

Research on a method for transforming speaker emotion in speech using time-domain speech decomposition techniques

Do Xuan Toan

Abstract: This study proposes a novel method for speaker emotion conversion in speech using time-domain speech decomposition techniques. Unlike conventional approaches that rely on spectral parameter manipulation, the proposed method decomposes the speech signal into distinct temporal components, including excitation, vocal tract, and aperiodic features, using a deep neural network-based decomposition framework. Emotion conversion is achieved by selectively modifying the temporal dynamics of these components, guided by a learned emotion mapping function. Objective evaluations, including mel-cepstral distortion and F0 root mean square error, demonstrate that the proposed method effectively transfers target emotional attributes while preserving speaker identity and speech naturalness. Subjective listening tests further confirm significant improvements in emotional expressiveness and overall quality compared to baseline voice conversion systems. The findings highlight the potential of time-domain decomposition for fine-grained, speaker-independent emotion conversion, offering a robust alternative to frequency-domain methods. This approach holds promise for applications in expressive speech synthesis, human-computer interaction, and assistive communication technologies.

Keywords: voice emotion conversion, speech decomposition, time-domain speech processing, speaker emotion transformation