Spatial Release from Speech-on-Speech Masking in Reverberation Evaluated Using the Portable Automated Rapid Test

Udbhav Singhal[1,3], Marusa Laure[2,3], Alyson Ditmore[4], William J. Bologna[4], Aaron R. Seitz[3], Frederick J. Gallun[6], G. Christopher Stecker[5], Norbert Kopčo[1]

[1]Institute of Computer Science, P.J. Šafárik University in Košice, Slovakia. [2]University of Maribor, Slovenia

[3]Department of Psychology, Center for Cognitive and Brain Health, Northeastern University, Boston, MA

[4]Department of Speech-Language Pathology and Audiology, Towson University, Towson, MD. [5]Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE

[6]Department of Otolaryngology, Oregon Health & Science University, Portland, OR

Introduction

- **Reverberation**, caused by reflections from room surfaces, distorts ITD (Interaural time differences) and ILD (Interaural level differences) cues, reducing speech intelligibility and localization accuracy (Shinn-Cunningham et al., 2005; Kuttruff, 2016; Ihlefeld & Shinn-Cunningham, 2011).
- **Spatial Release from Masking (SRM)** is used to evaluate the benefit of spatial separation for speech intelligibility. **SRM** is typically **reduced** by **~5 dB** in **reverberant** compared to **anechoic** environment (Bronkhorst & Plomp, 1992; Lavandier & Culling, 2007; Marrone et al., 2008, Momtaz et al., 2025).

Here, we use the Portable Automated Rapid Testing (PART) system to rapidly measure **SRM** for speech stimuli in anechoic and reverberant environments.

Objective: Examine SRM caused mostly by **informational masking** by employing a **closed-set corpus** with highly confusable words (Coordinate Response Measure, CRM) spoken by same-sex talkers with two **symmetrically placed maskers** (thereby minimizing energetic masking effects).

ADDITIONAL GOALS:

I. Evaluate **psychometric function slopes:** Determine whether reverberation alters the steepness of the psychometric functions differently for **co-located** versus **spatially separated** conditions.
II. Examine whether **spatial tuning** to target location improves over time.
III. Analyze **error patterns:** Identify how often/when listeners confuse target and masker speech (informational masking) vs. random errors.

Methods

Participants:

10 participants (age 22-32 years; 6 female) with self-reported normal hearing.

Stimuli and Setup:

- Coordinate Measure Response (CRM; Bolia et al., 2000) sentences presented virtually over headphones in **anechoic** or **reverberant** environment (reverberant room size: 10*10m, T_{60} = 0.4 s).
- CRM sentences in the form of **"Ready (CALLSIGN), go to (COLOR) (NUMBER) now"**, with 8 possible **call signs**: (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, Tiger), 4 **colors** (red, green, white, and blue), and 8 **numbers** (1 to 8); spoken by three male voices.
- ***Target***: Callsign **Charlie** presented at **0˚** azimuth (Fig 1) by a voice randomly selected on each trial (talker used on the previous trial excluded).
- **Maskers:** Two maskers A and B with two different callsigns (excluding Charlie) simultaneously presented at 0˚ in the **co-located** condition or at ±45˚ in the **separated** condition. Each masker used a different non-target voice.

Procedure:

- Apple iPad & Sennheiser HD 280 Pro headphones were used to run the study.
- Response on each trial was entered by pressing the correct color/number button on tablet screen. Response was correct only if both color & number were identified correctly. Feedback ("correct" or "incorrect") was provided.
- Adaptive scans (Lelo de Larrea-Mancera et al., 2024) were used to measure performance, changing target or masker level. Target-to-Masker Ratio (**TMR**s) ranged from **20 to -20 dB.** Starting masker level **50 dB SPL.** At least one component (target or masker) **remaining at 70 dB SPL** throughout each scan. There were five scans per condition: **First two scans** were **non-adaptive**, the **remaining** scans were **adaptive. Each scan** consisted of **9 trials**.
- The experiment was divided into **two sessions (Anech, Reverb)**, counterbalanced across subjects. **Each session** took about **20 mins.** For anechoic condition, **531 trials** were measured across participants. For reverberant condition **696 trials** were measured (26% trials in anechoic condition were excluded to coding error).

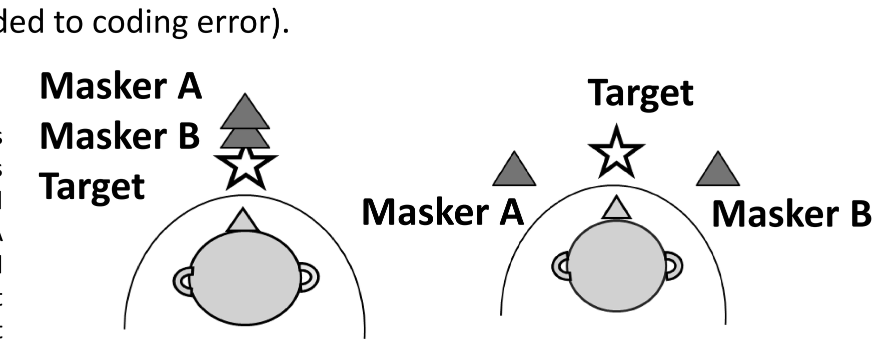

Figure 1: Spatial configurations used in experiment. Listener is facing 0°. In co-located configuration, target and maskers A and B are in front. In separated configuration, maskers A & B are at the ±45° azimuth, while target remains at the 0°.

Results

PSYCHOMETRIC FUNCTIONS

Psychometric functions were obtained by fitting a logistic model (sigmoid) to the correct color-and-number response data. The Speech Reception Thresholds (SRTs; Target-to-Masker ratio for 50% correct performance) and slopes were estimated from the functions.

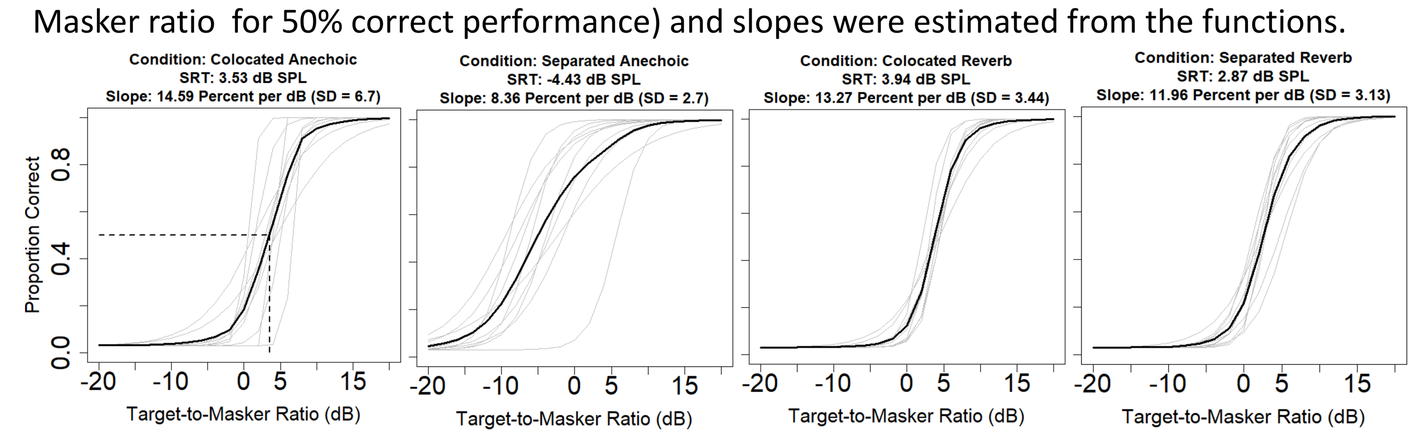

Figure 2: Psychometric function are shown for all combinations of spatial configurations and environments. Thick black line represents averaged psychometric functions derived from individual fits (thin lines). Dashed line shows the point from which SRT and slope was determined.

Results for each combination of configuration and environment are mostly consistent across participants (one outlier in anechoic separated condition).

SRT AND SRM

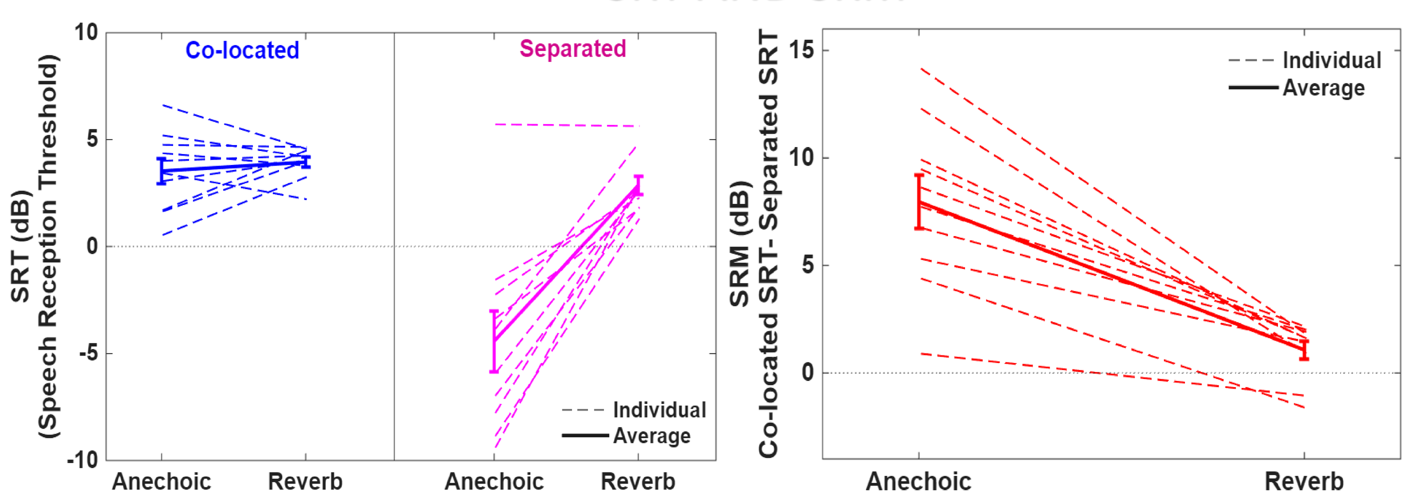

Figure 3:(a) Anechoic and Reverb Speech Reception Threshold (SRT) estimated for individual participants and averaged within location. (b) Spatial Release from Masking (SRM) estimated for individual participants and average within environment (anechoic and Reverb), Error bars show the standard error of the mean.

No effect of environment in collocated configuration (both thresholds around **3 dB**). SRM was **8 dB** in the anech. Env., reduced to only **1.1 dB** in reverberation (ANOVA with factors **configuration & environment** found signif. interaction, ***p* < .001**).

SRM was significantly **greater** in **anechoic vs. reverberant environment** (difference of 6.9 dB), indicating that **spatial separation provides more perceptual benefit in the absence of reverberation**.

SLOPES OF PSYCHOMETRIC FUNCTIONS

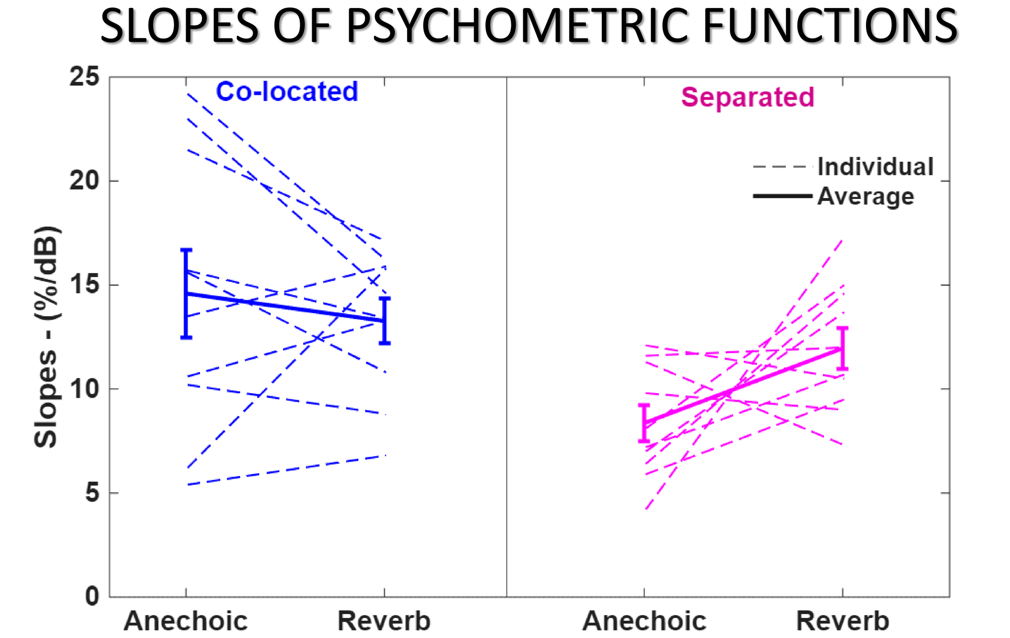

Figure 3: The slopes of the psychometric curve estimated for individual participants and averaged within location (co-located and separated). Error bars show the standard error of the mean.

The slopes of the psychometric curves were comparable in the co-located conditions **(anechoic: 14.6%/dB, reverberant 13.3%/dB),** becoming shallower in the separated conditions **(anechoic : 8.4%/dB vs 12%/dB in reverberation).**

A 2-way ANOVA showed a significant main effect of **configuration** (***p*** = **.021**) and **trend towards significant interaction** with environment (p = .056).

Shallower slopes indicate that more non-energetic factors are more dominant in determining the separated threshold, mostly in anechoic environment.

PERFORMANCE ANALYZED SEPARATELY FOR COLOR AND NUMBER WORDS AND DIFFERENT TMRS

To examine how performance varied over time, we analyzed performance separately for the two key words (Fig. 5a). To examine whether these temporal effects change at different TMRS, the data were further subdivided into 3 TMR bins (Fig. 5b).

A 3-Way ANOVA on **Fig 5 (a):**
The configuration (co-located, separated) factor interacted significantly with item type (color vs. number) (***p* = .002**) (also with environment).

Number performed **better** than **color** for **separated conf. (both Anechoic and Reverb)**. **Number** performed **worse or equal** to **color** for **co-located conf.**

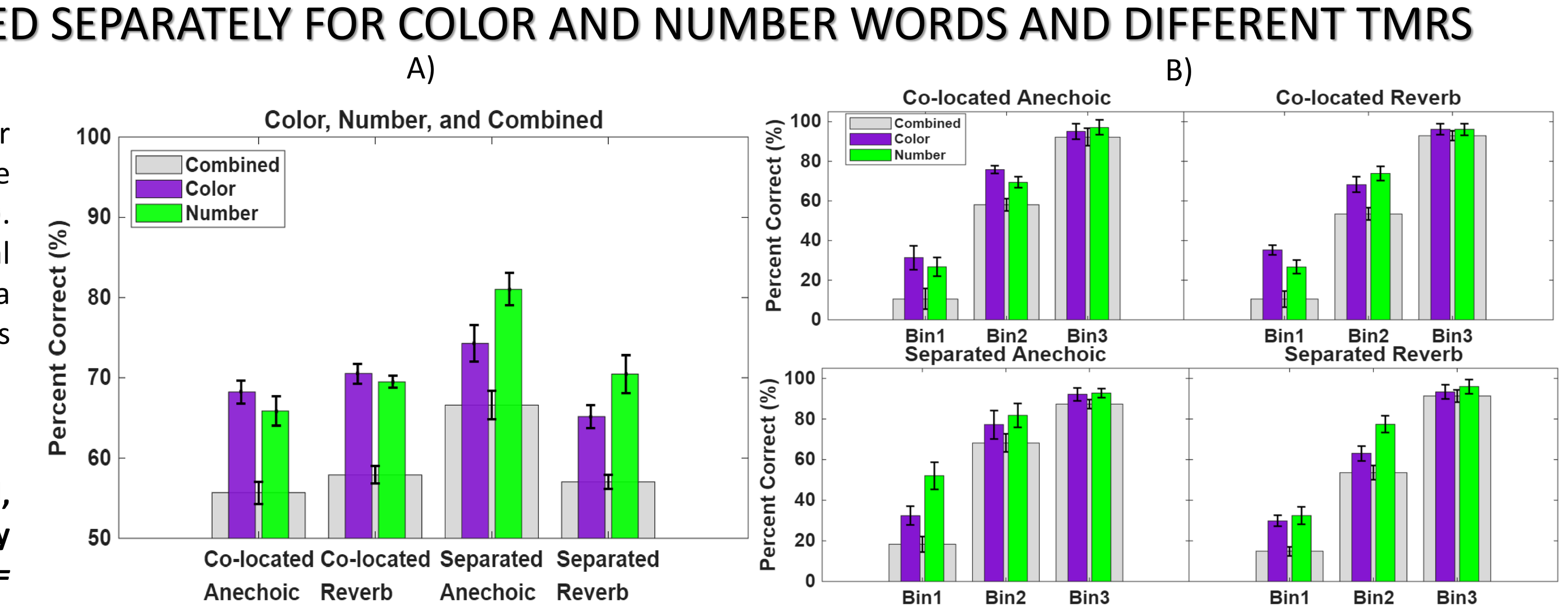

Figure 5: (a) Percent correct computed separately for **color** and **number** keywords, as well as for the combined color+number key phrase. (b) Same data subdivided by TMR (for each participant, **Bin 1** included all TMRs below SRT −1.5 dB, **Bin 2** covered SRT ±1.5 dB, and **Bin 3** included all TMRs above SRT +1.5 dB).

When Target is separated from Maskers (but not when they are co-located), item-to-item tuning to the target location/away from masker locations occurs. This effect is mostly visible at lower TMRs.

ERROR PATTERN: INTRUSIONS VS RANDOM RESPONSES

Intrusion occurs when response matches a masker. Intrusions were analyzed for both combined color+number key phrases and separately for color / number key words.

Fig 6a: For combined key phrase (gray), **Location** x **Environment** interaction comes out significant (***p* = .046**).
A 3-way ANOVA including item (color vs. number) only found a significant main effect of **location** (***p* = .012**).

Fewer combined intrusions observed in **separated anechoic** condition. When data analyzed by color and number, the **benefit of location significant for both environments (but small).**

Fig 6b: Fewer combined intrusions at higher TMRs.

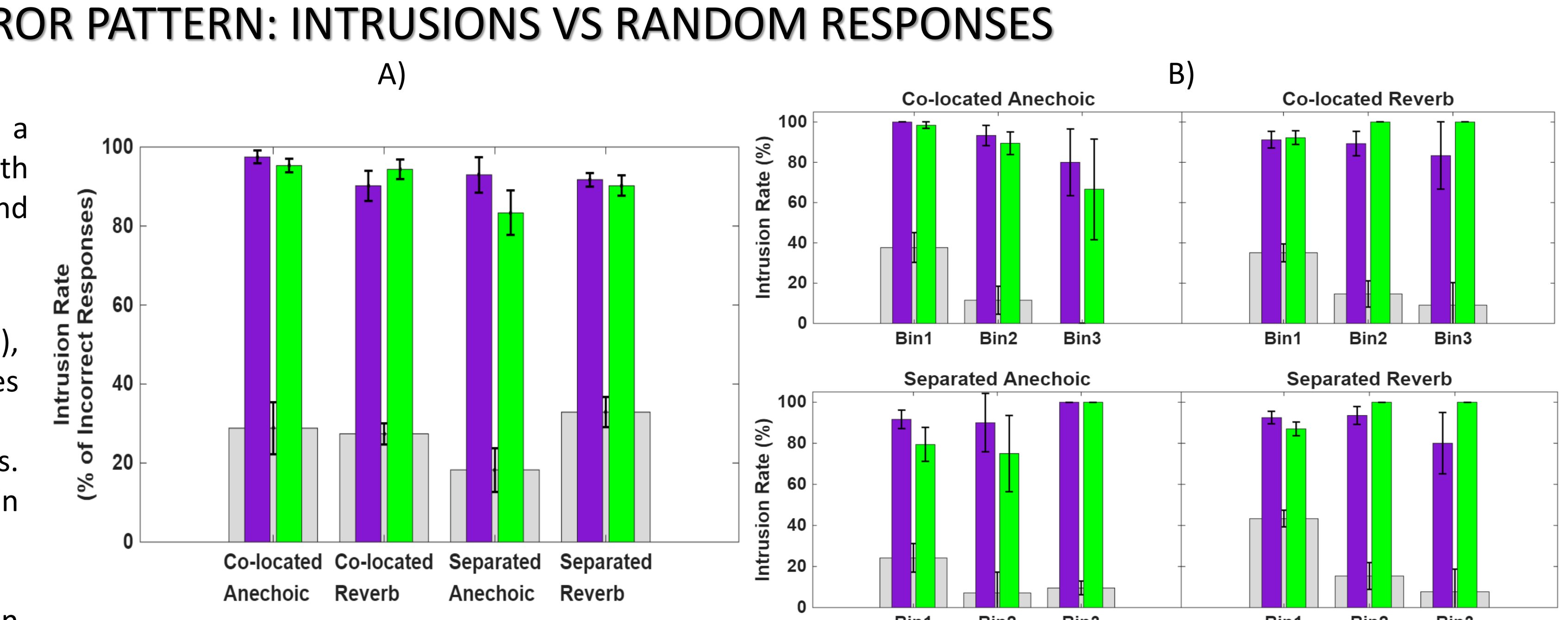

Figure 6: (a) Intrusions computed separately for color (purple) and number (green) key words; and combined (grey) key phrase. (b) Intrusions broken by TMR (like in Fig 5b)

In combined analysis, intrusion rate was lower in separated anechoic condition, suggesting that the spatial benefit is the strongest. This might be related to the difference between color and number intrusion rates in that condition. Also, the combined intrusions are most common at lowest TMRs.

Discussion

- SRM was **8 dB** in the anechoic environment, reduced to only **1.1 dB** in reverberation, consistent with previous studies.
- This was driven mainly by the **cost of reverberation** being **large for separated (7.3 dB)** and **small (0.4 dB) for co-located** configuration.
- **Shallower slopes** were observed in the **separated condition** (mainly in **anechoic**), indicative of informational masking dominating performance. However, intrusion rate was lowest in **separated anechoic** condition, so the confusions were not due to limits of spatial processing.
- **Performance** improved for the **second vs first word** in **separated conditions** in both rooms. This indicates that the spatial tuning in the separated condition occurs on the time scale of words (note that result is particularly robust because chance performance would predict an opposite trend). This tuning was mostly visible at low TMRs.
- **Intrusion rates** were **lowest** in **separated anechoic condition,** did not change over time, and decreased as TMR increased.
- These results suggest that, in scenarios with **high informational masking**, the benefit of **spatial separation** can be dramatically **reduced** in **reverberant vs anechoic environments**, possibly due to increased difficulty segregating the target from maskers in reverberation.
- The study showed that **PART** can be used to obtain measures of **SRM** consistent with published approaches.
- Future: use PART to study **SRM** across **large and varied populations** (e.g., online, using Prolific).

References and Acknowledgements

[Work supported by HORIZON-MSCA-2022-SE-01 Grant No. 101129903, APVV-23–0054, NIH R01-DC018166]

- Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. 107(2), 1065–1066.
- Bronkhorst, A., and Plomp, R. (1992). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing,"J. Acoust. Soc. Am. 92(6), 3132–3139.
- Esteban Sebastian Lelo de Larrea-Mancera, Tess Koerner, Eric C. Hoover, G. Christopher Stecker, Frederick J. Gallun, Aaron R. Seitz; Estimating sensory thresholds with the adaptive scan method of psychophysical testing. Proc. Mtgs. Acoust. 13 May 2024; 54 (1): 050002. https://doi.org/10.1121/2.0001939
- Ihlefeld, A., and Shinn-Cunningham, B. G. (2011). "Effect of source spectrum on sound localization in an everyday reverberant room," J. Acoust.Soc. Am. 130(1), 324–333.
- Kuttruff, H. (2016). Room Acoustics (CRC Press, Boca Raton, FL), p. 322.
- Marrone, N., Mason, C. R., and Kidd, G. (2008). "Tuning in the spatialdimension: Evidence from a masked speech identification task,"J. Acoust. Soc. Am. 124(2), 1146–1158
- Lavandier, M., and Culling, J. F. (2007). "Speech segregation in rooms:Effects of reverberation on both target and interferer," J. Acoust. Soc.Am. 122(3), 1713–1723.
- Momtaz, S., Williams, B. T., Miller, M. K., Bologna, W. J., Koerner, T. K., Lelo De Larrea-Mancera, E. S., Seitz, A. R., Gallun, F. J., & Stecker, G. C. (2025). Effects of reverberation and binaural sensitivity on spatial release from masking. The Journal of the Acoustical Society of America, 158(4). https://doi.org/10.1121/10.0039576
- Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (2005). "Localizing nearby sound sources in a classroom: Binaural room impulse responses," J. Acoust. Soc. Am. 117(5), 3100–3115.