Comments

Comment by Kollmeier:

You processed both the speech and the interferer through the same nonlinear system and derived the intelligibility at the output. Due to the nonlinearities included in the processing (such as, e.g., envelope and instantaneous frequency extraction) your method is significantly different from first processing speech and interferer separately and then mixing both before presenting the result to the human listener (a method described by Hohmann and Kollmeier 1995). Your current method might therefore primarily estimate the detrimental effect of your nonlinear processing algorithm on a linear mixture of signals rather than estimating the salience of the cues preserved in your procedure for speech perception. Did you test the intelligibility of your processed sentences in noise or other conditions after processing the target speech in isolation? Would you expect the same results for this condition?

Hohmann, V. and Kollmeier, B. (1995) The effect of multichannel dynamic compression on speech intelligibility. J. Acoust. Soc. Am. 97, 1191-1195.

Reply:

We agree that the non-linearity in processing would make our methods significantly different from yours. We did not test the intelligibility under conditions where the target and the masker were processed in isolation before they were mixed for presentation to the listeners. Had we done that, we would expect better performance than observed because of the additional nonlinear interaction between the target and the masker.

We disagree with the assertion that the order of processing implies that our results "estimate the detrimental effect of your nonlinear processing algorithm on a linear mixture of signals rather than estimating the salience of the cues preserved in your procedure for speech perception." In realistic listening situations, one would not be able to obtain a clean copy of the target or the masker but would need to deal with the mixed signal. The main point of our study is that auditory performance would be greatly improved if we could encode the fine structure, or the slowly-varying version of it, in this mixed signal.

Addendum by Zeng et al.:

It has been brought to our attention that Sheft and Yost (2001) also evaluated the relative contributions of envelope and phase modulations to auditory signal classification. They processed the auditory signal through a six-channel filterbank and extracted the envelope and phase modulation functions from each channel. By systematically manipulating the envelope filter and the carrier type and frequency, they found that signal classification depends on both low-rate amplitude and phase modulations.

Sheft, S. and Yost, W.A. (2001) Auditory abilities of experienced signal analysts. AFRL Prog. Rep. 1, contract no. SPO700-98-D-4002.