Influence Of A Microphone Array On Speech-On-Speech Masking Psychometric Functions

Sylvain Favrot[1], Christine R. Mason[1] and Gerald Kidd Jr.[1]

[1] *Boston University, Hearing Research Center, 635 Commonwealth Ave Boston, MA 02215, USA, E-Mail: sfav@bu.edu*

Introduction

Listeners with hearing loss often have a reduced ability to perceptually segregate and selectively attend to one specific talker among several competing talkers. One potential solution to this problem is to use beamforming microphone arrays, which selectively amplify the incoming sound from a specific (target) direction while attenuating unwanted sounds emanating from other directions (maskers). We have recently used speech-on-speech masking to investigate the performance of such arrays [1]. In that study, the psychometric functions relating speech intelligibility performance to target-to-masker ratio often exhibited a "plateau" or "dip" when the masker talkers were colocated with the target talker. This effect was highly subject dependent, and has also been reported in previous studies (e.g., [2], [3]) where it mainly appeared with one or two masker talkers uttering speech material that was very similar to the target. Importantly, this effect disappears when maskers are spatially separated from the target because spatial cues help the listener to segregate the maskers from the target reducing source confusions.

The aim of the present study was to investigate whether plateau effects also occur when listening through a highly directional microphone array. This seemed to be possible because a beamforming array produces a single channel output resulting in an auditory image similar to that which occurs when sources are colocated. To evaluate this possibility, it was necessary first to quantify the plateau effect and second to provide a measure of the intelligibility change with the spatial separation of the maskers when this effect occurs.

Methods

Four normal-hearing subjects (aged 19 to 21) participated in a speech-on-speech masking experiment. Their task was to identify the words spoken by a target talker in the presence of two simultaneous talkers uttering similar sentences (e.g., [1]). The target and speech maskers were from a closed-set laboratory-designed corpus ("BU corpus"; [4]) that consists of 5-word strings. On every trial the words were randomly selected from eight exemplars in each category; for example: "Sue found three red shoes." The target talker was identified by the name "Sue" with scoring based on correctly identifying individual words in the sentence.

In a large singled-walled IAC booth, five loudspeakers were arranged along a semicircle at 0º, ±15º and ±45º at a distance of 1.5 m from a KEMAR manikin. The target was presented from 0°. The maskers were either colocated with the target or were spatially separated by ±15º or ±45º. Each masker sentence was played at a nominal level of 61.5 dBA SPL. The target sentence level was chosen randomly on each trial from 10 different levels such that the resulting target-to-masker ratio (TMR) ranged from -30 to +15 dB with 24 repetitions of each TMR. Subjects were located in an adjacent booth and listened either through the KEMAR microphones to simulate "natural" binaural conditions (KEMAR) or through a prototype microphone array (MicArray). The array consisted of 8 cardioid microphones placed on a headband and provided a high degree of directional selectivity (for further details see [1]). For "MicArray" listening, the headband array was placed on the head of KEMAR and its output signal was played diotically through circumaural headphones to the subject. The array was always directed towards the target at 0º azimuth.

Results

Psychometric Functions

Figure 1 shows the proportion correct words as a function of TMR for individual subjects in each condition. Logistic functions were fit to each psychometric function (solid lines) using a least-squares approach.

For the KEMAR listening condition, subjects showed different amounts of deviation from the fitted function. In particular, subjects L152 and L155 exhibited a clear dip around 0 dB TMR in the colocated condition. For separated masker conditions, the data did not show such deviations from fitted functions. Similar general observations can be made for the MicArray conditions although the plateaus were apparent even for some separated masker conditions.

The plateau effect can be explained by considering that when the target and maskers were presented at the same level and from the same spatial location, the only difference at the listener's ears was the talker's voice. This makes the sources easily confused. With lower TMRs, the difference in level between the target and maskers provided an extra cue to segregate the target. This suggests that for moderate negative TMRs (here roughly between -15 to -5 dB), subjects could possibly "listen in the dips" of the masker envelopes.

In order to quantify the magnitude of the plateau effect, the mean square error (MSE) between the measured and the fitted curves from -20 to 0 dB TMR was computed (Fig. 2, left panel). Significantly larger errors for colocated masker conditions than for separated ones were apparent in most cases. In addition, for separated conditions, MSEs tend to be larger when listening to MicArray than when listening

through KEMAR. Unlike KEMAR, the microphone array does not process incoming sources differently except for their level and frequency response. This means that the previously proposed level cue argument could also explain the moderate plateaus for the separated masker conditions indicated for some subjects by the larger MSE.

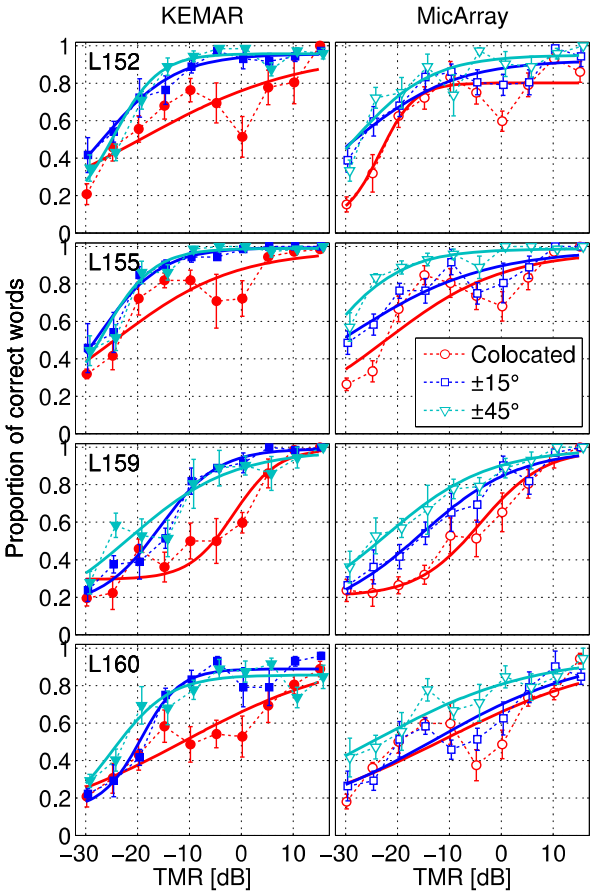

Figure 1: Proportion correct words as a function of TMR for the four subjects (rows) and two listening conditions (columns). The solid line shows the fitted logistic functions. Error bars represent the standard errors.

Spatial Benefit

Typically, spatial release from masking (computed as the threshold difference between colocated and separated conditions) is used to characterize the improvement in intelligibility that occurs when maskers are separated from the target. However, when the psychometric functions have a plateau, thresholds alone are inadequate to fully describe these effects. Therefore, analogous to the measure of performance advantage derived visually by Wightman et al. [3], the *spatial benefit* is defined here as (Se-Co)/(1-Co), where Se and Co are the averaged proportion correct for the separated and colocated conditions, respectively, for TMRs ranging between -10 to 0 dB.

The computed spatial benefits are shown in the right panel of figure 2. For natural listening conditions, spatial benefits were large and ranged between 0.6 and 0.8 for both masker separation angles. For MicArray listening conditions, spatial benefits for the ±15° angle were significantly lower than KEMAR for all subjects. However, for the ±45° separation, spatial benefits were on par with KEMAR conditions. This demonstrates that the spatial selectivity of this array yields performance that is similar to that observed with natural listening for the 45° separation. However, this microphone array is not spatially selective enough at 15° separation to equal natural listening performance under these conditions.

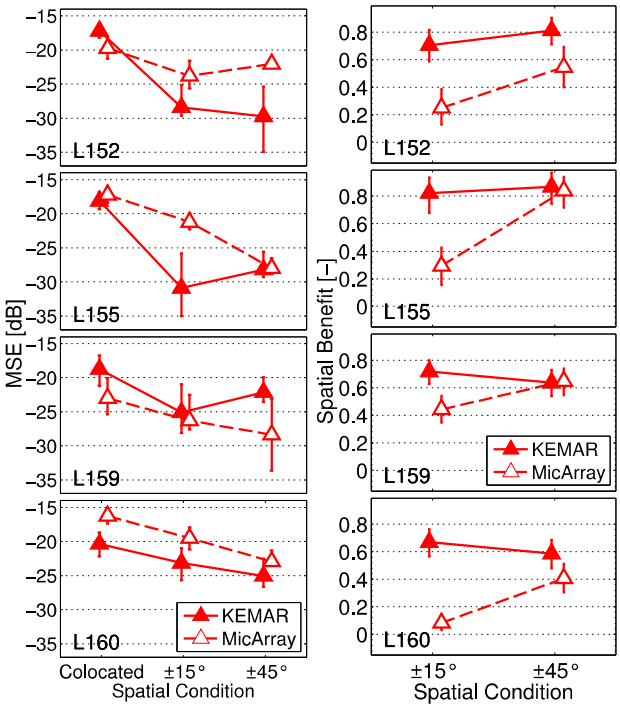

Figure 2 Mean square error (left) and Spatial Benefit (right). Error bars represent the 95% confidence interval.

Conclusion

MSE can be used to evaluate plateau effects in psychometric functions. These results revealed that the microphone array conditions show more deviations from the logistic function fits than natural listening conditions. Spatial benefit, as computed here, better described the improvement of intelligibility when spatially separating masker talkers for conditions producing plateaus.

The microphone array prototype provided spatial benefit that was similar to natural listening for maskers at ±45° azimuth.

Acknowledgments

Work supported by NIH/NIDCD and AFOSR.

References

[1] Favrot S., Mason C.R., Streeter T.M., Desloge J.G. and Kidd G.Jr. (2013). Performance of a highly directional microphone array in a multi-talker reverberant environment, ICA, Montreal, Canada, 19, 050145.

[2] Brungart, D.S., Simpson, B.D., Ericson, M.A., and Scott, K.R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers, JASA 110, 2527-2538.

[3] Wightman F., Kistler D., Brungart D. (2006). Informational masking of speech in children: auditory-visual integration, JASA 119, 3940-9.

[4] Kidd G.Jr., Best V. and Mason C.R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech, JASA 124, 3793-3802.