Using an intelligibility measure to create noise robust cepstral coefficients for HMM-based speech synthesis

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proc. LISTA Workshop

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Using an intelligibility measure to create noise robust cepstral coefficients for HMM-based speech synthesis

Cassia Valentini-Botinhao\textsuperscript{1}, Yan Tang\textsuperscript{2}, Junichi Yamagishi\textsuperscript{1}, Simon King\textsuperscript{1}

\textsuperscript{1} The Centre for Speech Technology Research, University of Edinburgh, UK
\textsuperscript{2} Language and Speech Laboratory, Universidad del País Vasco, Spain
C.Valentini-Botinhao@sms.ed.ac.uk, y.tang@laslab.org, jyamagis@inf.ed.ac.uk, Simon.King@ed.ac.uk

Abstract

The aim of this work is to increase intelligibility of HMM-based synthetic speech in noisy environments by modifying clean synthetic speech given that noise is known. For that purpose we need a measure for intelligibility of speech in noise that can automatically define the sort of modifications that we need to apply. In previous experiments\textsuperscript{[1]} we have observed that spectrum envelope modifications can have a significant positive impact on the intelligibility of HMM-generated synthetic speech in noise and that the Glimpse proportion measure (GP)\textsuperscript{[2]} is highly correlated with subjective scores under those circumstances.

We have then introduced a method for cepstral coefficient extraction that modifies spectrum envelope based on the GP measure. The GP accounts only for the effect of additive noise, not requiring a reference unmodified speech signal to produce an intelligibility prediction. To control the amount of distortions introduced by the modification we extract cepstral coefficients using an optimization criterion with two terms. The first term accounts for the minimization of the mismatch between natural speech periodogram and magnitude spectrum as modeled by cepstral coefficient, the current criterion used for cepstral coefficient extraction performed at the training stage of the HMM-based speech synthesis framework\textsuperscript{[3]}. The second term accounts for the maximization of an approximated analytical and differentiable version of the GP measure. Using this method we found significant intelligibility gains however not for all tested noise types which indicates that we need a more effective method for controlling distortions\textsuperscript{[4]}.

In this work we propose to limit the frequency resolution of the modifications, and therefore the amount of distortions, by altering only the first few cepstral coefficients, known to be responsible for the coarse frequency resolution of the spectrum. Fig.1 shows the long term average spectrum of original and modified speech, where we can see the effect that limiting the degrees of freedom has on the spectrum envelope. Listening experiments results as shown in Fig.2 indicates that when we modify less coefficients we can improve intelligibility even further.

1. References


