Understanding vcooded speech with two types of upward spectral frequency mismatch in cochlear implants vs. shrinking speakers
Alexis Hervais-Adelman, Annick Aubin-Pouliot, Melanie Knowles, Roy D Patterson
Centre for the Neural Basis of Hearing, Physiology, Development and Neuroscience Department, University of Cambridge, Downing Street, Cambridge, CB2 3EG United Kingdom.

1. Introduction

Cochlear implants (CIs) restore the sensation of hearing in profoundly deaf individuals by stimulating the auditory nerve with an electrode array placed in the cochlea. Due to the morphology of the cochlea, electrode arrays cannot be inserted to the apex. Thus, sounds transduced by the device stimulate regions of the cochlea that code for higher frequency signals than the source. This basalward shift induces a place frequency mismatch, which is thought to severely reduce the intelligibility of speech signals transduced by CIs.

Noise Vocoding is a technique used to simulate CI processors (Shannon et al., 1995) for normally-hearing listeners. Upward spectral envelope shifts can be applied to NV speech to simulate a basalward shift of the electrode array.

2. Results

Data averaged over participants and trials. Error bars represent +/-1 standard error of the mean.

- Significant main-effect of shift magnitude ($F(4,16)= 42.88$, $p<0.001$, partial eta-squared = 0.915). The greater the simulated shift, the more difficult the syllables are to correctly identify.
- Post-hoc pairwise comparisons showed that performance in the 4.8mm shift is significantly worse than any other condition, 4.8mm shift performance is significantly worse than no shift or 1.6mm.
- Significant main-effect of spectral detail ($F(3,12)= 333.64$, $p<0.001$, partial eta-squared = 0.988). Decreasing spectral detail renders syllables more difficult to identify.
- Post-hoc pairwise comparisons show that there is a significant difference in performance between all pairs of spectral detail conditions, except 4 vs. 12 channels which is marginally significant ($p=0.054$).
- Main effect of shift type is not significant ($p=0.086$).

Although there is no significant main effect of shift type, it is involved in significant interactions.

There is a significant interaction of spectral detail and shift type ($F(12,18)= 6.138$, $p=0.009$, partial eta-squared= 0.607). The graph shows the data collapsed over all shift magnitudes, and it can be clearly seen that at intermediate levels of spectral detail, syllable-identification performance is significantly more impaired by spectral degradation in the tonotopically-shifted condition than the scale-shifted condition.

There is also a marginally-significant ($p=0.052$) three-way interaction between shift type, shift magnitude and spectral detail, such that the effect of spectral detail on intelligibility is more deleterious as the magnitude of shift increases, and this effect is more pronounced in the tonotopic shift than scale-shift condition.

4. Conclusions

The data indicate that under certain circumstances, the tonotopic shift induced by the place frequency mismatch of CI electrode arrays at shallow insertion depths is more deleterious to speech intelligibility than are the scale shifts produced by artifically "shrinking" a speaker.

While tonotopic shifts are not encountered in natural listening environments, humans speakers of various sizes do exist, and the auditory system appears to have a mechanism to cope with this type of variability.

Re-predicting a finding by Fu & Shannon (1999), there is no interaction between the effect of spectral detail and shift magnitude on intelligibility, suggesting that even at low spectral resolution there are still cues that permit the normalisation of shifted speech.

Given that the auditory system is geared to handle scale shifts, if CIs could minimise the warping of the frequency axis induced by the tonotopic shift of the electrode array this could enhance intelligibility of CI speech.

Pre-shifting the spectral envelope of speech upwards, so that the disparity between analysis and synthesis filters in the cochlea occurs in a region where the tonotopic and log-frequency axes are less divergent could help in this regard, although information about speaker size would be useful.

References


