Published in final edited form as:
J Acoust Soc Am. 1988 September ; 84(3): 917–928.

Effects of noise on speech production: Acoustic and perceptual analyses

W. Van Summers, **David B. Pisoni**, **Robert H. Bernacki**, **Robert I. Pedlow**, and **Michael A. Stokes**
Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405

Abstract

Acoustical analyses were carried out on a set of utterances produced by two male speakers talking in quiet and in 80, 90, and 100 dB SPL of masking noise. In addition to replicating previous studies demonstrating increases in amplitude, duration, and vocal pitch while talking in noise, these analyses also found reliable differences in the formant frequencies and short-term spectra of vowels. Perceptual experiments were also conducted to assess the intelligibility of utterances produced in quiet and in noise when they were presented at equal S/N ratios for identification. In each experiment, utterances originally produced in noise were found to be more intelligible than utterances produced in the quiet. The results of the acoustic analyses showed clear and consistent differences in the acoustic–phonetic characteristics of speech produced in quiet versus noisy environments. Moreover, these acoustic differences produced reliable effects on intelligibility. The findings are discussed in terms of: (1) the nature of the acoustic changes that take place when speakers produce speech under adverse conditions such as noise, psychological stress, or high cognitive load; (2) the role of training and feedback in controlling and modifying a talker's speech to improve performance of current speech recognizers; and (3) the development of robust algorithms for recognition of speech in noise.

INTRODUCTION

It has been known for many years that a speaker will increase his/her vocal effort in the presence of a loud background noise. Informal observations confirm that people talk much louder in a noisy environment such as a subway, airplane, or cocktail party than in a quiet environment such as a library or doctor's office. This effect, known as the Lombard reflex, was first described by Etienne Lombard in 1911 and has attracted a moderate degree of attention by researchers over the years. The observation that speakers increase their vocal effort in the presence of noise in the environment suggests that speakers monitor their vocal output rather carefully when speaking. Apparently, speakers attempt to maintain a constant level of intelligibility in the face of degradation of the message by the environmental noise source and the corresponding decrease in auditory sidetone at their ears. Lane and his colleagues (Lane *et al*., 1970; Lane and Tranel, 1971) have summarized much of the early literature on the Lombard effect and have tried to account for a wide range of findings reported in the literature. The interested reader is encouraged to read these reports for further background and interpretation.

Despite the extensive literature on the Lombard effect over the last 30 years, little, if any, data have been published reporting details of the acoustic–phonetic changes that take place when a speaker modifies his vocal output while speaking in the presence of noise. A number of studies have reported reliable changes in the prosodic characteristics of speech produced in noise. However, very few studies have examined changes in the spectral properties of speech produced in masking noise.

In an earlier study, Hanley and Steer (1949) found that in the presence of masking noise, speakers reduce their rate of speaking and increase the duration and intensity of their utterances. In another study, Draegert (1951) examined the relations between a large number of physical measures of voice quality and speech intelligibility in high levels of noise and found a similar pattern of results. His interest was focused primarily on factors that correlated with measures of speech intelligibility rather than on a description of the acoustic–phonetic changes that take place in the speaker's speech. In addition to changes in duration and intensity, Draegert reported increases in vocal pitch and changes in voice quality due to a shift in the harmonic structure. The change in harmonic structure was shown by a difference in intensity between the low- and high-frequency components. To obtain these measures, the speech was bandpass filtered to obtain estimates of the locations of the major concentrations of energy in the spectrum. Unfortunately, no measurements of the size of the effects were reported in this article.

In another study on the intelligibility of speech produced in noise, Dreher and O'Neill (1957) reported that, when presented at a constant speech-to-noise ratio, speech produced by a speaker with noise in his ears is more intelligible than speech produced in quiet. This result was observed for both isolated words and sentences. In each case, for the noise condition, a broadband random noise source was presented over the speaker's headphones during production.

Related findings have been reported by Ladefoged (1967, pp. 163–165) in an informal study designed to examine how eliminating auditory feedback affects a speaker's speech. Auditory feedback was eliminated by presenting a loud masking noise over headphones at an intensity level that prevented the subject from hearing his/her voice even via bone conduction. Subjects read a prepared passage and also engaged in spontaneous conversation. According to Ladefoged, although subjects' speech remained intelligible, it became "very disorganized" by removal of auditory feedback through the presentation of masking noise. Of special interest to us was the observation by Ladefoged that the length and quality of many of the vowel sounds were affected quite considerably by the masking noise. Some sounds became more nasalized, others lost appropriate nasalization. Pitch increased and there appeared to be much less variability in the range of pitch. Ladefoged also noticed a striking alteration in voice quality brought about by the tightening of the muscles of the pharynx. These findings were summarized informally by Ladefoged in his book without reporting any quantitative data. To our knowledge, these results have never been published. Nonetheless, they are suggestive of a number of important changes that may take place when speakers are required to speak under conditions of high masking noise.

The Dreher and O'Neill (1957) results suggest that masking noise which does not eliminate auditory feedback to the subject may have a positive influence on speech intelligibility. On the other hand, the Ladefoged (1967) findings suggest that this may not be the case when environmental noise is so loud that all auditory feedback is eliminated.

The present investigation is concerned with the effects of masking noise on speech production. Our interest in this problem was stimulated, in part, by recent efforts of the Air Force to place speech recognition devices in noisy environments such as the cockpits of military aircraft. Although it is obvious that background noise poses a serious problem for the operation of any speech recognizer, the underlying reasons for this problem are not readily apparent at first glance. While extensive research efforts are currently being devoted to improving processing algorithms for speech recognition in noise, particularly algorithms for isolated speaker-dependent speech recognition, a great deal less interest has been devoted to examining the acoustic–phonetic changes that take place in the speech produced by talkers in high ambient noise environments. If it is the case, as suggested by the

published literature, that speakers show reliable and systematic changes in their speech as the noise level at their ears increases, then it would be appropriate to examine these differences in some detail and to eventually incorporate an understanding of these factors into current and future algorithm development. Thus the problem of improving the performance of speech recognizers may not only be related to developing new methods of extracting the speech signal from the noise but may also require consideration of how speakers change their speech in noisy or adverse environments.

As noted earlier, a search through the literature on speech communication and acoustic–phonetics published over the last 40 years revealed a number of studies on the effects of noise on speech production and speech intelligibility. While changes in duration, intensity, and vocal pitch have been reported, and while changes in voice quality have been observed by a number of investigators, little is currently known about the changes that take place in the distribution of spectral energy over time such as modifications in the patterns of vowel formant frequencies or in the short-term spectra of speech sounds produced in noise. The present investigation was aimed at specifying the gross acoustic–phonetic changes that take place when speech is produced under high levels of noise as might be encountered in an aircraft cockpit. We expected to find reliable changes in prosodic parameters such as amplitude, duration, and vocal pitch, which have previously been reported in the literature. We were also interested in various segmental measures related to changes in formant frequencies and in the distribution of spectral energy in the short-term spectra of various segments. These measures might reflect changes in the speaker's source function as well as the articulatory gestures used to implement various classes of speech sounds. In the present study, digital signal processing techniques were used to obtain quantitative measures of changes in the acoustic–phonetic characteristics of speech produced in quiet and in three ambient noise conditions. A second aspect of the study involved perceptual testing with these utterances to verify Dreher and O'Neill's earlier finding that speech produced in noise was more intelligible than speech produced in quiet when the two conditions were presented at equivalent S/N ratios (Dreher and O'Neill, 1957).

I. ACOUSTIC ANALYSES

A. Method

1. Subjects—Two male native English speakers (SC and MD) were recruited as subjects. SC was a graduate student in psychology and was paid $5.00 for his participation. MD was a member of the laboratory staff and participated as part of his routine duties. Both speakers were naive to the purpose of the study and neither speaker reported a hearing or speech problem at the time of testing. Both speakers served for approximately 1 h.

2. Stimulus materials—Stimulus materials consisted of the 15 words in the Air Force speech recognition vocabulary: the digits "zero," "one," "two," "three," "four," "five," "six," "seven," "eight," "nine"; and the control words "enter," "frequency," "step," "threat," and "CCIP." These words were typed into computer files and different randomizations of the list of 15 words were printed out for the subjects to read during the course of the experiment.

3. Procedure—Subjects were run individually in a single-walled sound-attenuated booth (IAC model 401 A). The subject was seated comfortably in the booth and wore a pair of matched and calibrated TDH-39 headphones. An Electrovoice condenser microphone (model C090) was attached to the headset with an adjustable boom. Once adjusted, the microphone remained at a fixed distance of 4 in. from the subject's lips throughout the experiment.

The masking noise consisted of a broadband white noise source that was generated with a Grason–Stadler noise generator (model 1724). The noise was low-pass filtered at 3.5 kHz, using a set of Krohn-Hite filters (model 3202R) with a roll-off of 24 dB per octave, and passed through a set of adjustable attenuators. The masking noise was presented binaurally through the headphones. Subjects wore the headphones during the entire experiment.

Subjects read the words on the test lists under four conditions: quiet, 80, 90, or 100 dB of masking noise in their earphones. The quiet condition measured from 33- to 37-dB SPL background noise with the attenuators set to their maximum setting. Measurements of the noise were made with a B&K sound level meter and artificial ear connected to the earphones.

After the headset was adjusted and the subject became familiar with the environment, a sheet of written instructions was provided to explain the procedures that would be followed. Subjects were informed that they would be reading English words from a list and that they should say each word clearly with a pause of about 1–2 s between words. They were told that masking noise at various levels of intensity would be presented over their headphones during the course of the experiment and that their task was to read each word as clearly as possible into the microphone. They were also told that the experimenter would be listening to their speech outside the booth while the recording was being made. Before the actual recordings were made, both subjects were given about 15 min of practice reading lists of the vocabulary with no noise over the headphones. This was done to familiarize the subjects with the specific vocabulary and the general procedures to be used in making the audiotapes.

Data were collected from subjects reading the lists under all four noise conditions. The noise levels were randomized within each block of four lists with the restriction that over the entire experimental session, every noise level was followed by every other noise level except itself. Subjects took about 40 s to read each list. After each list was read, the masking noise was turned off for about 40 s during which the subjects sat in silence. Each list of 15 words was read in each of the four masking conditions five times, for a total of 300 responses from each subject. Recordings were made on an Ampex AG-500 tape recorder running at $7\frac{1}{2}$ ips.

4. Speech signal processing—Productions of the digits “zero,” “one,” “two,” “three,” “four,” “five,” “six,” “seven,” “eight,” and “nine” were analyzed using digital signal processing techniques. These 400 utterances (ten words × five repetitions × four noise levels × two talkers) were digitized using a VAX 11/750 computer. The utterances were first low-pass filtered at 4.8 kHz and then sampled at a rate of 10 kHz using a 16-bit A/D converter (Digital Sound Corporation model 2000). Each utterance was then digitally edited using a cursor-controlled waveform editor and assigned a file name. These waveform files were then used as input to several digital signal processing analyses.

Linear predictive coding (LPC) analysis was performed on each waveform file. LPC coefficients were calculated every 12.8 ms using the autocorrelation method with a 25.6-ms Hamming window. Fourteen linear prediction coefficients were used in the LPC analyses. The LPC coefficients were used to calculate the short-term spectrum and overall power level of each analysis frame (window). Formant frequencies, bandwidths, and amplitudes were also calculated for each frame from the LPC coefficients. In addition, a pitch extraction algorithm was employed to determine if a given frame was voiced or voiceless and, for voiced frames, to estimate the fundamental frequency ($F0$).

Total duration for each utterance was determined by visual inspection and measurement from a CRT display that simultaneously presented the utterance waveform along with time-aligned, frame-by-frame plots of amplitude, *F*0 (for voiced frames), and formant parameters. Cursor controls were used to locate the onset and offset of each utterance. Following identification of utterance boundaries, a program stored the total duration, mean *F*0, and mean rms energy for each utterance. The onset and offset of the initial vowel of each utterance were also identified and labeled. For each utterance, mean formant frequencies from this vowel segment were also stored. In the case of the word "zero," the initial vowel /i/ could not be reliably segmented apart from the following voiced segments; thus the entire /ire/ segment was used as the initial vowel for this utterance. Similarly, for the utterances "three" and "four," the semivowel /r/ was included as part of the vowel during segmentation.

Finally, the peak amplitude frame (25.6-ms window) from the stressed vowel of each utterance was identified and a regression line was fit to the spectrum of this analysis frame. The slope of this regression line was taken as a measure of "spectral tilt," to quantify the relative distribution of spectral energy at different frequencies.

B. Results and discussion

The influence of ambient noise on various acoustic characteristics of the test utterances is described below. In each case, an analysis of variance was used to determine whether noise level had a significant effect on a given acoustic measure. Separate analyses were carried out for the two talkers. The analyses used word ("zero" through "nine") and noise level as independent variables. The presentation of results will focus on the effect of noise on the various acoustic measures. The "word" variable will be discussed only in cases where a significant word × noise interaction was observed.

1. Amplitude—Mean rms energy for utterances spoken at each noise level are shown for each talker in Fig. 1. The data are collapsed across utterances. For each talker, the measured amplitudes show a consistent increase with an increase in noise level at the talker's ears. The largest increase occurred between the quiet condition and the 80-dB noise condition. Analyses of variance revealed that, for each talker, noise level had a significant effect on amplitude [$F(3,160)=190.41$, $p < 0.0001$ for talker MD, and $F(3,160) =211.15$, $p < 0.0001$ for talker SC]. Newman–Keuls multiple range analyses revealed that, for each talker, each increase in noise led to a significant increase in amplitude (all ps < 0.01). For talker MD, there was also a significant word × noise interaction [$F(27,160) = 1.72$, $p < 0.03$]. For both speakers, the pattern of increased masking noise producing an increase in amplitude was present for every word. The word × noise interaction for speaker MD is due to variability across words in the amount of amplitude increase.

2. Duration—Mean word durations for utterances spoken at each noise level are shown for each speaker in Fig. 2. The data are again collapsed across utterances. The pattern is similar to that observed for amplitude: Word duration shows a consistent increase with each increase in noise at the speakers' ears. However, for speaker MD, the change in duration between the 80- and 90-dB conditions is very small (6 ms). For SC, there is only a slight (15-ms) change in duration across the 80-, 90-, and 100-dB noise conditions. Analyses of variance demonstrated that, for each speaker, noise had a significant effect on word duration [$F(3,160) = 23.08$, $p < 0.0001$ for speaker MD, and $F(3,160) = 25.31$, $p < 0.0001$ for speaker SC]. Newman–Keuls analyses revealed that, for speaker MD, word duration was significantly shorter in the quiet condition than in any of the other conditions (ps < 0.01), and significantly longer in the 100-dB condition than in the other conditions (p s < 0.01). Durations did not significantly differ in the 80- and 90-dB conditions for MD. For speaker

SC, Newman–Keuls tests revealed that duration in the quiet condition was significantly shorter than in the other three conditions (ps < 0.01), but that duration did not significantly vary among the 80-, 90-, and 100-dB noise conditions.

3. Fundamental frequency—Mean fundamental frequencies for utterances spoken at each noise level are plotted separately for each speaker in Fig. 3. The data demonstrate a larger change in *F*0 across noise conditions for speaker SC than for MD. For MD, *F*0 showed a small increase as the noise increased from quiet to 80 dB to 90 dB, followed by a slight drop in *F*0 between the 90- and 100-dB conditions. For SC, a large jump in *F*0 occurred between the quiet and 80-dB noise conditions, followed by small additional increases in F0 in the 90- and 100-dB conditions. Analyses of variance showed a significant effect of noise on *F*0 for each speaker [$F(3,160) = 3.53$, $p < 0.02$ for speaker MD, and $F(3,160) = 42.07$, $p < 0.0001$ for speaker SC]. Newman–Keuls analyses revealed a significant change in *F*0 between the quiet and 90-dB condition for speaker MD ($p < 0.05$). For speaker SC, the Newman–Keuls tests showed that *F*0 in the quiet condition was significantly lower than in any of the other noise conditions (ps < 0,01).

4. Spectral tilt—As mentioned earlier, a regression line was fit to the spectrum of a representative frame from each token. The peak amplitude frame from the initial vowel was identified and used for these measurements. The slope of the regression line was taken as a measure of "spectral tilt" to index the relative energy at high versus low frequencies. Mean spectral tilt values for utterances spoken at each noise level are plotted for each speaker in Fig. 4. For each speaker, there was a decrease in spectral tilt accompanying each increase in noise. This decrease in tilt reflects a change in the relative distribution of spectral energy so that a greater proportion of energy is located in the high-frequency end of the spectrum when utterances are produced in noise. Analyses of variance demonstrated a significant change in spectral tilt across noise conditions for each speaker [$F(3,160) = 56.82$, $p < 0.0001$ for speaker MD, and $F(3,160)= 23.85$, $p < 0.0001$ for speaker SC]. Newman–Keuls analyses revealed a significant decrease in spectral tilt with each increase in noise for speaker MD (ps < 0.01). For SC, spectral tilt was significantly greater in the quiet condition than in any of the other noise conditions (ps < 0.01). In addition, tilt was significantly greater in the 80-dB noise condition than in the 100-dB condition ($p < 0.05$).

On first examination, it appears that the decrease in spectral tilt observed in the high-noise conditions may be due to the increases in *F*0 also observed in these conditions. However, a close examination of these two sets of results suggests that the relative increase in spectral energy at high frequencies in the high-noise conditions is not entirely due to increases in *F*0. For speaker MD, *F*0 did not change a great deal across noise conditions (see Fig. 3); the change in *F*0 was significant only in the quiet versus 90-dB comparison. Yet each increase in noise led to a significant decrease in spectral tilt for speaker MD. For speaker SC, the 80- and 100-dB noise conditions did not differ in the analysis of *F*0, yet a significant decrease in spectral tilt was obtained between these two conditions.

5. Formant frequencies—The influence of masking noise level on vowel formant frequencies was analyzed next. Mean *F*1 and *F*2 frequencies from the initial vowel of each utterance were examined. Noise had a consistent effect on the formant data for speaker SC and a less consistent effect for speaker MD. Mean *F*1 frequencies for utterances produced in each noise condition are shown in Fig. 5. The data for speaker MD appear in the left-hand portion of the figure and the data for speaker SC appear in the middle of the figure. A significant main effect of noise on *F*1 frequency was observed for speaker SC [$F(3,160) = 14.91$, $p < 0.0001$], along with a marginally significant noise × word interaction [$F(27,160)=1.5$, $p < 0.07$]. For this speaker, *F*1 frequency tended to increase as the noise level increased. Newman–Keuls tests revealed that, for this speaker, *F*1 was significantly

lower in the quiet condition than in any of the other noise conditions. The marginally significant noise × word interaction for SC suggests that the pattern of an increase in *F*1 accompanying an increase in noise may not hold for all ten utterances. The consistency of this pattern can be seen by examining Fig. 6. This figure displays *F*1 and *F*2 frequency data for the quiet and 100-dB noise conditions for each of the ten utterances produced by SC. With the exception of the utterance "one," *F*1 was greater in the 100-dB condition than in the quiet condition for all utterances.

The main effect of noise and the noise × word interaction did not reach significance in the analysis of *F*1 frequency for speaker MD. As Fig. 5 shows, the change in mean *F*1 frequency across noise conditions was less than 3 Hz for this speaker. The *F*1 and *F*2 data for speaker MD are broken down by utterance in Fig. 7. Although the noise × word interaction was not significant for MD, the pattern of results shown in this figure suggests that the presence of masking noise may have produced a compacting, or reduction, in the range of *F*1 for this speaker. In the majority of cases, utterances with low *F*1 frequencies showed an increase in *F*1 in noise, while utterances with high *F*1 frequencies showed a decrease in *F*1.

The mean values shown in Figs. 3 and 5 demonstrate a striking similarity between the *F*0 data and the *F*1 data for each speaker. For speaker MD, there was little change in *F*0 across noise conditions and no significant influence of noise on *F*1 frequency. For speaker SC, both *F*0 and *F*1 were significantly higher in the 80-, 90-, and 100-dB noise conditions than in the quiet condition. These data suggest a close relationship between *F*0 and *F*1; apparently, an increase in fundamental frequency leads to an increase in *F*1. We carried out one additional analysis to further test this conclusion.

In order to determine whether *F*0 and *F*1 were, in fact, directly related, a second analysis was run on speaker SC's data. In this analysis, the effects of word and noise level on initial-vowel *F*1 frequency were again tested but with initial-vowel *F*0 entered as a covariate in the analysis. Mean *F*1 frequencies at each noise level based on the adjusted cell means from this analysis (in which *F*0 is covaried out) appear in the right-hand portion of Fig. 5. The results of this analysis were nearly identical to those observed in the original analysis off 1 frequency for SC. The main effect of noise on *F*1 frequency remained significant [$F(3,159) = 5.32$, $p < 0.0017$]. Also, as in the original analysis off *F*1 for speaker SC, the noise × word interaction fell short of significance [$F(27,159)=1.44$, $p < 0.09$]. Finally, Newman–Keuls tests comparing *F*1 frequencies in the various noise conditions revealed the identical pattern observed in the original analysis: *F*1 frequency was significantly lower in the quiet condition than in any of the other noise conditions. Thus, for speaker SC, it appears that noise had an influence on *F*1 frequency independent of its influence on *F*0.

Turning to the *F*2 data, masking noise did not produce a significant main effect on *F*2 frequency for speaker SC. However, a significant noise × word interaction was present [$F(27,160) = 1.92$, $p < 0.008$]. An examination of Fig. 6 suggests that the range of *F*2 frequencies was reduced in the presence of noise for speaker SC. Utterances containing high *F*2 frequencies showed a decrease in *F*2 in the 100-dB condition, while utterances with low *F*2 frequencies showed increases in *F*2 when noise was increased.

The main effect of noise and the noise × word interaction did not approach significance in the analysis of *F*2 frequency for speaker MD. An examination of Fig. 7 shows that, for most utterances, *F*2 showed little change between the quiet and 100-dB noise condition for this speaker.

Fundamental frequency, amplitude, and duration all tended to increase in the presence of noise. In addition, the results demonstrated consistent differences in the spectral characteristics of vowels produced in noise versus quiet. Vowels from utterances produced

in noise had relatively flat spectra, with a relatively large proportion of their total energy occurring in higher frequency regions. Vowels from utterances produced in quiet had steeper spectra with relatively little energy present in high-frequency regions. First formant frequencies also appeared to be influenced by the presence of noise for at least one speaker. For SC, *F*1 frequencies were higher for vowels from utterances produced in the three noise conditions than for vowels produced in the quiet. There was little change in *F*2 frequencies across noise conditions for either speaker.

The present results demonstrated several clear differences in the acoustic characteristics of speech produced in quiet compared to speech produced in noise. Previous research by Dreher and O'Neill (1957) suggests that the changes in the spectral and temporal properties of speech which accompany the Lombard effect improve speech intelligibility. We carried out two separate perceptual experiments to verify their earlier conclusions.

II. PERCEPTUAL ANALYSES—EXPERIMENT I

In experiment I, subjects identified utterances from the quiet condition and the 90-dB masking noise condition in a forced-choice identification task. Utterances from the quiet and 90-dB noise condition were mixed with broadband noise at equivalent S/N ratios and presented to listeners for identification. If Dreher and O'Neill's conclusion concerning the intelligibility of speech produced in noise versus quiet is correct, subjects should identify utterances produced in the 90-dB noise condition more accurately than utterances produced in the quiet condition.

A. Method

1. Subjects—Subjects were 41 undergraduate students who participated to fulfill a requirement for an introductory Psychology course. All subjects were native English speakers and reported no previous history of a speech or hearing disorder at the time of testing.

2. Stimuli—Stimulus materials were the tokens of the digits zero through nine, produced in quiet and 90 dB of masking noise by both talkers. For each talker, the five tokens of each word produced in each masking condition were used for a total of 100 utterances per masking condition (five tokens × ten digits × two talkers). All stimuli were equated in terms of overall rms amplitude using a program that permits the user to manipulate signal amplitudes digitally (Bernacki, 1981).

3. Procedure—Stimulus presentation and data collection were controlled by a PDP 11/34 computer. Stimuli were presented via a 12-bit digital-to-analog converter over matched and calibrated TDH-39 headphones. Wideband noise, filtered at 4.8 kHz, was mixed with the signal during stimulus presentation.

The 200 utterances from the quiet and 90-dB masking conditions were randomized and presented to subjects in one of three S/N conditions: − 5-, − 10-, and − 15-dB S/N ratio. The S/N ratio was manipulated by varying signal amplitude while holding masking noise constant at 85 dB SPL. Stimuli were presented at 70 dB SPL in the − 15-dB S/N condition, 75 dB SPL in the − 10-dB S/N condition, and 80 dB SPL in the − 5-dB S/N condition.

Subjects were tested in small groups in a sound-treated room and were seated at individual testing booths equipped with terminals interfaced to the PDP 11/34 computer. At the beginning of an experimental trial, the message "READY FOR NEXT WORD" appeared on each subject's terminal screen. The 85-dB SPL masking noise was presented over the headphones 1 s later. A randomly selected stimulus was presented for identification 100 ms

following the onset of masking noise. Masking noise was terminated 100 ms following stimulus offset. A message was then displayed on each subject's screen instructing the subject to identify the stimulus. Subjects responded by depressing one of the ten digit keys on the terminal keyboard. Subjects were presented with two blocks of 200 experimental trials. Within each block, each of the 200 utterances was presented once.

4. Design—All 200 test utterances were presented to each subject. Thus talker (MD or SC) and masking noise condition (quiet or 90 dB) were manipulated as within-subjects factors. The S/N ratio in the listening conditions was manipulated as a between-subjects factor. Subjects were randomly assigned to one of the three S/N conditions. Thirteen subjects participated in the − 15-dB S/N condition, 13 participated in the − 10-dB condition, and 15 participated in the − 5-dB condition.

B. Results

The percentage of correct digit responses is displayed separately by speaker (MD or SC), masking noise condition (quiet or 90 dB SPL), and S/N ratio (− 5, − 10, or − 15 dB) in Fig. 8. A three-way ANOVA was carried out on these data using speaker, masking noise, and S/N ratio as independent variables.

As expected, S/N ratio had a significant main effect on identification [$F(2,38) = 202.91$, $p < 0.0001$]. As shown in Fig. 8, performance was highest in the − 5-dB S/N condition, somewhat lower in the − 10-dB condition, and lowest in the − 15-dB condition. This pattern was observed for both talkers and for both the quiet and 90-dB noise conditions.

Turning to the main focus of the experiment, masking noise produced a significant main effect on identification [$F(1,38) = 162.75$,$p < 0.0001$]. Digits produced in 90 dB of masking noise were consistently identified more accurately than digits produced in the quiet regardless of talker or S/N ratio (see Fig. 8).

A significant interaction was observed between masking noise and S/N ratio [$F(2,38) = 11.04$, $p < 0.0003$). For each speaker, as S/N ratio decreased, the effect of masking noise on identification accuracy increased. Thus the difference in performance for digits produced in quiet versus 90-dB masking noise was smallest in the − 5-dB S/N condition, greater in the − 10-dB condition, and greatest in the − 15-dB condition. Apparently, the acoustic–phonetic differences between the utterances produced in quiet versus 90 dB of noise had a greater influence on intelligibility as the S/N ratio decreased.

A significant noise × talker interaction was also obtained [$F(1,38) = 5.68$, $p < 0.03$). At each S/N ratio, the influence of masking noise on identification accuracy was greater for speaker MD than for speaker SC.

III. PERCEPTUAL ANALYSE—EXPERIMENT II

In experiment I, digits produced in noise were recognized more accurately than digits produced in quiet. The consistency of this effect in experiment I is quite remarkable given that the stimuli were drawn from a very small, closed set of highly familiar test items. To verify that the results of experiment I were reliable and could be generalized, we replicated the experiment with a different set of stimuli drawn from the original test utterances.

A. Method

Experiment II was carried out with stimuli taken from the 100-dB masking condition. That is, the replication used the 200 stimuli from the quiet and 100-dB masking conditions. In this experiment, ten subjects participated in the − 15-dB S/N condition, nine subjects

participated in the − 10-dB condition, and ten subjects participated in the − 5-dB condition. All subjects were native speakers of American English and met the same requirements as those used in the previous experiment. All other aspects of the experimental procedure were identical to those of experiment I.

B. Results and discussion

The results of experiment II are shown in Fig. 9. Percent correct identification is broken down by speaker (MD or SC), masking noise (quiet or 100 dB SPL), and S/N ratio (− 5, − 10, or − 15 dB). As in the previous experiment, a three-way ANOVA was carried out on these data using talker, masking noise, and S/N ratio as independent variables.

Comparing the data shown in Figs. 8 and 9, it can be seen that the pattern of means obtained in the two experiments is nearly identical. The results of the ANOVA performed on the data from this experiment also replicate the results of the previous experiment. A significant main effect of S/N ratio was obtained. Identification accuracy decreased as S/N ratio decreased [$F(2,26) = 117.33$, $p < 0.0001$]. There was also a significant main effect of talker [$F(1,26) = 39.79$, $p < 0.0001$]. As in the first experiment, SC's tokens were identified more accurately than MD's tokens.

Each of the significant effects involving the masking noise variable reported in experiment I was also replicated. There was a significant main effect of masking noise [$F(1,26) = 249.84$, $p < 0.0001$]. Utterances produced in 100 dB of noise were more accurately identified than utterances produced in the quiet. Significant interactions were observed between masking noise and S/N ratio [$F(2,26) = 9.46$, $p = 0.0009$] and between masking noise and talker [$F(1,26) = 41.16$, $p < 0.0001$]. As in experiment I, the effect of masking noise on identification accuracy increased as S/N ratio decreased. Also replicating the results of experiment I, the effect of masking noise on performance was greater for talker MD than for talker SC.

The results of these perceptual experiments replicate the findings of Dreher and O'Neill (1957). In their earlier research, as in each of the perceptual experiments reported here, subjects were more accurate at identifying utterances originally produced in noise than utterances produced in quiet. This pattern was found for each talker's utterances and at each S/N ratio in the present experiments. Furthermore, in each experiment, the effect of masking noise on intelligibility increased as S/N ratio decreased. Thus differences in the acoustic–phonetic structure of utterances produced in quiet and utterances produced in noise had reliable effects on intelligibility. The magnitude of these effects increased as the environment became more severe (as S/N ratio decreased).

IV. GENERAL DISCUSSION

The results of the present acoustic analyses demonstrate reliable and consistent differences in the acoustic properties of speech produced in quiet environments and environments containing high levels of masking noise. The differences we observed in our analyses were not restricted only to the prosodic properties of speech such as amplitude, duration, and pitch, but were also present in measurements of vowel formant frequencies. Moreover, for both talkers, we observed substantial changes in the slopes of the short-term power spectra of vowels in these utterances, which were shifted upward to emphasize higher frequency components.

The changes in amplitude, fundamental frequency, and duration reported here were often fairly small across the different noise levels. In particular, in comparing the 80-dB and 100-dB conditions, the change in amplitude was about 2 dB for each speaker. This 2-dB increase

in the face of a 20-dB increase in masking noise is much smaller than would be predicted from previous research. Research using communication tasks involving talker–listener pairs have generally reported a 5-dB increase in signal amplitude for each 10-dB increase in noise (Lane *et al.*, 1970; Webster and Klump, 1962). The smaller differences observed in the present study suggest that masking noise may have a greater influence on speech in interactive communication tasks involving talker–listener pairs than in noninteractive tasks, such as the one used here, where no external feedback is available. Despite the magnitude of the observed differences, the findings are reliable and demonstrate important changes in speech produced in various noise conditions.

The results from the two perceptual experiments demonstrated that speech produced in noise was more intelligible than speech produced in quiet when presented at equal S/N ratios. Apparently, several acoustic characteristics of speech produced in noise, above and beyond changes in rms amplitude, make it more intelligible in a noisy environment than speech produced in the quiet. The present results also show that these acoustic differences play a greater and greater role as the S/N ratio decreases in the listener's environment.

The present findings replicate several gross changes in the prosodic properties of speech which have been previously reported in the literature (Hanley and Steer, 1949; Draegert, 1951). For one of our two speakers, the results also demonstrate a clear influence of masking noise on the formant structure of vowels. We believe that the present results have a number of important implications for the use of speech recognition devices in noisy environments and for the development of speech recognition algorithms, especially algorithms designed to operate in noise or severe environments.

In the recent past, a major goal of many speech scientists and engineers working on algorithm development has been to improve recognition of speech in noise (Rollins and Wiesen, 1983). Most efforts along these lines have involved the development of efficient techniques to extract speech signals from background noise (Neben *et al.*, 1983). Once the speech signal was extracted and the noise "stripped off" or attenuated, recognition could proceed via traditional pattern recognition techniques using template matching. Other efforts have attempted to solve the speech-in-noise problem by developing procedures that incorporate noise into the templates that is similar to the noise in the testing environment (Kersteen, 1982). By this technique, the signal does not have to be extracted from the noise; rather the entire pattern containing signal and noise is matched against the stored template.

This second technique, of incorporating noise into the templates, is accomplished by training the speech recognizer in a noisy environment so that noise along with speech is sampled on each trial. Kersteen (1982) reported success with this method of training; the highest recognition performance was produced when training and testing occurred in the same noise environment. Kersteen (1982) interpreted these results as demonstrating the importance of incorporating noise into the templates when noise is also present at testing.

An alternative explanation for the success of this training method is that the templates produced in noise capture acoustic characteristics of speech produced in noise that differ from those of speech produced in quiet. Unfortunately, little, if any, attention has been devoted to examining the changes in the speech signal that occur when a talker speaks in the presence of masking noise. The present findings demonstrate reliable differences in the acoustic–phonetic structure of speech produced in quiet versus noisy environments. Because of these differences, the problem of speech recognition in noise is more complicated than it might seem at first glance. The problem involves not only the task of identifying what portion of the signal is speech and what portion is noise but it also involves dealing with the

changes and modifications that take place in the speech signal itself when the talker produces speech in noise.

Any speech recognition algorithm that treats speech as an arbitrary signal and fails to consider the internal acoustic–phonetic specification of words will have difficulty in recognizing speech produced in noise. This difficulty should be particularly noticeable with the class of current algorithms that is designed around standard template matching techniques. These algorithms are, in principle, incapable of recovering or operating on the internal acoustic–phonetic segmental structure of words and the underlying fine-grained spectral changes that specify the phonetic feature composition of the segments of the utterance. Even if dynamic programming algorithms are used to perform time warping before pattern matching takes place, the problems we are referring to here still remain. Factors such as changes in speaking rate, masking noise, or increases in cognitive load may affect not only the fairly gross attributes of the speech signal but also the fine-grained segmental structure as well. Moreover, as far as we can tell, changes in speaking rate, effects of noise, and differences in cognitive load, to name just a few factors, appear to introduce nonlinear changes in the acoustic–phonetic realization of the speech signal. To take one example, it is a well-known finding in the acoustic–phonetic literature that consonant and vowel durations in an utterance are not increased or decreased uniformly when a talker's speaking rate is changed (see Miller, 1981, for a review). Thus simple linear scaling of the speech will not be sufficient to capture rate-related changes in the acoustic–phonetic structure.

The present findings are also relevant to a number of human factors problems in speech recognition. Both of the speakers examined in this study adjusted their speech productions in response to increased masking noise in their environment. These adjustments made the speech produced in noise more intelligible than speech produced in quiet when both were presented at equal amplitudes in a noisy environment. The speakers appeared to automatically adjust the characteristics of their speech to maintain intelligibility without having been explicitly instructed to do so. Presumably, the increase in intelligibility would have been at least as great if the speakers had been given such instructions. Given the currently available recognition technology, it should be possible to train human talkers to improve their performance with speech recognizers by appropriate feedback and explicit instructions.

In this regard, the present findings are related to several recent investigations in which subjects received explicit instructions to speak clearly (Chen, 1980; Picheny *et al*, 1985; Picheny *et al*, 1986). These studies on "clear speech" suggest that subjects can readily adjust and modify the acoustic–phonetic characteristics of their speech in order to increase intelligibility. Picheny *et al.* (1985) collected non-sense utterances spoken in "conversational" or "clear speech" mode. In conversational mode, each talker was instructed to produce the materials "in the manner in which he spoke in ordinary conversation." In clear speech mode, talkers were instructed to speak "as clearly as possible." Utterances produced in clear speech mode were significantly more intelligible than utterances produced in conversational mode when presented to listeners with sensorineural hearing losses. Chen (1980) reported the same pattern of results when "clear" and "conversational" speech was presented to normal-hearing subjects in masking noise.

Picheny *et al.* (1986) and Chen (1980) also carried out acoustic analyses to identify differences in the acoustic characteristics of clear and conversational speech. Many of the differences they identified are similar to those reported here. Specifically, longer segment durations, higher rms amplitudes, and higher $F0$ values were reported for clear speech versus conversational speech. These changes in amplitude, duration, and pitch are also

characteristic of speech that is deliberately emphasized or stressed by the talker (Lieberman, 1960; Klatt, 1975; Cooper *et al*, 1985). Thus clear speech, emphasized or stressed speech, and speech produced in noise all tend to show increases in these three prosodic characteristics.

The pattern of formant data shows less similarity between speech produced in noise and clear speech or emphasized (stressed) speech. Chen (1980) reported that in clear speech *F*1 and *F*2 moved closer to target values. This movement enlarges the vowel space and makes formant values for different vowels more distinct, a pattern that is also characteristic of stressed vowels (Delattre, 1969). Our vowel formant data do not display this pattern of change. In the present study, masking noise produced increases in *F*1 frequency for speaker SC but had little effect on formant frequencies for MD. Thus it appears that the presence of masking noise did not produce the same qualitative changes in production as instructions to speak clearly or to stress certain utterances. While several parallel changes occur in each case, a number of differences are also present in the data.

The literature on "shouted" speech also provides an interesting parallel to the present findings. Increases in fundamental frequency, vowel duration, and *F*1 frequency have all been reported for shouted speech (Rostolland and Parant, 1974; Rostolland, 1982a,b). In addition, spectral tilt is reduced in shouted speech (Rostolland, 1982a). Each of these findings is in agreement with the present data for speech produced in noise. Thus it appears that the differences between speech produced in quiet and speech produced in noise are similar in kind to the differences between spoken and shouted speech. However, for each of the variables mentioned above, the differences between shouted and spoken speech are greater in magnitude than the present differences between speech produced in quiet and speech produced in noise.

In the present investigation, we found that speech produced in noise was more intelligible than speech produced in quiet when presented at equal S/N ratios. It would, therefore, be reasonable to expect that shouted speech should also be more intelligible than conversational speech in similar circumstances. However, the literature reports exactly the opposite result: When presented at equal S/N ratios, shouted speech is less intelligible than conversational speech (Pickett, 1956; Pollack and Pickett, 1958; Rostolland, 1985). While our talkers were able to increase the intelligibility of their speech by making changes in speech production that appear similar in kind to those reported for shouted speech, the magnitude of these changes is much greater in shouted speech. The extreme articulations that occur in shouted speech apparently affect intelligibility adversely, perhaps introducing distortions or perturbations in the acoustic realizations of utterances (Rostolland, 1982a,b).

In addition to the recent work of Picheny *et al.* (1985, 1986) and Chen (1980) on clear speech, there is an extensive literature in the field of speech communication from the 1940s and 1950s that was designed to improve the intelligibility of speech transmitted over noisy communication channels. Instructions to talk loudly, articulate more precisely, and talk more slowly have been shown to produce reliable gains in speech intelligibility scores when speech produced under adverse or noisy conditions is presented to human listeners for perceptual testing (see, for example, Tolhurst, 1954, 1955). Unfortunately, at the present time, we simply do not know whether these same training and feedback techniques will produce comparable improvements in performance with speech recognition systems. It is clearly of some interest and potential importance to examine these factors under laboratory conditions using both speech recognizers and human observers. This line of research may yield important new information about the variability of different talkers and the "goat" and "sheep" problem discussed by Doddington and Schalk (1981). If we knew more precisely which acoustic–phonetic characteristics of speech spectra separate goats from sheep, we

would be in a better position to suggest methods to selectively modify the way talkers speak to speech recognizers through training and directed feedback (see Nusbaum and Pisoni, 1987). We consider this to be an important research problem that has been seriously neglected by engineers and speech scientists working on the development of new algorithms for speech recognition. The human talker is probably the most easily modified component of a speech recognition system. In addition to being the least expensive component to change or modify, it is also the most accessible part of the system. Thus substantial gains in performance in restricted task environments should be observed simply by giving talkers directed feedback about precisely how they should modify the way they talk to the system. To this end, during training, the recognition system could provide the talker with much more information than a simple yes/no decision about the acceptance or rejection of an utterance. There is every reason to believe that a talker's speech can be modified and controlled in ways that will improve the performance of speech recognizers, even poorly designed recognizers that use fairly crude template-matching techniques.

We should qualify these remarks by also noting that these expected gains in performance can only be realized by additional basic research on how humans talk to speech recognizers under a wider variety of conditions. The results reported in the present article demonstrate that talking in the presence of masking noise not only affects the prosodic aspects of speech but also the relative distribution of energy across the frequency spectrum and the fine-grained acoustic–phonetic structure of speech as revealed in the formant frequency data. If we knew more about the role of feedback in speech production, and if we had more knowledge about the basic perceptual mechanisms used in speech perception, we would obviously have a much stronger and more principled theoretical basis for developing improved speech recognition algorithms specifically designed around general principles known to affect the way humans speak and listen.

The present investigation has a number of limitations that are worth discussing in terms of generalizing the findings beyond the present experimental context. First, we used isolated words spoken in citation form. It is very likely that a number of additional and quite different problems would be encountered if connected or continuous speech were used for these tests. The effects we observed with isolated words may be even more pronounced if the test words are put into context or concatenated together into short phrases or sentences.

Second, the subjects in this experiment did not receive any feedback about the success or failure of their communication. They were simply told that the experimenter was listening and recording their utterances. Clearly, there was little incentive for the speaker to consciously change his speech even with masking noise present in the headphones. It seems reasonable to suppose that much larger changes might have been observed in the acoustic–phonetic properties of the utterances produced under masking noise if some form of feedback were provided to the talker to induce him to modify his articulations to improve intelligibility.

Finally, in these tests, no sidetone was provided to the talker through his headphones. In standard military communication tasks, sidetone is typically provided through the headphones and often serves as an important source of feedback that can modify the talker's speech output. As in the case of masking noise, it is not clear how auditory sidetone affects the acoustic–phonetic properties of talker's speech other than increasing or decreasing amplitude (see Lane and Tranel, 1971). Obviously, this is an area worthy of future research as it may be directly relevant to problems encountered in attempting to modify the way a talker speaks to a speech recognizer under adverse conditions. Thus automatic changes in the level of the sidetone may not only cause the talker to speak more loudly into the

recognizer, but may also help him to articulate his speech more precisely and, therefore, improve performance with little additional cost to the system.

V. CONCLUSIONS

The problem of recognizing speech produced in noise is not just a simple problem of detection and recognition of signals mixed in noise. Speakers modify both the prosodic and segmental acoustic–phonetic properties of their speech when they talk in noise. Consequently, important changes in the physical properties of the speech signal must be considered along with the simple addition of noise to the signal in solving the recognition problem.

The presence of masking noise in a talker's ears not only affects the prosodic attributes of speech signals but affects the segmental properties as well. Talkers not only speak louder and slower in noise, but they also raise their vocal pitch and introduce changes in the short-term power spectrum of voiced segments. Talkers also introduce changes in the pattern of vowel formant frequencies.

In trying to articulate speech more precisely under these adverse conditions, the talker introduces certain changes in the acoustic–phonetic correlates of speech that are similar to those distinguishing stressed utterances from unstressed utterances. The changes in the prosodic properties of speech which occur in noise are also similar to changes that occur when subjects are explicitly instructed to "speak clearly." However, the *F*1 and *F*2 data suggest that the changes in productions that subjects automatically make when speaking in noise are not identical to the changes that occur when subjects are given clear speech instructions or when subjects put stress or emphasis on particular utterances.

The results of this study, taken together with the earlier findings reported in the literature on improving the intelligibility of speech in noise, suggest that it may be possible to train talkers to improve their performance with currently available speech recognizers. Directed feedback could be provided to talkers about their articulation and how it should be selectively modified to improve recognition. If this type of feedback scheme were employed in an interactive environment, substantial gains might also be made in reducing the variability among talkers. Thus changes in a talker's speech due to high levels of masking noise, physical or psychological stress, or cognitive load could be accommodated more easily by readjustments or retuning of an adaptive system.

The present findings also suggest that the performance of current speech recognizers could be improved by incorporating specific knowledge about the detailed acoustic–phonetic changes in speech that are due to factors in the talker's physical environment such as masking noise, physical stress, and cognitive load. Some of these factors appear to introduce reliable and systematic changes in the speech waveform and, therefore, need to be studied in much greater detail in order to develop speech recognition algorithms that display robust performance over a wide variety of conditions.

Acknowledgments

The research reported here was supported, in part, by Contract No. AF-F-33615-86-C-0549 from Armstrong Aerospace Medical Research Laboratory, Wright–Patterson AFB, Ohio, and, in part, by NIH Research Grant NS-12179. This report is Tech. Note 88-01 under the contract with AAMRL. We thank Cathy Kubaska, Howard Nusbaum, and Moshe Yuchtman for numerous contributions in carrying out this project. We also thank Diane Kewley-Port for her help in collecting the speech from the two talkers, and Michael Cluff for his help in running subjects in the perceptual experiments.

References

Bernacki, B. Research on Speech Perception Progress Report No. 7. Bloomington, IN: Speech Research Laboratory, Indiana University; 1981. WAVMOD: A program to modify digital waveforms; p. 275-286.

Chen, FR. Unpublished Master's thesis. Cambridge, MA: Massachusetts Institute of Technology; 1980. Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level.

Cooper WE, Eady SJ, Mueller PR. Acoustical aspects of contrastive stress in question–answer contexts. J. Acoust. Soc. Am. 1985; 77:2142–2156. [PubMed: 4019901]

Delattre P. An acoustic and articulatory study of vowel reduction in four languages. lnt. Rev. Appl. Linguist. 1969; 7:295–325.

Doddington GR, Schalk TB. Speech recognition: Turning theory to practice. IEEE Spectrum. 1981; 18:26–32.

Draegert GL. Relationships between voice variables and speech intelligibility in high level noise. Speech Monogr. 1951; 18:272–278.

Dreher JJ, O'Neill JJ. Effects of ambient noise on speaker intelligibility for words and phrases. J. Acoust. Soc. Am. 1957; 29:1320–1323.

Hanley TD, Steer MD. Effect of level of distracting noise upon speaking rate, duration and intensity. J. Speech Hear. Disord. 1949; 14:363–368.

Kersteen, ZA. An evaluation of automatic speech recognition under three ambient noise levels. paper presented at the Workshop on Standardization for Speech I/O Technology, National Bureau of Standards; 18–19 March; Gaithersburg, MD. 1982.

Klatt DH. Vowel lengthening is syntactically determined in connected discourse. J. Phon. 1975; 3:129–140.

Ladefoged, P. Three Areas of Experimental Phonetics. London: Oxford U. P; 1967.

Lane HL, Tranel B. The Lombard sign and the role of hearing in speech. J. Speech Hear. Res. 1971; 14:677–709.

Lane HL, Tranel B, Sisson C. Regulation of voice communication by sensory dynamics. J. Acoust. Soc. Am. 1970; 47:618–624. [PubMed: 5439662]

Lieberman P. Some acoustic correlates of word stress in American English. J. Acoust. Sac. Am. 1960; 32:451–454.

Lombard E. Le signe de I'elevation de la voix. Ann. Mal. Oreil. Larynx. 1911; 37:101–119. (Cited by Lane and Tranel, 1971.).

Miller, JL. Effects of speaking rate on segmental distinctions. In: Eimas, PD.; Miller, JL., editors. Perspectives on the Study of Speech. Hillsdale, NJ: Erlbaum; 1981.

Neben G, McAulay RJ, Weinstein CJ. Experiments in isolated word recognition using noisy speech. Proc. Int. Conf. Acoust. Speech Signal Process. 1983:1156–1158.

Nusbaum HC, Pisoni DB. Automatic measurement of speech recognition performance: a comparison of six speaker-dependent recognition devices. Comput. Speech Lang. 1987; 2:87–108.

Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing I: lntelligibility differences between clear and conversational speech. J. Speech Hear. Res. 1985; 28:96–103. [PubMed: 3982003]

Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. J. Speech Hear. Res. 1986; 29:434–446. [PubMed: 3795886]

Pickett JM. Effects of vocal force on the intelligibility of speech sounds. J. Acoust. Soc. Am. 1956; 28:902–905.

Pollack I, Pickett JM. Masking of speech by noise at high sound levels. J. Acoust. Soc. Am. 1958; 39:127–130.

Rollins A, Wiesen J. Speech recognition and noise. Proc. Int. Conf. Acoust. Speech Signal Process. 1983:523–526.

Rostolland D. Acoustic features of shouted voice. Acustica. 1982a; 50:118–125.

Rostolland D. Phonetic structure of shouted voice. Acustica. 1982b; 51:80–89.

Rostolland D. Intelligibility of shouted voice. Acustica. 1985; 57:103–121.

Rostolland, D.; Parant, C. Physical analysis of shouted voice. paper presented at the Eighth International Congress on Acoustics; London. 1974.

Tolhurst, GC. Joint Project Rep. No. 35. Pensacola, FL: U.S. Naval School of Aviation Medicine, Naval Air Station; 1954 Nov 30. The effect on intelligibility scores of specific instructions regarding talking.

Tolhurst, GC. Joint Project Rep. No. 58. Pensacola, FL: U.S. Naval School of Aviation Medicine, Naval Air Station; 1955 Jul 30. The effects of an instruction to be intelligible upon a speaker's intelligibility, sound pressure level and message duration.

Webster JC, Klumpp RG. Effects of ambient noise and nearby talkers on a face-to-face communication task. J. Acoust. Soc. Am. 1962; 34:936–941.

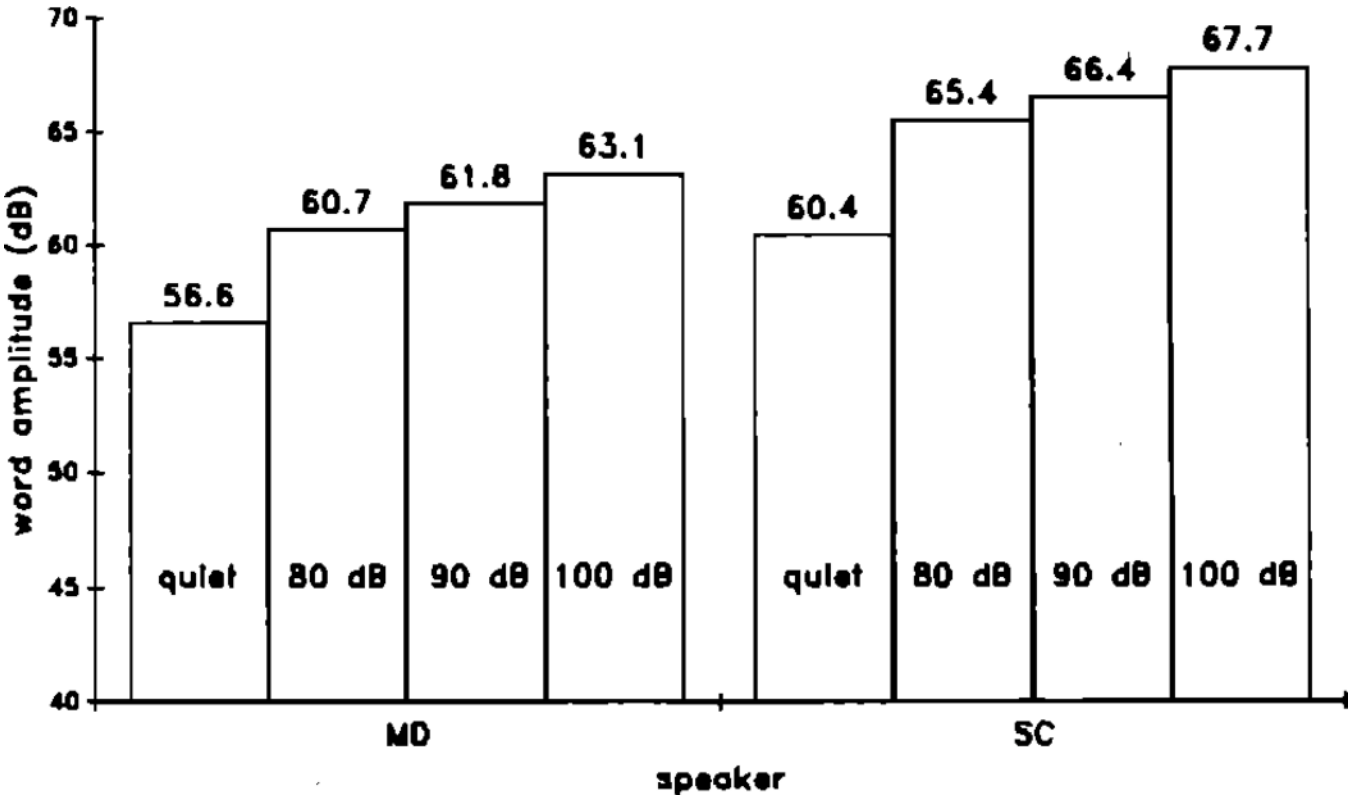

FIG. 1.
Mean rms amplitudes for words produced in quiet, 80, 90, and 100 dB of masking noise. Values are collapsed across utterances and presented separately for each speaker.

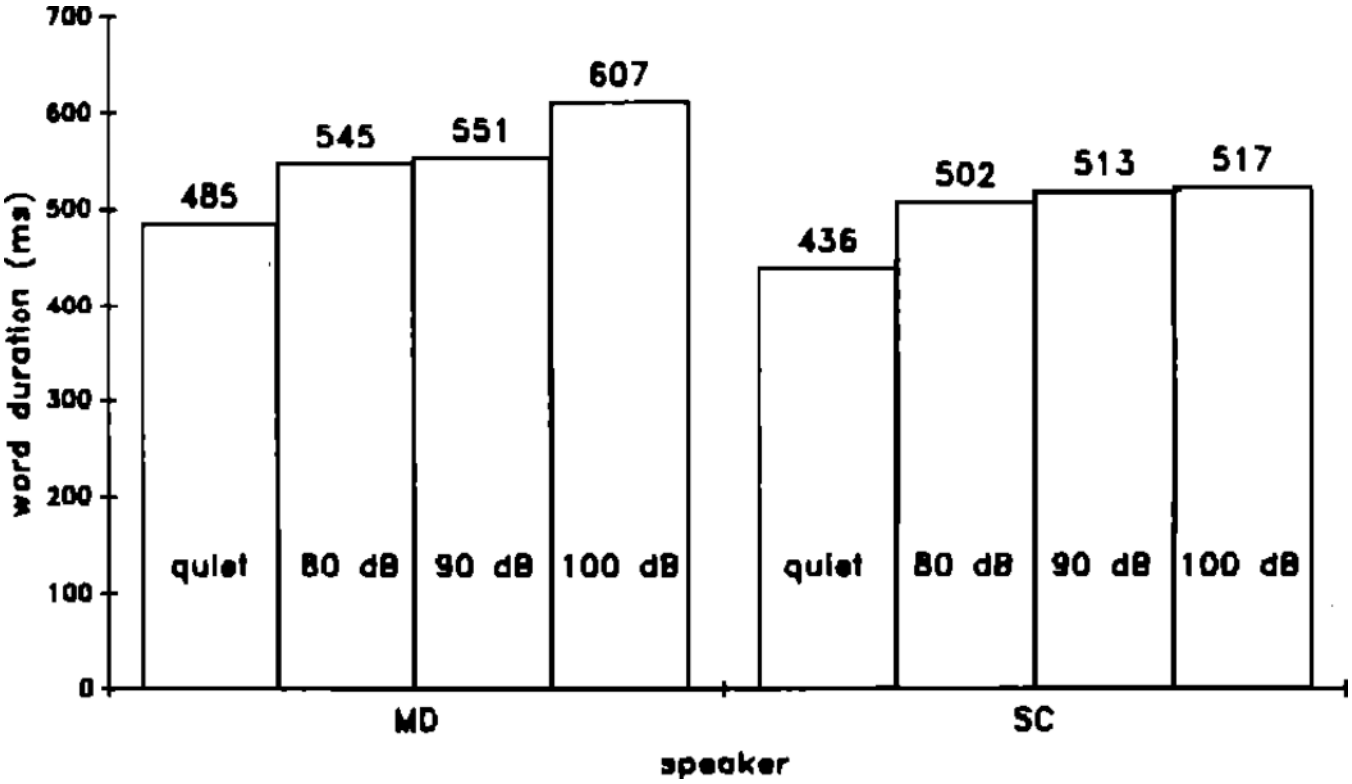

FIG. 2.
Mean durations for words produced in quiet, 80, 90, and 100 dB of masking noise. Values are collapsed across utterances and presented separately for each speaker.

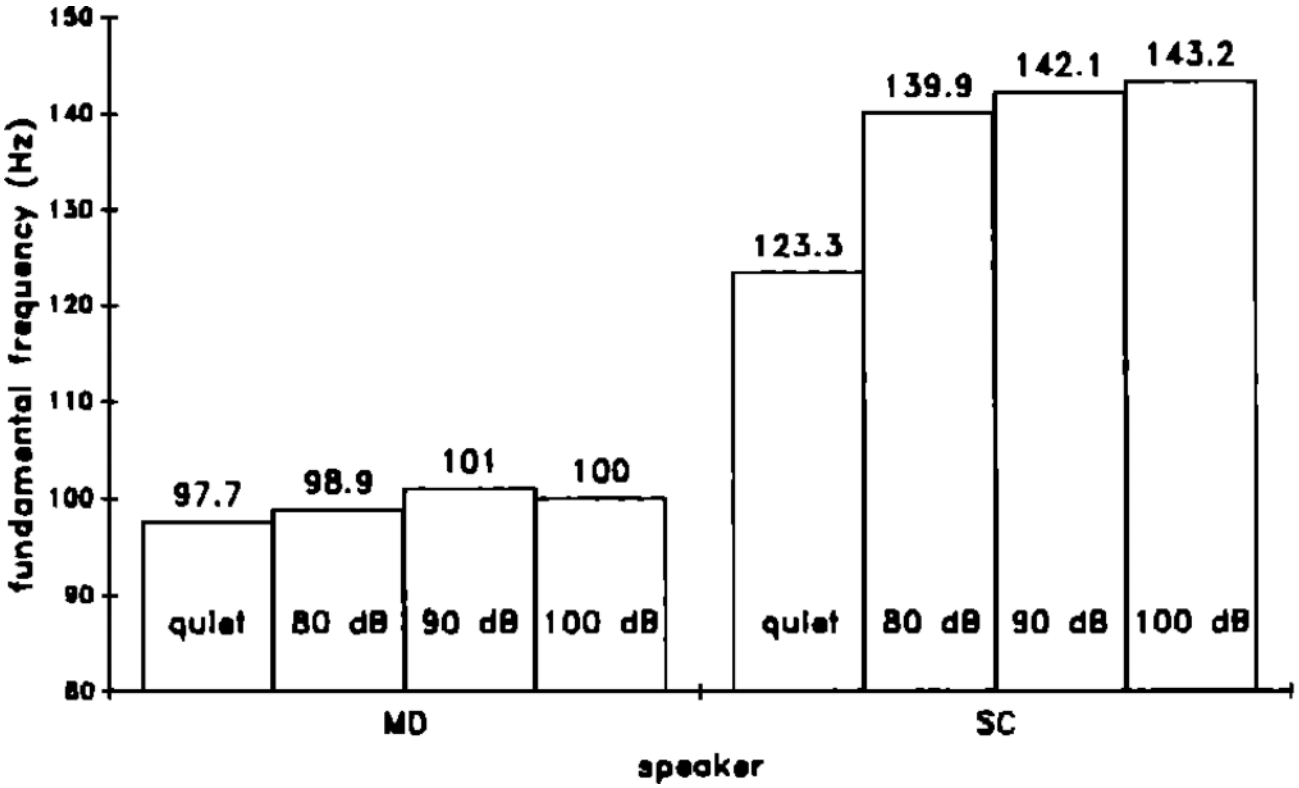

FIG. 3.

Mean fundamental frequency values for words produced in quiet, 80, 90, and 100 dB of masking noise. Values are collapsed across utterances and presented separately for each speaker.

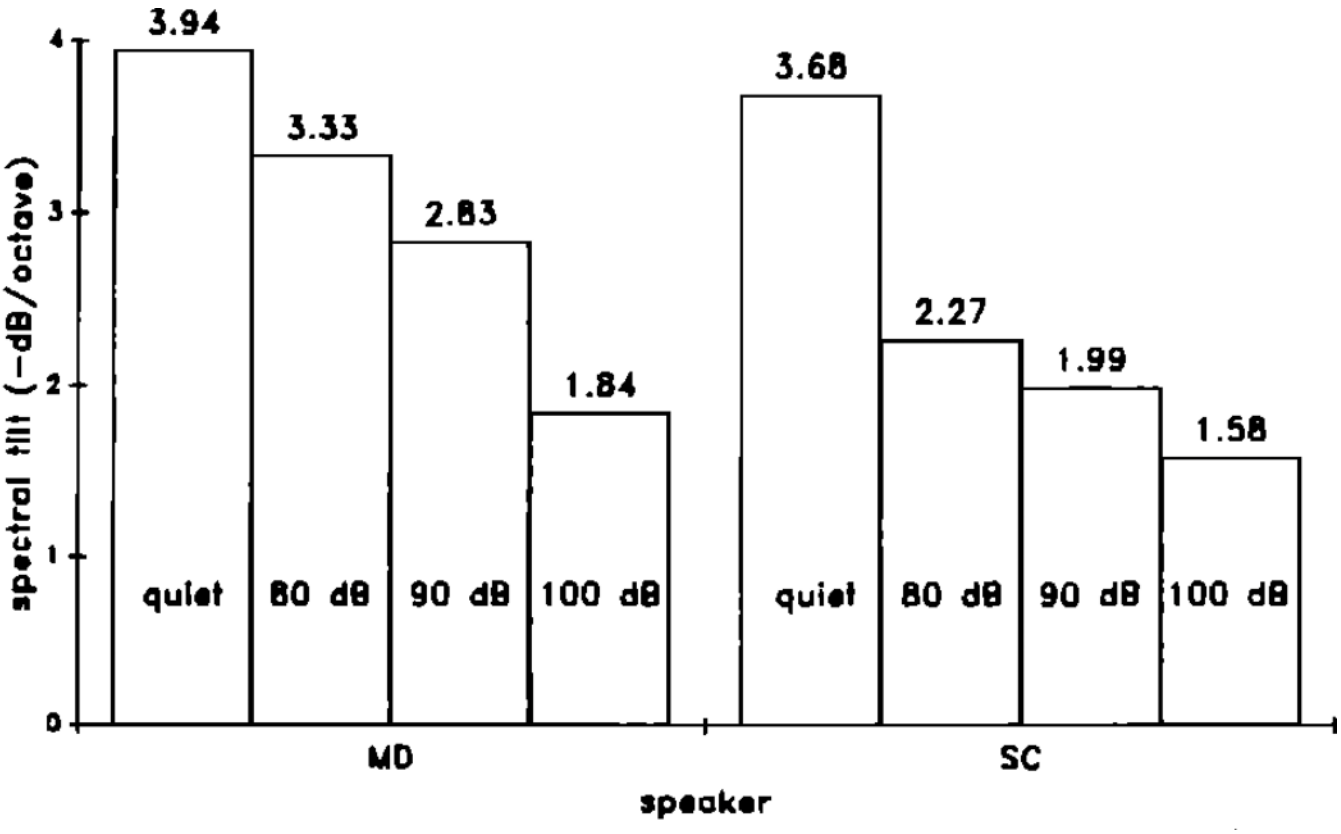

FIG. 4.
Mean spectral tilt values for words produced in quiet, 80, 90, and 100 dB of masking noise. Values are collapsed across utterances and presented separately for each speaker.

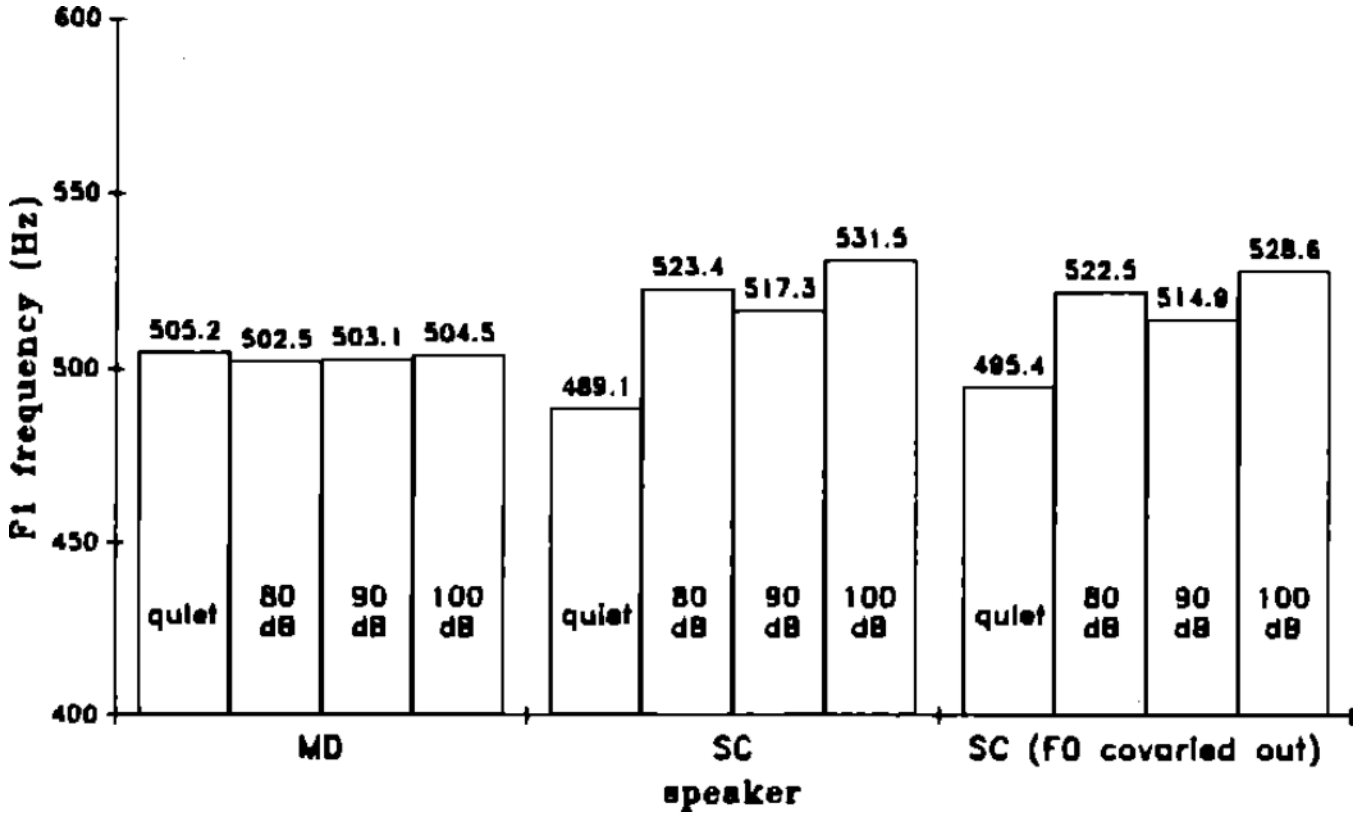

FIG. 5.

Mean first formant frequency values for words produced in quiet, 80, 90, and 100 dB of masking noise. Values are collapsed across utterances and presented separately for speaker MD, speaker SC, and speaker SC with *F*0 covaried out.

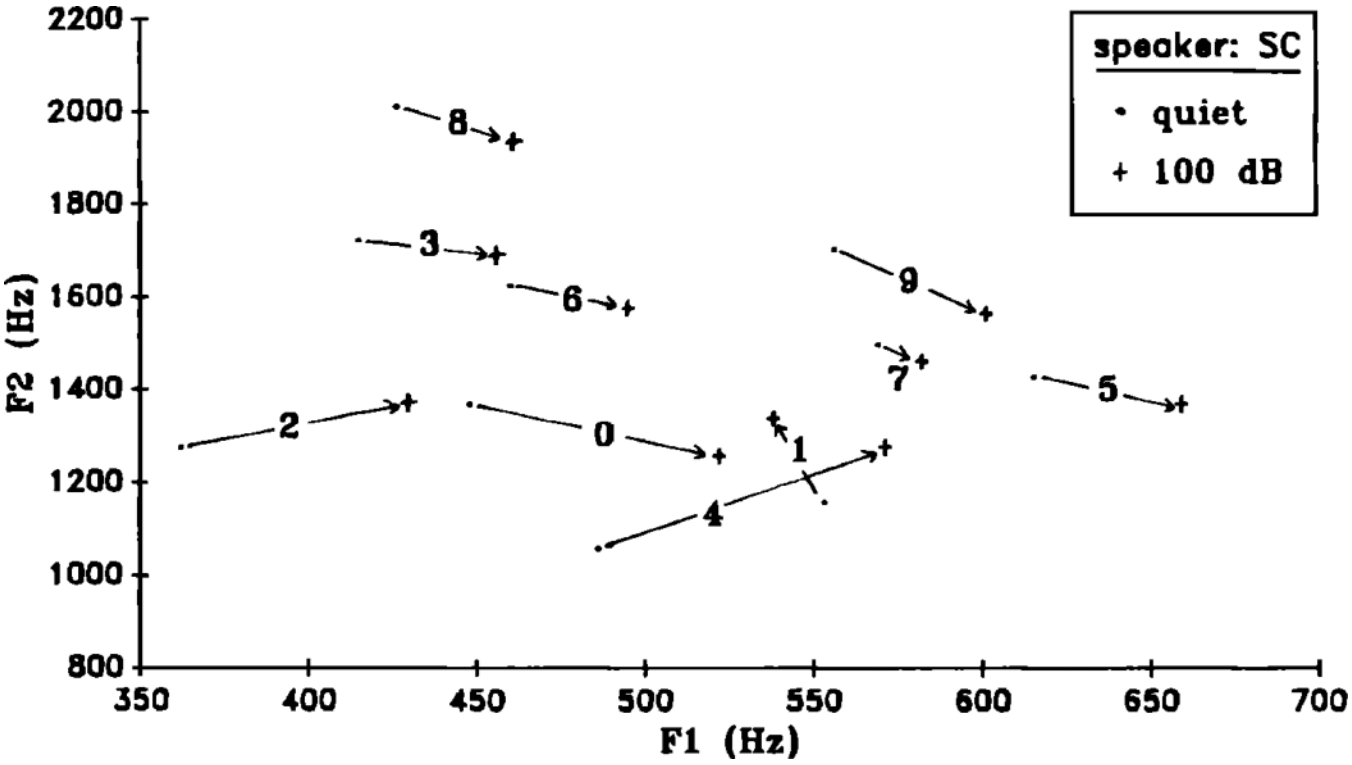

FIG. 6.
Mean first and second formant frequencies for words produced in quiet and 100 dB of masking noise by speaker SC. Values are presented separately for each utterance.

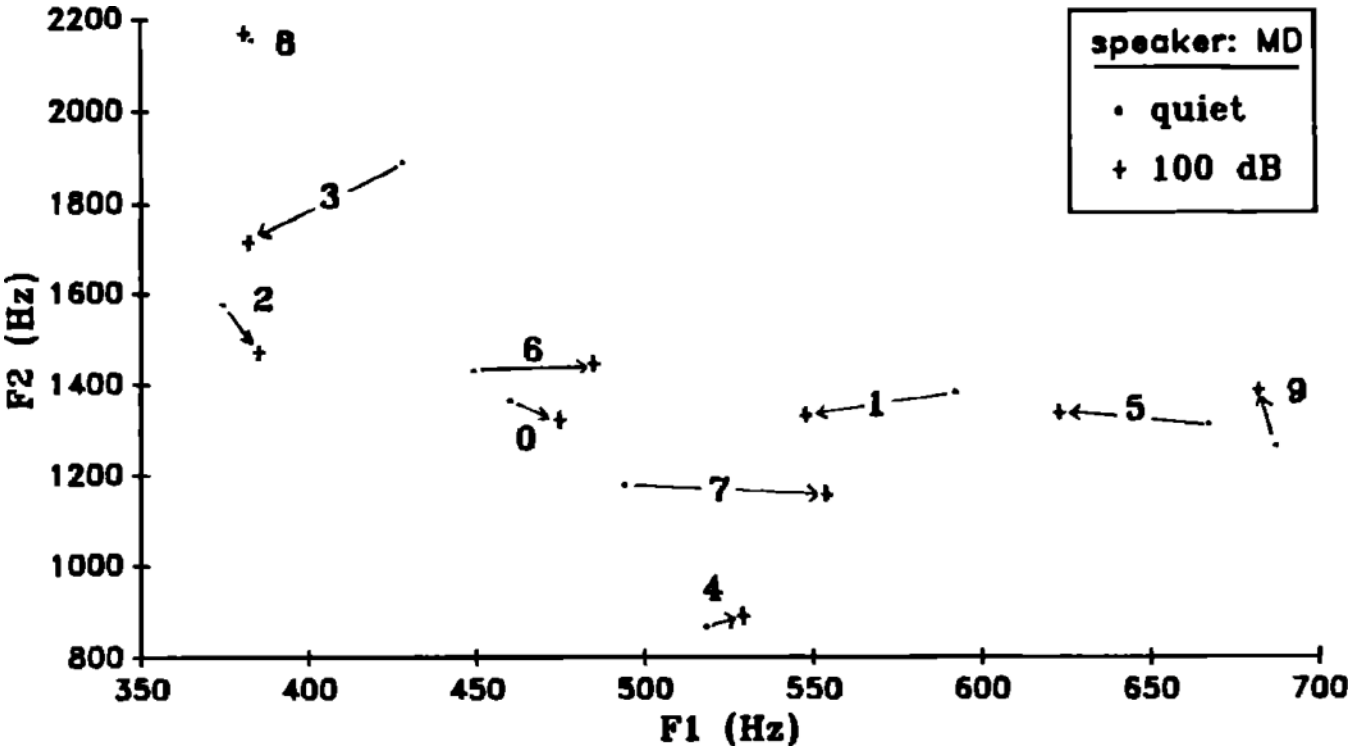

FIG. 7.
Mean first and second formant frequencies for words produced in quiet and 100 dB of masking noise by speaker MD. Values are presented separately for each utterance.

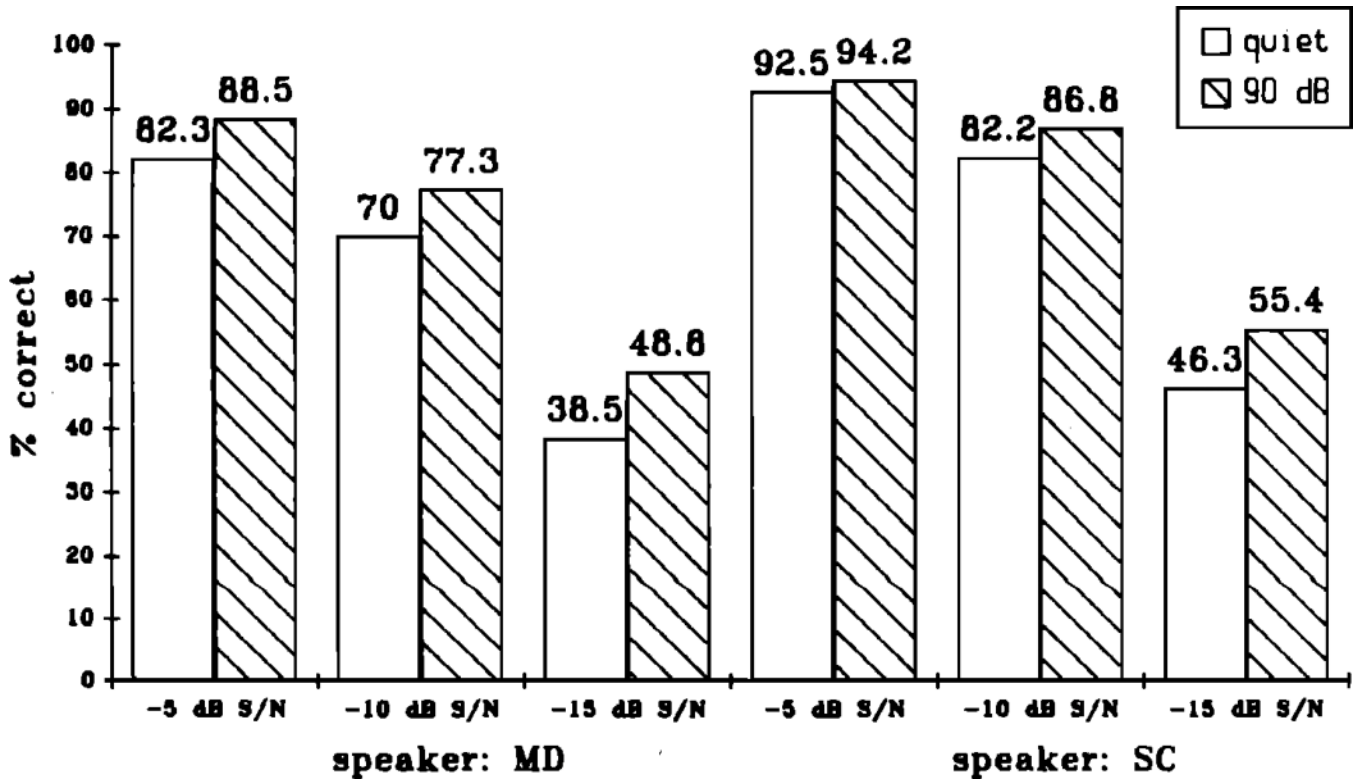

FIG. 8.
Intelligibility of words produced in quiet and 90 dB of masking noise (experiment I). Performance is broken down by S/N ratio and speaker.

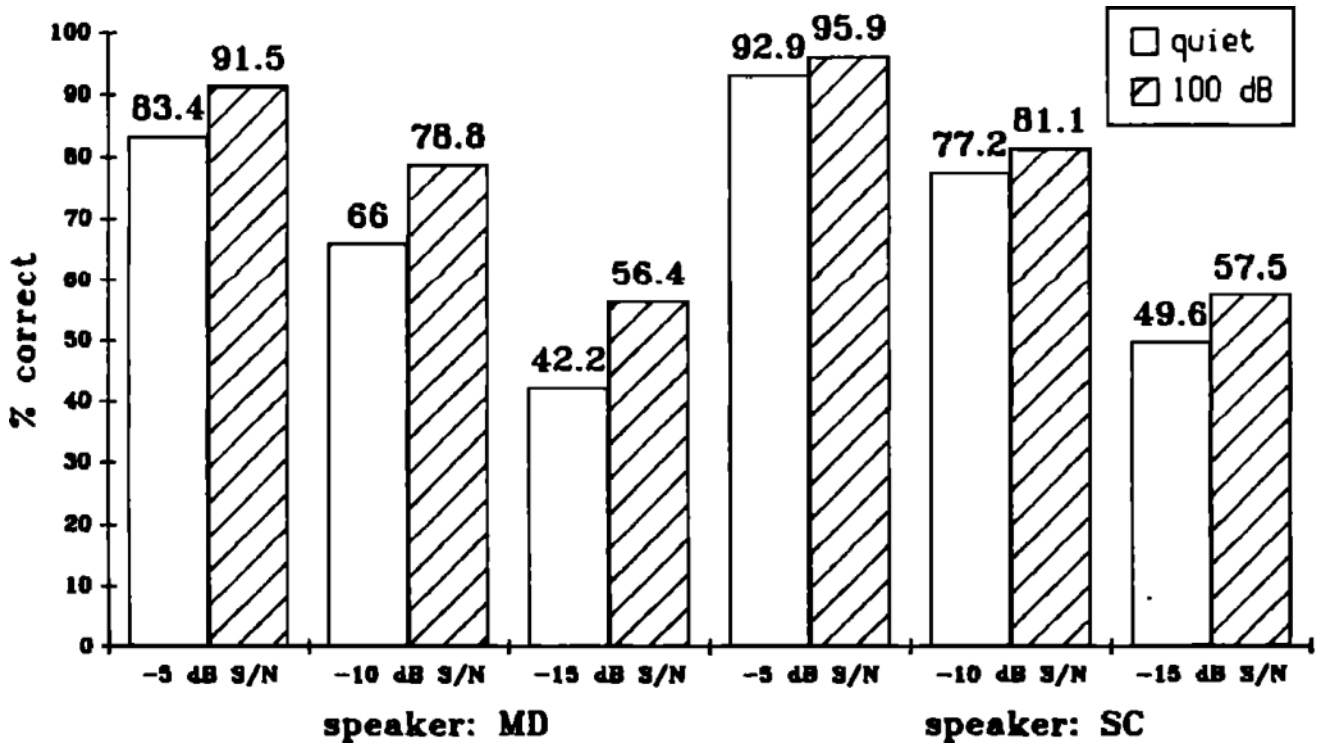

FIG. 9.
Intelligibility of words produced in quiet and 100 dB of masking noise (experiment II). Performance is broken down by S/N ratio and speaker.