The role of early and late reflections on spatial release from masking: Effects of age and hearing loss

Nirmal Kumar Srinivasan, Meghan Stansell, and Frederick J. Gallun

Citation: The Journal of the Acoustical Society of America **141**, EL185 (2017); doi: 10.1121/1.4973837
View online: https://doi.org/10.1121/1.4973837
View Table of Contents: http://asa.scitation.org/toc/jas/141/3
Published by the Acoustical Society of America

Srinivasan *et al.*: JASA Express Letters [http://dx.doi.org/10.1121/1.4973837] Published Online 1 March 2017

The role of early and late reflections on spatial release from masking: Effects of age and hearing loss[a)]

Nirmal Kumar Srinivasan,[b)] Meghan Stansell, and Frederick J. Gallun[c)]
National Center for Rehabilitative Auditory Research, Veterans Affairs Portland Health Care System, Department of Veterans Affairs, 3710 Southwest United States Veterans Hospital Road, P5-NCRAR, Portland, Oregon 97239, USA
NSrinivasan@towson.edu, Meghan.Stansell@va.gov, Frederick.Gallun@va.gov

Abstract: Early reflections have been linked to improved speech intelligibility, while later-arriving reverberant sound has been shown to limit speech understanding. Here, these effects were examined by artificially removing either early reflections or late reflections. Removing late reflections improved performance more for colocated than for spatially separated maskers. Results of a multiple regression analysis suggest that pure-tone average (PTA) is a significant predictor of spatial release from masking (SRM) in all acoustic conditions. Controlling for the effects of PTA, age is a significant predictor of SRM only when early reflections are absent.

[DDO]
Date Received: August 22, 2016 **Date Accepted:** December 27, 2016

1. Introduction

In real world listening situations, we listen to speech in the presence of sources of masking, and one major source of masking is distracting speech, while another is reverberation. The process of understanding target speech in the presence of masking speech involves separating the acoustic information of the target speech and ignoring the acoustic information of the masking speech. Target speech and masking speech can be segregated based on factors such as disparity in pitch, voice color, accent, phoneme rate, context (predictable versus unpredictable), the location of the target and maskers in the environment, or interaural time and level differences (Alain, 2007).

Reverberation is the persistence of sound in a listening environment after the direct sound has stopped producing and is caused by multiple reflections from surfaces in the environment. Reverberation is present in almost all everyday communication scenarios. Reduction in speech intelligibility due to reverberation is typically attributed to two effects: (1) alteration of the spectrotemporal characteristics within individual phonemes and (2) reduction of the low frequency envelope modulations essential for speech intelligibility (Nabelek *et al.*, 1989). Plomp (1976) investigated spatial release from masking (SRM) in different reverberant environments and concluded that SRM was inversely related to reverberation time, suggesting that distortions of the spatial cues can provide a third source of reduced speech intelligibility due to reverberation. While young normal hearing individuals can function in moderate amounts of reverberation with a minimal reduction in speech understanding, older individuals with and without hearing loss are often found to be more susceptible to the distortions of the speech signal (Helfer, 1992). Marrone *et al.* (2008) evaluated the interaction between aging, hearing loss, and reverberation on SRM and showed that listeners with hearing loss achieved even less SRM than normal-hearing listeners, both with and without reverberation. The effects of age on SRM were not statistically distinguishable from the effects of hearing loss in the Marrone *et al.* study, which only examined large spatial separation (±90°). Srinivasan *et al.* (2016) examined SRM for various spatial separations between the target and the maskers in an anechoic listening environment and

[a)]Portions of this work were presented at the 171st Acoustical Society of America meeting, May 2016, Salt Lake City, UT.
[b)]Current address: Department of Audiology, Speech-Language Pathology and Deaf Studies, Towson University, 8000 York Road, Towson, MD 21252, USA.
[c)]Also at: Department of Otolaryngology/Head and Neck Surgery, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, OR 97239, USA.

concluded that aging was the contributing factor predicting SRM at smaller separations whereas hearing loss was the contributing factor at larger separations.

A binaural room impulse response (BRIR) represents the acoustical transformation for a source of sound in a room as measured at each of the listener's ears, and therefore is influenced by factors such as the acoustic characteristics of the room, the position and orientation of the source, the position and orientation of the listener, and the locations of other sound reflective objects in the room. The sound reaching a listener's ear in a reverberant room consists of three main components: the direct sound (DS), the early reflections (ERs), and late reverberation (LR). The DS is the sound that reaches the listener before interacting with any surfaces and thus it is not influenced by the acoustic characteristics of the listening environment. ERs are defined as reflections that arrive immediately (<50 ms) after the direct speech. Perceptually, ERs are grouped with the direct speech to form a single percept (Watkins and Holt, 2000). While it could in theory disturb perception to group these reflections with the direct speech, studies have shown that strong early reflections improve speech intelligibility by increasing the loudness of direct sound (Soulodre *et al.*, 1989). These studies did not examine SRM, however. LR arrives after early reflections and is made up of a very large number of delayed and attenuated copies of the original signal with different amounts of amplitude and phase distortions. LR is not perceptually integrated with the direct sound and it interferes with the intelligibility of subsequent sounds (Boothroyd, 2004) due to the temporal distortions LR imposes on the speech signal (Plomp, 1976).

Nabelek and Robinette (1978) found that normal hearing listeners and listeners with hearing impairment benefit equally from a single added ER. Bradley *et al.* (2003) found ERs up to 50 ms after the direct sound could increase the effective signal-to-noise ratio by up to 9 dB and improve speech intelligibility for normal hearing and hearing impaired listeners. However, Bess (1999) found that the listeners with hearing impairment do not experience the benefits of ERs. This disagreement in the literature, as well as the limited information on the role of early reflections on SRM, motivates further examination of the role of early and late reflections on speech intelligibility for colocated and separated maskers by testing listeners with and without hearing impairment. As aging and hearing loss are not conditions to which listeners can be assigned and often present together, it is important to examine the relative effects of each by testing a sample varying in both age and hearing ability. Often the two cannot be separated completely in the groups of listeners included in the sample, but statistical methods can be used to further examine the differences among groups as was done in Srinivasan *et al.* (2016). Below, we report the results of an examination of spatial release from masking in the presence of full reverberation, early reflections, or late reverberation for listeners varying in age and hearing ability.

2. Methods

2.1 Listeners

Three listener groups were recruited based on age and hearing status. The young normal hearing (YNH) individuals (n = 11) had audiometric thresholds of ≤20 dB hearing level (HL) [hearing level re: ANSI 3.6-2004 (ANSI, 2004)] at all octave frequencies between 250 Hz and 8 kHz, whereas the older normal hearing (ONH) individuals (n = 15) had thresholds of 20 dB HL or better up to 2 kHz, but up to 25 dB HL at 4 kHz and up to 35 dB at 8 KHz. Older hearing impaired (OHI) individuals (n = 18) had thresholds between 5 dB HL and 45 dB HL at frequencies up to 2 kHz and thresholds of 15 to 65 dB HL at 4 kHz and thresholds of 10 to 80 dB HL at 8 kHz. Tympanometry was performed to rule out middle ear abnormalities and no more than one air-bone gap greater than 10 dB was present at octave frequencies from 500 Hz to 4 kHz. All listeners were in good health with no history of otological disorders. Also, all participants had scores of 24 or higher on the mini-mental state examination (Folstein *et al.*, 1975) to rule out dementia or any other cognitive impairment. None of the individuals in the OHI group used hearing aids for everyday listening, which is consistent with the no greater than moderate hearing loss present in this group.

2.2 Stimuli

All available sentences for three of the male talkers in the Coordinate Response Measure corpus (CRM; Bolia *et al.*, 2000) were used for the target and maskers. All sentences in the CRM have the form "Ready [CALL SIGN] go to [COLOR] [NUMBER] now." There are eight possible call signs (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, and Tiger), four colors (Blue, Red, White, and Green) and eight numbers (1–8). All sentences were bandpass filtered from 80 Hz to 8 kHz.

Room simulation. Virtual acoustic techniques were used to simulate the reverberant environment with the following dimensions: length: 5.7 m; width: 4.3 m; height: 2.6 m. The broadband (125–4000 Hz) reverberation time (T_{60}) was 500 ms. Room simulation techniques were identical to those described by Zahorik (2009) with the application of an additional equalization filter to correct for the loudspeaker response used in the head-related transfer function (HRTF) measurement procedures. The simulation method uses an image model (Allen and Berkley, 1979) to compute directions, delays, and attenuations of early reflections, which are then, along with the direct-path, spatially rendered using non-individualized HRTFs. Late reverberant energy was simulated statistically using exponentially decaying independent Gaussian noise samples in octave bands from 125 to 4000 Hz for each ear. Zahorik's (2009) matching procedure was used to match the spectral power in the six bands to the average power in the last 10 ms of the early reflections. Overall, this method of room simulation has been found to produce BRIRs that are reasonable physical and perceptual approximations of those measured in real rooms (Zahorik, 2009).

Signal processing techniques were employed to create three room impulse response (RIR) conditions: ALL, EARLY, and LATE. All three parts of the RIR (DS, ERs, and LR) were retained for the ALLcondition. DS and the first 50 ms of the RIR [images from Allen and Berkley's image model (1979)] were retained for the EARLY condition. DS and portions of RIR later than 50 ms [exponentially decaying independent Gaussian noise samples in octave bands determined by the Sabine equation (1922)] were retained for the LATE condition.

Within each simulated room, the target speech source was positioned at a distance of 1.4 m in front of the listener, who was located at the center of the room facing the sound source. Two spatial separations in azimuth between the target and masking speech sources were used: colocated (all three sentences presented from 0° azimuth) and spatially separated condition (target at 0°, symmetrical maskers at ±30°). Target and masking speech were convolved with the RIRs of different conditions for their appropriate locations relative to the listener.

2.3 Procedure

Listeners were seated in a sound-attenuating chamber at the National Center for Rehabilitative Auditory Research (NCRAR, Portland, OR, USA) and listened to speech stimuli presented over insert earphones (ER2; Etymotic Research, Elk Grove, IL, USA). On each trial, the listener was presented with a set of three simultaneous CRM sentences. The goal was to attend to the sentence identified by the call sign "Charlie" and ignore the two masking sentences. The target and masking talkers varied randomly from trial to trial. Brungart (2001) showed that there is no significant advantage to a listener when asked to identify any one of the CRM talkers specifically.

The target speech was presented to the listeners at 20 dB sensation level (SL) and was kept constant during the experiment. The masking sentences were presented at levels relative to the target and were appropriately scaled in SL to achieve a set of ten target-to-masker ratios (TMRs) varying in 2 dB steps from 10 to −8 dB. These TMRs were presented in a descending staircase with two presentations at each TMR as described in Gallun *et al.* (2013), and as described in that study, thresholds were estimated based on a linear transformation from number of correct responses (both color and number of the target call sign has to be correct to be counted as a correct response). This procedure is very rapid and provides fairly similar estimates of threshold in the co-located condition to what would be obtained with a longer adaptive tracking procedure and only slightly underestimates thresholds in the spatially separated conditions when threshold is near −10 or +10 dB (Gallun *et al.*, 2013).

Performance was measured in six conditions: colocated and spatially separated for all three BRIR manipulations. Responses were obtained using a touch screen located in front of the listener. Feedback was given after each presentation in the form of "correct" or "incorrect." Data collection was self-paced and listeners were instructed to take breaks whenever they felt the need. All procedures were approved by the VA Portland Health Care System Institutional Review Board and all listeners were monetarily compensated for their time. All stimulus presentation and data collection were implemented using MATLAB; statistical analyses were performed using SPSS version 22 (IBM Corporation, Armonk, NY, USA).

3. Results

A repeated-measures analysis of variance on mean TMR thresholds was conducted with spatial separations (0 and 30) and RIR conditions (ALL, EARLY, and LATE) as within-subjects factors and age (younger versus older) and hearing status (NH versus HI) as between-subject factors. Significant main effects were found for all the four factors [spatial separation: F (1, 41) = 347.50, $p < 0.001$; BRIR condition: F (2, 82) = 17.82, $p < 0.001$; age: F (1, 41) = 6.63, $p < 0.05$; pure-tone average (PTA): F (1, 41) = 147.84, $p < 0.001$]. TMR thresholds were significantly lower for the spatially separated conditions compared to the co-located conditions. Overall, the groups were ordered: YNH, ONH, and OHI (going from the lowest to the highest TMR threshold).

There was a significant interaction between RIR condition and spatial separation [F (2, 82) = 4.01, $p = 0.02$]. *Post hoc* analysis using paired sample t-tests and Bonferroni correction ($p < 0.05$) revealed that for the colocated condition, identification thresholds for the EARLY condition were significantly lower than ALL and LATE conditions and there was no significant difference between ALL and LATE conditions. This was true for all the three listener groups. However, for the spatially separated condition, the thresholds were ordered: EARLY, ALL, LATE (going from the lowest to highest TMR threshold). Also, there was a significant interaction between hearing loss and spatial separation [F (1, 41) = 75.03, $p < 0.001$] indicating that the difference in thresholds between people with normal hearing and impaired hearing was higher in the spatially separated condition as compared to colocated condition.

To further examine the trends suggested by the main effects of hearing loss, age, spatial condition, and RIR condition, as well as the significant interactions, age, and hearing loss were examined as continuous rather than categorical statistical entities. Figure 1 shows the relationships between age, PTA, and identification thresholds for the three RIR conditions in this experiment. Hearing loss was quantified by the PTA of audiometric thresholds for the octave frequencies 0.5, 1, 2, and 4 kHz. Correlations among age, PTA, and identification thresholds for the three different BRIR conditions were statistically significant ($p < 0.05$) at colocated and spatially separated listening conditions. One limitation of this approach, however, is a fairly high correlation [$r(42) = 0.55$, $p < 0.001$] between age and PTA. To address this, multiple regression approaches employed by Srinivasan *et al.* (2016) were used to remove the potential confounding correlation between age and PTA to evaluate the individual contributions of each of the confounding variables. Table 1 shows the proportion of variance accounted for and standardized regression coefficients for the predictor variables (age and PTA) for the multiple regression analyses predicting colocated and spatially separated thresholds at three RIR conditions. PTA, rather than age, was the significant predictor of colocated thresholds in all the three RIR conditions. However, unlike the colocated condition, both age and PTA were significant predictors of spatially separated thresholds in all the three RIR conditions.

In addition, in order to separate the overall performance differences among the three groups from the effects of reverberation on the presence and ability to use spatial cues, the colocated and spatially separated thresholds were subtracted to

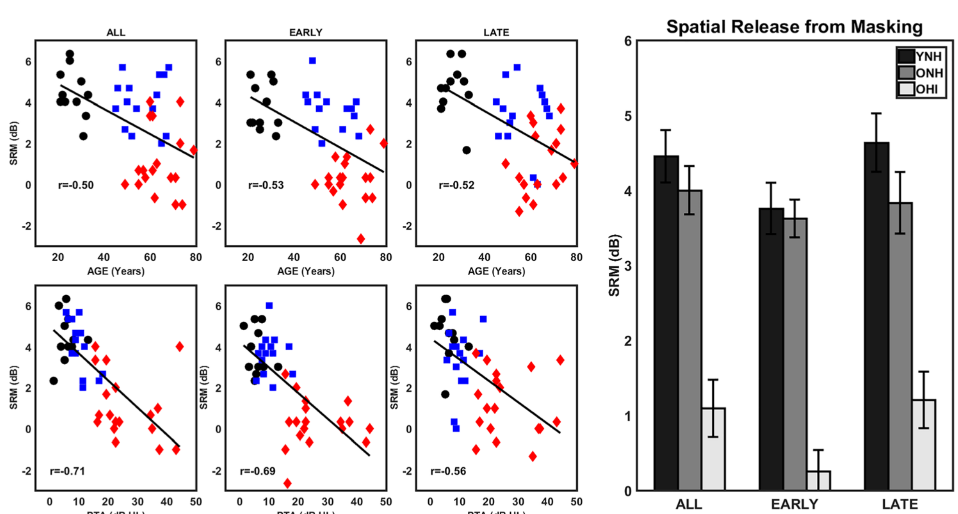

Fig. 1. (Color online) Left panel shows SRM for the three RIR conditions as a function of age (top row) and PTA (bottom row) for individual listeners. Within each panel, YNH are denoted by circles, ONH by squares, and OHI by diamonds. The solid lines are the least squares fits to the data. All correlations are significant at $p < 0.001$. Right panel shows the average SRM for the three RIR conditions as a function of listener groups. Error bars indicate ± 1 SEM.

Table 1. Standardized multiple regression coefficients for age and PTA predicting colocated thresholds, spatially separated thresholds, and SRM for the three BRIR conditions. All multiple regression models are significant at $p < 0.001$ level. Significant model contributors are indicated in bold font.

	Colocated			Separated			SRM		
BRIR condition	R^2	Age	PTA	R^2	Age	PTA	R^2	Age	PTA
All	0.38	0.20	**0.48**	0.76	**0.23**	**0.73**	0.52	−0.16	**−0.62**
Early	0.46	0.16	**0.57**	0.73	**0.24**	**0.70**	0.50	−0.22	**−0.57**
Late	0.40	0.05	**0.61**	0.72	**0.27**	**0.67**	0.38	**−0.31**	**−0.39**

produce a measure of SRM for the three RIR conditions. SRM for the three RIR conditions as a function of age and PTA is shown in the left panel of Fig. 2. The right panel of Fig. 2 displays the mean SRM (±1 standard error of the mean) for the three RIR conditions as a function of listener group. All correlations between SRM and age were statistically significant ($p < 0.05$) with the correlation values ranging from −0.50 to −0.53. Also, all correlations between SRM and PTA were statistically significant ($p < 0.05$) with the correlation values ranging from −0.56 to −0.71. Table 1 shows the amount of variance accounted for and standardized regression coefficients for the predictor variables (age and PTA) for the multiple regression analyses predicting SRM at three RIR conditions. PTA, rather than age, was the significant predictor of SRM in ALL and EARLY conditions. However, both age and PTA were significant predictors of SRM in LATE condition. While the effects of age were fairly consistent across all three RIR conditions, there was a reduction in the effectiveness of PTA as a predictor for the LATE condition, which appears to be driven primarily by the reduction in SRM for a few of the older participants with better PTAs.

4. Discussion and conclusion

The main goal of this study was to investigate the role of early and late reflections on spatial release from masking. All three listener groups tested in this experiment obtained release from masking in all three RIR conditions. Although all listeners benefited from spatially separating the maskers from the target, older hearing impaired listeners benefit dramatically less compared to older normal hearing and young normal hearing listeners. Also, reduced SRM in the early condition was due to the greater improvement in performance for the colocated condition than for the separated condition.

Multiple regression analyses predicting the amount of SRM using age and PTA indicated that age and PTA were significant predictors predicting SRM in LATE conditions, but that PTA alone predicted SRM in the EARLY and ALL conditions. These results suggest that aging plays a role in speech intelligibility in reverberant

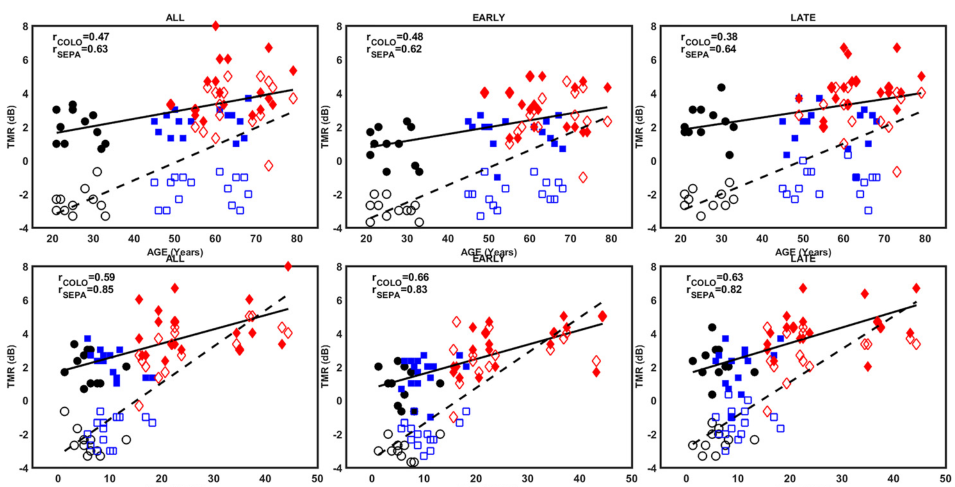

Fig. 2. (Color online) Identification thresholds at the three RIR conditions for colocated (filled symbols) and separated conditions (open symbols) as a function of age (top row) and PTA (bottom row) for individual listeners. Within each panel, YNH are denoted by circles, ONH by squares, and OHI by diamonds. The solid lines are the least squares fits to the colocated data and the dashed lines to the separated data. All correlations are significant at $p < 0.05$.

listening environments when all the early reflections are removed, reducing the effective signal-to-noise ratio between the direct speech and reverberation and making the signal less intelligible.

The results of this study indicate that PTA is a better predictor of SRM in ALL and EARLY than is age, but that the age plays a role in the LATE condition. One possible explanation of this finding could be that in the ALL and EARLY conditions, the strength of correlation between PTA and SRM was stronger than the correlation between age and SRM. Another possible explanation for the differential contributions of age and PTA to SRM in results could be, as explained in Srinivasan *et al.* (2016), the variability of hearing status of the listeners used in these experiments. PTA being a better predictor of SRM is consistent with the results of Srinivasan *et al.* (2016) and Glyde *et al.* (2013), both of which used listeners varying more substantially in age and PTA than in Gallun *et al.* (2013).

In summary, these results indicate that removing the late reflections in BRIR improves speech understanding most when spatial cues are absent. This is true for both normal hearing and hearing impaired listeners, across a wide range of ages. Reduced SRM due to aging, as distinct from hearing impairment, was only observed when early reflections were removed. If we could design dereverberation algorithms for hearing aids such that they remove the artifacts on the speech signal caused by the late arriving reflections while leaving the beneficial early reflections intact, we would predict that the performance of individuals using the algorithm would improve in reverberant listening conditions, especially when the additional cue of spatial separation is reduced. Such reduction could occur through the physical layout, spatial perception deficits of the listener, hearing aid processing, or a combination of all three. Also, the results suggest that it would be useful in the future to aim to recruit samples in which age and hearing loss are as uncorrelated as possible in order to better allow the statistical modeling to distinguish between these factors.

Acknowledgments

We would like to thank all of the participants who volunteered their time to be involved in this experiment. We are also grateful to Rachel Ellinger, Kasey Jakien, and Sean Kampel for their assistance with data collection and Samuel Gordon for engineering support. This work was supported by the NIH-NIDCD Grant No. R01 DC011828. The contents of this article are the private views of the authors and should not be assumed to represent the views of the Department of Veteran Affairs or the United States Government.

References and links

Alain, C. (**2007**). "Breaking the wave: Effects of attention and learning on concurrent sound perception," Hear Res. **229**, 225–236.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**(4), 943–950.

ANSI (**2004**). ANSI 3.6-2004, *American National Standard Specification for Audiometers* (American National Standards Institute, New York).

Bess, F. H. (**1999**). "Classroom acoustics: An overview," Volta Rev. **101**, 1–14.

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**, 1065–1066.

Boothroyd, A. (**2004**). "Room acoustics and speech perception," Semin. Hear. **25**, 155–166.

Bradley, J. S., Sato, H., and Picard, M. (**2003**). "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233–3244.

Brungart, D. (**2001**). "Evaluation of speech intelligibility with the coordinate response measure," J. Acoust. Soc. Am. **109**, 2276–2279.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (**1975**). "Mini-mental scale: A practical method for grading the cognitive state of patients for the clinician," J. Psychiat. Res. **12**, 198.

Gallun, F. J., Kampel, S. D., Diedesch, A. C., and Jakien, K. M. (**2013**). "Independent impacts of age and hearing loss on spatial release in a complex auditory environment," Front. Neurosci. **252**(7), 1–11.

Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (**2013**). "The effects of hearing impairment and aging on spatial processing," Ear. Hear. **34**(1), 15–28.

Helfer, K. (**1992**). "Aging and the binaural advantage in reverberation and noise," J. Speech Hear. Res. **35**, 1394–1401.

Marrone, N., Mason, C. R., and Kidd, G., Jr. (**2008**). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am. **124**, 1146–1158.

Nabelek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). "Reverberant overlap and self-masking in consonant identification," J. Acoust. Soc. Am. **89**, 1259–1265.

Nabelek, A. K., and Robinette, L. (**1978**). "Influence of the precedence effect on word identification by normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **63**, 187–194.

Plomp, R. (**1976**). "Binaural and monaural speech-intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica **34**, 201–211.

Sabine, W. C. (**1922**). *Collected Papers on Acoustics* (Harvard University Press, Cambridge, MA).

Soulodre, G. A., Popplewell. N., and Bradley, J. S. (**1989**). "Combined effects of early reflections and background noise on speech intelligibility," J. Sound Vib. **135**, 123–133.

Srinivasan, N., Jakien, K. M., and Gallun, F. J. (**2016**). "Release from masking for small spatial separations: Effects of age and hearing loss," J. Acoust. Soc. Am. **140**, EL73–EL78.

Watkins, A. J., and Holt, N. J. (**2000**). "Effects of a complex reflection on vowel identification," Acustica **86**, 532–542.

Zahorik, P. (**2009**). "Perceptually relevant parameters for virtual listening simulation of small room acoustics," J. Acoust. Soc. Am. **126**(2), 776–791.