Determining Confidence Measures on Fundamental Frequency Estimations

Boyuan Deng

August 2016

MSc. Thesis

Supervised by:
Denis Jouvet, Inria Nancy-Grand Est, France
Ingmar Steiner, Saarland University, Germany

Co-supervised by:
Yves Laprie, CNRS, France

Abstract

Fundamental frequency estimation acts as the basis of many applications related to speech or music, such as speaker recognition, development of speech synthesis voices and music information retrieval. Various algorithms working in the time or frequency domain have been developed in the past and work quite well on good-quality signals. They become more error prone on lower-quality signals but none of them indicate how reliable the estimation results are, which is a crucial metric if the overall system makes decisions based on probabilities or multiple estimation algorithms are combined using ensemble methods.

In our work, we tackle this problem by developing and evaluating confidence measures based on machine learning approaches for several existing fundamental frequency estimation algorithms. Such confidence measure can be seen as the probability of the estimation result to be correct. Sophisticated neural network architectures, including long short-term memory networks (LSTM), are chosen for the construction of these measures. And numerous features computed on the signals and generated during estimations are used as input.

Besides confidence measures, our work also illustrates some empirical analyses on the robustness of existing fundamental frequency estimation algorithms. Data from an audio corpus is artificially distorted in a quantitatively controlled way, by modifying signal levels or adding certain types of noise at given signal-to-noise ratios, etc. Changes in error rates as well as the different error types (deviation or error in voiced/unvoiced decisions) are visualized and the underlying causes are analyzed in detail. The conclusions drawn may help guide the choice of fundamental frequency estimation algorithms depending on situations and needs.