to decompose speech spectrograms into a dictionary of spectral bases and corresponding activation weights, which can be interpreted as representing phonetic content and prosodic or emotional characteristics, respectively. By training separate NMF dictionaries on neutral and target emotional speech from the same speaker, we develop a transformation framework that modifies the activation patterns of neutral speech to match those of the target emotion while preserving the original spectral bases. This allows for the transfer of emotional prosody without altering the underlying linguistic content. Experimental evaluations on a Vietnamese emotional speech corpus demonstrate that the proposed method effectively converts neutral speech into happy, sad, and angry expressions. Objective measures, including Mel-cepstral distortion and spectral centroid shifts, confirm significant acoustic changes, while subjective listening tests indicate a high degree of perceived emotional naturalness and speaker identity retention. The findings highlight the potential of NMF as a robust, unsupervised tool for emotion conversion in voice applications.

Keywords: voice emotion conversion, non-negative matrix factorization, speech processing, speaker emotion transformation