Attending to Space or Intensity Modulates Spatial Release from Informational Masking

Antje Ihlefeld[1], Sarah Chu[2], Barbara G. Shinn-Cunningham[3]

Hearing Research Center, Boston University, Boston MA 02215, USA,
Email: [1]antje1@gmail.com, [2]schu@mit.edu, [3]shinn@cns.bu.edu

Introduction

When trying to understand target speech in a background of perceptually similar speech maskers or when listening for a speech target with unknown acoustic features, informational masking (IM) can impair performance. IM causes target detection thresholds to be elevated relative to what traditional models of peripheral masking would predict (e.g., see [1]). Previous studies show that when target and masker are perceived at different locations, thresholds can improve, an effect known as spatial release from masking (SRM). Furthermore, even when target and masker are coming from the same location, release from IM can occur if the overall speaking level differs between the competing utterances. Previous studies show that the amount of SRM can depend on how similar the target and masker are in level. Here, we examined whether the amount of SRM depends on listeners attending to location versus intensity of the target.

Probabilistic Model

Previously we showed that a probabilistic model of response patterns can help tease apart the contributions of two distinct processes underlying IM ([2], Fig. 1). Our model posits that intelligibility in a selective speech identification task depends on low-level spectrotemporal continuity (short-term segmentation), on correctly joining short-term segments across spectrotemporal discontinuities (across-time linkage), and on the ability to properly select short-term segments and/or streams (selective attention). When IM is the primary form of interference, errors consist of the subject reporting either the masker message (henceforth, *masker* errors) or a combination of words from both the target and masker messages (henceforth, *mix* errors). In contrast, random guesses, where listener may report keywords that were not part of either the target or masker message (henceforth, *drop* errors), rarely occur. This is consistent with the idea that the spectrotemporal structure of speech stimuli used in these studies is rich enough for listeners to properly segment syllables from the acoustic mixture, and that errors arise from difficulties in selecting the correct syllables and/or streams of syllables.

For each cue condition, angular separation (AS), and level differences (LD) of the sources, the fixed probability of properly selecting a keyword is P_{SEL}, while the fixed probability of properly linking the words across time is P_{STR}. Conditioned on there being no drop errors, P_{SEL} and P_{STR} can be computed from the relative likelihoods of correct responses, masker errors, and mix errors (see Fig. 1).

Methods

To emphasize the effects of IM, the current study employs spectrally interleaved bandpass filtered target and masker speech that was derived from the Coordinate Response Measure corpus (see methods in the full-cue condition in [3]). Target and masker [<color> <number>] phrases were extracted from the original utterances by time windowing. <*Color*> was one of the set [white, red, blue, and green]. <*Number*> was one of the digits between one and eight, excluding the two-syllable digit seven. In each trial, two different [<color><number>] phrases were used as sources. The numbers and colors in the competing utterances were randomly chosen, but constrained to differ from each other in each trial. Subjects were instructed to report the target color and number based on intensity, location, or both intensity and location. Feedback was provided. Three consecutive blocks always had the same instructions; each session consisted of three blocks of each of the three cue-conditions. Each session consisted of nine blocks of 60 trials each. Seven normal-hearing subjects were paid for their participation in the experiment; each completed four sessions of the experiment. The data from the first session were discarded as practice.

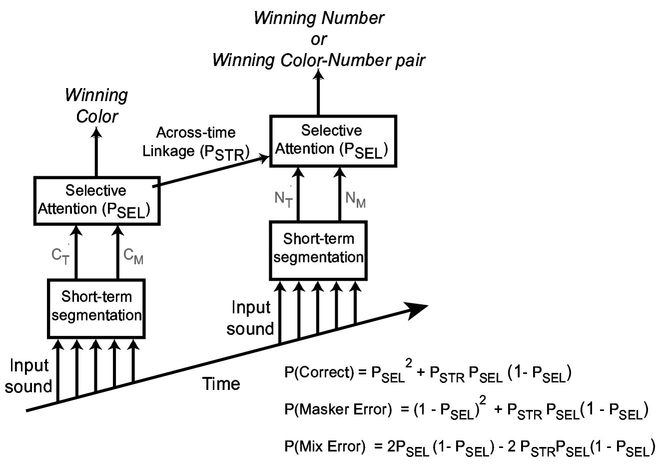

Figure 1: Decision-theory model to quantify the roles of selective attention (P_{SEL}) and across-time linkage (P_{STR}).

Target location was chosen randomly from trial to trial, and was equally likely to be any of 11 locations (±90˚, ±80˚, ±50˚, ±40˚, ±10˚, 0˚). On each trial, the angular separation between target and masker was randomly chosen (either 0˚, 10˚, or 90˚). Target and masker levels were between 50 dB and 80 dB SPL, roving randomly from trial to trial. On each trial, the level difference between target and masker was randomly chosen (either 0 dB, ±5 dB, or ±10 dB).

Results

When collapsed across all LD, percent correct performance improves as a function of AS for all cue conditions (Fig. 2A). However, the improvement is much smaller in the level condition than in the location and level-and-location conditions. The patterns of mix and drop errors are similar across all cue conditions, decreasing with increasing AS (Fig. 2B, D). In contrast, the pattern of masker errors is at best weakly affected by AS in the level condition, but decreases rapidly in the other two cue conditions.

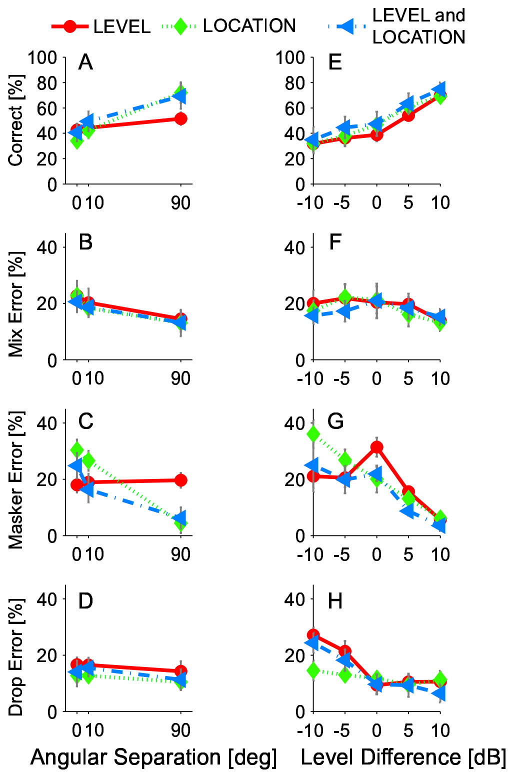

Figure 2: Across-subject mean performance as a function of AS (left column) and LD (right column). A, E. Percent correct. B, F. Mix Errors. C, G. Masker Errors. D, H. Drop Errors. Error bars show two standard errors of the mean.

Averaging the data across AS shows the effect of LD. As a function of the LD between the two utterances, percent correct performance is roughly similar across cue conditions (Fig. 2E). However, for negative LDs, performance decreases less steeply in the level cue condition than in the other two conditions. The pattern of masker errors differs across cue conditions. In the level condition, masker errors peak at 0 dB. In contrast, in the location condition, masker errors decrease monotonically as a function of LD. In the level-and-location condition, the rate of masker errors plateaus for negative LDs. Finally, more drop errors occur for negative level differences than for positive level differences, consistent with predictions made by traditional models of peripheral masking.

Model

To quantify the contributions of selective attention and across-time linkage, results were fit by the model parameters P_{SEL} and P_{STR} (Fig. 3). As a function of AS, the trajectories for the level condition are almost vertical (circles in Fig. 3 A), showing that increasing AS only affects P_{SEL} not P_{STR} when listeners are selecting the target based on its level. This suggests that indeed, when listeners are attending to level, AS does not affect the ability to select a target word and/or stream, but the ability to link words across time improves with increasing spatial separation when listeners are not attending to space. Finally, when listeners were attending to location or to level and location, the ability to select the target words and/or stream improves with increasing spatial separation.

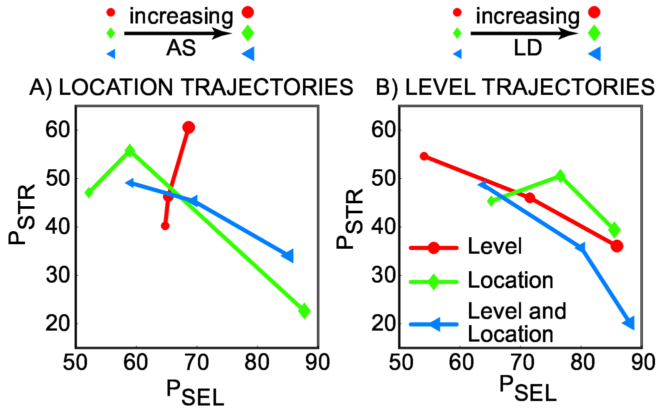

Figure 3: Parameter fits. A. as a function of AS. B. as a function of LD. Only [0, 5, 10 dB] LD are shown. Symbol size increases with increasing AS and LD (A and B, respectively).

When viewed as a function of LD, the parameter trajectories are fairly similar across cue conditions (Fig. 3 B): both P_{SEL} and P_{STR} increase with increasing LD. This is consistent with the idea that intensity cues can improve both across-time linkage and selective attention, whether or not listeners are attending to intensity.

Conclusions

Overall, subjects often mistook masker for target when attending to space and AS was small, or when attending to level and LD was small. Increasing AS can improve the ability to link keywords and select the target. The relative contributions of these mechanisms depend on what target attribute the listener attends. In contrast, LD can improve target selection and across-time linkage, whether or not listeners are attending to target intensity.

References

[1] Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G.: Note on informational masking. J Acoust Soc Am 113 (2003) 2984-2987.

[2] Ihlefeld, A.: Strategies of spatial listening for speech comprehension. Ph.D thesis, Boston University, 2007

[3] Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E.: Bottom-up and top-down influences on spatial unmasking. Acta Acustica 91 (2005). 967-979