PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Adaptation to room acoustics and its effect of speech understanding

Pavel ZAHORIK[1]

[1] University of Louisville, USA

ABSTRACT

Effective sound transmission between source and receiver is essential for good communication and sound quality in learning spaces. Of course, sound transmission can be significantly affected by the acoustics of the space. Under many circumstances, however, these acoustical effects have relatively minor perceptual consequences. This may be explained, in part, by listener adaptation to the acoustics of the listening environment. Here, evidence that room adaptation improves speech understanding is summarized. The adaptation, is rapid (around 1 second), strongest for rooms with reverberation times between 0.4 and 1 second, and observable for a variety of speech materials. It also appears to depend critically on the amplitude modulation characteristic of the signal reaching the ear. A better understanding of room adaptation effects can inform and contribute to methods for improving speech understanding and sound quality in rooms for both normally-hearing and hearing-impaired listeners.

Keywords: Speech Intelligibility, Room Acoustics, Virtual Auditory Space

1. BACKGROUND

It has long been known that speech understanding is degraded by reverberation (1). The degradation stems primarily from temporal distortion of the speech signal caused by the reverberation (2, 3), and is known to scale with the amount of reverberation (1, 4). Because most everyday communication situations involve sound transmission within a reverberant sound field, it is critically important to understand how and by what mechanisms speech understanding is impacted by reverberation. This is especially true, given that the negative effects of reverberation on speech understanding are exacerbated both by background noise (1, 5) – another ubiquitous property of everyday listening environments, and by hearing loss (6), where poor performance in reverberation is the most frequent complaint given by hearing aid users (7). Given these challenges, it is perhaps remarkable that for individuals with normal hearing, few communication problems are encountered in everyday reverberant environments. This suggests that processing in the normally functioning auditory system must effectively counteract the deleterious effects of reflected sound and reverberation, even though acoustically, these effects are clearly measurable and specific to a given listening environment and the spatial configuration of components in the communication chain.

The auditory system has a number of mechanisms that can immediately assist with speech understanding in reverberation, including the binaural system (8) and mechanisms related to the precedence effect (see 9, and 10 for reviews). Beyond these immediate effects, there is now emerging evidence that prior listening exposure to reverberation can provide an environmental listening context that renders speech as perceptually less reverberant (11, 12) and can result in objective improvements in speech intelligibility (13). This suggests that the processing of sound in reverberant space within the normally functioning auditory system may involve processes more complicated than previously thought. The goal of the present study is to determine the extent to which this environmental context effect depends on the acoustics of the listening space.

Watkins (11, 12) was the first to demonstrate an effect of listening context on speech perception in reverberation. He used target speech signals on an 11-point continuum from "sir" to "stir" embedded in a carrier phrase, and noted the point at which the speech percept changed from "sir" to "stir" – a categorical perception task. When both target and carrier phrase were presented in minimal reverberation, the change point was near the center of the continuum. When the target was presented in moderate reverberation, but the carrier phrase remained in minimal reverberation, the change point

[1] pavel.zahorik@louisville.edu

shifted toward “sir”. This can be explained by reverberant energy filling in the temporal gap following the stop consonant in “stir”, causing it to be perceived as more like “sir”. When both the target and the carrier phrase were presented in reverberation, the change point shifted back to where it was observed when both target and carrier were presented in minimal reverberation. This suggests that the reverberant carrier phrase provides contextual information that allows the auditory system to compensate for the effects of the reverberation on the target word. Watkins and his colleagues have interpreted this result as being consistent with a type of high-level perceptual constancy, similar to other well-known perceptual constancies in vision, such as brightness constancy or color constancy. They have also demonstrated the effect with additional speech continua (14) and non-speech contexts (15, 16), and have shown that the effect is driven primarily by the amplitude envelope of the speech signal reaching the listener (17). This latter result is appealing; because of its potential links to the modulation transfer function concept, which forms the basis for standard methods of predicting speech intelligibility in rooms, such as the speech-transmission index (STI, 18) and speech-intelligibility index (SII, 19).

Brandewie and Zahorik (13) reported context effects similar to those identified by Watkins and his colleagues, but using different methods. In their study, Brandewie and Zahorik (13) compared speech reception thresholds (SRTs) using the coordinate response measure (CRM, 20) in a background of spatially separate noise within a simulated reverberant room. Two different listening conditions were tested. In one condition, listeners were provided with consistent listening exposure to the same reverberant room, both within and across trials. In a second condition, consistent exposure to the room was disrupted by removing the CRM carrier phrase and by changing the room from trial to trial. SRTs were found to be 2-3 dB lower, on average, in the consistent exposure condition. This suggests that consistent environmental listening context in a reverberant room can facilitate speech understanding. Similar context effects were not observed when the test room was anechoic, suggesting that the context effects are specific to reverberant sound fields. Additional work has demonstrated that the effect generalizes to highly heterogeneous sentence materials (21), is fully activated by approximately 1 s of listening exposure (22), is strongest for rooms with moderate reverberation times ($0.4 \leq RT60 \leq 1$ s, 23), and is relatively independent of location within the room (24). The importance of the amplitude envelope in the room context effect has also been demonstrated using similar methods (25).

In addition to speech understanding, similar short-term adaptation to aspects of room acoustics has been observed in listening tasks such as: echo detection (26), sound localization (27, 28), judgments of loudspeaker sound quality (29), and amplitude modulation detection (30). In all cases, the adaptation results in what can be considered improvements in listening performance, thus clearly demonstrating the importance of this effect for natural, everyday listening that contains indirect sound. Although these effects may be specific examples of perceptual priming (31) in the auditory modality, it is at present not clearly understood what mechanisms or basic auditory capabilities underlie the adaptation. This is a critical gap in knowledge that is preventing further progress in this area.

2. NEXT STEPS

2.1 Theoretical Framework

Now that the existence of room adaptation effects have been clearly demonstrated for normal-hearing listeners, and certain basic characteristics of the effects have been identified, there is a critical need to reveal the basic auditory functions that underlie these effects and how they may interact with other known effects (e.g. Binaural Squelch) to facilitate improved speech understanding in reverberation. It is hypothesized that both monaural and binaural processing of the amplitude modulation (AM) information underlies these room adaptation effects through the conceptual framework shown in Figure 1. The basis for this framework is the modulation transfer function (MTF) concept (3), which, along with its binaural extension (32), has been shown to be highly effective at predicting speech intelligibility in complex listening situations including reverberant rooms with background noise. To extend the MTF-concept to explain room adaptation, we hypothesize an additional stage that provides compensation for the MTF of the room. Through the combination of room MTF estimation and compensation, information about the AM characteristics of the sound source may be gained over time, which is independent from the acoustical effects of the room. In general, this is a type of process that seeks inference regarding distal source properties in the face of degraded or underspecified proximal sensory input. Such processes are thought to support a variety of perceptual phenomenon, including the constancies of brightness (33), size (34), and loudness (35). Neural evidence demonstrating the reverberation-resistant coding of AM in the inferior colliculus (36, 37) supports the ideas of MTF estimation and compensation in our framework, as does the fact that room adaptation effects on speech understanding and perception have been shown to depend specifically on the time-varying amplitude envelope of the speech signal (17, 25). Further work will be needed to test hypotheses stemming from this conceptual framework.

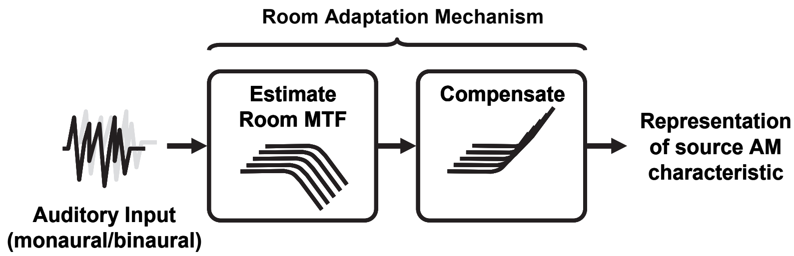

Figure 1 **-** Proposed conceptual framework to explain room adaptation.

2.2 Monaural Versus Binaural Contributions

A fundamental issue that must be addressed is the role of monaural versus binaural input. Work by Watkins (11) has clearly demonstrated monaural compensation for reverberation, and monaural cells in the inferior colliculus have been identified that are reverberation-resistant (36, 38). On the other hand, Brandewie and Zahorik (13) have concluded that room adaptation appears to require binaural input. This conclusion may be premature for at least two reasons, however. 1) Two of 14 listeners still showed adaptation effects, even for monaural input, and 2) monaural input was always tested for using the ear contralateral to the noise source (the “better” ear, in terms of signal-to-noise ratio), which may have influenced the results. Additional studies are therefore needed to clarify the role of monaural versus binaural input in the observed room adaptation effects.

2.3 Room MTF Development over Time

We have operationalized the conceptual framework described above for the case of monaural room impulse response (RIR) input by first passing this input through a gammatone filterbank (39) to simulate peripheral auditory filtering and then computing the MTF in each band from the RIR using a method described by Schroeder (5). We then examined the effect of exposure time by varying the length of the RIR used for analysis, ranging from the first 5 ms (near anechoic) to the entire RIR, which was approximately 1.8 s in this case. Results of this analysis are shown in Figure 2, where a clear development of the room MTF characteristic is evident over time. By 1000 ms of exposure, the room MTF is fully developed in all but the lowest modulation frequencies. We suggest that this strategy of analyzing the developmental time course of MTF estimates may help explain time-course results for room adaptation effects, where enhancement for a similar room was found to increase as a function of exposure time up to about 1 s and then plateau (22). This suggestion is based on the idea that compensation in our hypothesized framework would be incomplete without an accurate estimate of the room MTF, which is only available after sufficient exposure time. These preliminary data have important implications, because they suggest that the room adaptation effects might be explainable based on acoustical properties of the MTF that develop over time, and on relatively low-level auditory processes that encode these properties. This approach will be especially relevant because like the STI (3), it can be adapted to take running speech as an input. Further scientific study will be needed to definitively determine if such temporal development of the MTF is related to the effects of room adaptation on speech understanding.

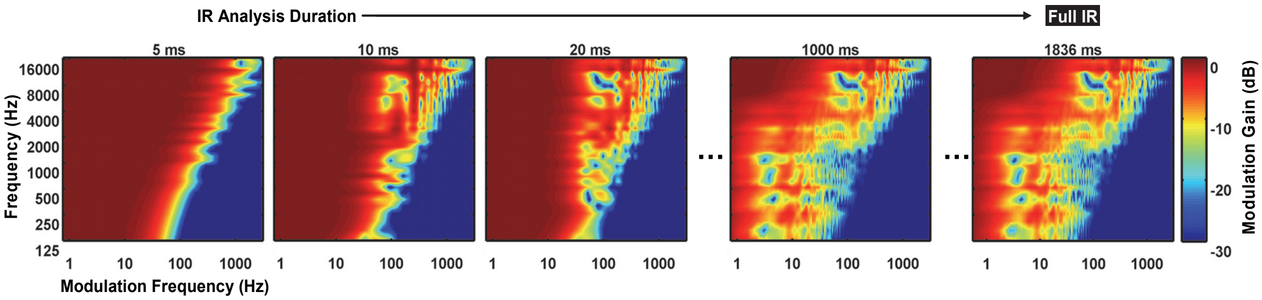

Figure 2 - Room modulation transfer functions (MTFs) as a function of auditory filter center frequency and analysis window length. MTFs were derived from the impulse response (IR) at the left ear for a frontal sound source (1.4 m distance) in a simulated reverberant room (broadband $RT60$ = 1.8 s).

2.4 Impact of Hearing Loss

Results of a preliminary study (40) using testing methods identical to those used by Brandewie and Zahorik (13) found that some hearing impaired users may also exhibits room adaptation effects related to speech understanding, although large individual differences were observed. These differences were not clearly related to the large differences in pure-tone audiometric thresholds for the sample of listeners tested. Therefore, additional study will be required to fully understand how hearing loss may impact room adaptation effects.

3. IMPLICATIONS FOR ROOM DESIGN/COMFORT

Although most natural listening situations would allow for listening experience with an acoustical environment of sufficient time to activate the described room adaptation effects, there are a number of important complexities related to this adaptation that are not yet known. For example, how might the psychophysical effects of room adaptation be impacted by other devices in the communication chain, such as sound reinforcement systems, sound assistive devices (e.g. hearing aids), or teleconferencing systems? As pointed out in the STI standard (18), such devices in the communication chain affect the MTF. To the extend that the MTF is involved in psychophysical room adaptation, such devices may impact adaptation. There is also the issue of different or conflicting room acoustic information potentially being conveyed through the device. This may also impact room adaptation beyond the MTF. Additional research will be needed to explore these implications.

ACKNOWLEDGEMENTS

This work was supported in part by the University of Louisville and the NIH (R01DC008168).

REFERENCES

1. Knudsen VO. The hearing of speech in auditoriums. Journal of the Acoustical Society of America. 1929;1(1):56-82.
2. Bolt RH, MacDonald AD. Theory of speech masking by reverberation. Journal of the Acoustical Society of America. 1949;21(6):577-80.
3. Houtgast T, Steeneken HJM. 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. 1985;77(3):1069-77.
4. Lochner JPA, Burger JF. The influence of reflections on auditorium acoustics. Journal of Sound and Vibration. 1964;1(4):426 - 54.
5. Nabelek AK, Mason D. Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms. Journal of Speech, Language, and Hearing Research. 1981;24(3):375-83.
6. Gelfand SA, Hochberg I. Binaural and monaural speech discrimination under reverberation. Audiology. 1976;15(1):72-84.
7. Johnson JA, Cox RM, Alexander GC. Development of APHAB norms for WDRC hearing aids and comparisons with original norms. Ear and Hearing. 2010;31(1):47-55.
8. Moncur JP, Dirks D. Binaural and monaural speech intelligibility in reverberation. Journal of Speech

and Hearing Research. 1967;10(2):186-95.
9. Litovsky RY, Colburn HS, Yost WA, Guzman SJ. The precedence effect. Journal of the Acoustical Society of America. 1999;106(4 Pt 1):1633-54.
10. Brown AD, Stecker GC, Tollin DJ. The precedence effect in sound localization. Journal of the Association for Research in Otolaryngology. 2015;16(1):1-28.
11. Watkins AJ. Perceptual compensation for effects of reverberation in speech identification. Journal of the Acoustical Society of America. 2005;118(1):249-62.
12. Watkins AJ. Perceptual compensation for effects of echo and of reverberation on speech identification. Acta Acustica united with Acustica. 2005;91(5):892-901.
13. Brandewie E, Zahorik P. Prior listening in rooms improves speech intelligibility. Journal of the Acoustical Society of America. 2010;128(1):291-9.
14. Beeston AV, Brown GJ, Watkins AJ. Perceptual compensation for the effects of reverberation on consonant identification: evidence from studies with monaural stimuli. Journal of the Acoustical Society of America. 2014;136(6):3072-84.
15. Watkins AJ, Makin SJ. Steady-spectrum contexts and perceptual compensation for reverberation in speech identification. Journal of the Acoustical Society of America. 2007;121(1):257-66.
16. Watkins AJ, Makin SJ. Perceptual compensations for reverberation in speech identification: Effects of single-band, multiple-band and wideband noise contexts. Acta Acustica united with Acustica. 2007;93:403-10.
17. Watkins AJ, Raimond AP, Makin SJ. Temporal-envelope constancy of speech in rooms and the perceptual weighting of frequency bands. Journal of the Acoustical Society of America. 2011;130(5):2777-88.
18. IEC-60268-16. Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index. Geneva: International Electrotechnical Commission, 2003.
19. ANSI-S3.5-1997. Methods for calculation of the speech intelligibility index. New York: American National Standards Institute, 1997.
20. Bolia RS, Nelson WT, Ericson MA, Simpson BD. A speech corpus for multitalker communications research. Journal of the Acoustical Society of America. 2000;107(2):1065-6.
21. Srinivasan NK, Zahorik P. Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. Journal of the Acoustical Society of America. 2013;133(1):EL33-9.
22. Brandewie E, Zahorik P. Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments. Journal of the Acoustical Society of America. 2013;134(2):EL265-70.
23. Zahorik P, Brandewie EJ. Speech intelligibility in rooms: Effect of prior listening exposure interacts with room acoustics. Journal of the Acoustical Society of America. 2016;140(1):74.
24. Brandewie EJ, Zahorik P. Speech intelligibility in rooms: Disrupting the effect of prior listening exposure. Journal of the Acoustical Society of America. 2018;143(5):3068.
25. Srinivasan NK, Zahorik P. Enhancement of speech intelligibility in reverberant rooms: role of amplitude envelope and temporal fine structure. Journal of the Acoustical Society of America. 2014;135(6):EL239-45.
26. Clifton RK, Freyman RL, Meo J. What the precedence effect tells us about room acoustics. Perception and Psychophysics. 2002;64(2):180-8.
27. Kopco N, Schoolmaster M, Shinn-Cunningham BG. Learning to judge distance of nearby sounds in reverberant and anechoic environments. Proceedings of the Joint Congress CFA/DAGA '04, Strasbourg, France. 2004.
28. Shinn-Cunningham BG. Learning reverberation: Considerations for spatial auditory displays. Proceedings of the International Conference on Auditory Display, Atlanta, GA. 2000:126-34.
29. Toole FE. Loudspeakers and rooms for sound reproduction - A scientific review. Journal of the Audio Engineering Society. 2006;54(6):451-76.
30. Zahorik P, Anderson PW. Amplitude modulation detection by human listeners in reverberant sound fields: Effects of prior listening exposure. Proceedings of Meetings on Acoustics. 2013;19(1).
31. Ochsner KN, Chiu CY, Schacter DL. Varieties of priming. Current Opinions in Neurobiology. 1994;4(2):189-94.
32. van Wijngaarden SJ, Drullman R. Binaural intelligibility prediction based on the speech transmission index. Journal of the Acoustical Society of America. 2008;123(6):4514-23.
33. MacLeod RB. An experimental investigation of brightness constancy. Archives of Psychology. 1932(135).

34. Holway AF, Boring EG. Determinants of apparent visual size with distance variant. American Journal of Psychology. 1941;54:21-37.
35. Zahorik P, Wightman FL. Loudness constancy with varying sound source distance. Nature Neuroscience. 2001;4(1):78-83.
36. Kuwada S, Bishop B, Kim DO. Azimuth and envelope coding in the inferior colliculus of the unanesthetized rabbit: effect of reverberation and distance. Journal of Neurophysiology. 2014;112(6):1340-55.
37. Slama MC, Delgutte B. Neural coding of sound envelope in reverberant environments. Journal of Neurophysiology. 2015;35(10):4452-68.
38. Kuwada S, Bishop B, Kim DO. Approaches to the study of neural coding of sound source location and sound envelope in real environments. Frontiers in Neural Circuits. 2012;6:Article 42, 1-12.
39. Slaney M. An efficient implementation of the Patterson-Holdsworth auditory filter bank. Apple Computer Technical Report #35, 1993.
40. Zahorik P, Brandewie E. Perceptual Adaptation to Room Acoustics and Effects on Speech Intelligibility in Hearing-Impaired Populations. Proceedings of Forum Acusticum. 2011:2167-72.