Accuracy of audiometric test room simulations of three real-world listening environments

Robyn M. Cox, Genevieve C. Alexander, and Izel M. Rivera

Citation: The Journal of the Acoustical Society of America **90**, 764 (1991); doi: 10.1121/1.401946
View online: https://doi.org/10.1121/1.401946
View Table of Contents: https://asa.scitation.org/toc/jas/90/2
Published by the Acoustical Society of America

Accuracy of audiometric test room simulations of three real-world listening environments

Robyn M. Cox
Memphis State University, and Department of Veterans Affairs Medical Center, Memphis, Tennessee 38101

Genevieve C. Alexander and Izel M. Rivera
Department of Veterans Affairs Medical Center, Memphis, Tennessee 38101

(Received 9 January 1991; accepted for publication 25 April 1991)

Hearing aid benefit depends primarily on the extent to which amplification facilitates speech understanding in typical everyday listening environments. In the hearing aid fitting process, improved speech understanding is often measured in an audiometric test room. However, because audiometric test rooms are smaller, quieter, and less reverberant than typical rooms, these data may not accurately predict speech understanding in daily life. This study was undertaken to evaluate the validity of three simulated real-world listening environments created in an audiometric test room. The three environments represented a typical living room, cocktail party, and classroom, respectively. Twenty normal-hearing subjects, listening monaurally, provided intelligibility scores for four phonetic contrasts produced by each of three normal talkers. Intelligibility obtained in the real environment was compared with that measured in the corresponding simulated environment. Results indicated that the relative intelligibility of talkers and phonetic contrasts remained essentially constant across each real-simulated environment pair, and that significant talker × contrast interactions seen in the real environments were usually reproduced in the simulated environments. However, there were somewhat fewer significant intelligibility differences in the simulated environments than in the real environments. Also, the intelligibility of one talker deteriorated more than expected in the simulated reverberant environment. Overall, the outcome suggested that these typical listening environments were rather accurately simulated (for monaural listening) in an audiometric test room using appropriate adjustments of presentation level, signal-to-babble ratio, and synthetic reverberation effects.

PACS numbers: 43.71.Es, 43.71.Gv, 43.66.Ts, 43.66.Yw

INTRODUCTION

For individuals with mild to severe hearing loss, hearing aid benefit is determined principally by the extent to which speech communication ability is improved in daily life (Barcham and Stephens, 1980; Golabek *et al.*, 1988; Hagerman and Gabrielsson, 1984). It is not surprising, therefore, that there has always been a lively interest in developing clinical procedures to predict the amount of improvement in speech understanding that will be associated with a hearing aid fitting. Nevertheless, numerous investigators have reported that clinical measurements of the intelligibility or quality of amplified speech are only marginally predictive of the benefit people will receive from their hearing aids in daily life (e.g., Kapteyn, 1977; Harris and Goldstein, 1979; Walden *et al.*, 1983). The apparent discrepancy between clinically measured benefit and real-world benefit must be due at least partially to inadequate clinical test methods.

One of the factors identified as contributing to the inaccuracy of clinical predictions of hearing aid benefit is the listening environment in which speech understanding tests are often conducted (e.g, Harris and Reitz, 1985). Hearing aids are typically selected and evaluated in audiometric test rooms. These small sound-treated rooms have an ambient noise level that is much lower than found in everyday listening. In addition, the reverberation time is much shorter (0.1–0.2 s) and the spacing of reflections is more dense than found in typical rooms. These conditions produce a listening environment that is probably never experienced by the hearing aid wearer outside of the audiology clinic.

These limitations have been recognized for many years and efforts have been made to compensate for them. The most common approach has been to add a competing noise or multitalker babble to the target speech to simulate a noisy environment such as a cocktail party. Because the cocktail party type of situation is often identified as the most problematic for hearing aid wearers, documentation of improved speech understanding in this situation is frequently a cornerstone of hearing aid selection procedures. Unfortunately, hearing aid users consistently report that, despite improvements measured in the clinical setting, they realize little or no benefit from amplification in noisy everyday environments (e.g., Scherr *et al.*, 1983; Walden *et al.*, 1984; Lynn and Lesner, 1990). This has led to an interpretive dilemma: Is the lack of reported benefit due to an inability of hearing aid wearers to judge benefit, or was the clinically measured benefit invalid due to inappropriate test methods? In a recent study, Cox and Alexander (1991) measured hearing aid benefit in real everyday environments, including a cocktail party type of setting. They reported that, when hearing aid wearers listened to speech in the real environment and used their hearing aids in their preferred manner, benefit was close to zero. This bolsters the validity of the self-assessed benefit data and suggests that some clinical methods used to predict hearing aid benefit in noisy settings may be invalid.

Reverberant listening environments also significantly degrade speech intelligibility and limit hearing aid benefit. The distortions associated with reverberation involve temporal smearing of speech elements, whereas those associated with background noise produce simultaneous masking. Because of this basic difference, it has been proposed that reverberation produces speech perception errors that are distinct from those produced by noise. Recent studies have tended to support this hypothesis (Nabelek and Dagenais, 1986; Nabelek *et al.*, 1989; Helfer and Wilber, 1990; Helfer and Huntley, 1989). In recognition of the unique nature of reverberant listening environments, attempts have been made to simulate them for clinical testing. Two approaches have been taken to this problem: (1) addition of reflective material to the walls of audiometric test rooms, and (2) electronic simulation in which reflections are artificially added to the speech signal before transduction.

Through a comparison of published studies, Nabelek and Robinette (1978) determined that reverberation effects were substantially exaggerated when the reverberation was produced by reflective surfaces in rooms similar in size to audiometric test rooms. They attributed this to the atypical distribution of early reflections in such small rooms. Thus, although audiometric test rooms can be made reverberant, the listening environment produced thereby is rarely experienced in daily listening.

Other researchers have evaluated the validity of electronically produced reverberation. Nabelek and Robinette (1978) found that electronically simulated early reflections produced less deterioration in speech intelligibility than nominally similar real reverberation. Irwin and McCauley (1987), on the other hand, reported that simulated reverberation produced more deterioration than real reverberation and that the difference between simulated reverberant conditions was less than the difference between real reverberant conditions. In the latter study it was noted that the electronically produced reflections were nonrandom and thus not similar to those produced in real rooms. Overall, these studies suggested that speech understanding measured in simulated reverberant environments in audiometric test rooms may not validly predict speech understanding in typical real-work reverberant conditions.

Although it has been evident for some time that hearing aid benefit varies as a function of listening environments, this issue has not been addressed in hearing aid fittings until recently. Technological advances have made it possible to produce hearing aids that respond adaptively to the acoustic environment using algorithms that are intended to reduce the deleterious effects of ambient noise and/or reverberation. Furthermore, multimemory hearing aids can now be programmed to perform as several distinctly different instruments, depending on the user's choice. These capabilities have created a demand for convenient and valid methods to predict speech understanding in typical real-world listening environments. Such methods are needed so that fitting strategies for these more sophisticated hearing instruments can be refined and individualized.

Based on these considerations, the present study was undertaken to evaluate the validity of simulated real-world listening environments created in an audiometric test room. The long-term goal was to develop environment simulations that would produce the same effects on speech understanding as produced by the real environments. In previous studies, we have defined and evaluated three basic listening environments for speech communication (e.g., Cox *et al.*, 1987; Cox and Alexander, 1991; Cox and Gilmore, 1990). Both theoretical considerations and the data of Walden *et al.* (1984) suggest that these three environments place distinctly different demands on the listener and together represent a large proportion of the everyday listening situations experienced by the typical hearing aid wearer. In the present study, understanding was measured for speech material recorded in each of the real environments and for the same speech material presented in the corresponding environments simulated in an audiometric test room. An analytic speech understanding test was used so that effects on intelligibility of specific phonetic contrasts could be compared in real and simulated environments. In addition, the target speech was produced by three talkers known to differ in intelligibility in real environments. The research questions were as follows. (1) Do the simulated environments degrade speech intelligibility to the same extent as the real environments? (2) Are patterns of talker intelligibility seen in the real environments maintained in the simulated environments? (3) Are patterns of phonetic contrast intelligibility seen in the real environments maintained in the simulated environments? (4) Are talker-contrast interactions in the simulated environments similar to those seen in the real environments?

I. METHOD

Cox *et al.* (1987) used the speech pattern contrast (SPAC) test (Boothroyd, 1985) to quantify the intelligibility of six normal talkers in four typical listening environments (including the three environments used in the present study). Master recordings of the SPAC test were generated by each talker in a quiet, nonreverberant environment (see Cox *et al.* for details of the recording process). In addition, a multitalker babble was recorded in the same environment from six different talkers reading simultaneously from different texts. The master recordings and multitalker babble were re-recorded in the real environments and subsequently presented to subjects via an insert earphone. The data obtained in that investigation were used to select the talkers and SPAC subtests for the present work. The master recordings and real-room recordings used by Cox *et al.* were also used in this study. Details are given below.

A. Test stimuli

The SPAC test is a 4AFC test, including four segmental subtests, each yielding two contrast scores (a total of eight contrast scores). Each subtest is composed of 12 test words.

Cox *et al.* (1987) found that four contrasts, produced by two subtests, significantly differentiated among talkers in the real environments. These two subtests were chosen for the present study. The contrasts were: initial consonant place (ICP), final consonant voicing (FCV), final consonant continuance (FCC), and final consonant place (FCP).

To maximize the likelihood that the talkers would produce natural-sounding speech, the SPAC test words were embedded in short sentences (reproduced in Cox *et al.*, 1987, Appendix A). These sentences were devised to present the items in a variety of contexts with respects to preceding and following phonemes, position of test item in the utterance, and length of utterance. Practice items were also constructed in a similar manner. The sentences were randomly assigned to the test items in each subtest. There are 12 forms of the SPAC test, each form having a different set of 12 correct responses in each subtest. Each talker recorded a different combination of four forms for the master recordings.

B. Listening environments

The three listening environments were designated A, B, and C. In each environment, the data of Pearsons *et al.* (1977) were used to determine appropriate speech and background noise levels as well as appropriate talker–listener distance. The levels and distances were those reported by Pearsons *et al.* to be maintained by talkers and listeners in everyday environments to allow essentially complete intelligibility for conversations in that setting.

Environment A represented a communication situation in which speech is at normal or casual conversational level and background noise and reverberation are relatively low. Examples of environment A include face-to-face conversation in a typical living room or quiet office. The target speech level was 55 dBA L_{eq} (L_{eq} = equivalent continuous level) and the background noise was delivered at 48 dBA L_{eq} (both measured beside the listener's ear).

Environment B represented a communication situation in which external environmental noise is relatively low but speech cues are reduced because of reverberation. Examples of environment B include listening as an audience member to a lecture delivered in an unamplified classroom, or communicating across a relatively large room. In this environment, the target speech was delivered at a level of 70 dBA L_{eq} at 1 m from the talker. However, because communication was occurring over a distance, the level measured beside the listener's ear was lower (63 dBA L_{eq} in this study). The background noise was delivered at 55 dBA L_{eq} at the listener's ear.

Environment C represented a communication situation where external environmental noise is relatively high and speech levels are somewhat raised. Examples of environment C include face-to-face communication at a social event with competing conversations nearby, or communication with a clerk in a busy store. In addition to the data of Pearsons *et al.* (1977), the report of Plomp (1977) was considered in selecting the speech-to-babble ratio (SBR) in this environment. The target speech level was 64 dBA L_{eq} and the background noise was delivered at 62 dBA L_{eq} (both measured beside the listener's ear).

TABLE I. Reverberation time (s) as a function of frequency for the two real rooms used to implement the three listening environments, for the environments simulated in an audiometric test room, and for the audiometric test room itself.

Freq. (kHz)	Envir. A and C Real	Envir. A and C Sim.	Envir. B Real	Envir. B Sim.	Audiometric test room
0.125	0.70	0.75	1.05	1.14	0.28
0.25	0.35	0.26	0.92	1.05	0.14
0.5	0.37	0.23	1.01	1.09	0.10
1.0	0.39	0.26	0.91	1.08	0.06
2.0	0.55	0.43	0.91	1.02	<0.05
4.0	0.57	0.37	0.85	0.96	<0.05

1. Real environments

Environments A and C were implemented in a 5.8 × 6.1 × 2.6-m room, which contained carpeting, window drapes, and upholstered furniture. The talker–listener distance was 1 m for environment A and 0.5 m for environment C (both well inside the critical distance of 2.7 m, estimated using the equation described by Peutz, 1971). Environment B was implemented in a classroom 18×6.1×3.2 m (ceiling lowered to 2.6 m in the rear 1/3 of the room). The room was uncarpeted with hard walls and acoustical tile ceiling. It contained classroom chairs and several tables. The talker–listener distance was 5 m (considerably beyond the estimated critical distance of 3.7 m). In all environments, both talker and listener were located toward the middle of the room.

Reverberation time (RT_{60}) was determined in each real environment by energizing the rooms using narrow bands of random noise. At each frequency, the RT_{60} was extrapolated from the reverberation decay in the -5- to -25-dB range. Spatial averaging was performed across three measurement positions. Table I gives the reverberation times as a function of frequency for these two rooms.

For each talker, the master recordings of two forms of the SPAC test were re-recorded in each of the real environments. A different combination of forms was used in each listening environment. The target speech was produced by a small loudspeaker (Realistic Minimus 7) designated as the "talker" and the background noise was generated by uncorrelated recordings of multitalker babble transduced by four identical small loudspeakers placed symmetrically around the listener at approximately the same distance as the talker and at azimuths of 45°, 135°, 225°, and 315°. The "listener" was a KEMAR manikin equipped with an ear-simulator coupler and microphone in one ear. The frequency response of the reproduction system was essentially flat from 100 Hz to 14 kHz when measured in a highly damped test room. The environmental recordings were made on magnetic tape with a flat frequency response in the 50-Hz to 10-kHz range (Panasonic AG6810 recorder). Additional details of the re-recording procedures may be found in Cox *et al.* (1987).

2. Simulated environments

The three real environments were simulated in a 1.9×1.8×1.9-m audiometric test room lined with soundabsorbing foam. Reverberation time as a function of frequency is given in Table I. Ambient noise in the test room was 53 dB(C)/19 dB(A). The real environments were simulated by processing the master SPAC test recordings and the multitalker babble recording using a two-channel electronic reverberator (Yamaha, Rev 5) and presenting the processed speech and babble to subjects listening in the sound field. The target speech was presented from a small loudspeaker (Realistic Minimus 7) located 1 m in front of the subject. A recording of the multitalker babble was split and delivered from four identical small loudspeakers mounted in the corners around the listener at azimuths of 45°, 135°, 225°, and 315°. Both speech and babble were presented at levels appropriate for the environment simulated, measured at the listener's position in the unobstructed sound field. The frequency response of the reproduction system was essentially flat from 150 Hz to at least 13 kHz.

In creating the reverberation effects in the simulated listening environments, no attempt was made to reproduce the microstructure of the reverberant effects in the real environments. Three factors were considered in the simulation: (1) RT_{60} as a function of frequency, (2) the level of reverberant effects compared to the level of the direct sound (the mix of direct and reverberant sound was selected using a scale from "0" to "10" in which, nominally, "0" yielded direct sound only and "10" yielded reverberant sound), and (3) appropriate spacing of early reflections (estimated from the size of the real rooms). Reverberation times for the simulated environments were quantified by observing the decay of abruptly terminated processed (i.e., artificially reverberated) noise bands at the subject's position in the sound field. As in the real environments, RT_{60} was extrapolated from the decay in the -5- to -25-dB range. These values are shown in Table I.

Although the real room used for environments A and C was about as reverberant as a typical living room, the location of the listener (well inside the critical distance) suggested that the level of reverberant effects would be very low relative to the direct sound. Spectrographic analyses of the real environment recordings confirmed this: there was no evidence of the prolongation of speech elements that is characteristic of reverberation effects. To simulate these environments, the reverberator was configured to produce and average RT_{60} of 0.4 s. As shown in Table I, measured RT_{60} averaged 0.38 s. However, to simulate a listening location inside the critical distance, the mix of direct and reverberant sound levels was adjusted to a scale value of "1." This reduced the reverberant effects to an almost inaudible level. As a result, speech intelligibility in simulated environments A and C was controlled mainly by the SBR.

The real room used for environment B had a mean reverberation time of 0.94 s. To simulate this environment, the reverberator was configured to an RT_{60} of 0.9 s. The measured RT_{60}s were somewhat longer than this, producing a mean simulated RT_{60} of 1.05 s. To simulate a listening condition outside the critical distance, the mix of direct and reverberant effects was adjusted to a scale value of "9." The first four reflections occurred 20, 36, 48, and 54 ms after the direct sound and at levels of -1, -4.6, -3.9, and -11.3 dB, respectively, relative to the direct sound. Later reflections were dense and random and decayed gradually in a manner similar to real rooms. The reverberant effect was distinctly audible.

C. Talkers

Three of the original six talkers were selected for this study. They were referred to as talkers 2, 4, and 5 by Cox *et al.* (1987) and they retain the same designations here. In the earlier study, talker 4 was significantly more intelligible than talker 5 in all environments. Talker 2 was significantly more intelligible than talker 5 in environments A and C but not in environment B. Therefore, talkers 4, 5, and 2 were chosen to represent high intelligibility, low intelligibility, and environment-dependent intelligibility, respectively.

Long-term average one-third octave speech band spectra for the three talkers are illustrated in Fig. 1. Articulation rates for the test sentences were 3.4, 3.3, and 3.6 syllables/s for talkers 2, 4, and 5, respectively. Talkers 2 and 5 were male, talker 4 was female.

1. Subjects

Fourteen males and six females served as subjects. All had hearing thresholds better than 25 dB HL (*re*: ANSI, 1989) in the range from 250–8000 Hz. Ages ranged from 16–38 years with a mean of 22 years old.

D. Procedure

Both the environmental recordings and the master recordings were low-pass filtered at 10 kHz, sampled at 20 kHz, and digitized with 12-bit resolution, for presentation to subjects.

To measure intelligibility in the real environments, the digitized environmental recordings were presented monaur-

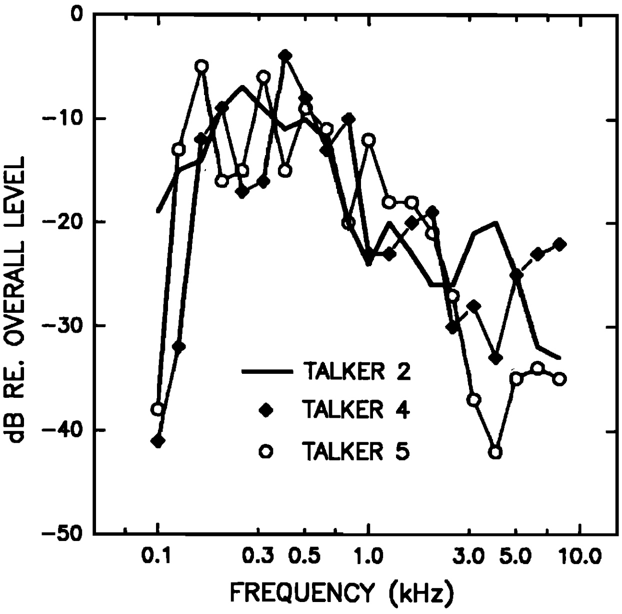

FIG. 1. Long-term 1/3-oct band speech spectra for the three talkers.

ally to subjects via an insert earphone (Etymotic Research ER-2) coupled to the ear canal using a compressible foam ear plug. Playback levels were calibrated using an ear simulator coupler. In the frequency range from 150 Hz–10 kHz, the spectrum and level of the signal delivered to the average subject were equal to those that would have occurred if the subject had been actually located in the environments where the recordings were made.

To measure intelligibility in the simulated environments, the multitalker babble and digitized master recordings were replayed, processed by the reverberator, and delivered in the sound field at the SBR experienced by the listeners in the corresponding real environment. The subject listened monaurally; the nontest ear was plugged.

Delivery and scoring of the test items were controlled by an IBM-AT class microcomputer system. For both real and simulated environment conditions, the subjects were seated in the audiometric test room, facing a 13-in. monitor screen. For each item, the four alternatives were displayed on the screen and the subject keyed in a response on a small handheld keypad.

Experimental variables were controlled as follows: presentation order of listening condition (real versus simulated environment), environment (A, B, C), and talker (2, 4, 5) were counterbalanced across subjects; a given subject heard the same environment-talker schedule in both listening conditions; for a given talker–environment combination, the same two forms were presented in both listening conditions but in reverse order. Half of the subjects listened with their right ear and half used their left ear. Each subject responded to two SPAC subtests under both listening conditions, for all talkers, in all environments. Data were collected in two sessions.

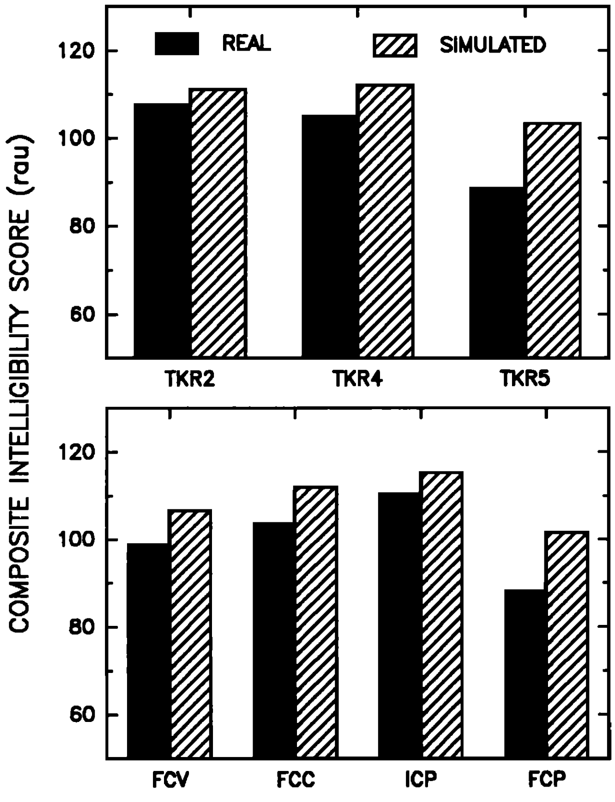

FIG. 2. Intelligibility scores obtained in the real and simulated environment A listening conditions. The upper panel depicts mean scores for each talker across all four contrasts. The lower panel gives mean scores for each contrast across all three talkers. TKR2 = talker 2, TKR4 = talker 4, TKR5 = talker 5, ICP = initial consonant place, FCV = final consonant voicing, FCC = final consonant continuance, FCP = final consonant place.

II. RESULTS

Each subject provided four contrast scores for each talker in each environment under both real and simulated environment listening conditions. The contrasts were: initial consonant place (ICP), final consonant voicing (FCV), final consonant continuance (FCC), and final consonant place (FCP). Each contrast score was based on two SPAC test forms, and thus, 24 items. Raw scores were proportions of correct responses. These were transformed into rationalized arcsine units (raus) before statistical analysis as described by Studebaker (1985). The scale of rationalized arcsine units extends from -23 to 123. In the range from about 12–88 rau, rau scores are very similar to the corresponding percentages.

In evaluating the data, it should be kept in mind that the listening environments were intended to be typical of those experienced in daily life. Speech levels and SBRs were known to produce full intelligibility for conversational speech among the normal talkers and normal-hearing listeners that served in the study. Thus all scores were relatively high. The aspect of the data that was under scrutiny was the absolute and relative similarity of scores obtained in the real and simulated versions of each environment.

A. Environment A

Figure 2 illustrates the composite intelligibility scores obtained in the real and simulated environment A listening conditions. The upper panel depicts scores for each talker across all four contrasts and the lower panel gives scores for each contrast across all three talkers. Both panels indicate that scores obtained in the simulated environment A were slightly higher than those obtained in the real environment A. These data were entered into a repeated-measures analysis of variance with three variables: listening condition (real versus simulated), talkers (three), and contrasts (four). To examine the effect of listening condition on scores, we were interested in the tests for the overall main effect of listening condition and the simple main effect of listening condition for each talker and each contrast. The analysis revealed that the difference between the overall composite scores of 100 rau in the real environment and 109 rau in the simulated environment was significant [$F(1,19) = 31.9$, $p < 0.001$]. In addition, the difference between listening conditions was significant ($p < 0.05$) for all talkers except talker 2 and for all contrasts except ICP. Despite the difference in overall scores between the real and simulated environments, Fig. 2 also shows that the patterns of mean intelligibility scores across talkers and across contrasts were essentially the same in both conditions. Talkers 2 and 4 were generally more intelligible than talker 5 in both listening conditions and the ranking of scores for the four contrasts was also the same in the two conditions.

The three-way interaction between listening conditions, talkers, and contrasts was also significant [$F(6,114) = 4.5$, $p < 0.001$]. Further analyses were performed using a separate analysis of variance for each of the two listening conditions. Both analyses produced a significant interaction between talkers and contrasts [real environment $F(6,114) = 9.2$, $p < 0.001$; simulated environment $F(6,114) = 4.9, p < 0.001$]. These were further explored using the Student–Neuman Keuls *post-hoc* procedure ($\alpha = 0.05$). The results are shown in Table II. Perusal of Table II reveals that both listening conditions produced a large number of significant differences between talkers for individual contrasts and between contrasts for individual talkers. For the most part, the ranking of scores for talkers and contrasts and the pattern of significant differences between scores were the same for the two listening conditions. However, the real environment produced a few more significant differences than were seen in the simulated environment. For example, for FCV there was a significant difference between talkers 4 and 2 in the real environment but not in the simulated environment. Similarly, talker 4 produced significantly lower scores for FCV than FCC in the real environment but not in the simulated environment. There were no significant differences seen in the simulated environment that were not also seen in the real environment.

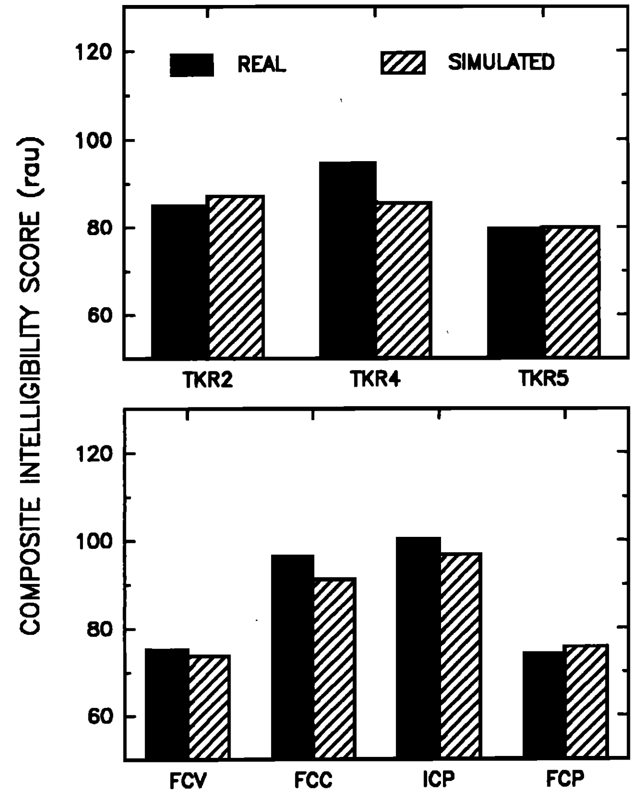

FIG. 3. Intelligibility scores obtained in the real and simulated environment B listening conditions. The upper panel depicts mean scores for each talker across all four contrasts. The lower panel gives mean scores for each contrast across all three talkers.

B. Environment B

Figure 3 illustrates the composite intelligibility scores obtained in the real and simulated environment B listening conditions. For most talkers and contrasts, differences in mean scores between the two listening conditions were small. A three-way analysis of variance (ANOVA) analogous to that performed for the environment A data confirmed that there was not a statistically significant difference between the overall scores of 84 rau in the simulated environment and 87 rau in the real environment. Further, there was no significant difference between scores in the two listening conditions for any contrast and the rankings of contrast intelligibility scores were similar in both listening conditions.

In the real environment, talker 2 was notably less intelligible than talker 4 and similar in intelligibility to talker 5. In the simulated environment, all three talkers were about equally intelligible. In other words, the intelligibility of talker 4 decreased in the simulated environment relative to the real environment. The test of simple main effects revealed that this change in score for talker 4 was significant [$F(1,76) = 7.4, p = 0.008$], whereas scores achieved for talkers 2 and 5 in the two listening conditions were not significantly different.

TABLE II. Results of *post-hoc* analyses of talker × contrast interactions in real and simulated environment A listening conditions. Scores increased from left to right. Underlining indicates scores that were not significantly different ($\alpha = 0.05$).

	Real environment				Simulated environment			
Contrast								
FCV	tkr5	tkr4	tkr2		tkr5	tkr2	tkr4	
FCC	tkr5	tkr4	tkr2		tkr5	tkr2	tkr4	
ICP	tkr5	tkr2	tkr4		tkr5	tkr2	tkr4	
FCP	tkr5	tkr4	tkr2		tkr5	tkr4	tkr2	
Talker								
tkr2	FCP	FCV	FCC	ICP	FCP	FCV	ICP	FCC
tkr4	FCP	FCV	FCC	ICP	FCP	FCV	FCC	ICP
tkr5	FCC	FCP	FCV	ICP	FCV	FCP	FCC	ICP

The three-way interaction (listening condition × talker × contrast) was not significant for these data. However, to allow comparison of the talker × contrast interactions in the real and simulated environments, further analyses were carried out using separate ANOVAs for each listening condition. Again, both analyses produced significant talker ×contrast interactions [real environment $F(6,114) = 5.7$, $p < 0.001$; simulated environment $F(6,114) = 6.1$, $p < 0.001$]. The results of *post-hoc* testing of these data are given in Table III. The Table reveals that the environment B listening conditions produced a large number of significant differences among talkers and contrasts. Again, the overall pattern of the results across talkers and contrasts was similar for both real and simulated environments. However, note that, in the real environment data, the intelligibility of contrasts produced by talker 4 was significantly higher than those produced by the other talkers for all four contrasts, whereas this did not occur in the simulated environment. This is consistent with the earlier analyses. The real environment again resulted in a few more significant differences

TABLE III. Results of *post-hoc* analyses of talker × contrast interactions in real and simulated environment B listening conditions. Scores increased from left to right. Underlining indicates scores that were not significantly different ($\alpha = 0.05$).

	Real environment				Simulated environment			
Contrast								
FCV	tkr5	tkr2	tkr4		tkr5	tkr2	tkr4	
FCC	tkr2	tkr5	tkr4		tkr5	tkr4	tkr2	
ICP	tkr5	tkr2	tkr4		tkr5	tkr4	tkr2	
FCP	tkr2	tkr5	tkr4		tkr2	tkr4	tkr5	
Talker								
tkr2	FCP	FCV	FCC	ICP	FCP	FCV	FCC	ICP
tkr4	FCP	FCV	FCC	ICP	FCP	FCV	FCC	ICP
tkr5	FCV	FCP	ICP	FCC	FCV	FCP	FCC	ICP

than seen in the simulated environment. There were no significant differences observed in the simulated environment that were not present in the real environment also.

C. Environment C

Figure 4 illustrates the composite intelligibility scores obtained in the real and simulated environment C listening conditions. As seen in environment B, there were minimal differences in mean scores for the two listening conditions within talkers and contrasts. A three-way ANOVA (variables = listening condition, talker, and contrast) confirmed that there was not a statistically significant difference between the overall scores of 102 rau in the simulated environment and 100 rau in the real environment. Further, there were no significant differences between listening conditions in scores for any talker or any contrast.

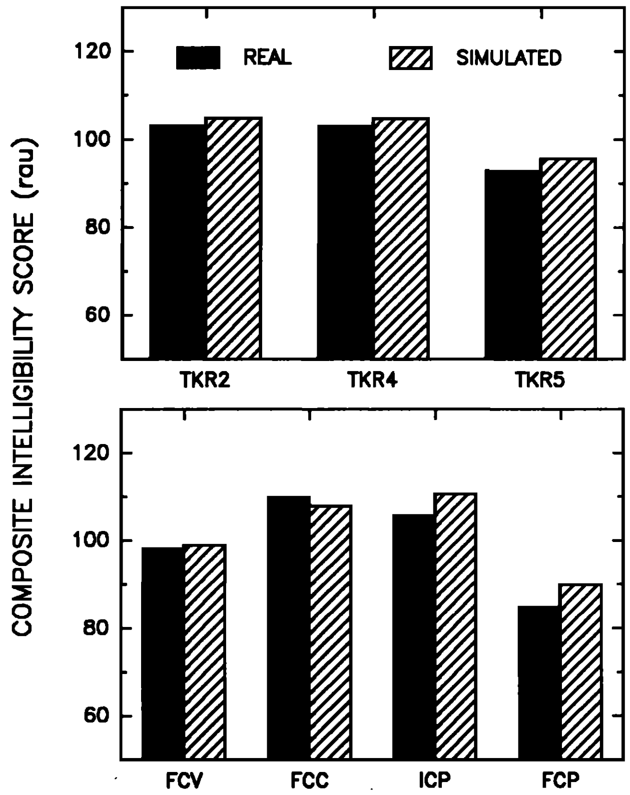

FIG. 4. Intelligibility scores obtained in the real and simulated environment C listening conditions. The upper panel depicts mean scores for each talker across all four contrasts. The lower panel gives mean scores for each contrast across all three talkers.

TABLE IV. Results of *post-hoc* analyses of talker × contrast interactions in real and simulated environment C listening conditions. Scores increased from left to right. Underlining indicates conditions for which scores were not significantly different ($\alpha = 0.05$).

	Real environment				Simulated environment			
Contrast								
FCV	tkr5	tkr2	tkr4		tkr5	tkr4	tkr2	
FCC	tkr5	tkr4	tkr2		tkr5	tkr2	tkr4	
ICP	tkr4	tkr5	tkr2		tkr2	tkr4	tkr5	
FCP	tkr5	tkr4	tkr2		tkr5	tkr2	tkr4	
Talker								
tkr2	FCP	FCV	ICP	FCC	FCP	FCV	FCC	ICP
tkr4	FCP	ICP	FCV	FCC	FCP	FCV	FCC	ICP
tkr5	FCP	FCV	FCC	ICP	FCV	FCP	FCC	ICP

Although the differences in intelligibility across talkers were relatively small in environment C, it can still be seen that, overall, talkers 2 and 4 were more intelligible than talker 5 in both real and simulated environments. Also, the rankings of mean contrast scores were similar in both conditions. The three-way interaction (listening condition × talker × contrast) was not significant. However, comparison of talker × contrast interactions in the real and simulated environments required separate ANOVAs for each listening condition. Again, significant interactions were seen between talker and contrast scores in both analyses [real environment $F(6,114) = 6.6$, $p < 0.001$; simulated environment $F(6,114) = 6.7$, $p < 0.001$]. The *post-hoc* test results for environment C data are given in Table IV.

Examination of Table IV substantiates the impression from Fig.4 that there were relatively small differences in intelligibility across talkers in this environment: Only two contrasts (FCV and FCP) produced significantly different scores for different talkers. The pattern was exactly duplicated in both real and simulated environments. Within talkers, the real environment produced a few significant differences between contrasts that were not seen in the simulated environment, but the general pattern of contrast intelligibility was similar in the two listening conditions. Environment C was the only environment to produce a significant difference in the simulated condition that was not present in the real condition: FCC versus ICP for talker 5.

III. DISCUSSION

The most global issue considered in evaluating the validity of the simulated environments was overall intelligibility: Were the simulations equal to the real settings in terms of the general ability to understand speech? Results showed that for environments B and C, the simulations were essentially equal to the real environments in this respect. However, we were surprised to note a discrepancy in difficulty between real and simulated environment A. Because this listening condition had a relatively good signal-to-babble ratio combined with little or no reverberation, we had assumed that it would be the easiest to simulate.

In an attempt to determine the basis for the difference in overall intelligibility between real and simulated environment A, we compared wideband spectrograms of utterances from both conditions. These analyses revealed that, despite the equivalent SBRs used in the two conditions, the recordings made in the real environment contained a somewhat higher level of background noise, spread across the entire spectrum. Because the noise was not present in the real environment, it must have been introduced by the recording equipment as a by-product of amplification required for the low levels in environment A. This effect was not as noticeable in recordings made in environments B and C because of the higher levels used in those settings. We hypothesize that this slight additional noise level produced the small but systematic difference in the overall intelligibility between the two environment A listening conditions. Despite this outcome, the general similarity in the pattern of results across real and simulated environments indicates that the simulation of environment A was fairly accurate.

Another factor under consideration was the relative intelligibility of different talkers in the real and simulated environments. Recall that the three talkers used in this study were chosen because, in the earlier study, one (talker 4) was highly intelligible in all three environments, another (talker 5) was relatively less intelligible in all three environments and the intelligibility of the third talker (2) varied across environments, being similar to that of talker 4 in environments A and C but more similar to talker 5 in environment B. The upper panels of Figs. 2–4 reveal that these patterns of talker intelligibility were observed in this study also for the real environments (solid bars) and in simulated environments A and C (hatched bars, Figs. 2 and 4). In simulated environment B (hatched bars, Fig. 3), the expected pattern of intelligibility was seen for talkers 2 and 5 but not for talker 4. In this environment, talker 4 was expected to be more intelligible than both 2 and 5. However, in simulated environment B, the intelligibility of all three talkers was about equal. Although talkers 2 and 5 maintained constant intelligibility across both listening conditions, the intelligibility of talker 4 was significantly poorer in the simulated reverberant environment than in the real reverberant environment. These results reveal that the method used to simulate environment B was not fully successful in the case of talker 4. Although the relative intelligibility of this talker's phonetic contrasts remained constant across the real and simulated environment B conditions (see Table III), the absolute intelligibility of talker 4 was not the same in the two listening conditions.

At present, we do not have an explanation for this outcome. It may be relevant to note that talker 4 was the only female talker and that her speech spectrum revealed relatively high levels in the 6- to 10-kHz frequency region (see Fig. 1). In the present investigation, the highest frequency region in which reverberation measurements were made was 4 kHz. This upper limit seemed appropriate because information in the frequency region above 4 kHz is relatively unimportant for intelligibility of average speech (Pavlovic, 1987). However, it is possible that this high-frequency region is important for intelligibility in individual talkers. Perhaps a significant difference between the two environment B listening conditions in processing high-frequency information was responsible for the result seen for talker 4. Further studies, employing reverberation measurements at higher frequencies will be necessary to explore this matter.

Comparison of the lower panels of Figs. 1–4 indicates that, overall, the relative intelligibility of the four phonetic contrasts remained quite constant across real and simulated listening conditions for all three environments. Furthermore, the analyses of talker × contrast interactions depicted in Tables II–IV indicated that, even at this detailed level of analysis, the intelligibility of individual contrasts in the speech of individual talkers remained notably consistent across the two listening conditions.

Except for the somewhat anomalous result for talker 4 in environment B, the main exceptions to the generally positive outcome of this study were a relatively small number of significant intelligibility differences that were seen in the real environments but not seen in the corresponding simulated environments. Irwin and McCauley (1987) reported that differences between simulated reverberant conditions were less than those between real reverberant conditions. The results of the present study were consistent with this observation and suggested that some intelligibility differences were less pronounced in all three of our simulated environments than they were in the corresponding real environments.

One factor that should be noted in assessing both the outcome and the implications of this investigation is the monaural nature of the listening task. Subjects listened monaurally, via earphone, in the real environment conditions and monaurally, with the nontest ear plugged, in the simulated environment conditions. Many hearing-impaired individuals wear only one hearing aid and in most cases they are essentially monaural listeners. Thus the conditions tested in this investigation are relevant to a large proportion of clinical hearing aid evaluations. However, extension of these findings to binaural listening conditions should be made with caution because of the potential for binaural interactions to differentially effect the SBR in real versus simulated environments.

Although the correspondence between real and simulated listening environments was less than perfect, the overall outcome of this investigation was encouraging regarding the feasibility of valid simulations of everyday listening environments in audiometric test rooms. These data indicate that, with appropriate adjustments of presentation level and signal-to-babble ratio, and with reverberation effects synthesized using relatively simple and widely available procedures, intelligibility distinctions that are observed in real environments usually can be validly reproduced in corresponding simulated environments. Thus score differences obtained in a simulated environment can be generalized with considerable confidence to the results that would be obtained for the same talker and the same speech material in the corresponding real environment. However, the fact that one talker's intelligibility underwent an unexpected deterioration in the simulated reverberant environment indicates that additional investigation is needed in this type of environment to more fully delineate possible interactions between talker intelligibility and simulation characteristics.

Finally, it should be recognized that, because the present study employed normal-hearing subjects, the results can be applied with full confidence only to other normal hearers. However, numerous studies in the literature suggest that normal-hearing and hearing-impaired listeners tend to respond in a similar way to intelligibility differences produced by, for example, different talkers (Harris *et al.*, 1961), different reverberation conditions (Helfer and Wilber, 1990), different distortions (Lawson and Chial, 1982) and different transmission systems (Gabrielsson, Schenkman, and Hagerman, 1988). These studies lend support to a hypothesis that the results for normal hearers in the present study would also apply to hearing-impaired listeners.

ACKNOWLEDGMENTS

This work was supported by funding from the Department of Veterans Affairs Rehabilitation Research and Development Service. Software to administer and score the SPAC test was developed by Robert M. Joyce.

ANSI. (**1989**). ANSI S3.6-1989. "Specification for Audiometers" (American National Standards Institute, New York).

Barcham, L. J., and Stephens, S. D. G. (**1980**). "The use of an open-ended problems questionnaire in auditory rehabilitation," Br. J. Audiol. **14**, 49–51.

Boothroyd, A. (**1985**). Measurement of speech feature perception in subjects with limited hearing," J. Acoust. Soc. Am. Suppl. 1 **77**, S107.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). "Intelligibility of average talkers in typical listening environments," J. Acoust. Soc. Am. **81**, 1598–1608.

Cox, R. M., and Gilmore, C. (**1990**). "Development of the profile of hearing aid performance (PHAP)," J. Speech Hear. Res. **33**, 343–357.

Cox, R. M., and Alexander, C. G. (**1991**). "Hearing aid benefit in everyday environments," Ear Hear. **12**, 127–139.

Gabrielsson, A., Schenkman, B., and Hagerman, B. (**1988**). "The effects of different frequency responses on sound quality judgements and speech intelligibility, "J. Speech Hear. Res. **31**, 166–177.

Golabek, W., Nowakowska, M. Siwiec, H., and Stephens, S. D. G. (**1988**). "Self-reported benefits of hearing aids by the hearing impaired," Br. J. Audiol. **22**, 183–186.

Hagerman, B., and Gabrielsson, A. (**1984**). "Questionnaires on desirable properties of hearing aids," Karolinska Inst. Rep. TA109.

Harris, J. D., Haines, H. L., Kelsey, P. A., and Clark, T. D. (**1961**). "The relation between speech intelligibility and the electro-acoustic characteristics of low-fidelity circuitry, "J. Aud. Res. **1**, 357–381.

Harris, R. W., and Goldstein, D. P. (**1979**). "Effects of room reverberation upon hearing aid quality judgements," Audiology **18**, 253–262.

Harris, R. W, and Reitz, M. L. (**1985**). "Effects of room reverberation and noise on speech discrimination by the elderly," Audiology **24**, 319–324.

Helfer, K. S., and Huntley, R. A. (**1989**). "Aging influences on consonant confusions in reverberation and noise," Paper presented at the national meeting of the American Speech–Language–Hearing Association, St Louis, MO; ASHA **31** (10), 185 (1989).

Helfer, K. S., and Wilber, L. A. (**1990**). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech Hear. Res. **33**, 149–155.

Irwin, R. J. and McCauley, S. F. (**1987**). "Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation," J. Acoust. Soc. Am. **81**, 1557–1565.

Kapteyn, T. S. (**1977**). "Factors in the appreciation of a prosthetic rehabilitation," Audiology **16**, 446–452.

Lawson, G. D., and Chial, M. R. (**1982**). "Magnitude estimation of degraded speech quality by normal- and impaired-hearing listeners," J. Acoust. Soc. Am. **72**, 1781–1787.

Lynn, J. M., and Lesner, S. A. (**1990**). "Comparison of hearing aid prescriptions: use, benefit and satisfaction," Audiol. Today **2** (2), 33.

Nabelek, A. K., and Robinette, L. (**1978**). "Reverberation as a parameter in clinical testing," Audiology **17**, 239–259.

Nabelek, A. K., and Dagenais, P. A. (**1986**). "Vowel errors in noise and in reverberation by hearing-impaired listeners," J. Acoust. Soc. Am. **80**, 741–748.

Nabelek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). "Reverberant overlap- and self-masking in consonant identification," J. Acoust. Soc. Am. **86**, 1259–1265.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1977**). "Speech levels in various noise environments," EPA Rep. No. 600/1-77-025.

Peutz, U. M. (**1971**). "Articulation loss of consonants as a criterion for speech transmission in a room," J. Audio Eng. Soc. **19**, 915–919.

Plomp, R. (**1977**). "Acoustical aspects of cocktail parties," Acustica **38**, 186–191.

Scherr, C. K., Schwartz, D. M., and Montgomery, A. A. (**1983**). "Follow-up survey of new hearing aid users," J. Acad. Rehabilitative Audiol. **16**, 202–209.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Walden, B. E., Schwartz, D. M., Williams, D. L., Holum-Hardegan, L. L., and Crowley, J. M. (**1983**). "Test of the assumptions underlying comparative hearing aid evaluations," J. Speech Hear. Disord. **48**, 264–273.

Walden, B. E., Demorest, M. E., and Helper, E. L. (**1984**). "Self-report approach to assessing benefit derived from amplification," J. Speech Hear. Res. **27**, 49–56.