J Am Acad Audiol 29:95–105 (2018)

Effects of Early- and Late-Arriving Room Reflections on the Speech-Evoked Auditory Brainstem Response

DOI: 10.3766/jaaa.16017

Rida Al Osman*
Christian Giguère*
Hilmi R. Dajani†

Abstract

Background: Room reverberation alters the acoustical properties of the speech signals reaching our ears, affecting speech understanding. Therefore, it is important to understand the consequences of reverberation on auditory processing. In perceptual studies, the direct sound and early reflections of reverberated speech have been found to constitute useful energy, whereas the late reflections constitute detrimental energy.

Purpose: This study investigated how various components (direct sound versus early reflections versus late reflections) of the reverberated speech are encoded in the auditory system using the speech-evoked auditory brainstem response (ABR).

Research Design: Speech-evoked ABRs were recorded using reverberant stimuli created as a result of the convolution between an ongoing synthetic vowel /a/ and each of the following room impulse response (RIR) components: direct sound, early reflections, late reflections, and full reverberation. Four stimuli were produced: direct component, early component, late component, and full component.

Study Sample: Twelve participants with normal hearing participated in this study.

Data Collection and Analysis: Waves V and A amplitudes and latencies as well as envelope-following response (EFR) and fine structure frequency–following response (FFR) amplitudes of the speech-evoked ABR were evaluated separately with one-way repeated measures analysis of variances to determine the effect of stimulus. Post hoc comparisons using Tukey's honestly significant difference test were performed to assess significant differences between pairs of stimulus conditions.

Results: For waves V and A amplitudes, a significant difference or trend toward significance was found between direct and late components, between direct and full components, and between early and late components. For waves V and A latencies, significant differences were found between direct and late components, between direct and full components, between early and late components, and between early and full components. For the EFR and FFR amplitudes, a significant difference or trend toward significance was found between direct and late components, and between early and late components. Moreover, eight, three, and one participant reported the early, full, and late stimuli, respectively, to be the most perceptually similar to the direct stimulus.

Conclusions: The stimuli that are acoustically most similar (direct and early) result in electrophysiological responses that are not significantly different, whereas the stimuli that are acoustically most different (direct and late, early and late) result in responses that are significantly different across all response measures. These findings provide insights toward the understanding of the effects of the different components of the RIRs on auditory processing of speech.

Key Words: auditory stressors, early reflections, full reflections, late reflections, reverberation, speech-evoked ABR

Abbreviations: ABR = auditory brainstem response; ANOVA = analysis of variance; EFR = envelope-following response; FFR = fine structure frequency–following response; FFT = fast Fourier transform; RIR = room impulse response; RMS = root-mean-square; RT = reverberation time; SIN = speech in noise

---

*School of Rehabilitation Sciences, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada; †School of Electrical Engineering and Computer Science, Faculty of Engineering, University of Ottawa, Ottawa, ON, Canada

Corresponding author: Rida Al Osman, School of Rehabilitation Sciences, Faculty of Health Sciences, University of Ottawa, Ottawa, ON K1H 8M5, Canada; E-mail: ralos010@uottawa.ca

The authors alone are responsible for the content and writing of this article.

INTRODUCTION

Speech communication is essential in human daily interaction, but it is often hindered by perceptual stressors such as background noise (Pichora-Fuller, 2003; Wagener and Brand, 2005), fast speech rate (Tun, 1998; Adams et al, 2012), reverberation (Neuman et al, 2010; Wróblewski et al, 2012), and transmission channel distortion (Houtgast and Steeneken, 1985; Assmann and Summerfield, 2004). In particular, room reverberation alters the acoustical properties of the speech signals reaching our ears in situations of face-to-face communication, which can affect speech understanding (Crandell and Smaldino, 2000; Bradley et al, 2003). Therefore, it is important to understand the consequences of reverberation on auditory processing of speech.

Reverberation is defined as the sum of all sound reflections off the walls and objects in an enclosed space (Blauert, 1997). The reflections arriving within the first 50 msec after the direct sound are often called early reflections and so are distinguished from late reflections (Bradley et al, 2003; Arweiler and Buchholz, 2011). The series of direct sound, early reflections, and late reflections are typically characterized by the room impulse response (RIR) (Blauert, 1997). In addition, the reverberation time (RT), defined as the time required for the sound level to decay by 60 dB after it has been turned off, is often used as a simple parameter to characterize room reverberation (Crandell and Smaldino, 2000). However, the overall effects of reverberation on speech intelligibility depend on many other factors, including the dimension of the room, the source–receiver distance, and the relative orientation of talkers and listeners (Blauert, 1997; Bradley et al, 2003). Boothroyd (2008) and Giguère (2013) discussed several of these factors in more detail in the context of speech understanding in children with hearing impairment.

From an acoustical standpoint, reverberation can cause degradation of speech signals, such as temporal smearing, filling dips and gaps in the temporal envelope, increasing the prominence of low-frequency energy, and flattening formant transitions (Assmann and Summerfield, 2004; Sayles and Winter, 2008). Such degradation of speech signals mainly results from masking of a speech segment by reflections of the segment itself, known as self-masking, and by reflections of previous segments, known as overlap-masking (Devore and Shinn-Cunningham, 2003). On the other hand, reverberation amplifies the speech signal level. Such amplification can assist with speech perception, particularly in instances where the direct sound is reduced (Bradley et al, 2003).

Behavioral studies investigating the effects of reverberation on speech perception reported degradation in identifying and discriminating consonantal features (Gelfand and Silman, 1979),vowels (Nábělek and Letowski, 1988), and poorer performance in sentence intelligibility tests (Hazrati and Loizou, 2012; Wróblewski et al, 2012). Various methods are used in such studies for creating noisy reverberant stimuli. Often, the RIRs of the environments to be simulated are convolved with sentence stimuli (recorded in anechoic conditions) to create the reverberant speech material, which is then normalized to equate with the nonreverberated stimuli. Noise is then added to the reverberated speech signals at fixed signal-to-noise ratios. This process negates some positive effects of reverberation, such as signal amplification due to the early reflections. Also, room reverberation is sometimes considered as a whole, whereas the various components of the RIR can produce different effects on speech perception. The lack of a common methodology, in the aforementioned studies, to create reverberant materials can result in various outcomes with respect to the effects of reverberation on speech perception

As noted earlier, the RIR is composed of various components: direct sound, early reflections, and late reflections. Arweiler and Buchholz (2011), Yang and Bradley (2009), and Bradley et al (2003) studied the components of the RIR and established the importance of early reflections using sentence and word intelligibility test scores, in conditions similar to real rooms for both hearing-impaired and nonimpaired participants. Results from their studies show that for situations where the direct sound is reduced, it is possible to understand the speech signal due to the presence of the early reflections. Specifically, Bradley et al (2003) showed that for young adult listeners, the added early reflections have the same effect as an increase in the level of the direct sound, leading to an increase of the signal-to-noise ratio up to 9 dB in some situations where the direct sound pathway is compromised. These studies show that the early reflection components of the RIR provide useful energy to speech perception in reverberant listening conditions. Furthermore, Bradley et al (2003) indicated that minimizing the RT by increasing room absorption can sometimes decrease speech intelligibility; instead, they proposed that RTs should be optimized to allow for better speech intelligibility. Indeed, standards such as ANSI S12.60-Part 1 (ANSI, 2010) and CSA Z412-00 (CSA, 2000) define optimal reverberation characteristics for proper acoustical design of classrooms and offices, respectively.

From an auditory processing standpoint, speech stimuli evoke both transient and steady-state brainstem responses. The transient response refers to the initial part of the auditory brainstem response (ABR) (typically <20 msec), reflecting the transmission delay between the ear and the rostral brainstem structures, in response to any type of stimulus (speech or nonspeech)

(Skoe and Kraus, 2010). The transient response has been used by clinicians and researchers as an objective tool for assessing the neural activity of the ascending auditory pathway between the cochlear nerve and the inferior colliculus in midbrain. The transient response is typically analyzed in the time domain, which involves measuring amplitudes and latencies of various peaks (most notably V and A) in the waveform. Previous studies (e.g., Laroche et al, 2013; Prévost et al, 2013) using a synthesized speech vowel (/a/) have demonstrated that background noise has an effect on waves V and A by increasing the latency and decreasing the amplitude. The steady-state response, on the other hand, is generated in response to a periodic speech stimulus, and provides information about the neural encoding of the periodic components of the stimulus. The steady-state response follows the speech stimulus acoustical features, including fundamental frequency and low-frequency formants (Skoe and Kraus, 2010). The steady-state response is typically analyzed in the frequency domain and can be categorized into the envelope-following response (EFR), representing the neural response that follows the speech envelope, and the fine structure frequency–following response (FFR), representing the neural response that follows the speech formants. The EFR characterizes the evoked response at F0 and its lower harmonics, while the FFR characterizes the evoked response at higher harmonics, typically in the region of the first formant, F1 (Aiken and Picton, 2008; Laroche et al, 2013).

Timing measures of the speech-evoked ABR (e.g., peak latency) provide information about the accuracy with which the brainstem nuclei synchronously respond to acoustic stimulation. A change on the order of a fraction of 1 msec in the transient speech-evoked ABR peak latency has been associated with difficulties in perceiving speech in noise (SIN) (Wible et al, 2004; Banai et al, 2009; Hornickel et al, 2009), and can be of clinical significance (Wible et al, 2004; Chandrasekaran and Kraus, 2010). Spectral magnitude measures of the steady-state speech-evoked ABR (e.g., root-mean-square amplitude) provide information about the robustness with which the brainstem nuclei respond to acoustic stimulation. A change of magnitude variability of $<0.01\ \mu V$ and $<0.001\ \mu V$ in the EFR and FFR amplitudes, respectively, has been associated with difficulties in perceiving SIN (Song et al, 2012; Anderson et al, 2013) and reverberation (Bidelman and Krishnan, 2010).

Given the impact of reverberation on real-world listening, and the increasing use of evoked responses to study auditory processing of more naturalistic signals, it is worthwhile to investigate the effects of reverberation on these responses. However, only a few researchers have investigated the effects of reverberation on speech stimuli using electrophysiological methods. Unlike behavioral studies where overlap-masking and self-masking effects of reverberation can be easily explored using multisyllabic words or long-duration sentences, studying reverberation electrophysiologically poses challenges due to the restriction of short stimulus duration. Bidelman and Krishnan (2010) produced reverberant stimuli by time domain convolution of a 250-msec synthetic vowel /i/ with RIRs recorded in a corridor at a distance of 0.63 m (RT = 0.7 sec), 1.25 m (RT = 0.8 sec), and 5 m (RT = 0.9 sec) from the sound source. Results indicated that the neural encoding of F0 in speech-evoked ABR was affected only under the strongest level of reverberation, whereas F1 encoding was degraded in all reverberation conditions, indicating a differential impact of reverberation on pitch- and formant-related information. Fujihira and Shiraishi (2015) produced reverberant stimuli by time convolution of a 170-msec synthetic /da/ with three RIRs (RT = 0.5, 1.0, and 1.5 sec) chosen from the environment/architectural acoustics “SMILE 2004” database (Architectural Institute of Japan, 2004). The authors reported significant correlation between word intelligibility scores and the spectral components of the FFR response near F1 under the effects of reverberation for elderly adults, but not between word recognition and the spectral components of the EFR response, suggesting that the intelligibility of reverberant speech is related to their ability to encode the temporal fine structure of speech.

It is important to note that in Bidelman and Krishnan (2010) and Fujihira and Shiraishi (2015), the methodological choices and/or study design promoted the detrimental contribution of reverberation on speech processing. In Bidelman and Krishnan (2010), the amplitude of the reverberant speech stimulus was normalized to the same overall level as the nonreverberated condition, thus negating the natural amplification that typically occurs in reverberant conditions. In Fujihira and Shiraishi (2015), the early reflection components from the RIR were removed, thus eliminating the benefit they may have had on speech perception. These studies were not designed to reveal the separate effects of early and late reflections on the speech-evoked EFR and FFR.

The purpose of the current study is to shed light on how various components (direct sound versus early reflections versus late reflections versus full RIR) of the reverberated speech are encoded in the auditory system using an electrophysiological test. Differences among these components are established with respect to measures of latency of waves V and A, amplitude of waves V and A, and amplitude of EFR and FFR. No other study, to our knowledge, has made this comparison using the speech-evoked ABR. Moreover, the effects of reverberation on waves V and A have not been previously reported.

Given the beneficial perceptual effects of early reflections and the detrimental effects of late reflections reported above, we hypothesized that the speech-evoked

ABR to the direct and early stimuli would be similar, whereas the responses to direct or early stimuli would be different from responses to late and full stimuli. More specifically, given the effects of late reverberation on speech signals, such as temporal smearing of the speech envelope, the speech-evoked ABR to the late and full reverberant stimuli would have longer latencies and reduced amplitudes for waves V and A, and reduced EFR and FFR amplitudes, compared to the response to the direct stimulus.

METHODS

Participants

Twelve participants (two males and ten females) participated in the study. The age of participants ranged from 21 to 31 yr (mean = 24). All participants had normal hearing thresholds, defined as ≤15 dB HL bilaterally at 0.25, 0.5, 0.75, 1, 2, and 4 kHz, and none of the participants reported a history of hearing- or ear-related issues. All participants provided their informed consent in compliance with a protocol approved by the University of Ottawa Research Ethics Board.

RIR

The overall effects of reverberation on speech intelligibility depend on many factors including the dimension of the room, the source–receiver distance, and the relative orientation of talkers and listeners (Blauert, 1997; Bradley et al, 2003). In our study, we designed a scenario where we presented no particular advantage for early or late reflections. Our design characterized a situation where the source and listener positions reflected an equal ratio of early and late reflections. This is in contrast to previous studies that promoted detrimental effects of reverberation.

The RIR was computed following the algorithm from McGovern (2003), which is based on the image method from Allen and Berkley (1979). This method consists of modeling reflections as if they originated from virtual locations behind room boundaries, and the contributions from all the virtual sources are summed. Each virtual source contributes a delayed impulse (echo), whose time delay is equal to the distance between the source and the listener divided by the speed of sound. The echo amplitude is inversely proportional to the distance traveled and proportional to the room reflection coefficient to the power of the number of reflections (Gardner, 2002).The algorithm of McGovern (2003) assumes that the room model is rectangular, making it easy to simulate using a computer program. Furthermore, it assumes that the room reflection coefficient is uniform throughout the room. The algorithm takes into account the number of virtual sources, the position of the virtual sources, position of the listener, delay and magnitude of each reflection, room dimensions (H × L × W), and average room reflection.

The RIR in this study was computed for a room of dimensions 3.5 × 11 × 14 m and for source and listener coordinate positions of (2 m, 2.5 m, 1.5 m) and (2 m, 8 m, 1.2 m), respectively. The source–listener distance (5.5 m) was selected to reflect an equal ratio of early and late energy. The early energy value was calculated as the sum of the amplitude components in RIR up to 50 msec following (but not including) the direct sound. Similarly, the late energy was calculated as the sum of the amplitude components in RIR following the early reflections (Figure 1).

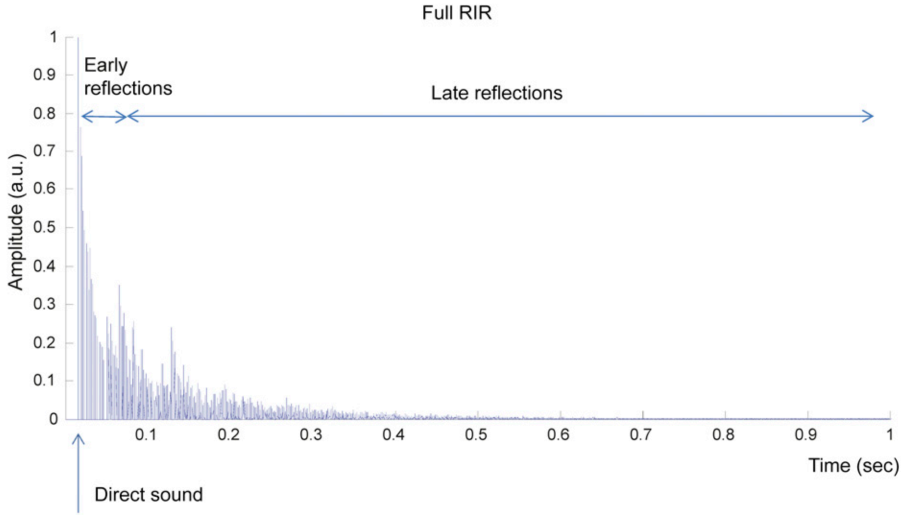

Figure 1. Representation of the full RIR showing the direct sound, early reflections, and late reflections. The RIR was produced for a room of dimensions 3.5 × 11 × 14 m and for source and listener coordinate positions of (2 m, 2.5 m, 1.5 m) and (2 m, 8 m, 1.2 m), respectively. The RIR waveform was rectified. a.u. = arbitrary units. (This figure appears in color in the online version of this article.)

Bradley (1986) indicated that the level of reverberation could vary in typical classrooms from 0.4 to 1.2 sec. The RT in this study was chosen as 1 sec, simulating a classroom near the upper end of this range. The average room reflection was calculated from the desired RT based on the Eyring formula (Irwin and Graf, 1979).

Stimuli

A 2-sec synthetic vowel /a/ was generated using formant synthesis (F0 = 0.1 kHz, F1 = 0.7 kHz, F2 = 1.22 kHz, F3 = 2.6 kHz) based on a simplified version of the Klatt synthesizer (Klatt, 1980; Laroche et al, 2013). The reverberant stimuli were created as a result of the convolution between the 2-sec synthetic vowel /a/ and each of the RIR components: direct sound, early reflections and late reflections, and the full RIR (Figure 1), as follows:

1. Direct component = vowel/a/*RIR direct sound
2. Early reflections = vowel/a/*RIR early reflections only (i.e., reflections within 50 msec following the direct sound)
3. Late reflections = vowel/a/*RIR late reflections only (i.e., reflections following the early reflections)
4. Full RIR = vowel/a/*full RIR (i.e., direct sound + early reflections + late reflections)

where the symbol * denotes convolution.

The reverberant stimuli were chopped between 1,200 msec and 1,500 msec to include self-masking from the initial portion of the vowel, thus modeling a situation with an ongoing speech stimulus. We selected 1,200 msec to ensure the room response had stabilized, given the 1-sec RT. The end result was four stimuli (direct, early, full, and late), each with a 300-msec duration generated at a sampling frequency of 48 kHz and with 16-bit resolution. The direct stimulus was presented at 76.9 dB SPL, as measured in an IEC 60318-4 ear simulator (IEC, 2010) (Product RA0045, G.R.A.S. Sound and Vibration A/S, Holte, Denmark). The levels of the reverberant stimuli (early at 79.4 dB SPL, full at 79.2 dB SPL, and late at 79.7 dB SPL) were not normalized to be consistent with real room environments. The different time-domain stimulus waveforms and amplitude spectra are shown in Figure 2. The spectral components of full were not always higher than those of direct, early, and late. We suspect that this is due to phase effects, which also affect the time-domain waveforms.

Recording and Analysis of Speech-Evoked ABR

Recordings of the speech-evoked ABR were performed in a shielded audiometric room at the University of Ottawa, in dim lighting. Responses were recorded with a vertical montage using BioMARK™ v.7.0.2 system (Biologic Marker of Auditory Processing, Bio-logic Systems Corp., Mundelein, IL) from the vertex (Cz), with the right earlobe as reference and the left earlobe as ground. Responses were recorded over 319.8-msec epochs (1,024 points/epoch corresponding to a sampling frequency of ~3202 Hz) and were amplified and filtered using an amplifier with a gain of 10,000 and a filter bandwidth extending from 30 to 1000 Hz. All electrode impedances were below 5 kOhm at 10 Hz.

To minimize artifacts, the participants were instructed to remain relaxed and to avoid abrupt movements while

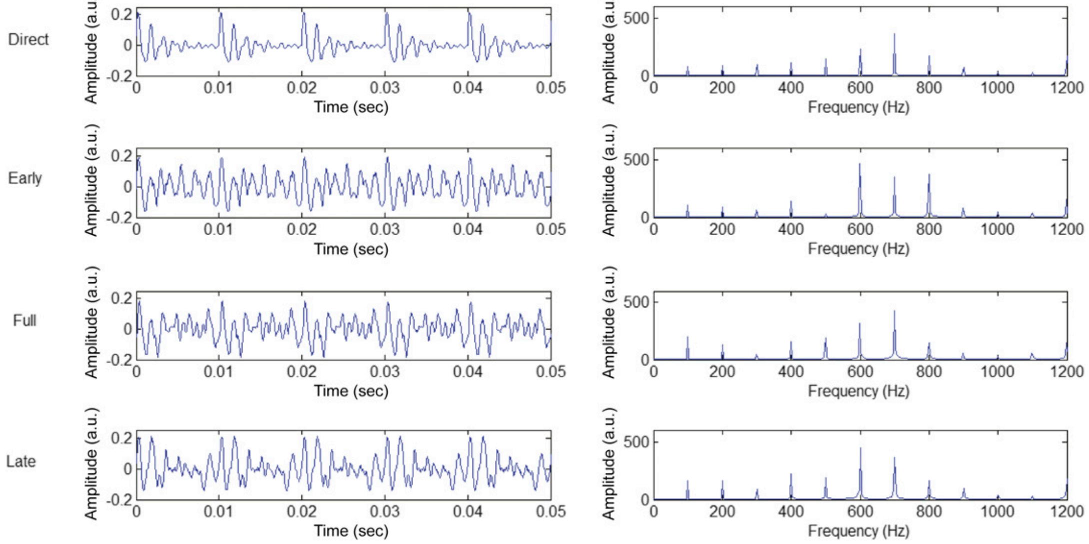

Figure 2. Time-domain waveforms and amplitude spectra of the four stimuli (direct, early, full, and late). Only the first 50 msec is shown. a.u. = arbitrary units. (This figure appears in color in the online version of this article.)

sitting in a comfortable reclining chair and watching a muted movie with subtitles. None of the participants reported to have fallen asleep. Also, recorded sweeps in which the response exceeded 23.8 μV were discarded. In addition, we followed the recommendations of the manufacturer of BioMARK™ regarding environmental electromagnetic noise reduction, such as turning off fluorescent lights when operating the equipment and making sure all the devices are connected to the isotransformer provided with the system.

The stimuli (direct, early, late, and full) were delivered using the Bio-logic insert-earphone of the BioMARK™ system, monaurally in the right ear and in alternating polarity, and were presented in a pseudorandomized order. For each of the stimuli, two sets of 1,500 ABR responses were recorded consecutively. The 1,500 artifact-free responses were added and divided by the total number of responses to minimize the influence of the cochlear microphonic and stimulus artifacts on the response (Aiken and Picton, 2008; Campbell et al, 2012), and to obtain response components related to the stimulus envelope. We term the responses calculated with the addition method as the envelope following responses (EFR). Similarly, the 1,500 artifact-free responses to the two polarities were subtracted from each other and divided by the total number of responses to obtain response components that follow the stimulus temporal fine structure (Aiken and Picton, 2008; Anderson et al, 2010). We term the responses calculated with the subtraction method as the FFRs. To test the absence of electromagnetic leakage, the electrodes were placed on one participant's scalp, and in all other aspects the experiments proceeded as normal, with the sole exception that instead of placing the insert earphone in the participant's right ear, it was inserted into the IEC 60318-4 ear simulator (IEC, 2010), which presents approximately the same acoustic load to the transducer as if the earphones were inserted in the ear. Spectral analysis of the recorded signal coherently averaged over 3,000 trials showed no significant components in the steady-state response, indicating no contamination of the recorded responses from the sound-generating equipment to the electrodes (Al Osman et al, 2016).

We analyzed both the transient and the steady-state components of the responses to the speech stimuli. The steady-state response follows the transient response and was taken to start at 19.8 msec for a duration of 300 msec. Figure 3 shows the time waveform and spectrum of the grand average of the EFR with the four stimuli (direct, early, late, and full).

For the transient response, the amplitude and latency of waves V and A were visually extracted from the time-domain EFR waveforms. The prominent onset wave "V" in the response starts 6–10 msec following the stimulus, reflecting the time delay to the upper auditory brainstem, and is followed by the negative wave "A" (Chandrasekaran and Kraus, 2010).

For the steady-state response, spectral analysis was performed using the Fourier transform in Matlab v.7.9 (MathWorks, Natick, MA) to extract the amplitudes of different harmonics from the EFR and FFR. The EFR amplitude was computed from the fundamental and harmonics below the first formant in the EFR spectrum, as the square root of the combined power of F0, 2F0, and 3F0. The FFR amplitude was computed from the harmonics around the first formant F1 in the FFR spectrum, as the square root of the combined power of 5F0 to 9F0. The harmonic amplitudes correspond to the height of the peak in the EFR and FFR root-mean-square amplitude spectrum (in μV).

Waves V and A amplitudes and latencies as well as EFR and FFR amplitudes were evaluated separately with one-way repeated measures analyses of variance

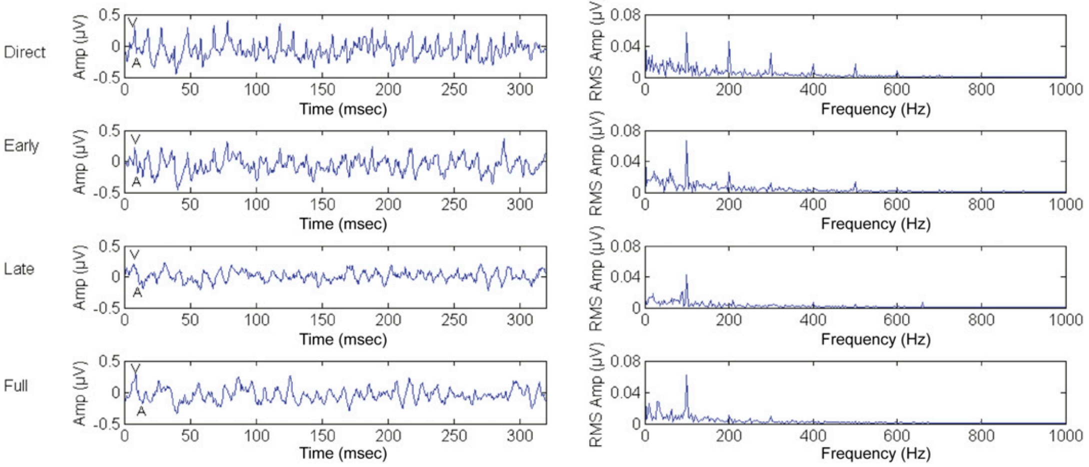

Figure 3. Grand average over all participants of the EFR of the speech-evoked ABR in time domain and frequency domain with the four stimuli (direct, early, late, and full). Amp = amplitude. (This figure appears in color in the online version of this article.)

(ANOVAs), using SPSS version 18.0 (SPSS Inc., Chicago, IL) to determine the effect of stimulus. Post hoc comparisons using Tukey's honestly significant difference test were performed to assess significant differences between pairs of stimulus conditions.

Behavioral Test

The participants were instructed to listen, monaurally in the right ear via a shielded earphone and at the same levels as in the electrophysiological experiment, to the direct sound and subsequently listen to the early, late, and full stimuli. They were then asked to select, on a computer screen, their choice of sound (early, late, or full) that most resembles the direct sound. The direct sound was presented first, and then the other three sounds were presented in pseudorandomized order. This test was repeated three times.

RESULTS

Transient Response

The waves V and A amplitudes and latencies across all stimuli (direct, early, full, and late) are shown in Figure 4. The ANOVA results show that there is a main effect of stimulus with all of these four measures: V amplitude ($p < 0.05$), A amplitude ($p < 0.05$), V latency ($p < 0.001$), and A latency ($p < 0.001$). Post hoc tests were used to determine the pairwise differences among the different stimulus conditions. For transient response amplitudes, a significant difference or trend toward significance was found between direct and late (V amplitude [$p < 0.05$], A amplitude [$p < 0.05$]), between direct and full (V amplitude [$p < 0.05$], A amplitude [$p = 0.85$]), and between early and late (V amplitude [$p < 0.05$], A amplitude [$p < 0.05$]), among the possible six pairwise stimulus comparisons. For transient response latencies, significant differences were found between direct and late (V latency [$p < 0.01$], A latency [$p < 0.001$]), between direct and full (V latency [$p < 0.01$], A latency [$p < 0.01$]), between early and late (V latency [$p < 0.01$], A latency [$p < 0.01$]), and between early and full (V latency [$p < 0.05$], A latency [$p < 0.01$], among the possible six pairwise stimulus comparisons.

Steady-State Response

The EFR and FFR amplitudes across all stimuli (direct, early, full, and late components) are shown in Figure 5. The ANOVA results show that there is a main effect of stimulus with these two measures: EFR ($p < 0.05$), FFR ($p < 0.05$). Post hoc tests showed that there was a significant difference or trend toward significance only between direct and late (EFR [$p = 0.05$], FFR [$p < 0.05$]), and between early and late (EFR [$p < 0.05$], FFR [$p = 0.05$]), among the possible six pairwise stimulus comparisons.

Behavioral Response

Eight of the twelve participants reported the early reflection stimulus as perceptually the most similar to the direct sound (six participants in all three trials and two in two of three trials). Three participants reported the full stimulus (two in all three trials and one in two of three trials), and only one reported the late reflection stimulus (two of three trials) as the most similar to the direct stimulus.

DISCUSSION

In this study, we explored the effects of the different components of the RIR, namely the direct sound, the early and late reflections, and the full RIR on auditory processing of speech using speech-evoked ABR and a behavioral test. Our study modeled a situation where the source and listener positions reflected an equal ratio of early and late energy. The 300-msec stimuli were extracted from the convolution of a long vowel with the different components of the RIR or the full RIR to

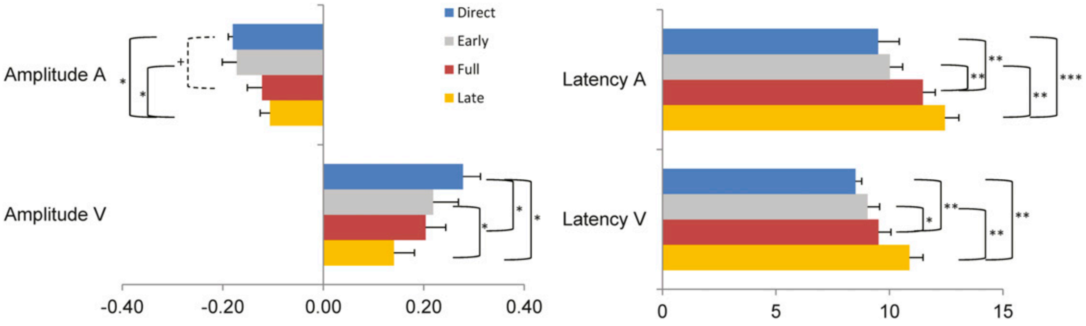

Figure 4. Average of individual latencies (msec) and amplitudes (μV) of waves A and V for each stimulus (direct, early, full, and late). Error bars indicate standard errors of the mean. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$, +$p < 0.1$. (This figure appears in color in the online version of this article.)

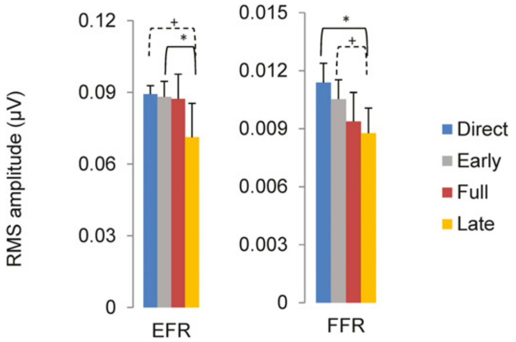

Figure 5. Average of EFR amplitudes (combination of the amplitudes at F0, 2F0, and 3F0 in the EFR spectrum) across participants, and of FFR amplitudes (combination of the amplitudes at 5F0 to 9F0 in the FFR spectrum) across participants, for each stimulus (direct, early, full, and late). Error bars indicate standard errors of the mean. $^{*}p < 0.05$, $+p < 0.1$. (This figure appears in color in the online version of this article.)

model the effects of reverberation on an ongoing speech signal.

Transient and steady-state speech-evoked ABR were recorded. Our study revealed four key findings (Figures 4 and 5):

1. Responses to direct and early components are not significantly different or trending toward significance.
2. Responses to direct/early and late components are all significantly different or trending toward significance.
3. Responses to full and late components are not significantly different or trending toward significance.
4. Responses to direct/early and full components present a mixed picture with most transient response measures, showing a significant effect or trend toward significance, whereas steady-state responses do not.

Regarding finding (1), waves V and A amplitudes and latencies, and EFR and FFR amplitudes, all showed no statistically significant difference nor a trend toward significance between direct and early reflection components. Furthermore, the results from the behavioral test corroborated these findings. Eight participants reported that the early stimulus is the most perceptually similar to the direct sound. These electrophysiological and perceptual results are consistent with findings from Bradley et al (2003) and Roman and Woodruff (2013) who investigated the effects of reverberation on the intelligibility of speech with normal hearing adults. The authors reported that early reflections occurring within the first 50 msec are beneficial to speech perception and can be integrated by the auditory system with the direct sound.

Regarding finding (2), all transient and steady-state response measures showed statistically significant differences between direct and late reflection components, with the exception of the EFR amplitude, which showed a nearly statistical significance ($p = 0.05$). This corroborates the results from the behavioral test where only one participant reported that the late stimulus was the most perceptually similar to the direct sound. Again, these results are also consistent with findings from Bradley et al (2003) and Roman and Woodruff (2013), who indicated that late reflections have a detrimental effect on speech perception and cannot be integrated with the direct sound.

All transient and steady-state response measures also showed a statistically significant difference between early and late reflection components, despite their equal energy, with the exception of FFR amplitude, which showed a nearly statistical significance ($p = 0.05$). These findings are compatible with those from Bradley et al (2003), Boothroyd (2008), Yang and Bradley (2009), and Roman and Woodruff (2013), which indicated that early reflections possess useful energy while late reflections possess detrimental energy. This makes it likely that electrophysiological responses to these two types of reflections would also differ.

Regarding finding (3), all transient and steady-state response measures showed neither a statistically significant difference nor a trend toward significance between full and late components. Furthermore, for finding (4), waves V and A latencies and wave V amplitude showed a statistically significant difference between direct and full components, while wave A amplitude and EFR and FFR amplitudes showed no difference. Likewise, waves V and A latencies showed a difference between early and full components, whereas waves V and A amplitudes and EFR and FFR amplitudes showed no difference. Still, the full stimulus always produced a response in between that of the direct/early and late component stimuli for all ABR measures, as shown in Figures 4 and 5. This intermediate result for the full response is not surprising, given that there were equal amounts of early and late reflection energy in the full stimulus in this study. It can be hypothesized that using higher or lower ratios of early to late reflection energy would steer the full response closer to that of the early or late responses respectively in all ABR measures.

It should be noted that as a result of the convolution process and choice of room reverberation parameters in this study, the direct stimulus was presented at a lower level (76.9 dB SPL) than the reverberant stimuli, early (79.4 dB SPL), full (79.2 dB SPL), and late (79.7 dB SPL); however, it resulted in responses that were higher in amplitude and shorter in latency than responses to the full and late reflection stimuli and not significantly different from responses to the early reflection stimulus. This result can be attributed to the smearing effects of reverberation that blur the speech signal, resulting in a degraded ABR. The smearing effects in the reverberant stimuli (early, full, and late) are shown in Figure 2, where the direct stimulus shows the strongest amplitude modulation compared to the reverberant stimuli (early, full, and late).

Our results can be contrasted with the findings of Bidelman and Krishnan (2010). These authors did not decompose the RIR into direct, early, full, and late but rather compared direct (or anechoic) with full RIR at various RTs (0.6, 0.7, and 0.8 sec). They indicated that the neural encoding of F1 was more affected by reverberation than the encoding of F0 in speech-evoked ABR, suggesting a differential impact on the source-filter components of speech. We also found that the size EFR amplitude at F0 was hardly affected by the full RIR reverberant condition compared to the direct component. Furthermore, the FFR amplitude at F1 decreased in size in the full RIR condition compared to direct component (Figure 5), as in Bidelman and Krishnan (2010); however, post hoc tests showed that this was not statistically significant. In our study, we modeled a situation where the source and listener positions reflected an equal ratio of early and late energy, whereas Bidelman and Krishnan (2010) modeled situations with increasing distances (up to 5 m for RT = 0.8 sec). It is understood that at larger distances, the energy of late reflections becomes increasingly higher than the energy of the early reflections. Therefore, methodological differences in the ratio of late to early reflections between the two studies likely influenced the effects of the full response on the steady-state speech-evoked ABR at F1. Previous studies on the effects of reverberation (Bidelman and Krishnan, 2010; Fujihira and Shiraishi, 2015) did not measure the transient ABR. We found waves A and V latencies and amplitudes to be very sensitive to the effects of reverberation, with all measures showing a significant difference or trend between direct and full RIR (Figure 4), possibly due to differences in the stimulus envelope (Figure 2).

A correspondence between speech-evoked ABR and SIN perception has been reported in several studies. For instance, Kraus and her colleagues recorded speech-evoked ABR in six-talker babble noise using the syllable /da/, and measured SIN perception using the hearing-in-noise test, with various age groups: children (Anderson et al, 2010), young adults (Anderson and Kraus, 2010; Song et al, 2011), and older adults (Anderson and Kraus, 2012). The authors reported that all participants with poor SIN perception had deficits in the subcortical spectral and temporal representation, including low-frequency spectral magnitudes and the timing of transient response peaks, compared to participants with good SIN perception. Furthermore, Hornickel et al (2009) evaluated the relationship between the subcortical representation of stop-consonant timing and SIN perception. They recorded speech-evoked ABR to the syllables /ba/, /da/, and /ga/ in children with a wide range of reading abilities. They reported that subcortical differentiation of responses to these three speech stimuli was found to be correlated with reading ability and SIN perception, with better performers showing greater differences among brainstem responses to the three syllables. In addition, a correspondence between speech-evoked ABR and word intelligibility test performance was found in the context of reverberated stimuli (Fujihira and Shiraishi, 2015).

Overall, our study provides insights toward understanding the effects of the different components of the RIR on auditory processing of speech. The results show that stimuli that are acoustically most similar (direct and early) result in electrophysiological responses that are not significantly different. However, stimuli that are acoustically most different (direct and late, early and late) result in responses that are significantly different across all response measures. Meanwhile, none of the response measures with the full stimulus were significantly different from those with the late stimulus, while some of the response measures with this stimulus were different from those obtained with direct and early stimuli. In part, this may be explained by the equal proportion of early and late energy in the reverberation scenario under study. Moreover, complex interference patterns among the different components of the full RIR could have boosted or reduced certain frequency regions of the speech signals, which in turn could have amplified or attenuated the neural responses in the auditory system. To capture the essence of full versus direct/early/late comparisons, various source-receiver configurations need to be established, simulating short and long distances, and employing various RTs. The method and tools used in this study could be helpful when designing future experiments on electrophysiological responses under different reverberant conditions.

CONCLUSION

This is the first study to investigate how various components (direct sound versus early reflections versus late reflections versus full RIR) of reverberated speech are encoded in the auditory system using the speech-evoked ABR. It shows that effects of the different components of the RIR can be studied electrophysiologically, thereby providing a means other than perceptual studies to investigate the effect of room acoustics on auditory processing. The amplitude of the reverberant stimuli in our study was not normalized, as in previous studies on reverberation with speech-evoked ABR, thus allowing for the natural amplification that typically occurs in reverberant conditions.

Our results showed that responses to the direct and early reflections component are similar with respect to measures of waves V and A amplitudes and latencies as well as EFR and FFR amplitudes. Also, our results showed that responses to the late reflections component are different from responses to the direct or early reflections component. These findings provide insights toward understanding the effects of the different components of the RIRs on auditory processing of speech, and are consistent with the notion that the direct sound and early reflections constitute useful energy, while the late reverberation constitutes detrimental energy. As such, the study provides support at the auditory processing level that reverberation should be optimized (i.e., increase early reflections and decrease late reflections) rather than simply minimized in practical situations. Future work should explore the correlation between reduced/increased speech-evoked ABR measures (EFR and FFR amplitudes, and waves V and A amplitude and latencies) and perceptual studies of room acoustics.

This study was conducted for one specific room and one specific source–receiver set-up. The RIR was created with the image source method for a rectangular room with an RT of 1.0 sec. A vowel /a/ was convolved with different components of RIR (direct, early, late, and full). The convolved speech stimuli were delivered monaurally to the participants' right ear. To generalize the effects of reverberation on the speech-evoked ABR, additional acoustic scenarios should be studied where direct, early, and late reflections are presented in different combinations and proportions. As such, supplementary scenarios need to be investigated with different source-receiver set-ups, room shapes, and RTs. Moreover, different populations should be investigated in a real acoustic environment that includes background noise and a richer set of speech stimuli, delivered binaurally, especially with children, given the important effect of reverberation on classroom learning.

REFERENCES

Adams EM, Gordon-Hickey S, Morlas H, Moore R. (2012) Effect of rate-alteration on speech perception in noise in older adults with normal hearing and hearing impairment. *Am J Audiol* 21(1):22–32.

Aiken SJ, Picton TW. (2008) Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245(1–2):35–47.

Al Osman R, Giguere C, Dajani H. (2016) Effects of stimulus rate and noise on speech-evoked auditory brainstem responses. *Can J Speech Lang* 40(2):104–119.

Allen JB, Berkley DA. (1979) Image method for efficiently simulating small-room acoustics. *J Acoust Soc Am* 65(4):943–950.

American National Standards Institute (ANSI). (2010) *Acoustical Performance Criteria, Design Requirements and Guidelines for Schools, Part 1: Permanent Schools. ANSI S12.60–2009/Part 1.* New York, NY: ANSI.

Anderson S, Kraus N. (2010) Objective neural indices of speech-in-noise perception. *Trends Amplif* 14(2):73–83.

Anderson S, Kraus N. (2012) cABR: a neural probe of speech-in-noise processing. Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research, Nyborg, Denmark.

Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N. (2013) Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance. *J Speech Lang Hear Res* 56(1):31–43.

Anderson S, Skoe E, Chandrasekaran B, Zecker S, Kraus N. (2010) Brainstem correlates of speech-in-noise perception in children. *Hear Res* 270(1–2):151–157.

Architectural institute of Japan. (2004) Sound material in living environment "SMILE 2004". Tokyo: Gilhodo Shuppan Co. Ltd.

Arweiler I, Buchholz JM. (2011) The influence of spectral characteristics of early reflections on speech intelligibility. *J Acoust Soc Am* 130(2):996–1005.

Assmann P, Summerfield Q. (2004) Perception of speech under adverse conditions. In: Popper AN, Fay RR, eds. *Speech Processing in the Auditory System.* 1st ed., Chapter 5. New York, NY: Springer New York, 231–309.

Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. (2009) Reading and subcortical auditory function. *Cereb Cortex* 19(11): 2699–2707.

Bidelman GM, Krishnan A. (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res* 1355:112–125.

Blauert J. (1997) *Spatial Hearing: The Psychophysics of Human Sound Localization*. Revised Edition. Cambridge, MA: MIT press.

Boothroyd A. (2008) The acoustic speech signal. In: Madel and Flexer, eds. *Pediatric Audiology*, Chapter 17. New York, NY: Thieme, 159–167.

Bradley JS. (1986) Speech intelligibility studies in classrooms. *J Acoust Soc Am* 80(3):846–854.

Bradley JS, Sato H, Picard M. (2003) On the importance of early reflections for speech in rooms. *J Acoust Soc Am* 113(6): 3233–3244.

Campbell T, Kerlin JR, Bishop CW, Miller LM. (2012) Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear Hear* 33(1):144–150.

Canadian Standards Association (CSA). (2000) *CSA-Z412 Guideline on Office Ergonomics*. Toronto, ON: CSA International.

Chandrasekaran B, Kraus N. (2010) The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47(2):236–246.

Crandell CC, Smaldino JJ. (2000) Classroom acoustics for children with normal hearing and with hearing impairment. *Lang Speech Hear Serv Sch* 31(4):362–370.

Devore S, Shinn-Cunningham B. (2003) Perceptual consequences of including reverberation in spatial auditory displays. Proceedings of International Conference on Auditory Displays, Boston, MA.

Fujihira H, Shiraishi K. (2015) Correlations between word intelligibility under reverberation and speech auditory brainstem responses in elderly listeners. *Clin Neurophysiol* 126(1):96–102.

Gardner WG. (2002) Reverberation algorithms. In: Kahrs M, Brandenburg K, eds. *Applications of Digital Signal Processing to Audio and Acoustics*, Chapter 3. Norwell, MA: Kluwer Academic Publishers, 85–131.

Gelfand SA, Silman S. (1979) Effects of small room reverberation upon the recognition of some consonant features. *J Acoust Soc Am* 66:22–29.

Giguère C. (2013) Understanding hearing loss: implications for speech perception. In: Fitzpattrick EM, Doucet SP, eds. *Audiologic Rehabilitation for Children with Hearing Loss: From Infancy to Adolescence*, Chapter 2. New York, NY: Thieme Medical Publishers, 18–28.

Hazrati O, Loizou PC. (2012) The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners. *Int J Audiol* 51(6):437–443.

Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. (2009) Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci USA* 106(31):13022–13027.

Houtgast T, Steeneken HJ. (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.

International Electrotechnical Commission. (2010) *IEC 60318-4. Electroacoustics: Simulators of the Human Head and Ear. Part4: Occluded-Ear Simulator for the Measurement of Earphones Coupled to the Ear by Means of Ear Inserts*. Geneva, Switzerland: IEC.

Irwin JD, Graf ER. (1979) *Industrial Noise and Vibration Control*. Englewood Cliffs, NJ: Prentice-Hall.

Klatt HD. (1980) Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am* 67:971–995.

Laroche M, Dajani HR, Prévost F, Marcoux AM. (2013) Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. *Ear Hear* 34(1):63–74.

McGovern SG. (2003) A Model for Room Acoustics. https://www.mathworks.com/matlabcentral/fileexchange/5116-room-impulse-response-generator/content/rir.m. Accessed January 2, 2013.

Nábělek AK, Letowski TR. (1988) Similarities of vowels in nonreverberant and reverberant fields. *J Acoust Soc Am* 83(5):1891–1899.

Neuman AC, Wróblewski M, Hajicek J, Rubinstein A. (2010) Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear Hear* 31(3):336–344.

Pichora-Fuller MK. (2003) Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *Int J Audiol* 42(Suppl 1):S59–S67.

Prévost F, Laroche M, Marcoux AM, Dajani HR. (2013) Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clin Neurophysiol* 124(1):52–60.

Roman N, Woodruff J. (2013) Speech intelligibility in reverberation with ideal binary masking: effects of early reflections and signal-to-noise ratio threshold. *J Acoust Soc Am* 133(3):1707–1717.

Sayles M, Winter IM. (2008) Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron* 58(5):789–801.

Skoe E, Kraus N. (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31(3):302–324.

Song JH, Skoe E, Banai K, Kraus N. (2011) Perception of speech in noise: neural correlates. *J Cogn Neurosci* 23(9):2268–2279.

Song JH, Skoe E, Banai K, Kraus N. (2012) Training to improve hearing speech in noise: biological mechanisms. *Cereb Cortex* 22:(5):1180–1190.

Tun PA. (1998) Fast noisy speech: age differences in processing rapid speech with background noise. *Psychol Aging* 13(3):424–434.

Wagener KC, Brand T. (2005) Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters. *Int J Audiol* 44(3):144–156.

Wible B, Nicol T, Kraus N. (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* 67(3):299–317.

Wróblewski M, Lewis DE, Valente DL, Stelmachowicz PG. (2012) Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear Hear* 33(6):731–744.

Yang W, Bradley JS. (2009) Effects of room acoustics on the intelligibility of speech in classrooms for young children. *J Acoust Soc Am* 125(2):922–933.