

A Novel Trajectory-based Spatial-Temporal Spectral Features for Speech Emotion Recognition

張鈞閔 Chun-Min Chang
國立清華大學電機工程學系
Department of Electrical Engineering
National Tsing Hua University
cmchang@gapp.nthu.edu.tw

林維誠 Wei-Cheng Lin
國立清華大學電機工程學系
Department of Electrical Engineering
National Tsing Hua University

李祈均 Chi-Chun Lee
國立清華大學電機工程學系
Department of Electrical Engineering
National Tsing Hua University
clee@ee.nthu.edu.tw

Abstract

Speech is one of the most natural form of human communication. Recognizing emotion from speech continues to be an important research venue to advance human-machine interface design and human behavior understanding. In this work, we propose a novel set of features, termed trajectory-based spatial-temporal spectral features, to recognize emotions from speech. The core idea centers on deriving descriptors both spatially and temporally on speech spectrograms over a sub-utterance frame (e.g., 250ms) - an inspiration from dense trajectory-based video descriptors. We conduct categorical and dimensional emotion recognition experiments and compare our proposed features to both the well-established set of prosodic and spectral features and the state-of-the-art exhaustive feature extraction. Our experiment demonstrate that our features by itself achieves comparable accuracies in the 4-class emotion recognition and valence detection task, and it obtains a significant improvement in the activation detection. We additionally show that there exists complementary information in our proposed features to the existing acoustic features set, which can be used to obtain an improved emotion recognition accuracy.

Keywords: emotion recognition, speech processing, spatial-temporal descriptors, Mel-filter bank energy