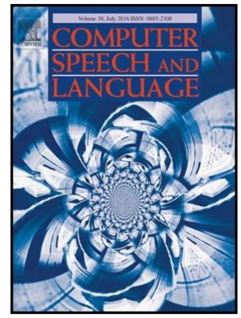
Accepted Manuscript



Title: Estimating acoustic speech features in low signal-to-noise ratios using a statistical framework

Author: Philip Harding, Ben Milnercor

PII: S0885-2308(16)30027-4
DOI: http://dx.doi.org/doi: 10.1016/j.csl.2016.08.001
Reference: YCSLA 787

To appear in: *Computer Speech and Language*

Received date: 22-2-2016
Revised date: 15-6-2016
Accepted date: 4-8-2016

Estimating acoustic speech features in low signal-to-noise ratios using a statistical framework

Philip Harding, Ben Milnercor*

School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

*Corresponding author

Email address: b.milner@uea.ac.uk (Ben Milner)

Highlights

- We propose an integrated statistical framework to estimate acoustic speech features
- Speaker and noise adaptation methods are developed and applied to noise–free models
- Compared to other methods, the proposal is more accurate particularly at low SNRs

Abstract

Accurate estimation of acoustic speech features from noisy speech and from different speakers is an ongoing problem in speech processing. Many methods have been proposed to estimate acoustic features but errors increase as signal-to-noise ratios fall. This work proposes a robust statistical framework to estimate an acoustic speech vector (comprising voicing, fundamental frequency and spectral envelope) from an intermediate feature that is extracted from a noisy time-domain speech signal. The initial approach is accurate in clean conditions but deteriorates in noise and with changing speaker. Adaptation methods are then developed to adjust the acoustic models to the noise conditions and speaker. Evaluations are carried out in stationary and nonstationary noises and at SNRs from -5dB to clean conditions. Comparison with conventional methods of estimating fundamental frequency, voicing and spectral envelope reveals the proposed framework to have lowest errors in all conditions tested.

Keywords: Voicing, Fundamental frequency, Spectral envelope, Noise adaptation, Speaker adaptation

1. INTRODUCTION

Acoustic speech features take many forms and include parameters such as voicing, fundamental frequency, spectral envelope, formant frequencies and voice activity. Excitation features, such as voicing and fundamental frequency, are used in many speech processing applications and include, for example, speech coding, enhancement, noise estimation, automatic