Speech intelligibility in rooms: Disrupting the effect of prior listening exposure

Eugene J. Brandewie^{a)}

Center for Applied and Translational Sensory Science, University of Minnesota, Minneapolis, Minnesota 55455, USA

Pavel Zahorik^{b)}

Department of Otolaryngology and Communicative Disorders, University of Louisville, Louisville, Kentucky 40292, USA

(Received 21 April 2016; revised 26 April 2018; accepted 2 May 2018; published online 23 May 2018)

It has been demonstrated that prior listening exposure to reverberant environments can improve speech understanding in that environment. Previous studies have shown that the buildup of this effect is brief (less than 1 s) and seems largely to be elicited by exposure to the temporal modulation characteristics of the room environment. Situations that might be expected to cause a disruption in this process have yet to be demonstrated. This study seeks to address this issue by showing what types of changes in the acoustic environment cause a breakdown of the room exposure phenomenon. Using speech carrier phrases featuring sudden changes in the acoustic environment, breakdown in the room exposure effect was observed when there was change in the late reverberation characteristics of the room that signaled a different room environment. Changes in patterns of early reflections within the same room environment did not elicit breakdown. Because the environmental situations that resulted in breakdown also resulted in substantial changes to the broadband temporal modulation characteristic of the signal reaching the ears, results from this study provide additional support for the hypothesis that the room exposure phenomenon is linked to the temporal modulation characteristics of the environment. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5038278>

[VB]

Pages: 3068–3078

I. INTRODUCTION

Reverberation is known to be detrimental to speech intelligibility (Knudsen, 1929). This is due to its temporal smearing effects, which result in both “overlap masking” of one speech token onto subsequent speech tokens, and a reduction of the inherent temporal fluctuations (amplitude modulation, AM) in the speech signal (Bolt and MacDonald, 1949; Houtgast and Steeneken, 1985; Nábelek *et al.*, 1989). However, in moderate amounts of reverberation, speech intelligibility is largely unaffected for most normal-hearing listeners. There is growing evidence that normal-hearing listeners have perceptual mechanisms that compensate for the influence of reverberation on the speech signal (Beeston *et al.*, 2014; Brandewie and Zahorik, 2010, 2013; Srinivasan and Zahorik, 2013, 2014; Ueno *et al.*, 2005; Watkins, 2005; Watkins *et al.*, 2011), and the compensation appears to depend strongly on the temporal modulation characteristics of the preceding input (Srinivasan and Zahorik, 2014; Watkins, 2005; Watkins *et al.*, 2011). This compensation process, referred to here as the room exposure effect, has been shown to build up quickly, taking less than 1 s of listening exposure time to the room environment to achieve maximal compensation (Beeston *et al.*, 2014; Brandewie and

Zahorik, 2013). However, it is currently unclear whether and how the compensation might breakdown in response to new room acoustic information.

The current study is an extension of work by Brandewie and Zahorik (2010), where the effect of acoustic contexts on changes in overall speech intelligibility was measured. In that study, the authors used speech materials from the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000) in a virtual room simulation. In their “no carrier” (NC) condition, no room context was provided prior to a speech target consisting of a color and number (i.e., “Green three”). To minimize room exposure in the NC condition, the authors switched the virtual room environment from trial to trial. In the “sentence carrier” (SC) condition, two sentences of speech material were presented in the same reverberant context with the target color and number embedded at the end of the second sentence. Exposure was maximized in the SC condition by presenting the same reverberant environment across a block of trials. Intelligibility was assessed at nine signal-to-noise ratios (SNRs) using a spatially separated broadband masker, also present in the virtual room environment. Psychometric functions were fit to the data for each condition and speech reception thresholds (SRTs) were estimated as the SNR where performance was at the midpoint of the function. It was found that SRTs were significantly lower in the SC condition compared to the NC condition. This effect was not found in an anechoic environment or under monaural listening for most participants. Together this

^{a)}Electronic mail: ebrandew@umn.edu

^{b)}Also at: Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY 40292, USA.

suggests that prior binaural exposure to the reverberant room environment aided speech intelligibility. Differences in SRTs were 2.68 dB on average, which translated to a mean improvement of about 18% in speech intelligibility. This difference in intelligibility was attributed to the effect of prior exposure to the room acoustics.

It has been suggested that room exposure effects may show some similarities to the precedence effect in sound localization (Brandewie and Zahorik, 2010), where the arrival of the direct waveform dominates the perceived location of a sound source in an acoustically reflective environment (Wallach *et al.*, 1949). More specifically, the precedence effect has also been shown to strengthen, or build up with repeated exposure to the sound source and reflection (Freyman *et al.*, 1991). This buildup of precedence is rapid (Freyman *et al.*, 1991), appears to be mediated by cortical-level auditory processing (Grantham, 1996) but is subconscious and automatic (Clifton and Freyman, 1997). Many of these observations seem consistent with those observed in the room exposure effects on speech understanding (Brandewie and Zahorik, 2010), particularly the buildup in compensation over comparatively brief exposure times (Beeston *et al.*, 2014; Brandewie and Zahorik, 2013).

The processes of auditory object formation and auditory streaming also show patterns of buildup, although typically at time scales longer than those observed for precedence effect buildup (Bregman, 1978, 1990). Brown and Stecker (2013) have argued that the fusion of the direct waveform with early reflections, as takes place during the precedence effect, may be another example of auditory object formation and streaming, where the auditory system binds reverberant reflections together into a coherent single object at a single perceived location. Perhaps such object formation processes relate to the effect of prior room exposure on speech perception and speech understanding.

An additional important observation in both the precedence effect and auditory object formation/streaming is that the perceptual facilitation that results from buildup can be rapidly degraded in response to new perceptual input that is inconsistent with the input that resulted in the buildup. Such “breakdown” in the precedence effect was initially shown to occur when abrupt changes are made to the spatial configuration of sound source and echo (Clifton, 1987). Subsequent studies have shown that breakdown can also result from other changes, such as the relative delay between source and echo (Clifton *et al.*, 1994), and changes to source/echo spectral characteristics (McCall *et al.*, 1998). It has also been shown with auditory streaming that sudden changes in the stimulus presentation can cause a resetting, or breakdown, of the buildup of segregation (Haywood and Roberts, 2013; Roberts *et al.*, 2008; Rogers and Bregman, 1993, 1998). To the extent that room exposure effects on speech understanding are related to the precedence effect and/or auditory object formation/streaming, one might expect to find similar evidence of breakdown following abrupt changes to the room acoustic environment. Such evidence of breakdown has yet to be shown empirically, however, and the precise aspects of room acoustic change that may elicit breakdown are unknown.

It has been previously demonstrated that the room exposure effect depends strongly on the temporal envelope of the preceding sound (Srinivasan and Zahorik, 2014; Watkins, 2005; Watkins *et al.*, 2011). It is also well known that changes to the reverberation characteristics of a room, particularly reverberation time, can strongly influence the way in which the room distorts the temporal modulation characteristic of sound reaching a receiving point, and that these distortions often result in reduced speech understanding (Houtgast and Steeneken, 1985). Thus, a compensation mechanism that restores the distortions to the speech temporal envelope caused by reverberation could be ecologically advantageous, and is supported by physiological evidence showing that certain populations of neurons in the inferior colliculus are resistant to the temporal distortions of reverberation (Kuwada *et al.*, 2012, 2014; Slama and Delgutte, 2015). To the extent that the room exposure effect also exhibits breakdown, then one might expect the breakdown to be driven by sudden changes to the temporal modulation characteristic of the room, given that this characteristic is most likely involved in the buildup of the effect.

On the other hand, it has been noted that sudden changes in the spatial location of a stimulus and/or its acoustic reflection(s) can cause a breakdown in both precedence (Clifton, 1987) and auditory streaming (Rogers and Bregman, 1998). Therefore, it is conceivable that a sudden change in the early reflection spatial pattern could also cause a breakdown in room exposure buildup. Although Watkins *et al.* (2011) demonstrated that vocoded speech still provided consistent monaural room exposure effects, which suggests that the early reflection spatial pattern plays little role in the buildup of room exposure, it is still possible that the perceived locations of reflections affect buildup. If so, then a sudden change in these perceived spatial aspects would predict breakdown similar to that observed for auditory streaming or precedence.

The purpose of this study is to (a) determine whether the room exposure effect breaks down in response to new acoustic information about the environment, and (b) explore the roles of temporal and spatial aspects of the new acoustic information implicated in breakdown. The general testing paradigm including the use of virtual auditory space techniques is similar to previous work (Brandewie and Zahorik, 2010). Experiment 1 determines whether a sudden change in the temporal modulation characteristics of the acoustic environment resulting from a change imposed on late reverberant energy elicits breakdown. Experiment 2 investigates whether a sudden change to the spatial patterns of early reflections causes breakdown.

II. EXPERIMENT 1: INCONGRUENT LATE REVERBERATION PATTERNS

A. Methods

1. Listeners

Seventeen listeners (12 female) ages 18–25 yr were recruited for this experiment. All had normal hearing as verified by pure-tone air-conductive audiometric testing

[thresholds of 25 dB hearing level (HL) or better at octave frequencies from 250 to 8000 Hz in both ears] and were fluent in English. Listeners were paid for their participation. All 17 listeners participated in the R1 conditions and 11 of these listeners participated in the R3 conditions. All procedures involving human subjects were approved by the University of Louisville Institutional Review Board.

2. Room modeling

Virtual acoustic space techniques were used to simulate the room environments tested in this experiment. The techniques were identical to those described by Brandewie and Zahorik (2010) and are described in more detail in Zahorik (2009). This simulation technique has been found to produce binaural room impulse responses (BRIRs) that are reasonable physical and perceptual approximations to those measured in a real room (Zahorik, 2009).

Five rooms were simulated: R1, R2, R3, R4, R5. The size and dimensions of the rooms were identical [width (x): 4.257 m, length (y): 5.673 m, height (z): 2.583 m]. Only the absorptive properties of the simulated surfaces were varied between environments. The energy absorption coefficients (α) used in the model for each reverberant room are presented in Table I. Since the room simulation technique treats early reflections and late reverberation separately, there are

TABLE I. Energy absorption (α) input parameters for early and late energies for the room simulation model and resulting room acoustic parameters: T_{60} , C_{50} , and broadband interaural cross-correlation coefficient (IACC), and speech transmission index (STI) for the speech source position (0 deg azimuth, 1.4 m distance) estimated from model BRIR outputs for each simulated reverberant room (R1–R5). Note the decrease in STI with increasing reverberation. This is indicative of increases in temporal envelope distortion (decrease in temporal modulation depth) with increasing reverberation.

Parameter	Center Frequency (Hz)	R1	R2	R3	R4	R5
Early α	Broadband	0.390	0.290	0.124	0.064	0.041
Late α	125	0.533	0.400	0.171	0.089	0.057
	250	0.400	0.300	0.129	0.067	0.043
	500	0.400	0.300	0.129	0.067	0.043
	1000	0.400	0.300	0.129	0.067	0.043
	2000	0.293	0.220	0.094	0.049	0.031
T_{60} (s)	4000	0.267	0.200	0.086	0.044	0.029
	Broadband	0.316	0.488	1.216	2.379	2.966
	125	0.364	0.322	0.776	1.439	2.655
	250	0.293	0.401	0.933	1.717	2.751
	500	0.328	0.394	0.914	1.808	2.748
C_{50} (dB)	1000	0.368	0.421	0.957	1.888	3.015
	2000	0.317	0.510	1.234	2.399	3.430
	4000	0.096	0.496	1.273	2.526	1.581
	Broadband	25.8	17.2	5.6	0.7	-6.5
	125	8.5	16.0	0.1	-1.2	-23.5
IACC	250	11.1	17.7	7.4	0.6	-17.8
	500	11.1	15.9	4.3	0.25	-19.1
	1000	8.9	14.4	3.0	-1.2	-21.5
	2000	27.5	16.1	4.5	-0.1	-3.7
	4000	42.3	19.5	8.3	3.5	20.4
STI	Broadband	0.90	0.89	0.45	0.33	0.13
		0.932	0.929	0.723	0.596	0.563

separate sets of coefficients for each portion of the simulation, and only the late reverberation simulates any frequency-dependent absorption effects.

The absorption coefficients for R2 were identical to those from Zahorik (2009) and Brandewie and Zahorik (2010), and were designed to approximate a real, moderately reverberant room (a large office room) with gypsum board walls, hence the somewhat greater absorption at low frequencies. The amount of reverberant energy for the other reverberant rooms was controlled by setting the absorption coefficients to be multiplicative factors of the coefficients from R2. These values were chosen to create a continuum of rooms varying in reverberation times (T_{60}). Common room acoustic parameters are shown in Table I for each simulated room. The parameters were computed from the BRIRs based on ISO-3382 (1997). Speech transmission index (STI) values were also computed for each BRIR using methods described in IEC-60268-16 (2003) and Schroeder (1981).

Each simulated environment presented a speech stimulus simulated to be 1.4 m directly in front of the listener (0° azimuth angle) and a broadband Gaussian noise masker simulated to be 1.4 m directly opposite the listener's right ear (90° azimuth). The masker always preceded the speech by 150 ms, during which the masker's amplitude linearly increased from zero to full-scale. The masker was present throughout the speech and ended (without ramping) with the speech stimulus. No attempt was made to equalize sound levels across the room environments, therefore the more reverberant rooms produced greater at-the-ear sound levels than less reverberant rooms, as would occur naturally.

3. Speech corpus

Speech materials for this study were from the CRM corpus (Bolia *et al.*, 2000). Each speech sentence in the corpus has the format "Ready <Call Sign> go to <Color> <Number> now." The corpus features eight talkers (four male, four female), eight call signs (Charlie, Ringo, Laker, Hopper, Arrow, Tiger, Eagle, Baron), four colors (Blue, Red, White, Green), and eight numbers (1–8). All combinations of talkers, call-signs, colors, and numbers were used in these experiments.

4. Room exposure conditions

Three conditions were created that varied in the length and content of the speech carrier phrase that preceded the target phrase. The "no carrier" conditions (NC) limited prior exposure to the listening environment by presenting listeners with only the color/number target without a preceding carrier phrase. The "congruent carrier" conditions (CC) provided additional room exposure by presenting listeners with a two-sentence carrier phrase preceding the color/number target. The NC and CC conditions were identical to the two conditions of Brandewie and Zahorik (2010), although in the previous study the CC condition was designated SC ("sentence carrier"). The "incongruent carrier" conditions (IC) were similar to the CC conditions, except that a switch was made to another room environment immediately prior to the presentation of the color/number target. The carrier phrase was

convolved with one BRIR and the target phrase convolved with a different BRIR. The point in the sentence at which a switch in the environment occurred is illustrated in Fig. 1. This change was performed by replacing the waveform of the target portion of the CC phrase with that of a different convolved waveform. Therefore, this change occurred immediately without any “spill-over” of the previous environment’s reverberant tail into the target phrase, but rather had a reverberant tail of the new environment.

The switch points were selected individually for each of the 2048 sentences of the CRM corpus. These selections were made on the original (anechoic) CRM waveforms prior to any convolution with BRIRs. An algorithm was used to aid the selection process by finding the nearest zero-crossing at the selected point. That selected sample was then saved in a data file that corresponded with each individual CRM waveform. These selection points were used to extract the color/number targets for the NC conditions while also providing the switch points for IC conditions.

5. Design and procedure

Listeners were tested in a paradigm similar to that used in a previous study (Brandewie and Zahorik, 2010). Listeners were tested at nine signal-to-noise ratios (SNRs): -28 to $+4$ dB in 4 dB steps to ensure that psychometric functions could be fit to all room environments. These were the same SNR values used by Brandewie and Zahorik (2010) and Zahorik and Brandewie (2016). SNR was manipulated by adjusting the gain of the speech target prior to convolution with the BRIRs. The masker level was fixed. Target color and number, and the SNR were selected at random for each trial.

Each listener participated in blocks (of 54 trials each) of NC, CC, and IC conditions. The primary room environments tested were R1 and R3, which represented reverberation times ($T_{60} \sim 0.3$ s, and ~ 1.2 s, respectively) of typical indoor environments. All listeners participated in five blocks of the CC condition for each room environment (R1 and R3) in which the room environment was kept constant across the block of trials. Maintaining the same room across a block of trials was meant to maximize room exposure in the CC conditions. This methodology was identical to that used by Brandewie and Zahorik (2010).

The listeners also completed 15 blocks of the NC condition, in which the room environment was randomly varied from trial to trial. This manipulation was expected to minimize carry-over effects from exposure to a particular room

environment in the NC conditions. These blocks always included three room environments, but the specific room environments varied with each listener (see below). Only NC trials with R1 or R3 targets were analyzed and compared.

Listeners also participated in 10 blocks of the IC condition in which the carrier phrase was always presented in the R2 ($T_{60} \sim 0.5$ s) room environment, but the target was presented in a separate room environment. Much like the NC condition, the target phrase was presented in a randomly varied room environment from trial to trial. Due to an unfortunate change in methodology during the data collection phase of this study, the particular room environments that were presented varied with each listener. For six of the listeners the IC phrases switched from R2 to either R1 or R5, for the other 11 listeners this change was from R2 to R1 or R3. Additionally, for the 11 listeners that had R3 in their IC condition, the NC condition included rooms R1, R2, and R3, randomly selected for each trial. For the other six listeners, the NC condition included rooms R1, R2, and R5. It has been shown that highly reverberant rooms, such as R5 ($T_{60} \sim 3.0$ s), fail to show any evidence of the buildup of room exposure (Zahorik and Brandewie, 2016). Therefore, room R5 was eliminated from the testing conditions during data collection and resulted in the different methodology for the two groups. This difference is further addressed in Sec. II B.

Listeners were seated in a sound-attenuating chamber (Acoustic Systems, Austin, TX). All stimuli were presented over equalized headphones (Beyerdynamic DT-990 Pro) at a moderate level [70 dB sound pressure level (SPL) peak at the entrance to the ipsilateral ear for $+4$ dB SNR]. The listener’s task was to select the appropriate color and number combination using a computer mouse on a graphical interface. Feedback as to whether the response was correct was provided after every trial. All stimulus presentation and response collection was implemented using custom software in a MATLAB environment (Mathworks Inc., Natick, MA).

6. Data analysis

For each listener, the proportion of correct responses, PC, was computed separately for each room and experimental condition at all SNRs. To be considered a correct response, both the color and number responses had to match the target. Logistic functions were then fit to the PC data using a maximum-likelihood algorithm (psignifit toolbox [ver. 2.56], Wichmann and Hill, 2001) to approximate the

FIG. 1. (Color online) Illustration of the speech carrier phrase and target position for congruent carrier (CC) and incongruent carrier (IC) conditions, including the point at which a change occurs in the acoustic environment in the IC conditions. No carrier (NC) conditions only presented the color and number target in isolation. An example waveform demonstrates the instantaneous transition between room environments in the IC carrier phrase and the target phrase.

psychometric function for each listener in a given condition. The lower asymptote, δ , of the function is set to the chance performance level of 1/32 (3.125%) in this task. Threshold, α , slope, β , and upper asymptote, $1-\lambda$, parameters were estimated using the maximum-likelihood procedure. The threshold parameter, α , represents the SNR that corresponds to the midpoint of the function, which will vary slightly across fits depending on the estimated value of λ . This fitting procedure is similar to that used by Brandewie and Zahorik (2010), however, here the upper asymptote parameter, $1-\lambda$, is estimated by maximum-likelihood procedures rather than set to 1.0. The inclusion of this parameter accounts for “lapses” in subject attention to the task and results in better estimates of psychometric function threshold and slope (Wichmann and Hill, 2001). For more details on how functions were fit to the data, see Zahorik and Brandewie (2016).

Goodness of fit was evaluated by using a deviance statistic, D , which is defined as $2 \log (L_{\max}/L)$, where $L_{\max}/L$ is a likelihood for the saturated model that has as many estimated parameters as data points relative to the best-fitting model with 3 estimated parameters (α , β , and λ). Function fits were excluded based on poor fits to the data from both overdispersion ($p > 0.975$) and underdispersion ($p < 0.025$) (Wichmann and Hill, 2001). This exclusion criterion is identical to that used by Zahorik and Brandewie (2016).

B. Results and discussion

Due to poor function-fits (overdispersion), data from three listeners were excluded from the final results of R1, while data from four listeners were excluded from R3. All remaining fits can be assumed to be reasonable approximations to the data ($0.025 \leq p \leq 0.975$) with a mean deviance, $D = 7.80$, $SD = 3.34$. In order to focus solely on the changes in the room exposure effect the data were analyzed and compared using the SNR threshold values (α) alone. For more details on function fitting and how changes in room environments intrinsically affect threshold and slope, we refer readers to Zahorik and Brandewie (2016).

To determine if the changes in the methodology of the NC blocks affected performance, a two-sample t -test was performed on the R1 NC thresholds between listeners that had R3 versus those that had R5 in their NC test blocks. There was no significant difference between the two groups [$t(15) = 2.13$, $p = 0.38$]. Therefore, the data from both these groups were combined for further analysis.

Mean SNR thresholds for each room environment in each condition are presented in Fig. 2 with errors bars representing the standard error of the mean. The results for 14 participants are included in the R1 data, and seven participants are included in the R3 data. Significant differences ($p < 0.05$) between experimental conditions are shown in Fig. 2 (indicated with an asterisk).

In the R1 environment, a repeated-measures (within-subjects) analysis of variance confirmed a significant main effect of stimulus condition, [$F(2,26) = 21.28$, $p < 0.001$, $\eta^2 = 0.62$]. *Post hoc* comparisons were made between conditions (using Bonferroni corrections). There was a significant improvement in mean threshold between the

NC ($M = -18.69$ dB, $SD = 1.23$) and CC conditions ($M = -20.19$ dB, $SD = 1.44$) [$t(13) = 4.40$, $p < 0.005$, Cohen's $d = 1.12$]. This effect was expected and replicates the advantage of a matching acoustic context (Brandewie and Zahorik, 2010), and therefore demonstrates the buildup phenomenon of the room exposure effect. Performance in the IC condition ($M = -18.11$ dB, $SD = 1.47$) was worse (-0.58 dB) than the NC condition; however, this difference was not statistically significant [$t(16) = 2.36$, $p = 0.10$, Cohen's $d = 0.62$]. Performance in the IC condition was also significantly worse than the CC condition [$t(13) = 5.92$, $p < 0.005$, Cohen's $d = 1.62$], suggesting a breakdown of the room exposure effect. All together these results clearly demonstrate both buildup and breakdown of room exposure effects in R1.

In the R3 environment, results were similar to R1. There was a significant effect of stimulus condition, [$F(2,12) = 39.98$, $p < 0.001$, $\eta^2 = 0.87$]. *Post hoc* comparisons revealed a significant difference in mean threshold between the NC ($M = -10.01$ dB, $SD = 1.56$) and CC conditions ($M = -13.41$ dB, $SD = 1.81$) [$t(6) = 11.63$, $p < 0.001$, Cohen's $d = 2.01$] demonstrating the buildup effect in R3. Although the mean threshold in the IC condition appears somewhat better than the NC condition, this difference did not reach statistical significance ($M = -11.55$ dB, $SD = 1.93$) [$t(6) = 3.16$, $p = 0.06$, Cohen's $d = 0.88$]. Additionally, thresholds with the IC carrier were significantly worse than the CC carrier [$t(6) = 5.58$, $p < 0.005$, Cohen's $d = 0.99$]. Therefore, we can conclude that both buildup and breakdown phenomena were also observed in R3.

Taken together, the results from R1 and R3 both demonstrate clear buildup and breakdown of the room exposure effect in response to changing late reverberant energy, and resultantly, the temporal modulation characteristic of the room. Acoustical analyses clearly show changes in the STI as a function of the reverberation manipulation (see Table I). The buildup effect sizes varied by room environment, where the effect in R3 (3.40 dB) is shown to be larger than in R1 (1.50 dB). This interaction between the effect of prior exposure and the STI was expected and has been explored in more detail in previous work (Zahorik and Brandewie, 2016), and is consistent with other work that shows causal links between the temporal modulation characteristics of the room and the room exposure effect (Srinivasan and Zahorik, 2014; Watkins *et al.*, 2011). The breakdown is quite robust in R1, showing NC-like performance in the IC condition. The magnitude of breakdown in R3 was somewhat less, although still statistically significant. It is possible that this minor imbalance in breakdown was caused by the amount of reverberation preceding the environment switch in the IC condition. For IC in R1, the switch was from a more reverberant room (R2) to a less reverberant room (R1), whereas for IC in R3, the switch was from a less reverberant room (R2) to a more reverberant room (R3). In the latter case, it is possible that some of the buildup to R2 transferred to R3, and therefore somewhat attenuated the breakdown effect. Regardless, it is clear that breakdown of the room exposure effect exists, and results from changes to the temporal

FIG. 2. Mean speech reception thresholds (SRTs) from experiment 1 for the NC, CC, and IC conditions for room environments R1 ($n=14$, $T_{60}=0.32$ s) and R3 ($n=7$, $T_{60}=1.22$ s). In the IC conditions, the preceding carrier phrase was always presented in the R2 environment ($T_{60}=0.49$ s). Error bars represent the standard error of the mean. Asterisks indicate significant differences ($p < 0.05$) evaluated using within-subjects comparisons (see text for details). Decreased SRT from NC to CC indicates a buildup of the room exposure effect. Increased SRT between CC and IC indicates a breakdown of the room exposure effect.

modulation characteristics of the room that are consistent with a new room environment.

III. EXPERIMENT 2: INCONGRUENT EARLY REFLECTION PATTERNS

A. Methods

1. Listeners

Ten paid normal-hearing listeners (six female) ages 19–31 yr participated in this experiment with the same requirements as experiment 1. None of these listeners participated in experiment 1.

2. Room configurations

Identical virtual auditory space techniques were used to simulate three of the reverberant environments from experiment 1 (R2, R3, and R4). New to this experiment was a manipulation of the spatial configuration of the source and masker within each simulated room. Three configurations (A, B, and C), as shown in Fig. 3, were tested in each room.

Configuration A and the corresponding absolute positions of the speech, masker, and listener are identical to that used by Brandewie and Zahorik (2010) and in experiment 1. In the other configurations (B, C), the relative position of the speech source and masker to the listener remained unchanged (i.e., the speech was always presented 1.4 m directly in front of the listener, and the masker was always presented 1.4 m directly opposite the listener's right ear); however, the absolute positions of the speech stimuli and masker were altered.

Unlike experiment 1, the SNR between the speech stimuli and the noise masker was held at a constant -13 dB. Prior to convolution with the BRIRs, the speech and noise stimuli had equal RMS amplitudes and SNR was manipulated by digitally reducing the gain of the speech signal before convolution. Based on the psychometric functions of experiment 1, it was concluded that more rapid speech intelligibility assessments could be made at a single SNR as long as comparisons between conditions were limited to the same room environment. Therefore, instead of measuring SNR thresholds, this experiment measured changes in PC between

FIG. 3. Experiment 2 setup showing source and listener configurations within the virtual room listening environments (not to scale) for the three configurations tested (A, B, C). The relative position of the speech source, noise masker, and listener are shown. The listener is always directly facing the speech source in the virtual environment (0-deg azimuth) with the noise masker to the listener's right side (90-deg azimuth).

FIG. 4. (Color online) Binaural room impulse responses (BRIRs) (left ear, 0 degree azimuth only) for each room configuration for the R4 room environment are displayed. Differences can be seen in the arrival time and intensity of the early reflections (particularly between 10 and 30 ms as shown in inset boxes). The beginning portion (direct waveform) is identical for all three configurations, but this is visually masked by configuration C in this figure.

stimulus conditions. This methodology is similar to previous work demonstrating the buildup phenomenon of the room exposure effect (Brandewie and Zahorik, 2013).

For the room configuration manipulation of this experiment, it is important to verify that changes in the room configuration resulted in minimal changes to the temporal modulation characteristics of the room while clearly changing the arrival times and directions of the early reflections. The BRIRs of the left-ear channel (0 deg azimuth) in R3 are presented in Fig. 4 for each configuration (A, B, C). Visual inspection of these impulse responses reveals that the overall energy decay represented by the envelope of the late reverberant tails appears quite similar for each configuration. Table II presents common room acoustic parameters computed from the BRIRs for each room environment. In agreement with visual inspection of the impulse responses, the reverberation times (T_{60}) for each configuration in a given room environment are quite similar. Additionally, in Fig. 4 it is clear that the temporal position and amplitude of the early reflection patterns greatly differ between room configurations, especially between 10 and 30 ms.

To further demonstrate that temporal information is only mildly affected by changes in configuration, the STI was computed for each room configuration and displayed in Table II. These values further demonstrate that changing the

TABLE II. Broadband reverberation times (T_{60}) and clarity indices (C_{50}), speech transmission indices (STI), and broadband interaural cross-correlation (IACC) for each room and configuration.

Room	R2			R3			R4		
Conf.	A	B	C	A	B	C	A	B	C
T_{60} (s)	0.49	0.47	0.47	1.22	1.15	1.15	2.38	2.27	2.27
C_{50} (dB)	13.86	14.67	14.23	3.95	4.77	4.27	-0.79	0.37	-0.09
IACC	0.906	0.909	0.908	0.536	0.609	0.585	0.330	0.400	0.384
STI	0.929	0.924	0.926	0.723	0.738	0.709	0.596	0.604	0.566

room configuration resulted in only small differences in the transmission of temporal information when compared to the calculated differences between room environments. These differences are much smaller than those observed across rooms in experiment I (Table I). All together, these observations indicate that manipulation of the spatial position of the speech and masker created a substantial difference in the early reflection patterns while maintaining nearly identical late reverberation decay tails.

3. Design and procedure

In this experiment, the same three conditions as in experiment 1 were tested (NC, CC, and IC). However, in the IC conditions a switch was made to another spatial configuration (A, B, or C) in the same room environment. This manipulation resulted in a sudden change in the order and location of early reflections while maintaining the same overall temporal modulation characteristics of the room. Twelve condition combinations were tested for each room environment. This included the three NC conditions (one for each configuration), three CC conditions (one for each configuration), and six IC conditions (for each odd pairing of room configurations). However, data for the NC and CC conditions for configuration A were originally collected as part of another study (Brandewie and Zahorik, 2013) in which the same listeners participated and were therefore not included in the blocks of trials for this experiment. These data are reproduced here for comparison. Listeners completed 25 blocks, each containing 60 trials [two repetitions for each combination of condition (10) and room environment (3)].

B. Results and discussion

All individual PC scores were converted to rationalized arcsine units (RAU) to address the non-uniformity of variance near ceiling and floor performance (Studebaker, 1985). All statistical analysis and numerical computations were conducted in RAUs, and then transformed back to PC for display purposes. PC for each room environment in each condition is presented in Fig. 5.

Since each configuration had two effective IC conditions (e.g., B carrier to A target and C carrier to A target in the A configuration), a mean IC score was computed for each listener in each configuration and room for further statistical analysis. A paired-samples t test confirmed no significant differences between these two scores across all configurations and rooms [$t(89) = 2.69, p = 0.99$], providing support for using mean IC scores in subsequent analysis. In agreement with experiment 1, statistical analysis was performed separately for each room environment in which an analysis of variance was performed with three levels of configuration (A, B, C) and three levels of condition (NC, CC, IC) in a repeated-measures design.

In the R2 environment, the analysis of variance found no significant interaction between room configuration and condition [$F(3.76, 33.80) = 0.34, p = 0.84, \text{partial } \eta^2 = 0.03$ (Huynh-Feldt corrected due to sphericity violation)]. There was a significant main effect of stimulus condition, [$F(1.92,$

FIG. 5. (Color online) Mean proportion correct (PC) by condition for each room configuration (panel) and room environment (symbol) for experiment 2. The right panel displays scores averaged across the three target room configurations (A, B, C). The type of symbol indicates the room environment (circles, R2; squares, R3; triangles, R4). Error bars represent the standard error of the mean. Significant differences ($p < 0.05$) between carrier conditions (CC, IC) and corresponding NC conditions in the same room environment are indicated by an asterisk above the symbols for the averaged data (right panel). Breakdown effects are defined as significant differences between CC and IC conditions. No significant breakdown effects were observed. See text for details on the statistical analysis.

17.25) = 6.871, $p < 0.01$, $\text{partial } \eta^2 = 0.43$ (Huynh-Feldt corrected)]. *Post hoc* comparisons were made between conditions (using Bonferroni corrections). There was no significant difference in PC between NC ($M = 86.23$, $SD = 8.99$) and CC conditions ($M = 92.99$, $SD = 9.46$) [$p = 0.06$, Cohen's $d = 0.73$]. This result suggests no significant room exposure effects in the R2 environment. This point is further discussed below. There was a significant difference in mean PC between the NC and IC conditions ($M = 91.79$, $SD = 5.28$) [$p < 0.05$, Cohen's $d = 0.75$], suggesting that performance significantly improved with the incongruent carrier. There was no significant difference between CC and IC performance [$p = 1.00$, Cohen's $d = 0.15$], suggesting no observed breakdown effects. There was also no significant main effect of room configuration [$F(1.37, 12.30) = 2.284$, $p = 0.15$, $\text{partial } \eta^2 = 0.20$].

In the R3 environment, a repeated-measures analysis of variance found no significant interaction between room configuration and condition [$F(4, 36) = 0.93$, $p = 0.46$, $\text{partial } \eta^2 = 0.09$ (no sphericity violation in R3)]. There was a significant main effect of stimulus condition, [$F(2, 18) = 35.28$, $p < 0.01$, $\text{partial } \eta^2 = 0.80$]. *Post hoc* comparisons (with corrections) revealed a significant difference in mean PC between the NC ($M = 49.94$, $SD = 10.77$) and CC conditions ($M = 63.84$, $SD = 10.12$) [$p < 0.01$, Cohen's $d = 1.33$], demonstrating the room exposure effect. There also was a significant difference between NC and IC ($M = 63.13$, $SD = 8.77$) conditions [$p < 0.01$, Cohen's $d = 1.34$], but no difference between CC and IC conditions [$p = 1.00$, Cohen's $d = 0.07$]. Together, these results suggest that there were no observed breakdown effects in R3. There was, however, a significant effect of room configuration [$F(2, 18) = 72.25$, $p < 0.01$, $\text{partial } \eta^2 = 0.89$]. Since the effect of room configuration on speech intelligibility was not of primary interest to this study, detailed *post hoc* comparisons for this effect

are not provided here. However, further discussion and a possible explanation for this outcome is given below.

The results in R4 were similar to those in R3. A repeated-measures analysis of variance found no significant interaction between room configuration and condition [$F(4, 36) = 1.10$, $p = 0.37$, $\text{partial } \eta^2 = 0.11$ (no sphericity violation in R4)]. There was a significant main effect of stimulus condition, [$F(2, 18) = 52.39$, $p < 0.01$, $\text{partial } \eta^2 = 0.85$]. *Post hoc* comparisons (with corrections) revealed a significant difference in mean PC between the NC ($M = 28.44$, $SD = 10.78$) and CC conditions ($M = 46.95$, $SD = 9.54$) [$p < 0.01$, Cohen's $d = 1.82$], demonstrating the room exposure effect. There also was a significant difference between NC and IC ($M = 45.51$, $SD = 7.70$) conditions [$p < 0.01$, Cohen's $d = 1.82$], but no difference between CC and IC conditions [$p = 1.00$, Cohen's $d = 0.17$]. Together, these results also suggest that there were no observed breakdown effects in R4. There was also a significant effect of room configuration in R4 [$F(2, 18) = 40.74$, $p < 0.01$, $\text{partial } \eta^2 = 0.82$].

The results clearly demonstrate that there was *no observable breakdown* between the CC and the IC conditions in each room environment tested. The results of this experiment suggest that it is unlikely that the buildup of room exposure relies on using specific spatial information in the early reflections.

It had previously been suggested (Brandewie and Zahorik, 2010) that the effect of room exposure might rely on a pattern-matching model similar to that which has been suggested for the precedence effect (Blauert and Col, 1992). An internalization of the spatial pattern of reflections could allow for comparisons of new inputs to either enhance (buildup) or suspend (breakdown) the model. However, these results do not support this hypothesis, since sudden changes in this spatial information were ineffective in

eliciting a breakdown. Ecologically, it makes sense for the auditory system to not rely on specific spatial cues, because in everyday situations, the relative positioning of the sound sources and the listener can be in constant flux. Such a pattern-matching model would be burdened with constantly updating the internal model of the acoustic environment with every movement of the source or listener. It is also important to point out that the absence of breakdown effects in this experiment are consistent with the hypothesis that room exposure effects are elicited by the temporal modulation characteristics of the room environment, since room temporal modulation contributions were essentially constant between room configurations.

The magnitude of the observed room exposure effect varied with the room environment, with the greatest effects seen in the R4 environment. This is in agreement with previous research, in which rooms with greater reverberation show an increased effect of buildup of room exposure (Zahorik and Brandewie, 2016). The smaller (and largely non-significant) effects observed in R2 may be due to ceiling effects at the higher performance levels. Previous work (Brandewie and Zahorik, 2010) suggests that at these high performance levels the measurable difference between NC and CC conditions is greatly reduced.

There were significant main effects of room configuration in R3 and R4. The three room configurations were not equated for intelligibility, and the overall intelligibility appeared to rely on the target's configuration. *Post hoc* comparisons in R3 demonstrated that configuration C had significantly worse performance than configuration A [$p < 0.01$, Cohen's $d = 0.91$] and B [$p < 0.01$, Cohen's $d = 1.67$]. Configuration B demonstrated significantly higher overall performance than A [$p < 0.01$, Cohen's $d = 0.98$]. The results in R4 were qualitatively similar to R3. The data shown in Fig. 5 illustrates these differences. These differences likely were a result of an interaction between intense early reflections (whether from the noise masker or the target speech) and which ear was affected by them. Due to the binaural nature of the configurations, the left ear, which was away from the masker, likely benefited from acoustic head-shadow effects (Plomp, 1976). This effect likely interacted with the absolute placement of the masker in our configurations, such that configuration C had high intensity early reflections for the masker, whereas configuration B had high intensity early reflections for the target. It is likely that this complex interaction decreased the at-the-ear SNR for configuration C, while it effectively increased the SNR for configuration B. As one might expect, these acoustic changes were more pronounced in the more reverberant rooms (R3 and R4), and likely contributed to the significant effect of target configuration in these environments.

IV. GENERAL DISCUSSION

This study has demonstrated significant buildup effects in agreement with previous research (Brandewie and Zahorik, 2010); and this study demonstrated that the buildup of room exposure can fail, or breakdown, when presented with an acoustic context consistent with a physically

different room, with different characteristics of late reverberation and therefore different temporal modulation characteristics (experiment 1). Additionally, breakdown effects have been shown to be absent with changes in the spatial positioning of the listener within the same room environment, furthering the hypothesis that these effects are related to the temporal modulation characteristics of the speech signal (experiment 2).

A. Room constancy

Although speech stimuli have been primarily used to study the room exposure phenomenon, it must be noted that the room exposure phenomenon may not be speech specific, but could be a general phenomenon related to constancy effects in hearing (Watkins and Makin, 2007). It is conceivable that high-level components of the cortical auditory system attempt to maintain the perception of room constancy: the impression that auditory objects are uncolored by the environment itself, similar in concept to color constancy in the visual system (see Foster, 2011).

Another useful analogue to the visual system is the concept of two "streams" of information: "what" and "where" pathways (Alain *et al.*, 2001). The room exposure phenomenon could be viewed as the room constancy what pathway companion to the precedence effect. Whereas the precedence effect "decolorizes" the room's effect on localization perception, the where element; the room exposure phenomenon decolorizes the environment's effect on temporal modulations to enhance the what pathway processing of object identification. Together, both processes may serve as components of a larger room constancy effect on auditory objects.

B. Potential role of interaural coherence

One additional mechanism that might help trigger the breakdown of room exposure effects is interaural coherence. It is clear that reverberation time (T_{60}) and its accompanying changes in temporal modulation properties do not completely describe the underlying perceptual quality of reverberation. Other spatial factors, such as interaural coherence, or interaural cross-correlation (IACC), also clearly contribute to perceived reverberation (Beranek, 2004). There is also known to be a significant binaural advantage for speech understanding in reverberant environments (Moncur and Dirks, 1967; Nábelek and Robinson, 1982). Additionally, from a physiological standpoint, Slama and Delgutte (2015) have argued that the reverberant coding advantage shown for neurons in the inferior colliculus of the rabbit, cannot be explained only by temporal modulation characteristics of the signal at the ear, but is likely also influenced by the ongoing fluctuations in IACC. Therefore, it is possible that changes in this binaural aspect of reverberation also contributes to the room exposure effect and may help act as a triggering mechanism for eliciting its breakdown. Indeed, in experiment 1, where IACC changed with the room manipulation, breakdown in the room exposure effect was observed. In experiment 2, where minimal changes in IACC were observed for different spatial configurations

within the same room, no breakdown was observed. Future studies will be needed to assess the independent contribution of IACC, since in this study all measures of room reverberation (e.g., T_{60} , C_{50} , and STI) were highly correlated.

C. Potential role of auditory objects

Although it is still not clear what the relationship is between the room exposure effect and auditory object formation, the auditory streaming hypothesis would predict better performance in all cases where the preceding speech is less reverberant than the target, since this would allow inherent talker cues (e.g., fundamental frequency, vocal tract length, prosody, etc.) to be clearer and more available to the listener, allowing a more coherent auditory object to be formed. It has been shown that these talker characteristics can elicit the formation of coherent auditory objects, even when spatial cues are removed (Allen *et al.*, 2008). In contrast to this idea, the data presented here demonstrated intelligibility was best when the room context was matched between carrier and target phrases, regardless of how reverberant the preceding context was compared to the target. It is still possible that these processes work in a serial fashion, however, where auditory object formation is preceded by the room exposure effect. Room exposure effects may help facilitate the buildup of auditory objects by allowing the inherent talker cues to be more easily processed in the presence of reverberation. Clearly, additional work is required to explore how these two processes may interact.

D. Limitations

While the purpose of the noise masker used in these experiments was primarily to limit ceiling-level performance and to enable measurable improvements, it likely had some influence on the perception of the room environment. Since the broadband noise also participated in the abrupt room/configuration changes, it may have contributed to the buildup and breakdown of room exposure effects. Some studies of room exposure effects have examined changes in consonant perception (Ueno *et al.*, 2005; Watkins, 2005) or speech intelligibility (Srinivasan *et al.*, 2016; Srinivasan and Zahorik, 2014) with only the speech stimulus present in the acoustic context. The results of studies using noise maskers (Brandewie and Zahorik, 2010, 2013; Srinivasan and Zahorik, 2013; Zahorik and Brandewie, 2016), including this one, have produced results both qualitatively and quantitatively similar to those without. Noisy and complex environments are quite commonplace in modern life, therefore it would not be surprising that a mechanism that compensates for the influence of the environment on perception would be robust to the influence of multiple sound sources (including noise-like sources). In addition, while this phenomenon has been exclusively studied with measurements of speech perception or intelligibility, it would be beneficial to understand if this is a general phenomenon of room exposure on subsequent perception, or if it is a characteristic unique to the domain of speech processing.

In these experiments, the speech and noise stimuli were abruptly switched into the new room environment (or spatial

configuration); the natural reverberant decay of the previous room was halted and replaced with a decay envelope matching the new environment. This manipulation would have disrupted the naturally occurring decay tails and may have influenced the results. The purpose of the manipulation was to ensure that a more reverberant carrier phrase would not inherently mask the following target phrase. The abrupt change ensured that the amount of this overlap masking would be identical for the target phrase regardless of which environment the preceding carrier was presented. The point at which the switch occurred was set to be just prior to the start of the color target word, when reverberant decay tails would be minimum. This unnatural shift in the environment, however, may limit the interpretation of these results, as the shift itself may be involved in inducing the breakdown phenomenon.

V. CONCLUSIONS

Results from this study both confirm that consistent prior listening exposure to a reverberant room can facilitate speech understanding, and more importantly, demonstrate that the facilitation can be reset by abruptly changing the temporal modulation properties of the room environment. Similar facilitation resets were not observed for spatial changes of the target/masker configuration within the same room environment that imposed minimal temporal modulation change. Overall, these results suggest that the room exposure effect for speech understanding is room-specific, but location-general.

ACKNOWLEDGMENTS

The authors wish to thank Noah Jacobs and Gina Collecchia for their help with data collection. Work supported by NIH-NIDCD Grant No. R01DC008168.

- Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. (2001). "‘What’ and ‘where’ in the human auditory system," *Proc. Natl. Acad. Sci. U.S.A.* **98**(21), 12301–12306.
- Allen, K., Carlile, S., and Alais, D. (2008). "Contributions of talker characteristics and spatial location to auditory streaming," *J. Acoust. Soc. Am.* **123**(3), 1562–1570.
- Beeston, A. V., Brown, G. J., and Watkins, A. J. (2014). "Perceptual compensation for the effects of reverberation on consonant identification: Evidence from studies with monaural stimuli," *J. Acoust. Soc. Am.* **136**(6), 3072–3084.
- Beranek, L. (2004). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York).
- Blauert, J., and Col, J.-P. (1992). "Irregularities in the precedence effect," *Adv. Biosci.* **83**, 531–538.
- 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.
- Bolt, R. H., and MacDonald, A. D. (1949). "Theory of speech masking by reverberation," *J. Acoust. Soc. Am.* **21**(6), 577–580.
- Brandewie, E., and Zahorik, P. (2010). "Prior listening in rooms improves speech intelligibility," *J. Acoust. Soc. Am.* **128**(1), 291–299.
- Brandewie, E., and Zahorik, P. (2013). "Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments," *J. Acoust. Soc. Am.* **134**(2), EL265–EL270.
- Bregman, A. S. (1978). "Auditory streaming is cumulative," *J. Exp. Psychol.* **4**(3), 380–387.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sounds*. (MIT Press, London).

- Brown, A. D., and Stecker, G. C. (2013). "The precedence effect: Fusion and lateralization measures for headphone stimuli lateralized by interaural time and level differences," *J. Acoust. Soc. Am.* **133**(5), 2883–2898.
- Clifton, R. K. (1987). "Breakdown of echo suppression in the precedence effect," *J. Acoust. Soc. Am.* **82**(5), 1834–1835.
- Clifton, R. K., and Freyman, R. L. (1997). "The precedence effect: Beyond echo suppression," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 233–255.
- Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (1994). "Listeners' expectations about echoes can raise or lower echo threshold," *J. Acoust. Soc. Am.* **95**(3), 1525–1533.
- Foster, D. H. (2011). "Color constancy," *Vision Res.* **51**(7), 674–700.
- Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (1991). "Dynamic processes in the precedence effect," *J. Acoust. Soc. Am.* **90**(2 Pt. 1), 874–884.
- Grantham, D. W. (1996). "Left–right asymmetry in the buildup of echo suppression in normal-hearing adults," *J. Acoust. Soc. Am.* **99**(2), 1118–1123.
- Haywood, N. R., and Roberts, B. (2013). "Build-up of auditory stream segregation induced by tone sequences of constant or alternating frequency and the resetting effects of single deviants," *J. Exp. Psychol.* **39**(6), 1652–1666.
- Houtgast, T., and Steeneken, H. J. M. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* **77**(3), 1069–1077.
- IEC-60268-16 (2003). "Sound system equipment—Part 16: Objective rating of speech intelligibility by speech transmission index" (International Electrotechnical Commission, Geneva, Switzerland).
- ISO-3382 (1997). "Acoustics—Measurement of room acoustic parameters" (International Organization for Standardization, Geneva, Switzerland).
- Knudsen, V. O. (1929). "The hearing of speech in auditoriums," *J. Acoust. Soc. Am.* **1**, 56–82.
- Kuwada, S., Bishop, B. B., and Kim, D. O. (2012). "Approaches to the study of neural coding of sound source location and sound envelope in real environments," *Front. Neural Circuits* **6**, 42.
- Kuwada, S., Bishop, B., and Kim, D. O. (2014). "Azimuth and envelope coding in the inferior colliculus of the unanesthetized rabbit: Effect of reverberation and distance," *J. Neurophysiol.* **112**(6), 1340–1355.
- McCall, D. D., Freyman, R. L., and Clifton, R. K. (1998). "Sudden changes in spectrum of an echo cause a breakdown of the precedence effect," *Percept. Psychophys.* **60**(4), 593–601.
- Moncur, J. P., and Dirks, D. (1967). "Binaural and monaural speech intelligibility in reverberation," *J. Speech Lang. Hear. Res.* **10**(2), 186–195.
- Nábělek, A. K., Letowski, T. R., and Tucker, F. M. (1989). "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.* **86**(4), 1259–1265.
- Nábělek, A. K., and Robinson, P. K. (1982). "Monaural and binaural speech perception in reverberation for listeners of various ages," *J. Acoust. Soc. Am.* **71**(5), 1242–1248.
- 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**, 200–211.
- Roberts, B., Glasberg, B. R., and Moore, B. C. J. (2008). "Effects of the build-up and resetting of auditory stream segregation on temporal discrimination," *J. Exp. Psychol.* **34**(4), 992–1006.
- Rogers, W. L., and Bregman, A. S. (1993). "An experimental evaluation of three theories of auditory stream segregation," *Percept. Psychophys.* **53**(2), 179–189.
- Rogers, W. L., and Bregman, A. S. (1998). "Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness," *Percept. Psychophys.* **60**(7), 1216–1227.
- Schroeder, M. R. (1981). "Modulation transfer functions: Definition and measurement," *Acta Acust. Acust.* **49**(3), 179–182.
- Slama, M. C. C., and Delgutte, B. (2015). "Neural coding of sound envelope in reverberant environments," *J. Neurosci.* **35**(10), 4452–4468.
- Srinivasan, N. K., Tobey, E. A., and Loizou, P. C. (2016). "Prior exposure to a reverberant listening environment improves speech intelligibility in adult cochlear implant listeners," *Cochlear Implants Int.* **17**(2), 1–7.
- Srinivasan, N. K., and Zahorik, P. (2013). "Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences," *J. Acoust. Soc. Am.* **133**(1), EL33–EL39.
- Srinivasan, N. K., and Zahorik, P. (2014). "Enhancement of speech intelligibility in reverberant rooms: Role of amplitude envelope and temporal fine structure," *J. Acoust. Soc. Am.* **135**(6), EL239–EL245.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Ueno, K., Kopčo, N., and Shinn-Cunningham, B. (2005). "Calibration of speech perception to room reverberation," presented at the *Forum Acusticum*, Budapest.
- Wallach, H., Newman, E. B., and Rosenweig, M. R. (1949). "The precedence effect in sound localization," *Am. J. Psychol.* **62**(3), 315–336.
- Watkins, A. J. (2005). "Perceptual compensation for effects of reverberation in speech identification," *J. Acoust. Soc. Am.* **118**(1), 249–265.
- Watkins, A. J., and Makin, S. J. (2007). "Perceptual compensation for reverberation in speech identification: Effects of single-band, multiple-band and wideband noise contexts," *Acta Acust. Acust.* **93**(3), 403–410.
- Watkins, A. J., Raimond, A. P., and Makin, S. J. (2011). "Temporal-envelope constancy of speech in rooms and the perceptual weighting of frequency bands," *J. Acoust. Soc. Am.* **130**(5), 2777–2788.
- Wichmann, F. A., and Hill, N. J. (2001). "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.* **63**(8), 1293–1313.
- Zahorik, P. (2009). "Perceptually relevant parameters for virtual listening simulation of small room acoustics," *J. Acoust. Soc. Am.* **126**(2), 776–791.
- Zahorik, P., and Brandewie, E. J. (2016). "Speech intelligibility in rooms: Effect of prior listening exposure interacts with room acoustics," *J. Acoust. Soc. Am.* **140**(1), 74–86.