

Dual Stage Probabilistic Voice Activity Detector

Ivan Tashev¹⁾, Andrew Lovitt²⁾, and Alex Acero¹⁾

¹⁾Microsoft Research, ²⁾Microsoft Corporation
{ivantash, anlovi, alexac}@microsoft.com

Abstract

Voice activity detectors (VAD) are integral part of the modern speech processing, speech enhancement and speech encoding systems. One of the major problems in practical realizations is to achieve robust VAD in conditions of background noise. Most of the statistical model-based approaches employ the Gaussian assumption in the discrete Fourier transform (DFT) domain, which deviates from the real observation. In this paper, we propose a class of VAD algorithms based on several statistical models of the probability density functions of the magnitudes. In addition, we evaluate several approaches for combining the likelihoods for each frequency bin for estimation of the likelihood for the entire frame. A data corpus with in-car noise is then used to evaluate the VAD and the results are discussed.