Published in final edited form as:

Proc Meet Acoust. 2011 April 1; 129(4): 2487. doi:10.1121/1.3588198.

Adaptation to Room Acoustics Using the Modified Rhyme Test

Eugene Brandewie* and Pavel Zahorik

Abstract

The negative effect of reverberant sound energy on speech intelligibility is well documented. Recently, however, prior exposure to room acoustics has been shown to increase intelligibility for a number of listeners in simulated room environments. This room adaptation effect, a possible extension of dynamic echo suppression, has been shown to be specific to reverberant rooms and requires binaural input. Because this effect has been demonstrated only using the Coordinated Response Measure (CRM) corpus it is important to determine whether the increase in intelligibility scores reported previously was due to the specific nature of the CRM task. Here we demonstrate a comparable room-acoustic effect using the Modified Rhyme Test (MRT) corpus in multiple room environments. The results are consistent with the idea that the room adaptation effect may be a natural phenomenon of listening in reverberant environments.

1. Introduction

Recent research has shown that when listeners are given prior exposure to the acoustics of the listening environment, objective improvements in speech intelligibility are observed (Brandewie & Zahorik, 2008; Zahorik & Brandewie, 2009; Zahorik, Brandewie, & Sivonen, 2009) and subjective shifts in the perception of certain speech sounds result (Watkins, 2005a, 2005b). While the mechanisms underlying these effects are not well understood, the intelligibility improvements have been shown to be absent in anechoic space and to rely on binaural input (Brandewie & Zahorik, 2010). This process may therefore be related to the buildup of echo suppression, which is a binaural phenomenon that has been well demonstrated with less complex stimuli, such as clicks and a single simulated echo (Freyman *et al.*, 1991), as well as with more complex stimuli, such as speech materials and multiple reflections (Djelani & Blauert, 2001).

Because the improvements in speech intelligibility with prior room exposure have been exclusively explored using the Coordinated Response Measure (CRM), a phonetically restrictive, closed-set speech corpus (Bolia *et al.*, 2000), it is important to determine whether other, more varied, speech materials produce similar results. The purpose of this experiment is to determine if intelligibility improvements in reverberant sound fields result using a testing paradigm essentially identical to Brandewie & Zahorik (2010), but with a large-set, phonetically diverse speech corpus, the Modified Rhyme Test (MRT) (House *et al.*, 1965).

2. Methods

2.1 Subjects

Nine female and five male listeners, ages 19 – 56 years (average: 29 years) participated in the experiment. All had normal hearing, as verified by audiometric screening at 15 dB HL.

2.2 Speech Corpus

Speech materials were taken from the Modified Rhyme Test corpus (House et al, 1965). Recordings were made by an American-English male talker in a double-walled sound isolation chamber (Acoustic Systems) using a compression microphone (Rode NT1A) at a distance of about 27 cm. Output levels were approximately 75 dB SPL at this distance. The speech materials were recorded digitally at a 96 kHz sampling frequency with 24-bit resolution. Speech materials were then high-pass filtered (3rd-order Butterworth) at 90 Hz to remove low-frequency environmental noise before being down-sampled to 48 kHz (24 bit) prior to convolution with room impulse responses (see below).

2.3 Room Simulation

Three room environments were simulated using virtual auditory space techniques. The *reverberant* room had a broadband (125 – 4000 Hz) reverberation time (T60) of 0.42 s while the *anechoic* room had a reverberation time near 0 s. An additional room, (T60 = 0.32 s), was used in the “No Carrier” condition to limit room exposure (see below). All rooms had identical dimensions of $5.7 \times 4.3 \times 2.6$ m, but differed in the absorptive properties of the simulated surfaces. Each room was simulated using a simple model of a binaural room impulse response (BRIR) using a technique detailed in Zahorik (2009). The model was constructed using an image-model (Allen & Berkley, 1979) to simulate early reflections and a statistical model for the late reverberant energy. This model has been shown to produce results that are perceptually similar to those derived from a real room, (Zahorik, 2009).

In all rooms, the target speech was simulated at a spatial location directly in front of the listener at a distance of 1.4 m, and a broadband noise masker was presented at a simulated location opposite the listener's right ear, also at 1.4 m. Signal-to-noise ratio (SNR) varied from –23 to +17 dB in 5 dB steps by adjusting the level of the speech signal. A diagram of the listening configuration is shown in Fig. 1.

2.4 Design

Two conditions were tested based on the amount of prior room exposure, as shown in Fig. 1. The “No Carrier” condition (NC) limited room exposure by presenting only a single target word from the MRT corpus and selecting one of the three simulated rooms at random from trial-to-trial within a block. The “Sentence Carrier” condition (SC) enhanced room exposure by presenting listeners with a two-sentence speech carrier prior to the target word and by maintaining a constant room environment across a block of trials.

Each block contained 54 trials. Most subjects completed 15 blocks of the NC condition and 5 blocks for the SC condition. This resulted in 30 trials for each combination of SNR, condition, and room environment. Three listeners (ZAJ, ZAS, ZAT) were unable to complete the SC condition due to scheduling and time constraints.

2.5 Procedure

All sounds were presented over equalized Beyerdynamic DT-990-Pro headphones using a Digital Audio Labs CardDeluxe for D/A conversion (24-bit, 44.1 kHz) at moderate level (approximately 65 dB SPL) within a double-walled sound isolation chamber (Acoustic Systems). Listeners were presented with six rhyming words, one of which was the target word. All words were presented on buttons within a graphical user interface. The listener's task was to select the target word from the list of presented words. Listeners entered their responses by clicking on the appropriate button. All signal processing and data collection applications were implemented using Matlab (Mathworks, Inc.).

2.6 Data Analysis

The proportion of correct responses (PC) was computed separately for each listener in all combinations of exposure condition, room, and SNR. Logistic functions of the following form were fit to the data using a maximum likelihood procedure (Wichmann & Hill, 2001a),

$$\text{Est.}PC(x) = (1 - \delta) \times \frac{1}{1 + \exp(-(\alpha - x)/\beta)} + \delta$$

where α is the estimated threshold parameter, β is the estimated slope parameter, and δ is chance-performance level (1/6 in the MRT task). Ninety-five percent confidence intervals were obtained for each fitted function at a threshold value of $PC = 0.583$ using a bootstrapping procedure (Wichmann & Hill, 2001b).

3. Results

Overall, the data were well-approximated by the logistic functions ($R^2 = 0.83$ in all cases with a median R^2 of 0.95). A summary of the threshold results for all listeners is shown in Fig. 2. Five out of the eleven listeners who participated in both conditions demonstrated a decreased speech reception threshold (SRT) in the SC condition compared to the NC condition, indicative of a room adaptation effect (Brandewie & Zahorik, 2010). However, six listeners show the opposite effect: an increased SRT in the SC condition. On average, no improvement in speech intelligibility was observed with the additional room exposure provided in the SC condition. This can be seen on the right side of Fig. 2 with the function fits for the data pooled across all listeners (AVG).

4. Discussion

There are a number of potentially important differences between the current study using the MRT task and the original study by Brandewie & Zahorik (2010) using the CRM task that might explain the absence of the improvement effect for the majority of listeners tested here. The first difference is the increase in the variability in the MRT task compared to the CRM task. The increase in between-subject threshold variability is evident in Table 1 where threshold variance (dB) in both the MRT and CRM tasks are displayed. In the anechoic room, the MRT threshold variability is nearly double that of the CRM. In the reverberant room, this difference is even greater. The within-subject variability is also greater in the MRT task, as evidenced by the larger average beta parameters and smaller average slopes displayed in Table 1.

The difference in slopes is also evident in a comparison of Fig. 3 and Fig. 4 which display the psychometric functions for a single representative listener (LRK) in both the MRT (Fig. 3) and CRM (Fig. 4) tasks. The steep slopes from the CRM task gave support for only using the estimated threshold values when analyzing differences between the NC and SC conditions. However, the gradual slopes observed using the MRT suggest that this analysis may not be applicable with this task due to less accuracy in the threshold estimates.

It is likely that task differences are responsible for the increase in response variability observed in the MRT task. Performance in this task largely relies on the perception of a single phoneme, whereas performance in the CRM task is mostly due to the identification of specific speech tokens within a small set size. The MRT task contained 274 possible target words. The CRM task, however, was limited to combinations of 4 colors and 8 numbers, creating 32 possible targets. Over the entire length of each study, listeners in the CRM task would have had eight times more exposure to the individual speech materials than

those in the MRT task. These differences likely contribute to the shallower slopes seen in this study.

There were also somewhat different procedures used for the presentation of the masker in the current study using the MRT than in our past study with the CRM (Brandewie & Zahorik, 2010), as shown in Fig. 5. For the CRM task, the color and number targets were imbedded within the second carrier phrase (displayed by the black waveform within the green waveform). However, for the MRT task, the target word was isolated, starting about 2 s after the second carrier phrase. It is still uncertain how long the perceptual buildup to the room acoustics lasts. It is possible that the room buildup rapidly decays in the absence of acoustic stimulation in some listeners. The noise masker may also contribute to room buildup since it too was spatialized in the virtual environment. For the CRM task, the masker (shown in red in Fig. 5) was present throughout the entire speech materials, while in the MRT task the masker was not present until just prior to the speech target. The less exposure time to the masker in the MRT task may have also contributed to the lack of buildup in some listeners.

Another possibility for the lack of buildup in some listeners relates to the component of reverberant listening that elicits better performance over time. The NC condition of the CRM task contained three reverberant rooms varying in their reverberation times presented randomly across a block of trials. For the MRT task, an anechoic room replaced one of the reverberant rooms in the NC blocks. It is possible that a presentation of the anechoic speech signal in the same block of trials provided additional information about the speech signal, the same information that the buildup phenomenon perhaps allows one to extract from speech signals degraded by reverberation. It is possible that this information is related to the amplitude modulation characteristics of speech, which are known to be attenuated in reverberant rooms (Houtgast & Steeneken, 1985). If the room adaptation effect is related to the extraction of the modulation pattern in the speech signal, then one would expect to see no difference between the NC and SC conditions when the NC condition included anechoic speech from the same talker. This would be especially evident in the MRT task which had only a single talker for the entire experiment.

To examine this hypothesis, the data from the reverberant room in the NC condition were split into two listener groups, those with positive buildup effects (NC threshold > SC threshold, $n = 6$) and those with negative effects (SC threshold > NC threshold, $n = 5$). The data were pooled for each of these groups into two subsets, one for trials in which the anechoic room preceded the test room and another in which the other reverberant room preceded the test room. Logistic functions were fit to the pooled data for each subset using the same method described previously. For the negative effect group, there was a 2.13 dB difference between thresholds in the reverberant test room when there was an anechoic preceding trial versus when there was a reverberant preceding trial (-12.030 vs. -9.903 SRT). Such a difference may indicate that individuals in this group were able to take advantage of the preceding anechoic signal to improve performance in the reverberant test room. The positive effect group, however, showed little difference (-0.16 dB) between the two situations (-10.606 vs. -10.767 SRT). This group seemed to perform each trial as if it was a separate event, and by doing so, did not benefit from the addition of the anechoic speech signal in the NC condition. Perhaps both groups found benefit in the increased room exposure provided by the carrier phrase in the SC condition, however, the negative effect group found greater benefit in the NC condition with what was essentially an anechoic carrier phrase from the preceding trial. This difference between the two groups might explain why the overall pooled data show no difference between the NC and SC conditions. Such a finding might also be an important indication that the buildup phenomenon extracts a

characteristic that is more easily attainable with anechoic speech, such as the modulation patterns produced by a particular talker.

5. Conclusions

Evidence for a speech intelligibility improvement effect with prior room exposure using the Modified Rhyme Test was observed, but only for a portion of the listeners. Differences in judgment variability and methodology may be related to the absence of the effect in the other listeners.

Acknowledgments

The authors wish to thank Regina Collecchia for assistance in data collection. Work supported in part by the NIH-NIDCD (Grant No. R01DC008168).

References

- Allen JB, Berkley DA. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*. 1979; 65(4):943–950.
- 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):1065–1066. [PubMed: 10687719]
- Brandewie E, Zahorik P. Adaptation to room acoustics and the effect on speech intelligibility. *Abstracts of the Midwinter Research Meeting / Association for Research in Otolaryngology*. 2008; 31:296.
- Brandewie E, Zahorik P. Prior listening in rooms improves speech intelligibility. *Journal of the Acoustical Society of America*. 2010; 128(1):291–299. [PubMed: 20649224]
- Djelani T, Blauert J. Investigations into the build-up and breakdown of the precedence effect. *Acta Acust. Acust.* 2001; 87:253–261.
- Freyman RL, Clifton RK, Litovsky RY. Dynamic processes in the precedence effect. *Journal of the Acoustical Society of America*. 1991; 90:874–884. [PubMed: 1939892]
- House A, Williams C, Hecker M, Kryter K. Articulation testing methods: Consonantal differentiation with a closed response set. *Journal of the Acoustical Society of America*. 1965; 37:158–166. [PubMed: 14265103]
- Houtgast T, Steeneken HJM. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.* 1985; 77:1069–1077.
- Knudsen VO. The hearing of speech in auditoriums. *Journal of the Acoustical Society of America*. 1929; 1:56–82.
- Watkins AJ. Perceptual compensation for effects of echo and of reverberation on speech identification. *Acta Acustica united with Acustica*. 2005a; 91:892–901.
- Watkins AJ. Perceptual compensation for effects of reverberation in speech identification. *Journal of the Acoustical Society of America*. 2005b; 118:249–262. [PubMed: 16119347]
- Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics*. 2001a; 63(8):1293–1313. [PubMed: 11800458]
- Wichmann FA, Hill NJ. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*. 2001b; 63(8):1314–1329. [PubMed: 11800459]
- Zahorik P. Perceptually relevant parameters for virtual listening simulation of small room acoustics. *Journal of the Acoustical Society of America*. 2009; 126(2):776–791. [PubMed: 19640043]
- Zahorik P, Brandewie E. Room adaptation effects on speech intelligibility as a function of room reverberation time. *Abstracts of the Midwinter Research Meeting / Association for Research in Otolaryngology*. 2009; 32:145.
- Zahorik, P.; Brandewie, E.; Sivonen, VP. Spatial hearing in reverberant rooms and effects of prior listening exposure. Paper presented at the International Workshop on Principles and Applications of Spatial Hearing (IWPASH); Zao, Miyagi, Japan. 2009.

Fig. 1.

Experimental setup showing source/listener configuration within the virtual room listening environment as well as target and masker signal details. The “No Carrier” (NC) and “Sentence Carrier” (SC) conditions were designed to either minimize or maximize listening exposure to the room prior to the target speech word. Room reverberation time was manipulated by changing the absorptive properties of the surfaces within the virtual room.

Fig. 2. Speech reception thresholds for NC (blue circles) and SC (green triangles) conditions in the reverberant room. 95% confidence intervals are displayed for each threshold estimate. Listeners are rank ordered from left to right by effect size (NC threshold – SC threshold). Effect size is also displayed numerically above each listener code. Three listeners (ZAJ, ZAS, & ZAT) did not participate in the SC condition. Threshold values based on function fits to the data pooled across all listeners are shown on right (AVG).

Fig. 3.

Proportion of correct speech target identifications as a function of signal-to-noise ratio and estimated psychometric function for a single listener (LRK) in the reverberant room in the MRT task. Speech reception thresholds estimates (with 95% confidence intervals) are displayed for each fitted function. Alpha (a) and beta (b) parameters for each function are also displayed.

Fig. 4.

Same as Figure 3, except data points and fitted functions correspond to performance in the CRM task. This is the same listener (LRK) as shown in Figure 3. Data originally published in Brandewie & Zahorik (2010).

Fig. 5. Examples of the signal waveforms (reverberant room) for the MRT and CRM tasks in the SC condition (right channel only).

MRT and CRM Comparisons: Table of average thresholds, threshold variance, average beta parameters, and average slope values for both the NC and SC condition in both the anechoic and reverberant rooms for the MRT and CRM tasks. Average effect sizes (NC threshold – SC threshold) for the reverberant room are presented in the right column for both the MRT and CRM tasks. CRM data originally published in Brandewie & Zahorik (2010).

Table 1

	Anechoic		Reverberant		
Condition:	NC	SC	NC	SC	Effect Size
MEAN THRESHOLD (dB):					
MRT:	-15.18	-14.79	-10.53	-10.87	0.14
CRM:	-19.88	-20.72	-13.39	-16.13	2.74
THRESHOLD VARIANCE (dB):					
MRT:	5.25	5.26	5.47	9.57	
CRM:	3.13	3.06	0.78	0.99	
BETA PARAMETER:					
MRT:	6.53	6.20	5.66	6.86	
CRM:	3.57	3.82	3.31	3.62	
SLOPE (P(C)/dB):					
MRT:	0.0406	0.0435	0.0386	0.0376	
CRM:	0.0717	0.0666	0.0791	0.0721	