Review Article

Adaptation to Reverberation for Speech Perception: A Systematic Review

Avgeris Tsironis[1], Eleni Vlahou[1] (iD)**, Panagiota Kontou[1], Pantelis Bagos[1] and Norbert Kopčo[2]** (iD)

Abstract

In everyday acoustic environments, reverberation alters the speech signal received at the ears. Normal-hearing listeners are robust to these distortions, quickly recalibrating to achieve accurate speech perception. Over the past two decades, multiple studies have investigated the various adaptation mechanisms that listeners use to mitigate the negative impacts of reverberation and improve speech intelligibility. Following the PRISMA guidelines, we performed a systematic review of these studies, with the aim to summarize existing research, identify open questions, and propose future directions. Two researchers independently assessed a total of 661 studies, ultimately including 23 in the review. Our results showed that adaptation to reverberant speech is robust across diverse environments, experimental setups, speech units, and tasks, in noise-masked or unmasked conditions. The time course of adaptation is rapid, sometimes occurring in less than 1 s, but this can vary depending on the reverberation and noise levels of the acoustic environment. Adaptation is stronger in moderately reverberant rooms and minimal in rooms with very intense reverberation. While the mechanisms underlying the recalibration are largely unknown, adaptation to the direct-to-reverberant ratio-related changes in amplitude modulation appears to be the predominant candidate. However, additional factors need to be explored to provide a unified theory for the effect and its applications.

Introduction

The sound travels from a source to a receiver through one direct and multiple indirect paths that are created as the sound reflects off various surfaces in the environment. These time-delayed, scaled copies of the direct sound are added to the overall signal and produce reverberation (Assmann & Summerfield, 2004; Beeston et al., 2014). Reverberation affects temporal and spectral features of the signal that reaches the ears by attenuating its amplitude modulation (AM), prolonging the energy peaks and masking the energy dips (Assmann & Summerfield, 2004; Nielsen & Dau, 2010; Shinn-Cunningham, 2003). It is ubiquitous in real-world listening and it impacts nearly all aspects of auditory processing, including sound localization, sound externalization, stream segregation, and speech intelligibility (e.g., Best et al., 2020; Culling et al., 2003; Gelfand & Silman, 1979; Helfer & Huntley, 1991; Knudsen, 1929; Nábělek et al., 1989; Shinn-Cunningham, 2000; Zahorik et al., 2005).

In reverberant sound fields, the reflections arrive at the ears from multiple directions, interfering with the direct sound and distorting interaural time and level differences, the binaural cues that are used for sound localization (Devore & Delgutte, 2010). In such environments, the perceived source location is dominated by the first arriving waveform, as can be illustrated for short click stimuli by the "precedence effect" (Litovsky et al., 1999). For longer stimuli consisting of multiple clicks, the dominance of sound onsets for spatial processing is exhibited as an increase in the perceptual weight of the early clicks relative to later arriving sound that is more degraded by reverberation (see, e.g., Stecker & Moore, 2018).

[1]Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece
[2]Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University, Košice, Slovakia

Corresponding author:
Eleni Vlahou, Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece.
Email: evlahou@gmail.com

Reverberation can be both detrimental and beneficial for spatial hearing. On the one hand, it can degrade directional sound localization accuracy (Shinn-Cunningham, 2000). On the other hand, it can serve as a distance cue, improving the accuracy of distance judgements (Zahorik et al., 2005). Importantly, sustained exposure to a mildly reverberant room over the period of hours leads to improvements in both directional accuracy and distance perception (Shinn-Cunningham, 2000), illustrating that adaptation to reverberation can improve spatial processing, albeit at a time scale longer than typically used in speech perception studies.

In the context of speech processing, early reflections that reach the listener within the first 50 ms after the direct sound increase the effective signal-to-noise ratio and boost intelligibility, for both normal-hearing and hearing-impaired listeners (Bradley et al., 2003). Conversely, late reverberation, especially at severe levels, degrades intelligibility (e.g., Gelfand & Silman, 1979; Knudsen, 1929; Reinhart & Souza, 2018).

Numerous studies have investigated the effects of reverberation, presented either alone or in combination with noise, on different classes of speech sounds. There is substantial variability in results ranging from minimal to strong disruptions in perception. Speech sounds with short and rapidly transient spectra are affected more severely (Assmann & Summerfield, 2004). Among consonants, the stops are particularly susceptible to disruption, as they contain periods of low energy and transient energy bursts, and reverberation fills in the silent gap during stop closure (Gelfand & Silman, 1979; Helfer, 1994). On the other hand, sibilant fricatives, characterized by strong energy at higher frequencies, are a class of sounds resilient to the effects of reverberation, while the perception of low-energy non-sibilant fricatives is deteriorated (Assmann & Summerfield, 2004; Gelfand & Silman, 1979). The place of the sound within a word also has a strong effect. Consonants in word-final position are affected more relative to word-initial position, due to overlap-masking from the energy of the preceding segment (e.g., Gelfand & Silman, 1979; Knudsen, 1929). The perception of vowels with longer steady-state energy is well retained by normal-hearing listeners, while perception can be degraded for diphthongs with rapidly changing formant transitions or for monophthongal vowels, for which segmental duration creates a phonemic difference (Assmann & Summerfield, 2004; Osawa et al., 2021; 2018).

The negative effects of reverberation are more pronounced for nonnative listeners, for children and older adults, and for individuals with hearing difficulties (Assmann & Summerfield, 2004; Lecumberri et al., 2010; Reinhart & Souza, 2018). Older listeners without significant peripheral hearing loss experience a decline in the perception of reverberant speech (Helfer & Huntley, 1991). Furthermore, the smearing of the temporal envelope induced by reverberation can be detrimental to people with cochlear implants, for whom even small amounts of reverberation deteriorate performance (Poissant et al., 2006). For nonnative listeners, signal degradations due to reverberation interact with imperfect linguistic knowledge, significantly degrading performance (Lecumberri et al., 2010; Nábĕlek & Donahue, 1984; Takata & Nábĕlek, 1990). However, there is some evidence that nonnative listeners might benefit from experiencing novel sounds in different rooms during implicit phonetic training (Vlahou et al., 2019).

In summary, research over the past several decades suggests that reverberation affects many aspects of spatial hearing and speech intelligibility in both positive and negative ways. It can be particularly detrimental for some populations, such as nonnative listeners and hearing-impaired individuals. On the other hand, for normal hearing adults in moderately reverberant environments, the perceptual impact of reverberation is negligible. People quickly adapt to room acoustics and communicate without experiencing difficulties or even noticing signal degradations. This phenomenon illustrates "phonetic perceptual constancy," akin to loudness constancy in audition and color, shape, and brightness constancy in vision (Assmann & Summerfield, 2004; Stecker & Hafter, 2000; Watkins & Makin, 2007; Watkins et al., 2011; Zahorik & Wightman, 2001).

To achieve phonetic perceptual constancy, the auditory system must recalibrate its processing of speech stimuli in each new reverberant environment, compensating for the specific distortions caused by the reverberant energy. While the factors and mechanisms of this calibration process are largely unknown, it has attracted increased research interest in the past couple of decades. Recent studies, using acoustic environments with varying levels of reverberation and diverse speech stimuli and tasks, have produced robust and consistent evidence that the reverberation of the preceding acoustic context can facilitate or disrupt subsequent speech perception (e.g., Brandewie & Zahorik, 2010; Beeston et al., 2014; Vlahou et al., 2021; Watkins, 2005b, 2011). A few studies have begun to investigate perceptual mechanisms that might underlie the effect (e.g., Srinivasan & Zahorik, 2014; Stilp et al., 2016; Watkins et al., 2011; Zahorik & Anderson, 2013). In parallel with psychophysical studies, human and animal neuroimaging experiments have revealed neural components that potentially support this complex adaptive ability of the auditory system (e.g., Devore & Delgutte, 2010; Devore et al., 2009; Fuglsang et al., 2017; Ivanov et al., 2022; Slama & Delgutte, 2015).

Here we present a systematic review of studies examining recalibration of speech perception after prior exposure to consistent or inconsistent reverberation. First, we present the different approaches used to quantify adaptation and to manipulate the consistency in the reverberation of the carrier and target speech. Next, we summarize key findings, both overall and in relation to the speech units investigated, as well as the time course of the phenomenon. We also review research investigating adaptation in nonnative listeners and hearing-impaired individuals. Then, we outline some of the perceptual and neurophysiological mechanisms purported to

underlie the effect. Lastly, we recommend key areas for future research, including the development of a unified theory that integrates the various contributing mechanisms to the effect, and the design of effective applications for adaptation to reverberation in augmented and virtual reality displays.

Methods

We chose to conduct a systematic review, as our approach aligns well with the guidelines for systematic reviews outlined by Munn et al. (2018), aiming to synthesize existing knowledge based on specific questions and inclusion criteria (see below), discuss the different methods used to measure adaptation, and provide insights to guide further research. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for the search methodology, screening strategy, and inclusion/exclusion criteria (Page et al., 2021).

Search and Selection Process

To search the available literature, we used the SCOPUS database. We included journal and conference articles, reviews, books, and book chapters. Studies in languages other than English were not considered. The search was conducted by two independent researchers (A.T. and E.V.) and was completed by January 24, 2022. The search terms "adaptation" OR "calibration" OR "learning" OR "compensation" OR "exposure" AND "reverbera*" OR "room" AND "speech" OR "phoneme" OR "sentence" OR "consonant" were entered into the title, abstract, and keyword fields. This search returned 651 results (see Supplemental material S1 and S2 for detailed search criteria and results).

Ten additional studies were identified based on citation searches and the authors' personal knowledge. After removal of one duplicate, studies were screened for eligibility by the same two researchers, based on the title and abstract. Inclusion criteria were that the articles had to be original studies or reviews that examined adaptation to reverberation for speech perception. Studies with no human participants investigating automatic speech recognition and signal processing dereverberation techniques were excluded. After discussion and agreement on any discrepancies, 578 studies were rejected during screening, leaving 72 studies for full-text reading. Many studies have investigated the detrimental effects of reverberation on speech perception (e.g., Culling et al., 2003; Helfer, 1994; Helfer & Huntley, 1991; Nábělek et al., 1989; see Assmann & Summerfield, 2004 for a review). Here, we only included studies that specifically examined how listeners adapt to reverberant speech, utilizing information from exposure to the immediately preceding room. Based on this criterion, the final set of 23 studies was determined. The screening and selection process is detailed in the PRISMA flowchart (Figure 1).
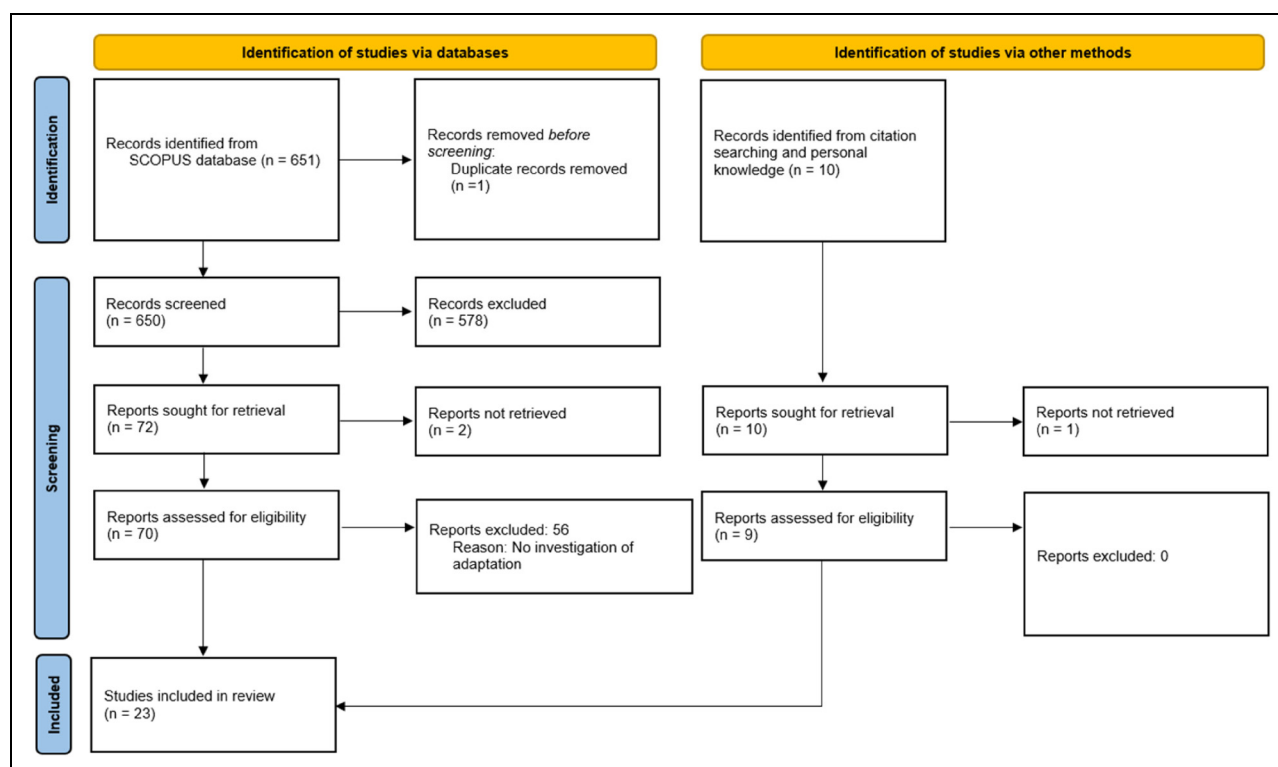

Figure 1. PRISMA flowchart illustrating the selection, screening, and inclusion/exclusion process of the review.

Results

Studies investigating adaptation to reverberant speech exhibit important differences and similarities. On the one hand, different labs have used different methods to quantify adaptation, employing diverse perceptual tasks, target speech units (phonemes, words, or sentences), monaural or binaural stimuli, noise-masked or unmasked speech, etc. This diversity precludes direct comparisons and quantitative synthesis across studies. On the other hand, there are many commonalities, including comparing the effect of matched versus mismatched preceding environment and examining the time course of adaptation (e.g., Beeston et al., 2014; Srinivasan & Zahorik, 2014; Vlahou et al., 2021). Here we first present a review of the main manipulated parameters, performance measures, and experimental setups used in the reviewed studies. Then, we present our systematic review of the 23 selected articles, comparing their results and identifying the main differences in terms of how reverberation was manipulated, whether the stimuli were masked by noise, what speech units were investigated, the time course of adaptation, the mechanisms of adaptation, and comparing native normal-hearing listeners to other listener groups.

The studies examined in this review are performed in virtual environments. In recent years, virtual sound presentation techniques have become essential tools in psychoacoustic research and hearing aid development. Without these techniques, many of the reviewed studies would be very challenging, if not impossible. Virtual sound presentation techniques enable the precise presentation of reverberant acoustic environments and allow researchers to create controlled, realistic acoustic conditions, which are vital for studying complex auditory phenomena (Kirsch et al., 2021). Most of the reviewed studies consist of trials in which the to-be-identified target stimulus is presented after a carrier stimulus. The adaptation is demonstrated by comparing performance in a consistent condition versus a no-carrier or an inconsistent condition. In these conditions, respectively, the reverberation of the carrier matches that of the target speech, the target is presented alone, or the carrier and the target reverberation differ. Modified/improved performance in the consistent condition is taken as evidence that listeners exploit information from the acoustic properties of the carrier to recalibrate speech perception. In some studies there is no carrier–target distinction within a trial; rather, adaptation is demonstrated by comparing performance between a consistent, "blocked" condition, in which the same reverberation is used for all trials within a block, and an inconsistent, "unblocked" condition, in which the reverberation of each trial within the block varies randomly (Osawa et al., 2021; Srinivasan & Zahorik, 2013, 2014; Srinivasan et al., 2016).

The inconsistency of the reverberation simulation has been achieved by two types of manipulations. Either different rooms were used for the carrier and the target/for each trial within a block (e.g., Brandewie & Zahorik, 2018;

Srinivasan & Zahorik, 2013; Vlahou et al., 2021), or different source-listener distances within the same room were used (studies by Watkins and colleagues, see Table 1).

Another distinction is whether monaural or binaural simulation is used. In binaural conditions, an important factor is whether the Binaural Room Impulse Response (BRIR) is recorded using the listener's own body or a standardized manikin. In all the studies reported here, non-individualized BRIRs were used. In virtual environments, using non-individualized BRIRs has the additional benefit that all the listeners hear the same identical stimuli. While studies by Watkins and colleagues have demonstrated robust adaptation using monaural stimuli, most researchers have employed binaural stimuli. Brandewie and Zahorik (2010) report limited adaptation with monaural presentation, while Garofolo et al. (2005b) report stronger adaptation in monaural presentation conditions. This discrepancy might be caused by differences in the experimental design, noise masking (see below), and speech stimuli across studies, or it might reflect the activation of different compensation mechanisms for binaural and monaural presentation. This issue is further discussed in the Conclusions section.

Yet another important difference concerns whether carrier and target stimuli are masked by noise. Zahorik and colleagues have used reverberation in combination with spatialized Gaussian noise. Introducing noise has two important benefits: it makes the task more difficult, effectively reducing ceiling effects in performance. It also makes the task more ecologically realistic, as everyday listening environments typically contain both noise and reverberation. On the other hand, the unique effects of reverberation and the listeners' compensation mechanisms might differ between noise-masked and unmasked conditions.

The investigation of adaptation to reverberation has spanned different speech units, ranging from individual phonemes and syllables (e.g., Beeston et al., 2014; Osawa et al., 2021; Vlahou et al., 2021) to ecologically realistic variable sentences (e.g., Srinivasan & Zahorik, 2014). The former approach allows for a more rigorous control over the effects of reverberation on different sounds, while the latter one better approximates the highly heterogeneous real-world speech communication.

Finally, various performance measures have been used in the adaptation studies, including improvement in speech reception thresholds (SRTs; Brandewie & Zahorik, 2010, 2013; Zahorik & Brandewie, 2016), information transfer rate (e.g., Vlahou et al., 2021), shifts in phoneme category boundaries (e.g., Watkins, 2005b; Watkins et al., 2011), or the reweighting of acoustic cues critical for phoneme perception (Stilp et al., 2016).

Reviewed Studies

Table 1 summarizes the studies included in the review. The table is primarily organized by the characteristics of each

Table I. Characterization of Studies Examining Adaptation to Reverberation for Speech Perception Included in the Review.

Study	Participants	Reverberation manipulation	Noise mask	Presentation mode	Target speech	Task	Performance measure
Watkins (2005a)	6, NH, N/F	Distance and room	N	Mono, bin	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Watkins (2005b)	24, NH, N/F	Distance and room	N	Mono, bin	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Watkins and Makin (2007)	36, NH, N/F	Distance	N	Mono	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Nielsen and Dau (2010)	18, NH, N/F	Distance	N	Diotic	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Watkins et al. (2010a)	12, NH, N/F	Distance	N	Mono	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Watkins et al. (2010b)	12, NH, N/F	Distance	N	Mono	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Watkins et al. (2011)	18, NH, N/F	Distance	N	Mono	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Watkins and Raimond (2013)	6, NH, N/F	Distance	N	Mono	[sir]-[stir] test words	Phoneme ID 2-AFC	Categorization shifts
Beeston et al. (2014)	160, NH, N/F	Distance	N	Mono	[sir], [skur],[spur], [stir] test words, 5 vowels	Phoneme ID 4-AFC	ITR
Vlahou et al. (2021)	18, NH, N/F	Room	N	Bin	VC syllables, vowel /a/+16 consonants, (k, t, p, f, g, d, b, v, ð, m, n, ŋ, z, θ, s, and ʃ)	Phoneme ID 16-AFC	Proportion correct, ITR
Longworth-Reed et al. (2009)	10, NH, N/F	Room	N	Bin, diotic	HINT	Word ID	Proportion correct
Brandewie and Zahorik (2010)	14, NH, N/F	Room	Y	Mono, bin	CRM	Select correct color/number	Proportion correct, SRTs
Brandewie and Zahorik (2011)	14, NH, N/F	Room	Y	Bin	MRT	Select target word from list of words	Proportion correct, SRTs
Brandewie and Zahorik (2013)	16, NH, N/F	Room	Y	Bin	CRM	Select correct color/number	Proportion correct
Srinivasan and Zahorik (2011)	21, NH, N/F	Room	Y	Bin	SPIN	Type last word of every sentence	Proportion correct
Srinivasan and Zahorik (2013)	60, NH, N/F	Room	Y	Bin	PRESTO	Type all words from sentence	Proportion correct
Srinivasan and Zahorik (2014)	30, NH, N/F	Room	N	Bin	PRESTO	Type all words from sentence	Proportion correct
Zahorik and Brandewie (2016)	49, NH, N/F	Room	Y	Bin	CRM	Select correct color/number	Proportion correct, SRTs

Table 1. Continued.

Study	Participants	Reverberation manipulation	Noise mask	Presentation mode	Target speech	Task	Performance measure
Brandewie and Zahorik (2018)	27, NH, N/F	Room	Y	Bin	CRM	Select correct color/number	Proportion correct, SRTs
Stilp et al. (2016)	63, NH, N/F	Room	N	Diotic	Synthetic vowel continuum (/i/-/u/)	Phoneme ID—2-AFC	Acoustic cue reweighting
Osawa et al. (2021)	10 NNS, 11 NS (all NH)	Distance and room	N	Bin	Japanese vowel contrast (/ie/-/iie/)	Phoneme ID—2-AFC	Categorization shifts
Zahorik and Brandewie (2011)	14 NH, 12 HI (all N/F)	Room	Y	Bin	CRM	Select correct color/number	SRTs
Srinivasan et al. (2016)	6 CI users	Room	N	Better ear	IEEE, TIMIT	Repeat back all words from sentence	Proportion correct

NH = normal-hearing; N/F = native or fluent speakers; HI = hearing impaired; CI = cochlear implant; NS = native speakers; NNS = nonnative speakers; mono = monaural; bin = binaural; HINT = Hearing in Noise Test (Nilsson et al., 1994); CRM = Coordinate Response Measure (Bolia et al., 2000); MRT = Modified Rhyme Test (House et al., 1965); SPIN = Speech Perception In Noise (Kalikow et al., 1977); PRESTO = Perceptually Robust English Sentence Test Open-set database (Gilbert et al., 2013); IEEE = The Institute of Electrical and Electronics Engineers sentence corpus (IEEE, 1969); TIMIT = DARPA TIMIT acoustic–phonetic continuous speech corpus (Garofolo et al., 1993); n-AFC task = n-alternative forced choice task; ITR = information transfer rate; SRTs = speech reception thresholds; ID = identification.

study outlined above, mainly the method of reverberation manipulation, noise masking, monaural or binaural presentation, lexical units for the target speech, and differences in tasks and performance measures. For the effect of strong or weak reverberation on adaptation, the studies varied in parameters and reverberation levels, making it challenging to summarize these effects in a table. Most research reviewed here has been conducted by Zahorik and colleagues and Watkins and colleagues (10 and 8 studies, respectively, out of the 23 studies reported here), with the remaining studies performed by other groups. In the following sections, the studies are compared based on the different characteristics.

Reverberation Manipulation and Effect of Masking

The main distinction between the studies reviewed here is whether the reverberation was manipulated by changing the distance from which the carrier and target were simulated or by changing the room in which the sources were simulated. A secondary prominent distinction is whether the target speech was masked by noise or not.

Manipulating Source-Listener Distance. In studies by Watkins and colleagues, listeners performed a phoneme identification test, identifying test words as "sir" or "stir". Tests words were drawn from an 11-step continuum between /sir/-/stir/, created by amplitude modulating tokens of "sir" to receive the temporal envelope of "stir" at various modulation depths (e.g.,

Watkins, 2005b). Test words were embedded in a context phrase ("OK, next you'll get [test word] to click on"). Both the context and the test words were convolved with room impulse responses recorded at a near distance (source at 0.32 m from the listener) with low reverberation, and a far distance (source at 10 m from the listener), with high reverberation. When the context was near and the test word was far, reverberation from the test-word filled the gap in its temporal envelope, masking an important cue for the identification of /t/, and participants tended to hear more "sir" responses and shift the category boundary. However, when the context was also simulated at a far distance, matching the test word's distance, listeners tend to hear "stir" again, shifting the category boundary back. In this design, while improvement in speech perception was not explicitly measured, adaptation was expressed as shifts in phoneme category boundary, as a function of the carrier distance.

In a series of experiments, this finding was replicated and extended across various conditions: for rooms with different sizes and geometry, under both normal and fast speech rates, for steady-spectrum noise-contexts with rapidly varying temporal envelopes, and for noise-vocoded speech stimuli (Watkins, 2005a, 2005b; Watkins & Makin, 2007; Watkins & Raimond, 2013; Watkins et al., 2011). Importantly, this design does not appear to rely on binaural input, as compensation with monaural speech appears to be as effective, or even stronger (Watkins, 2005a, 2005b) than with binaural speech. These findings were interpreted as evidence of a

monaural "extrinsic" compensation mechanism that is informed by the level of reverberation of the context. Later studies showed that, in addition to information from the preceding context, there are also important "intrinsic" sources of information that can facilitate adaptation. Specifically, information from the test-word itself, such as from its reverberation tail, plays a significant role in adaptation (Beeston et al., 2014; Watkins & Raimond, 2013). These intrinsic cues can help listeners adapt to reverberation even in the absence of an extrinsic context (preceding speech) (Beeston et al., 2014; Watkins & Raimond, 2013).

Manipulating Rooms and Speech Masking. Other groups have investigated adaptation to reverberation when the carrier and target are simulated in different rooms. Zahorik and colleagues have investigated this type of adaptation using different methods. In the majority of their studies, the source-listener distance was fixed, with the simulated speech source placed in front of the listener at 1.4 m and a spatially separated noise masker, also at 1.4 m, directly opposite to the listener's right ear (90° azimuth angle). In a consistent condition, the simulated room remained constant, both within a trial, where the target speech is preceded by a carrier from the same room, and throughout a block of trials, thus maximizing consistent exposure. This condition was compared against a no-carrier condition, where the target speech was presented without a preceding carrier and the simulated target room changed randomly from trial to trial, or against an inconsistent condition, where the target was preceded by a speech carrier from a different room (Brandewie & Zahorik, 2018). Using this and similar designs, a series of studies have consistently shown that, after brief prior exposure to consistent reverberation, participants improve speech perception relative to the no-carrier or inconsistent conditions.

An important aspect of this paradigm is that adaptation appears to require binaural information. In one study, participants showed an 18% improvement in word recognition after prior exposure to a consistent simulated room, compared to a no-carrier condition (Brandewie & Zahorik, 2010, Exp. 1). However, in an identical experiment with monaural input (with the right-ear signal digitally removed and the left-ear signal contralateral to the masker retained; Exp. 3), only two of the 14 participants showed improvement. The opposite pattern was observed in Watkins (2005b), where more robust adaptation was observed with monaural input. It is unclear whether this discrepancy is a result of distinct monaural versus binaural adaptation mechanisms that operate in the different paradigms, or whether it is a result of differences in the experimental setup and tasks. The concurrent presentation of spatialized noise in Brandewie and Zahorik (2010), alongside the primary task of speech recognition, introduces additional factors related to sound localization and spatial

unmasking (Beeston et al., 2014), further complicating the interpretation of the results.

Several experiments have investigated the effects of different carrier characteristics in this paradigm, and the magnitude of adaptation under diverse noise and reverberation levels. Exposure to inconsistent reverberation within a trial (i.e., when a preceding speech carrier is from a different simulated room than the target speech) can significantly degrade performance compared to a consistent condition, causing it to reach baseline levels, where the target speech is presented alone. There is some evidence that the drop in performance between consistent and inconsistent conditions is larger when the reverberation of the preceding carrier is more intense than the target's reverberation (Brandewie & Zahorik, 2018). In this paradigm, adaptation is most robust for moderate target reverberation conditions ($T_{60} \sim 1$ s), leading to an approximately 20% improvement in intelligibility (Zahorik & Brandewie, 2016). However, as the level of target reverberation increases, the adaptation effect becomes weaker, becoming negligible in strongly reverberant rooms with T_{60} at 3 s (Zahorik & Brandewie, 2011; 2016).

Another study used a similar experimental design except that no masking noise was applied (Vlahou et al., 2021). Two environments with strong levels of reverberation (broadband T_{60}'s of 2.5 s and slightly over 3 s, respectively) were used for targets. The same two environments, as well as an anechoic environment, were used for the carrier. In this study, the effects of a consistent preceding carrier were compared against different types of inconsistent carriers, including a no-context baseline, in which the target speech was presented without any carrier, an anechoic carrier, and a carrier presented in a different simulated room. In general, the effect of consistent reverberation was significant, but fairly small in this study, 5%–7% for the less reverberant room, whereas for the more reverberant room the effect was negligible, on the order of 1%–2%. This result partially corroborates the finding that in very strong reverberation adaptation is attenuated (Zahorik & Brandewie, 2011, 2016), while showing that effective adaptation to reverberation is possible even for T_{60} of 2.5 s. The disruptive effect of the anechoic and different-reverberant carrier was fairly similar in the Vlahou et al. study, with slight tendency to be larger for the anechoic carrier. On the other side, in the Brandewie and Zahorik (2018) study the disruptive effects were larger for the carrier with more reverberation.

Speech Units and Presentation Mode

The adaptation has been investigated for a range of speech units, from phonemes to sentences, and for both monaural and binaural presentation levels. Most studies investigating adaptation at the phoneme level have used consonants as target speech, with only a few studies using vowels (Osawa et al., 2021; Stilp et al., 2016). In studies by

Watkins and colleagues, adaptation has been repeatedly demonstrated for the unvoiced plosive /t/ within [sir]-[stir] test words, presented mostly monaurally. Beeston et al. (2014) extended this design, demonstrating monaural adaptation for two additional unvoiced stops (/p/, /k/). This study also introduced more variability by incorporating multiple speakers and featuring a greater number of vowels in the test words, although listeners were specifically tasked with identifying the heard consonant. The focus on unvoiced plosives differing in place of articulation is motivated by the fact that these features are more severely degraded by reverberation (Assmann & Summerfield, 2004; Gelfand & Silman, 1979). However, it's unclear whether this type of adaptation generalizes to different speech units that are also affected (e.g., low energy fricatives, nasals; Gelfand & Silman, 1979; Helfer & Huntley, 1991; Nábělek et al., 1989), and to what extent it affects everyday speech. Still, the examined consonants account for more than 10% of phonemes encountered in everyday discourse (Beeston et al., 2014), corroborating the ecological validity of the effect.

Vlahou et al. (2021) investigated 16 consonants (k, t, p, f, g, d, b, v, ð, m, n, ŋ, z, θ, s, and ʃ), each preceded by the same vowel and using a binaural presentation mode. The study analyzed both accuracy of individual consonant identification and phonetic category identification using information transfer analysis (Miller & Nicely, 1955; Shannon, 1948). It showed that the manner of articulation was the feature with the most robust improvement in the consistent condition. The effect was consistent across both simulated rooms, but it was restricted to stop consonants. There was also a significant improvement for voicing, but only in one of the simulated rooms.

Rather than phonemes and phonetic categories, Zahorik and colleagues have used words and sentences as target speech, drawn from various speech corpora, mostly using binaural presentation. The Coordinate Response Measure corpus (CRM; Bolia et al., 2000) was used in 5 out of the 10 studies by Zahorik and colleagues reported in Table 1. The CRM is a closed-set corpus, where participants choose their response from a limited set of predefined options. Sentences follow a structure ("Ready [call sign] go to [color] [number] now"), with the call sign known in advance and the participant selecting the correct color–number combination from eight numbers and four colors. The CRM has been used widely in speech-on-speech intelligibility research. However, since its linguistic variation and vocabulary size are limited (Eddins & Liu, 2012; Jakien et al., 2017), researchers have also used other corpora. One study from this lab (Brandewie & Zahorik, 2011) that used the Modified Rhyme Test (House et al., 1965) reported no improvement, on average, after prior exposure to consistent reverberation. Other studies have used subsets from corpora such as the Hearing in Noise Test (HINT; Nilsson et al., 1994; used in Longworth-Reed et al., 2009), and

material with rich linguistic and indexical variability from the Speech Perception In Noise (SPIN; Kalikow et al., 1977; used in Srinivasan & Zahorik, 2011), and the Perceptually Robust English Sentence Test Open-set database (PRESTO; used in Srinivasan & Zahorik, 2013, 2014). Finally, the IEEE corpus (IEEE, 1969) and TIMIT sentences (Garofolo et al., 1993) were also used in Srinivasan et al. (2016). Higher benefit of adaptation was observed for IEEE sentences than for TIMIT sentences, likely due to the more heterogeneous characteristics of the TIMIT corpus, requiring listeners to adjust to various parameters such as multiple talkers, regional dialects, and speaking rates.

Overall, these studies demonstrate that the adaptation to reverberation is observed at a range of speech units in both monaural and binaural stimulus presentation. However, the adaptation effect was not observed for all speech corpora, and it also depended on the mode of presentation, illustrating a complex relationship between these factors.

Time Course of Adaptation

The temporal dynamics of adaptation to reverberation can be examined across various timescales, ranging from milliseconds and seconds to more extended periods, spanning days or even a lifetime of exposure to environmental regularities and speech sounds (e.g., Traer & McDermott, 2016).

Beeston et al. (2014) examined the time course of monaural adaptation using phrases that contained a single test syllable preceded by a sequence of context words. The context was split into two parts, the first presented at a near source-listener distance (0.32 m), and the second, preceding the test word, at a far distance (10 m). Examining performance on the subsequent far test words, they found that as exposure to the far portion of the carrier increased from 0 to 500 ms, participants made fewer phoneme misclassifications. These findings suggest that the effect is fast enough to build up across half a second. However, due to methodological constraints the maximum duration of consistent exposure did not exceed 500 ms, thus it is not possible to determine the time needed for performance to plateau in this paradigm.

Rapid timescales were also reported from several studies with binaural stimuli in masked and unmasked conditions. In Vlahou et al. (2021) the length of the preceding carrier was manipulated such that in one condition it contained two syllables and in another four syllables (~800 and ~1600 ms, respectively). There was no evidence that consistent exposure over ~1 s improved phoneme identification. Longworth-Reed et al. (2009) compared the first and last 10 sentences within blocks that provided consistent exposure to a listening environment and showed improved word recognition by approximately 6% for a binaural condition with time-forward reverberation. However, other studies which have partitioned the data showed no further improvement after the first partition that they examined (first 18 trials in

Brandewie & Zahorik, 2010; first 6 sentences in Srinivasan & Zahorik, 2013; and first 5 sentences in Srinivasan et al., 2016). Using varying levels of reverberation and different signal-to-noise ratios (SNRs at −13 and −18 dB) Brandewie and Zahorik (2013) created six conditions in which they varied the length of the speech carrier phrase that preceded the target phrase, from 0 to 2.7 s. The duration of exposure required for intelligibility improvement to asymptote increased with SNR, from just 850 ms for the lower to 2.7 s for the higher SNR (Brandewie & Zahorik, 2013). These results suggest that adaptation to reverberant speech can be fully developed within 1 s in some conditions (Vlahou et al., 2021; Zahorik, 2019), but the precise timescale can vary widely for different levels of reverberation and noise.

While research on spatial hearing indicates that localization performance in a real room can continue to improve after several hours (Shinn-Cunningham, 2000), there is a lack of data exploring such long-term effects on speech perception. In a pilot study from our lab, using the experimental design from Vlahou et al. (2021) we examined the effects of continued exposure to simulated rooms across three 1 h sessions. Results from four participants showed no evidence of improved phoneme identification compared to the baseline session (data not shown). However, the sample size in this study was very small and only one room with intense reverberation was used, which has been shown to attenuate adaptation (Vlahou et al., 2021; Zahorik & Brandewie, 2011, 2016).

Adaptation in Nonnative and Hearing-Impaired Individuals

Two studies examined adaptation in hearing-impaired listeners (Brandewie & Zahorik, 2011; Srinivasan et al., 2016). Only one study specifically examined adaptation for nonnative listeners (Osawa et al., 2021). For the remaining reported studies, participants were normal-hearing listeners, and, when this information is reported, either native, or nonnative but fluent speakers of the target language.

There is some evidence that hearing-impaired individuals, who are particularly affected by the negative effects of reverberation, can also benefit from prior consistent exposure. Zahorik and Brandewie (2011) examined normal hearing listeners and listeners with sensorineural hearing loss of varying severity. They found that, although the SRTs from the hearing-impaired group were elevated compared to the normal hearing group, improvement due to consistent exposure was similar across groups. Consistent with the previous report of Zahorik and Brandewie (2016), the effect was strongest for the environments with modest reverberation, while little improvement was observed for anechoic rooms or rooms with more intense reverberation. Another study showed that cochlear implant users, who heard sentences presented to their self-reported best ear, were also able to significantly improve intelligibility when consistent reverberation was provided (Srinivasan et al., 2016).

To our knowledge, only one study has explicitly examined the effects of consistent versus inconsistent reverberation for nonnative listeners. Osawa et al. (2021) exposed native and nonnative listeners to tokens from a Japanese vowel length contrast along a durational continuum from /ie/ to /iie/. Reverberation adds a tail to the sound's offsets, elongating the perceived duration and thus obscuring a critical cue for the distinction of length contrasts (Osawa et al., 2018; 2021). The sounds were presented in anechoic and simulated rooms, in a "blocked" condition, where the same room was used throughout a block of trials, and an "unblocked" condition, where the simulated room varied randomly in each trial. Results showed that native listeners' categorization responses were unaffected by whether the room changed from trial to trial or remained consistent throughout the block. Nonnative listeners, on the other hand, changed their categorization responses significantly, increasing the long vowel responses in the unblocked condition for the more reverberant room. These results highlight nonnative listeners' sensitivity to variations in room acoustics, suggesting that inconsistent reverberation might be more disruptive for this population.

Perceptual and Neurophysiological Mechanisms

What mechanisms drive adaptation to reverberation for speech processing? Which aspects of the room acoustics and the speech signal do people use to recalibrate speech perception? T_{60} and the direct-to-reverberant energy ratio (DRR) are two parameters that have been dominant in the acoustic characterization of the environments used in the adaptation studies. The studies primarily manipulating the target-listener distance in a fixed room essentially manipulated the DRR while keeping the T_{60} constant (e.g., Beeston et al., 2014; Watkins & Makin, 2007; Watkins et al., 2011), while the studies switching the room primarily manipulated the T_{60} while largely disregarding the DRR, even though that also could vary as the room switched (e.g., Brandewie & Zahorik, 2010, 2018; Vlahou et al., 2021). The two measures are in general correlated, as a larger T_{60} means more reverberant energy and thus, on average, a lower DRR. However, since DRR is distance dependent and T_{60} is not, the fundamental question is whether the brain's adaptation to reverberation aims to compensate for changes in DRR or in T_{60}. In typical listening situations, as in a conversation with multiple talkers and other sources, the distances between the listener and the talkers randomly vary when there are multiple talkers. Thus, the adaptation to DRR would need to occur each time a new talker takes a turn in a conversation, that is, on the order of seconds. On the

other hand, the T_{60} stays constant in that scenario and thus the adaptation to it can take place over much longer time scales, corresponding to minutes or hours that a listener typically spends in one room. Thus, it is likely that adaptation to DRR would need to occur on the time scale of seconds, and it looks like no long-term room learning would be beneficial in such a scenario as the target-listener distances, and thus the DRR that the listener needs to adapt to, can change at any time to any value from a continuum. On the other hand, learning the distance-invariant effects of the room reverberation on the stimuli, that is, learning how to adapt to a room with a given T_{60}, might be much more beneficial on a longer time scale corresponding to how long a listener stays in one room. Moreover, since the listeners are commonly present in the same room repeatedly, such learning can even proceed over multiple visits. The results from the studies reviewed here suggest that most of the adaptation effects are fast, possibly supporting the DRR being the dominant parameter. However, since longer-term room learning effect are observed, for example, in distance perception in which they must be based on T_{60} (as the DRR-to-distance mapping must change for every room), it is still possible that such learning would generalize to speech perception, providing benefits in some specific conditions.

Importantly, while DRR is a convenient acoustic measure to characterize the reverberation effects on received sounds, it is unlikely that the brain can directly extract it from the stimuli as it would require deconvolving the BRIR from the heard stimuli and separating it into the direct and reverberant parts (Rakerd et al., 1999). However, several other measures are correlated with the DRR, including the AM (Zahorik et al., 2011), the early-to-late power ratio (Bronkhorst & Houtgast, 1999), frequency-to-frequency variation (Kopčo & Shinn-Cunningham, 2011), and interaural cross-correlation (Larsen et al., 2008), either systematically increasing or decreasing with reverberation. Thus, any of these parameters might be the ones actually extracted and adapted to instead of the DRR.

While a comprehensive understanding of the adapted cues is still lacking, recent studies have revealed several potential mechanisms that enable adaptation through (a) temporal envelope processing (e.g., Srinivasan & Zahorik, 2014; Watkins & Makin, 2007; Watkins et al., 2011; Zahorik, 2019), (b) acoustic cue reweighting (Stilp et al., 2016), and (c) tuning to statistical regularities of the reverberation (Traer & McDermott, 2016). In parallel, neurophysiological studies have examined neural compensatory mechanisms that support adaptation across different areas in the brain (e.g., Barzelay et al., 2023; Fuglsang et al., 2017; Ivanov et al., 2022; Slama & Delgutte, 2015). In the following, these mechanisms are described in more detail.

Temporal Envelope Processing

Both the temporal envelope, that is, the slow variations in narrowband amplitude over time, and the temporal fine structure, that is, the rapid oscillations with rate near the center frequency of the band, carry important information for speech perception (Moore, 2008). And, while both these characteristics can be degraded by reverberation (e.g., Watkins et al., 2011), converging evidence suggests that adaptation relies primarily on information obtained from the temporal envelope and persists even when fine-structure cues become unavailable (Srinivasan & Zahorik, 2014; Watkins & Makin, 2007; Watkins et al., 2011). For example, Watkins et al. (2011) reported that a noise-vocoded-speech carrier, which preserved the temporal envelope but not the fine structure, induced a similar amount of adaptation as a normal speech carrier containing both cues. Srinivasan and Zahorik (2014) exposed listeners to two types of chimeric stimuli: one in which the envelope was convolved with reverberant BRIR while the fine-structure was convolved with an anechoic HRTF, and one with the BRIR and HRTF reversed. The adaptation was observed only in the reverberant envelope condition, indicating the critical role of the temporal envelope.

What is less clear is which specific aspects of the temporal envelope are essential for adaptation. Zahorik (2019) proposed a conceptual model based on the modulation transfer function (MTF), a measure that quantifies the preservation of modulation depth in an enclosure and forms the basis of the Speech Transmission Index (STI; Houtgast & Steeneken, 1985). According to the model, adaptation is driven via monaural and binaural processing of AM information in a room. Estimation of the room MTF is followed by adaptation, that is, restoration of the reverberation-induced AM attenuations. This process is rapid, fully developed after approximately 1 s of consistent exposure to the room, and it might not be subject to further improvement (Zahorik, 2019). Behavioral findings from Zahorik and colleagues provide support for this framework. Exposure to consistent room reverberation results in improved AM detection thresholds, while exposure to variable rooms results in AM thresholds predicted by the room MTF (Zahorik & Anderson, 2013; Zahorik et al., 2012). Thus, enhanced AM sensitivity after consistent exposure counteracts the modulation depth reductions caused by reverberation and improves speech perception.

MTF-based accounts assume that the MTF can be perfectly extracted from the room. While this is feasible using analytical measurement techniques, when the probing signal is speech, it may be more difficult to accurately estimate the MTF, due to interactions between the modulation characteristics of the speech signal and those of the room (e.g., Payton et al., 2002). A further challenge to the MTF-based accounts is that adaptation appears to be critically sensitive to the time-direction of reverberation. For example, when reverberation is time-reversed, preceding the direct path energy, adaptation breaks down even though the modulation relative to a time-forward condition is approximately the same (MTF and STI almost identical;

Longworth-Reed et al., 2009). These and similar results prompted Watkins to suggest that a critical temporal envelope cue is the prominence of the tails at sound offsets and at spectral transitions in auditory filters (Watkins et al., 2011). The importance of time-direction in auditory perceptual constancy phenomena has also been observed in loudness judgments tasks, in which listeners perceive stimuli with a slow attack and fast decay as being louder than temporally reversed versions of them, even though the energy is the same in both conditions (Stecker & Hafter, 2000). The observed perceptual suppression of the tail at the ends of sounds likely results from auditory perceptual constancy mechanisms interpreting it as an acoustic by-product of reverberation and effectively disregarding it to rely on the distal properties of the sound source (Stecker & Hafter, 2000; Watkins et al., 2011).

Nielsen and Dau (2010) argued that a forward modulation masking mechanism, not associated with reverberation, could explain the findings of Watkins (2005b). Specifically, the carrier with low reverberation contains stronger modulations and masks the modulations present in the highly reverberant target. To test this hypothesis, they repeated basic aspects of the experiment by Watkins (2005b), introducing more carriers, including non-reverberated modulated and unmodulated speech-shaped noise. They showed that, relative to the two non-reverberant carriers, the modulated noise carrier tended to produce a shifted boundary in the "stir" responses compared to the same unmodulated carrier. This suggests that the proposed effect relates to the modulation content of the carrier, rather than its reverberation. However, subsequent experiments challenged this account (Beeston et al., 2014; Watkins & Raimond, 2013). For example, while the forward modulation masking hypothesis predicts that removing a preceding sentence would either reduce the masking of subsequent sounds, or cause no effect if masking was minimal, results showed the opposite pattern: when test words with strong reverberation were preceded by a silent context, more confusions were observed than when preceded by a carrier with matched strong reverberation (Beeston et al., 2014). These results suggest that at least part of the explanation must be attributed to information extracted from the level of reverberation of the carrier and target words.

Acoustic Cue Reweighting

Stilp et al. (2016) have drawn on concepts and findings from research on *spectral calibration*, whereby listeners perceptually suppress stable spectral cues in an acoustic environment, and give more weight to varying, more informative cues (e.g., Alexander & Kluender, 2010; Kiefte & Kluender, 2008). Reverberation introduces predictable spectrotemporal alterations to speech sounds, for example, by smearing across time spectral peaks that are useful for distinguishing a phoneme. Stilp et al. (2016) hypothesized that in such

conditions, acoustic cue reweighting will be even stronger, that is, there will be even stronger de-weighting of the stable cues, and increased reliance on non-predictable cues, compared to a condition without reverberation. To test this, they first estimated the relative perceptual weight of the second formant (F2) and the spectral tilt for the identification of isolated target vowels varying from /i/ to /u/. Next, they introduced precursor sentences filtered such that energy was enhanced near the center frequency of the second formant (F2) of the upcoming target vowel. As expected, this manipulation induced perceptual re-calibration such that listeners decreased perceptual weight for F2 and increased the weight for spectral tilt. Importantly, when simulated reverberation was applied to the same sentences, which spread the stable spectral energy for F2 across time, reweighting was even stronger. Unlike temporal envelope processing, this type of compensation does not appear to rely on reverberation tails, or reverberation per se, as removing the tails or presenting a tone that matched the target vowel's F2 instead of reverberation also induced cue reweighting. Overall, this mechanism takes an information-processing perspective on adaptation that emphasizes the unpredictable, information-bearing cues in the acoustic environment (Kluender et al., 2019; Stilp, 2020).

Tuning to Statistical Regularities of the Reverberation

The studies presented so far suggest that experience within a particular acoustic environment benefits speech perception in that environment. However, in everyday communication listeners encounter numerous acoustic spaces, with vastly different geometries, surface materials, and configurations. Are there structured components to this variability that listeners could leverage to separate the contributions of environmental filters and sound sources? Traer and McDermott (2016) conducted a large-scale statistical analysis of naturally occurring BRIRs, drawing random samples from the distribution of acoustic environments in which listeners typically spend their time. They analyzed 271 impulse responses of these acoustic spaces, including city streets, restaurants, parks, and offices. Their analyses showed that impulse responses were characterized by robust statistical regularities: (a) a transition from high kurtosis, produced by sparse early reflections, to Gaussian statistical properties within ~50 ms of the direct sound arrival, (b) an exponential decay of the reverberant tail, (c) frequency-dependent decay rates, and (d) decay rates that are more frequency-dependent in stronger reverberation. These characteristics were qualitatively similar for both indoor and outdoor spaces. Importantly, perceptual experiments revealed that listeners relied heavily on these regularities. For example, when the source was convolved with synthetic impulse responses that violated the statistical constraints, for example, by exhibiting a linear rather than exponential decay in the reverberant tail, listeners easily detected that what they heard was

"unnatural". Also, when listeners had to discriminate between sounds, they were less able to do so when the sources were convolved with atypical impulse responses.

These results suggest that, underlying the vast diversity in the acoustic spaces which people encounter daily, there are tight statistical regularities on which listeners rely. Although this study did not explicitly address adaptation to reverberant speech, it provides insights into how listeners adapt to diverse acoustic spaces. For example, it shows that a large majority of the indoor reverberant spaces have T_{60} below 1 s, suggesting that the reverberation adaptation mechanism should be preferably tuned to such T_{60}'s to optimize for the most common environments, consistent with the behavioral results reviewed here. Also, being able to capitalize on priors means that the perceptual system does not need to start from scratch in every new acoustic space, and that it can adjust rapidly and flexibly in unfamiliar and novel spaces, as long as these do not violate prior constraints. Brief exposure to a particular room could further refine these priors, allowing more effective speech recalibration.

Neural Mechanisms

Many studies have attempted to shed light on the neural mechanisms that support sound localization and speech recognition in reverberation (e.g., Barzelay et al., 2023; Devore & Delgutte, 2010; Devore et al., 2009; Ivanov et al., 2022; Kim et al., 2015; Kuwada et al., 2014; Slama & Delgutte, 2015). Here, we briefly summarize some of the studies that examine neural adaptation to reverberant speech.

Animal studies show advanced temporal coding of AM in reverberation in the inferior colliculus of unanesthetized rabbits (e.g., Kuwada et al., 2014; Slama & Delgutte, 2015). These studies show that, while reverberation degrades the temporal coding of AM, for most neurons the amount of degradation is less pronounced than the AM attenuation in the stimulus (Kuwada et al., 2014; Slama & Delgutte, 2015). Further, Slama and Delgutte (2015) reported that, in a subset of neurons, the temporal coding of AM was better for reverberant stimuli than for anechoic stimuli with equivalent modulation depth at the ear.

Recent research has centered on mechanisms that enable reverberation-invariant neural representations at the level of the auditory cortex (Fuglsang et al., 2017; Ivanov et al., 2022; Mesgarani et al., 2014). Ivanov et al. (2022) found that in anesthetized ferrets, neurons in the auditory cortex adapt to reverberation by increasing the latency of inhibitory components in their spectro-temporal receptive fields, consistent with predictions of a normative linear dereverberation model. Mesgarani et al. (2014) employed stimulus reconstruction techniques to derive the spectrographic representations of stimuli from neural responses across different conditions including anechoic, noisy, and reverberant environments. They showed that reconstructed spectrograms from responses of neural populations in the primary auditory cortex of awake ferrets resembled the spectrogram of the clean signal (devoid of noise or reverberation) more closely than the spectrograms of noisy or reverberant signals. A dynamic nonlinear model that combined synaptic depression and gain normalization was able to best account for the results. It is unclear whether functionally similar mechanisms are present at subcortical areas that provide input to the cortex. A recent study that used stimulus reconstruction techniques reported no evidence for a reverberation compensation mechanism in the IC of unanesthetized rabbits (Barzelay et al., 2023).

Finally, a study by Fuglsang et al. (2017) examined envelope tracking of attended versus unattended speech streams in human participants in complex listening situations with multiple talkers and reverberation. Results showed that envelope tracking of the attended speech was robust to distortions across all conditions, even in strong reverberation. Decoding of the unattended talker, on the other hand, deteriorated in strong reverberation. Importantly, for the attended talker the neural responses to highly reverberant speech resembled the original clean signal more than the distorted signal that was actually presented to the participants. These results suggest that, in real-life acoustic situations with multiple talkers and reverberation, selective attention modulates the cortical entrainment of speech envelope and might promote the formation of reverberation-robust neural representations of speech.

Conclusions and Directions for Future Research

Reports on the effects of reverberation on speech intelligibility can be traced back to nearly a century ago (e.g., Knudsen, 1929), but it was only over the past two decades that researchers have begun to elucidate how listeners adapt to room acoustics to recalibrate speech perception. The goal of this review was to summarize the current state of this research. A consistent picture that emerges under a wide range of experimental procedures, spanning diverse speech stimuli and tasks, is that listeners rapidly and efficiently exploit information from the preceding acoustic context to improve speech perception in reverberation.

Various characteristics of the preceding room acoustics can profoundly affect the buildup, or disruption, of the adaptation, which depends on source-listener distance and the correlated acoustic measure of DRR (e.g., Watkins, 2005b). It appears to be strongest in moderately reverberant target rooms (T_{60}'s between 0.4 and 1 s), diminishing at larger T_{60}'s (Brandewie & Zahorik, 2013; Vlahou et al., 2021; Zahorik & Brandewie, 2016; Zahorik, 2019). Less emphasis has been given to the disruptive effects of inconsistent carriers relative to the beneficial effects of consistent carriers (e.g., Brandewie & Zahorik, 2018; Vlahou et al., 2021). Inconsistent carriers can significantly disrupt performance, even below a baseline condition where the target speech is

presented alone (Vlahou et al., 2021), but the magnitude of the disruption can vary depending on the characteristics of the carrier and target (Vlahou et al., 2021; Brandewie & Zahorik, 2018). In typical everyday communication, rooms do not change abruptly; therefore, a sudden change to a different simulated room represents a violation of expectations that the perceptual system must overcome (Traer & McDermott, 2016). The impact of inconsistent carriers becomes particularly pertinent in AR/VR applications and poses a challenge in delivering consistent reverberant speech congruent with real environments (Best et al., 2020).

It is unclear whether adaptation relies on monaural or binaural input. Watkins and colleagues have repeatedly demonstrated robust adaptation in conditions with monaural presentation of speech (e.g., Beeston et al., 2014; Watkins, 2005a, 2005b; Watkins et al., 2011), while Zahorik and colleagues have shown very limited benefits without binaural presentation (Brandewie & Zahorik, 2010). The reason for this discrepancy remains unclear. Binaural and monaural presentation is likely to activate different compensation mechanisms. While energetically, monaural and binaural reverberation processing might be similar, and studies suggest that monaural reverberation information is sufficient, for example, for distance perception (Kopčo & Shinn-Cunningham, 2011), binaural processing interacts with reverberation processing in a time-dependent manner. For example, the interaural cross-correlation decreases over time for a stimulus in reverberation (Vlahou et al., 2021), which might result in improved ability of binaural processing to act on the initial, correlated portions of each utterance, but less so on the later, uncorrelated portions. In everyday listening, listeners regularly encounter noisy environments, with multiple talkers speaking simultaneously. In such conditions, binaural input might be necessary. More generally, reverberation is perceived binaurally in all real environments. Clearly, this is an area where more research is needed.

Several studies reviewed here have used concurrent presentation of spatialized noise with the speech stimuli (Zahorik and colleagues, see Table 1). On the one hand, this configuration more accurately reflects everyday listening environments, which commonly include both noise and reverberation. On the other hand, consonant perception shows different patterns of errors under conditions of noise, reverberation, or noise and reverberation; for example, while word-final stop consonants are particularly affected by reverberation, word-final fricatives are affected more by noise (Helfer, 1994; Helfer & Huntley, 1991). Further, this configuration might have introduced additional factors, in addition to the primary task of speech recognition, such as sound localization and spatial unmasking (Beeston et al., 2014), making the interpretation of these results challenging. For example, since the target and masker were at different locations, the mechanisms of spatial release from masking (SRM) are likely to have contributed to target speech identification (Bronkhorst, 2000). And, since the amount of SRM decreases with reverberation (Leclère et al., 2015), SRM can differentially influence the observed effects in different rooms in the noise-masking studies, possibly interacting with any reverberation compensation mechanism (Vlahou et al., 2021).

Adaptation to reverberation has been examined for different speech units, from phonemes and syllables (e.g., Watkins, 2005a, 2005b) to ecologically realistic variable sentences (e.g., Srinivasan & Zahorik, 2014). At the segmental level, there is evidence that some of the sounds that are more severely affected by reverberation can be improved with prior consistent exposure (Beeston et al., 2014; Vlahou et al., 2021). Moreover, the effect of reverberation is much more pronounced for the final than initial consonants within a word (Vlahou et al., 2021). The effect persists for phonetically balanced words in closed-set corpora with limited vocabulary size, like in the CRM, to highly heterogeneous material as in PRESTO and TIMIT databases (Garofolo et al., 1993; Gilbert et al., 2013). The open set studies make it difficult to identify whether adaptation has a more pronounced impact on specific speech units and phonetic features, since studies using this paradigm have focused on words and sentences rather than individual phonemes. On the other hand, by incorporating both diverse material and noise, this design better mirrors real-world listening.

Adaptation to reverberation is not specific to speech perception. For example, a recent study showed that, while reverberation affects the identification of material such as wood, metal, and glass, the effect is smaller when listeners are exposed to consistent reverberation compared to when the reverberation varies randomly (Koumura & Furukawa, 2017). Shinn-Cunningham (2000) also showed continuous improvement in sound localization after continuous exposure to consistent reverberation in a room. However, in contrast to Shinn-Cunningham's (2000) study, which reveals a more nuanced learning process for sound localization, research on speech perception consistently indicates a rapid timescale. Various studies demonstrate both monaural and binaural adaptation occurring in less than a second (e.g., Brandewie & Zahorik, 2013; Beeston et al., 2014; Vlahou et al., 2021), although in more challenging acoustic environments characterized by stronger reverberation and spatialized noise it might require exposure three times as long (Brandewie & Zahorik, 2013). However, there are important differences between the tasks of sound localization and speech perception. Primarily, localization is inaccurate compared to speech perception, especially in the distance dimension considered in the Shinn-Cunningham study. In speech perception, near-perfect accuracy is common in everyday communication in one's native language, except when the environment is noisy or otherwise challenging. Thus, there is much more room for improvement in localization over long periods of time, while speech perception needs to be accurate very quickly.

Preliminary simulations from Zahorik (2019) suggest that MTF estimation becomes fully developed after approximately 1 s of exposure. This could potentially explain the observed timescale of adaptation, given that estimating and enhancing AMs in a room appears to be a critical component of the adaptation process (Zahorik, 2019; Zahorik & Anderson, 2013). However, this modeling has not been performed on more challenging conditions like, for example, those with strong reverberation and no masking noise of Vlahou et al. (2021).

An important area for future research concerns how adaptation proceeds in situations with multiple talkers, in which listeners' ability to cope with reverberation is much more affected compared to situations with just one voice (Culling et al., 2003) and in which spatial attentional selection of the target is often dynamic (Best et al., 2008). In such scenarios, reverberation disrupts intelligibility by degrading the target speech and also by decorrelating the signal at the two ears from the interferer, thus reducing the ability of the auditory system to take advantage of spatially separated sources (Lavandier & Culling, 2008). More insights on these topics would be particularly informative as they relate to everyday listening situations and slow adaptation on the time scale of seconds has been observed in particular in the attention studies.

Reverberation profoundly shapes the perceived ambiance of a listening space, by increasing the perceived spaciousness of a room, and by enhancing the subjective realism and externalization experienced in simulated auditory environments—an aspect crucial for immersive applications (Best et al., 2020; Shinn-Cunningham, 2000). Reverberation poses challenges for individuals with hearing impairments, and it impacts the performance of automatic speech recognition devices (Yoshioka et al., 2012). Further investigation is warranted to understand how listeners benefit from consistent room exposure and counteract the disruptive effects of inconsistent carriers when processing speech in reverberation. Such insights are crucial for advancing immersive AR/VR applications and developing prosthetic devices for the hearing impaired (Mason & Kokkinakis, 2014; Reinhart et al., 2016) as such devices can present stimuli in an environment inconsistent with the current listening environment.

Finally, main questions to be addressed in future research on adaptation to reverberation for speech perception include the following: (1) what acoustic characteristics of reverberation are used in the adaptation process and how are they estimated; (2) is the adaptation room-specific (e.g., based on T_{60}) or distance-specific (e.g., based on DRR); (3) can a unified theory of adaptation to reverberation be developed that would incorporate the hypothesized mechanisms of adaptation and provide predictions for the available data; (4) how to design communication and prosthetic applications that allow adaptation to reverberation to enhance communication rather than disrupting it, for example, when listening to natural speech mixed with speech delivered via a hearing aid, a cochlear implant, or an virtual/augmented reality device.

ORCID iDs

Eleni Vlahou 🆔 https://orcid.org/0000-0003-2497-577X
Norbert Kopčo 🆔 https://orcid.org/0000-0001-5901-6355

Supplemental Material

Supplemental material for this paper is available online.

References

Alexander, J. M., & Kluender, K. R. (2010). Temporal properties of perceptual calibration to local and broad spectral characteristics of a listening context. *Journal of the Acoustical Society of America*, *128*(6), 3597–3613. https://doi.org/10.1121/1.3500693

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In S. Greenberg, W. A. Ainsworth, A. N. Popper, & R. R. Fay (Eds.), *Speech processing in the auditory system* (pp. 231–308). Springer. https://doi.org/10.1007/0-387-21575-1_5

Barzelay, O., David, S., & Delgutte, B. (2023). Effect of reverberation on neural responses to natural speech in rabbit auditory midbrain: No evidence for a neural dereverberation mechanism. *eNeuro*, *10*(5), ENEURO.0447-22.2023. https://doi.org/10.1523/eneuro.0447-22.2023

Beeston, A. V., Brown, G. J., & Watkins, A. J. (2014). Perceptual compensation for the effects of reverberation on consonant identification: Evidence from studies with monaural stimuli. *Journal of the Acoustical Society of America*, *136*(6), 3072–3084. https://doi.org/10.1121/1.4900596

Best, V., Baumgartner, R., Lavandier, M., Majdak, P., & Kopčo, N. (2020). Sound externalization: A review of recent research. *Trends in Hearing*, *24*, https://doi.org/10.1177/2331216520948390

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences*, *105*(35), 13174–13178. https://doi.org/10.1073/pnas.0803718105

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal*

of the Acoustical Society of America, 107(2), 1065–1066. https://doi.org/10.1121/1.428288

Bradley, J. S., Sato, H., & Picard, M. (2003). On the importance of early reflections for speech in rooms. *Journal of the Acoustical Society of America, 113*(6), 3233–3244. https://doi.org/10.1121/1.1570439

Brandewie, E. J., & Zahorik, P. (2010). Prior listening in rooms improves speech intelligibility. *Journal of the Acoustical Society of America, 128*(1), 291–299. https://doi.org/10.1121/1.3436565.

Brandewie, E., & Zahorik, P. (2011). Adaptation to room acoustics using the modified rhyme test. *Proceedings of Meetings on Acoustics. Acoustical Society of America, 129*, 2487. https://doi.org/10.1121/1.3588198

Brandewie, E. J., & Zahorik, P. (2013). Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments. *Journal of the Acoustical Society of America, 134*(2), EL265–EL270. https://doi.org/10.1121/1.4816263

Brandewie, E. J., & Zahorik, P. (2018). Speech intelligibility in rooms: Disrupting the effect of prior listening exposure. *Journal of the Acoustical Society of America, 143*(5), 3068–3078. https://doi.org/10.1121/1.5038278

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica, 86*(1), 117–128.

Bronkhorst, A., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature, 397*(6719), 517–520. https://doi.org/10.1038/17374

Culling, J. F., Hodder, K. I., & Toh, C. Y. (2003). Effects of reverberation on perceptual segregation of competing voices. *Journal of the Acoustical Society of America, 114*(5), 2871–2876. https://doi.org/10.1121/1.1616922

Devore, S., & Delgutte, B. (2010). Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: Influences of interaural time and level differences. *Journal of Neuroscience, 30*(23), 7826–7837. https://doi.org/10.1523/JNEUROSCI.5517-09.2010

Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., & Delgutte, B. (2009). Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain. *Neuron, 62*(1), 123–134. https://doi.org/10.1016/j.neuron.2009.02.018

Eddins, D. A., & Liu, C. (2012). Psychometric properties of the coordinate response measure corpus with various types of background interference. *Journal of the Acoustical Society of America, 131*(2), EL177–EL183. https://doi.org/10.1121/1.3678680

Fuglsang, S. A., Dau, T., & Hjortkjaer, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage, 156*, 435–444. https://doi.org/10.1016/j.neuroimage.2017.04.026

Garofolo, J., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., & Zue, V. (1993). *DARPA TIMIT acoustic–phonetic continuous speech corpus*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.4930

Gelfand, A. S., & Silman, S. (1979). Effects of small room reverberation upon the recognition of some consonant features. *Journal of the Acoustical Society of America, 66*(1), 22–29. https://doi.org/10.1121/1.383075

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability

sentence recognition test. *Journal of the American Academy of Audiology, 24*(01), 26–36. https://doi.org/10.3766/jaaa.24.1.4

Helfer, K. S. (1994). Binaural cues and consonant perception in reverberation and noise. *Journal of Speech and Hearing Research, 37*(2), 429–438. https://doi.org/10.1044/jshr.3702.429

Helfer, K. S., & Huntley, R. A. (1991). Aging and consonant errors in reverberation and noise. *Journal of the Acoustical Society of America, 90*(4), 1786–1796. https://doi.org/10.1121/1.401659

House, A. S., Williams, C., Heker, M. H., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America, 37*(1), 158–166. https://doi.org/10.1121/1.1909295

Houtgast, T., & Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America, 77*(3), 1069–1077. https://doi.org/10.1121/1.392224

IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17*(3), 225–246. https://doi.org/10.1109/TAU.1969.1162058

Ivanov, A. Z., King, A. J., Willmore, B. D. B., Walker, K. M. M., & Harper, N. S. (2022). Cortical adaptation to sound reverberation. *eLife, 11*, e75090. https://doi.org/10.7554/eLife.75090

Jakien, K. M., Kampel, S. D., Stansell, M. M., & Gallun, F. J. (2017). Validating a rapid, automated test of spatial release from masking. *American Journal of Audiology, 26*(4), 507–518. https://doi.org/10.1044/2017_AJA-17-0013

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*(5), 1337–1351. https://doi.org/10.1121/1.381436

Kiefte, M., & Kluender, K. R. (2008). Absorption of reliable spectral characteristics in auditory perception. *Journal of the Acoustical Society of America, 123*(1), 366–376. https://doi.org/10.1121/1.2804951

Kim, D., Zahorik, P., Carney, L. H., Bishop, B. B., & Kuwada, S. (2015). Auditory distance coding in rabbit midbrain neurons and human perception: Monaural amplitude modulation depth as a cue. *Journal of Neuroscience, 35*(13), 5360–5372. https://doi.org/10.1523/JNEUROSCI.3798-14.2015

Kirsch, C., Poppitz, J., Wendt, T., van de Par, S., & Ewert, S. D. (2021). Spatial resolution of late reverberation in virtual acoustic environments. *Trends in Hearing, 25*, 233121652110549. https://doi.org/10.1177/23312165211054924

Kluender, K. R., Stilp, C. E., & Llanos, F. (2019). Longstanding problems in speech perception dissolve within an information-theoretic perspective. *Attention, Perception, & Psychophysics, 81*(4), 861–883. https://doi.org/10.3758/s13414-019-01702-x

Knudsen, V. O. (1929). The hearing of speech in auditoriums. *Journal of the Acoustical Society of America, 1*(1_Supplement), 30. https://doi.org/10.1121/1.1901869

Kopčo, N., & Shinn-Cunningham, B. G. (2011). Effect of stimulus spectrum on distance perception for nearby sources. *Journal of the Acoustical Society of America, 130*(3), 1530–1541. https://doi.org/10.1121/1.3613705

Koumura, T., & Furukawa, S. (2017). Context-dependent effect of reverberation on material perception from impact sound. *Scientific Reports, 7*(1), 16455. https://doi.org/10.1038/s41598-017-16651-4

Kuwada, S., Bishop, B., & Kim, D. O. (2014). Azimuth and envelope coding in the inferior colliculus of the unanesthetized rabbit: Effect of reverberation and distance. *Journal of Neurophysiology*, *112*(6), 1340–1355. https://doi.org/10.1152/jn.00826.2013

Larsen, E., Iyer, N., Lansing, C. R., & Feng, A. S. (2008). On the minimum audible difference in direct-to-reverberant energy ratio. *Journal of the Acoustical Society of America*, *124*(1), 450–461. https://doi.org/10.1121/1.2936368

Lavandier, M., & Culling, J. F. (2008). Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer. *Journal of the Acoustical Society of America*, *123*(4), 2237–2248. https://doi.org/10.1121/1.2871943

Leclère, T., Lavandier, M., & Culling, J. F. (2015). Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation. *Journal of the Acoustical Society of America*, *137*(6), 3335–3345. https://doi.org/10.1121/1.4921028

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *Journal of the Acoustical Society of America*, *106*(4), 1633–1654. https://doi.org/10.1121/1.427914

Longworth-Reed, L., Brandewie, E., & Zahorik, P. (2009). Time-forward speech intelligibility in time-reversed rooms. *Journal of the Acoustical Society of America*, *125*(1), EL13–EL19. https://doi.org/10.1121/1.3040024

Mason, M., & Kokkinakis, K. (2014). Perception of consonants in reverberation and noise by adults fitted with bimodal devices. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1512–1520. https://doi.org/10.1044/2014_JSLHR-H-13-0127

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences*, *111*(18), 6792–6797. https://doi.org/10.1073/pnas.1318017111.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, *27*(2), 338–352. https://doi.org/10.1121/1.1907526

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, *9*(4), 399–406. https://doi.org/10.1007/s10162-008-0143-x

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Nábělek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and nonnative listeners. *Journal of the Acoustical Society of America*, *75*(2), 632–634. https://doi.org/10.1121/1.390495

Nábělek, A. K., Letowski, T. R., & Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *Journal of the Acoustical Society of America*, *86*(4), 1259–1265. https://doi.org/10.1121/1.398740

Nielsen, J. B., & Dau, T. (2010). Revisiting perceptual compensation for effects of reverberation in speech identification. *Journal of the Acoustical Society of America*, *128*(5), 3088–3094. https://doi.org/10.1121/1.3494508

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Osawa, E., Arai, T., & Hodoshima, N. (2018). Perception of Japanese consonant-vowel syllables in reverberation: Comparing nonnative listeners with native listeners. *Acoustical Science and Technology*, *39*(6), 369–378. https://doi.org/10.1250/AST.39.369.

Osawa, E., Hui, J. C. T., Hioka, Y., & Arai, T. (2021). Effect of prior exposure on the perception of Japanese vowel length contrast in reverberation for nonnative listeners. *Speech Communication*, *134*, 1–11. https://doi.org/10.1016/j.specom.2021.07.009

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & , … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*, n71, https://doi.org/10.1136/bmj.n71

Payton, K. L., Chen, S., & Braida, L. D. (2002). Comparison of approaches to estimate the speech modulation transfer function. *The Journal of the Acoustical Society of America*, *111*(5_Suppl), 2431–2431. https://doi.org/10.1121/1.4778339

Poissant, S. F., Whitmal, N. A., & Freyman, R. L. (2006). Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. *Journal of the Acoustical Society of America*, *119*(3), 1606–1615. https://doi.org/10.1121/1.2168428

Rakerd, B., Hartmann, W. M., & McCaskey, T. L. (1999). Identification and localization of sound sources in the median sagittal plane. *Journal of the Acoustical Society of America*, *106*(5), 2812–2820. https://doi.org/10.1121/1.428129

Reinhart, P. N., & Souza, P. E. (2018). Listener factors associated with individual susceptibility to reverberation. *Journal of the American Academy of Audiology*, *29*(01), 73–82. https://doi.org/10.3766/jaaa.16168

Reinhart, P. N., Souza, P. E., Srinivasan, N. K., & Gallun, F. J. (2016). Effects of reverberation and compression on consonant identification in individuals with hearing impairment. *Ear and Hearing*, *37*(2), 144–152. https://doi.org/10.1097/AUD.0000000000000229

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shinn-Cunningham, B. G. (2000). Learning reverberation: Considerations for spatial auditory displays. In P. Cook (Ed.), *Proceedings of the international conference on auditory display* (pp. 126–134). Georgia Institute of Technology.

Shinn-Cunningham, B. G. (2003). Acoustics and perception of sound in everyday environments. In N. Bianchi-Berthouze (Ed.), *Proceedings of the 3rd international workshop on spatial media* (pp. IWSM03-1–IWSM03-9). Springer.

Slama, M. C. C., & Delgutte, B. (2015). Neural coding of sound envelope in reverberant environments. *The Journal of Neuroscience*, *35*(10), 4452–4468. https://doi.org/10.1523/JNEUROSCI.3615-14.2015

Srinivasan, N. K., Tobey, E. A., & Loizou, P. C. (2016). Prior exposure to a reverberant listening environment improves speech intelligibility in adult cochlear implant listeners. *Cochlear Implants International*, *17*(2), 98–104. https://doi.org/10.1080/14670100.2015.1102455

Srinivasan, N., & Zahorik, P. (2011). The effect of semantic context on speech intelligibility in reverberant rooms. *Journal of the Acoustical Society of America*, *12*(4_Supplement), 060001. https://doi.org/10.1121/1.3588998

Srinivasan, N. K., & Zahorik, P. (2013). Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. *Journal of the Acoustical Society of America*, *133*(1), EL33–EL39. https://doi.org/10.1121/1.4771978

Srinivasan, N. K., & Zahorik, P. (2014). Enhancement of speech intelligibility in reverberant rooms: Role of amplitude envelope and temporal fine structure. *Journal of the Acoustical Society of America*, *135*(6), EL239–EL245. https://doi.org/10.1121/1.4874136

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *Journal of the Acoustical Society of America*, *107*(6), 3358–3368. https://doi.org/10.1121/1.429407

Stecker, G. C., & Moore, T. M. (2018). Reverberation enhances onset dominance in sound localization. *Journal of the Acoustical Society of America*, *143*(2), 786–793. https://doi.org/10.1121/1.5023221

Stilp, C. E. (2020). Acoustic context effects in speech perception. *WIRES Cognitive Science*, *11*(1), e1517. https://doi.org/10.1002/wcs.1517

Stilp, C. E., Anderson, P. W., Assgari, A. A., Ellis, G. M., & Zahorik, P. (2016). Speech perception adjusts to stable spectrotemporal properties of the listening environment. *Hearing Research*, *341*, 168–178. https://doi.org/10.1016/j.heares.2016.08.004

Takata, Y., & Nábĕlek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America*, *88*, 663–666. https://doi.org/10.1121/1.399769

Traer, J., & McDermott, J. H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences*, *113*(48), E7856–E7865. https://doi.org/10.1073/pnas.1612524113

Vlahou, E., Seitz, A. R., & Kopčo, N. (2019). Nonnative implicit phonetic training in multiple reverberant environments. *Attention, Perception, & Psychophysics*, *81*(4), 935–947. https://doi.org/10.3758/s13414-019-01680-0

Vlahou, E., Ueno, K., Shinn-Cunningham, B., & Kopčo, N. (2021). Calibration of consonant perception to room reverberation. *Journal of Speech Language and Hearing Research*, *64*(8), 2956–2976. https://doi.org/10.1044/2021_jslhr-20-00396

Watkins, A. J. (2005a). Listening in real-room reverberation: Effects of extrinsic context. In D. Pressnitzer, A. de Cheveigné, S. McAdams, & L. Collet (Eds.), *Auditory signal processing: Physiology, psychoacoustics, and models* (pp. 422–427). Springer. https://doi.org/10.1007/0-387-27045-0_52

Watkins, A. J. (2005b). Perceptual compensation for effects of reverberation in speech identification. *Journal of the Acoustical Society of America*, *118*(1), 249–262. https://doi.org/10.1121/1.1923369.

Watkins, A. J., & Makin, S. J. (2007). Steady-spectrum contexts and perceptual compensation for reverberation in speech identification. *Journal of the Acoustical Society of America*, *121*(1), 257–266. https://doi.org/10.1121/1.2387134

Watkins, A. J., Makin, S. J., & Raimond, A. P. (2010a). Constancy in the perception of speech when the level of room-reflections varies. In J. Buchholz, T. Dau, J. Dalsgaard, & T. Poulsen (Eds.), *Binaural processing and spatial hearing. ISAAR – international symposium on auditory and audiological research* (pp. 371–380). The Danavox Jubilee Foundation.

Watkins, A. J., & Raimond, A. P. (2013). Perceptual compensation when isolated test words are heard in room reverberation. In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), *Basic aspects of hearing: Physiology and perception* (pp. 193–201). Springer.

Watkins, A. J., Raimond, A., & Makin, S. J. (2010b). Room reflections and constancy in speech-like Sounds: within-band effects. In E. Lopez-Poveda, A. Palmer, & R. Meddis (Eds.), *The Neurophysiological Bases of Auditory Perception* (pp. 439–447). Springer. https://doi.org/10.1007/978-1-4419-5686-6_41

Watkins, A. J., Raimond, A. P., & Makin, S. J. (2011). Temporal envelope constancy of speech in rooms and the perceptual weighting of frequency bands. *Journal of the Acoustical Society of America*, *130*(5), 2777–2788. https://doi.org/10.1121/1.3641399

Yoshioka, T., et al. (2012). Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition. *IEEE Signal Processing Magazine*, *29*(6), 114–126. https://doi.org/10.1109/MSP.2012.2205029

Zahorik, P. (2019). Adaptation to room acoustics and its effect of speech understanding. Proceedings of the 23rd International Congress on Acoustics, Aachen, Germany, 9–13 September.

Zahorik, P., & Anderson, P. W. (2013). Amplitude modulation detection by human listeners in reverberant sound fields: Effects of prior listening exposure. *Proceedings of Meeting on Acoustics: Acoustical Society of America*, *19*, 050139. https://doi.org/10.1121/1.4800433

Zahorik, P., & Brandewie, E. (2011). Perceptual adaptation to room acoustics and effects on speech intelligibility in hearing-impaired populations. *Proceedings of Forum Acusticum*, 2167–2172. https://pubmed.ncbi.nlm.nih.gov/23455358

Zahorik, P., & Brandewie, E. (2016). Speech intelligibility in rooms: Effect of prior listening exposure interacts with room acoustics. *Journal of the Acoustical Society of America*, *140*(1), 74–86. https://doi.org/10.1121/1.4954723

Zahorik, P., Brungart, D., & Bronkhorst, A. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica United With Acustica*, *91*(3), 409–420. http://repository.tudelft.nl/view/tno/uuid%3A0258e01b-1126-4289-8d8b-b9cc394c55d8/

Zahorik, P., Kim, D. O., Kuwada, S., Anderson, P. W., Brandewie, E., Collecchia, R., & Srinivasan, N. (2012). Amplitude modulation detection by human listeners in reverberant sound fields: Carrier bandwidth effects and binaural versus monaural comparison. *Proceedings of Meetings on Acoustics. Acoustical Society of America*, *15*, 050002. https://doi.org/10.1121/1.4733848

Zahorik, P., Kim, D. O., Kuwada, S., Anderson, P. W., Brandewie, E., & Srinivasan, N. (2011). Amplitude modulation detection by human listeners in sound fields. *Proceedings of Meetings on Acoustics. Acoustical Society of America*, *12*, 50005–50010. https://doi.org/10.1121/1.3656342

Zahorik, P., & Wightman, F. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience*, *4*(1), 78–83. https://doi.org/10.1038/82931