Published in final edited form as:

J Acoust Soc Am. 1993 May ; 93(5): 2962–2973.

Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences

Scott E. Lively, David B. Pisoni, W. Van Summers, and Robert H. Bernacki

Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405

Abstract

The present investigation examined the effects of cognitive workload on speech production. Workload was manipulated by having talkers perform a compensatory visual tracking task while speaking test sentences of the form “Say hVd again.” Acoustic measurements were made to compare utterances produced under workload with the same utterances produced in a control condition. In the workload condition, some talkers produced utterances with increased amplitude and amplitude variability, decreased spectral tilt and $F0$ variability and increased speaking rate. No changes in $F1$, $F2$, or $F3$ were observed across conditions for any of the talkers. These findings indicate both laryngeal and sublaryngeal adjustments in articulation, as well as modifications in the absolute timing of articulatory gestures. The results of a perceptual identification experiment paralleled the acoustic measurements. Small but significant advantages in intelligibility were observed for utterances produced under workload for talkers who showed robust changes in speech production. Changes in amplitude and amplitude variability for utterances produced under workload appeared to be the major factor controlling intelligibility. The results of the present investigation support the assumptions of Lindblom's [“Explaining phonetic variation: A sketch of the H&theory,” in *Speech Production and Speech Modeling* (Kluwer Academic, The Netherlands, 1990)] H&H model: Talkers adapt their speech to suit the demands of the environment and these modifications are designed to maximize intelligibility.

INTRODUCTION

The present investigation examined the effects of cognitive workload on speech production. Cognitive workload is defined as the information processing load placed on the human operator while performing a particular task (Sheridan and Stassen, 1979). We equate information processing load with the amount of attention that must be directed to a task and assume that cognitive workload increases with task difficulty. It is well known that humans have only a limited capacity of attentional resources and that performance suffers when subjects are required to simultaneously engage in two attention-demanding tasks (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). In the present investigation, we divided talkers' attention between two equally important tasks: Subjects performed a compensatory visual tracking task while producing test utterances of the form “Say hVd again.” Acoustic measurements of the utterances produced under cognitive load were compared to measurements of tokens produced in a control condition. In another experiment, we examined the intelligibility of the utterances produced in each condition in order to determine how effectively talkers compensated for the increased workload.

Although a few previous studies have examined the effects of workload on the acoustic characteristics of speech, several important limitations of these investigations need to be addressed. First, workload has typically been confounded with psychological stress (Hecker *et al.*, 1970; Scherer, 1979; Tolkmitt and Scherer, 1986). For example, several studies have examined flight recorders recovered from aircraft crashes (Brenner *et al.*, 1979; Brenner *et*

al., 1985; Brenner and Shipp, 1988), while others have performed acoustic analyses on utterances collected from pilots and crews engaged in flight simulations (Jones, 1990). In another study, Streeter *et al.* (1983) analyzed recordings of phone conversations between a system operator and his superior just prior to the New York City blackout of 1977. In each of these conditions, speakers are under high levels of psychological stress, in addition to being under heavy cognitive workload. Moreover, it is difficult to quantify precisely how heavy each talker's workload was at the time the recording was made. In the present study, we attempted to minimize the effects of psychological stress by focusing subjects' attention on a compensatory tracking task that increased in difficulty over time. Unlike many other studies that have purported to study the effects of cognitive workload, we were able to quantify and manipulate the degree of workload that subjects were under when they produced the test utterances.

Second, the test environment and the set of utterances to be produced has typically not been controlled or manipulated in any systematic way in previous studies. Although a limited set of repeated utterances may be available from sources such as flight recorders (Brenner *et al.*, 1985), larger and more controlled speech samples can be collected in the laboratory environment. Hansen (1988) and Jones (1990) addressed this issue by collecting utterances from talkers in more constrained environments. Hansen (1988) limited his vocabulary to 35 words commonly used in aircraft communication and analyzed a number of acoustic parameters in his study on speech produced under simulated stress and workload. Jones (1990) also recorded multiple tokens of a small set of phrases from crews involved in flight simulations. In the present investigation, subjects participated in a very constrained, attention demanding task and produced numerous tokens of ten different vowels that were embedded in a carrier phrase. This procedure controlled for both the environmental conditions and the test vocabulary and permitted us to make reliable measurements that could be compared across different experimental conditions.

Finally, previous studies that have examined the effects of cognitive workload on speech production have focused on only a narrow range of acoustic parameters. Williams and Stevens (1969), Kuroda *et al.* (1976), and Tolkmitt and Scherer (1986) all found increases in fundamental frequency to be positively correlated with changes in talkers' psychological stress and workload levels. Changes in the period-to-period pitch variability or "jitter" in *F0* were also associated with increases in fundamental frequency. Brenner *et al.* (1985) analyzed the speech of pilots involved in crashes and found decreases in jitter in speech recorded during the emergency situation. Although changes in glottal source characteristics may provide a useful set of cues to assess the mental state of the talker, supralaryngeal adjustments may also occur under workload (Hansen, 1988). Changes in formant frequencies between workload and control conditions may provide additional information about the state of the talker. In the present study, we examined the acoustic correlates of both glottal and supralaryngeal adjustments, and changes in articulatory timing, as a function of cognitive workload.

In addition to improving the methodology of previous studies, it is important to place the present investigation in the context of a theoretical framework that relates speech production to speech perception. Lindblom (1990) has recently described a model that centers around the claim that talkers modify their speech in order to suit the demands of the environment. In his hypospeech and hyperspeech model (H&H model), Lindblom argues that the goal of the speaker is to maximize discriminability within the constraints of the communication environment. Speakers adjust their speech by using "signal-complementary processes" that balance the demands of "system-oriented control" and "output-oriented control." Signal-complementary processes are used to assess the listener's information that is independent of the speech signal. For example, speakers may monitor factors such as the listener's

background knowledge and the ambient listening conditions. Talkers modify their speech on the basis of these factors in order to increase the efficiency of their communication. System-oriented control refers to the demands that are put on the articulators during production, while output-oriented control refers to the goal-directed, communicative nature of speech.

According to the model, speakers attempt to balance the demands put on the articulatory system with the need to communicate efficiently under many different environmental conditions. Lindblom (1990) suggests that signal-complementary processes are dynamic and that they cause changes in the balance of system-oriented control and output-oriented control. For example, when output constraints are low (e.g., speaking in a quiet environment), speakers devote few resources to system-oriented control. Lindblom refers to this condition as “hypospeech” and characterizes it as an “economical” form of behavior. He suggests that articulators deviate little from neutral positions during hypospeech. Similarly, when the demands on output system control are high (e.g., discriminability and intelligibility must be maximized), then demands on system control are also high. Under these conditions, Lindblom suggests that subjects engage in hyperspeech. Lindblom argues that hyperspeech is characterized by “plasticity.” Articulators can be flexibly reorganized to meet the constraints of the production environment. According to the H&H model, speakers move along a continuum from hypospeech to hyperspeech when the demands of the communication environment change.

The relevance of the H&H model to speech produced in severe environments is clear. The goal of the speaker and the listener is to communicate quickly and efficiently. Under conditions of high cognitive workload, however, an additional constraint is added: Talkers must simultaneously perform an attention-demanding task while producing speech. According to Lindblom's model, conditions of high cognitive workload represent a situation in which output-oriented constraints are severe. This suggests that speakers should respond by increasing system-oriented controls in order to maintain intelligibility. The model makes two predictions under these conditions. First, utterances produced under workload should show adaptations to the workload environment, relative to a quiet control condition. These adaptations might include increases in amplitude in the workload condition. Talkers may also decrease the duration of their utterances in order to return their attention to the workload task. Second, these adaptations should enhance the intelligibility of utterances produced during workload, relative to utterances produced in a control condition. Within the framework of the H&H model, tokens produced under workload are an example of hyperspeech, whereas tokens produced in the control condition are an example of hypospeech.

In order to test the predictions of the H&H model, we conducted an experiment that examined both production and perception. In the production portion of the experiment, five talkers participated in a compensatory visual tracking task while producing speech. Acoustic analyses were then performed on these utterances and on utterances collected from the same subjects in a control condition. Next, we presented vowel tokens excised from utterances produced in the workload and control conditions to a panel of listeners for closed-set perceptual identification. According to the predictions of the H&H model, utterances produced under the workload condition should be more intelligible than utterances produced in the control condition. This increase in intelligibility reflects speakers' adaptations to the severe workload environment.

A. The compensatory tracking task

In the present experiment, we manipulated cognitive workload by using a compensatory visual tracking task that was originally described by Jex *et al.* (1966). The tracking task, referred to hereafter as the “JEX task,” involved manipulating a joystick in order to keep a

pointer centered between two boundaries on a CRT screen. The program controlling the task randomly deflected the pointer away from the center of the monitor and the subject was required to compensate for the movement by manipulating a joystick. The movement of the pointer is governed by the first-order unstable system $\dot{m} = \lambda m + K_c \lambda F$, where m is the position of the pointer along the horizontal plane, λ represents the degree of instability of the system, and K_c is a constant that mediates the degree of pointer movement. Here, F is a forcing function supplied by the subject that prevents the pointer from settling in the center position (Jex, 1979). Here, λ is used as the measure of workload in the present investigation and is measured in radians/second.

During training with the JEX task, λ was initially set a low level ($\lambda = 1.5$), indicating little attentional demand, and was increased until subjects lost control of the pointer and crashed it into a side boundary. After several days of practice, subjects' maximum λ levels were recorded. During testing, subjects performed the JEX task at 80% of their maximum level attained during training and also produced speech. We assumed that this high level of cognitive workload would be sufficient to induce robust acoustic and prosodic changes in the subjects' utterances, relative to a control condition, while still permitting the subject to carry out the task.

I. METHOD

A. Subjects

Five male native speakers of English were recruited as subjects. Three subjects (ME, TG, and EG) were psychology graduate students who were paid for their participation. Two subjects (MC and SL) were members of the laboratory staff who participated as part of their routine duties. All subjects were naive to the purpose of the study. None of them reported any history of hearing or speech impairments at the time of testing.

B. Procedure

Subjects were run individually in a single-walled sound-attenuated booth (IAC model 401A). The subject was seated comfortably facing a video screen interfaced to a Commodore 64 personal computer, which was used to control the JEX task and display the test utterances. The subject wore a headset fitted with an Electrovoice condenser microphone (model C090) attached to an adjustable boom. Once adjusted, the microphone remained at a fixed distance of 4 in. from the subject's lips. Subjects wore the headset continuously during training sessions on the JEX task, as well as during the experimental sessions in which the test utterances were collected.

Subjects were trained on the JEX task for several days. When a subject was able to consistently perform the task at a fairly high level of difficulty ($\lambda > 3.0$), the production component was introduced and recordings of the speech samples were made. Subjects were given one block of practice trials with $\lambda = 1.5$ at the beginning of each experimental session. Following the practice block, the difficulty of the JEX task was increased to 80% of the level attained by the subject during training without the speech production component. The difficulty of the JEX task was adjusted between blocks when subjects were either unable to successfully perform both tasks or when subjects did not crash the pointer into the boundaries during an entire block. The λ levels for each subject in each block of JEX trials are displayed in Appendix A. The scale constant (K_c) was held at 3.0 throughout each experimental session.

During the experiment, subjects simultaneously performed the JEX task and repeated the test phrases presented on the video screen. All experimental materials consisted of /h/-vowel-/d/ utterances embedded in the sentence frame: "Say hVd again." The English

vowels /æ ɔ̃ ɛ i ɪ u ʌ ə ər/ appeared in the hVd context. Test phrases were only presented while the subject was performing the JEX task at the 80% level of maximum difficulty. Test phrases were presented every 5 s. If the subject lost control of the JEX pointer, λ was reduced to 1.5 and then gradually increased to the 80% level. No sentences were presented until subjects had again reached the 80% level.

For each subject, speech samples were collected in two experimental sessions. Sessions consisted of four blocks of 20 trials. The ten phrases were presented twice within each block. Mispronounced phrases were presented until the subject correctly produced the token. Blocks of trials alternated between JEX and control conditions. During JEX blocks, subjects performed the tracking task and produced the test utterances displayed on the screen. During control blocks, subjects repeated the test utterances without performing the tracking task. For each subject, a total of eight tokens of each phrase was produced in each condition.¹

C. Speech signal processing

Test utterances were sampled and digitized on-line by a VAX 11/750 computer during all experimental sessions. Productions 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 sampled into a separate waveform file and then analyzed using digital signal processing techniques.

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 then used to calculate the short-term spectrum and overall power level of each analysis frame. Formant frequencies, bandwidths, and amplitudes were also calculated for each frame from the LPC coefficients (Markel and Gray, 1976). In addition, a pitch extraction algorithm (Noll, 1968) was employed to determine if a given frame was voiced or unvoiced and, for the voiced frames, to estimate the fundamental frequency (F_0).

Total duration for each phrase was determined by visual inspection of a CRT display that simultaneously presented the speech waveform and 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 carrier phrase. The onset and offset of the /h/ frication, vowel, and /d/ closure segments from the hVd portion of each utterance were also identified and labeled. Following identification of phrase and segment boundaries, duration and rms energy for each phrase and segment were saved in a data file. Fundamental frequency and formant frequency information were also stored for the phrase and for the vowel of the hVd portion of the utterance.

II. RESULTS AND DISCUSSION

The effects of cognitive workload on several acoustic measurements of the test utterances are summarized below. In each case, an analysis of variance was used to determine whether workload had a significant effect on a given acoustic measure. Separate analyses were carried out for each talker. Test utterance and workload condition (JEX or control) were the independent variables in each analysis. The presentation of results will focus on the effect of workload on the various acoustic measures. The test utterance variable will be ignored because no significant interactions of workload with test utterance type were obtained.

A. Amplitude

The upper panel of Fig. 1 shows amplitude averaged over entire phrases for utterances from the JEX and control conditions. The data are plotted separately for each talker. The lower

panel of the figure shows the amplitude of each of the segments comprising the hVd context. Amplitudes of the /h/ frication, vowel, and /d/ closure portion of each hVd utterance are shown for the five talkers in each workload condition. An asterisk above a pair of bars indicates a significant difference between values in the JEX and control conditions for a particular talker.

As shown in Fig. 1, amplitude was greater in the workload condition than the control condition for three of the five talkers. The pattern is very consistent for talkers ME, SL, and TG: Each had significantly higher amplitudes in the JEX condition for the entire phrase, as well as for the /h/, hVd vowel, and /d/ closure segments.² Talker MC showed significantly higher amplitudes in the JEX condition for the vowel and the /d/ closure segments, but also demonstrated higher amplitude in the control condition for the /h/ segment [$F_{/h/}(1,113) = 11.39, p < 0.01$; $F_v(1,113) = 4.72, p < 0.05$, $F_{/d/}(1,113) = 24.70, p < 0.01$]. Talker EG did not demonstrate a reliable increase in amplitude during workload productions at the phrase level. However, an increase in amplitude was observed in the control utterances for the /h/ segment of talker EG's utterances [$F(1,60) = 5.89, p < 0.05$].

B. Amplitude variability (across utterances)

The increase in mean amplitude was also accompanied by an increase in variability from one utterance to the next. Figure 2 shows the standard deviations of phrase and segmental amplitudes across utterances for each talker in each condition. Over the entire phrase, four of the five talkers demonstrated significant changes in amplitude variability between the control and the workload conditions. Of these four talkers, three showed increases in variability from the control condition to the workload condition [$t_{MC}(9) = 2.74, p < 0.05$; $t_{ME}(9) = 3.76, p < 0.01$; $t_{SL}(9) = 3.04, p < 0.05$]. The fourth talker produced the opposite pattern: Less amplitude variability was present in the workload condition than in the control condition [$t_{TG}(9) = -2.29, p < 0.05$]. Talker EG, who did not show any changes in amplitude across conditions, did not demonstrate any reliable changes in amplitude variability either.

As the lower panel of Fig. 2 shows, the pattern described for the phrase data was replicated in the amplitude variability of the vowels and /d/ closure segments in the hVd contexts. In each case, three of the five talkers had greater amplitude variability in the workload condition. For the vowels, the effect was statistically significant for each of these talkers [$t_{MC}(9) = 3.67, p < 0.01$; $t_{ME}(9) = 3.42, p < 0.01$, $t_{SL}(9) = 2.34, p < 0.05$]. For /d/ closures, however, the effect was significant only for one talker [$t_{ME}(9) = 4.86, p < 0.01$]. Talker TG demonstrated more variability in the control condition than in the workload condition for both the /h/ and vowel segments [$t_{/h/}(9) = -3.71, p < 0.01$; $t_v(9) = -3.52, p < 0.01$]. This pattern replicated his data from the entire phrase.

C. Spectral tilt

Increases in vocal amplitude are often correlated with changes in “spectral tilt.” High-amplitude utterances generally have flatter spectra. More high-frequency energy is present in high-amplitude utterances than in low-amplitude utterances. We examined the long-term spectra of the vowels in hVd contexts in each condition to determine if the increases in amplitude observed in the JEX task were also correlated with changes in spectral tilt. Figure 3 shows the differences in energy between vowels from the workload and control conditions across 40-Hz linear frequency bands.

A positive slope in these figures indicates that the difference in energy between workload and control conditions increases with frequency: Vowels in the workload condition have more high-frequency energy than the same vowels in the control condition. As in the amplitude results, four of the five subjects showed a consistent pattern (talkers EG, MC,

ME, SL). However, the four talkers who demonstrated changes in spectral tilt across conditions were not the same four who showed differences in amplitude in the previous analysis. Subject EG, who did not show any significant differences in amplitude between workload and control conditions, showed one of the clearest cases of changes in spectral tilt.

Overall, then, the data suggest that the workload task produces effects on spectral tilt that are not always correlated with changes in overall amplitude. Hansen (1988) has provided additional evidence that decreases in spectral tilt under workload are not necessarily linked to increases in amplitude. For both the JEX task and the dual task examined by Hansen (1988), spectral tilt decreased under workload without any associated change in amplitude.

D. Fundamental frequency

Following other researchers who have measured the effects of workload on speech (Brenner *et al.*, 1985; Hansen, 1988; Jones, 1990), we also analyzed the fundamental frequency of the phrases and the vowels in hVd contexts collected in each condition. Figure 4 shows mean F_0 values for each talker for the phrase and vowels in hVd contexts in each condition. Only talker MC increased F_0 during the JEX task for the entire carrier phrase [$F(1,113)=12.09$, $p<0.01$]. Two talkers showed an increase in F_0 for the vowels in the hVd context [$F_{MC}(1,113)=12.10$, $p<0.01$; $F_{SL}(1,134)=17.34$, $p<0.01$].

The absence of a consistent effect of workload on mean F_0 values across subjects was also reported by Hansen (1988). He examined speech produced while performing the JEX task and speech produced while performing a dual task requiring the performance of two simultaneous tracking operations. Neither of these workload conditions produced consistent effects on mean F_0 . In contrast, previous studies examining the effects of emotional stress on speech production have generally reported increases in mean F_0 when subjects experience increased levels of psychological stress (Williams and Stevens, 1969; Kuroda *et al.*, 1976).

E. Fundamental frequency variability (within utterances)

Although consistent changes in mean F_0 were not observed in the workload condition for our talkers, a different attribute of the F_0 data did show a fairly consistent pattern. Figure 5 shows the standard deviations of the frame-by-frame F_0 values of each phrase and the vowel from the hVd context in the JEX and control conditions. As shown in the figure, three subjects demonstrated a significant reduction in F_0 variability over the entire phrase when performing the JEX task [$F_{EG}(1,60)=4.58$, $p<0.05$; $F_{MC}(1,113)$, $p<0.01$; $F_{TG}(1,138)=6.51$, $p<0.05$]. One other subject (ME) displayed the same general pattern, although the differences were not statistically significant. For the vowels in the hVd contexts, only one speaker demonstrated a reduction in F_0 variability between the control and workload conditions [$F_{MC}(1,113)=32.39$, $p<0.01$].

Given that F_0 variability decreased for the entire phrase, but not for the vowel in the hVd context, the pattern is more likely due to a flattening of the overall F_0 contour. Rather than decreasing period-to-period F_0 variability or “vocal jitter,” talkers under workload apparently produce the whole phrase using a monotone pitch. Brenner *et al.* (1979) reported a similar decrease in frequency modulation when subjects were required to perform a speeded arithmetic task. This change in F_0 modulation could be responsible for the decrease in F_0 variability reported here.

F. Duration

Figure 6 shows the effect of cognitive workload on phrase and segmental durations. Four of the five talkers had significantly shorter overall phrase durations under workload

$[F_{EG}(1,60)=13.20, p<0.01; F_{MC}(1,113)=104.55, p<0.01; F_{ME}(1,140)=8.69, p<0.05; F_{TG}(1,138)=45.24, p<0.01]$. One talker demonstrated the opposite pattern $[F_{SL}(1,134)=5.37, p<0.005]$. This talker also produced the smallest change in phrase duration across conditions.

Segmental durations also decreased while performing the JEX task for some subjects. Two talkers produced significantly shorter /h/ segments in the workload condition than in the control condition $[F_{EG}(1,60)=13.20, p<0.01; F_{MC}(1,113)=91.49, p<0.01]$. Here /d/ closures were also significantly shorter for three of the five talkers $[F_{MC}(1,113)=32.38, p<0.01; F_{ME}(1,140)=50.20, p<0.01; F_{TG}(1,138)=5.37, p<0.05]$. The decrease in duration observed for the entire phrase, the /h/ frication, and the /d/ closure replicate results described by Hecker *et al.* (1968). Vowel duration in the hVd context was affected less consistently by the workload condition. Two talkers produced shorter vowels in the workload condition $[F_{MC}(1,113)=27.12, p<0.01; F_{TG}(1,138)=5.80, p<0.05]$ and one produced significantly longer vowels $[F_{SL}(1,134)=10.31, p<0.01]$.

G. Formant frequencies

Workload did not produce a significant change in the frequencies or bandwidths of the first three formants of the vowels produced by any of the five talkers. Thus it appears that the effects of increased workload have their primary locus on the subglottal and laryngeal source-related functions and speech timing rather than on the supralaryngeal control of speech production. While we observed an increase in speaking rate, we did not find a change in formant frequencies which might have suggested vowel reduction. Had the relative timing of the articulators changed, as opposed to the absolute timing, we might have observed a reduction of vowel targets that is typically associated with increases in speaking rate (Lindblom, 1963).

III. PERCEPTUAL IDENTIFICATION

Taken together, the results of the acoustic analyses indicated that talkers produced utterances under workload that were greater in amplitude and amplitude variability and shorter in duration than utterances produced in the control condition. It should be noted, however, that changes were talker-specific: We did not obtain consistent results across all five talkers for any of the variables that we examined.

Changes in amplitude and duration represent adaptations to the workload environment that are consistent with Lindblom's (1990) H&H model. Recall that the model assumed that speakers attempt to balance the demands of the production environment and the articulators in order to maximize discriminability. In the case of mental workload, subjects appeared to have shortened the duration of their utterances in order to devote more attention to the compensatory tracking task. Changes in amplitude may represent adjustments designed to maximize intelligibility.

We tested this prediction concerning amplitude in a perceptual identification experiment. Vowel tokens were excised from the hVd portion of each phrase in the workload and control conditions and were presented to a panel of listeners for closed-set identification in noise. Signal-to-noise ratio was manipulated as a between-subjects variable and workload was manipulated as a within-subjects variable. In addition, we manipulated rms amplitude as a between-subjects variable: Listeners in one condition heard vowels that were equated for amplitude, while listeners in another condition heard vowels that were not equated for rms amplitude. If speakers attempt to maximize intelligibility under workload conditions by increasing amplitude, then we would expect that utterances produced in the JEX condition should be more intelligible than utterances produced in the control condition. However, if

we equate all tokens for amplitude, the advantage in intelligibility for tokens produced under workload should be reduced or eliminated.

IV. METHOD

A. Subjects

Subjects who served as listeners in the perceptual identification experiment were 76 native speakers of English. All were undergraduates enrolled in an introductory psychology course at Indiana University and were given course credit for their participation. No subjects reported any history of a speech or hearing impairment at the time of testing.

B. Stimuli

400 vowel tokens were excised from the “Say HVD again” utterances described above. Each of the five talkers contributed 80 stimuli for use in the perceptual identification experiment (10 vowels [æ ɔ ɛ i ɪ u ɑ ʌ ə ər] × 4 tokens of each vowel × 2 workload conditions). This set of 400 tokens was then duplicated. The duplicate set of tokens was equated for rms amplitude using signal processing software running a PDP 11/34 laboratory computer. Thus, two sets of tokens were prepared for each talker: One set was equated for amplitude, while the other set was left in its original format.

C. Procedure

Subjects were tested in groups of 5 or fewer. Each subject was seated in a small, sound-attenuated cubicle, equipped with TDH-39 matched and calibrated headphones and an ADM-5 CRT terminal. The terminal was interfaced to a PDP 11/34 laboratory computer that controlled stimulus presentation and collected subjects' responses.

Subjects participated in a ten alternative forced-choice identification task. On each trial, a visual prompt was presented on the CRT monitor to alert the subject to the upcoming stimulus. Approximately 500 ms after the visual prompt was displayed, a randomly selected auditory stimulus was presented. The stimulus was masked by randomly generated broadband noise (Grason-Stadler 1724 Noise Generator). The masking signal began 100 ms prior to the onset of the stimulus and ended 100 ms after the offset of the stimulus. Subjects recorded their responses by pressing the appropriately labeled keys on the keyboard, which were labeled with the corresponding hVd word (had, hawed, head, heed, heard, hid, hod, hood, hud, who'd). Subjects were informed at the beginning of the experiment that they would be hearing only the vowel segments of the words, and that they were to use the words on the ten keys to record their responses. A single experimental session consisted of 400 trials. A 2-min break was given after the first 200 trials. Each experimental session lasted approximately 45 min.

Vowel type, talker, and workload condition were manipulated as within-subjects variables. Signal-to-noise ratio was manipulated as a between-subjects variable. One-third of the subjects were presented with tokens at a 0-dB S/N, one-third at -5 dB, and one-third at -10 dB. Signal and noise levels were originally set to 80 dB SPL. Signal levels were then attenuated via a set of programmable attenuators. Signal level was calibrated using a control token of the vowel /ɑ/ from talker EG. Amplitude was also manipulated as a between-subjects variable: Half of the subjects heard tokens that were equated for rms amplitude, while the other half heard tokens that were not equated for amplitude.

V. RESULTS AND DISCUSSION

In order to facilitate the presentation of the results, only main effects and interactions involving workload will be reported. To permit comparisons with the earlier acoustic analysis, a separate analysis of variance was conducted on the accuracy data collected in the perceptual identification experiment for each talker. Signal-to-noise ratio and amplitude condition were treated as between-subjects factors in each analysis. Workload condition and vowel type were treated as within-subjects factors. Significant interactions were analyzed using Tukey's HSD tests.

The main effect of most interest is workload. Figure 7 shows intelligibility as a function of talker and workload condition. Intelligibility increased 2%–5% for three of the five talkers from the control condition to the workload condition [$F_{MC}(1,70)=28.02, p<0.01$; $F_{ME}(1,70)=8.63, p<0.01$; $F_{SL}(1,70)=12.39, p<0.01$]. For talker TG, vowels produced in the control condition were more intelligible than vowels produced in the workload condition [$F(1,70)=11.13, p<0.01$].

The three talkers who demonstrated more intelligible utterances in the workload condition also showed reliable increases in the amplitude and amplitude variability of the hVd vowel segments produced under workload. Talker TG, who produced more intelligible vowels in the control condition, had higher amplitudes in the workload condition, but demonstrated significantly more amplitude variability in the control condition.

The advantage in intelligibility for tokens produced under workload can be localized by examining the effects of equating all of the tokens for amplitude. Tokens from the workload condition that were not equated for amplitude tended to be presented at a level above the desired signal-to-noise ratio, while tokens from the control condition were presented at a level below the target level. However, when tokens were equated for amplitude, the variability across conditions was removed and all tokens were presented at the same signal-to-noise ratio. The argument that changes in segmental amplitude were responsible for changes in intelligibility is supported by the interaction of workload condition with amplitude equalization condition. Three of our five talkers demonstrated this interaction [$F_{MC}(1,70)=6.27, p<0.05$; $F_{ME}(1,70)=4.64, p<0.05$; $F_{SL}(1,70)=5.82, p<0.05$]. Although Tukey's tests failed to localize the source of the interactions for any of the talkers, the overall pattern is clear. As shown in the top panel of Fig. 8, large differences in intelligibility were observed across workload conditions for tokens that were not equated for amplitude. As shown in the bottom panel, the differences in intelligibility between workload conditions were reduced for the items that were equated for amplitude. These findings suggest that the differences in amplitude and amplitude variability play an important role in controlling intelligibility in the workload condition.

Although an overall advantage in intelligibility was observed for tokens produced under workload, the effect was not very consistent for individual vowels across talkers. A significant interaction of workload condition with vowel type was observed for each of our talkers [$F_{EG}(9,630)=2.30, p<0.05$; $F_{MC}(9,630)=5.09, p<0.01$; $F_{ME}(9,630)=3.07, p<0.01$; $F_{SL}(9,630)=3.29, p<0.01$; $F_{TG}(9,630)=3.57, p<0.01$]. For talkers MC, ME, and SL, significant increases in perceptual identification accuracy were found in the workload condition for /ɛ ə/. Significant decreases in intelligibility were observed for /ɔ/ in the workload condition for talkers MC, ME, and TG. Increases in intelligibility were observed for /ʌ/ for MC and ME, but a decrease was found for TG. Less consistent changes were obtained for /æ i/. No changes in intelligibility were observed for any of the talkers for the vowel /u/. This pattern of perceptual data does not map on to the acoustic data as well as the

pattern observed for the main effect of workload. Recall that in the acoustic data, we failed to find any reliable interactions of workload and vowel type.³

VI. GENERAL DISCUSSION

In the Introduction, we outlined two important aspects of Lindblom's (1990) H&H model. First, it was assumed that speakers modify their speech to suit the demands of the production environment. When output constraints are high (e.g., conditions of high cognitive workload), speakers increase demands on the articulators. Similarly, when few constraints are placed on the output (e.g., speech produced in a quiet environment), speakers relax the demands on the articulators. Second, it was assumed that the purpose of these modifications was to maximize the discriminability of utterances, within the constraints of the production environment.

We tested these assumptions in the present experiment by placing subjects under cognitive workload while having them produce test utterances. We assumed that performing a compensatory visual tracking task while speaking would increase the demands on the output system and would lead to quantifiable acoustic and prosodic changes in the speech signal. Furthermore, we assumed that if speakers did modify their speech under workload, these adaptations would enhance speech intelligibility.

The results of the present study were consistent with the predictions of the H&H model. Talkers tended to produce shorter utterances with higher amplitude, more amplitude variability and flatter spectra in the workload condition. Inconsistent changes in the variability of *F0* were also observed. In a subsequent perceptual identification experiment, we found that vowel tokens produced in the workload condition were more intelligible than utterances produced in the control condition for speakers who demonstrated increased amplitude and amplitude variability during production under workload. Thus the present results indicate that talkers did modify their speech under workload and that these changes led to increases in intelligibility.

Despite the support our findings give to the H&H model, several important limitations of the present study should be acknowledged. First, the acoustic analyses demonstrated a great deal of intersubject variability. Although consistent trends were observed for some variables, in many cases only one or two subjects demonstrated significant changes from the control condition to the workload condition. This observation suggests that adaptations to the workload environment may be idiosyncratic and that more consistent changes may be observed only under more extreme conditions.⁴

Second, the magnitude of the increase in intelligibility between conditions needs to be addressed. In the present experiment, we observed a 2%–5% increase in intelligibility for utterances produced in the workload environment. Several factors may have contributed to this small increase. First, talkers were never informed that the intelligibility of their utterances would be assessed. More dramatic acoustic changes might have occurred, had subjects been attempting to communicate with a human listener (Lane and Tranel, 1971; Lane *et al.*, 1970). Second, we collected productions in a very benign environment. If the demands of the workload task had been greater, or if there had been more serious consequences of failure, greater acoustic changes and differences in intelligibility might have been observed. Third, we tested intelligibility using only isolated vowels. It is well known that identification accuracy of steady-state vowels is poor (Strange *et al.*, 1976; Strange *et al.*, 1979). If we had used different stimulus materials, such as CVCs or words, we might have observed larger differences in intelligibility between the workload and the control conditions.

Despite these limitations, the present study supports the assumptions of Lindblom's H&H model and also replicates findings from studies that have been conducted on speech in other types of severe environments. The present acoustic and perceptual results compare favorably with findings from studies that have examined speech produced under psychological stress, speech produced in noise, speech produced under instructions to speak clearly, and shouted speech.

As mentioned in the Introduction, workload and psychological stress have often been equated in studies that have examined speech produced in severe environments. The attentional demands placed on speakers under workload and psychological stress are similar: Typically, speakers must perform an attention demanding task while simultaneously producing speech. The acoustic correlates of psychological stress are also similar to those of speech produced under cognitive workload. In general, speech produced under psychological stress tends to be generated with higher amplitude and higher *F0* and less variability in the fundamental frequency (Brenner and Shipp, 1988; Brenner *et al.*, 1985). However, the effects of psychological stress on duration are inconsistent within and across studies. For example, Streeter *et al.* (1983) reported that one of their speakers increased durations under stress, while a second talker decreased durations (see, also, Brenner and Shipp, 1988).

The results of the present study are similar to findings reported from studies examining the acoustic correlates of psychological stress. The strong similarity between the present findings and results concerning the effects of stress suggests that one component of workload may, in fact, be increased psychological stress and that workload and stress may be very difficult to unconfound experimentally in the laboratory (Sheridan and Stassen, 1979). However, it should be noted that workload is more easily quantified than psychological stress when defined in terms of attentional demands. Psychological stress may be inferred from physiological measures or from self-report forms. In contrast, workload can be quantified in terms of the attentional demands of the subjects' task.

Several studies have also examined the effects of noise in the environment on the acoustic correlates of speech production. It is well known that speakers will increase their vocal effort in the presence of background noise (Lane and Tranel, 1971; Hanley and Steer, 1949; Lane *et al.*, 1970). This effect, known as the Lombard reflex, has been attributed to an attempt on the part of the speaker to maintain a constant level of intelligibility, despite degradation of the signal by environmental noise. The Lombard reflex is characterized by several consistent acoustic and prosodic changes in the speech signal. First, utterances produced in noise are longer in duration than utterances produced in quiet conditions (Dreher and O'Neill, 1957; Pisoni *et al.*, 1985). Second, speakers increase overall amplitude in the presence of background noise (Dreher and O'Neill, 1957). Third, fundamental frequency increases under noisy conditions. Finally, speech produced in noise tends to exhibit flatter spectra with more high-frequency energy than speech produced in a quiet environment. Perceptually, speech produced in noise also tends to be more intelligible than speech produced in a quiet environment (Summers *et al.*, 1988).

The differences between the acoustic correlates of speech produced in noise and speech produced under workload conditions may be due in part to the differing attentional demands of the two conditions. Under workload conditions, speakers are simultaneously engaged in an attention-demanding task, in addition to producing speech. When talkers speak in noise, however, the attentional focus of the speaker is on communicating information to a listener. As a consequence of these differences in task demands and attentional focus, talkers appear to produce utterances in the two conditions with differing acoustic properties. Whereas subjects in a workload experiment may only increase amplitude in order to increase

intelligibility, speakers in Lombard experiments produce utterances with longer durations, increased amplitude, and decreased spectral tilt.

The present findings also are relevant to recent investigations that have examined “clear speech” (Chen, 1980; Lindblom *et al.*, submitted; Pichney *et al.*, 1985; Pichney *et al.*, 1986). In these experiments, subjects are given explicit instructions to “speak as clearly as possible.” Speech produced under these conditions is typically compared to speech produced under normal conversational instructions. Clear speech shows several robust acoustic and prosodic changes, relative to conversational speech: Clear speech tends to be produced with higher amplitude, increased $F0$, and longer durations. Furthermore, the vowel space is enlarged by moving $F1$ and $F2$ closer to their target values (Chen, 1980). Not surprisingly, speech produced under instructions to speak clearly also tends to be more intelligible than conversational speech (Lindblom *et al.*, submitted; Pichney *et al.*, 1985, 1986).

The attentional demands of speech produced under instructions to speak clearly and speech produced under workload differ greatly. In experiments that examined the acoustic correlates of clear speech, subjects were explicitly instructed to speak in a manner that would maximize intelligibility. Attention is not concurrently focused on a secondary task in this condition. In contrast, speakers under workload must simultaneously attend to two equally important tasks. As a consequence the two types of speech show different types of adaptations. In the case of clear speech, intelligibility may be enhanced by increasing the size of the vowel space, lengthening the duration of utterances and increasing their amplitude. For speech produced under workload, speakers increase intelligibility by increasing amplitude and amplitude variability.

In each of the cases we have described above, an increase in amplitude has been observed. Another form of speech that also demonstrates increased amplitude is shouted speech. Several other acoustic changes are also apparent in shouted speech. For example, $F0$ and vowel duration tend to increase, while spectral tilt tends to decrease (Rostoland, 1982 a,b). In contrast to other types of speech produced in severe environments, the intelligibility of shouted speech decreases relative to utterances produced in a conversational manner (Pickett, 1956; Pickett and Pollack, 1958; Rostoland, 1985). Acoustic and temporal cues may change the speech signal in such a way that important perceptual cues are lost or distorted, rather than enhanced (Rostoland 1982a,b).

Within the context of Lindblom's (1990) H&H model, each of the cases of speech produced in severe environments represents a change from hypospeech to hyper-speech. When the constraints of the production environment are increased due to workload, stress, noise, or special instructions, talkers tend to increase the demands on their articulators. In general, these demands are satisfied by increases in amplitude and changes in $F0$ and duration. Table I summarizes the acoustic correlates of a number of types of speech produced in severe environments. As this table makes clear, consistent changes across studies are often difficult to find. This suggests that talkers adapt to severe environments in idiosyncratic ways. The manner in which speakers change from a hypospeech mode to a hyperspeech mode is open to a great deal of individual variability and warrants further investigation.

Although a great deal of individual variability is observed in the acoustic analyses of speech produced in severe environments, a consistent trend was observed for increases in intelligibility, relative to normal conversational speech. These findings are consistent with Lindblom's assumption that talkers change their speech in response to the environment in order to maximize discriminability.

In summary, the results of the present investigation demonstrate a number of changes in the acoustic-phonetic properties of speech produced under cognitive workload. These acoustic

changes have observable perceptual consequences for listeners. The present findings add to a growing body of literature showing that talkers modify their speech in response to both physical and cognitive demands of their immediate environments. Although some changes are produced consistently across talkers, many changes appear to be talker-specific. However, in each case, the effect of these adaptations appears to increase the intelligibility of speech produced in severe environments. These findings are consistent with Lindblom's (1990) H&H Model, which assumes that talkers adapt their speech to suit the demands of the environment and that these modifications are designed to maximize discriminability and intelligibility.

Acknowledgments

This work was supported by a contract from Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, Contract No. AF-F-33615-86-C-0549 to Indiana University, Bloomington, Indiana. The authors wish to thank Rich Lippman at MIT Lincoln Laboratories for his technical assistance. They also thank Robert Porter and an anonymous reviewer for their helpful comments on a previous version of this report. Send correspondence to David B. Pisoni, Speech Research Laboratory, Indiana University, Bloomington, IN 47405.

APPENDIX

The JEX λ levels during each block of the workload condition are shown in Table AI.

TABLE AI

	JEX 1	JEX 2	JEX 3	JEX 4
EG	...	...	3.45	3.45
MG	3.45	3.45	3.45	3.45
ME	3.50	3.50	3.50	3.50
SL	3.34	3.34	3.34	3.34
TG	3.20	3.00	3.20	3.10

¹For talker EG, we report data from experimental session 2 only. The level of JEX task difficulty used in experimental session 1 was apparently too low for this subject. He performed the task without crashing the pointer into the outer boundaries and did not appear to be under any workload. JEX task difficulty was increased in session 2. With this increase in difficulty, performance on the JEX task was similar to the performance of the other subjects. The exclusion of the data from session 1 for EG left 4 tokens of each utterance available from each condition.

²Phrase amplitude: [$F_{ME}(1,140)=22.15, p<0.01$; $F_{SL}(1,134)=11.15, p<0.01$; $F_{TG}(1,138)=6.09, p<0.05$].

/h/ amplitude: [$F_{ME}(1,140)=18.96, p<0.01$; $F_{SL}(1,134)=5.11, p<0.05$; $F_{TG}(1,138)=11.81, p<0.01$].

V amplitude: [$F_{ME}(1,140)=29.66, p<0.01$; $F_{SL}(1,134)=15.54, p<0.01$; $F_{TG}(1,138)=7.99, p<0.01$].

Closure amplitude: [$F_{ME}(1,140)=35.39, p<0.01$; $F_{SL}(1,134)=3.91, p<0.05$; $F_{TG}(1,138)=25.05, p<0.01$].

³In addition to the main effect for workload on intelligibility and the consistent interactions of workload with amplitude condition and workload with vowel, we also observed several interactions that were not consistent across talkers. First, workload interacted with signal-to-noise ratio in the perceptual identification data for two talkers [$F_{MC}(2,70)=5.09, p<0.01$; $F_{TG}(2,70)=8.85, p<0.01$]. Intelligibility decreased as a function of signal-to-noise ratio for both talkers. Small increases in the magnitude of the workload effect were observed at the -5 dB signal-to-noise ratio, relative to the 0- and -10-dB conditions. Talker MC showed an advantage for tokens produced in the workload condition while talker TG showed an advantage for utterances from the control condition.

A more complex interaction among the workload condition, vowel type, and signal-to-noise ratio variables was significant for two talkers [$F_{EG}(18,630)=1.74, p<0.05$; $F_{TG}(18,630)=1.99, p<0.01$]. Talker EG showed changes for only two of 30 possible paired comparisons across workload conditions within vowels and signal-to-noise ratios. In both cases, /æ ə/ produced under workload were significantly more intelligible than the same vowels produced in the control condition. For talker TG, four vowels (/æ ə ʌ ɔ/) showed changes in intelligibility within a signal-to-noise ratio across workload conditions. In three of the four cases, vowels produced in the control condition were more intelligible than vowels produced in the workload condition.

An interaction between workload, vowel, and amplitude equalization condition was also observed for the same two talkers [$F_{EG}(9,630)=2.03, p<0.05$; $F_{TG}(9,630)=3.22, p<0.01$]. In the unequated amplitude condition, talker EG showed an advantage for one vowel produced in the control condition (/əʁ/) and two vowels produced in the workload condition (/æ i/). No significant differences between workload conditions were observed for any vowels in the equated amplitude condition. Talker TG showed an intelligibility advantage for two control vowels in the unequated condition (/ɑ əʁ/) and two control vowels in the equated condition (/ʌ ə/). In addition, workload productions of /æ/ were more intelligible than control productions in the equated amplitude condition.

⁴Within the context of Lindblom's (1990) H&H model, it is not surprising that modifications are idiosyncratic. Lindblom explicitly states that his model is designed to describe the behavior of an *ideal* speaker, not that of any particular individual.

References

- Brenner M, Branscomb HH, Schwarz GE. Psychological stress evaluator—Two tests of a vocal measure. *Psychophysiology*. 1979; 16:351–357. [PubMed: 461663]
- Brenner M, Shipp T. Vocal stress analysis. *Mental-State Estimation* 1987. 1988:363–376. NASA Conference Publication 2504.
- Brenner, M.; Shipp, T.; Doherty, E.; Morrissey, P. Voice measures of psychological stress: Laboratory and field data. In: Titze, I.; Scherer, R., editors. *Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control*. The Denver Center for the Performing Arts; Denver: 1985. p. 239-248.
- Chen, F. Unpublished Master's thesis. Massachusetts Institute of Technology; Cambridge, MA: 1980. Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level.
- Dreher JJ, O'Neill JJ. Effects of ambient noise on speaker intelligibility for words and phrases. *J. Acoust. Soc. Am.* 1957; 29:1320–1323.
- Griffin G, Williams C. The effects of different levels of task complexity on three vocal measures. *Aviat Space Environ. Med.* 1987; 58:1165–1170. [PubMed: 3426490]
- Hanley T, Steer M. Effect of level of distracting noise upon speaking rate, duration, and intensity. *J. Speech Hear. Disord.* 1949; 14:363–368.
- Hansen, JHL. Unpublished doctoral dissertation. Georgia Institute of Technology; 1988. Analysis and compensation of stressed and noisy speech with application to robust automatic recognition.
- Hecker MHL, Stevens KN, von Bismarck G, Williams CE. Manifestations of task-induced stress in the acoustic speech signal. *J. Acoust. Soc. Am.* 1970; 44:993–1001. [PubMed: 5683666]
- Jex, HR. A proposed set of standardized sub-critical tasks for tracking workload calibration. In: Moray, N., editor. *Mental Workload*. Plenum; New York: 1979. p. 179-188.
- Jex HR, McDonnell JD, Phatak AV. A 'critical' tracking task for manual control research. *IEEE Trans.* 1966; HFE-7:138–145.
- Johnson, K.; Pisoni, D.; Bernacki, R. Research on Speech Perception Progress Report No. 15. Indiana University; Bloomington: 1989. Final report to the NTSB on the speech produced by the Captain of the Exxon Valdez.
- Jones, W. Unpublished Master's thesis. A&M University; Texas: 1990. An evaluation of voice stress analysis techniques in a simulated AWACS environment.
- Kuroda I, Fujiwara O, Okamura N, Utsuki N. Method for determining pilot stress through analysis of voice communication. *Aviat. Space Environ. Med.* 1976; 47:528–533. [PubMed: 1275844]
- 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]
- Lindblom B. Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 1963; 35:1773–1781.
- Lindblom, B. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, WJ.; Marchal, A., editors. *Speech Production and Speech Modeling*. Kluwer Academic; The Netherlands: 1990. p. 403-439.
- Lindblom B, Brownlee S, Davis B, Moon S-J. Speech transforms. *Speech Commun.* submitted.
- Markel, J.; Gray, A, Jr.. *Linear Prediction of Speech*. Springer-Verlag; New York: 1976.
- Noll A. Short-time spectrum and "cepstrum" techniques for vocal pitch detection. *J. Acoust. Soc. Am.* 1968; 36:296–302.

- Pichney M, Durlach N, Braida L. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J. Speech Hear. Res.* 1985; 28:96–103. [PubMed: 3982003]
- Pichney M, Durlach N, Braida L. 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 J. Effects of vocal force on the intelligibility of speech sounds. *J. Acoust. Soc. Am.* 1956; 28:905–908.
- Pisoni DB, Bemacki RH, Nusbaum HC, Yuchtman M. Some acoustic-phonetic correlates of speech produced in noise. *Proc. Int. Conf. Acoust. Speech Signal Process.* 1985:1581–1584.
- Pollack I, Pickett J. Masking of speech by noise at high sound levels. *J. Acoust. Soc. Am.* 1958; 39:127–130.
- Rostoland D. Acoustic features of shouted voice. *Acustica.* 1982a; 50:118–125.
- Rostoland D. Phonetic structure of shouted voice. *Acustica.* 1982b; 51:80–89.
- Rostoland D. Intelligibility of shouted voice. *Acustica.* 1985; 57:103–121.
- Scherer, KR. Nonlinguistic vocal indicators of emotion and psychopathology. In: Valdman, A., editor. *Emotions in Personality and Psychopathology.* Academic; New York: 1979. p. 493-529.
- Schneider W, Shiffrin R. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 1977; 84:1–66.
- Sheridan, T.; Stassen, H. Definitions, models and measures of human workload. In: Moray, N., editor. *Mental Workload.* Plenum; New York: 1979. p. 219-234.
- Shiffrin R, Schneider W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* 1977; 84:127–190.
- Strange W, Edman TR, Jenkins JJ. Acoustic and phonological factors in vowel identification. *J. Exp. Psych.: Human Perform. Percept.* 1979; 5:643–656.
- Strange W, Verbrugge RR, Shankweiler DP, Edman T. Consonant environment specifies vowel identity. *J. Acoust. Soc. Am.* 1976; 60:213–224. [PubMed: 956528]
- Streeter LA, MacDonald NH, Apple W, Kxauss RM, Galotti KM. Acoustic and perceptual indicators of emotional stress. *J. Acoust. Soc. Am.* 1983; 73:1354–1360. [PubMed: 6853847]
- Summers W, Pisoni D, Bernacki R, Pedlow R, Stokes M. Effects of noise on speech production: Acoustic and perceptual analyses. *J. Acoust. Soc. Am.* 1988; 84:917–928. [PubMed: 3183209]
- Tolkmitt FJ, Scherer KR. Effect of experimentally induced stress on vocal parameters. *J. Exp. Psych.* 1986; 12:302–312.
- Williams CE, Stevens KN. On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Med.* 1969; 40:1369–1372.

FIG. 1.

Phrase amplitude (upper panel) and segmental amplitudes for “Say hVd again” utterances produced in JEX and control conditions. The * symbol appears between mean values that are significantly different. Values are collapsed across utterances and presented separately for each talker.

FIG. 2.

Between-utterance standard deviations for phrase amplitudes (upper panel) and segmental amplitudes. The * symbol appears between mean values that are significantly different. Values are collapsed across utterances and presented separately for each talker.

FIG. 3.

Mean difference in energy between utterances produced in the JEX and control conditions across frequency bands. Values are collapsed across utterances and presented separately for each talker. For clarity of presentation, the traces for talkers EG, ME, MC, and SL have been elevated by 2.5, 5, 7.5, and 10 dB, respectively.

FIG. 4.
Mean fundamental frequency values for utterances produced in JEX and control conditions. The * symbol appears between mean values that are significantly different. Values are collapsed across utterances and presented separately for each talker.

FIG. 5. Mean within-utterance *F0* standard deviations for utterances produced in JEX and control conditions. The * symbol appears between mean values that are significantly different. Values are collapsed across utterances and presented separately for each talker.

FIG. 6.

Mean phrase duration (upper panel) and mean segmental duration values for utterances produced in JEX and control conditions. The * symbol appears between mean values that are significantly different. Values are collapsed across utterances and presented separately for each talker.

FIG. 7. Perceptual identification accuracy as a function of talker and workload condition. The * symbol appears between mean values that are significantly different. Values are collapsed across signal-to-noise ratio, vowel type, and amplitude normalization condition.

FIG. 8.

The interaction of workload condition and amplitude condition for each talker (unleveled—not equated for amplitude; leveled—equated for amplitude).

\$watermark-text

\$watermark-text

\$watermark-text

TABLE I

Effects of noise, stress and workload on selected characteristics of speech (adapted from Johnson *et al.*, 1989). ↑=consistent increase; ↓=consistent decrease; ↕=Inconsistent Increase; ↓=inconsistent decrease; ↕=increases and decreases; and NC=no change.

Condition	F0	SD F0	Jitter	Tilt	Duration	Intensity	Formants
Noise ^a	↑	↑		↓	↑	NC	F1 ↓
Noise ^b	↑			↓	↑	↑	F1 ↓
Stress ^c	↑		↓		↓	↑	
Stress ^d	↑		↓			NC	
Stress ^e	↕			↕	↑		
Perceived Stress ^e	↑				↑	↑	
Workload ^a	↑	↑		↓	NC	↑	F1&F2↑
Workload ^f	↑		↓				
Workload ^g	↑	↑			↓	↑	
Workload ^h	↑	↓	NC	↓	↓	↑	NC

^aHansen (1988).

^bSummers *et al.* (1988).

^cBrenner and Shipp (1988).

^dBrenner *et al.* (1985).

^eStreeter *et al.* (1983).

^fJones (1960).

^gGriffin and Williams (1987).

^hPresent data.