SPATIAL UNMASKING FOR NEARBY SPEECH SOURCES IN A SIMULATED ANECHOIC ENVIRONMENT

3aPP2

Jason Schickler[1,2], Norbert Kopčo[2,3], Barbara Shinn-Cunningham[1,2,3] and Ruth Litovsky[1,2]

[1]Hearing Research Center, Departments of [2]Biomedical Engineering and [3]Cognitive and Neural Systems, Boston University

1. BACKGROUND AND MOTIVATION

Angular separation of target and masker
- improves detectability of target
- improves speech reception of target

Studies of spatial advantage in speech reception
- fix target ahead
- angularly displace masker

Two peripheral factors identified:
- head shadow ("better-ear") advantage
- binaural advantages

Targets and maskers may be within arm's length
- large energy ef fects
- large (x-frequency) interaural dif ferences
  (head shadow + relative distance to ears)

Current study, measure speech thresholds for
- near (15 cm) & far (1 m) target and masker
- traditional configurations with target ahead
- masker ahead and target moving
- masker and target to side

Goals: to analyze acoustics, measure spatial effects, and lay groundwork for future studies in complex listening environments.

What are spatial advantages for near sources?
- · How large are energy effects?
- · How large is better-ear advantage?
- · How well can current models predict results?

Supported by AFOSR grant F49620-98-1-0108

2. METHODS

Procedures adapted from Hawley, Litovsky, and Culling (1999)

Simulate nearby sources using HRTFs (see below)
- Target (T): IEEE sentences
- Masker (M): speech-shaped noise

For each spatial configuration
- identify better ear (not always nearest target)
- equate energy of masker at better ear
- adaptively vary target level to find threshold
  (threshold defined as 60% words correct)

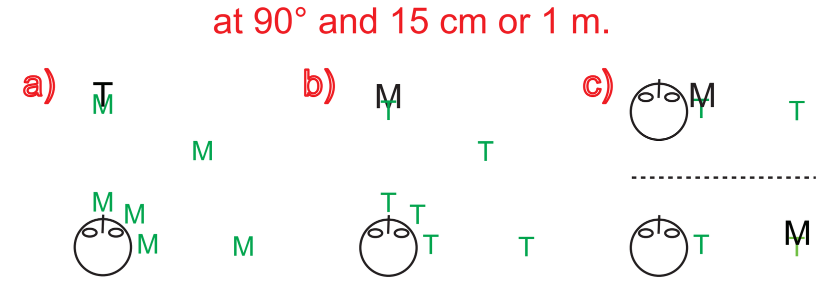
Adaptive runs in random order

Each condition tested at least 3x
(repeated until threshold stderr < 1 dB)

Subjects
- four normal-hearing undergraduates

15 spatial configurations were tested
- 6 with target fixed in front
- 5 additional with masker fixed in front
- 4 with target and masker to side

Figure 1: Spatial configurations of target (T) and masker (M). Conditions included: a) T fixed (0°, 1 m) and M at each of 6 locations, b) M fixed (0°, 1 m) and T at each of 6 locations, and c) T and M at 90° and 15 cm or 1 m.



3. SPATIAL SIMULATION: SPHERICAL-HEAD HRTFS

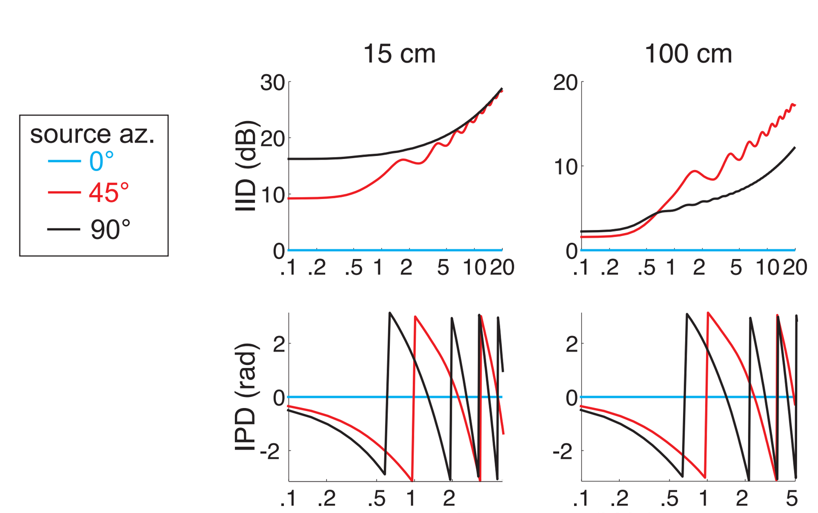
Spherical-head HRTFs were used to simulate stimuli for all subjects (e.g., see Shinn-Cunningham, Santarelli, and Kopčo, 2000)

For relatively low frequencies (as in speech signals), these HRTFs are similar to both KEMAR HRTFs (see Brungart and Rabinowitz, 1999) and individually measured HRTFs (see companion poster by Kopčo and Shinn-Cunningham, 2000).

HRTFs for six spatial locations were generated:
- 0 °, 45°, and 90° azimuth
- 15 cm and 1 m from the head

HRTFs for nearby sources differ from "traditional" HRTFs in that they contain:
- large changes in total energy with distance
- large interaural intensity dif ferences (IIDs) at all frequencies for nearby, lateral sources
- distance-dependent interaural dif ferences

Figure 2: Interaural differences as a function of frequency for the spherical-head HRTFs. Top panels show interaural intensity differences or IIDs (0.1 - 20 kHz), bottom panels show interaural phase differences or IPDs (0.1 - 5 kHz). Near distance (15 cm) on left, far distance (1 m) on right.



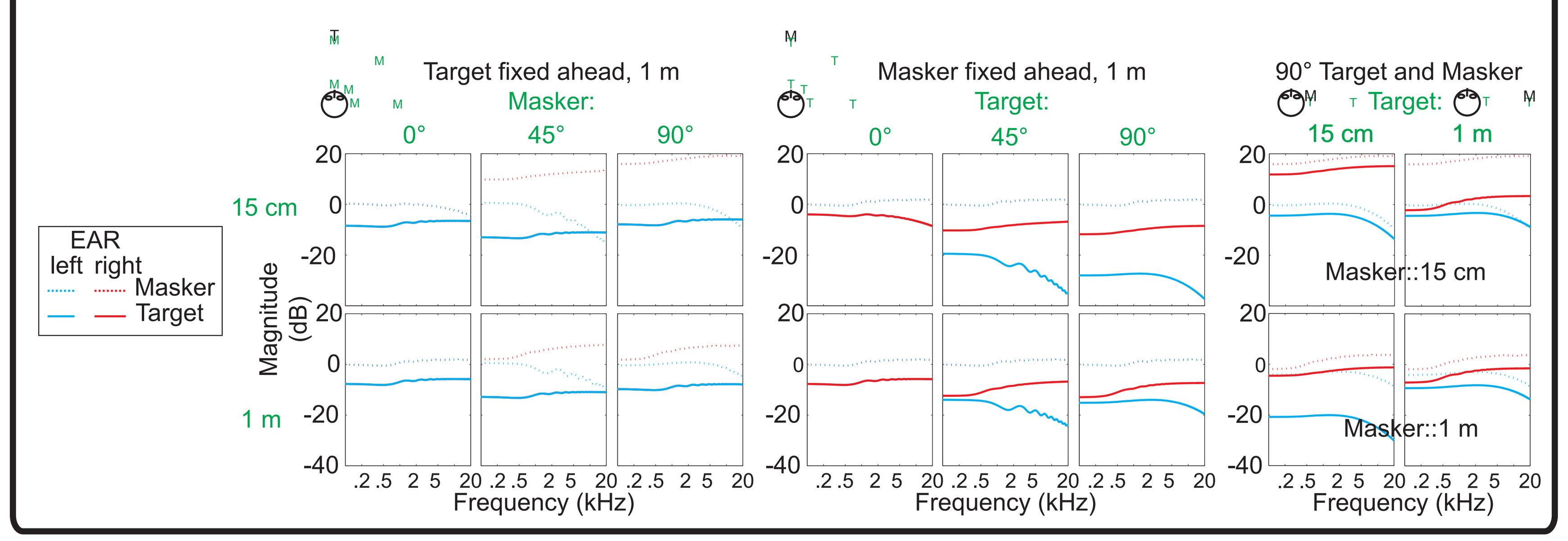
4. TARGET AND MASKER SPECTRA AT THRESHOLD

Graphs show target and masker spectral levels at left and right ears *at the measured threshold* (based on HRTFs only, not shaped by speech spectrum)
- · masker levels (dashed lines) *at the better ear* were roughly equated in the experiment
- · target levels (solid lines) give the measured levels at which 60% words were correct

Masker levels at the better ear are roughly equal (by design). Target levels at the better ear are similar across conditions, but are somewhat higher for the lateral conditions. Levels at the worse ear vary dramatically with condition.

Although better-ear threshold target levels are comparable across conditions,
- this analysis removes large energy effects that arise when a fixed-level source moves
- some variations are still evident

Figure 3: Target (T, solid lines) and Masker (M) spectral levels at threshold for left (blue) and right (red) ears (averaged across subjects). Cartoons in each set of panels show which spatial configurations are plotted in the set of panels.

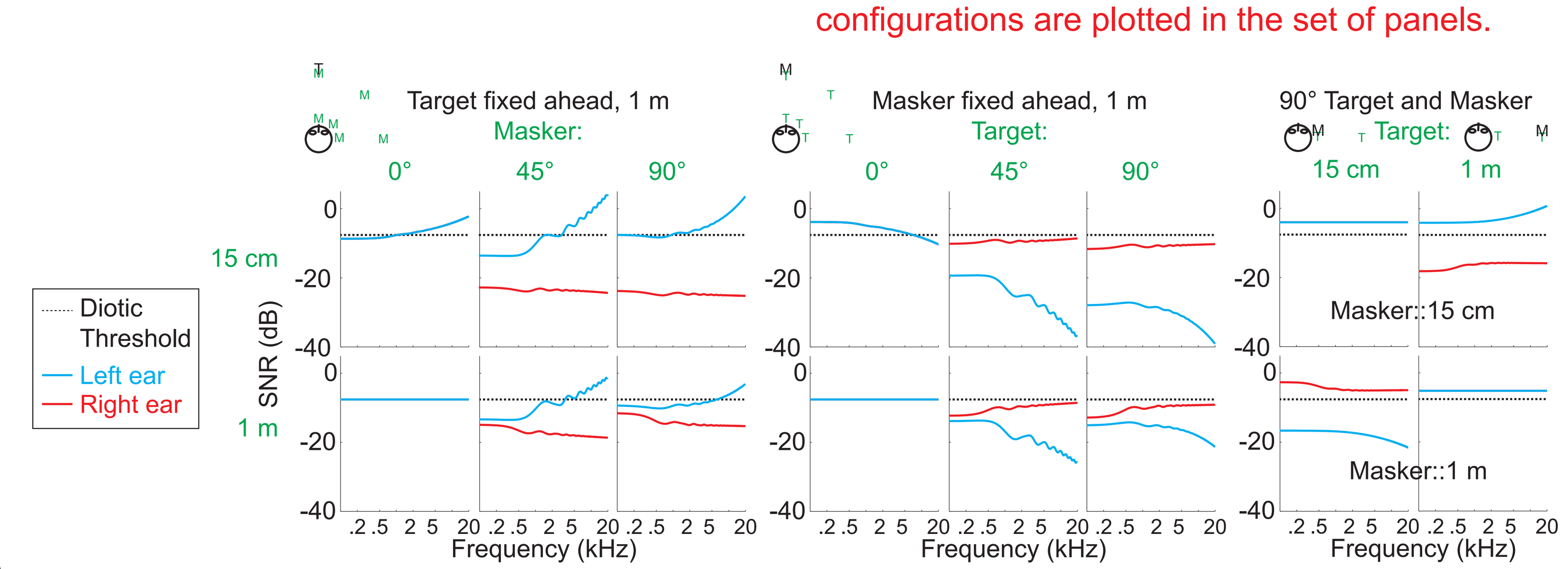


5. FREQUENCY-DEPENDENT SNR AT THRESHOLD

When T and M are co-located at (0°, 1 m), interaural differences are zero
- Signal-to-noise (SNR) is constant with freq
- provides a baseline measure at threshold (see dashed line in plots)
- lower SNR implies spatial advantage
- higher SNR implies spatial disadvantage

For T fixed ahead, SNR at better ear is at or below diotic reference (indicating a spatial advantage beyond energy effects). In some conditions (e.g., when T and M are at 90° az), SNR at better ear is larger than for diotic reference.

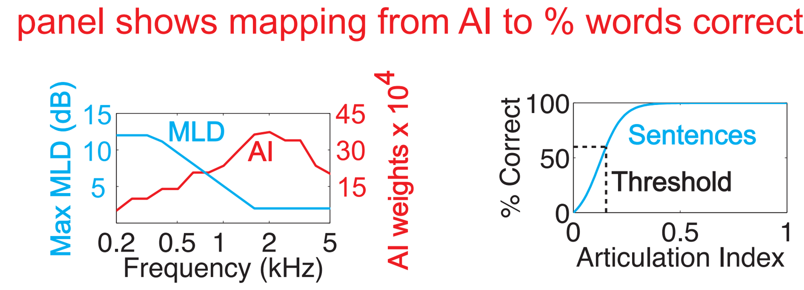
Figure 4: Signal-to-Noise Ratio (SNR) of Target to Masker for both ears as a function of frequency at threshold. Cartoons show which spatial configurations are plotted in the set of panels.



6. MODEL ANALYSIS: ZUREK (1993)

Zurek's model of speech intelligibility for binaural listening was used to predict results
- calculates audibility of signal in each 1/3-octave of frequency from better ear
- adds binaural contribution in each frequency
- weights each frequency by its importance
- sums to form Articulation Index (AI)
- maps AI to predicted % correct (sentences)

Figure 5: Binaural AI model assumptions (Zurek, 1993). Left panel shows maximal binaural contribution to SNR at each frequency (blue) and relative weight of each frequency (red). Right panel shows mapping from AI to % words correct.

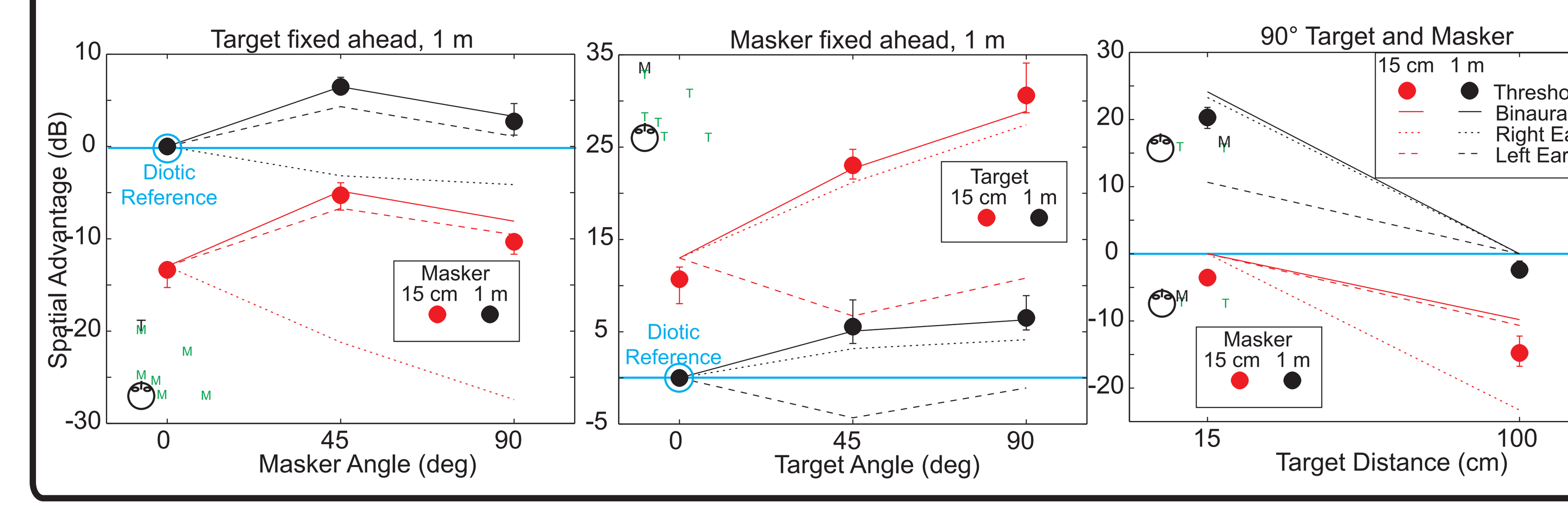


7. SPATIAL ADVANTAGE FOR SPEECH RECEPTION

Both overall energy effects and the better ear advantage are very large for nearby sources. Binaural contributions are small, but important.

The Zurek (1993) model captures gross energy effects, but there are some differences in predictions and results. These discrepancies tend to occur in conditions involving near and/or lateral sources. While one prediction (T at 1 m, 0°, M at 15 cm, 90°) could be improved by reducing the binaural model contribution, some errors are inherent in the better-ear model predictions.

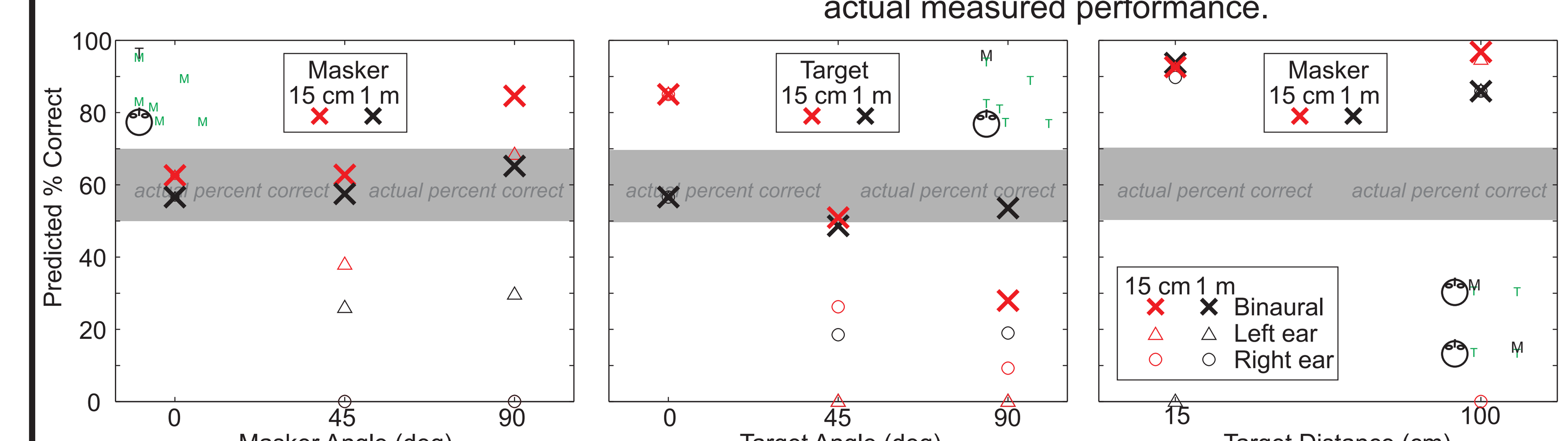
Figure 6: Spatial advantage (energy a target emits at threshold for a constant-energy masker) and model predictions, relative to diotic reference. Positive values are decreases in emitted target energy. Solid symbols are x-subj averages (bars show range of subject results). Lines are model predictions: solid is for full model; dashed lines are right and left ear without binaural processing. The difference between the solid and higher dashed line gives the predicted binaural contribution. The difference between the dashed lines is the predicted better-ear advantage.



8. PREDICTED INTELLIGIBILITY AT THRESHOLD

Figure 7: Predicted % correct word scores from model using SNRs and binaural cues present at threshold (measured 60% correct in gray range).

What appear to be small discrepancies in predicted thresholds result in large errors in % correct. For six conditions, predicted performance is substantially better (i.e., for all conditions with lateral T and M and two conditions involving nearby T or M), but for another condition, predicted performance is substantially worse than actual measured performance.



9. CONCLUSIONS

Spatial effects for nearby sources are large
- overall effects range from -15 dB to +30 dB relative to diotic, 1 m condition
- predicted "better ear" advantage up to 25 dB
- predicted binaural advantage small (< 2 dB)

Compared to Zurek model, actual performance differed from predictions for a subset of the conditions that were unique to our study (i.e., for nearby sources and/or lateral targets)

Further experiments should be performed to
- confirm effects of lateral targets and large IIDs on speech reception thresholds
- explicitly measure monaural thresholds to verify binaural and better-ear effects

10. REFERENCES

Brungart, D. S. and W. M. Rabinowitz (1999). "Auditory localization of nearby sources I: Head-related transfer functions." Journal of the Acoustical Society of America 106(3): 1465-1479.

Hawley, M. L., R. Y. Litovsky, and J. Culling. (2000). The "cocktail party problem" with four types of maskers: Speech, time-reversed speech, speech-shaped noise, or modulated speech-shaped noise. ARO MidWinter Meeting, St. Pete. Beach, FL.

Kopčo, N. and B. G. Shinn-Cunningham (2000). "Spatial unmasking of nearby narrowband sources in a simulated anechoic environment." Journal of the Acoustical Society of America 107(5): 2849.

Shinn-Cunningham, B. G., S. Santarelli, and N. Kopčo. (2000). "Tori of confusion: Binaural localization cues for sources within reach of a listener." Journal of the Acoustical Society of America 107(3): 1627-1636.

Zurek, P. M. (1993). Binaural advantages and directional effects in speech intelligibility. Acoustical Factors Affecting Hearing Aid Performance. G. Studebaker and I. Hochberg. Boston, MA, College-Hill Press.