Intelligibility of average talkers in typical listening environments

Robyn M. Cox
Department of Audiology and Speech Pathology, 807 Jefferson Avenue, Memphis State University, Memphis, Tennessee 38105

Genevieve C. Alexander and Christine Gilmore
Memphis V. A. Medical Center, 1030 Jefferson Avenue, Memphis, Tennessee 38104

(Received 1 October 1986; accepted for publication 23 December 1986)

Intelligibility of conversationally produced speech for normal hearing listeners was studied for three male and three female talkers. Four typical listening environments were used. These simulated a quiet living room, a classroom, and social events in two settings with different reverberation characteristics. For each talker, overall intelligibility and intelligibility for vowels, consonant voicing, consonant continuance, and consonant place were quantified using the speech pattern contrast (SPAC) test. Results indicated that significant intelligibility differences are observed among normal talkers even in listening environments that permit essentially full intelligibility for everyday conversations. On the whole, talkers maintained their relative intelligibility across the four environments, although there was one exception which suggested that some voices may be particularly susceptible to degradation due to reverberation. Consonant place was the most poorly perceived feature, followed by continuance, voicing, and vowel intelligibility. However, there were numerous significant interactions between talkers and speech features, indicating that a talker of average overall intelligibility may produce certain speech features with intelligibility that is considerably higher or lower than average. Neither long-term rms speech spectrum nor articulation rate was found to be an adequate single criterion for selecting a talker of average intelligibility. Ultimately, an average talker was chosen on the basis of four speech contrasts: initial consonant place, and final consonant place, voicing, and continuance.

PACS numbers: 43.71.Es, 43.71.Gv, 43.66.Ts [DW]

INTRODUCTION

People obtain hearing aids principally to improve their understanding of speech in everyday listening situations (Barcham and Stephens, 1980; Hagerman and Gabrielsson, 1984). Therefore, a major objective in hearing aid selection is to choose an instrument that will result in the greatest possible improvement in speech comprehension. To this end, the results of tests assessing speech recognition with each of several hearing aids often determine which instrument is ultimately recommended.

It has been amply demonstrated that the intelligibility of speech test material is fundamentally dependent upon the talker (e.g., Palmer, 1955; Weinhouse and Miller, 1963; Williams and Hecker, 1968; Kruel *et al.*, 1969; Hood and Poole, 1980). It has also been shown that the intelligibility of speech produced by a particular talker is significantly affected by instructions to speak in a normal/conversational manner versus a precise/clear manner (Tolhurst, 1957; Picheny *et al.*, 1985). Furthermore, Witter and Goldstein (1971) and Cox and McDaniel (1984) reported that the intelligibility of hearing-aid-processed speech interacted with talker with the result that the hearing aid identified as producing the most intelligible speech differed for different talkers.

Taken together, these reports lead to the conclusion that when a speech intelligibility test is used to determine either (a) which of several hearing aids provides the most improvement in speech understanding, or (b) the absolute amount of improvement provided by a particular instrument, the outcome will depend partly on the characteristics of the talker used to record the test materials and on the manner in which the materials are generated (conversational or clear). It is perhaps surprising that relatively little investigative attention has been paid to the issues involved in selection of talkers for speech intelligibility tests.

When a new intelligibility test is reported, the talker is typically described as lacking a pronounced regional accent: There are no other criteria commonly applied in talker selection. However, if the test results are used to predict hearing aid benefit for understanding everyday speech, it would appear, from the studies cited above, that care should be taken to assure that the talker's speech is average in intelligibility and that the speech itself is delivered in a conversational/normal manner when the test materials are recorded.

A review of the literature failed to reveal any investigations that provided analytic data describing the intelligibility, for normal hearing listeners, of normal talkers producing conversational speech in everyday listening settings. Hence, the present study was undertaken in an attempt to generate these data. The ultimate purpose in developing a description of the intelligibility of normal talkers was to provide a basis for selecting a talker of average intelligibility. In future work the chosen talker will produce speech materials for a speech intelligibility test to be used to quantify hearing aid benefit. This ultimate goal influenced some of the decisions made in designing the present study.

The research questions were:

(1) Are there significant intelligibility differences

 0001-4966/87/051598-11$00.80

among normal talkers when all are producing conversational speech in typical listening environments?

(2) Is there an interaction between talker intelligibility and listening environment? In other words, if a given talker is highly intelligible in a quiet living room, will that talker also be highly intelligible in other types of settings (e.g., in a classroom setting)?

(3) How intelligible are the phonetic features of normal conversational speech for normal hearers in typical listening settings?

(4) Is there an interaction between talker intelligibility and particular speech features? In other words, if a given talker obtains a relatively low score for intelligibility of one speech feature (e.g., final consonant voicing), will that talker also display relatively low intelligibility for all other speech features?

I. METHOD

A. Talkers

Three male and three female talkers were studied. They were chosen to satisfy the following criteria: (a) absence of unusual or atypical speech characteristics, (b) no pronounced regional dialect, and (c) ability to read prepared material in a manner similar to their spontaneous speech.

Each talker's long-term rms 1/3-oct-band speech spectrum is shown in Fig. 1. For each talker, a 1-min sample of continuous speech was analyzed to generate these spectra (Hewlett-Packard signal analyzer, model 3561A). The spectra have been normalized for overall level. In addition, each talker's articulation rate for the test sentences (syllables/s) is given in Fig. 1. These data were obtained using a digital spectrograph (Voice Identification, Inc., model RT1000). Each sentence was spectrographically displayed and the length of the sentence was determined from the beginning of the first syllable to the end of the last syllable. The articulation rate was computed as the average across 12 sentences. The range of 3.3–3.7 syllables/s is very similar to the range of 3.0–3.9 syllables/s reported by Picheny *et al.* (1986) for similar conversationally produced speech material.

B. Test stimuli

Talker intelligibility was quantified using the four segmental subtests of the speech pattern contrast (SPAC) test (Boothroyd, 1985a). This test yields the following contrast scores: (1) vowel height (high/low); (2) vowel place (front/back); (3) initial consonant voicing (voiced/voiceless); (4) final consonant voicing (voiced/voiceless); (5) initial consonant continuance (stop/continuant); (6) final consonant continuance (stop/continuant); (7) initial consonant place (more labial/more velar); and (8) final consonant place (more labial/more velar). In addition, the average of all eight contrast scores (the composite score) may be obtained. Boothroyd (1985b) reported a high correlation between the composite score and intelligibility of sentences produced by the same talker.

Each form of the SPAC test was composed of four subtests, each consisting of 12 test words and yielding two contrast scores. A complete form consisted of 48 test items and yielded the eight contrast scores noted above. It is a four-alternative forced-choice test and each response is scored for two different phonetic contrasts. For example, in the subtest that evaluates initial consonant place and final consonant

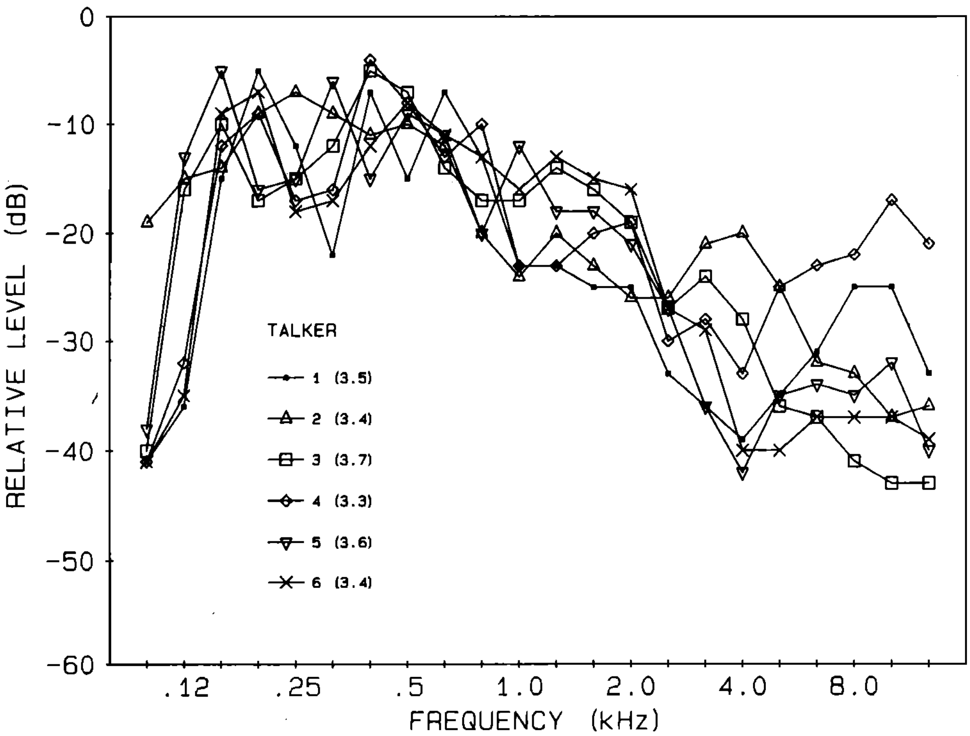

FIG. 1. Long-term rms 1/3-oct-band speech spectrum for each of the six talkers. Articulation rate (syllables/s) is given in parentheses for each talker.

place, one test word is "did" and the four alternatives include "did," "big," "bid," and "dig." A response of "did" is scored correct for both contrasts, the response "bid" is correct for final consonant place only, the response "dig" is correct for initial consonant place only, and the response "big" is incorrect for both contrasts. Thus, in a single form, each contrast score is based on 12 utterances. A complete list of the stimulus items is given in Boothroyd (1985b).

The SPAC test words were embedded in the sentences shown in Appendix A. These sentences were devised to present the items in a variety of contexts (as occurs in everday speech) with respect to preceding and following phonemes, position of test item in the utterance, and length of utterance. For each subtest, the 12 sentences were randomly assigned to the 12 test items. Each form (consisting of four subtests) was preceded by four practice items. Sentences for the practice items are also given in Appendix A.

1. Production of master recordings

There are 12 different forms of the SPAC test. The forms differ by selecting a different alternative as the stimulus word. Four forms of the test were generated by each talker in a sound treated audiometric room ($1.9 \times 1.8 \times 1.9$ m). Each talker recorded a different combination of four forms. Since each form consists of 48 utterances, each talker generated a total of 192 sentences. The speech was transduced by a 12.7-mm microphone (ACO, model 7013) located 30–40 cm from the talker's mouth, and recorded on magnetic tape (Panasonic AG6810 recorder). The effect of room acoustics on the recorded signal was negligible. In addition, samples of spontaneous speech and continuous reading were obtained from each talker. Finally, a 5-min segment of six-talker babble was recorded in the same room using the same instrumentation. This multitalker babble was used as the background noise in the environmental recordings. The long-term rms 1/3-oct-band spectrum of the multitalker babble is shown in Fig. 2. Also shown in Fig. 2 is the idealized speech spectrum given in ANSI S3.5-1969. The

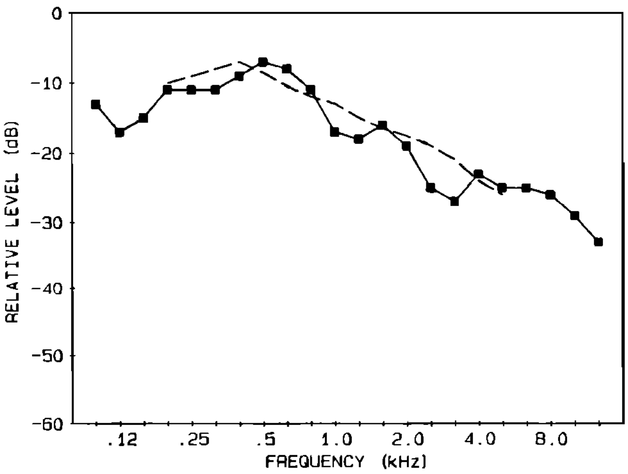

FIG. 2. Long-term rms 1/3-oct-band spectrum for the multitalker babble used in this investigation (filled squares) and the rms 1/3-oct-band idealized speech spectrum derived from ANSI S3.5-1969 (dashed line).

rms difference between the ANSI idealized speech spectrum and the multitalker babble spectrum used in this study was 2.9 dB.

C. Listening environments

Four basic listening environments were defined for evaluation. Both theoretical considerations and the data reported by Walden *et al.* (1984) suggest that these 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 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 distances. These investigators reported measurements of speech, background noise levels, and talker–listener distances that were maintained by talkers and listeners in everyday environments to allow essentially complete intelligibility for conversations in that setting.

Environment A represented face-to-face conversation in a typical living room or quiet office. The level of the primary message (sentences containing the SPAC items) was 55-dB L_{eq} (equivalent continuous dBA level) and the background noise (multitalker babble) was delivered at 48-dB L_{eq} (both measured beside the listener's ear). Talker–listener distance was 1 m. This environment was simulated in a 5.8- $\times$ 6.1-$\times$ 2.6-m room for which reverberation times as a function of frequency are given in Table I. This room contained carpeting, window drapes, and upholstered furniture.

Environment B represented listening as an audience member to speech delivered in an unamplified classroom, theater, church, etc. The primary message was delivered at a level of 70-dB L_{eq}, measured 1 m from the talker. The talker–listener distance was 5 m (the primary message level at the listener's ear was 63.5-dB L_{eq}). The competing babble was delivered at 55-dB L_{eq} beside the listener's ear. This environment was simulated in a room $18 \times 6.1 \times 3.2$ m (ceiling at rear 1/3 of room lowered to 2.6 m). This room, for which reverberation times are also shown in Table I, was uncarpeted and contained no significant wall covering. It held nonupholstered classroom chairs and several tables.

Environment C represented face-to-face conversation at a social event or in a public place with numerous people present. In addition to the data of Pearsons *et al.* (1977), the

TABLE I. Reverberation time (s) as a function of frequency for the two rooms used to simulate the four typical listening environments.

Freq. (Hz)	Envir. A and C1	Envir. B and C2
125	0.70	1.05
250	0.35	0.92
500	0.37	1.01
1000	0.39	0.91
2000	0.55	0.91
4000	0.57	0.85

report of Plomp (1977) was considered in selecting the speech-to-babble (S/B) ratio in this environment. The primary message level was 64-dB L_{eq} and the babble was delivered at 62-dB L_{eq} (both measured beside the listener's ear). The talker–listener distance was 0.5 m. This environment was simulated in both of the rooms described above. The two versions of this environment are referred to as environment C1 (average reverberation such as might be encountered at a social event in a private home) and environment C2 (longer reverberation such as might be encountered in a restaurant or church social).

The reverberation times measured in the two rooms used as listening environments were very similar to those reported by Formby (1977) for three living rooms and a large classroom, respectively, and by Nábělek and Pickett (1974) for typical small to medium sized rooms.

D. Environmental recordings

The master recordings of the SPAC subtests were rerecorded in each of the basic listening environments (A, B, C1, and C2). Both the SPAC items and the multitalker babble were replayed on audiocassette recorders (Tascam 122 and Nakamichi BX-300, respectively), amplified (Crown D-75 amplifiers), and transduced by small loudspeakers (Radio Shack Minimus-7). The frequency response of this reproduction system, measured in a highly damped audiometric test room, was flat, ± 5 dB from 100 Hz to 14 kHz. Five loudspeakers were used, one designated as the "talker" and four others which produced uncorrelated multitalker babble. Studebaker (1985) has reported that signals produced by loudspeakers of this size have essentially the same dispersion characteristics as the human voice whereas larger loudspeakers do not.

The "listener" was a KEMAR manikin equipped with a Zwislocki-type ear simulator coupler (Industrial Research Products, model DB-100) terminated with a 12.7-mm microphone (ACO, model 7013). The talker loudspeaker was located at a 0-deg azimuth to the manikin. In each environment, the babble loudspeakers were arranged around the manikin (approximate distances from the manikin were: environment A, 1 m; environment B, 4 m; environment C1, 0.75 m; environment C2, 4 m). During recordings, the output from the manikin's "eardrum" was amplified using a precision sound level meter with associated preamplifier (Larson·Davis model 800 B), and recorded on magnetic tape (Panasonic AG6810 recorder).

In each environment, two forms of the SPAC subtests were recorded for each talker. A different combination of two forms was used in each environment. For example, one talker's master recordings included forms E, F, G, and H. For this talker combination, EH was rerecorded in environment A, FH was used in environment B, GF was used in environment C1, and EF was used in environment C2. Different combinations of two forms were used in the reduced intelligibility conditions, described below; all forms were used an equal number of times overall.

The recordings of the six talkers were presented at an equal integrated rms level. To adjust the SPAC items appropriately for each environment, the recording for talker #1 was adjusted to achieve the primary message level (in dB L_{eq}) for that environment. The remaining talkers were presented at the same overall level as talker #1. Hence, the spectra shown in Fig. 1 accurately portray the relative levels of the various talkers in the environmental recordings.

The two SPAC forms recorded for each talker in each basic listening environment will be referred to as "typical intelligibility" conditions because they were adjusted to S/B ratio values that are maintained by normal talkers and listeners to yield essentially full intelligibility for conversations in these environments (according to the data of Pearsons *et al.*, 1977). In addition, the remaining two SPAC forms for each talker were recorded in the same settings but with the background noise level increased sufficiently to reduce all contrast scores to less than 100%. The amount of increase in background noise necessary to achieve this differed across the four environments and was selected empirically in each environment. In environments A, B, C1, and C2, the babble was increased 10, 10, 7, and 3 dB, respectively. These four conditions will be referred to as "reduced intelligibility" conditions. The inclusion of these conditions was necessary because it was anticipated that some speech features, notably vowel height, vowel place, and consonant voicing, would be completely intelligible in the typical intelligibility conditions in some environments. The reduced intelligibility conditions were employed in an attempt to decrease the ceiling effect for these highly intelligible speech features and, therefore, provide more information about differences among talkers.

In summary, for each talker, two forms of the SPAC subtests were recorded in each of the four environments with S/B ratio adjusted for typical intelligibility. In addition, two different forms of the SPAC subtests were recorded in each environment with the S/B ratio adjusted for reduced intelligibility.

E. Subjects

Four groups of ten subjects responded to the test recordings—one group for each environment. All subjects were young adults who passed a hearing screening (250–8000 Hz) at 15 dB HL. Some subjects served in more than one group.

F. Procedure

The environmental recordings were replayed monaurally to subjects (Tascam 122 audiocassette player) with the output transduced by an insert earphone (Etymotic Research, model ER-2) coupled to the ear canal using a compressible foam earplug. 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 environment where the recording was made. Calibration of the playback levels was achieved with the ER-2 earphone attached to a Zwislocki-type ear simulator coupled to a sound level meter (Larson·Davis 800B). The levels were adjusted to be equal to those at the manikin's eardrum in the original environment. The frequency response of the record/playback system (from the manikin's eardrum to the average subject's eardrum) was flat, ± 2 dB, from 100 Hz to at least 11 kHz.

Delivery and scoring of the test items were controlled by an Apple IIe microcomputer system. For a given environment, each subject responded to all conditions, thus auditing 24 SPAC forms (6 talkers $\times$ 2 S/B ratios $\times$ 2 forms). Presentation of typical and reduced intelligibility conditions was randomized. The two forms for a given condition were treated as a single test and were presented consecutively. Scores for the two forms were combined so that all contrast scores for a given talker were based on responses to 24 utterances by that talker.

II. RESULTS

The obtained data were proportions of correct responses for each 24-item subtest (two 12-item forms combined). These proportions were arcsin transformed before statistical analyses. For descriptive purposes, the data were corrected for guessing as described by Boothroyd (1985b) and expressed in a form similar to percentages. Thus a score of 0% signifies performance at chance level (negative corrected scores occur when performance is poorer than chance level) and a score of 100% represents perfect intelligibility.

A. Intertalker differences in overall intelligibility

In each environment, each talker's overall intelligibility for both typical and reduced intelligibility conditions was quantified using the SPAC composite score. These data were subjected to a three-factor analysis of variance, split-plot design (4 environments $\times$ 6 talkers $\times$ 2 S/B ratios). All main effects and interactions were significant ($p < 0.02$). Tests of simple simple main effects were performed to examine (a) intertalker differences within each S/B ratio condition and (b) talker–environment interactions.

Figure 3 shows the average composite score for each talker in the typical intelligibility condition in each environment. Although it is not obvious by examining the averaged scores, individual contrast scores of 100% were not uncommon in the typical intelligibility listening conditions, particularly for the more intelligible talkers. This would be expected to reduce the composite score differences among talkers. Nevertheless, tests of simple simple main effects for these data revealed that talker #4 was significantly more intelligible than talkers #1 and #5 in all environments and talker #2 was significantly more intelligible than talkers #1 and #5 in all environments except B ($p < 0.05$). Talkers #3 and #6 were not consistently differentiated from either the high-intelligibility talkers (#2 and #4) or the low-intelligibility talkers (#1 and #5).

Figure 4 shows the mean composite score for each talker in the reduced intelligibility condition in each environment. The important feature of this figure is the difference between talkers within each environment. (It should be remembered that the S/B ratios were decreased by different amounts in the four environments to achieve the reduced intelligibility conditions. Hence, no significance should be attached to the relationships between environments in Fig. 4.) Comparison with Fig. 3 suggests that, as expected, the differences between talkers were usually greater in the reduced intelligibility conditions. However, tests of simple simple main effects exploring intertalker differences in the reduced intelligibility conditions produced an outcome identical to that for the typical intelligibility conditions: Talker #4 was significantly more intelligible than talkers #1 and #5 in all environments, and talker #2 was significantly more intelligible than talkers #1 and #5 in all environments except B ($p < 0.05$). Again, talkers #3 and #6 were not consistently differentiated from either the high-intelligibility talkers (#2 and #4) or the low-intelligibility talkers (#1 and #5).

B. Talker–environment interactions

The data shown in Figs. 3 and 4 indicate that the most intelligible talker overall (#4) was highly intelligible in all environments and under both typical and reduced intelligibility conditions. Similarly, the least intelligible talkers (#1 and #5) maintained their low intelligibility in all conditions. To further explore the ordering of talker intelligibility in the different environments, Pearson product-moment correlation coefficients were determined for each pair of en-

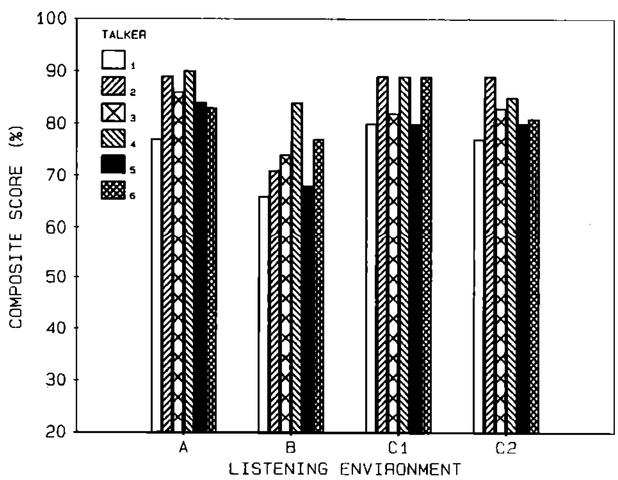

FIG. 3. Average composite intelligibility score for each talker in each typical listening environment.

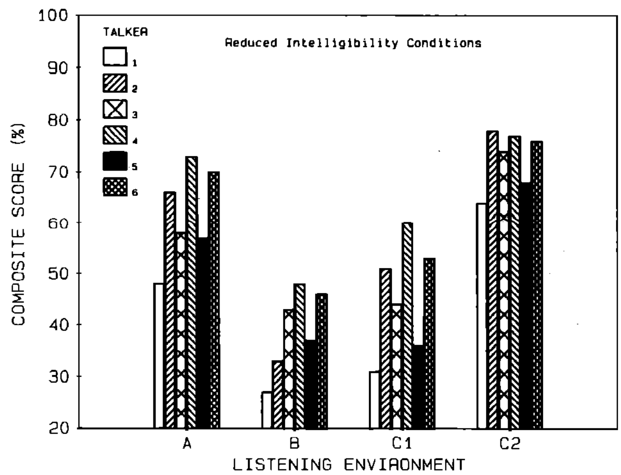

FIG. 4. Average composite intelligibility score for each talker in the reduced intelligibility condition in each environment.

vironments using the composite scores for each talker in the reduced intelligibility conditions. The results are given in Table II. High correlations were observed between environments A, C1, and C2, whereas all correlations involving environment B were considerably lower. Inspection of the data revealed that this outcome was attributable to the sharply reduced intelligibility of talker #2 in environment B. Although talker #2 was relatively highly intelligible in three of the four environments, his intelligibility was relatively poor in environment B (this effect can also be seen in the typical intelligibility conditions). Of the six talkers studied, only talker #2 gave this kind of result. The other five talkers maintained their relative intelligibility across all environments and in both typical and reduced intelligibility conditions.

C. Intelligibility of speech features in typical listening environments

Figure 5 shows the mean score obtained for each phonetic contrast in each typical listening environment. Results for all talkers are averaged in this figure. Even though the typical listening conditions employed primary and competing message levels that provide essentially full intelligibility for conversations for normal listeners, it is evident from Fig. 5 that not all phonetic features were fully intelligible. Also, features that were relatively less intelligible in one environment tended to be less intelligible in the other environments as well. These data were subjected to repeated measures analysis of variance (contrasts × environments). All main effects and interactions were statistically significant ($p < 0.01$). *Post hoc* testing was performed to explore intelligibility differences among the eight phonetic features in each environment. These analyses revealed that initial consonant continuance (icc) and final consonant place (fcp) were significantly less intelligible than vowel height and place (vht and vpl), and final consonant continuance (fcc) in all environments ($p < 0.01$). Initial and final consonant voicing (icv and fcv) and initial consonant place (icp) were not consistently differentiated from the other contrasts.

Figure 5 also illustrates that for every feature the poorest intelligibility score was obtained in environment B, the classroom. Apparently, environment B was the least intelligible setting of all in spite of the fact that the S/B ratio measured at the manikin's ear in this environment was the best of the four (+ 8.5 dB compared to + 7 dB in environment A and 12 dB in environments C1 and C2). *Post hoc* testing revealed that the mean scores obtained in environment B were significantly poorer than those obtained in the other

TABLE II. Pearson product-moment correlation coefficients between each pair of environments, derived from the composite intelligibility scores for each talker in the reduced intelligibility conditions.

	Envir. B	Envir. C1	Envir. C2
Envir. A	0.78	0.98	0.91
Envir. B		0.77	0.66
Envir. C1			0.93

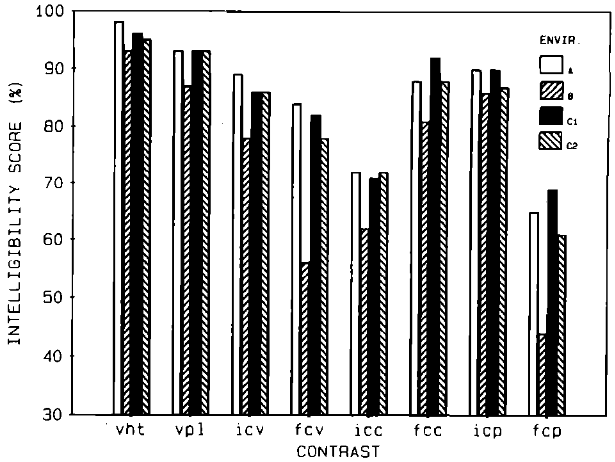

FIG. 5. Average intelligibility score obtained for each phonetic contrast in each typical listening environment. Data for all talkers are combined.

three environments for all contrasts except vowel height (vht), initial consonant continuance (icc), and initial consonant place (icp) ($p < 0.05$).

In environments A, C1, and C2, the range of mean intelligibility scores for individual features never exceeded 10% and was usually less than 5%. *Post hoc* testing revealed that the mean scores obtained in these three environments were not significantly different for any contrast ($p > 0.05$). This outcome confirms the *a priori* supposition that these environments were about equally intelligible.

D. Talker–feature interactions in typical listening environments

To examine interactions between talkers and feature intelligibility, the mean intelligibility score for each phonetic contrast was determined for each talker in each typical listening environment. These data are shown in Fig. 6. For each environment, the contrast scores were subjected to repeated measures analysis of variance (contrasts × talkers). All main effects and interactions were significant ($p < 0.01$). Tests of simple main effects were performed to explore the differences among talkers for each contrast. Because of the large number of *post hoc* tests, a significance level of $p < 0.01$ was adopted.

Examination of the four panels of Fig. 6 reveals that the overall pattern in each environment was consistent across talkers: All talkers displayed relatively high intelligibility for vowel height and vowel place and relatively poor intelligibility for final consonant place and initial consonant continuance. At a more microscopic level, however, numerous significant differences among talkers emerged. In all environments, at least five of the eight contrasts significantly differentiated among the talkers. In Fig. 6, the contrasts that did not differentiate among talkers are denoted with a filled square.

Within the contrasts showing significant differences among talkers, a pattern of one or two most intelligible (but not different) talkers, and one or two least intelligible (but not different) talkers, was commonly seen. Three contrasts (fcv, fcc, and icp) differentiated significantly among the

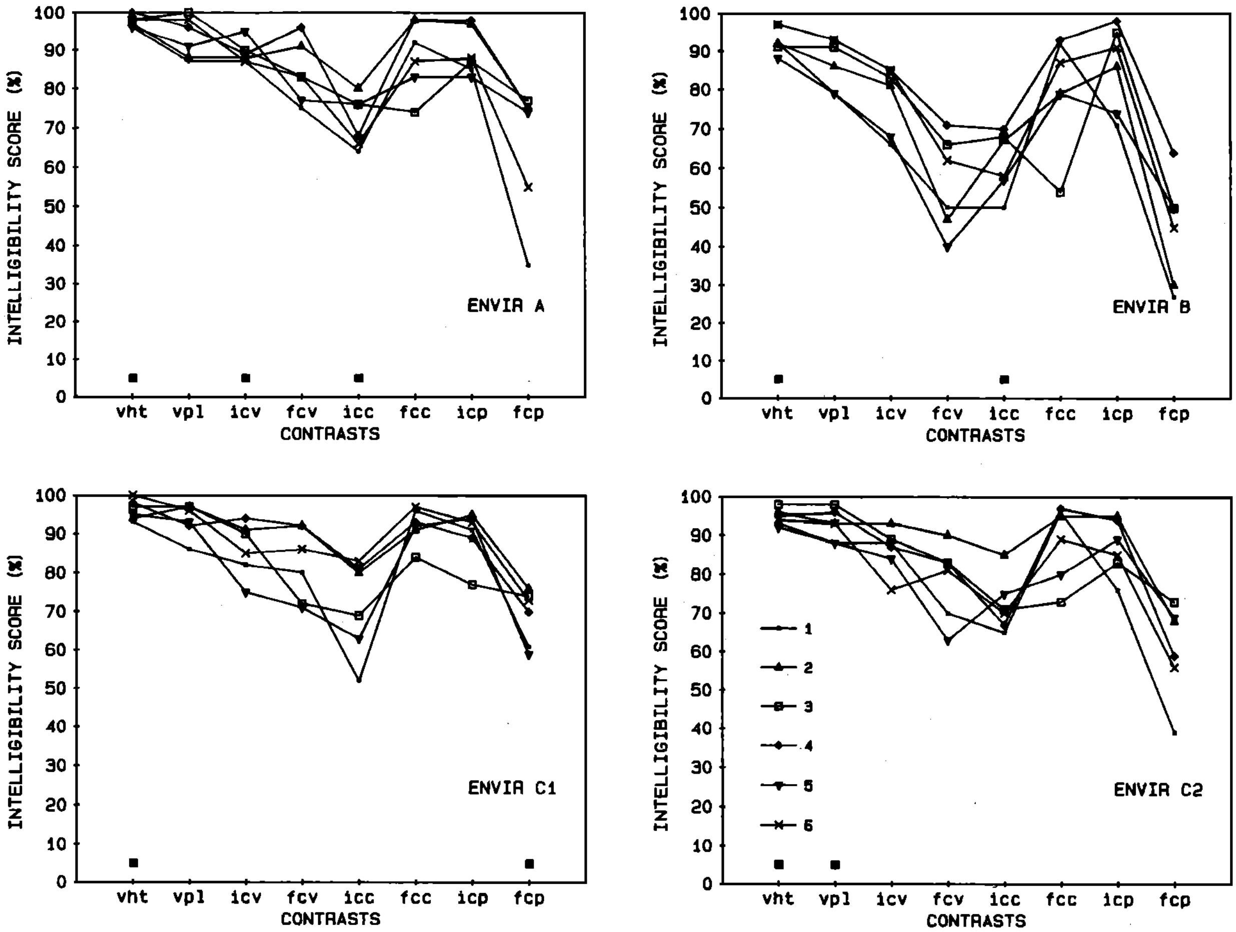

FIG. 6. Intelligibility scores for each phonetic contrast for each talker in each typical listening environment. The filled squares at the bottom of each panel indicate the contrasts for which the differences among talkers were not statistically significant.

talkers in all four environments. Three other contrasts (vpl, icv, and fcp) differentiated among the talkers in three of the four environments. Details of the *post hoc* analyses for the significant contrasts are given in Appendix B.

Examination of Fig. 6 reveals that a talker yielding a high score for one contrast did not necessarily yield a high score for other contrasts in which significant intertalker differences were found. For example, in environment A, talker #3 scored significantly higher than the lowest scorers for vowel place (vpl) and final consonant place (fcp) but significantly lower than the highest scores for final consonant continuance (fcc). This pattern can be observed for talker #3 in all four environments. A second example can be seen in the data for talker #1. This talker yielded relatively poor intelligibility for most of the contrasts that differentiated among talkers. However, her scores for final consonant continuance (fcc) were invariably among the highest. These observations support the conclusion that interactions between talkers and feature intelligibility are frequently seen in normal conversational speech produced in typical listening environments.

As discussed earlier, the data shown in Fig. 4 illustrate that talkers generally maintained their relative overall (composite) intelligibility across environments. The data shown in Fig. 6 (and Table BI) provide an opportunity to assess the extent to which the talkers' relative intelligibility for individual phonetic contrasts was maintained across environments. Examination of these data indicates that talkers who scored in the most intelligible group for a particular contrast in one environment tended to appear in the most intelligible group for that same contrast in the other environments where that contrast differentiated significantly among talkers. Similarly, talkers scoring in the least intelligible group for a particular contrast tended to maintain relatively low intelligibility for that contrast across all environments. There were, however, a few occasions when a talker who scored in the lowest intelligibility group for a particular contrast in one environment also scored in the highest intelligibility group for that same contrast in another environment.

III. DISCUSSION

Although it has been well established that talkers differ in the intelligibility of their speech, previous work has tended to concentrate on intelligibility differences for normal hearers listening to faint, masked, or distorted monosyllabic word lists. It would be reasonable to postulate that these intertalker differences would not be seen in typical listening situations where conversations are fully intelligible for normally hearing persons. However, the results of this investigation clearly indicate that intelligibility differences among normal talkers do persist in typical listening settings and for conversationally produced speech (Fig. 3). This outcome indicates that the talker for a test of everyday speech understanding should be carefully selected to produce speech having average intelligibility.

Figure 3 shows that overall intelligibility of speech contrasts was better than 75% for all talkers in environments A, C1, and C2, but was considerably poorer than this in environment B. Because the listening conditions were intentionally specified at S/B ratios that are reported to allow essentially complete intelligibility for everyday conversations, it was expected that overall intelligibility scores would be high. The relatively poor intelligibility performance in environment B was not anticipated. This environment, the simulated classroom, was fundamentally different from the other three environments in that the listener was separated from the talker by more than the critical distance [estimated as 3.7 m using the formula described by Peutz (1971)]. Hence, a high proportion of reverberant sound was present in the speech reaching the listener's ear. The S/B ratio reported for this type of environment by Pearsons *et al.* (1977) was determined by measuring the signal levels maintained by teachers relative to the background noise generated by the pupils (or other sources). Perhaps the S/B ratio maintained by the teachers was less than really necessary to achieve good intelligibility. If so, this would explain the poorer scores obtained in this study in the classroom environment. In any case, it appears that the S/B ratio reported for classrooms by Pearsons *et al.* was not overly pessimistic since a recent report by Markides (1986) noted even poorer S/B ratios in classroom settings.

Two environments, C1 and C2, simulated face-to-face conversations in a social setting with numerous people present. Two versions of environment C were included because previous investigations have shown that under some circumstances the combination of background noise and reverberation results in lower speech intelligibility than background noise alone (e.g., Finitzo-Hieber and Tillman, 1978). In this study, the social setting was simulated in rooms having different reverberation times to explore the possibility that conversationally produced speech would be less intelligible in the room with the longer reverberation time. As Figs. 3 and 5 reveal, this did not occur; the longer reverberation time in environment C2 did not result in significantly lower intelligibility scores than seen in environment C1 when S/B ratio and talker/listener distance were adjusted to simulate those observed in similar daily life settings (a different outcome would be expected if the talker/listener distance was greater). However, it is noteworthy that mean scores for final consonant voicing (fcv), final consonant continuance (fcc), and initial and final consonant place (icp and fcp) were all lower (though not significantly so) in environment C2 than in C1.

Five of the six talkers maintained their relative overall intelligibility in each of the four basic listening environments (Fig. 4 and Table II). This outcome suggests that a talker who produces average speech in one environment may usually be expected to produce average speech in other typical environments. However, there was on exception: The intelligibility of talker #2's speech was disproportionately reduced in environment B. The anomalous outcome for talker #2 in environment B suggests that certain voices may be more than usually susceptible to the degrading effect of reverberation. An examination of Fig. 1 does not reveal any striking differences between talker #2 and the other talkers (in long-term rms speech spectrum or articulation rate) that would clearly account for this result. It is noteworthy, however, that talker #2's speech spectrum level in the 100 Hz 1/3-oct band is about 20 dB higher than that of the other talkers. It is conceivable that this unusually high level of low-frequency energy combined with the relatively long low-frequency reverberation time in environment B to produce a particularly degraded speech signal. This suggestion implies that as talker–listener distance approaches or exceeds the critical distance, interaction between a talker's long-term average speech spectrum and room reverberation may influence speech intelligibility in a way that is not easily predicted from that talker's intelligibility in less reverberant listening environments.

Many studies of hearing impaired persons and of normally hearing individuals listening to degraded speech have established that vowel perception is the least affected by the degrading condition, followed, in turn, by the voicing, manner, and place features (e.g., Miller and Nicely, 1955; Oyer and Doudna, 1959; Boothroyd, 1984). The results of this investigation show that feature perception follows the same pattern for normal hearers listening to conversationally produced speech in typical listening environments. In all environments, the vowel score (mean of vht and vpl) was the highest, followed, in turn, by voicing (mean of icv and fcv), continuance (mean of icc and fcc), and place (mean of icp and fcp). In addition, previous studies have indicated that features are better received in the word initial position (e.g., Owens and Schubert, 1968; Levitt and Resnick, 1978). This pattern was also observed in this study: Across all features the word initial score exceeded the word final score in each environment.

There was, however, one noteworthy anomaly in the data: The scores for initial consonant continuance (icc) were consistently lower than those for final consonant continuance (fcc). The most likely explanation for this unexpected outcome appears to be in the SPAC test items used for these contrasts: The subtest for final consonant continuance is composed of 12 items contrasting either d/z or t/s. The subtest for initial consonant continuance contains four items contrasting either d/z or t/s and eight items contrasting either p/f or b/v. Correct recognition of continuance for these items is dependent on perception of the duration of the high-frequency nonperiodic component. However, the group of phonemes d/t/z/s is considerably more powerful in natural speech than the group p/f/b/v (Fletcher, 1953). Hence, one would expect contrasts involving p/f or b/v to be less discriminable than d/z or t/s in conversationally produced speech. This result would probably not be observed if the test lists were recorded in a manner that equalized the intensity of the test items.

Across all talkers, the overall pattern of contrast intelligibility was similar to that shown in previous investigations.

However, as Fig. 6 illustrates, individual talkers varied significantly from each other, even in the highly intelligible environment simulating an average living room (environment A). In addition, since there were numerous significant interactions between talkers and contrast intelligibility, it is clear that a talker who exhibits average intelligibility overall may produce some phonetic contrasts with high intelligibility and others with low intelligibility.

A. Selecting an average talker

A major purpose of this investigation was to provide a basis for selecting a talker of average intelligibility. The results indicate that the characteristics of an average talker are not easily described. Examination of Fig. 1 reveals that there were no clear differences in the speech spectra of the individual talkers that could be used to predict their intelligibility. For example, it would be reasonable to postulate that a talker's speech level in the high frequencies would be predictive of his/her overall intelligibility. Of the six talkers studied, the most intelligible talker (#4) did indeed display the greatest speech energy above 4 kHz. However, the least intelligible talker (#1) ranked second in this respect, suggesting that average high-frequency energy *per se* is not necessarily predictive of intelligibility. Hence, an average long-term speech spectrum would not be a valid criterion for selection of an average talker.

It is well established that the temporal characteristics of speech are related to its intelligibility with slower speech being more intelligible than faster speech. Among the talkers studied here, the most intelligible talker was the one with the slowest speech (3.3 syll/s). However, the talker with the fastest speech (talker #3) was not the least intelligible. These results do not support the use of articulation rate as a single criterion for selecting a talker of average intelligibility.

Furthermore, in selecting an average talker, it is clearly not sufficient to choose a talker whose SPAC composite intelligibility score is average for the test environment since this by no means ensures that the individual speech features will be produced with average intelligibility.

As a result of these considerations, it was decided that the average talker would be selected on the basis of his/her intelligibility for several speech contrasts. Four contrasts were chosen: initial consonant place (icp) and final consonant voicing, continuance, and place (fcv, fcc, and fcp). Three contrasts (icp, fcv, and fcc) were included because they differentiated significantly among talkers in all four listening environments. Final consonant place (fcp) was also included because it differentiated among talkers in three of the four environments, and because the place feature is the most susceptible to speech degradation resulting from hearing impairment (thus it seems particularly important that speech intended for quantifying speech recognition ability of the hearing impaired should be average in production of this contrast). Vowel place (vpl) and initial consonant voicing (icv) were excluded even though they significantly differentiated among talkers in three of the four environments because vowel perception and word initial voicing are known to be rather highly intelligible even to persons with severe hearing losses (see Revoile and Pickett, 1982, for a review). Vowel height (vht) and initial consonant continuance (icc) were excluded because they did not strongly differentiate among talkers.

To determine the average talker of the six included in this investigation, the rms deviation of the talker's score from the average score for contrasts fcv, fcc, fcp, and icp was derived for each talker in each environment. The talker with the smallest rms deviation was considered the average talker overall. Using this method, talker #6 was the average talker in all four listening environments. For this talker, a female, the rms deviation was 3%–5% in the different environments. The rms deviations for the other talkers ranged from 4.6% to 16% across environments.

ACKNOWLEDGMENTS

Appreciation is expressed to Arthur Boothroyd for providing information on the use of the SPAC test. R. Sherbecoe, J. Matesich, and M. Joyce generated the software to present and score the SPAC test. The administrative assistance of Kay M. Pusakulich is gratefully acknowledged. This work was supported by Veterans Administration Rehabilitation Research and Development Project #344.

APPENDIX A: SENTENCES USED TO PRESENT SPAC TEST AND PRACTICE ITEMS

Practice items:

Is_____up there now?
Pick out_____please.
I want you to choose_____right now.
Show me_____.

Test items:

Can you find_____now?
The next word is_____.
Identify_____please.
Choose the word_____next.
Look for_____on the screen.
Show me_____next.
Find the word_____there.
I'd like you to pick_____now.
You should make_____your next choice.
Do you see_____up there?
Point out the word_____.
You can choose next.

APPENDIX B: DETAILED RESULTS OF *POST HOC* TESTS OF TALKER × CONTRAST INTERACTIONS

TABLE BI. Results of *post hoc* analyses of intellgibility data for the phonetic contrasts that differentiated significantly ($p < 0.01$) among the six talkers in the typical listening environments. Underline indicates the talkers for whom intelligibility was not significantly different. The letters vpl = vowel place, icv = initial consonant voicing, fcv = final consonant voicing, icc = initial consonant continuance, fcc = final consonant continuance, icp = initial consonant place, and fcp = final consonant place.

Contrast	Environment	Talkers					
vpl	A	1	2	5	4	6	3
vpl	B	5	1	2	3	4	6

TABLE BI. (*Continued.*)

Contrast	Environment	Talkers					
vpl	C1	1	4	5	2	6	3
icv	B	1	5	2	3	4	6
icv	C1	5	1	6	2	3	4
icv	C2	6	5	1	3	4	2
fcv	A	5	1	6	3	2	4
fcv	B	5	2	1	6	3	4
fcv	C1	5	3	1	6	4	2
fcv	C2	5	1	6	3	4	2
icc	C1	1	5	3	2	4	6
icc	C2	1	4	6	3	5	2
fcc	A	3	5	6	1	4	2
fcc	B	3	2	5	6	1	4
fcc	C1	3	2	4	5	1	6
fcc	C2	3	5	6	2	1	4
icp	A	6	3	1	5	2	4
icp	B	1	5	2	6	3	4
icp	C1	3	4	1	6	5	2
icp	C2	1	3	6	5	4	2
fcp	A	1	6	2	4	5	3
fcp	B	1	2	6	3	5	4
fcp	C2	1	6	4	2	5	3

ANSI (**1969**). ANSI S3.5-1969, "American National Standard Method for the Calculation of the Articulation Index" (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. (**1984**). "Auditory Perception of Speech Contrasts by Subjects with Sensorineural Hearing Loss," J. Speech Hear. Res. **27**, 134–144.

Boothroyd, A. (**1985a**). "Measurement of Speech Feature Perception in Subjects with Limited Hearing," paper presented at the 109th meeting of the Acoustical Society of America, Austin, TX [J. Acoust. Soc. Am. Suppl. 1 **77**, S107 (1985)].

Boothroyd, A. (**1985b**). "Evaluation of Speech Production of the Hearing Impaired: Some Benefits of Forced-Choice Testing," J. Speech Hear. Res. **28**, 185–196.

Cox, R. M., and McDaniel, D. M. (**1984**). "Intelligibility Ratings of Continuous Discourse: Application to Hearing Aid Selection," J. Acoust. Soc. Am. **76**, 758–766.

Finitzo-Heiber, T., and Tillman, T. W. (**1978**). "Room Effects on Monosyllabic Word Discrimination Ability for Normal and Hearing Impaired Children," J. Speech Hear. Res. **21**, 440–448.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York), p. 84.

Formby, C. (**1977**). "The Effect of Environment on the Directional Performances of Directional Hearing Aids," M. A. thesis, Memphis State University.

Hagerman, B., and Gabrielsson, A. (**1984**). "Questionnaires on Desirable Properties of Hearing Aids," Karolinska Institute Rep. TA109, Stockholm.

Hood, J. D., and Poole, J. P. (**1980**). "Influence of the Speaker and Other Factors Affecting Speech Intelligibility," Audiology **19**, 434–455.

Kruel, E. J., Bell, D. W., and Nixon, J. C. (**1969**). "Factors Affecting Speech Discrimination Test Difficulty," J. Speech Hear. Res. **12**, 281–287.

Levitt, H., and Resnick, S. B. (**1978**). "Speech Reception by the Hearing Impaired: Methods of Testing and the Development of New Tests," in *Sensorineural Hearing Impairment and Hearing Aids*, Scandinavian Audiology, Supplementum No. 6, edited by C. Ludvigsen and J. Barfod, Chap. II, pp. 107–130.

Markides, A. (**1986**). "Speech Levels and Speech-to-noise Ratios," Br. J. Audiol. **20**, 115–120.

Miller, G. A., and Nicely, P. E. (**1955**). "An Analysis of Perceptual Confusion Among Some English Consonants," J. Acoust. Soc. Am. **27**, 338–352.

Náběलek, A., and Pickett, J. M. (**1974**). "Binaural Speech Perception Through Hearing Aids Under Noise and Reverberation with Normal and Hearing Impaired Listeners," J. Speech Hear. Res. **17**, 724–739.

Oyer, J. H., and Doudna, M. A. (**1959**). "Structural Analysis of Word Responses Made by Hard of Hearing Subjects on a Discrimination Test," Arch. Otolaryngol. **70**, 357–364.

Owens, E., and Schubert, E. D. (**1968**). "The Development of Constant Items for Speech Discrimination Testing," J. Speech Hear. Res. **11**, 656–667.

Palmer, J. M. (**1955**). "Effects of Speaker Differences on Intelligibility of Phonetically Balanced Word Lists," J. Speech Hear. Disord. **20**, 192–195.

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 Engin. Soc. **19**, 915–919.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking Clearly for the Hard of Hearing I: Intelligibility Differences Between Clear and Conversational Speech," J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech," J. Speech Hear. Res. **29**, 434–446.

Plomp, R. (**1977**). "Acoustical Aspects of Cocktail Parties," Acustica **38**, 186–191.

Revoile, S. G., and Pickett, J. M. (**1982**). "Speech Perception by the Severely Hearing-Impaired," in *Deafness and Communication*, edited by D. G. Sims, G. G. Walter, and R. L. Whitehead (Williams & Wilkins, Baltimore, MD), Chap. 2, pp. 25–39.

Studebaker, G. A. (**1985**). "Directivity of the Human Vocal Source in the Horizontal Plane," Ear Hear. **6**, 315–319.

Tolhurst, G. C. (1957). "The Effects of an Instruction to be Intelligible Upon a Speaker's Intelligibility, Sound Pressure Level, and Message Duration," U. S. Naval School of Aviation Medicine, Project NM 001 104 500, Rep. No. 58, Pensacola, FL.

Walden, B. E., Demorest, M. E., and Hepler, E. L. (1984). "Self-Report Approach to Assessing Benefit Derived from Amplification," J. Speech Hear. Res. 27, 49–56.

Weinhouse, I., and Miller, M. (1963). "Discrimination Scores for Two Lists of Phonetically Balanced Words," J. Aud. Res. 3, 9–14.

Williams, C. E., and Hecker, M. H. (1968). "Relation Between Intelligibility Scores for Four Test Methods and Three Types of Speech Distortion," J. Acoust. Soc. Am. 44, 1002–1006.

Witter, H. L., and Goldstein, D. P. (1971). "Quality Judgements of Hearing Aid Transduced Speech," J. Speech Hear. Res. 14, 312–322.