PROPAGATION OF STATISTICAL INFORMATION THROUGH NON-LINEAR FEATURE EXTRACTIONS FOR ROBUST SPEECH RECOGNITION

R. F. Astudillo\textsuperscript{1}, D. Kolossa\textsuperscript{2}, R. Orglmeister\textsuperscript{3}

Electronics and Medical Signal Processing Group, TU Berlin, 10587 Berlin, Germany
Email: \textsuperscript{1}ramon@astudillo.com, \textsuperscript{2}d.kolossa@ee.tu-berlin.de, \textsuperscript{3}reinhold.orglmeister@tu-berlin.de

Abstract

Automatic speech recognition systems often rely on statistical noise suppression methods to increase their recognition performance in non-stationary noisy environments. However, even with a good approximation of the noise power spectrum, the estimated clean signal contains residual noise along with artifacts introduced by speech estimation inaccuracies. In this paper, we show that this can be compensated by propagating a measure of the uncertainty of estimation through the feature extraction process and combining it with missing feature techniques directly in the feature domain. In the suggested method, each Fourier coefficient estimated by the Ephraim-Malah Filter\cite{1} is first replaced by a complex Gaussian distribution with mean equal to its estimated value and variance proportional to the uncertainty derived from the estimation process. By using a combination of analytical and Pseudo Monte Carlo methods, this statistical information is then efficiently propagated through the feature extraction into the feature domain, were we are able to combine it with the statistical information of the trained speech recognizer, for example by using modified imputation\cite{2} or uncertainty decoding\cite{3}, and thus reduce the effect of inexact clean signal estimation. We are therefore not limited to spectral domain as in many customary missing feature methods, thus more suitable domains e.g. MFCC or PLP-RASTA can be used.

References:


Key Words: Unscented Transform, Complex Gaussian Distribution, Rice Distribution, MMSE-LSA Bayesian Estimation, Missing Feature Techniques