

Published in final edited form as:
Proc Meet Acoust. ; 12: .

The effect of semantic context on speech intelligibility in reverberant rooms

Nirmal Srinivasan* and Pavel Zahorik

Abstract

Although it is well known that semantic context affects speech intelligibility and that different reverberant rooms affect speech intelligibility differentially, these effects have seldom been studied together. Revised SPIN sentences in a background of Gaussian noise in simulated rooms with reverberation time (T60) of 1 and 0.25 s were used. The carrier phrase and the target word of the speech stimuli were manipulated to be either in the same room or in different rooms. As expected, intelligibility of predictable sentences was higher compared to unpredictable sentences—the context effect. The context effect was higher in the low-reverberant room as compared to the high-reverberant room. When the carrier phrase and target words were in different rooms, the context effect was higher when the carrier phrase was in the low-reverberant room and target word in the high-reverberant room. For predictable sentences, changing the target word from high-reverberation to low reverberation with a high reverberant carrier increased intelligibility. However, with a low-reverberant carrier and different rooms for the target word, there was no change in intelligibility. Overall, it could be concluded that there is an interaction between semantic context and room acoustics for speech intelligibility.

1. Introduction

It is universally accepted that linguistic context can affect the perception of speech. It is also well-known that altering speech signals by factors such as reverberation and noise can reduce speech intelligibility (Nabelek & Mason, 1981). Although both noise and reverberation can degrade speech signals in isolation, these distortions occur simultaneously in most listening environments and are often more detrimental than the sum of the component distractions (Nabelek, 1988). The accuracy of speech perception not only depends on the sensory data generated from the stimulus itself but also from the context within which the stimulus occurs. Words are recognized more easily when they are presented in sentences rather than in isolation or in carrier phrases (Boothroyd, 1985; Miller, Heise, & Litchen, 1951, O'Neil, 1957), and words with high frequency of occurrence in spoken and written language are more easily recognized than are words with a low frequency of occurrence (Giolas & Epstein, 1963; Savin, 1963).

Previous research has shown that prior exposure to the acoustics of the listening environment can improve speech intelligibility ((Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2009; Zahorik, Brandewie, & Sivonen, 2009). However, the cumulative effect of reverberation, noise, and semantic context on room adaptation and speech intelligibility has not been studied. This study investigates the effect of semantic context on speech intelligibility and room adaptation effects for different reverberant and noisy rooms.

2. Methods

2.1 Subjects

Twenty one normal-hearing (pure-tone air-conductive thresholds ≤ 20 dB HL from 250 – 8000 Hz) adults (7 male, 14 female, age range 19 – 56 years) participated in the experiment. The subjects were paid for their participation. None had prior experience with the speech corpus materials used in the study. All procedures involving human subjects were reviewed and approved by the University of Louisville Institutional Review Board.

2.2 Sound Field Simulation

Virtual acoustic techniques were used to simulate two listening environments. The more reverberant room (MR) had a broadband (125 – 4000 Hz) reverberation time, T_{60} , of approximately 1 s while the less reverberant room (LR) had a, T_{60} , of approximately 0.25 s. The room had identical dimensions of $5.7 \times 4.3 \times 2.6$ m (L \times W \times H), but differed in the absorptive properties of the simulated surfaces. Room simulation techniques were based on methods described by Zahorik (2009). Briefly, the simulation methods used an image model (Allen and Berkley, 1979) to compute the directions, delays, and attenuations of the early reflections, which were then, along with the direct-path, spatially rendered using non-individualized head-related transfer functions (HRTFs). The late reverberant energy was simulated statistically using exponentially decaying independent Gaussian noise samples in octave bands from 125 – 4000 Hz for each ear. Overall, this simplified method of room simulation has been found to produce binaural room impulse responses (BRIRs) that are reasonable physical and perceptual approximations of those measured in real rooms (Zahorik, 2009).

In both the 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 to listener's right ear, also at a distance of 1.4 m. The signals were presented at three signal-to-noise ratios (SNRs): -8 dB, -6 dB, and -4 dB. A graphical representation of the listening configuration is shown in Figure 1.

2.3 Speech Stimuli

Speech materials were taken from the Speech Perception In Noise (SPIN) set sentences (Kalikow, Stevens, & Elliot, 1977). The test consists of 8 lists of 50 sentences with the target word being the last word of every sentence. The speech material was naturally spoken by a female speaker with a General American English dialect in a sound-treated acoustic chamber. The sentences were recorded with a Crown CM312 microphone situated at a distance of 2 cm from the side of the mouth. The recorded sentences were digitized with a sampling rate of 44100 samples per second and quantized at 16 bits per sample. Speech materials were low-pass filtered (3rd-order Butterworth) at 90 Hz to remove any environmental noises and were normalized for same RMS level (64 dB SPL) prior to convolution with room impulse responses.

Two conditions were tested based on the characteristics of the carrier phrase and the target word, as shown in Figure 1. The "Congruent" condition enhanced prior room exposure by having both the carrier phrase and the target word in a constant room environment across all the trials in a block. The "Incongruent" condition limited prior room exposure by having the carrier phrase and the target word in different rooms. Both Congruent and Incongruent sentences were presented in MR and LR conditions. In the CongruentMR condition, both the carrier phrase and target word were presented in the more reverberant room. In the CongruentLR condition, both the carrier phrase and target word were presented in the less reverberant room. In the IncongruentLR condition, the carrier phrase was presented in MR

room and the target word was presented in LR room. In the IncongruentMR condition, the carrier phrase was presented in LR room and the target word was presented in MR room.

2.4 Design and Procedure

Each listener was tested in two of the three SNRs. Each listener performed one experimental session which lasted approximately one hour. During that session, listeners were presented with 400 sentences in 8 blocks of 50 sentences. Every block of trials consisted of 25 predictable last word sentences and 25 unpredictable last word sentences at a single SNR. The listeners always heard speech stimuli in congruent conditions followed by incongruent conditions. The order of presentation was randomized within every individual block. Each listener completed 8 blocks of trials (CongruentMR, CongruentLR, IncongruentMR, and IncongruentLR at two SNRs) and listened to 200 sentences at each SNR (50-CongruentMR, 50-CongruentLR, 50-IncongruentMR, and 50-IncongruentLR). Each target word was repeated twice, both in predictable and unpredictable sentences, in the experiment. However, the same target word did not repeat in a single block of 50 sentences.

The listeners were seated in a sound-attenuating chamber (Acoustic Systems, Austin, TX, custom double walled) and listened to speech via 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 68 dB SPL). Listeners were instructed to type in the last word of every sentence. No feedback was provided at the end of each trial. Data collection was self-paced, and listeners were instructed to take breaks whenever they felt fatigued. All stimulus presentation and data collection was implemented using MATLAB software (Mathworks, Inc.).

2.5 Dependent Measure

Speech intelligibility represented the proportion of final word of the sentences that were correctly identified. Note that identifying the last word to be a plural when it was actually singular and vice versa was considered to be an incorrect response.

3 Results

Figure 2 displays the effect of semantic context and SNR on speech intelligibility for congruent and incongruent conditions. Since each listener listened to two of the three SNRs, the intelligibility presented here is the mean from 14 listeners. Table 1 shows the effect of semantic context and SNR on speech intelligibility during congruent and incongruent listening conditions.

For predictable sentences, when the carrier sentence was in the MR room, there was a 30% increase in intelligibility when the target word was presented in incongruent condition as compared to the congruent condition. This increase in intelligibility could be attributed to two factors: 1.) contextual cues from the predictable nature of the sentences presented and 2.) LR room for the target word. However, when the carrier sentence was in the less reverberant room, there was no difference in speech intelligibility between congruent and incongruent listening conditions. This overall drop in intelligibility between congruent and incongruent conditions for predictable sentences in the MR room implied that semantic context of speech had a higher effect than room adaptation on speech intelligibility in reverberant rooms.

For unpredictable sentences, the intelligibility was higher when the target word was in the less reverberant room as compared to the more reverberant room irrespective of the congruency condition. However, the difference in intelligibility between the congruent and incongruent conditions was higher when the target word was in the less reverberant room.

Also, consistent with past research, predictable sentences had higher intelligibility compared to unpredictable sentences and a decrease in intelligibility was noted as the listening environment became more reverberant.

Figure 3 displays the same data as shown in Figure 2, but plotted differently to highlight the effect of adapting to the listening environment on speech intelligibility for congruent and incongruent conditions. Table 2 shows the effect of room adaptation on speech intelligibility during congruent and incongruent listening conditions.

Unpredictable sentences showed room adaptation effects. Intelligibility was higher when both carrier sentences and target words were presented in the same room (congruent) as compared to when they were presented in different rooms (incongruent). This improvement in intelligibility based on prior room exposure was approximately 10% for the MR room and approximately 5% for the LR room. Predictable sentences did not show any room adaptation effects. Intelligibility was best when the target word was presented in the less reverberant room irrespective of the congruency conditions, which is consistent with the idea that the contextual information is more difficult to encode in the more reverberant room.

4. Conclusions

Semantic context of speech had a facilitative effect on speech intelligibility in reverberant rooms. This effect was larger in the MR room as compared to the LR room. Also, prior exposure to listening environment improves intelligibility of speech with low semantic context by an average of 10% for the more reverberant environment and by about 5% for the less reverberant environment. Overall, semantic context has a much greater effect on speech intelligibility than prior exposure to listening environment.

Acknowledgments

The authors wish to thank Regina Collecchia for assistance in data collection. This research was supported by 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.
- Boothroyd A. Evaluation of speech production of the hearing impaired: Some benefits of forced-choice testing. *Journal of Speech and Hearing Research*. 1985; 28:186–196.
- 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]
- Giolas TG, Epstein A. Comparative intelligibility of word lists and continuous discourse. *Journal of Speech and Hearing Research*. 1963; 12:349–358. [PubMed: 14071872]
- Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*. 1977; 61 (5):1337–1351. [PubMed: 881487]
- Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*. 1951; 41:329–335. [PubMed: 14861384]
- Nabelek AK. Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *Journal of Acoustical Society of America*. 1984; 84 (2):476–484.
- Nabelek AK, Mason D. Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms. *Journal of Speech and Hearing Research*. 1981; 24:375–383. [PubMed: 7300279]
- O’Neil JJ. Recognition and intelligibility of test materials in context and isolation. *Journal of Speech and Hearing Disorders*. 1957; 22:87–90.

- Savin HB. Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*. 1963; 35 (2):200–206.
- 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.

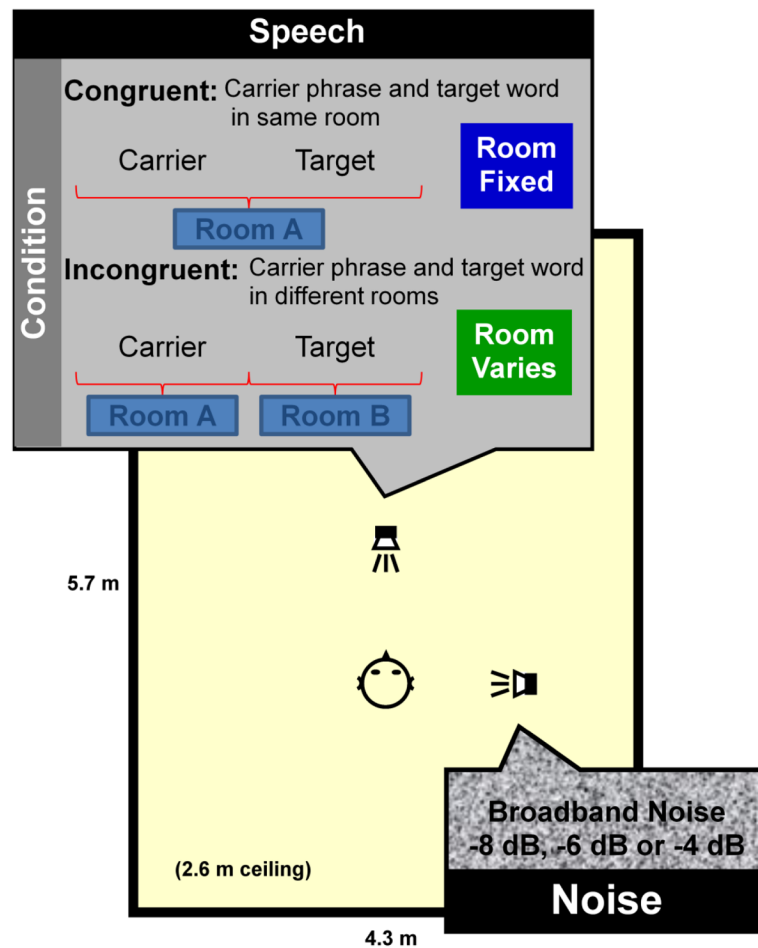


Figure 1.

Experimental setup showing source/listener configuration within the virtual listening environment as well as target and noise signal details. The congruent and incongruent conditions were designed to either minimize or maximize listening exposure to the room prior to the target word. Room reverberation time was manipulated by changing the absorptive properties of the surfaces of the virtual room.

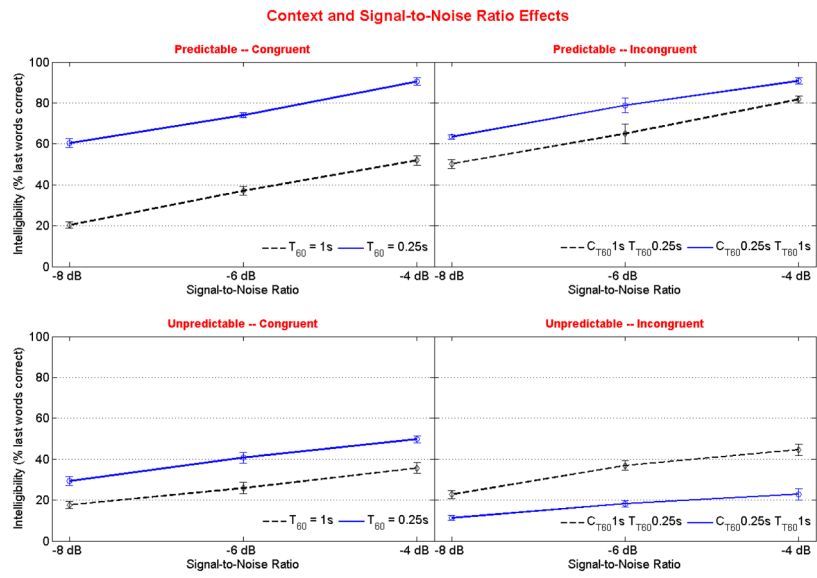


Figure 2. Effect of semantic context and SNR on speech intelligibility. Error bars show 95% confidence intervals for each mean.

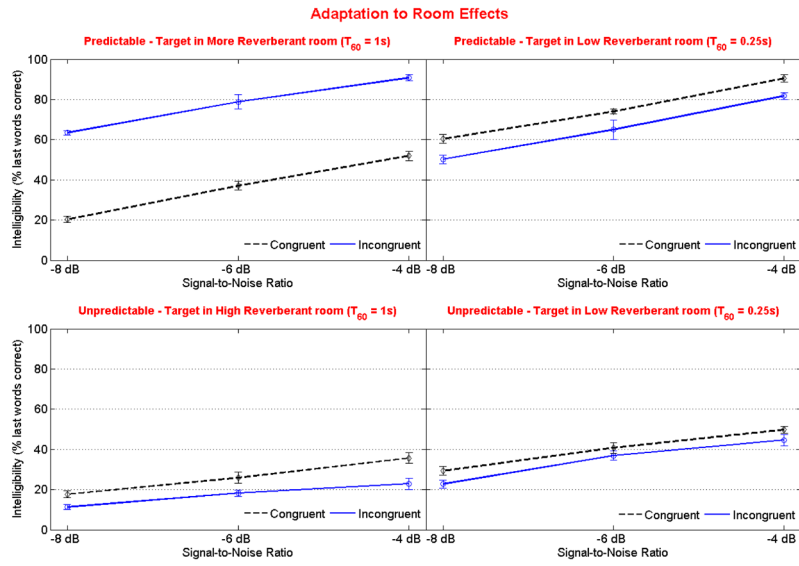


Figure 3. Effect of adaptation to the room on speech intelligibility. Error bars show 95% confidence intervals for each mean.

Table 1

Individual means of speech intelligibility for congruent and incongruent conditions showing the effect of semantic context and SNR.

	SNR	Carrier sentence $T_{60}: 1s$			Carrier sentence $T_{60}: 0.25s$		
		Target $T_{60}: 1s$	Target $T_{60}: 0.25s$	Difference	Target $T_{60}: 1s$	Target $T_{60}: 0.25s$	Difference
Predictable	-4 dB	52	82	30	91	90	-1
	-6 dB	37	65	28	79	74	-5
	-8 dB	20	50	30	63	60	-3
Unpredictable	-4 dB	36	45	9	23	50	27
	-6 dB	26	37	11	18	41	23
	-8 dB	18	23	5	11	29	18

Table 2

Individual means of speech intelligibility for congruent and incongruent conditions showing the effect of room adaptation.

	SNR	Target Word T ₆₀ : 1s			Target Word T ₆₀ : 0.25s		
		Congruent	Incongruent	Difference	Congruent	Incongruent	Difference
Predictable	-4 dB	52	91	-39	90	82	8
	-6 dB	37	79	-42	74	65	9
	-8 dB	20	63	-43	60	50	10
Unpredictable	-4 dB	36	23	13	50	45	5
	-6 dB	26	18	8	41	37	4
	-8 dB	18	11	7	29	23	6