Effects of periodicity and vocoding in modelling the intelligibility of speech in background noise

Kurt Steinmetzger and Stuart Rosen
UCL Speech, Hearing and Phonetic Sciences, London, UK

Introduction

Auditory models concerned with predicting the intelligibility of speech presented in background noise are usually evaluated with masks that vary widely regarding both their envelope fluctuations and spectral properties, for example speech-shaped noise and interfering talkers. Here we took the opposite route and investigated how model predictions are affected by a single acoustic factor, the presence or absence of periodicity in both masker and target speech. Secondly, little is known about how acoustic degradations of the target speech itself affect the performance of common auditory models. We thus additionally lowered the spectral resolution of the target speech using a channel vocoder.

Methods

Periodicity in the target speech was either discarded using a noise-vocoder or synthesised with a pulse train following the natural F0-contour. Harmonic complexes based on the F0-contours of real speech were used as periodic masks, while speech-shaped noise was used as aperiodic masker. A number of existing models – extended speech intelligibility index (ESII; Rhebergen & Versfeld, 2005), speech-based speech transmission index (sSTI; Goldsworthy & Greenberg, 2004), multi-resolution speech-based envelope power spectrum model (mr-sEPSPM; Jørgensen et al., 2013), and short-time objective intelligibility measure (STOI; Taal et al., 2011) – were used to predict previously obtained behavioural data (Steinmetzger & Rosen, 2015).

Results

In line with the data, all four models predicted that listeners would benefit from periodicity in the masker, but the size of this effect was consistently underestimated (Figure 2). Secondly, although all models correctly predicted that preserving periodicity information in the target speech did not have an effect on intelligibility rates, only the STOI was able to account for the substantially lower intelligibility of vocoded speech (Figure 3).

Discussion

These results suggest that spectral information is not sufficiently represented in current auditory models and that additional processing steps may be needed. Additionally, a correlation-based back end, as used in the STOI, appears to be a promising technique that seems to account for a wider range of conditions than traditional SNR-based measures.

References


Acknowledgements

This project has been funded with support from the European Commission under Contract FP7-People-2011-290000 and the Medical Research Council (Grant Number G1001255). We thank Johannes Zaar, Torsten Dau, and the rest of the DTU Hearing Systems group for advice and code.