Comments

Comment by Kollmeier:

The template-match localization model that you suggest is very close to recent models by the Blauert group and some recent work from our group (Nix and Hohmann 1999). There is also a close relation to Articulation Index-type models if the direct-to-reverberation ratio is used instead of the SNR. Would you therefore expect that your model would also describe speech intelligibility in reverberant conditions if appropriate frequency weighting factors are employed?

Nix, J. and V. Hohmann (1999) Statistics of binaural Parameters and Localization in Noise. In: T. Dau, V. Hohmann, B. Kollmeier (Eds), *Psychophysics, Physiology and Models of Hearing.* Singapore, World Scientific Publishing Co. pp. 263-266.

Reply:

The extended version of this paper, as presented at Dourdan, described our view that in highly reverberant environments the perception of location tends to be dominated by interaural level differences (ILDs). The template-matching localization model in question imagines that the listener has templates for ILDs, one for each one-third-octave band. Templates are straight lines relating ILD to azimuth, and the model assumes that the slopes of the lines are learned by the listener with experience in the room. Apart from the linearization of templates and the restriction to ILDs, this model resembles the Bayesian classification model by Nix and Hohmann. Our template model proved to be an adequate proof of the ILD concept in that it successfully accounted for localization accuracy as a function of direct to reverberant ratio and as a function of noise cutoff frequency.

The ILD templates are time independent. They are best fits to averaged spectral responses, measured over intervals that are long compared to room reverberation times. Thus, the model assumes that listeners use temporally-averaged ILDs as localization cues (Hartmann and Constan 2002). Although we cannot do it yet, we would expect eventually to be able to predict the slopes of the ILD templates from the direct to reverberant ratios in the corresponding bands.

Professor Kollmeier has asked whether our template model can be related to speech intelligibility in rooms. Yes, perhaps. In the limit of very long reverberation time reverberated speech would resemble steady background noise, and a simple signal to noise ratio would apply. In more usual cases, a key element in speech intelligibility, as measured by articulation index and as predicted by the speech transmission index (Houtgast and Steeneken 1985), is the modulation transfer function (Tchorz and Kollmeier 1999). Such temporal characteristics of a room cannot be calculated from the averaged spectral response because there is inadequate phase information.

Hartmann, W.M. and Constan, Z.A. (2002) Interaural level differences and the level-meter model. J. Acoust. Soc. Am. 112, 1037-1045.

Houtgast, T. and Steeneken, H.J.M. (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J. Acoust. Soc. Am. 77, 1069-1077.

Tchorz, J. and Kollmeier, B. (1999) A model of auditory perception as a front end for automatic speech recognition. J. Acoust. Soc. Am. 106, 2040-2050.