Recognition of Speech Produced in Noise

Andrea L. Pittman
Terry L. Wiley
University of Wisconsin–Madison

A two-part study examined recognition of speech produced in quiet and in noise by normal hearing adults. In Part I 5 women produced 50 sentences consisting of an ambiguous carrier phrase followed by a unique target word. These sentences were spoken in three environments: quiet, wide band noise (WBN), and meaningful multi-talker babble (MMB). The WBN and MMB competitors were presented through insert earphones at 80 dB SPL. For each talker, the mean vocal level, long-term average speech spectra, and mean word duration were calculated for the 50 target words produced in each speaking environment. Compared to quiet, the vocal levels produced in WBN and MMB increased an average of 14.5 dB. The increase in vocal level was characterized by increased spectral energy in the high frequencies. Word duration also increased an average of 77 ms in WBN and MMB relative to the quiet condition. In Part II, the sentences produced by one of the 5 talkers were presented to 30 adults in the presence of multi-talker babble under two conditions. Recognition was evaluated for each condition. In the first condition, the sentences produced in quiet and in noise were presented at equal signal-to-noise ratios (SNR_E). This served to remove the vocal level differences between the speech samples. In the second condition, the vocal level differences were preserved (SNR_P). For the SNR_E condition, recognition of the speech produced in WBN and MMB was on average 15% higher than that for the speech produced in quiet. For the SNR_P condition, recognition increased an average of 69% for these same speech samples relative to speech produced in quiet. In general, correlational analyses failed to show a direct relation between the acoustic properties measured in Part I and the recognition measures in Part II.

KEY WORDS: speech perception, speech acoustics, background noise, competing message

The presence of a competing acoustic signal during communication often interferes with the perception of speech. This is particularly true for persons with sensorineural hearing loss who are more susceptible to the deleterious effects of noise (Walden, Prosek, & Worthington; 1975). Studies have shown that word recognition in noise or competing message can differ for listeners with and without hearing loss as well as for listeners with different types and degrees of hearing loss (Beattie, 1989; Walden, Demorest, & Hepler, 1984; Wilson, Zizz, Shanks, & Causey, 1990). Although these differences represent the influence of noise on the perception of speech, they do not clarify the influence of noise on the *production* of speech or the subsequent *perception* of that speech. It is well-established that the acoustic properties of speech produced in noise are significantly different from those produced in quiet (Amazi & Garber, 1982; Junqua, 1993; Letowski, Frank, & Caravella, 1993; Summers, Pisoni, Bernaki, Pedlow, & Stokes, 1988; Tartter, Gomes, & Litwin, 1993; Webster & Klump, 1962). The relation between these properties and the perception of speech, however, remains unclear.

Speech Production in Noise

The acoustic characteristics of speech produced in noise typically are determined by recording speech stimuli from a single talker in a quiet environment and again in the presence of noise or competing message. Although many acoustic characteristics have been examined, increases in vocal level, changes in spectral composition, and increases in word duration have been reported most consistently. Summers et al. (1988) examined the acoustic properties of the digits “one” through “nine” produced by two men in quiet and in three levels of noise. Significant increases in vocal levels were observed for each talker in each noise level (an average of 4.5, 6.0, and 6.9 dB in 80, 90, and 100 dB SPL noise, respectively). The slope of a regression line, fitted to the data points of an amplitude-by-frequency analysis of the speech stimuli, was significantly steeper for speech produced in noise than in quiet. The steeper slope indicated a significant increase in amplitude for higher frequencies relative to the lower frequencies for speech spoken in noise. The mean word duration for each talker also increased significantly, from 461 ms in quiet to 524 ms in 80 dB SPL white noise, and increased further with each increase in noise level.

Tartter et al. (1993) examined the vocal levels of two women who produced the digits “zero” through “nine” in the presence of white noise. They reported an average increase of 1.0, 2.6, and 3.7 dB in 35, 60, and 80 dB SPL noise, respectively. As in the Summers et al. (1988) study, the slope of an amplitude-by-frequency analysis was calculated for each of the speech samples. Significant increases in high-frequency energy were reported for the 60 dB SPL noise condition relative to the lower noise level of 35 dB SPL. A significant increase in word duration (from an average of 343 ms in quiet to 530 ms in 80 dB SPL noise) also was reported.

Junqua (1993) examined the vocal levels of five men and five women who produced several subsets of speech materials (digits, monosyllabic words, bisyllabic words, and letters) in 85 dB SPL white noise. Average vocal level increases of 18.2 and 12.6 dB were reported for the men and women talkers, respectively. No significant shifts in spectral composition were found. An increase in phoneme duration also was reported, although no values were provided.

Letowski et al. (1993) evaluated the vocal levels of running speech produced by five men and five women in the presence of multi-talker babble, traffic noise, and wide band noise presented at 70 and 90 dB SPL. They reported significant increases in vocal level between quiet and both noise levels; however, no significant differences in vocal levels were found across the three noise types. An analysis of the amplitude of 20 frequencies taken from the long-term spectrum of speech indicated significantly larger increases in amplitude for frequencies ≥630 Hz. A measure of words per minute revealed no significant differences in speech rate between the quiet and the three noise conditions. Although no significant differences in the acoustic characteristics of running speech were found for the speech produced in each competitor, long-term spectral analyses may not have been sensitive to changes in individual words—particularly those important for perception. Materials produced in competitors that differ in spectral and semantic content (e.g., wideband noise vs. multi-talker babble) may not differ acoustically over the long term, although it is possible that differences may be observed for individual words. If so, the perception of speech produced in noise may be influenced by these acoustic changes.

The results of these studies suggest that (a) both men and women increase their vocal levels as a function of noise level; (b) the amplitude of mid- to high-frequency energy increases more than that for lower frequencies; (c) speech rates of men and women are similar in noise; and (d) vocal level, spectral composition, and word duration do not appear to be influenced by the spectral content of the noise when measured over the long term.

Perception of Speech Produced in Noise

The recognition of speech produced in quiet and in noise has been compared in only a few published studies and with conflicting results (Dreher & O’Neill, 1957; Junqua, 1993; Summers et al. 1988). Junqua (1993) reported significant decreases in the recognition of digits, monosyllabic words, bisyllabic words, and letters produced in 85 dB SPL white noise relative to the same stimuli produced in quiet. The Junqua report, however, did not provide details regarding how the stimuls levels were set for various conditions.

Dreher and O’Neill (1957), on the other hand, reported significantly higher recognition scores (an average of 27%) for spondees spoken in 70 dB SPL white noise than for the same spondees spoken in quiet. Summers et al. (1988) also reported significantly higher recognition scores (an average of 6%) for monosyllabic digits produced in 90 dB SPL white noise than for the same digits produced in quiet. It is important to note that stimulus levels in the Dreher and O’Neill (1957) study were not equalized before presentation, unlike the stimulus levels in the Summers et al. (1988) report. This may account, in part, for the difference in recognition scores across these two studies.

In summary, there are few data available regarding the recognition of speech spoken in noise even though a considerable portion of everyday communication takes place in the presence of a competitor. Further, the outcomes of these studies have been ignored in terms of clinical applications. If speech recognition is influenced by speech production, which is in turn influenced by a competing noise, this would be an important consideration to the face validity of speech-recognition measures used clinically. Although previous studies describe the differences in recognition between speech produced in quiet and in noise, it is not clear whether the magnitude of the differences warrants the use of environment-specific speech materials in a clinical setting. The most important consideration is whether recognition in noise is significantly underestimated using currently available speech materials. The studies reviewed above indicate that recognition of speech produced in quiet is generally poorer than for speech produced in noise. Unfortunately, those differences were evaluated only for a small number of stimuli not typically used in an audiological evaluation and were limited to a noise background not typical of everyday communication.

This study determined the recognition of speech produced in quiet and in two types of noise. In Part I, speech samples spoken in quiet and in two noise conditions were used to determine if the type of noise significantly influenced production. In Part II, the speech samples from one talker (exhibiting the average acoustic characteristics of speech spoken in noise) were selected and presented to a group of listeners under two listening conditions. In the first condition, the vocal level differences between the samples were removed by presenting each at a signal-to-noise ratio (SNR) that equated the overall presentation level. This determined the degree to which recognition may be underestimated in clinically derived measures. In the second condition, the same speech samples were presented with the vocal level differences preserved. Recognition in this condition may more accurately reflect the perception of speech in noisy environments. The recognition scores from these two conditions were then analyzed with respect to the acoustic characteristics measured in Part I to determine the influence of these characteristics on perception.

Part I: Development of Speech Material

Method

Participants

Five women between the ages of 19 and 28 years participated as talkers. All had hearing thresholds ≤20 dB HL at audiometric frequencies 0.25, 0.5, 1, 2, 4, and 8 kHz in each ear and normal middle-ear function as determined by tympanometry (Roup, Wiley, Safady, & Stoppenbach, 1998). All five women were native speakers of American English with no noticeable regional dialects.

Materials

Fifty low-predictability (LP) sentences from the Speech in Noise (SPIN) Test were used (Kalikow, Stevens, & Elliot, 1977). These sentences consisted of a unique, although ambiguous, carrier phrase (e.g., "He would not think about the...") followed by a unique target word (*crack*). The structure of the sentences provided no contextual information with which to predict the final target word during the recognition task (Part II). This required the listener to rely on the acoustic information rather than the semantic content of the sentence. Five practice sentences began each list to allow the talker to adjust to each speaking environment. Three randomizations of the 50 sentences were constructed, one for each speaking environment.

Procedure

Each talker was seated in a sound-treated room with a head-worn microphone (Shure, SM10A) placed 1 inch from the lips, out of the breath stream. Each talker read the 50 LP sentences first in quiet, then in the presence of wide band noise (WBN), and again in the presence of meaningful multi-talker babble (MMB).[1] The competitors were delivered binaurally at 80 dB SPL through insert earphones (Etymotic, ER-3A). The WBN was generated by an audiometer (GSI, 16), and the MMB was routed through the audiometer from a cassette tape player (Nakamichi, CR-2A). The presentation level and spectra of each competitor were confirmed for both insert earphones with acoustic measures in a 2-cc coupler. The earphones were removed for the quiet speaking environment. The overall noise level in the sound-treated room in the quiet environment was 16 dB SPL.

To encourage each talker to speak in a manner that would maximize recognition, an assistant wearing headphones was seated outside the window of the sound booth and instructed to write the final word of each sentence. Each talker was told that the listener was unable to see the features of her face and was instructed to speak clearly, to read the sentences in order, and to wait for the listener to look up from the response sheet before proceeding. The talker was unable to see the written responses. Unknown to the talker, all sentences were

---

[1] The MMB competitor contained independent conversations of three men and three women recorded separately and then mixed to produce a multi-talker competitor. Semantic information was preserved in that portions of each conversation could be selectively followed. It was produced by G. Donald Causey in 1979 at the Biocommunications Laboratory at the University of Maryland.

digitally recorded (Tascam, DA-P1) at a 44.1 kHz sampling rate for later analyses. It was felt that the talker might artificially alter her vocal effort if she were aware of the recording. Each talker was informed of the recording at the completion of the session, and each agreed to have her speech samples included in the study.

Acoustic Analyses

The sentences produced by each talker in each speaking environment were low-pass filtered at 10 kHz and digitized using a 16-bit A/D converter. The target word within each of the 50 sentences was extracted, concatenated, and saved in 15 separate speech samples (5 talkers × 3 speaking environments). The boundaries of each target word were visually determined using a digital audio editor (Syntrillium Software Corp., CoolEdit). Using digital signal processing techniques, long-term-average speech spectra (LTASS) were measured in $^{1}/_{3}$-octave bands for each 50-word speech sample. A 1000-Hz reference tone of a known SPL and voltage was pre-recorded on each digital audiotape and used to calculate the level of each $^{1}/_{3}$-octave band as well as the overall vocal level (in dB SPL). To describe the spectral composition of each speech sample with a single number, the slope of a regression line (in dB SPL/kHz) was fitted to 14 of 15 data points representing the amplitude of each $^{1}/_{3}$-octave band frequency. The 15th frequency band was not included because of the limited bandwidth of the earphones used in Part II of this study. The average duration (in ms) was calculated for the 50 words measured in each speech sample.

Results and Discussion

The LTASS for each talker and speaking environment, as well as the spectra calculated for all five talkers (lower right panel), are shown in Figure 1. The solid, dashed, and dotted lines represent speech produced in quiet, WBN, and MMB, respectively. The bottom right panel shows the average spectra of each speaking environment for all five talkers. In general, the spectra for speech produced in both WBN and MMB exhibited higher overall levels than the spectra for speech produced in quiet. Talkers 4 and 5 exhibited the smallest differences in level between the speaking environments, whereas Talkers 2 and 3 exhibited the largest differences. Small differences between the spectra for the WBN and MMB speaking environments are apparent for Talkers 2, 3, and 5, but not for Talkers 1 and 4.

Vocal Levels

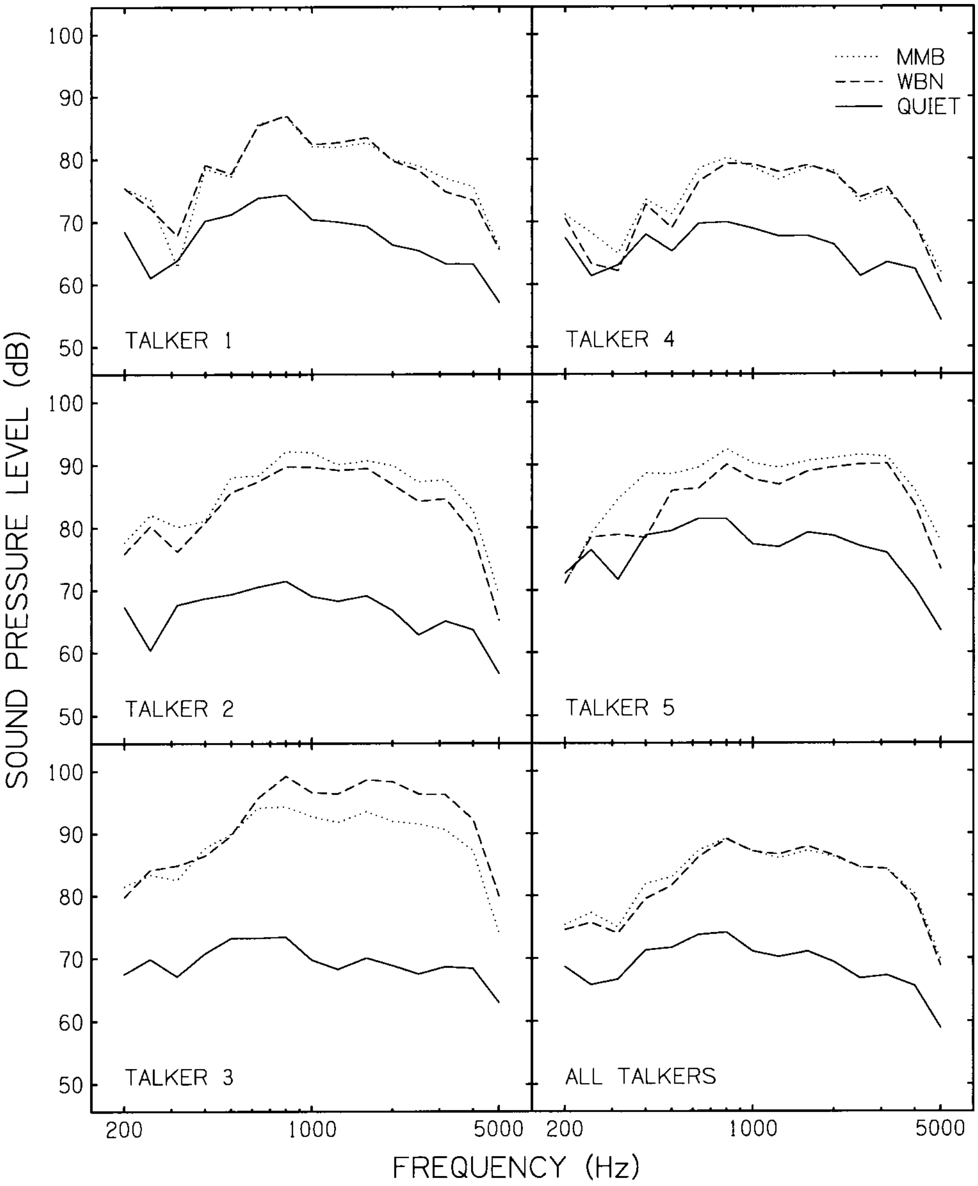
The vocal levels for each talker and speaking condition are shown in the top panel of Figure 2. Also shown are the mean vocal levels (and one standard deviation) for each speaking environment. Relative to quiet, vocal levels increased an average of 14.5 dB in noise. A one-way ANOVA with repeated measures and planned orthogonal contrasts confirmed that the vocal levels of the speech spoken in noise were significantly higher than those spoken in quiet [$F(2, 8) = 17.7$, $p = 0.001$, $\omega^2 = 0.47$; Quiet vs. WBN: $t^2_{Dunn}(3, 8) = 26.4$, $p < 0.001$; Quiet vs. MMB: $t^2_{Dunn}(3, 8) = 26.4$, $p < 0.001$]. The vocal levels of speech produced in WBN and MMB did not differ significantly. These results are consistent with those of Junqua (1993), who reported an average increase of 15 dB for speech produced in 85 dB SPL white noise. These results are somewhat higher than the 4.6 dB increase in 80 dB SPL white noise reported by Summers et al. (1988) and the 3.7 dB increase reported by Tartter et al. (1993). The reason for these differences in vocal level is unclear but may reflect individual differences among talkers. It is possible that the two talkers in each study exhibited small increases in vocal level similar to Talkers 4 and 5 in the present study (9 dB in 80 dB SPL WBN).

The absolute vocal levels found in the present study also are somewhat higher than those reported in previous studies. This is likely due to differences in the distance of the talker from the recording microphone. For example, the microphone in the present study was positioned 1 inch from the talker's mouth, whereas in Letowski et al. (1993) and Summers et al. (1988), the microphones were 12 and 4 inches from the talkers, respectively. Using the inverse square law to estimate vocal levels at a microphone distance of 1 inch, the levels in quiet for the Letowski and Summers studies are equivalent to 84 and 71 dB SPL, respectively, which are somewhat similar to the average vocal level of 82 dB SPL in the present study.

Spectral Composition

The slope values for each speaking condition are shown in the middle panel of Figure 2 as a function of talker. Also shown are the mean slope values (and one standard deviation) for each speaking environment. The positive slope values for the two noise environments indicate an increase in high-frequency energy for speech produced in noise. For example, the mean amplitude at 2.5 kHz for all talkers shown in Figure 1 increased an average of 18 dB compared to an average increase of only 7 dB at 0.2 kHz. A one-way ANOVA with repeated measures revealed a significant difference between the slope values for the speech samples produced in each speaking environment [$F(2, 8) = 19.338$, $p < 0.001$, $\omega^2 = 0.48$]. Planned orthogonal contrasts revealed a significant difference between the values for speech produced in quiet and in both noise environments [Quiet vs. WBN:

Figure 1. Long-term average spectra of speech spoken in quiet, in wide band noise (WBN), and in meaningful multi-talker babble (MMB) for each of the five talkers, with the combined spectra of all five talkers in the lower right panel.

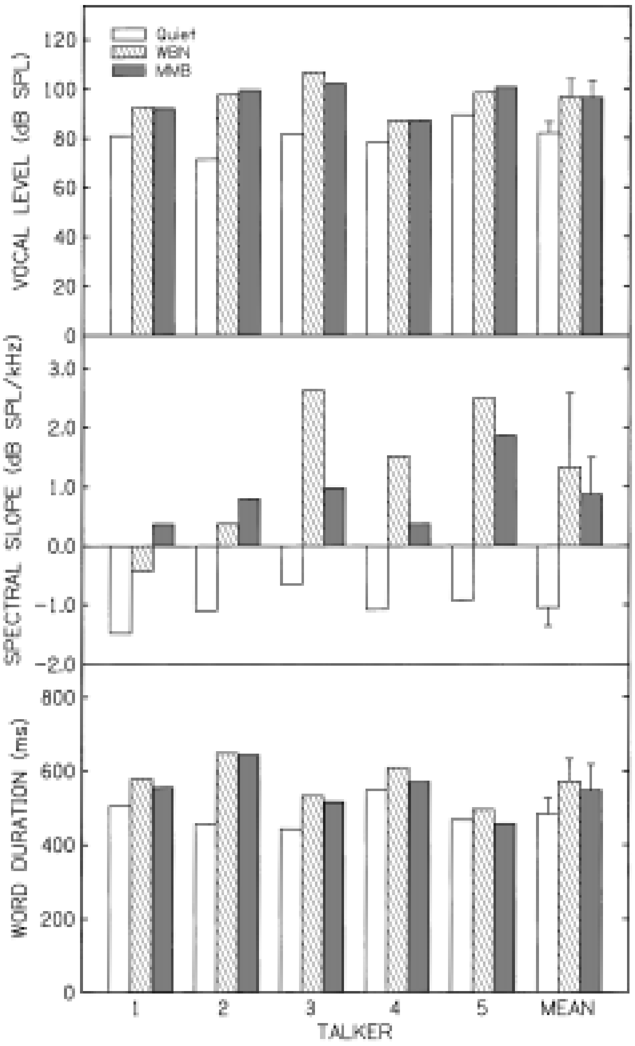


$t^2_{Dunn}(3, 8) = 33.375$, $p < 0.001$; Quiet vs. MMB: $t^2_{Dunn}(3, 8) = 23.838$, $p < 0.001$]. However, no difference was found between the slope values for the speech produced in WBN and in MMB [$t^2_{Dunn}(3, 8) = 0.801$, $p = 0.528$]. These results are consistent with Summers et al. (1988) and Tartter et al. (1993), who also reported significant increases in high-frequency energy for speech spoken in noise. The lack of differences between WBN and MMB in vocal level and slope are consistent with Letowski et al. (1993), who suggested that the overall level of a competitor, rather than its spectral content, determines changes in the acoustic characteristics of speech.

Target Word Duration

The average target word durations for each speaking environment are shown by talker in the bottom panel of Figure 2. Also shown are the mean word durations (and one standard deviation) for each speaking environment. Relative to speech spoken in quiet, target word duration increased an average of 88 and 65 ms in WBN and in MMB, respectively. A one-way ANOVA with repeated measures revealed significant differences in word duration for the speech spoken in quiet and in noise [$F(2, 8) = 6.7$, $p = 0.021$, $\omega^2 = 0.20$].

Figure 2. Mean vocal levels in dB SPL (top panel), spectral slope in dB SPL/kHz (middle panel), and word duration in ms (bottom panel) for the 50 target words produced in quiet, in wide-band noise (WBN), and in meaningful multi-talker babble (MMB) for each talker. Group means and ±1 standard deviation are shown for each speaking environment at the right.



Planned orthogonal contrasts revealed significantly longer target word durations for the speech spoken in noise relative to speech in quiet [Quiet vs. WBN: t^2_{Dunn} (3, 8) = 12.2, $p = 0.002$; Quiet vs. MMB: t^2_{Dunn}(3, 8) = 6.7, $p = 0.014$], although no difference in word duration was observed between WBN and MMB [t^2_{Dunn}(3, 8) = 0.8, $p = 0.527$]. These results are consistent with Summers et al. (1988), who reported an average increase of 60 ms in word duration under similar conditions, but are somewhat shorter than the 185-ms increase reported by Tartter et al. (1993).

To determine whether the insert earphones caused an occlusion effect that was not present in the quiet speaking environment, Talker 1 was asked to return for further testing. She read 10 of the original 50 sentences under three conditions: (1) wearing the insert earphones with no noise input, (2) wearing the insert earphones with 80 dB SPL of WBN, and (3) in quiet without the insert earphones. The sentences were analyzed as described above. Although the acoustic characteristics of the speech spoken in 80 dB SPL noise were similar to those measured previously for this talker, no significant differences were found between the speech spoken in the two quiet environments. This suggests that the insert earphones did not create an occlusion effect that might have affected speech production.

In summary, the acoustic characteristics of the speech materials in the present study appear to be consistent with those of previous studies. Relative to speech spoken in quiet, speech spoken in noise demonstrated significant increases in vocal level, spectral slope, and word duration. The speech samples of Talker 1 were used for a recognition task described in Part II because the acoustic characteristics of her speech were closest to the average of all five talkers. In addition, the sentences produced by this talker contained no errors, whereas the other four talkers occasionally misread one or two nontarget words.

Part II: Recognition

Part II of this study compared the recognition of speech produced in quiet and in noise. Like previous studies of this kind, the speech samples produced in quiet and in noise were presented at SNRs that equated overall vocal levels (SNR_E). Unlike previous studies, the speech samples also were presented at SNRs that preserved these vocal-level differences (SNR_P). In this way, the influence of the spectral and temporal changes in the speech stimuli could be evaluated independent of, and then in combination with, the additional contribution of increased vocal level. Based on the work of Summers et al. (1988) and Dreher and O'Neill (1957), one would expect higher recognition scores for the speech produced in WBN and MMB than for the speech produced in quiet. In addition, one would expect no difference in recognition between the speech produced in WBN and MMB, because no significant acoustic differences were observed.

Method

Participants

Twenty-seven women and 3 men between the ages of 18 and 30 years served as listeners. Each participant had hearing thresholds in the test ear ≤10 dB HL at audiometric frequencies 0.25, 0.5, 1, 2, and 4 kHz and ≤15 dB HL at 8 kHz. Hearing levels in the nontest ear were ≤20 dB HL at audiometric frequencies 0.25 through 8 kHz. The ear with the lowest thresholds was chosen as the test ear. In cases of equal thresholds in both ears, the test ear was alternated across listeners. All listeners exhibited normal middle ear function bilaterally based on tympanometry (Roup et al., 1998).

Speech Materials

The 50 sentences produced in quiet and in the two noise conditions by Talker 1 were digitally extracted from the original recording to remove extraneous utterances. The sentences within each condition were randomized and recorded onto a compact disk at a sampling rate of 22.05 kHz. A 4-s gap was inserted between each sentence to allow time for a written response. Two 1-kHz calibration tones also were recorded. The first was equal in average RMS level to the sentences produced in quiet, and the second was equal to the average RMS level of the speech produced in WBN and MMB. Separate calibration tones were not necessary for the WBN and MMB sentences, because the overall level of the two samples differed by less than 1 dB. No attempt was made to equalize the RMS levels of the target words within each sentence. This enabled preservation of the acoustic characteristics unique to each speaking environment, including variations in vocal level.

Procedure

Each 50-sentence speech sample was presented with the MMB competitor at 0, –5, and –10 dB SNRs. The level of the speech remained constant, and the level of the competitor changed according to the SNR. There were two listening conditions. In the first condition (SNR_E), the levels of the three 50-sentence speech samples were equated by presenting each at the same SNR. The recognition scores would therefore reflect the influence of all acoustic differences between the samples, except vocal level. In the second condition (SNR_P), the level differences between the three 50-sentence speech samples were preserved so that recognition scores would reflect the influence of all the acoustic differences, including vocal level. This was accomplished by setting the noise level equal to that of the speech produced in quiet and then presenting the speech produced in WBN and MMB 11 dB higher, which is equivalent to the increase in vocal level for this talker. Presentation of the speech material at 0 and –5 dB SNRs was discontinued for the SNR_P condition after the results of the first five listeners revealed ceiling effects.

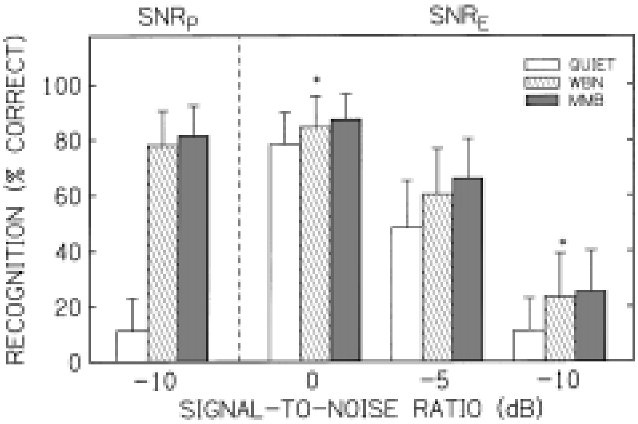
Each 50-sentence sample was presented monaurally through a TDH-50 earphone at 60 dB SL relative to the pure-tone threshold at 1 kHz. Each participant was instructed to write the final word of each of the sentences. Testing was conducted in 2 one-hour sessions. The first five listeners responded to a total of 18 lists of 50 sentences each (3 speech samples × 3 SNRs × 2 listening conditions). The number of lists was reduced to 12 when the 0 and –5 dB SNRs were discontinued from the SNR_P condition. To reduce learning effects, the speech samples, SNRs, and listening conditions were randomized; and each participant was familiarized with the sentences at a +30 dB SNR before testing. Recognition scores were examined to confirm that performance did not improve significantly between the first and last presentation of the sentences.

Results and Discussion

Part II was conducted to determine if speech produced in quiet and in the two noise environments differed in terms of recognition. Mean scores for the SNR_E and SNR_P conditions are shown in Figure 3 as a function of SNR and listening condition. The data from one of the 30 listeners were corrupted for the 0 and –10 dB SNR_E conditions and could not be used (indicated with asterisks). Before statistical analyses, scores were transformed into rationalized arcsine units (RAU) so that the variances would be homogenous across the range of scores (Studebaker, 1985).

For the SNR_E condition, recognition of speech produced in noise was an average of 15% higher (at –5 dB SNR) than that of the speech produced in quiet. A one-way

Figure 3. Mean recognition in percent correct (and ±1 standard deviation) for speech produced in quiet, in wide-band noise (WBN), and in meaningful multi-talker babble (MMB) presented under two listening conditions: SNR_P, for which vocal level differences were preserved, and SNR_E, for which vocal levels differences were removed. Asterisks indicate those conditions with only 29 listeners; all other values were obtained for 30 listeners.



ANOVA with repeated measures revealed significant differences in recognition for the speech samples at each SNR [0 dB SNR: $F(2, 56) = 7.4$, $p = 0.001$, $\omega^2 = 0.07$; –5 dB SNR: $F(2, 58) = 27.5$, $p < 0.001$, $\omega^2 = 0.11$; –10 dB SNR: $F(2, 56) = 21.0$, $p < 0.001$, $\omega^2 = 0.13$]. Planned orthogonal contrasts, listed in Table 1, revealed significantly better recognition for both samples of speech spoken in noise compared to speech spoken in quiet ($p < 0.01$). No significant differences were found for the speech spoken in WBN and MMB at 0 and –10 dB SNR. However, recognition of MMB was significantly higher than that of WBN at –5 dB SNR ($p = 0.05$).

In the SNR_P condition, differences in recognition performance were greatest at –10 dB SNR; scores for speech spoken in noise were an average of 69% higher than those for speech spoken in quiet. A one-way ANOVA with repeated measures revealed significant differences among the three speech samples [$F(1.7, 48.2) = 529.7$, $p < 0.001$, $\omega^2 = 0.78$; degrees of freedom were adjusted to compensate for a lack of sphericity[2]]. Planned orthogonal contrasts revealed significantly higher recognition scores for speech spoken in WBN and MMB than in speech spoken in quiet [Quiet vs. WBN: t^2_{Dunn} (3, 58) = 753.5, $p < 0.001$; Quiet vs. MMB: t^2_{Dunn} (3, 58) = 833.3, $p < 0.001$], although no differences in recognition scores were found between WBN and MMB [t^2_{Dunn} (3, 58) = 2.0, $p = 0.123$].

Because recognition scores were higher for speech spoken in WBN and MMB than for speech spoken in quiet, additional analyses were performed to determine if the acoustic characteristics measured in Part I influenced performance. Specifically, the percentage of listeners able to correctly identify each target word was calculated for the quiet, WBN, and MMB speech samples. The differences in percent between the two noise conditions and the quiet condition were calculated (WBN-quiet and MMB-quiet, respectively). These values quantified the magnitude of improvement between the target words spoken in noise and in quiet. In the same way, difference values also were calculated for the peak RMS level (dB SPL), spectral slope (dB SPL/kHz), and duration (ms) of each target word. This was done for the 0, –5, and –10 dB SNR_E conditions. Recall that the SNRs of the speech samples in the SNR_E condition were equated so that the large vocal level differences would be removed. However, because no attempt was made to equalize the RMS levels of each target word for the recognition task, some level differences between the words remained. Correlation coefficients were computed to determine the relation between the changes in performance and the changes in the three acoustic characteristics.

Pearson's product-moment correlations are listed in Table 2. In separate analyses, the increased vocal level and spectral slope observed for the speech produced in noise (WBN and MMB) correlated significantly with increases in recognition ($p < 0.01$). However, the effects were small for all but the peak RMS level for the WBN speech sample at –5 dB SNR. Interestingly, this highest correlation conflicts with the results of the recognition task described earlier, for which significantly better performance was observed for the MMB speech sample than for the WBN speech sample at –5 dB SNR. Overall, these results suggest that increases in vocal level and spectral composition do not completely account for the observed increases in recognition.

Table 1. Planned orthogonal contrasts (t^2_{Dunn}) by SNR for the speech produced in quiet, in wide-band noise (WBN), and in meaningful multi-talker babble (MMB) presented in the SNR_E condition. Asterisks indicate significant contrasts ($p \leq 0.05$).

SNR	Contrast	*df*	*t*	*p*
0	Quiet vs. WBN	3, 56	6.5	0.001*
	Quiet vs. MMB	3, 56	14.4	<0.001*
	WBN vs. MMB	3, 56	1.5	0.214
–5	Quiet vs. WBN	3, 58	24.2	<0.001*
	Quiet vs. MMB	3, 58	52.8	<0.001*
	WBN vs. MMB	3, 58	52.7	0.05*
–10	Quiet vs. WBN	3, 56	26.8	<0.001*
	Quiet vs. MMB	3, 56	35.5	<0.001*
	WBN vs. MMB	3, 56	0.5	0.613

Table 2. Pearson's product-moment correlation coefficients (*r*) relating the differences in recognition and acoustic characteristics for each target word produced in quiet and in wideband noise (WBN-quiet) and meaningful multi-talker babble (MMB-quiet) presented at 0, –5, and –10 dB SNR in the SNR_E condition. Significant correlations ($p < 0.01$) are indicated by asterisks.

	Speaking condition	
	WBN-quiet	MMB-quiet
Peak RMS levels		
0 dB SNR	0.57*	0.16
–5 dB SNR	0.74*	0.40*
–10 dB SNR	0.61*	0.33*
Spectral Slope		
0 dB SNR	0.45*	0.26
–5 dB SNR	0.51*	0.32*
–10 dB SNR	0.41*	0.24
Word Duration		
0 dB SNR	0.11	0.32*
–5 dB SNR	–0.02	0.01
–10 dB SNR	–0.03	0.11

---

[2] A lack of sphericity indicates that the variances of all possible comparisons (quiet, WBN, MMB scores) were not equal, which may inflate the Type I error rate. An adjustment to the degrees of freedom using the Greenhouse-Geisser method was made to maintain a rejection rate of 5%.

General Discussion

In this study, speech produced in quiet and in two noise conditions (Part I) was presented to listeners in a recognition paradigm using two SNR conditions (Part II). The acoustic analyses of speech produced in the two noise types revealed significant increases in vocal level, spectral composition, and word duration as compared to speech produced in quiet. Interestingly, the acoustic analyses revealed no differences between the speech produced in the two noise types (WBN and MMB) despite the spectral and semantic differences of these competitors. In terms of recognition, scores were an average of 69% higher for the speech produced in noise when the vocal level differences between the speech samples were preserved (SNR_P) and an average of 15% higher when the vocal level differences were removed (SNR_E). No significant differences in recognition were found for the speech produced in WBN and in MMB except at –5 dB SNR_E, where a significant increase of 6% was observed for speech produced in MMB. In general, these results suggest that the recognition of speech produced in noise was significantly better than that for speech produced in quiet and that the spectral and semantic content of the WBN and MMB competitors did not appear to differentially influence the production of speech or the subsequent perception of that speech.

The results of this study are consistent with those of Summers et al. (1988). In that study, digits produced in broadband noise and in quiet were presented in a paradigm similar to the –10 dB SNR_E condition in the present study for an average increase in recognition of 6%. Dreher and O'Neill (1957) reported increases of 27% for spondees produced in white noise and presented at +4 dB SNR. Junqua (1993), on the other hand, reported significant decreases in the perception of digits and monosyllabic and bisyllabic words produced in noise for some listeners. Unfortunately, these results cannot be compared directly because of insufficient information provided by Junqua regarding methodology and statistical significance. Recall, however, that Junqua reported no significant differences in the spectral composition of his speech materials, which may explain the discrepancy between the results of that study and those of the present study.

In general, higher recognition scores were observed for speech produced in noise relative to speech produced in quiet in the SNR_P condition. The increased performance was likely due to the improved signal-to-noise ratio provided by the large increases in vocal level. Although the effects were smaller, significant increases in recognition also were observed when the overall vocal level of the speech samples was equated (SNR_E condition). The residual difference in performance for this condition suggests that additional variations in the acoustic characteristics of each target word (other than overall vocal level) may have contributed to the observed differences in performance. Further analyses of each target word revealed significant correlations between performance and two acoustic characteristics (vocal level and spectral composition). However, the effects were small and inconsistent. Overall, these results suggest that there is not a simple relation between vocal level or spectral composition of individual words and recognition. Rather, recognition is more likely the result of complex interactions between these and other acoustic characteristics that were not examined.

Implications

The results of this study have implications for at least two areas of clinical audiology. First, Wiley and Page (1997) argued that, among other things, speech perception tasks should provide results that can be applied to rehabilitation efforts, such as amplification, and the prediction of communication difficulties in everyday listening situations. The results of Part I suggest that the acoustic characteristics of speech spoken in noise are significantly different from those for speech spoken in quiet. These characteristics, therefore, should be considered when using hearing aid prescriptive procedures. For example, many hearing aid prescriptive methods use the long-term spectrum of speech produced in quiet as a reference for all incoming signals (Byrne & Dillon, 1986; Cox & Moore, 1988; Schwartz, Lyregaard, & Lundh, 1988). Hearing aid manufacturers and others recommend a decrease in low-frequency gain and an increase in high-frequency gain for the best perception of speech in noisy environments (Martin, 1996). Although this practice may reduce the effects of upward spread of masking, the results of this study suggest that smaller adjustments may be necessary. Talkers will naturally speak louder in noisy conditions and therefore reduce low-frequency and increase high-frequency energy. If the parameters of a hearing aid are set without this consideration, the acoustic properties of speech may be overcorrected and, in some cases, perception may actually be degraded (e.g., the hearing aid may be forced to operate in saturation). It is important to remember, however, that the talkers in the present study were specifically instructed to speak clearly to a listener. Whether this is fully representative of speech in a typical noise environment is unknown.

The results of Part II suggest that speech-recognition tasks used clinically are of limited value for predicting communication difficulties in everyday situations that involve noise or competing speech because these tasks use speech samples recorded in quiet. The absence of a relation between recognition and the most robust acoustic differences between these speech samples suggests that it may not be possible to predict accurately speech recognition in noise through simple modifications of speech produced in quiet (e.g., increasing the SNR or shaping the frequency response). Rather, these results suggest the need to develop speech samples for recognition tests that incorporate the acoustic characteristics of actual speaking environments, including those with background noise. In this way, the effects of hearing loss on speech recognition can be determined more accurately by closely imitating common communication environments under controlled conditions.

Acknowledgments

We would like to acknowledge Ray Kent, Dolores Vetter, Keith Kluender, and Cynthia Fowler for their insightful contributions to this project and Patricia Stelmachowicz for her many helpful comments on earlier versions of this manuscript. This work was funded in part by a grant from the Ventry and Friedrich Memorial Funds of the American Speech-Language-Hearing Foundation.

References

Amazi, D. K., & Garber, S. R. (1982). The Lombard Sign as a function of age and task. *Journal of Speech and Hearing Research, 25,* 581–585.

Avaaz Innovations. (1995). *Computerized Speech Research Environment (CRSE) v4.5 User's Guide.* London, Ontario, Canada: Author.

Beattie, R. (1989). Word recognition functions for the CID W-22 Test in multi-talker noise for normally hearing and hearing-impaired subjects. *Journal of Speech and Hearing Disorders, 54,* 20–32.

Byrne, D., & Dillon, H. (1986). The National Acoustics Laboratories' (NAL) new procedure for selection the gain and frequency response of a hearing aid. *Ear and Hearing, 7,* 257–265.

Cox, R. M., & Moore, J. N. (1988). Composite speech spectrum for hearing aid gain prescriptions. *Journal of Speech and Hearing Research, 31,* 102–107.

Dreher, J. J., & O'Neill, J. J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *Journal of the Acoustical Society of America, 29,* 1320–1323.

Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America, 93,* 510–524.

Kalikow, D. N., Stevens, K. N., & Elliot, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337–1351.

Kryter, K. D. (1962). Methods for calculation and use of the articulation index. *Journal of the Acoustical Society of America, 34,* 1689–1697.

Martin, R. L. (1996). How to shape amplified sound to help patients hear in background noise. *The Hearing Journal, 49,* 49–50.

Letowski, T., Frank, T., & Caravella, J. (1993). Acoustical properties of speech produced in noise presented through supra-aural earphones. *Ear and Hearing, 14,* 332–338.

Roup, C. M., Wiley, T. L., Safady, S. H., & Stoppenbach, D. T. (1998). Tympanometric screening norms for adults. *American Journal of Audiology, 7,* 55–60.

Schwartz, D. M., Lyregaard, P. E., & Lundh, P. (1988, February). Hearing aid selection for severe-to-profound hearing loss. *The Hearing Journal,* 13–17.

Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech and Hearing Research, 28,* 455–462.

Summers, W. V., Pisoni , D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustical and perceptual analyses. *Journal of the Acoustical Society of America, 84,* 917–928.

Tartter, V. C., Gomes, H., & Litwin, E. (1993). Some acoustic effects of listening to noise on speech production. *Journal of the Acoustical Society of America, 94,* 2437–2440.

Walden, B. E., Prosek, R. A., & Worthington, D. W. (1975). *The prevalence of hearing loss within selected U.S. Army branches* (Interagency No. IAO 4745, August, 31, 1-95). Washington, DC: U.S. Army Medical Research and Development Command.

Webster, J. C., & Klumpp, R. G. (1962). Effects of ambient noise and nearby talkers on a face-to-face communication task. *Journal of the Acoustical Society of America, 34,* 936–912.

Wiley, T. L., & Page, A. L. (1997). Summary: Current and future perspectives on speech perception tests. In L. L. Mendel & J. L. Danhauer (Eds.), *Audiologic evaluation and management and speech perception assessment* (pp. 201–210) San Diego, CA: Singular Publishing

Wilson, R. H., Zizz, C. A., Shanks, J. E., & Causey, G. D. (1990). Normative data in quiet, broadband noise, and competing message for the Northwest University Auditory Test No. 6 by a female speaker. *Journal of Speech and Hearing Disorders, 55,* 771–778.

Received May 9, 2000

Accepted March 1, 2001

DOI: 10.1044/1092-4388(2001/038)

Contact author: Andrea Pittman, PhD, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE 68131. Email: pittmana@boystown.org