75-21,198

SPENCER, Linda Ellis, 1945-
AN INVESTIGATION OF FACTORS CONTRIBUTING TO LISTENER RATINGS OF LOUDNESS AND EFFORT FOR SELECTED SPEECH SAMPLES.

The University of Oklahoma, Ph.D., 1975
Speech Pathology

Xerox University Microfilms, Ann Arbor, Michigan 48106

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN INVESTIGATION OF FACTORS CONTRIBUTING TO LISTENER RATINGS OF LOUDNESS AND EFFORT FOR SELECTED SPEECH SAMPLES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

LINDA ELLIS SPENCER

Oklahoma City, Oklahoma

1975

AN INVESTIGATION OF FACTORS CONTRIBUTING TO

LISTENER RATINGS OF LOUDNESS AND EFFORT

FOR SELECTED SPEECH SAMPLES

APPROVED BY

Donald T Counihan

Walter L. Cullinan

Floyd W. Emanuel

Thomas E. Hutchinson

R B Deal

DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

The author wishes to express her gratitude to Donald T. Counihan, Ph.D., for his guidance in the planning, executing, and writing of this dissertation. Appreciation is also expressed to the members of the dissertation committee, Walter L. Cullinan, Ph.D., and Floyd W. Emanuel, Ph.D., Department of Communication Disorders; Thomas E. Stokinger, Ph.D., Chief of Audiology and Speech Pathology Services, Veterans' Administration Hospital; and Roy B. Deal, Jr., Ph.D., Department of Biostatistics and Epidemiology, for their helpful advice and suggestions.

Appreciation is extended to Donald E. Parker, Ph.D., Department of Biostatistics and Epidemiology, for his assistance in the statistical analysis of the data. Thanks are expressed to H. Andrew Leeper, Ph.D., Oklahoma State University, for helpful suggestions during the planning and execution of the investigation.

The author also wishes to thank Steffi B. Resnick, Ph.D., City University of New York, for her assistance in the instrumental design. Thanks are due to John W. Spencer, Ph.D., Ms. Karen Carothers, Mr. Bill Lovallo, and Ms. Kay Ahaus, who served as experimental subjects, and to the individuals who served as judges.

Special gratitude is expressed to my husband, John, who provided counsel and encouragement throughout this endeavor, and to our friends for their understanding.

This research was supported in part by a Veterans' Administration Trainseeship from the Department of Communication Disorders, University of Oklahoma Health Sciences Center.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS--Continued

AN INVESTIGATION OF FACTORS CONTRIBUTING TO LISTENER RATINGS OF LOUDNESS AND EFFORT FOR SELECTED SPEECH SAMPLES

CHAPTER I

INTRODUCTION

In recent years, investigators have attempted to define the input-output relationships that exist for various sensory transducers as they are revealed by changes in the behavior of human subjects (28, 29). From such studies has come a better understanding of the interactions between the physical properties of stimuli and the percepts that arise from sensory experience with them. The researcher in speech has been an active participant in this line of inquiry based on his interest in understanding the relationships between the physiologic and acoustic events of speech and the perception of these events by listeners. One area in which this interest has been focused concerns the relationship between the intensity of vocal production and the degree of loudness perceived by listeners. An understanding of this relationship could be expected to shed light on the strategies employed by humans in the production and perception of speech.

While intensity-loudness relationships for pure-tone stimuli have been relatively well defined (28, 29), the same is not true for speech signals. Lane, Catania, and Stevens (16) report that when speakers are asked to rate the loudness of their own vowel productions, obtained using the method of magnitude production, loudness increased in relation to sound pressure level at a faster rate than that reported for listeners rating the loudness of pure tones. To explain this phenomenon, the authors hypothesized that speakers use cues to the physical magnitude of the speech level which arise from kinesthetic feedback from the speech production mechanism, in addition to the intensity of the signal. These cues may reflect the amount of effort used to produce speech at different intensities.

As part of the same experiment, subjects were asked to rate the loudness of their own recorded vowel productions. Their loudness ratings resembled those of subjects rating the loudness of pure tones. The authors concluded that the subjects who rated the speech samples depended on the physical magnitude (intensity) of the vowel stimuli rather than on cues to the amount of effort used to produce the vowel. In a subsequent study, however, Mendel et al. (21) reported systematic differences in loudness ratings of vowels and other types of speech stimuli, depending upon whether or not changes in signal intensity were accompanied by changes in speaker production effort.

When the complexity of the speech signal increases, that is, when vowels and consonants are combined into syllables or larger speech units, the loudness function reported by listeners is similar to that reported by speakers when judging the loudness of their own vowel productions (1, 2, 3, 9, 15). It is thought that relevant cues to the loudness of speech are provided by spectral changes in the acoustic signal which accompany variations in the effort used by the speaker to produce a particular speech sample at different intensities (7, 9).

Acoustic parameters which have been found to contribute to listeners' perceptions of vocal effort include speech power, fundamental vocal frequency, and phonetic quality (18). In addition, it has been demonstrated that the complexity of the speech sample affects loudness ratings assigned by listeners (8, 15, 16, 22). However, there have been few attempts to investigate, in a single study with uniform speech samples, the effects of those factors which could contribute significantly to listeners' ratings of speech loudness and effort. Information gained from such an investigation could contribute materially to an understanding of the speech perception process. Further, such an investigation could be expected to provide direction for future studies of the relationship between speech sound production and its perceived loudness and effort. It is the purpose of the present study, therefore, to investigate the effects of production intensity, playback intensity, fundamental vocal frequency change accompanying alterations in production effort, vowel identity, and type of speech sample on listeners' perceptions of speech loudness and effort.

CHAPTER II

REVIEW OF THE LITERATURE

Studies of the loudness of speech have focused on the psychological correlates of changes in the intensity of isolated vowels (1, 4, 15, 16, 18, 21, 31), consonant-vowel syllables (1, 15, 17, 21, 23), and connected sentences (8). The results of these studies suggest that speech loudness is not dependent upon a single acoustic cue, but, rather, depends upon complex relationships among a number of acoustic cues. No investigators have attempted to determine in a single study the effects of a number of these variables which individually have been shown to provide pertinent information about the loudness of speech. The present investigation considered the effects of vocal intensity, playback intensity, type of speech sample (vowel or sentence), and variations in fundamental vocal frequency accompanying alterations in speaking level (vowels) upon listener ratings of speech loudness and speech effort. Literature reviewed as background for this study is reported under the major headings: (a) physiologic correlates of speech loudness, and (b) psychoacoustic studies of speech loudness.

Physiologic Correlates of Speech Loudness and Effort

When a speaker increases his speaking level, he may do so by using greater "vocal effort." That is, he may increase the amount of work done by the physiological mechanisms normally used for speech production. Models of the speech production processes postulate an interaction between pulmonic (subglottal) pressure and pressures within the oral cavity (lingual and intra-oral pressures). A number of investigators (10, 15, 17, 19, 20, 23) have attempted to define "vocal effort" in terms of interactions among these physiological correlates of speech production.

The impetus for this line of investigation was provided by the results of a study by Lane, Catania, and Stevens (16) who related a speaker's estimates of the loudness of his own vowel productions (the autophonic response) to the 1.1 power of the sound pressure of the vowel production. Moreover, they reported that when auditory feedback to the speaker was reduced, the form and slope of the loudness function was altered little. They concluded that physiological feedback from the act of vocalization, that is, the speaker's subjective estimate of his own vocal effort, was more important as a source of information about speaking level than was the loudness of the signal itself.

Subsequently, a number of investigators have attempted to define the physiological correlates of vocal effort. Ladefoged and McKinney (15) concluded that the work performed on air during speech production was approximately proportional to the square of the subglottal presure. This finding suggests that the amount of work done in providing pulmonic air needed to support the vibration of the vocal folds is a primary physiological correlate of speaking effort.

This relationship does not appear to hold over the entire range of vocal fundamental frequencies, however. Isshiki (14) reported that air flow through the glottis increases when speakers phonate at higher rather than lower pitches. He hypothesized further that glottal resistance is probably of primary importance in regulating vocal intensity at low pitches, while at high pitches air flow rate assumes greater importance. Thus, even though subglottal pressure was reported by Ladefoged and McKinney to increase linearly with vocal intensity, the interaction of the physiological voice production mechanism with the rising air stream appears to be non-uniform over the entire pitch range.

Further, it appears that intensity regulation may be subject to some degree of speaker-to-speaker variability. Re-examination of graphs presented by Isshiki reveals some variability in air flow rates at each pitch attributable to the individual speakers. More recently, Scharf (26) asked three speakers to produce a single vowel (/ɑ/) at a "comfortable" speaking level. A different intensity was recorded for each speaker. When the same speakers were asked to phonate at different self-perceived levels of vocal effort, uniform changes in speaking intensity were not observed from speaker to speaker. These findings suggest that individual speakers probably do not utilize the same amount of physiological effort to phonate at the same intensities.

Recently, investigators have attempted to define the relationship of the supraglottal articulators to pulmonary function as part of a sensory feedback system used in the perception of speaking effort. Malecot (19, 20), for example, has speculated that physiological feedback from the interior of the oral cavity, based on intraoral air pressure, regulates the lenis/fortis (voiced/voiceless) contrast noted among English consonants by a number of other investigators (2, 6, 12, 30).

Further evidence that measures of intraoral air pressure are related to ratings of speech effort has been provided by the results of a number of recent studies. Ringel *et al.* (23) demonstrated that speakers are able to scale the degree of effort used in consonant production, and show that measures of intraoral air pressure relate well to changes in the speaker's articulatory effort. In a subsequent study, Brown and Brandt (9) tested speakers' abilities to control the intensities of speech utterances both with and without auditory feedback. They reported that when speakers received binaural masking, measures of SPL and intraoral pressure varied little from measures obtained during a condition of no masking. In other words, the speakers appeared to be using sensory-motor information in addition to auditory feedback in controlling and maintaining constant speech intensity.

More recently, Leeper and Noll (17) found that measures of perceived vocal effort obtained using the methods of magnitude production and magnitude estimation were power functions of intraoral air pressure, lingual (tactile) pressure, and sound pressure level (SPL). The investigators speculated that force applied to the oral mucosa excites kinesthetic, touch, and, perhaps, pressure senses. That listeners were able to rate the amount of effort used by the speakers implies the presence of cues to this percept within the acoustic signal.

The results of these studies suggest that vocal effort is a result of the work done by the entire speech production mechanism. Moreover, mechanical pressures generated throughout the vocal tract may provide important sensory feedback which plays a primary role in controlling vocal effort.

Psychoacoustic Studies of Speech Loudness and Effort

Recent literature in the area of speech production has suggested that speakers use physiologic feedback from the degree of vocal effort as a cue to speech loudness. It remains to be determined, however, whether a listener bases his estimates of the loudness of speech on estimates of the vocal effort used by the speaker, or whether he relies on the intensity of the signal, as in listeners' ratings of the loudness of pure tones (19, 28), or some interaction between these two parameters. The motor theory of speech perception suggests that listeners impose a unique perceptual set on speech as opposed to non-speech stimuli, presumably based on their knowledge of the mechanics of speech production (27). Thus, it would be logical to suspect that the psychophysical scaling of the loudness of speech might be different from the scaling of the loudness of pure tones.

Lane, Catania, and Stevens (16) determined that the loudness function for a speaker's ratings of his own vocal productions (the autophonic response) differed from the loudness function for vowels heard by listeners, suggesting that speakers and listeners respond to different arrays of cues to speaking level. The listeners were allowed to hear the isolated vowel samples under two intensity conditions. First, vowels were heard at playback intensities which corresponded to the original production intensities. Presumably, this condition provided acoustic cues to production level (effort) as well as sound pressure level (SPL) differences among the samples. Second, a single vowel production was duplicated a number of times and reproduced to a panel of listeners at a number of playback intensities, providing SPL differences among samples which represented a single production level. Under these two conditions of vowel presentation, listener loudness ratings for the isolated vowel /ɑ/ were related to SPL by a power function of 0.7. In contrast, the autophonic response was related to SPL by a power function of 1.1. Lane, Catania, and Stevens concluded that the speakers used cues to vocal effort to monitor the levels of their own voices, whereas listeners responded only to the SPL differences among the samples.

Subsequent investigations have failed to confirm Lane, Catania, and Stevens' conclusion and have suggested, in fact, that differences in physiological effort generate acoustic cues which may affect listeners' judgments of speech loudness. The nature of these effort-related spectral alterations is as yet unknown, as is the type of speech sample (vowel, syllable, or connected speech) in which the alterations occur. There exists, however, a body of recent literature which bears on these questions.

Lehiste and Peterson (18) have suggested that changes in speaking effort are signaled by alterations in the vowel spectrum. It has been reported previously (5, 13, 24, 25) that greater power is associated with vowels produced with greater mouth opening. That is, intrinsic vowel amplitude appears to be related to the interaction between the glottal sound source and any impedance imposed by the articulatory configuration. Lehiste and Peterson reported that when several isolated vowels were produced at equal intensity, those vowels having lower intrinsic amplitudes and, thus, requiring greater vocal effort to produce at the same intensity as the more powerful vowels, sounded louder to listeners. In contrast, when the same vowels were produced at subjectively equal levels of speaking effort they sounded about equally loud. The listeners noted differences in vocal quality among the vowel samples produced at equal intensity which, the authors speculated, were related to differences in vocal effort and may have influenced the loudness ratings.

Further evidence that listeners may use cues to effort within the vocal signal is offered by Ladefoged and McKinney (15). They reported the loudness of CV syllables to be an exponential function both of SPL (1.2) and of subglottal pressure (2.0). Since SPL was found to be a linear function of subglottal pressure, with an exponent of 0.6, it was reasoned that ". . . in the case of speech sounds, loudness is directly related to physiological effort . . . ." It might be speculated that the cues to physiological effort were conveyed to the listeners primarily by the intensity of the signal. However, some aspect of the spoken signal other than SPL also may have influenced loudness ratings. Ladefoged and McKinney reported that words spoken with equal subglottal pressure were equally loud, even though their intensities were not always equal, a finding which is consistent with previously cited reports by Lehiste and Peterson. Subsequently, Allen (1) replicated Ladefoged and McKinney's study, concluding that ". . . both intensity and effort were used as relevant cues for the loudness judgments of these subjects."

Mendel <u>et al</u>. (21) reported that loudness functions for speech and non-speech stimuli differed from each other. Moreover, when listeners heard only SPL differences among samples having equal production effort, the loudness functions were similar to those for non-speech stimuli. However, when listeners heard samples having different production levels, the speech loudness function was different from that produced when listeners heard only playback level difference. This suggests that listeners made use of unspecified cues to loudness in addition to signal intensity. Their results also suggest that these cues were available to listeners in rating isolated vowels as well as in syllables and in connected speech samples.

The availability of cues to production effort in connected sentences was studied by Brandt, Ruder, and Shipp (8) as well, who investigated listeners' perceptions of vocal loudness and vocal effort. They recorded a single sentence spoken at eight different intensities over a 35 dB range and then rerecorded the stimulus samples under two experimental conditions. First, they selected one sample which originally had been produced at 80 dB SPL and rerecorded it at eight different intensities which corresponded to the eight levels utilized in the initial recording. Among these rerecorded samples, they reasoned, vocal effort was constant while the intensity (SPL when replayed to listeners) of the spoken samples varied. Under the second experimental condition, they rerecorded the original eight stimuli at one intensity, thus allowing vocal effort to vary while intensity (SPL when replayed to listeners) was controlled at one level. In addition, samples whose playback intensities corresponded to the original production intensities were provided. All of these samples were randomized and made into a single listening tape. Listeners then rated each of the samples for loudness and then for effort. When both intensity and effort cues were available to listeners, loudness was proportional to SPL to the 1.12 power, while the effort function had a slope of 0.89. When intensity cues alone were available and effort was constant, the exponents of the loudness and effort functions were diminished to 0.92 and 0.38, respectively. With effort varying and intensity remaining constant, the slope of the effort function was 0.57, while the slope of the loudness function was 0.40. These relationships of listener judgments to sound pressure level suggest that listeners could differentiate cues to effort from those which signaled loudness. Also, examination of the data suggested to these experimenters that intensity differences (SPL at playback) must be available to listeners making loudness judgments of speech, while both intensity differences at playback and differences in speaker production levels must be available for listeners to hear changes in production effort.

A number of alterations in the acoustic signal probably influence listeners' ratings of loudness and/or effort. Fundamental vocal frequency may be one such factor. In studies by Lane _et al_. (16) and by Ladefoged and McKinney (15) increases in speaking effort reportedly were accompanied by increases in fundamental vocal frequency. To investigate the possibility that systematic alterations in vocal frequency provide pertinent cues to speaking effort, Moll and Peterson (22) collected vowel samples with fundamental vocal frequency held constant as speaking level increased. Their loudness functions were less steep than those reported by Lane, Catania, and Stevens (16), who allowed listeners to hear fundamental vocal frequency alterations accompanying increases in speaking intensity.

Moll and Peterson's conclusion that the alteration in fundamental vocal frequency associated with changes in speaking intensity is important to loudness judgments has been corroborated to some extent by the results of a more recent study by Wright and Colton (31). For both magnitude estimation and magnitude production tasks, the authors demonstrated that rated vocal effort for the vowel /ɑ/ is a power function both of frequency level (FL) and SPL. Also, they noted that a change in the slope of the vocal effort function occurred at the speaker's most comfortable effort level, perhaps corresponding to a change in the operational mode of the vocal system which occurred at that point.

More recently, Bernstein (4) reported a similar relationship between vocal effort, and judged loudness of the vowel /ɑ/. When SPL, formant frequencies, duration, and fundamental frequency of the vowel stimuli were controlled, the author noted that increasing vocal effort did not systematically increase loudness. In fact, vowels produced by a male and a female speaker at the lower end of a range of intensities often sounded louder to listeners, though vocal effort would appear to have been less. The author concluded that a highly pertinent cue to vocal loudness may be produced by an inefficient mode of phonation which occurs when an individual attempts to phonate at a pitch or intensity which falls outside a range in which he phonates most comfortably. Spectral analysis of the vowels produced in this study indicated that loudness changes may be predicted from the amount of spectral energy around 3000 Hz.

Wave-form analysis of connected speech samples has provided additional evidence of a number of spectral cues to speech loudness which may be generated by alterations in speaking effort. Brandt, Ruder, and Shipp (8) reported that spectral analysis of the samples originally produced by speakers at eight different intensities revealed changes in articulatory patterns as well as alterations in the amount and distribution of energy throughout the spectrum which accompanied increases in speaking intensity. Most notable were alterations in the amplitudes of consonants and vowels and in the amount of noise energy, especially at low vocal intensities. The latter finding is consistent with reports by Austin and Emanuel (3) of high correlations between listener judgments of vowel roughness and vocal effort for vowels and measures of inharmonic energy in the vowel spectra. Additionally, these measures may correspond to the differences in vocal quality reported by Lehiste and Peterson (18) among vowels produced with different degrees of vocal effort.

Additional spectral analysis of their sentence samples by Brandt, Ruder, and Shipp revealed a systematic increase in total bandwidth with increases in vocal intensity. Based on loudness data, the authors concluded that stimulus bandwidth provided the primary cue to the perception of vocal effort. Results of a subsequent study (7) using bandpass filtering indicated that increased bandwidth appears to be important to the perception of loudness but not to vocal effort. The authors hypothesize that increases in vocal fundamental frequency as well as intensity are important to effort perception, while intensity and bandwidth changes contribute to the perception of vocal loudness.

Many of the investigations reviewed in this section have dealt with relationships between the intensity of speech samples and their rated loudness or effort. Allen (1) has suggested that this approach is too simplistic to offer real insight into the way in which a listener estimates the loudness of speech. In addition to SPL, a listener appears to use a multiplicity of acoustic cues to speaking level which are related to the amount of effort used by the speaker. The nature and importance of these cues cannot be assessed by the use of investigative techniques of limited scope. Thus, there appears to be a need for a systematic and controlled investigation of the relationships between listener judgments of speech loudness and effort for more than a single type of speech sample, with consideration of the relative importance of a variety of factors which may influence the ratings. The present investigation was undertaken with this need in mind.

CHAPTER III

DESIGN OF THE INVESTIGATION

A major objective of this study was to evaluate, for both isolated-vowel and connected-speech test samples, effects of vocal production intensity on listener judgments of vocal loudness and vocal effort. Additionally, effects of the phonetic context and the fundamental vocal frequency of isolated vowel productions on listener judgments of vowel loudness and effort were studied. The specific research questions investigated are listed below. The first three questions were investigated both for isolated vowel and sentence test samples, but the last two were investigated for isolated vowels only.

1. What are the effects of vocal (production) intensity on listeners' perceptions of vocal loudness and vocal effort?

2. What are the effects of the intensity at which speech samples are heard on listeners' perception of vocal loudness and vocal effort?

3. What are the effects of type of speech sample (sentence or vowel) on listeners' perceptions of vocal loudness and vocal effort?

4. What are the effects of vowel identity (/ɑ/ or /i/) on listeners' perceptions of vocal loudness and vocal effort?

5. What are the effects of (a) "natural" vocal frequency variations and (b) a controlled vocal frequency on listeners' perceptions of vocal loudness and vocal effort?

The selection of subjects, the experimental apparatus, and the procedures followed in the collection of the data are presented in the following sections.

Subjects

The subjects for this study were two adult males and two adult females; all were between the ages of twenty-five and thirty-three years. Each subject presented normal speech and vocal quality, as determined by a speech pathologist. Also, within each sex group, the two individuals selected as subjects presented closely similar habitual fundamental vocal frequencies.

Speech Samples

The subjects individually produced seven-second isolated samples of two test vowels, /ɑ/ and /i/, under two "conditions" of frequency control (described below). Additionally, each subject produced a test sentence ("According to the present information the profits are high.") which was used in a study by Brandt et al. (1969). Each test speech sample was produced at each of three intensities: 65, 75, and 85 dB re 0.0002 dyne/cm^2 (SPL), at a mouth-to-microphone distance of four inches.

The two "vocal-frequency-conditions" for vowel productions were as follows. For the first (Frequency Condition I), fundamental vocal frequency was not controlled; rather, it was permitted to vary "naturally" with changes in vocal intensity. For the second (Frequency Condition II), subjects were required to match the fundamental vocal frequency of each test vowel production as nearly as possible to that of a pure tone supplied as a reference.

The "target" frequency for Frequency Condition II was specified separately for each of the two sex groups, as follows. First, an estimate was made of each subject's pitch range by determining the highest and lowest pure tones that each could match during a sustained (five seconds) production of the vowel /ɑ/. Next, a frequency representing the thirty per cent point (above the lowest frequency) in the subject's fundamental vocal frequency range was specified as the "target". A within-sex mean was then obtained of the "target" frequencies for the individual subjects. The mean frequency obtained (212 Hz for the females; 173 Hz for the males) was taken as the "target" frequency to be matched by the two subjects in each sex group during Frequency Condition II.

Instrumentation

This study required six instrumental systems: (a) a signal system, (b) an audio recording system, (c) an intensity control system, (d) a frequency control system, (e) an audio playback system, and (f) a calibration system.

Description

Signal system. A simple electro-mechanical cam timer, activated by the experimenter, controlled the illumination of panel lights used to signal subjects to begin and terminate test vowel phonations.

Audio recording system. The audio recording system, used to obtain magnetic tape recordings of the test samples, consisted of a sound level meter (General Radio, Type 1551-C) with an attached non-directional piezoelectric ceramic microphone (General Radio, PZT Type 1560-P3); a dual-channel magnetic tape recorder (Ampex, Model AG 440);

and, a monitoring amplifier (Bruel and Kjaer, Type 2606).

The PZT microphone was designed to have a flat frequency response (±1 dB) from 20 to 8000 Hz when placed at a 70° angle of incidence to the sound source, and a sensitivity of -60.3 dB re 1v/microbar. When appropriately placed in a sound field, the sound level meter should indicate the sound pressure level at its PZT microphone with an average signal-to-noise ratio in octave bands from 20 to 10,000 Hz of at least 66 dB. The tape recorder was designed to have a flat frequency response (± 2 dB) from 40 to 12,000 Hz with a signal-to-noise ratio of at least 65 dB at a tape speed of 15 inches per second. Figure 1 shows that the output from the sound level meter could be switched either to the input of the tape recorder or, as needed, to other devices.

Intensity control system. In data collection, the output of the tape recorder was led directly to the input of an amplifier (Bruel and Kjaer, Type 2606) which served as a vocal-intensity-monitoring meter. This instrument indicated vocal intensities with a signal-to-noise ratio of at least 55 dB. The amplifier had an averaging time constant of approximately 100 msec, and designed accuracy within ± 0.5 dB up to crest factor 10.

Fundamental frequency control system. As illustrated in Figure 1, the frequency control system consisted of two sections. The first section incorporated an audio oscillator (Hewlett-Packard, Model 200 ABR) connected through a matching transformer to a high-quality loudspeaker. A normally open switch between the transformer and the loudspeaker could be closed to present the subject a tone produced by the oscillator as a frequency reference.

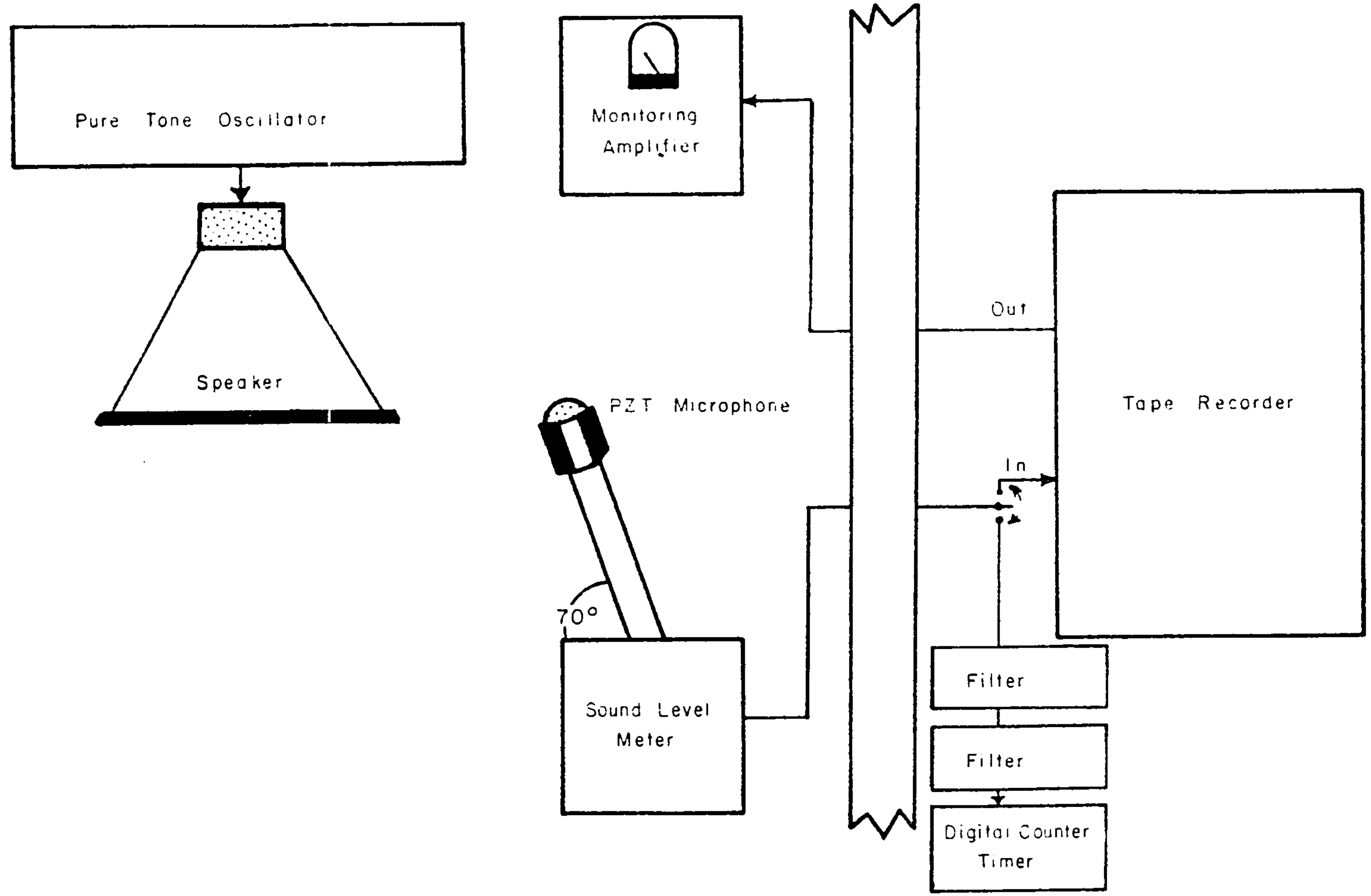

Figure 1.--Simplified diagram of the audio recording system.

The second section of the fundamental frequency control system consisted of two variable electronic filters (Spencer-Kennedy Laboratories, Model 302) and a digital counter-timer (Transistor Specialties, Inc., Model 361). The two variable electronic filters were arranged in series to provide a 36 dB per octave rejection rate beyond selected cut-off frequencies. This section was used by the experimenter to measure the fundamental frequency of the speaker's vowel productions under Frequency Condition II. The output of the sound level meter which was incorporated as part of the audio recording system, could be led directly to the second section of the frequency control apparatus by closing a single-pole-double-throw switch inserted in the circuit (see Figure 1).

Playback system. The playback system, used to present the recorded samples to listeners for judgment, consisted of Channel I of each of two dual-channel magnetic tape recorders (Ampex, Model AG 440), an attenuator (Hewlett-Packard, Model 350-D), an auxiliary amplifier (Channel II amplifier of one of the dual-channel tape recorders), a monitoring amplifier (Bruel and Kjaer, Type 2606), and a loudspeaker (Altec, Model 844A).

The channel-one outputs of the two tape recorders were led to a switch, which permitted the experimenter to select one output to be led to the attenuator. The attenuated signal was led to the auxiliary amplifier and then to the loudspeaker. Before the tapes were played to listeners, the monitoring amplifier was inserted into the circuit between the switch and the attenuator to balance the outputs of the two tape recorders.

Calibration system. Components employed in instrument calibration included a pure tone oscillator (Hewlett-Packard, Model ABR 200) which drove a loudspeaker, a sound level meter (General Radio, Type 1551-C), and its attached non-directional piezoelectric ceramic microphone (General Radio, PZT Type 1560-P3), and a monitoring amplifier (Bruel and Kjaer, Type 2606). A simplified diagram of the calibration system is presented in Figure 2.

Calibration

The magnetic tape recorder was inspected and adjusted by an audio engineer prior to data collection. The voltmeter of the monitoring amplifier was calibrated to indicate vocal intensities of 65, 75, and 85 dB SPL. To calibrate this meter, a 1000 Hz reference tone produced by the pure tone oscillator was led to a loudspeaker. The PZT microphone of the sound level meter was placed at a 70^{o} angle of incidence to and one inch in front of the loudspeaker in an acoustically isolated room. The intensity of the pure tone was adjusted until it produced a 75 dB SPL sound level meter deflection. The output of the sound level meter was then connected directly to the input of the tape recorder and the input amplifier of the recorder was adjusted for a -10 dB deflection of its VU meter in response to the 75 dB SPL input. The output of the recorder was then led to the monitoring amplifier and the input potentiometer of the amplifier was adjusted for a 75 dB deflection on the voltmeter of the amplifier.

Next, the 60 dB range and then the 80 dB range on the sound level meter were selected, and the intensity of the pure tone was adjusted until 65 dB SPL and 85 dB SPL readings were obtained, respectively. The 65 dB and 85 dB signals produced deflections of the tape recorder's

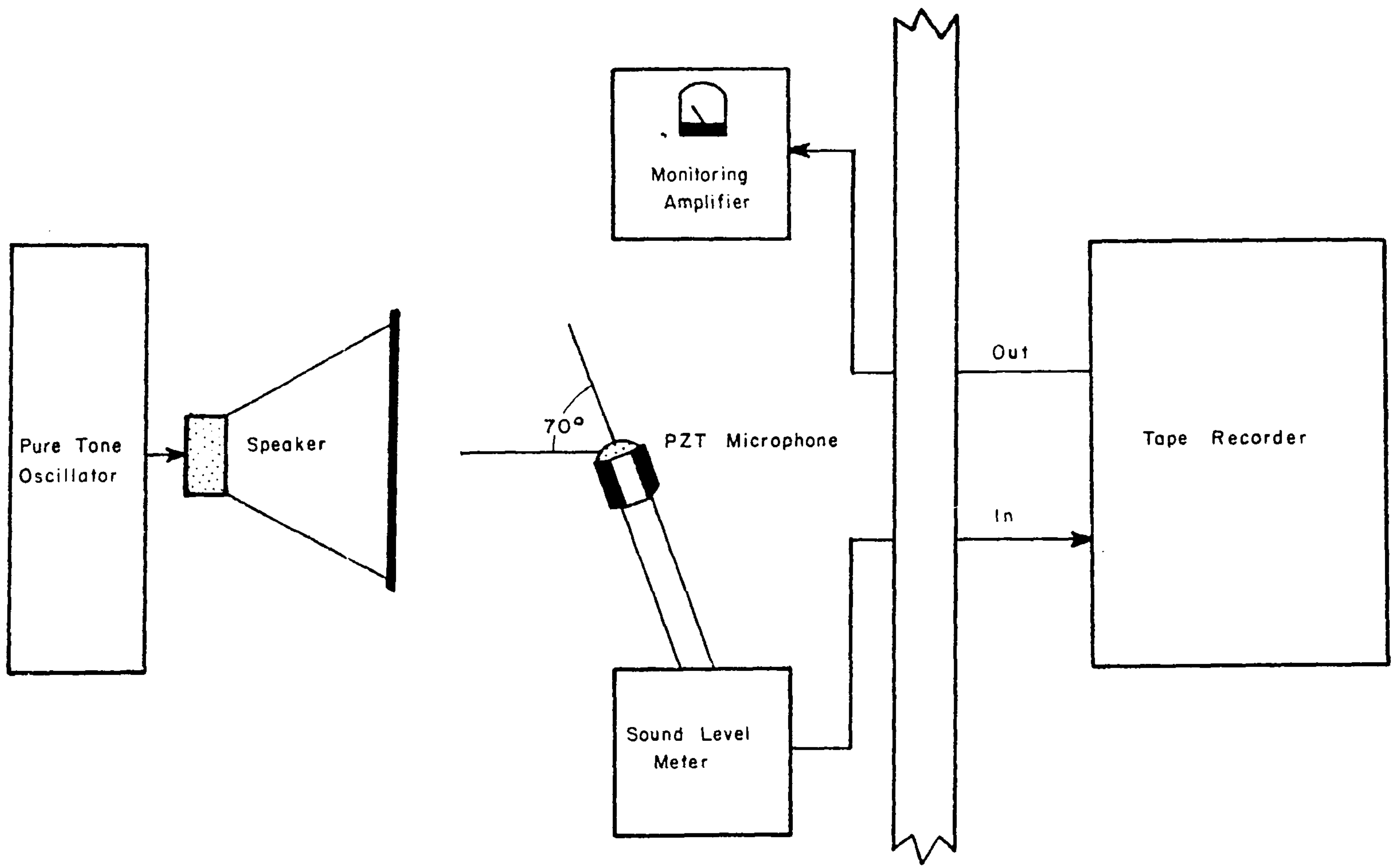

Figure 2.--Simplified diagram of the calibration system

VU meter and the monitoring amplifier's voltmeter were the same as that for a 75 dB tone, thus confirming that the voltage delivered by the sound level meter was the same for tones of 65, 75, and 85 dB SPL. A single reading could therefore be marked on the voltmeter of the monitoring amplifier as the intensity level which each subject was to maintain during production of all test samples.

Procedures

The experimental procedures in this study included: (a) recording the subjects' productions of the test samples, (b) analyzing the intensities of the recorded vowel and sentence samples, (c) analyzing the fundamental frequencies of the isolated vowel samples, and (d) presenting the recorded vowel and sentence productions to the judges who individually rated each test production for loudness and effort.

Recording Procedure

All test speech samples were collected in an acoustically isolated two-room suite with a low ambient noise level in the Speech and Hearing Center, University of Oklahoma Health Sciences Center. During recording, the subject and examiner were in separate rooms of the two-room suite, but an intercom system allowed the examiner to communicate with the subject. The test room contained the subject's chair, the sound level meter with its attached PZT microphone, the vocal-intensity-monitoring amplifier, and the signal lights used to control the initiation and termination of each test vowel phonation. The adjoining (equipment) room contained the tape recorder, the filter system, the digital counter-timer, and the cam timer which controlled the activation of the signal lights.

To obtain the spoken samples, each subject was initially familiarized with the experimental procedures and was then seated in the examination chair. The head rest on the chair was adjusted so that head movement was minimized during recording. The sound-level-meter microphone was placed at a 70° angle of incidence to and four inches in front of the subject's mouth. The monitoring amplifier was positioned to allow the subject to observe readily the intensity of each test phonation.

Test samples were collected in two recording sessions. During the first session, both sentences and isolated vowels (Frequency Condition I) were recorded. Isolated vowels only (Frequency Condition II) were recorded in the second session. This fixed order of vowel-frequency conditions was chosen to avoid for samples produced under Frequency Condition I a possible carry-over "influence" of the fundamental vocal frequency control imposed for Frequency Condition II.

During the first recording session, each subject practiced producing the test samples at 65, 75, and 85 dB SPL. Subjects were instructed to use "natural" patterns of rate and stress in producing the sentences at those intensities. The needle on the monitoring amplifier's voltmeter was allowed to swing past the intensity mark during production of stressed syllables and to fall to zero after the word "information," which marked the end of the phrase "According to the present information the profits are high." For the remainder of the passage, subjects were instructed to maintain the needle, as nearly as possible, at the intensity mark. For vowel production subjects were required to maintain the position of the monitoring amplifier's voltmeter needle at the intensity mark (± 1 dB) throughout most of the seven-second production.

During the second recording session, each subject practiced producing the vowel samples at the required frequency (212 Hz for the females; 173 Hz for the males) as well as at the required duration (seven seconds) and intensities (65, 75, and 85 dB SPL). To aid in producing samples at the required frequency, the subject supplied himself with a pure tone reference by depressing a push-button switch, thus connecting the loudspeaker and the audio oscillator. The frequency of the pure tone was the predetermined "target" frequency at which the subject was to produce all test vowels. During the practice session, the output of the sound level meter was connected to an acoustic filtering system. The filters, connected in series, were adjusted to pass an approximately 10-Hz band centered at the "target" frequency. The signal was led to the counter-timer which displayed the frequency of the filtered wave digitally. When it appeared that the subject was reliably producing test vowels at the required frequency, the reference tone was switched off, and the sound level meter was reconnected to the recording system.

Upon completion of the above described training, the test samples were recorded. The order of production of all test samples at each of the three intensities was counterbalanced for all four subjects during each of the two recording sessions. If the subject did not produce the appropriate sample, or did not maintain the required intensity (± 1 dB) or fundamental frequency (± 2 Hz in Frequency Condition II), the trial was repeated until an acceptable performance was achieved.

Intensity Analysis

To aid in the selection of vowel and sentence samples which met the production criteria set for this experiment, the intensities of all sample productions were analyzed following recording by reproducing them and noting the intensity of each sample as indicated by the monitoring amplifier. For each subject, one sample of the sentence and one of each of the two test vowels (Frequency Condition I) which were within 1 dB of the experimental intensities were retained for incorporation into an experimental listening tape. Also, for each subject, several vowel samples (Frequency Condition II) which were within 1 dB of the experimental intensities were analyzed further to determine the fundamental frequency of each. Of these samples, one sample of each of the two test vowels was chosen for the listening tape. The procedure used for fundamental frequency analysis is presented in detail below.

Fundamental Frequency Analysis

The vowel samples produced by each subject in Frequency Condition II which were within 1 dB of the experimental intensities were analyzed by acoustically filtering each sample to determine the fundamental frequency. The acoustic filters, connected in series, were adjusted to pass an approximately 10-Hz band centered at the "target" frequency (173 for males; 212 Hz for females). The signal was led to the counter-timer which digitally displayed the frequency of the filtered wave. For each speaker, one acceptable sample of each vowel at each intensity was selected.

Construction of Listening Tapes

To produce data pertinent to the research questions for this study, it was necessary for each acceptable sentence and vowel test sample to be reproduced for listeners at each of three playback intensities representing a 20-dB range. For that purpose, three copies of each test sample were produced by tape dubbing. Two separate listening tapes (one for sentences and another for vowels) were then prepared by splicing all of the test samples (three replicates of each sentence or vowel test production) in random order. Tape I contained 144 vowel samples, while Tape II contained 36 sentence samples. The psychophysical scaling method of magnitude estimation was used to obtain listener ratings of loudness and effort; thus, one test sample of each type (sentence or vowel) which had been produced by one speaker at 75 dB SPL was chosen by the investigator as a "standard" sample. This method of "listening tape" preparation is similar to that used by Brandt _et al_. (8) in a similar study.

Rating Procedure

A total of 23 listeners, all professional staff members or graduate students in speech pathology at the University of Oklahoma Health Sciences Center, independently assessed the recorded samples for loudness and for effort. No single rating session exceeded one and one-half hours duration.

The judgments were made in an acoustically-isolated room with the listeners, no more than six at a time, seated in a semicircle facing the loudspeaker. Two tape recorders, an attenuator, and a monitoring amplifier were placed in the adjoining control room. The listening tapes were reproduced using one tape recorder, while the "standard" tape sample was reproduced using the other. A switch inserted in the circuit allowed the experimenter to select the output of either of the two tape recorders. The output voltage level of the tape recorders was first balanced using the monitoring amplifier. Then, from the switch, the signal was led to the attenuator, which stepped the signal down by 0, 10, or 20 dB, according to the randomization schedule used to prepare each test tape. The attenuated signals were then introduced into an amplifier and boosted by a constant amount so that all of the samples were clearly audible to the listeners. Care was taken not to overdrive the tape recorder amplifiers. Although the relative production intensity differences among test samples were maintained at playback, no attempt was made to duplicate exactly the true production intensities of 65, 75, and 85 dB SPL.

Prior to the rating of the sentence and the vowel samples for loudness and for effort, the listeners were allowed to practice the direct magnitude estimation procedure by rating the durations of segments of a tape-recorded 1000 Hz tone. After the examiner answered questions about the rating procedure, the listeners made judgments of the loudness of each sample. Following a short rest period, they made estimates of the amount of effort used to produce the same samples. In each rating session, the "standard", a sample which had been produced by one subject at 75 dB SPL, was reproduced with 10 dB of attenuation, so that its intensity at playback was in the middle of the range of playback intensities used in this experiment. The "standard", assigned the value of 100, was presented six times before the listeners began the rating task and after every fourth sample to be rated. The listeners rated the magnitudes of all test samples in relation to the standard. Copies of the instructions given to the listeners for the practice rating session and for the loudness and effort rating sessions for sentences and for vowels are included in Appendices A through C.

To obtain a sufficiently large number of loudness and effort ratings for the sentences and for the vowels, four different groups of listeners (no more than six listeners per group) were used. However, a question arose concerning the extent to which the ratings made by the individual groups were comparable to each other. To answer this question, a single "standard" group of three listeners, each having demonstrated satisfactory test-retest reliability in rating one of the groups of samples for loudness or for effort, rated the remainder of the test samples. Spearman rank order correlations (Rho) were computed for the ratings of this "standard" group and each of the other groups of listeners. The magnitudes of the correlations thus obtained, ranging from 0.96 to 0.93, suggested that the four individual groups were comparable to each other, since their ratings were comparable to those of the "standard" group of listeners. The correlation coefficients obtained showing the correlations among ratings from the groups of listeners are presented in Appendix D.

CHAPTER IV

RESULTS AND DISCUSSION

Results

This study was designed to investigate the effects of a number of variables on listener judgments of loudness and effort for selected vowel and sentence samples. Four normal-speaking adults, two females and two males, produced the vowels /ɑ/ and /i/, as well as a single sentence, at three intensities. The vowels were produced normally, with fundamental frequency free to vary as vocal intensity was varied, and then with fundamental frequency controlled at a predetermined level. The sentences and the vowels were recorded onto magnetic tape, then reproduced and randomized onto separate listening tapes. Each sample was played at three different intensities over a 20-dB range. Thus, the relationships among speaker production intensities and the playback levels at which the listeners heard the samples were varied systematically. The psychophysical method of magnitude estimation was used to obtain listener ratings of the loudness of each sample and the amount of effort used by the individual speakers. The intra-class correlation coefficient (11) was utilized to provide an index of the reliability of listener judgments.

Vowels

For each vowel sample, the geometric mean was obtained (over all listeners) for each of the two response variables (loudness and effort). Possible differences among the geometric means were initially examined using an analysis of variance with factorial arrangement of treatments, in which the main factors were production intensity, playback intensity, and subjects. Separate analyses were performed for each test vowel (/ɑ/ or /i/) in each fundamental frequency condition (frequency varying with alterations in production intensity or held constant at a predetermined level). The results of these initial statistical analyses are shown in Appendix D. The analyses revealed large subject-by-production intensity interactions. The decision was made, therefore, to examine the data on a subject-by-subject basis. Another preliminary statistical analysis suggested that loudness and effort ratings were not different for subjects of different sex. Therefore, in further analysis of the data, the factor of speaker sex was not considered.

In a further evaluation of the listener ratings, an analysis of variance with factorial arrangement of treatments was utilized in which the factors were production intensity, playback intensity, vowel (/ɑ/ or /i/), and frequency condition (frequency varying with alterations in production intensity or held constant at a predetermined level). Separate analyses were performed for each subject for loudness and for effort. The pooled sums of squares for all interactions of greater than two factors were used as an estimate of experimental error. A significance level of 0.05 was selected for this investigation.

The findings of this study are organized and presented according to the following major headings: (a) reliability of loudness and effort ratings, (b) loudness ratings, and (c) effort ratings. Within the discussions of loudness ratings and effort ratings, the effects of vowels, frequency conditions, production intensity, and playback level are considered.

In reporting the findings relating to playback level effects for both the vowel and the sentence data, reference is made to playback levels of 65, 75, and 85 dB. These level notations do not refer to the actual SPL at which the samples were played back for rating by the listeners. Rather, they designate only the relative playback levels of the samples. During playback of the samples to the listeners, an effort was made to present all samples within a comfortable listening range. Thus, a sample played back at "75 dB" should be regarded as being at a level 10 dB greater than a sample played back at "65 dB", and so on. No absolute playback levels are intended in these presentations.

Reliability of listener ratings. Table 1 presents, separately for loudness and for effort ratings, estimates of interjudge reliability obtained using the intraclass correlation procedures described by Ebel (11) using the formula which adjusts for between-judge variance. The overall reliability of listener ratings was 0.78 for loudness and 0.75 for effort. Reliability coefficients of 0.82 and 0.80 were obtained for loudness and for effort ratings, respectively, for those vowel samples which were replayed for listeners at playback levels which corresponded to the original speaker production intensities. That is, samples produced at 65 dB SPL were the least intense at playback, while samples produced at 85 dB SPL were the most intense at playback, and so on.

TABLE 1

INTRA-CLASS CORRELATION COEFFICIENTS FOR LISTENER RATINGS OF LOUDNESS AND OF EFFORT USED TO PRODUCE VOWEL SAMPLES

	Loudness	Effort
Overall	0.78	0.75
Playback Intensity and Production Intensity Covary	0.82	0.80
Playback Intensity Varies, Production Intensity Constant	0.82	0.82
Production Intensity Varies, Playback Intensity Constant	0.58	0.54

Reliability coefficients of 0.82 were found both for loudness and for effort ratings of those vowel samples, at each production intensity, which were replayed to listeners at each of the three playback levels. In this particular analysis, responses to a sample produced at 75 dB SPL, for example, and replayed at all three playback levels, were examined. Analysis of responses to those vowel samples which were produced at all three production intensities but replayed at constant playback levels resulted in reliability coefficients of 0.58 for loudness ratings and 0.54 for effort ratings. It should be noted that three different analyses were performed of the same group of ratings (loudness or effort) to obtain these reliability coefficients.

The magnitudes of the reliability coefficients indicate that ratings are more reliable when listeners hear playback level differences among the samples, regardless of the relation of the playback level to the original production intensity. Reliability was lower, however, for ratings of samples which did not present listeners with intensity differences at playback.

<u>Loudness ratings</u>. The results of statistical analyses of the loudness ratings of the individual vowel samples are presented, independently for each of the four speakers, in Tables 2 through 5. Examination of these tables shows that production intensity and playback intensity main effects are significant for all speakers. In addition, the vowel main effect is significant for Speaker A, while the frequency condition main effect is significant for Speakers B, C, and D. For Speaker A, the vowel-by-production intensity interaction is significant, while for Speakers B and C the frequency condition-by-production intensity interaction is significant. All other main effects and interactions are not significant at the 0.05 level of confidence.

Significant main effects and interactions for production intensity and playback level are illustrated graphically (see Figures 3 through 12). To illustrate the effect of production intensity, without regard to playback intensity, loudness ratings ($\log_{10}$) of the vowel samples averaged over the three playback levels are presented for each speaker as functions of the original production intensities (Figures 4 through 7). Similarly, loudness ratings ($\log_{10}$) of vowels averaged over the three production levels are presented as functions of relative playback level, providing a graphic display of the effect of playback level, without regard to original production intensity (Figures 8 through 12). Those graphs presenting production intensity effects will also be used to demonstrate interactions involving vowels, production intensities, frequency conditions, and production intensities. In all of the

TABLE 2

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS OF VOWELS PRODUCED BY SPEAKER A

Source	df	ms	F
Production Intensities (A)	2	0.0141	4.2857*
Playback Intensities (B)	2	0.6769	205.7447*
AB	4	0.0040	1.2158
Vowels (C)	1	0.0666	20.2432*
AC	2	0.0279	8.4800*
BC	2	0.0021	< 1
Frequency Condition (D)	1	0.0018	< 1
AD	2	0.0051	1.5500
BD	2	0.0027	< 1
CD	1	0.0066	2.0061
Residual	64	0.00329	

* $p \leq 0.05$

TABLE 3

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS OF VOWELS PRODUCED BY SPEAKER B

Source	df	ms	F
Production Intensities (A)	2	0.1935	58.8146*
Playback Intensities (B)	2	0.5760	175.0760*
AB	4	0.0034	1.033
Vowels (C)	1	0.0002	< 1
AC	2	0.0018	< 1
BC	2	0.0001	< 1
Frequency Condition (D)	1	0.0324	9.848 *
AD	2	0.0228	6.930 *
BD	2	0.0006	< 1
CD	1	0.0018	< 1
Residual	64	0.00329	

* $p \leq 0.05$

TABLE 4

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS OF VOWELS PRODUCED BY SPEAKER C

Source	df	ms	F
Production Intensities (A)	2	0.0336	10.213*
Playback Intensities (B)	2	0.7062	214.650*
AB	4	0.0068	2.066
Vowels (C)	1	0.0001	< 1
AC	2	0.0012	< 1
BC	2	0.0012	< 1
Frequency Condition (D)	1	0.0225	6.8389*
AD	2	0.0273	8.2979*
BD	2	0.0060	1.8237
CD	1	0.00025	< 1
Residual	64	0.00329	

* $p \leq 0.05$

TABLE 5

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS OF VOWELS PRODUCED BY SPEAKER D

Source	df	ms	F
Production Intensities (A)	2	0.0165	5.015*
Playback Intensities (B)	2	0.7485	227.5076*
AB	4	0.0025	< 1
Vowels (C)	1	0.0003	< 1
AC	2	0.0018	< 1
BC	2	0.0021	< 1
Frequency Condition (D)	1	0.1368	41.580 *
AD	2	0.0027	< 1
BD	2	0.0006	< 1
CD	1	0.0006	< 1
Residual	64	0.00329	

* $p \leq 0.05$

figures, Frequency Condition I refers to the condition in which the fundamental frequency of the vowel was allowed to vary with changes in speaker production intensity. In Frequency Condition II, fundamental vocal frequency was held constant at a predetermined level.

Vowel effects on loudness ratings. Tables 2 through 5 show that the vowel main effect is significant for one of the four speakers (Speaker A). Examination of the mean loudness ratings for this speaker, averaged over playback levels and plotted as a function of production intensity (see Figure 4), shows that /ɑ/ tends to be rated louder than /i/. For Speaker A, the interactions involving vowels with playback intensities and frequency conditions were not significant, suggesting that the loudness difference between the two vowels, indicated by the significant vowel main effect, holds similarly in the two fundamental frequency conditions and at the three playback intensities. However, a significant vowel-by-production intensity interaction exists, indicating that the relationship between the means for the two vowels differs at two or more production intensities. This significant interaction will be examined further later in this section.

Fundamental frequency effects on loudness ratings. Tables 2 through 5 also show that a significant frequency main effect is found for three of the four speakers (Speakers B, C, and D), but not for Speaker A. For Speakers B and C, however, significant frequency condition-by-production intensity interactions are found, suggesting that the relationship between the means in the two frequency conditions differs from one production intensity to another. This significant interaction is described later in this section. The absence of significant interactions involving frequency conditions for Speaker D suggests that the differences in the mean loudness ratings for the two frequency conditions, described in the frequency main effect, obtains for this speaker at all production and playback intensities and for the two vowels. Figure 7 shows for Speaker D that the mean ratings in Frequency Condition I, where fundamental frequency was free to vary with alterations in production intensity, exceed those in Frequency Condition II where fundamental frequency was controlled at a single level.

Production intensity effects on loudness ratings. The production intensity main effects are significant for each of the four speakers, suggesting that listeners made use of cues provided by this factor in rating the loudness of the vowel samples. The means involved in the production intensity main effects for each of the four speakers are plotted in Figure 3. Figure 3 shows a trend toward increased loudness ratings with increased production intensity for each speaker. For three of the four speakers (A, C, and D), each 10 dB increment in production intensity results in relatively small increments in rated loudness. Speaker B evidences the greatest loudness increment from the lowest (65 dB) to the highest (85 dB) production intensity.

The absence of significant interactions involving production and playback levels suggests that the relationship between the mean loudness ratings for the various production intensities evidenced in the production intensity main effect, does not differ significantly at the playback levels employed in this experiment.

The presence of significant interactions between vowel and production intensity for Speaker A and between frequency conditions and

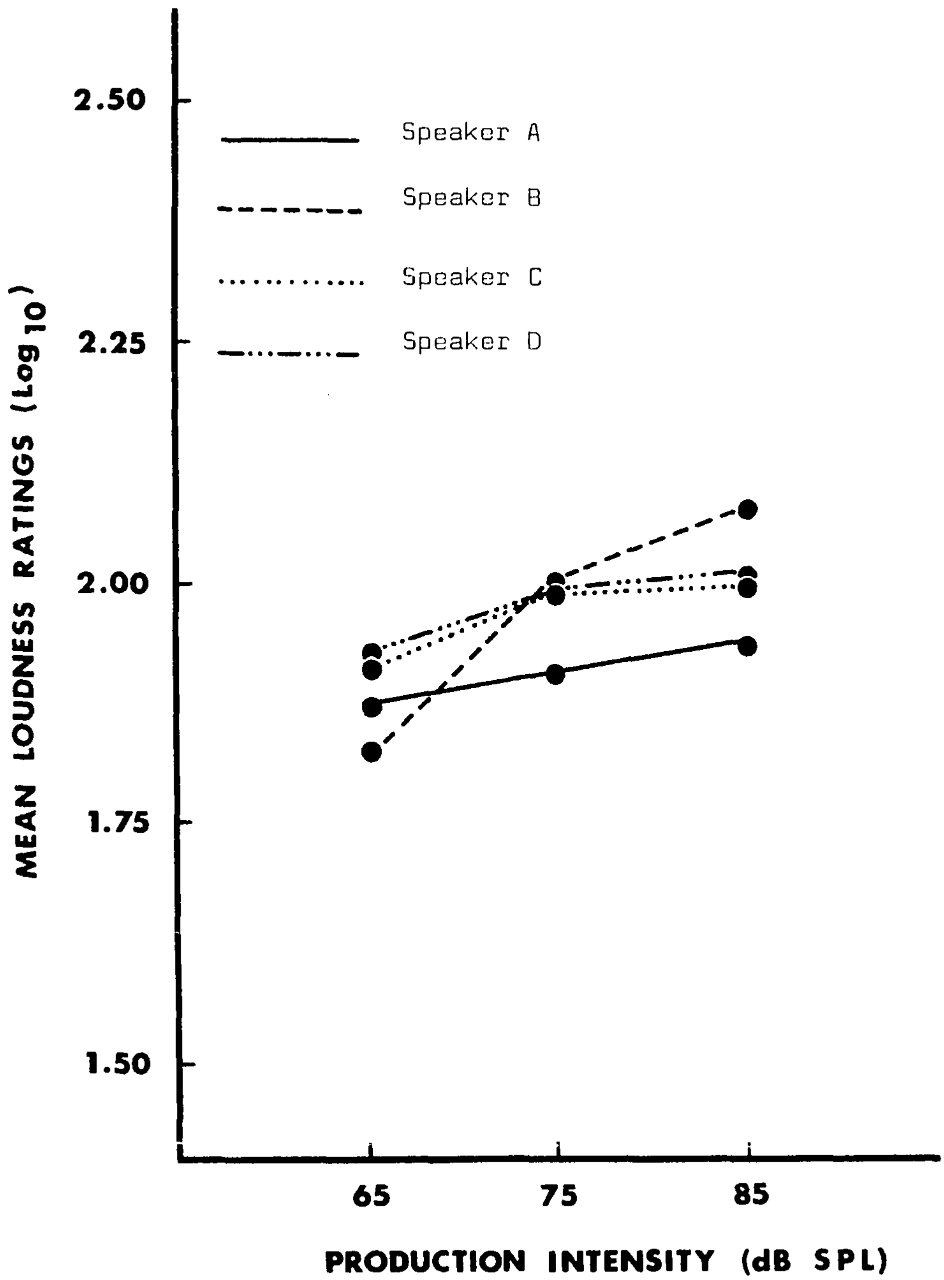

Figure 3.--Mean loudness ratings ($\log_{10}$) for Speakers A, B, C, and D, at each of three production intensities. The geometric means of listener ratings have been averaged over vowels, frequency conditions, and playback levels.

production intensity for Speakers B and C suggests that the effects of alterations in production intensity upon loudness ratings differ for the vowels for Speaker A and for frequency conditions for Speakers B and C.

The significant vowel-by-production intensity interaction for Speaker A is illustrated graphically in Figure 4. This figure presents mean loudness ratings as a function of production intensity, individually for the vowels /ɑ/ and /i/, in Frequency Conditions I and II. Inspection of the plotted means reveals that the mean loudness ratings for Speaker A, derived over all three playback levels, vary as a function of the vowel and the intensity at which it was produced. Figure 4 shows that the vowel /ɑ/ is judged to be louder than /i/ at two intensities, 65 dB and 75 dB, while the two vowels are judged to be about equally loud at 85 dB. The trends within these data also suggest that, for the vowel /ɑ/, increases in production intensity tend to result in a slight increase in loudness from 65 dB to 75 dB and a decrease in loudness ratings from 75 to 85 dB. For /i/, loudness ratings changed little from 65 dB to 75 dB, but increased from 75 dB to 85 dB. For Speakers B, C, and D (Figures 5 through 7) the effects of alterations in production intensity on loudness ratings do not appear to be related to the vowel produced.

As previously reported, significant frequency main effects are found for three of the four speakers studied, Speakers B, C, and D. For two of these speakers, B and C, a significant frequency condition-by-production intensity interaction exists. These significant interactions are displayed graphically in Figures 5 and 6. Inspection of these figures indicates that the effects of frequency condition on loudness

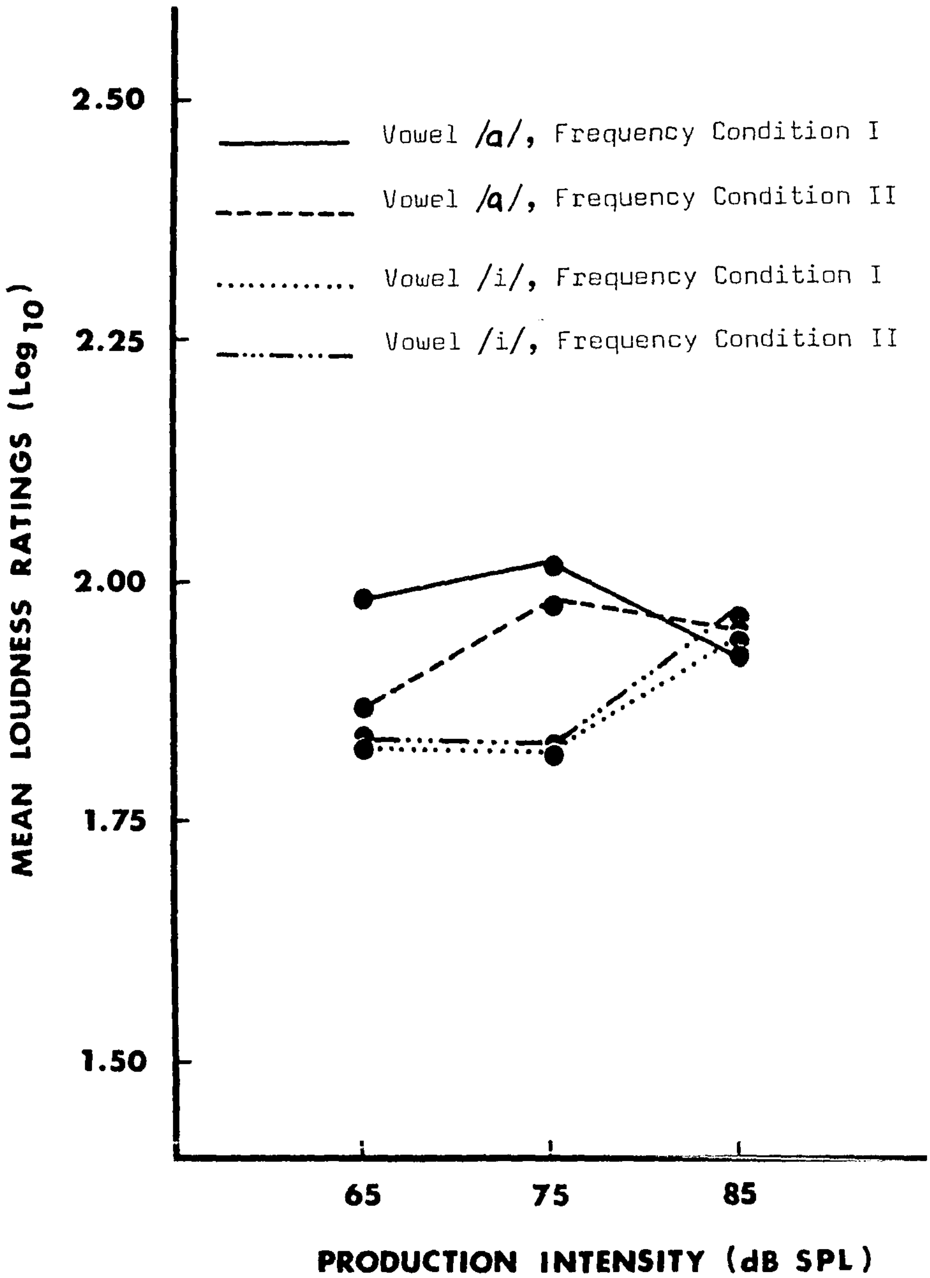

Figure 4.--Mean loudness ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker A under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

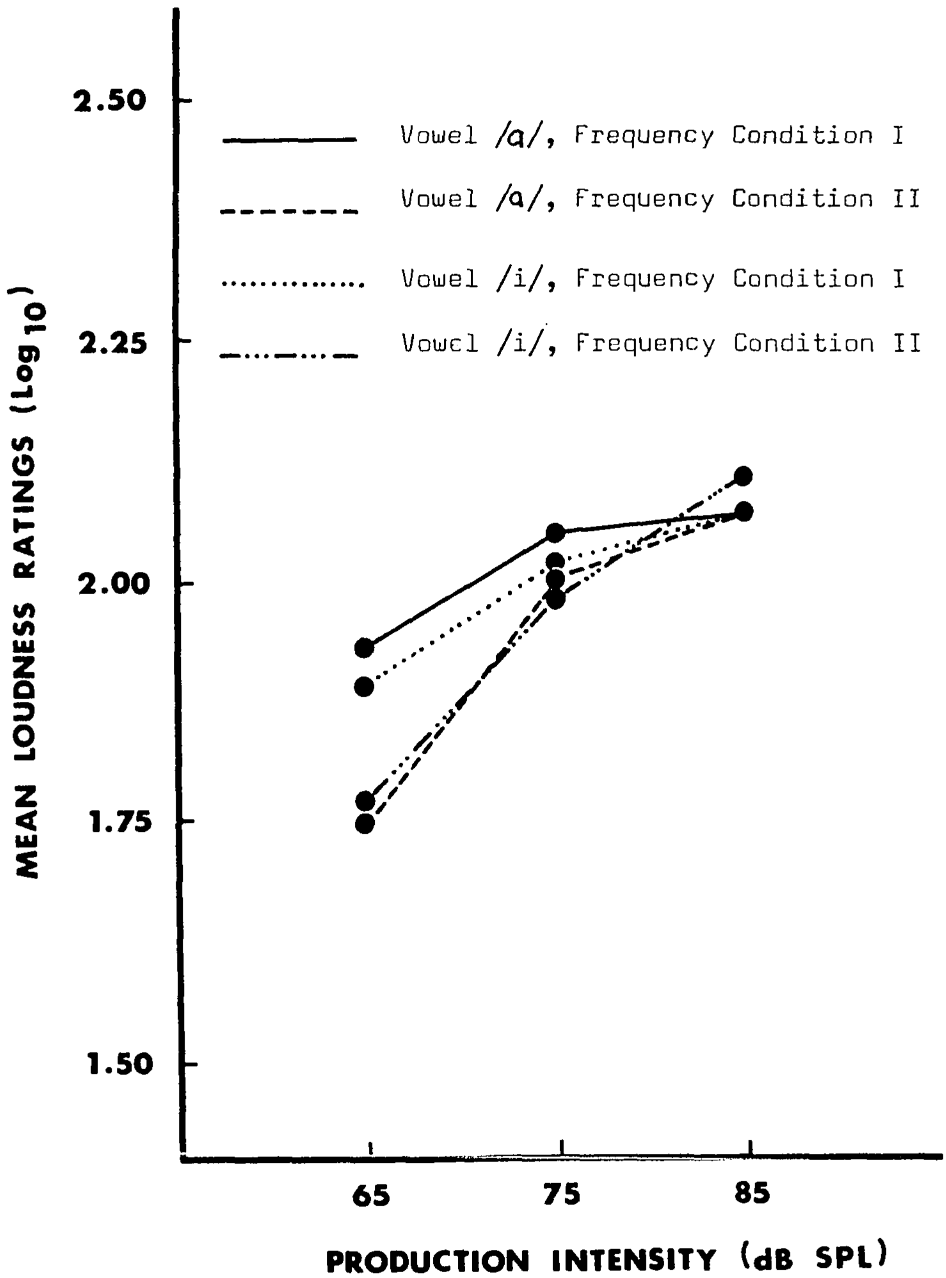

Figure 5.--Mean loudness ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker B under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

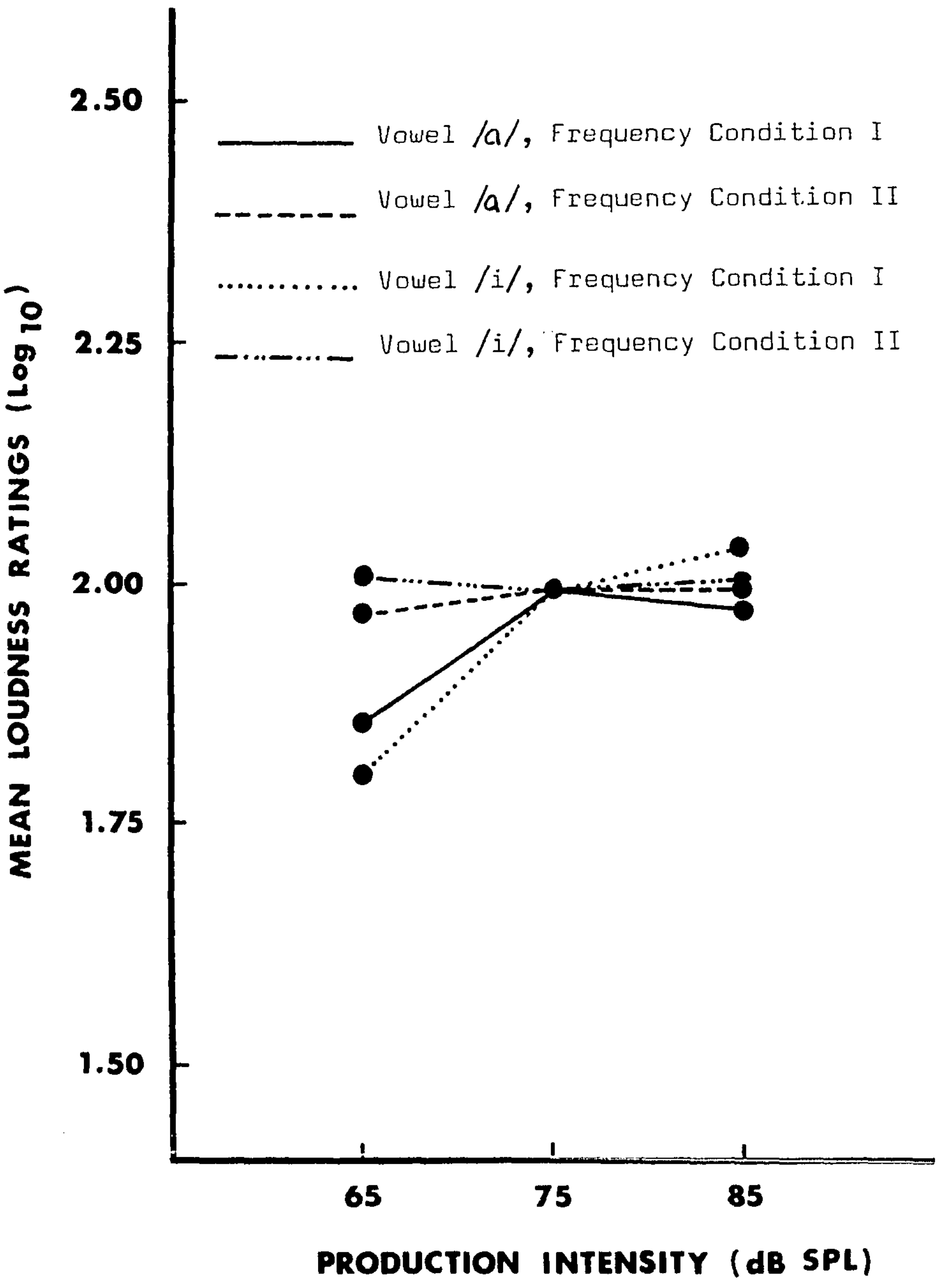

Figure 6.--Mean loudness ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker C under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

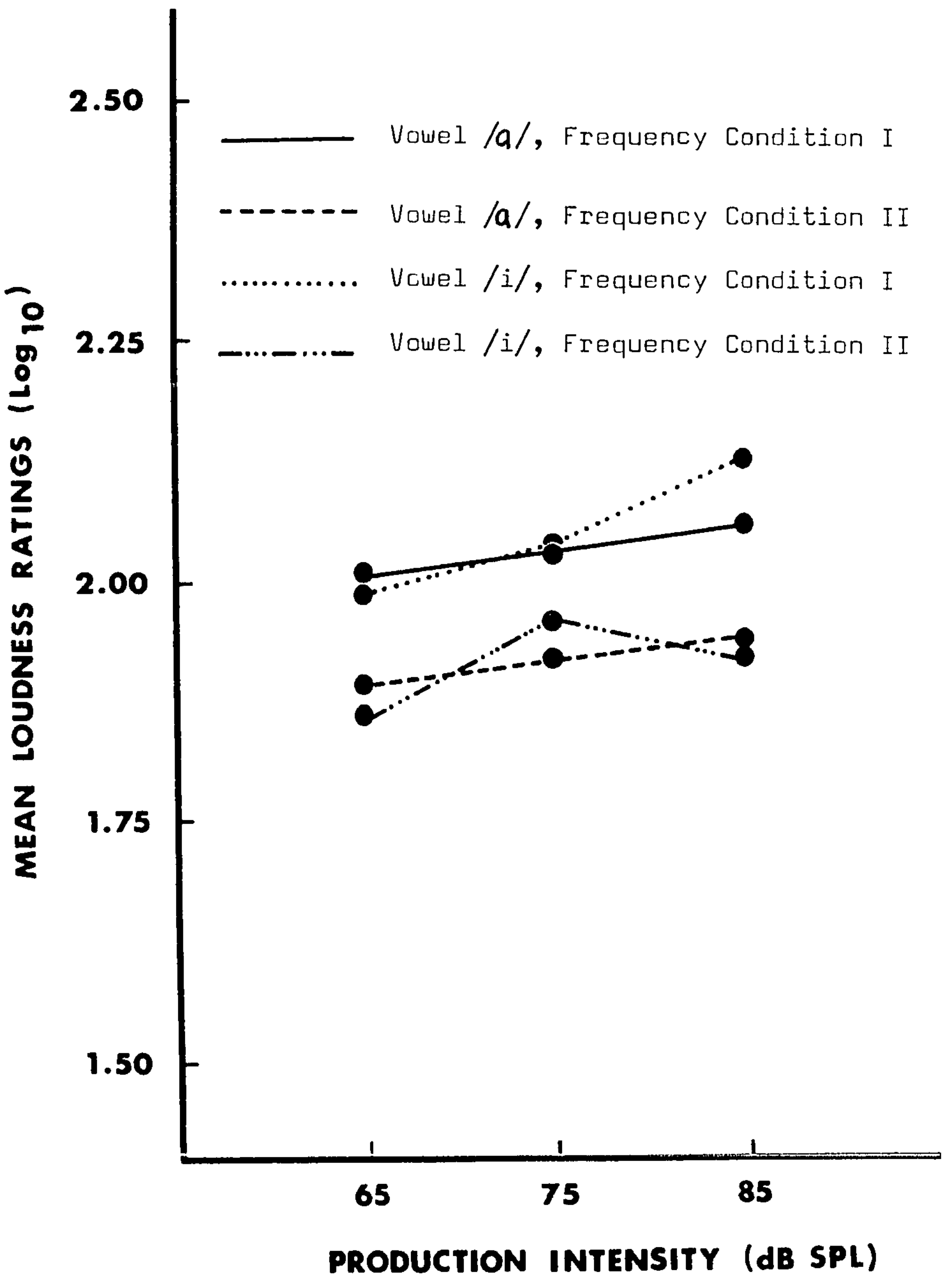

Figure 7.--Mean loudness ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker D under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

ratings appear primarily at the 65 dB production intensity. At the 75 and 85 dB levels, no consistent pattern of difference is apparent between frequency conditions.

It is interesting that, while Speakers B, C, and D exhibit frequency condition effects, these effects are not the same for the three speakers. For example, at the 65 dB production intensity level, vowels produced by Speaker B under conditions where fundamental frequency was free to vary with production intensity changes (Frequency Condition I) were rated louder on the average than the same vowel produced with fundamental frequency constant (Frequency Condition II). At the 65 dB level, however, vowels produced by Speaker C were rated louder in Frequency Condition II, than in Frequency Condition I. For Speaker D (Figure 7), vowels produced in Frequency Condition I were rated louder than in Frequency Condition II at all production intensities.

From the statistical analyses and the plots of means in Figures 3 through 7, it seems reasonable to conclude that for the present speaker group, increases in production intensity tend to be associated with increases in mean loudness ratings at each of the three playback levels. It seems apparent, however, that the extent to which this is true varies from speaker to speaker. Increments in rated loudness are not consistently seen between each of the 10-dB production intensity increments employed in the study, and in some instances, declines in rated loudness were observed. The effects of vowels and frequency conditions, as well, showed considerable speaker-to-speaker variability, suggesting that caution must be observed in generalizing about relationships between production intensity and rated loudness across speakers.

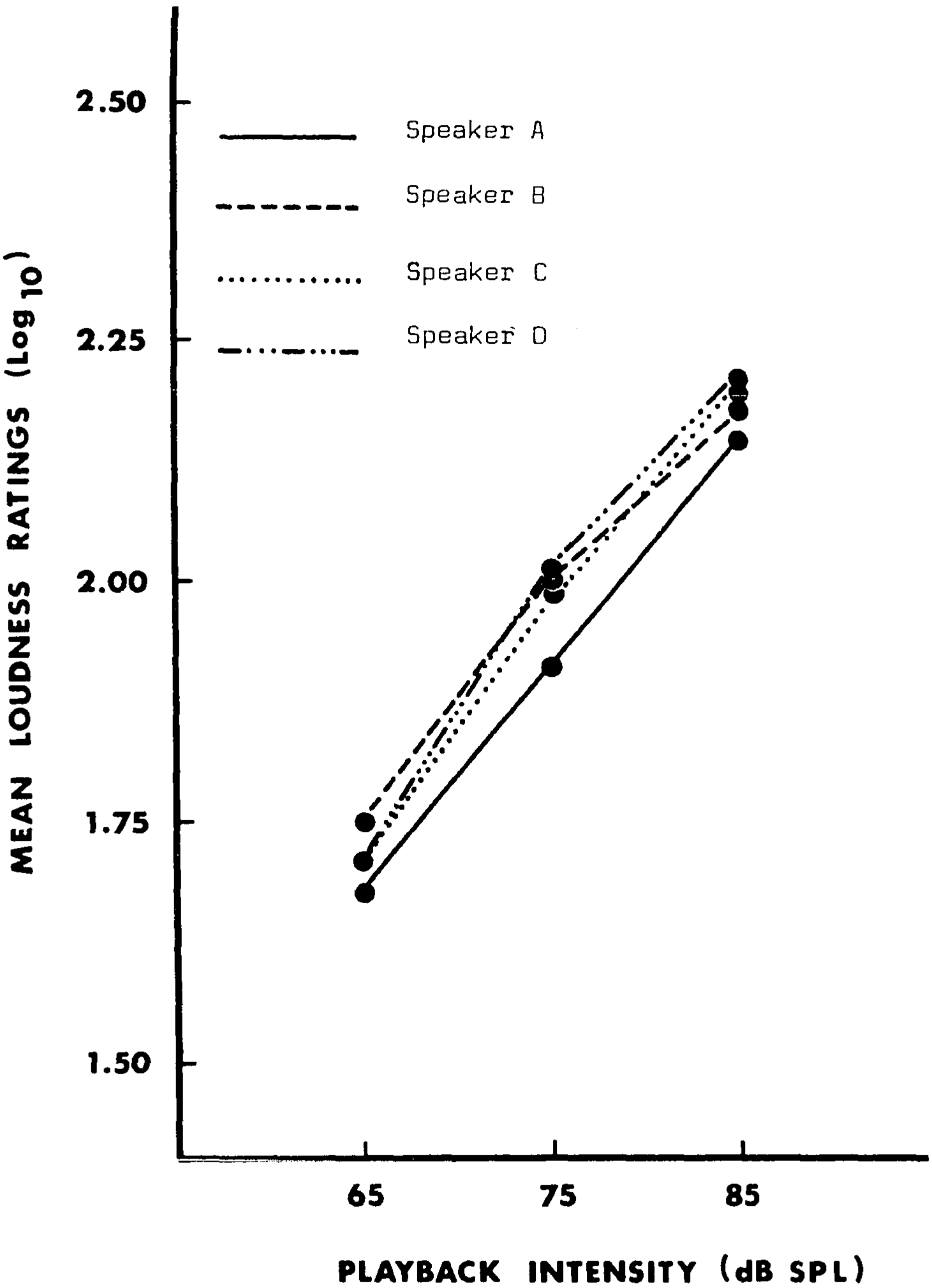

Figure 8.--Mean loudness ratings ($\log_{10}$) for Speakers A, B, C, and D, at each of three playback levels. The geometric means of listener ratings have been averaged over vowels, frequency conditions, and production intensities.

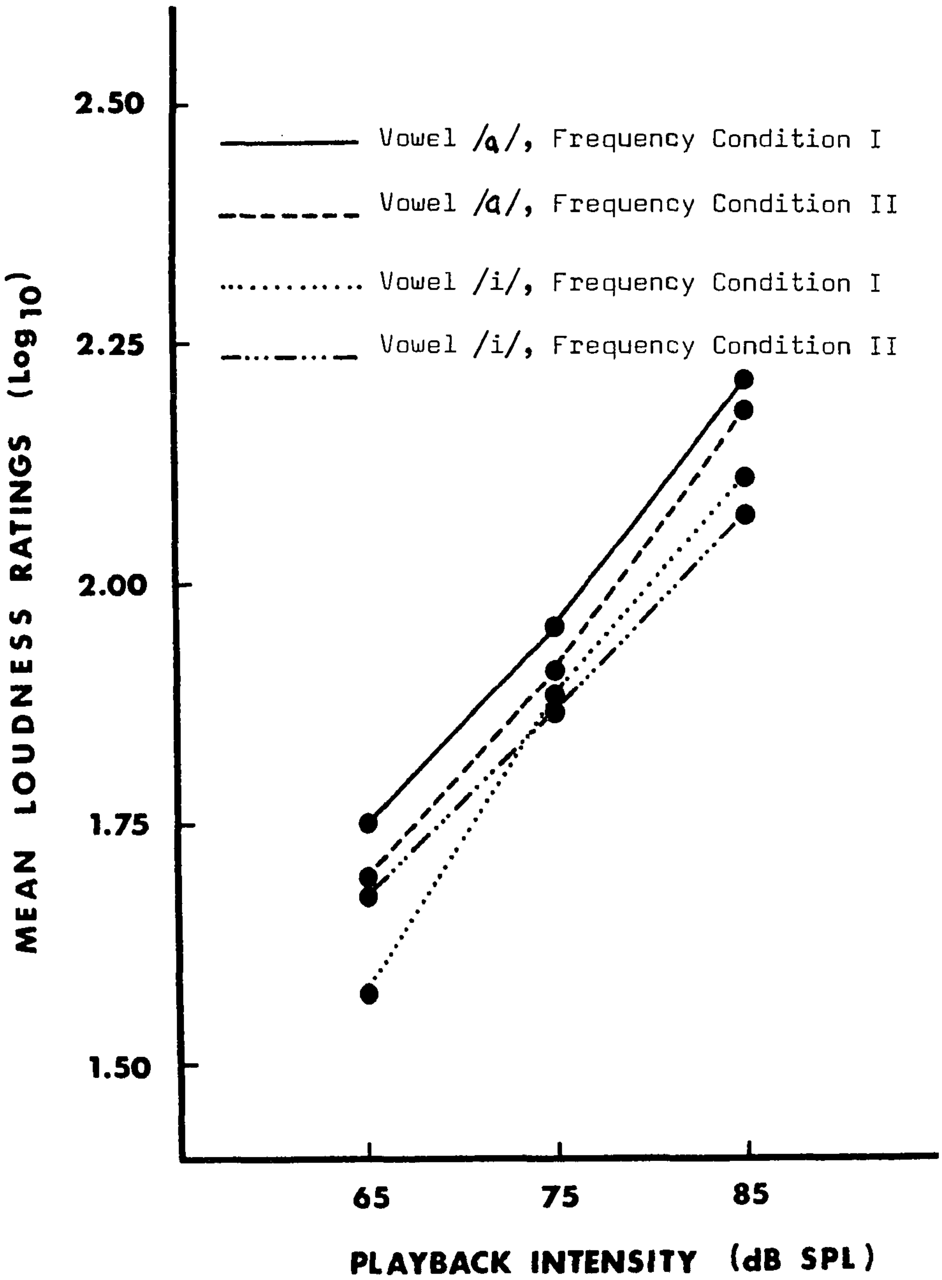

Figure 9.--Mean loudness ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker A under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

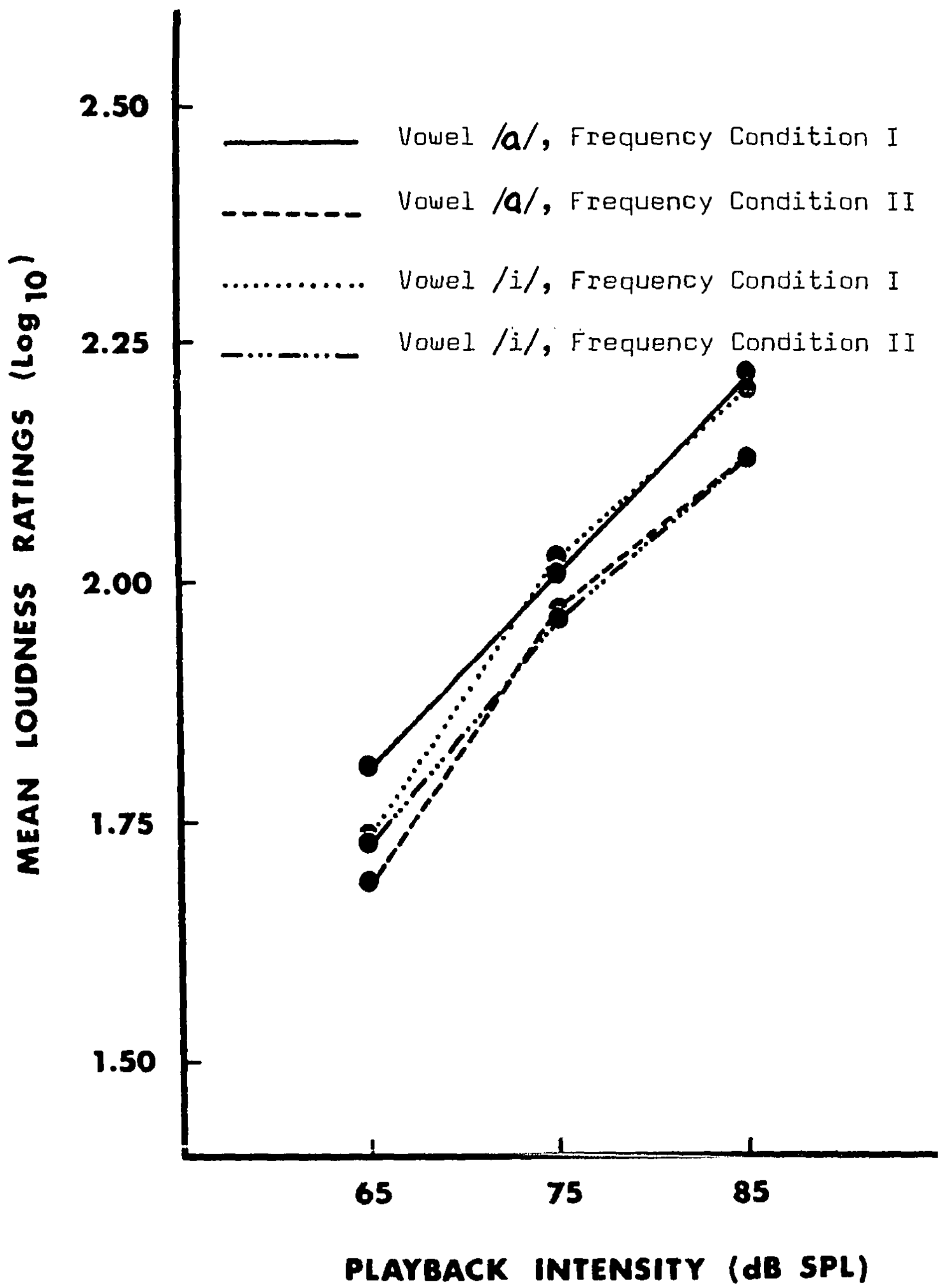

Figure 10.--Mean loudness ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker B under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

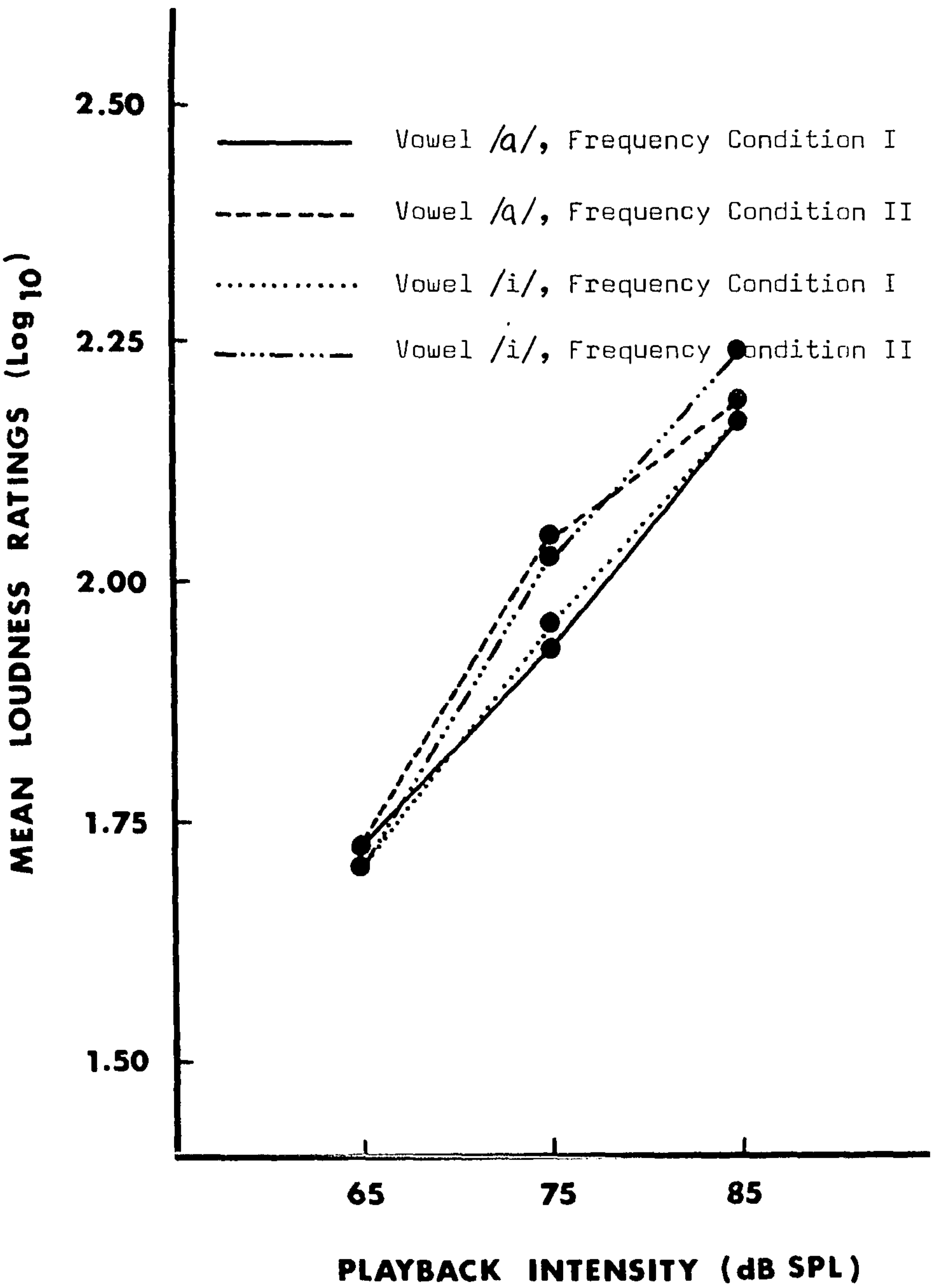

Figure 11.--Mean loudness ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker C under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

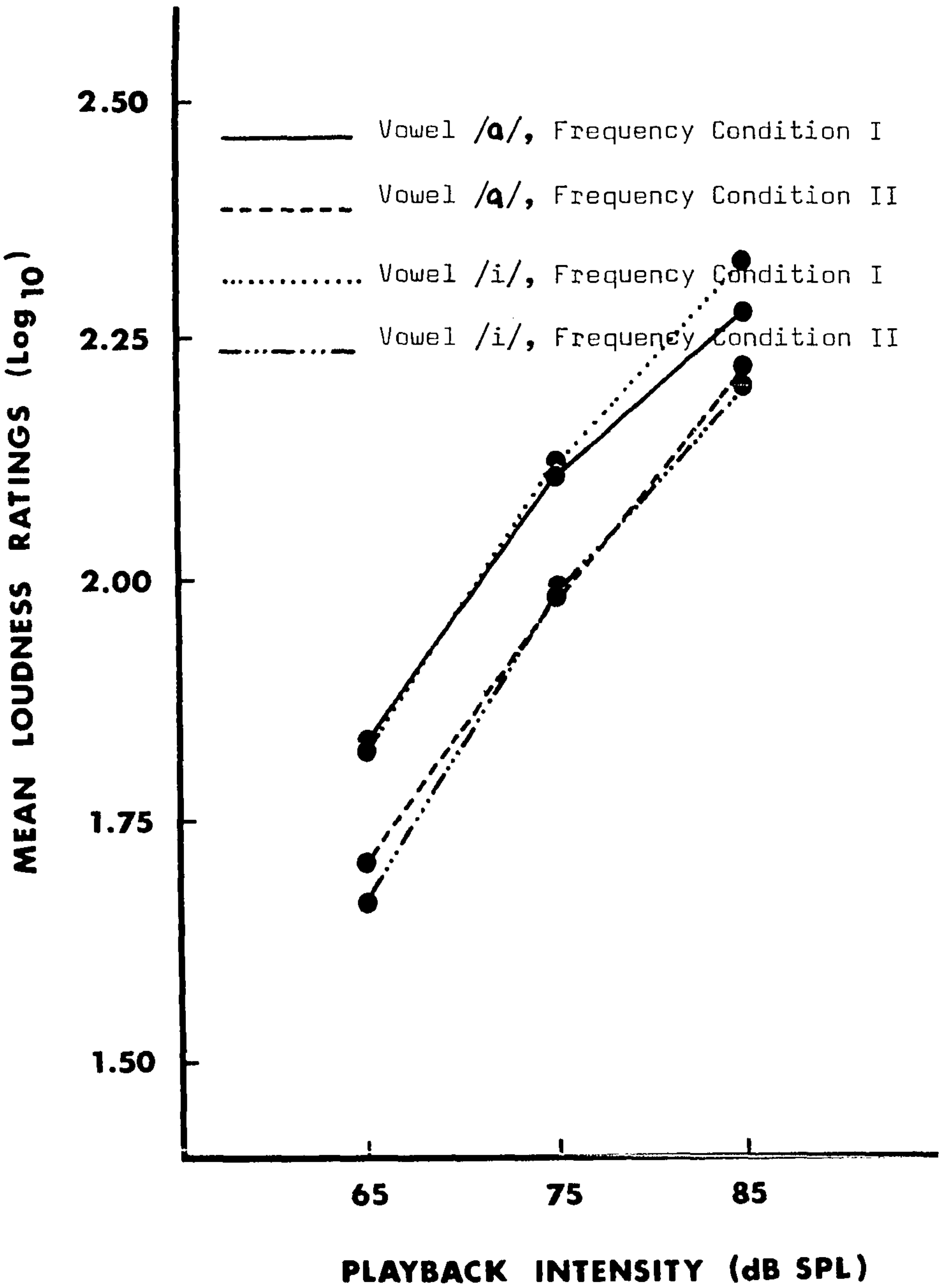

Figure 12.--Mean loudness ratings (log) for vowels /ɑ/ and /i/ produced by Speaker D under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

Playback intensity effects on loudness ratings. Examination of Tables 2 through 5 indicates that the playback intensity main effect is significant for each of the four speakers, and that none of the interactions involving playback intensity, vowels, and frequency conditions is significant. This suggests that the effect of playback intensity on mean loudness ratings is similar for both vowels, in both frequency conditions, and at each production level for each speaker. The significant playback intensity main effect, presented in Figure 8, reveals an almost linear increase in rated loudness for each speaker as playback intensity is increased from 65 dB to the 85 dB level.

From the statistical analysis and the plots of means in Figures 8 through 12, it is reasonable to state that under the conditions of this experiment, the level at which vowel samples are played to listeners is a primary determinant of the loudness of that sample as it is reported by listeners, independent of the vowel produced, frequency condition, or production intensity. The absence of significant playback-by-production intensity interactions for any of the four speakers suggests that the effects of changes in playback levels on the perception of loudness by listeners are similar regardless of the intensity at which the vowel is produced. Comparison of Figures 8 through 12 with Figures 3 through 7 shows that loudness increases more rapidly and more consistently as a function of playback level than production intensity. This suggests that playback level provides more definitive cues to the loudness of vowel samples across speakers than the intensity at which the vowel is produced.

Effort ratings. The results of the analyses of variance of the effort rating data are presented in Tables 6 through 9. Examination of these tables reveals that the production intensity and playback intensity main effects are significant for all four speakers. For Speaker A, the vowel main effect is significant, while for Speakers B, C, and D the frequency condition main effects are significant. The production intensity-by-playback intensity and production intensity-by-frequency condition interactions for Speaker B and the production intensity-by-vowel, playback intensity-by-vowel, and production intensity-by-frequency condition interactions for Speaker C are significant. All other main effects and interactions are not significant at the 0.05 confidence level.

All main effects and interactions will be illustrated graphically following the procedures established in discussions of loudness ratings of these samples. Effort ratings of the vowels averaged over the three playback levels are presented as functions of the intensities at which they were produced by the speakers, providing a graphic display of the effects of production intensity, without regard to the intensities at which the listeners actually heard the samples. Similarly, to illustrate the effects of the three playback intensities without regard to production intensities, effort ratings of the vowel samples, averaged over the three production levels, are presented for each speaker as functions of the intensities at which the listeners actually heard them.

Vowel effects on effort ratings. Tables 6 through 9 show that, as in the previous analysis of loudness ratings, a significant vowel main effect is found for only one of the four speakers (Speaker A). Examination of the mean effort ratings, averaged over playback levels

TABLE 6

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWELS PRODUCED BY SPEAKER A

Source	df	ms	F
Production Intensities (A)	2	0.0444	7.0076*
Playback Intensities (B)	2	0.8962	141.447 *
AB	4	0.0066	1.0417
Vowel (C)	1	0.1011	15.9566*
AC	2	0.0078	1.2310
BC	2	0.0102	1.6099
Frequency Condition (D)	1	0.0252	3.9773
AD	2	0.0078	1.2310
BD	2	0.0003	< 1
CD	1	0.0048	< 1
Residual	64	0.0063	

* $p \leq 0.05$

TABLE 7

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWELS PRODUCED BY SPEAKER B

Source	df	ms	F
Production Intensities (A)	2	0.2646	41.7618*
Playback Intensities (B)	2	0.6666	105.2094*
AB	4	0.2643	41.7144*
Vowel (C)	1	0.021	3.3144
AC	2	0.0006	< 1
BC	2	0.00015	< 1
Frequency Condition (D)	1	0.072	11.3637*
AD	2	0.0234	3.6932
BD	2	0.0051	< 1
CD	1	0.0009	< 1
Residual	64	0.0063	

* $p \leq 0.05$

TABLE 8

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWELS PRODUCED BY SPEAKER C

Source	df	ms	F
Production Intensities (A)	2	0.0810	12.784 *
Playback Intensities (B)	2	0.8919	140.768 *
AB	4	0.0048	< 1
Vowel (C)	1	0.0030	< 1
AC	2	0.0324	5.1137*
BC	2	0.0254	4.0089*
Frequency Condition (D)	1	0.0312	4.924 *
AD	2	0.0729	11.5058*
BD	2	0.0060	< 1
CD	1	0.0003	< 1
Residual	64	0.0063	

* $p \leq 0.05$

TABLE 9

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWELS PRODUCED BY SPEAKER D

Source	df	ms	F
Production Intensities (A)	2	0.08919	12.9263*
Playback Intensities (B)	2	1.1160	176.138 *
AB	4	0.0057	< 1
Vowels (C)	1	0.0000	< 1
AC	2	0.0108	1.6572
BC	2	0.0060	< 1
Frequency Condition (D)	1	0.3604	55.3035*
AD	2	0.0171	2.6989
BD	2	0.0015	< 1
CD	1	0.0141	2.2254
Residual	64	0.0063	

* $p \leq 0.05$

and plotted as a function of production intensity (see Figure 14), shows that effort ratings for /ɑ/ tend to exceed those for /i/. For Speaker A, none of the interactions involving vowels, frequency conditions, and playback intensities is significant, suggesting that the effort relationship between these vowels does not differ significantly in the two frequency conditions or at the various playback levels. Interestingly, Speaker C evidences significant interactions between vowel and production intensity and between vowel and playback level in the absence of a significant vowel main effect. These significant interactions will be examined later.

Fundamental frequency effects on effort ratings. Tables 6 through 9 show that a significant frequency main effect exists for three of the four speakers (Speakers B, C, and D). For Speakers B and D, the absence of significant interactions involving frequency conditions suggests that the relationship between mean effort ratings described in the frequency main effect does not differ significantly for the vowels or in the various production and playback conditions. Inspection of Figure 15 shows that for Speakers B and D, vowels produced in Frequency Condition I tended to receive higher effort ratings than vowels produced in Frequency Condition II. For Speaker C, however, a significant frequency condition-by-production intensity interaction was found, indicating that the relationship between the mean effort ratings for the two frequency conditions does differ significantly at one or more of the production intensities. This significant interaction will be described later.

Production intensity effects on effort ratings. Tables 6 through 9 show that the production intensity main effect is significant for each of the four speakers. The means involved in this main effect for each speaker are presented in Figure 13. Inspection of this figure shows a trend toward increased effort ratings for Speakers B, C, and D as production intensity is increased from 65 dB to 85 dB. For Speaker A, effort ratings increase from 65 dB to 75 dB, but remain relatively constant from 75 dB to 85 dB. Of the four speakers, Speaker B displays the most marked increase in rated effort over this range of production intensities.

For Speakers A and D, interactions involving production and playback intensity, vowels, and frequency conditions are not significant, indicating that the production intensity effect defined for these speakers in Figure 13 does not differ significantly at the various playback intensities, in the two frequency conditions, or for the two vowels.

Speakers B and C, however, evidence significant interactions between production intensity and frequency conditions. Means involved in these interactions are plotted for Speaker B in Figure 15 and for Speaker C in Figure 16. In describing the frequency main effect, it was noted that for Speaker B the mean effort ratings in Frequency Condition I exceeded those in Frequency Condition II. Inspection of Figure 15 suggests that this relationship obtained primarily at the 65 dB and 75 dB production intensities. At the 85 dB level, no consistent difference between the means for the two frequency conditions is apparent. As seen in the description of the frequency condition main effect, Speaker C differs from Speaker B in that mean effort ratings are greater in Frequency Condition II than in Frequency Condition I. Figure 16 suggests

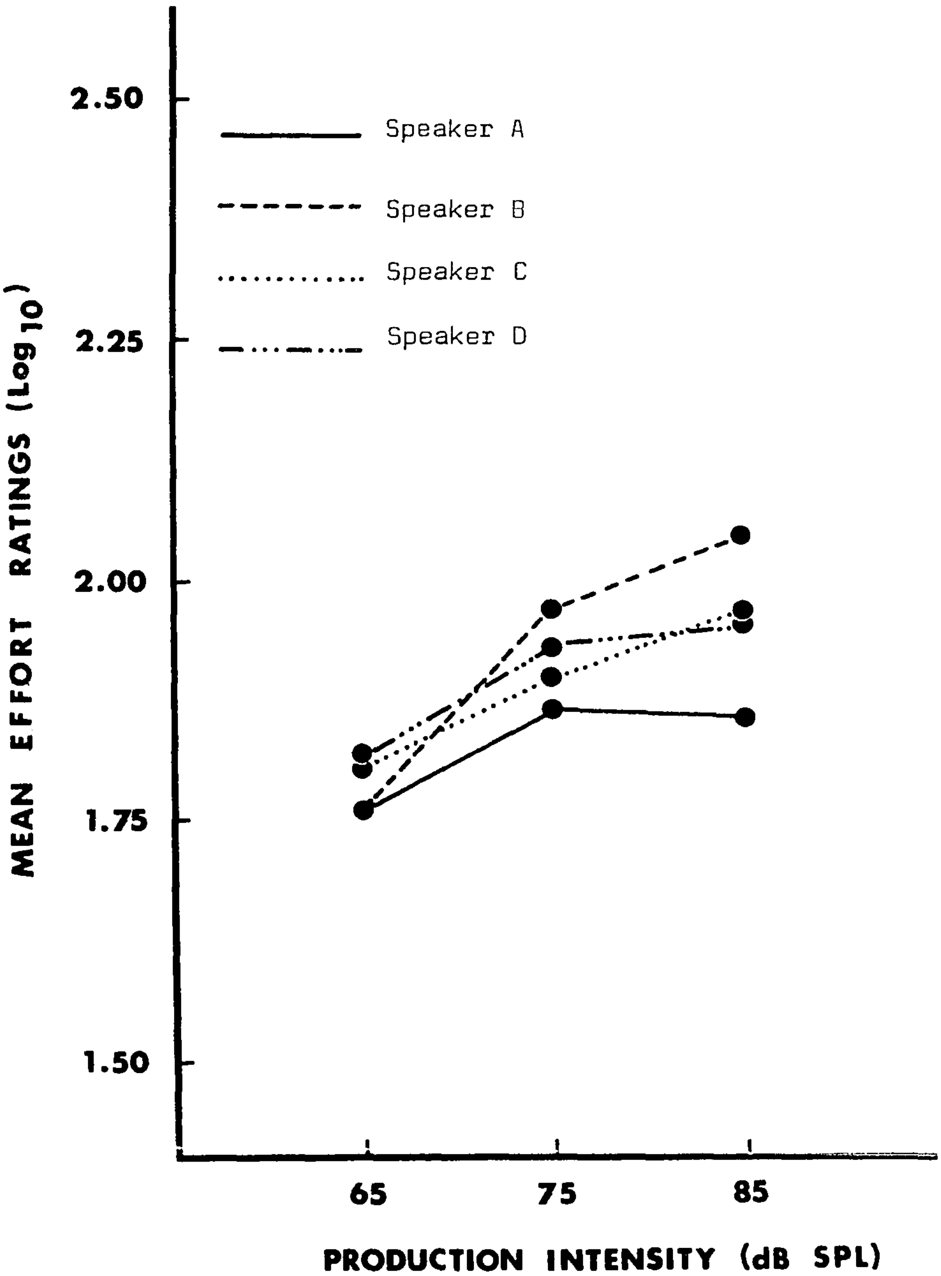

Figure 13.--Mean effort ratings ($\log_{10}$) for Speakers A, B, C, and D, at each of three production intensities. The geometric means of listener ratings have been averaged over vowels, frequency conditions, and playback levels.

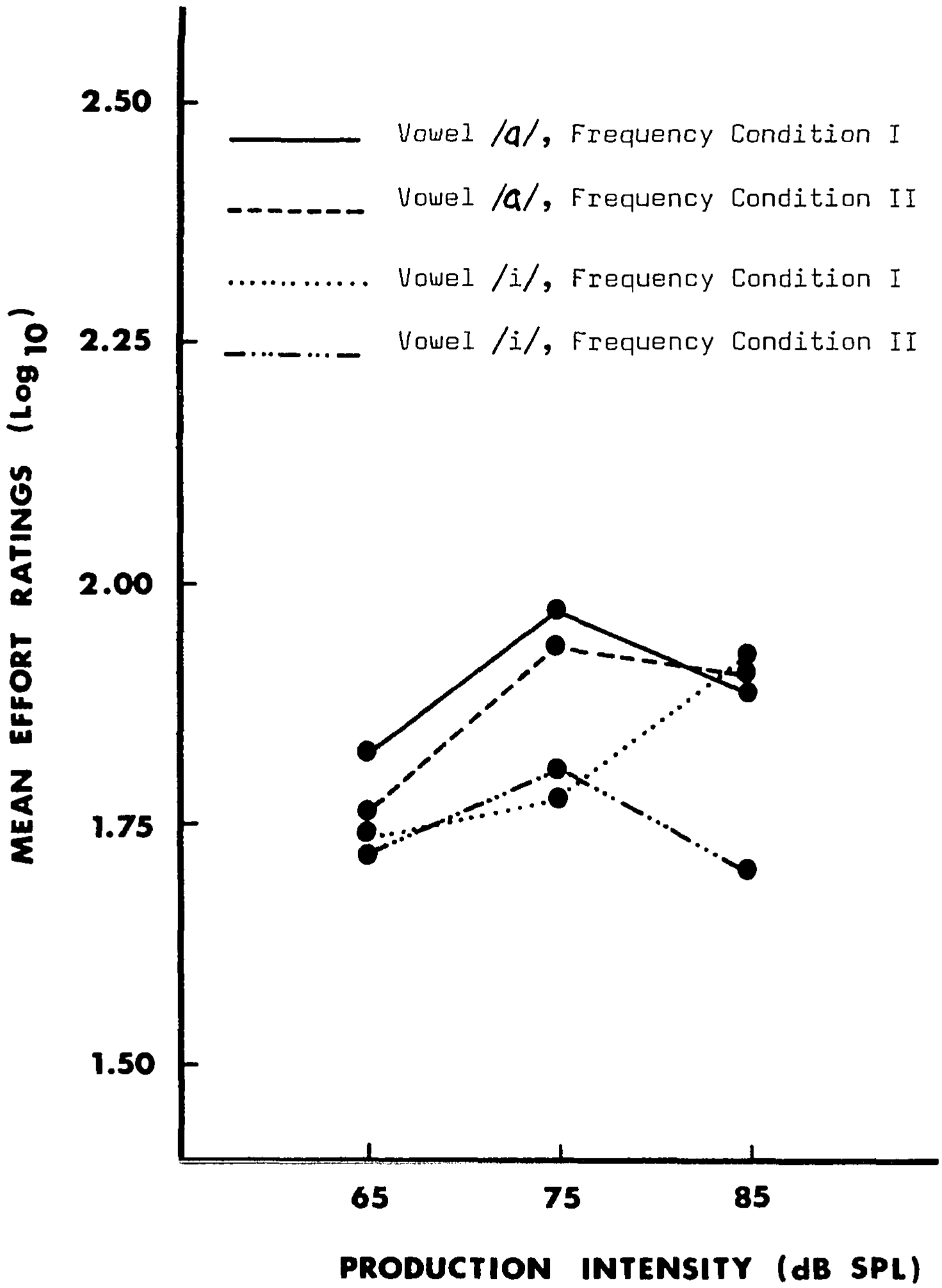

Figure 14.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker A under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

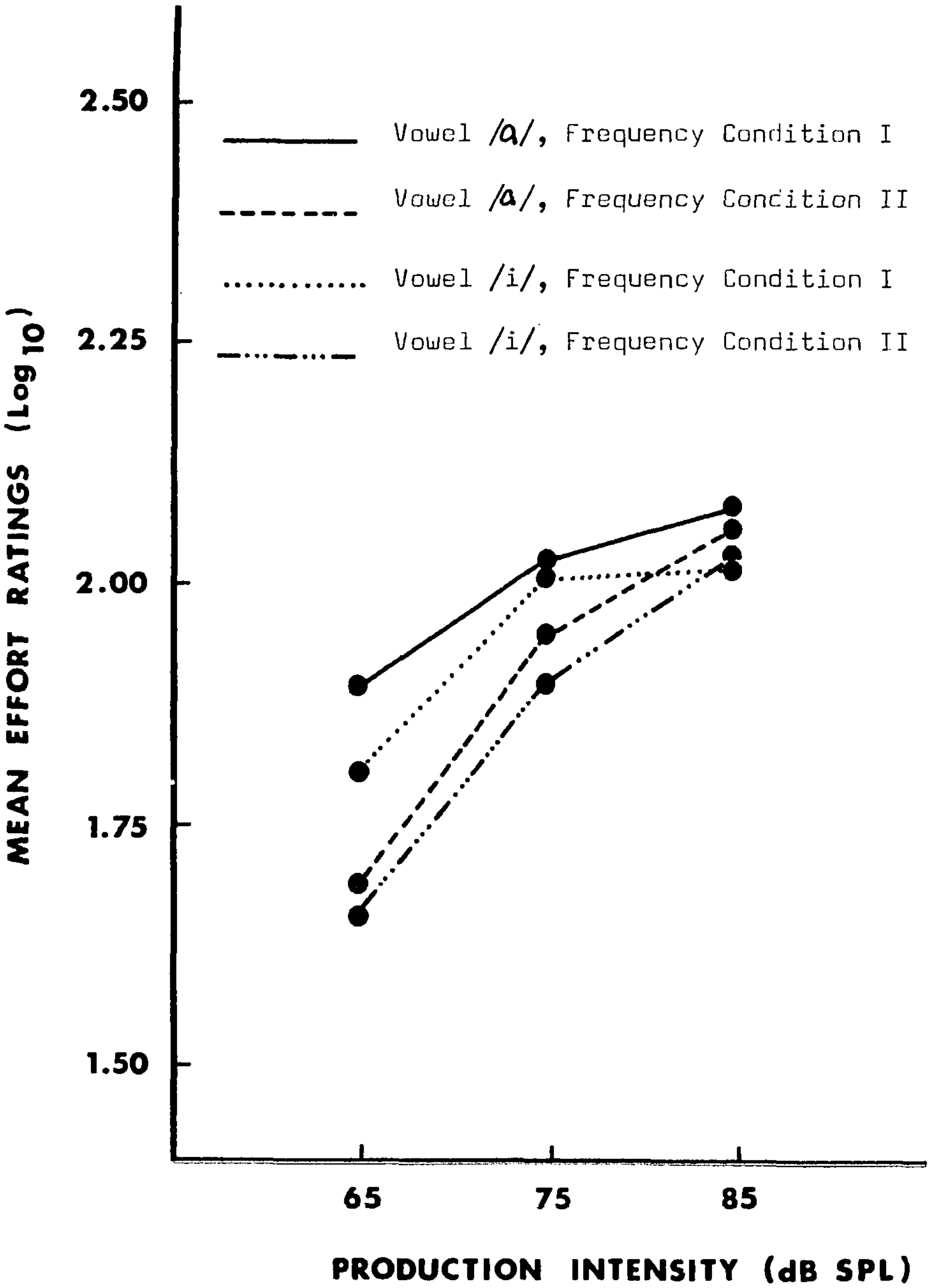

Figure 15.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker B under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

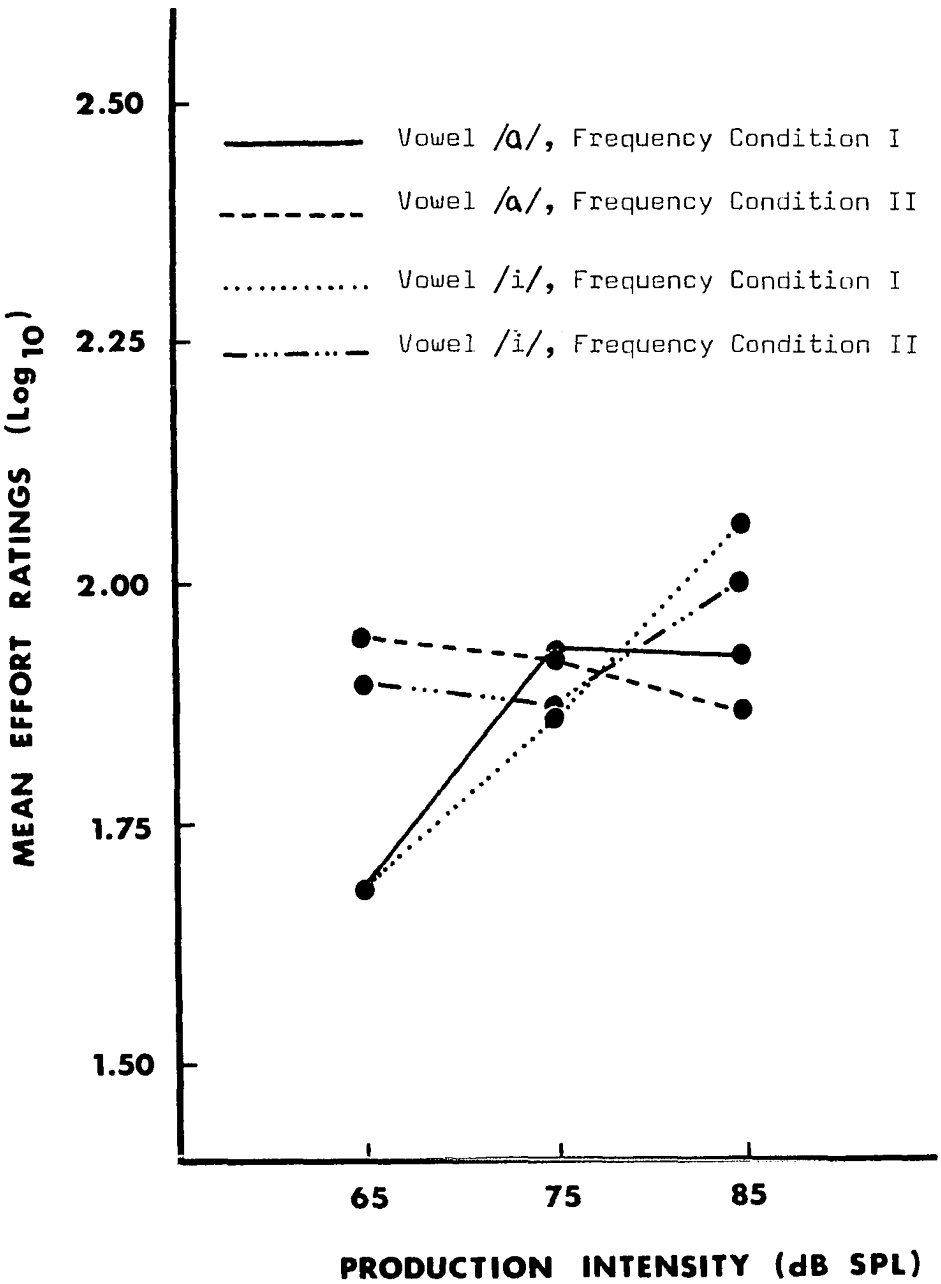

Figure 16.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker C under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

that for Speaker C consistent differences between the means for the two frequency conditions occur primarily at the 65 dB production intensity level; at the 75 dB and 85 dB levels no consistent difference can be observed. In addition, it is apparent that increasing production intensity results in greater increments in rated effort for Speaker C in Frequency Condition I than in Frequency Condition II; for Speaker B, an opposite relationship exists.

Tables 6 through 9 show that, for Speaker C, a significant production intensity-by-vowel interaction exists. Inspection of the plot of means for Speaker C in Figure 16 shows that the vowel /i/ is associated with greater effort than /ɑ/ at the 85 dB level, but that consistent differences in the means for the vowels are not apparent at either the 65 dB or 75 dB levels. It may also be seen that the vowel /i/ shows a somewhat greater increase in rated effort than /ɑ/ as production intensity is increased from 65 dB to 75 dB.

From the statistical analyses and the plots of means in Figures 13 through 17, it seems reasonable to conclude that, for the present speakers, increases in production intensity tend to be associated with increased effort ratings at each of the three playback levels. It is apparent, however, as observed in analyses of loudness ratings, that increments in rated effort are not consistently seen between each 10 dB increment in production intensity and that, in some instances, declines in rated effort occur even though production intensity is increased. A comparison of the plots of effort ratings in Figures 13 through 17 with those in Figures 3 through 7 shows a substantial similarity in the patterns of change in both effort and loudness ratings that occur with

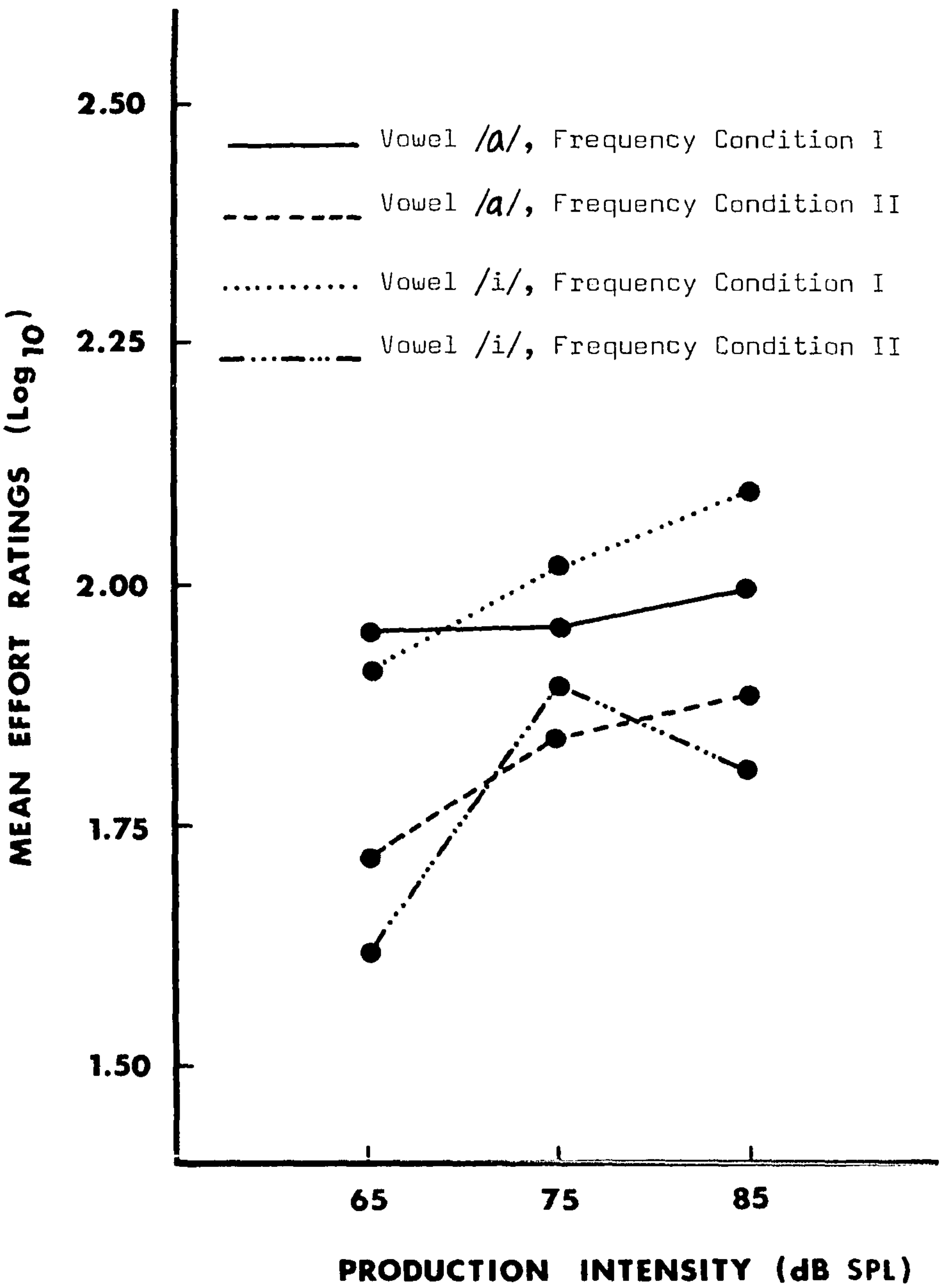

Figure 17.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker D under two frequency conditions. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

changes in production level.

Consistent with the findings for loudness ratings of vowels, considerable speaker-to-speaker variability in effort ratings was noted. It is of interest that the speaker who showed a significant vowel effect for loudness ratings also showed vowel effects for effort ratings (Speaker A). The remaining three speakers showed frequency effects for effort ratings as well as for loudness ratings. Only Speaker C showed a vowel effect for effort ratings but not for loudness ratings.

Playback intensity effects on effort ratings. Tables 6 through 9 indicate that the playback intensity main effect is significant for each of the four speakers. The means involved in this main effect are presented for each speaker in Figure 18. Examination of this plot of means reveals that for all speakers mean effort ratings increase almost linearly as playback intensity is increased in 10 dB steps from 65 dB to 85 dB. For Speakers A and D, none of the interactions involving playback intensity is significant suggesting that relationships between mean effort ratings and playback intersity, shown in Figure 18, do not differ significantly at the various production intensity levels, in the two frequency conditions, or for the two vowels.

Tables 6 through 9 also indicate a significant vowel-by-playback intensity interaction for Speaker C. Examination of the means plotted in Figure 21 shows that while the means for /ɑ/ appear somewhat greater than those for /i/ at the 65 dB level, the means for /i/ tend to exceed those for /ɑ/ at the 85 dB level. No clear pattern of difference is apparent at the 75 dB level. As was seen in the production intensity-by-vowel interaction, greater increments in rated effort appear to occur

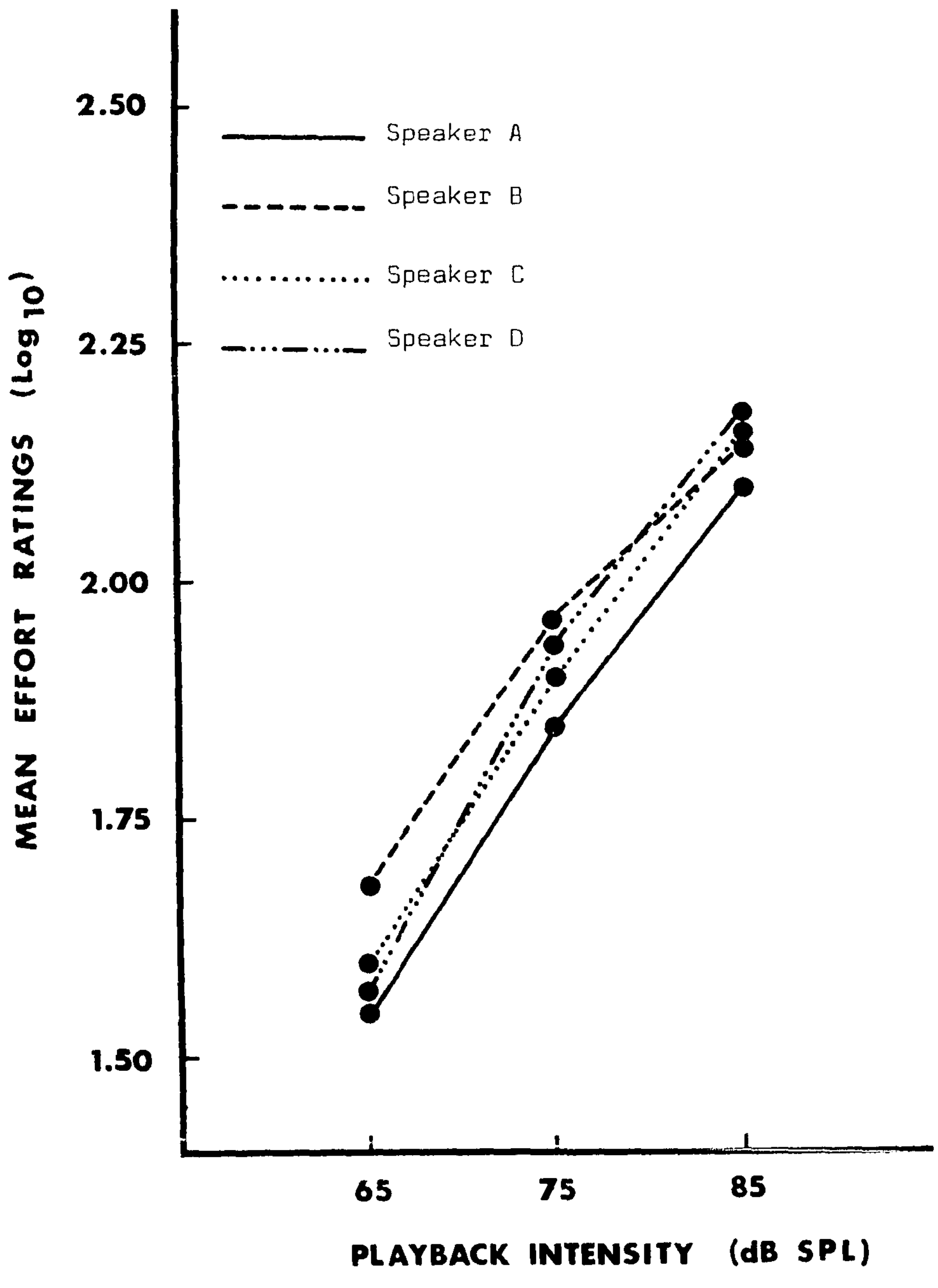

Figure 18.--Mean effort ratings ($\log_{10}$) for Speakers A, B, C, and D, at each of three playback levels. The geometric means of listener ratings have been averaged over vowels, frequency conditions, and production intensities.

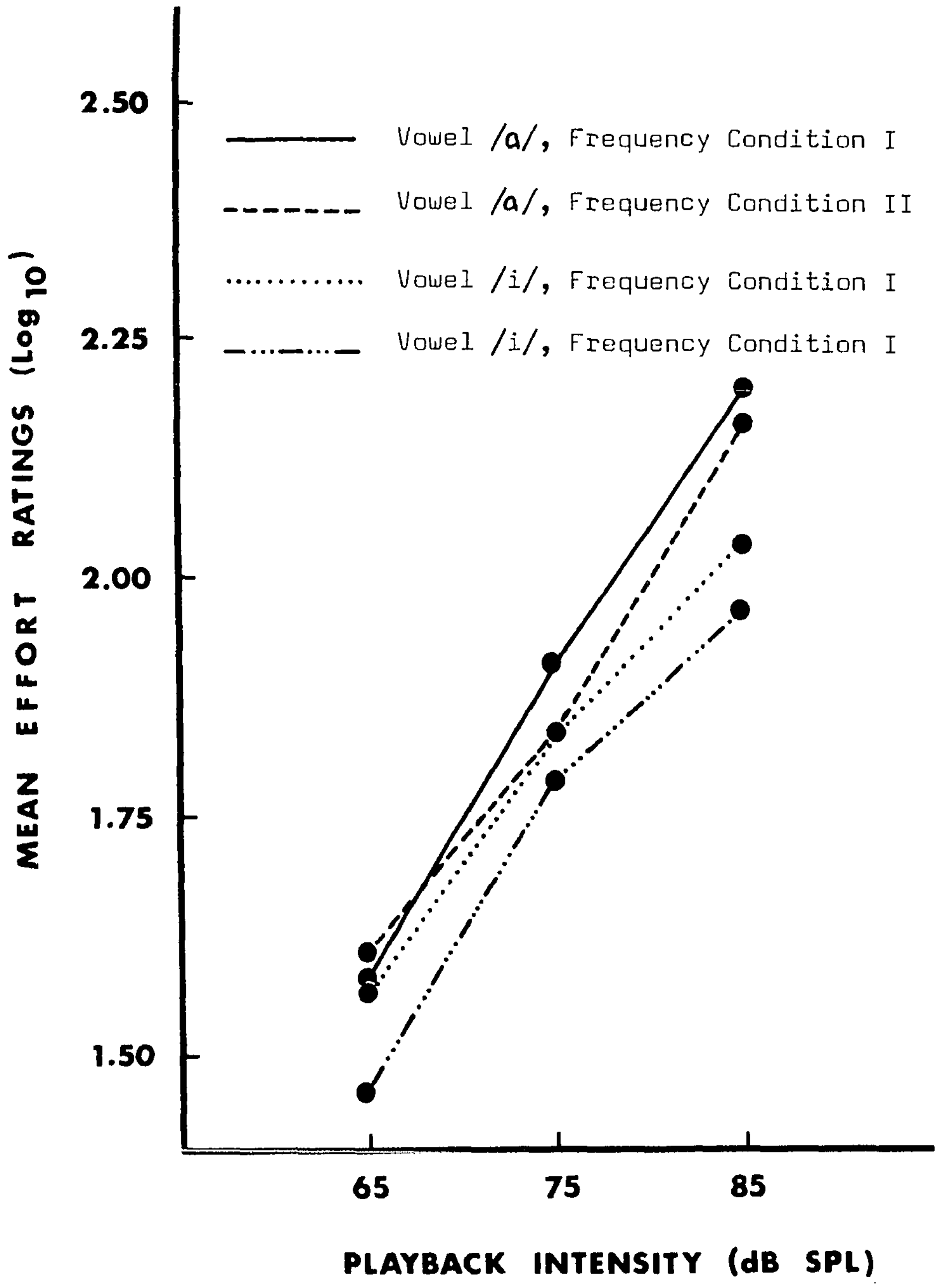

Figure 19.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker A under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

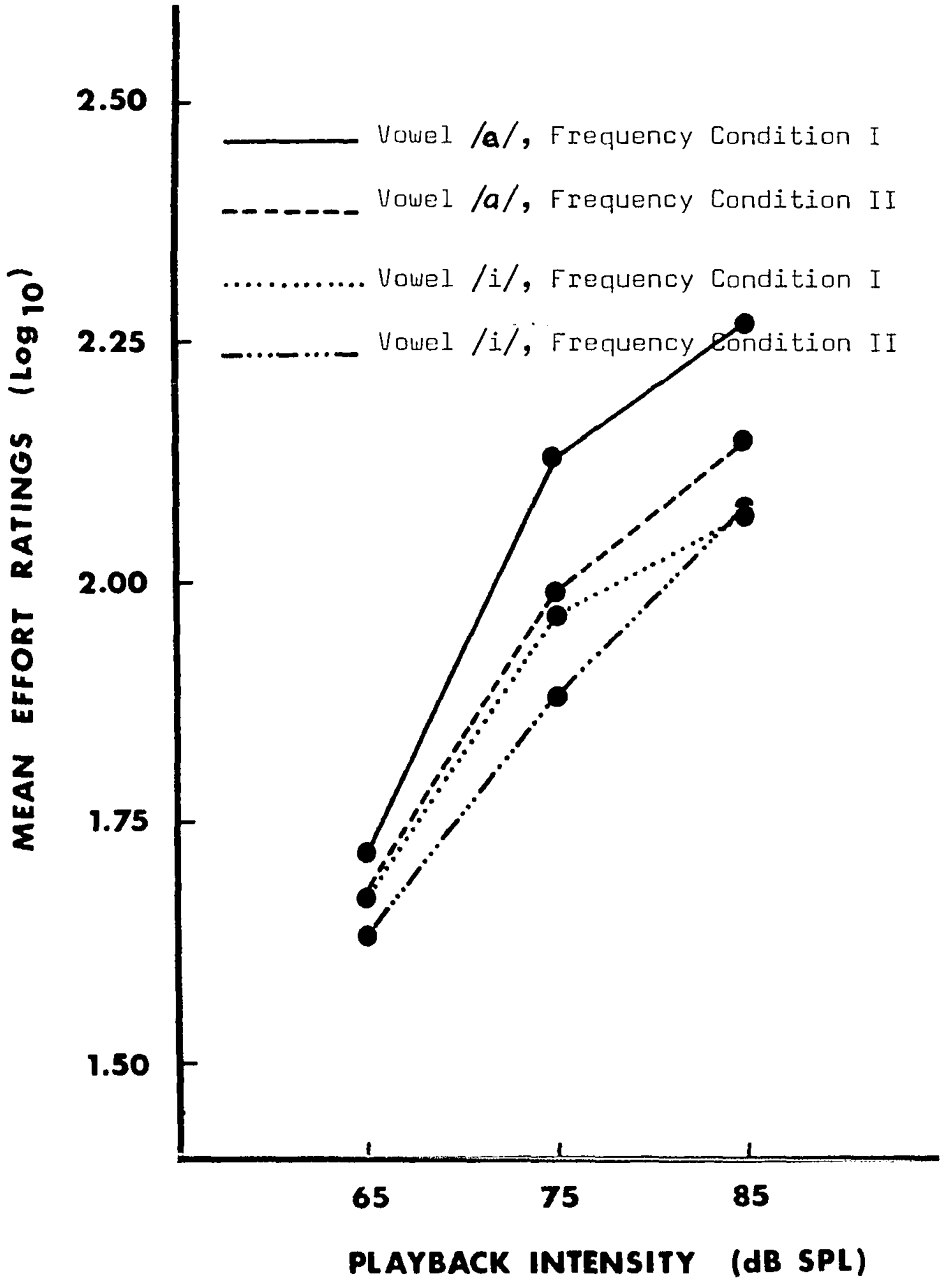

Figure 20.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker B under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

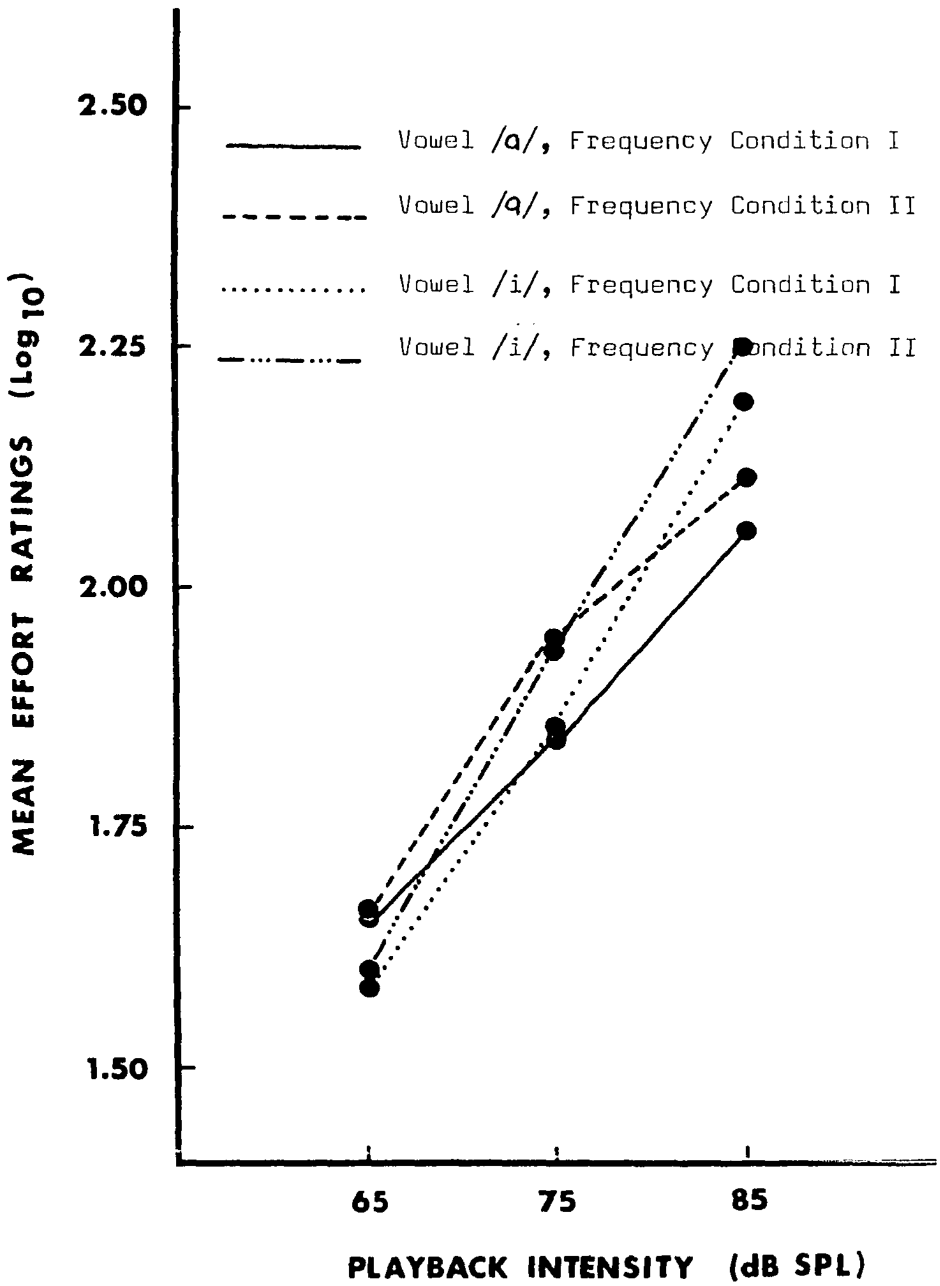

Figure 21.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker C under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

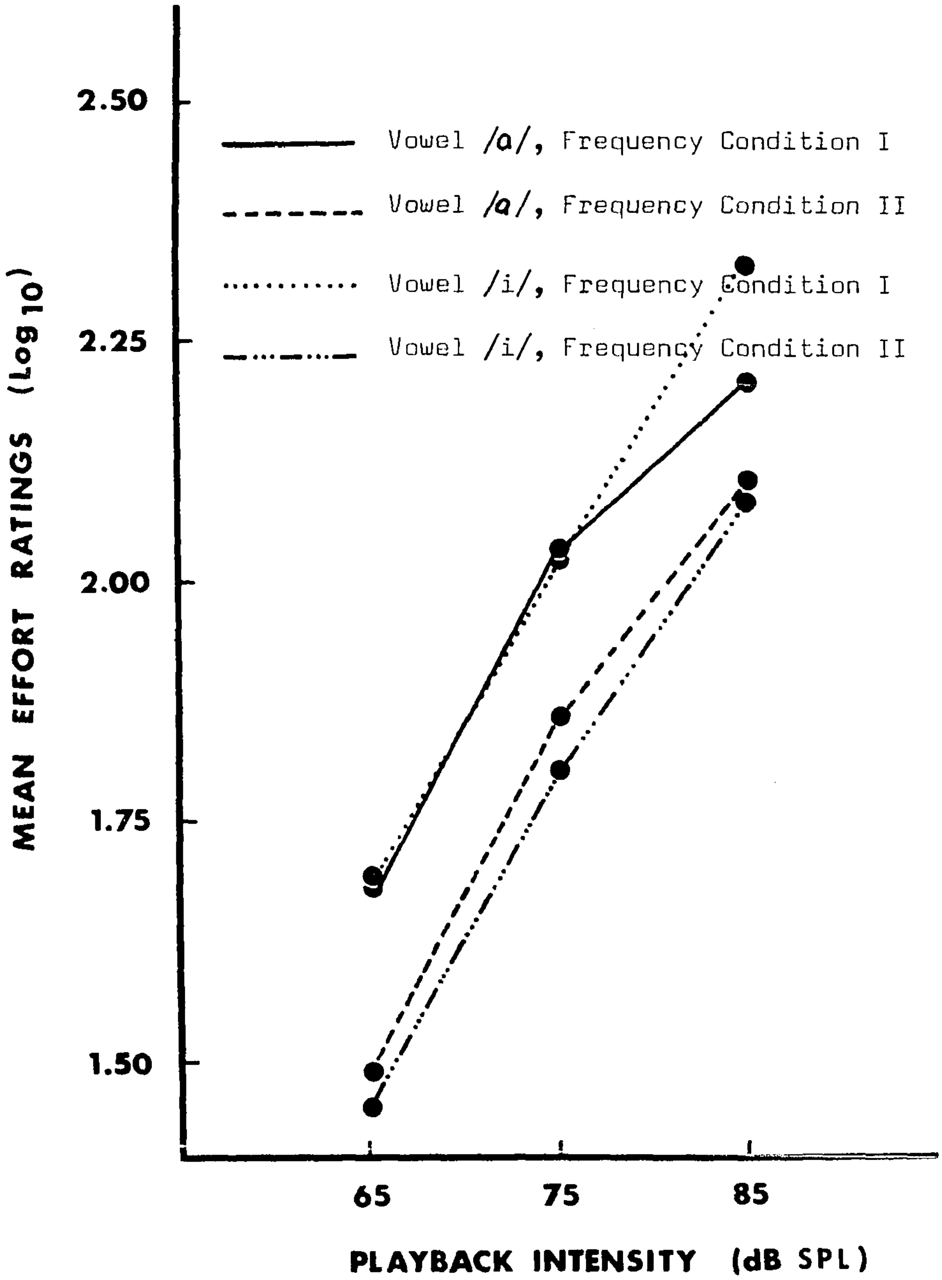

Figure 22.--Mean effort ratings ($\log_{10}$) for vowels /ɑ/ and /i/ produced by Speaker D under two frequency conditions. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

for the vowel /i/ than for /ɑ/ as playback intensity is increased from 65 dB to 85 dB.

In Tables 6 through 9, it may be seen that a significant production intensity-by-playback intensity interaction exists for Speaker B. This finding suggests that there is a differential effect of playback levels on the mean effort ratings assigned at one or more of the three production intensity levels for this speaker. Figure 23 provides a graphic display of the means involved in this interaction. Inspection of this figure suggests that a major source of this interaction lies in the greater difference between mean effort ratings at the 65-dB and 75-dB production levels than between the 75-dB and 85-dB production levels at the lowest and at the highest playback intensities.

The effects of playback level on ratings of effort are similar to those observed in the analysis of loudness ratings; that is, increases in rated effort occur almost linearly with increases in playback level for all speakers. The range of the effort ratings is also somewhat greater than the range of the loudness ratings from the lowest to the highest playback levels. An interesting difference between the analyses of effort and loudness ratings is seen in the presence or absence of interactions involving playback and production intensity. It may be recalled that no interactions involving production and playback intensities were observed in analysis of the loudness data. In ratings of effort, however, one of the four speakers, Speaker B, evidenced a significant interaction between production and playback levels. Similarly, playback intensity was not found to interact with either vowel or frequency condition in the analysis of the loudness data, yet an

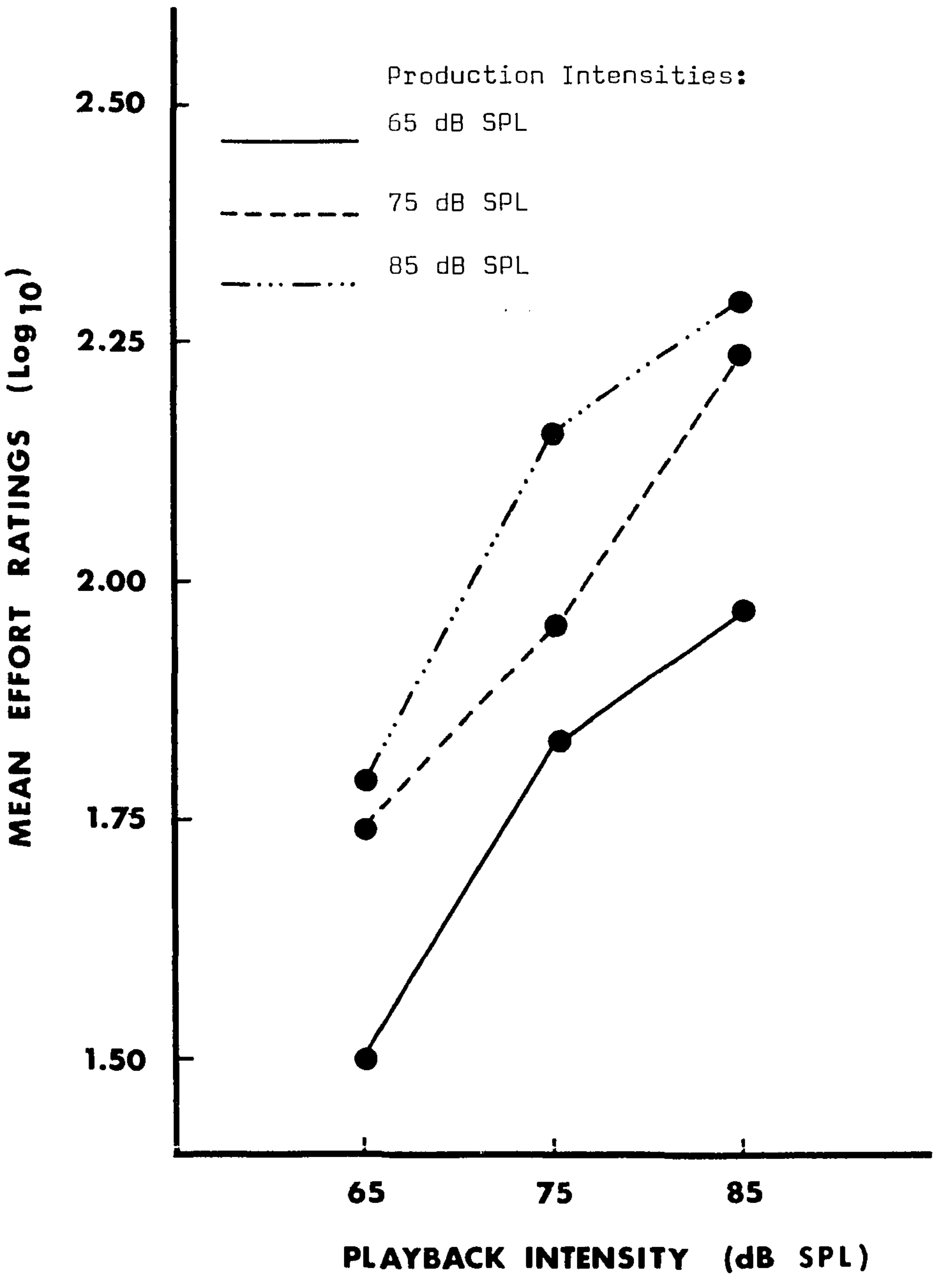

Figure 23.--Mean effort ratings ($\log_{10}$) for all vowels produced by Speaker B. Mean ratings for each of three production intensities are plotted as functions of playback intensity.

interaction between vowel and playback intensity was found in analysis of the effort ratings for Speaker C. Comparison of Figures 18 through 22 with Figures 13 through 17 shows that effort increases more steeply and more consistently as a function of playback level than production intensity. This suggests that playback level provides more definitive cues to the loudness of vowel samples across speakers than the intensity at which the vowel is produced.

Discussion

In the present investigation, effort and loudness rating data were examined separately for each speaker. While the present subject sample included two male and two female speakers, no attempt was made to undertake a detailed study of differences related to speaker sex. This decision was based on a preliminary statistical analysis of the data which indicated that effort and loudness ratings, under the conditions of the experiment, did not vary with sex. The data were analyzed, however, to determine the effect of variations in production intensity, playback level, the vowel produced, and frequency condition (frequency constant or frequent varying) for each speaker.

The present findings support the notion that increases in production intensity are associated with increases in the amount of effort and loudness perceived by listeners. It seems clear, however, that increments in either loudness or effort are not consistently seen for each speaker with each 10 dB increment in production intensity. Further, for both effort and loudness, declines in ratings are sometimes observed even though production intensity is increased. The variability that exists among speakers, as well as that seen from one production intensity to another for an individual speaker, suggests that caution should be observed in generalizing about the relationships among effort and loudness ratings and production intensity.

The findings of the present study relating to isolated vowels indicate that there is considerable variability in effort-loudness relationships from speaker to speaker. This finding was anticipated. It seems reasonable to assume that speakers do not employ identical strategies in regulating vocal intensity. Speakers may vary, for example, in the extent to which they rely on increased respiratory effort, increased vocal tension, and/or vocal tract adjustments in increasing vocal intensity. It is also possible that the extent to which each of these intensity regulating mechanisms is employed may vary in the same speaker from trial to trial. If it is assumed that these physiological events generate the acoustic cues that are available to the judges, there is reason to expect that effort and loudness ratings might differ substantially from speaker to speaker as production intensity is varied. This trend was noted in the data obtained in this experiment.

The present findings are consistent with those recently reported by Scharf (26). Scharf's data suggest that individual speakers do not use identical degrees of effort when phonating at equal intensities, and that speaking effort does not increase uniformly with uniform increments in speaking intensity. Thus, it is quite possible that differences in findings across studies employing single speakers as the subject sample (8, 15, 16, 18, 21, 22) may relate as much to differences among the speakers studied as to differences in experimental method.

Present findings show that increases in playback level resulted in nearly linear increases in rated effort and loudness for all speakers, in contrast to the irregular patterns of change in rated effort and loudness that occurred with changes in production intensity. It is also apparent that the slopes for effort and loudness functions plotted according to playback level are markedly steeper than the slopes for effort and loudness plotted according to production level. These findings suggest that playback level variations provide more definitive cues to and more systematic changes in rated effort and loudness than do variations in production intensity.

The present findings also show that, except for effort ratings for one of the four speakers studied, the effects of playback and production intensity variations on effort and loudness ratings are independent of each other. This suggests that listeners tend to respond, for the most part, to playback level variations without regard to the intensity at which the sample was produced and, conversely, to production intensity variations independent of the playback level at which the sample is heard.

These findings are in substantial agreement with those of Brandt et al. (8) who reported, on the basis of a study in which the speech sample was a sentence, that loudness and effort ratings are affected by the intensities at which samples are heard, as well as by cues generated by differences in production levels, aside from playback level. Lane, Catania, and Stevens (16), however, felt that speech loudness is based solely upon SPL differences among speech samples, a finding which is not completely supported by the results of the present investigation. In contrast to Lane et al., a number of subsequent investigators (1, 5, 18) hypothesized that listeners make judgments of the physiological effort used by a speaker to produce a sample of a given intensity when they estimate speech loudness. These investigators, however, asked listeners to make loudness ratings of speech samples which incorporated a range of SPL differences among them. Thus, even though Ladefoged and McKinney (15) and Allen (1) report the existence of a linear relationship between loudness and a physiological correlate, subglottal pressure, there is no evidence that listeners employed in those studies relied upon acoustic cues other than intensity alone. However, the results of the present investigation, as well as findings reported by Brandt et al. (8), suggest that increasing speaking level results in increased speaking intensity and spectral alterations within the spoken sample itself which can be utilized, independent of each other, by listeners.

The findings of the present study also suggest that consistent vowel effects are not observed across the speakers studied. For only one of the four subjects (Speaker A) did effort and loudness ratings vary significantly according to the vowel produced. In this instance, the vowel /ɑ/ received significantly higher loudness and effort ratings than the vowel /i/. In one other speaker (Speaker C), effort but not loudness ratings differed significantly according to the vowel. In this instance, the vowel effect varied complexly as a function of both production and playback intensity level. These findings differ from those reported by Lehiste and Peterson (18) who report that vowels with lesser intrinsic intensity are rated louder than vowels with greater intrinsic intensity. In their study, it was postulated that the differences in the rated loudness of intrinsically weak and strong vowels relate to the unequal amounts of effort required to produce each type of vowel at a uniform production intensity. In the one instance where a vowel effect was observed in the present study, a vowel with greater intrinsic intensity, /ɑ/, was rated louder and more effortful than a vowel with lesser intrinsic intensity, /i/. The present findings suggest, again, that there may be considerable hazard in extrapolations of findings from studies of single speakers.

In the present study, the effect of frequency condition (frequency constant or frequency varying) was significant for three of the four speakers studied. For two of three speakers, the effect of frequency condition was found to vary significantly as a function of production intensity. These findings suggest that, while rated effort and loudness are influenced by frequency condition (frequency constant or frequency varying) in the present study, the effects of frequency condition are not seen in all speakers studied nor are they the same for all speakers where frequency condition is a significant factor. In a study of a single speaker, Moll and Petersen (22) reported that listeners heard greater changes in vowel loudness when they were permitted to hear frequency differences among vowels produced at different intensities than when vowels were produced at a uniform frequency over the same intensity range. In the present experiment, considering the three speakers who displayed frequency condition effects, only one speaker (Speaker C) displayed steeper loudness and effort functions when frequency was allowed to vary than when frequency was held constant, and this was true only between two of the three production levels. The findings of the present study relating to vowel and frequency condition effects suggest that frequency condition tended to influence effort and loudness ratings more frequently than did the vowel produced. It is apparent, however, that some speakers may evidence a vowel effect without a frequency condition effect or the reverse. It is also clear that no consistent pattern emerges that defines the behavior of all speakers. These findings again suggest that substantial speaker-to-speaker variation occurs in relation to the effects of frequency condition on perceived effort and loudness.

Sentences

A second experiment was performed to determine the extent to which listeners' ratings of loudness and effort for sentence samples are dependent upon speaker differences, production intensity and playback level. For each sentence sample, the geometric mean of all listener ratings was obtained for each of the two response variables (loudness and effort). The geometric means were examined using an analysis of variance with factorial arrangement of treatments, in which the factors were production intensity, playback level, and speakers. The analyses were performed separately for loudness and for effort ratings of the same sentences. Though significant main effects and interactions involving speakers are observed (Tables 12 and 13), it was not feasible to perform the statistical analyses separately for the individual speakers, as was done for isolated vowels, because only one sentence production per speaker at each production intensity was available. Therefore, the sentence data represent results found for the four subjects combined. Descriptions of these data within the text, however, will consider the effects of the experimental variables separately for the individual subjects.

For purposes of discussion, this section will be organized as follows: (a) reliability of listener ratings of loudness and effort for sentences; (b) loudness ratings of the sentences; and (c) effort ratings of the sentences.

Reliability of listener ratings. Table 10 presents, separately

TABLE 10

INTRA-CLASS CORRELATION COEFFICIENTS FOR LISTENER RATINGS OF LOUDNESS AND OF EFFORT USED TO PRODUCE SENTENCE SAMPLES

	Loudness	Effort
Overall	0.99	0.94
Playback Intensity and Production Intensity Covary	0.99	0.97
Playback Intensity Varies, Production Intensity Constant	0.99	0.92
Production Intensity Varies, Playback Intensity Constant	0.93	0.95

for loudness and for effort ratings, estimates of interjudge reliability obtained using the intraclass correlation coefficient (11). Overall correlations were 0.99 for loudness ratings and 0.94 for effort ratings of the same sentence samples. For three individual modes of sample analysis, both for loudness ratings and for effort ratings of the same samples, all reliability coefficients were 0.92 or above. The magnitude of these coefficients indicates a high degree of reliability of listener judgments of these sentence samples.

Comparison of the reliability coefficients presented for sentences with those for vowels in Table 1 suggests that listeners rate effort and loudness more reliably for sentence than for vowel samples. It is possible that the availability of consonant-related effort and loudness cues contributes to the increased reliability of ratings of sentence samples.

Loudness ratings. The results of the statistical analysis of the loudness ratings for the sentences produced by the four subjects used as speakers in this study are presented in Table 11. Examination of this table shows that the speaker, production intensity, and playback intensity main effects are significant. In addition, the speaker-by-production intensity interaction is significant. All other main effects and interactions are not significant at the 0.05 level of confidence.

Following the format established for presentation of data for the vowel samples, all significant main effects and interactions will be presented graphically. Loudness ratings ($\log_{10}$), averaged over playback intensities, will be presented as a function of production intensity. In addition, loudness ratings, averaged over production intensities, will be presented as functions of playback intensity. The findings of this section of the study will be organized as follows: (a) production intensity effects on loudness ratings, and (b) playback intensity effects on loudness ratings. The speaker main effect and the speaker-by-production intensity interaction will be discussed within these headings, as appropriate.

Production intensity effects on loudness ratings. The significant production intensity main effect shown in Table 11 suggests that

TABLE 11

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS OF SENTENCES PRODUCED BY FOUR SPEAKERS

Source	df	ms	F
Production Intensities (A)	2	0.0346	9.416*
Playback Intensities (B)	2	0.7018	190.615*
AB	4	0.0018	< 1
Speakers (C)	3	0.0328	8.915*
AC	6	0.0116	3.166*
BC	6	0.0085	2.318
Residual	12	0.0036	

* $p \leq 0.05$

the mean loudness ratings, averaged over all speakers and playback levels, increases as production intensity increases from 65 dB to 85 dB. The presence of a significant speaker-by-production intensity interaction, however, shows that production intensity effects on loudness ratings differ for the four speakers. The means involved in this interaction, averaged over all playback levels, are plotted in Figure 24. Inspection of this figure reveals two relatively distinct patterns, one for Speakers A and C and one for Speakers B and D. Mean loudness ratings for Speakers A and C increase substantially from the 65 dB to the 75 dB level and then increase much less from 75 dB to 85 dB. Speakers B and D evidence a decline in mean loudness ratings as production level is increased from 65 dB to 75 dB and a slight increase in rated loudness between 75 dB and 85 dB. Speakers A, B, and C show an overall increase in rated loudness from the lowest to the highest production intensity level, while Speaker D does not. It may also be noted that Speaker D is rated consistently louder than the other speakers at each of the three production levels. It seems reasonable to conclude from these findings that while mean loudness ratings for the speakers as a group tend to increase with production intensity, this relationship is not consistently seen for all of the speakers studied. The extent and direction of changes in rated loudness with each 10-dB increment in production intensity varies for individual speakers.

Playback intensity effects on loudness ratings. The significant playback intensity main effect shown in Table 11 is displayed graphically for each speaker in Figure 25. In this figure, the mean loudness ratings, averaged over all production levels, are presented for

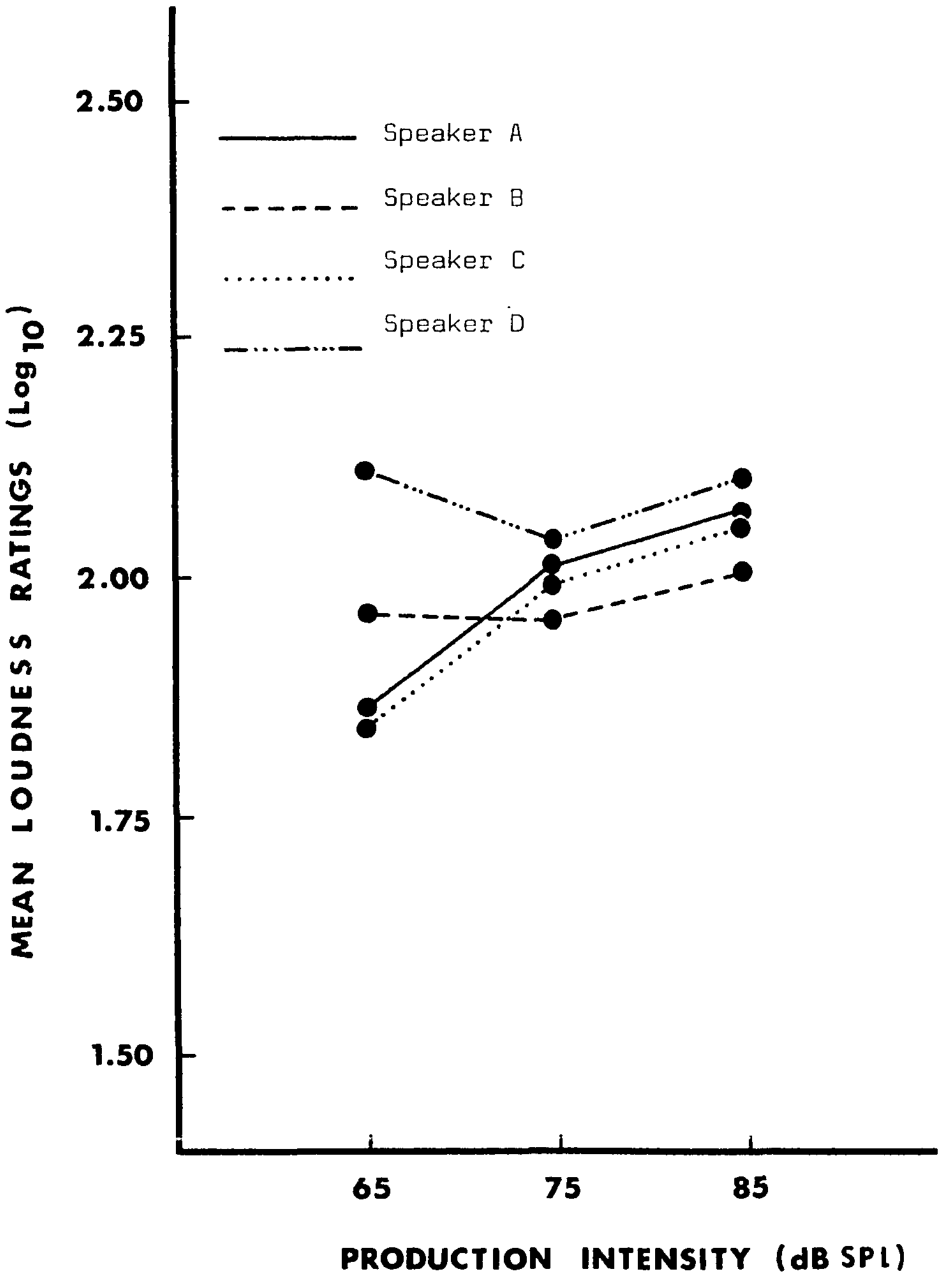

Figure 24.--Mean loudness ratings ($\log_{10}$) for sentences produced by Speakers A, B, C, and D. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

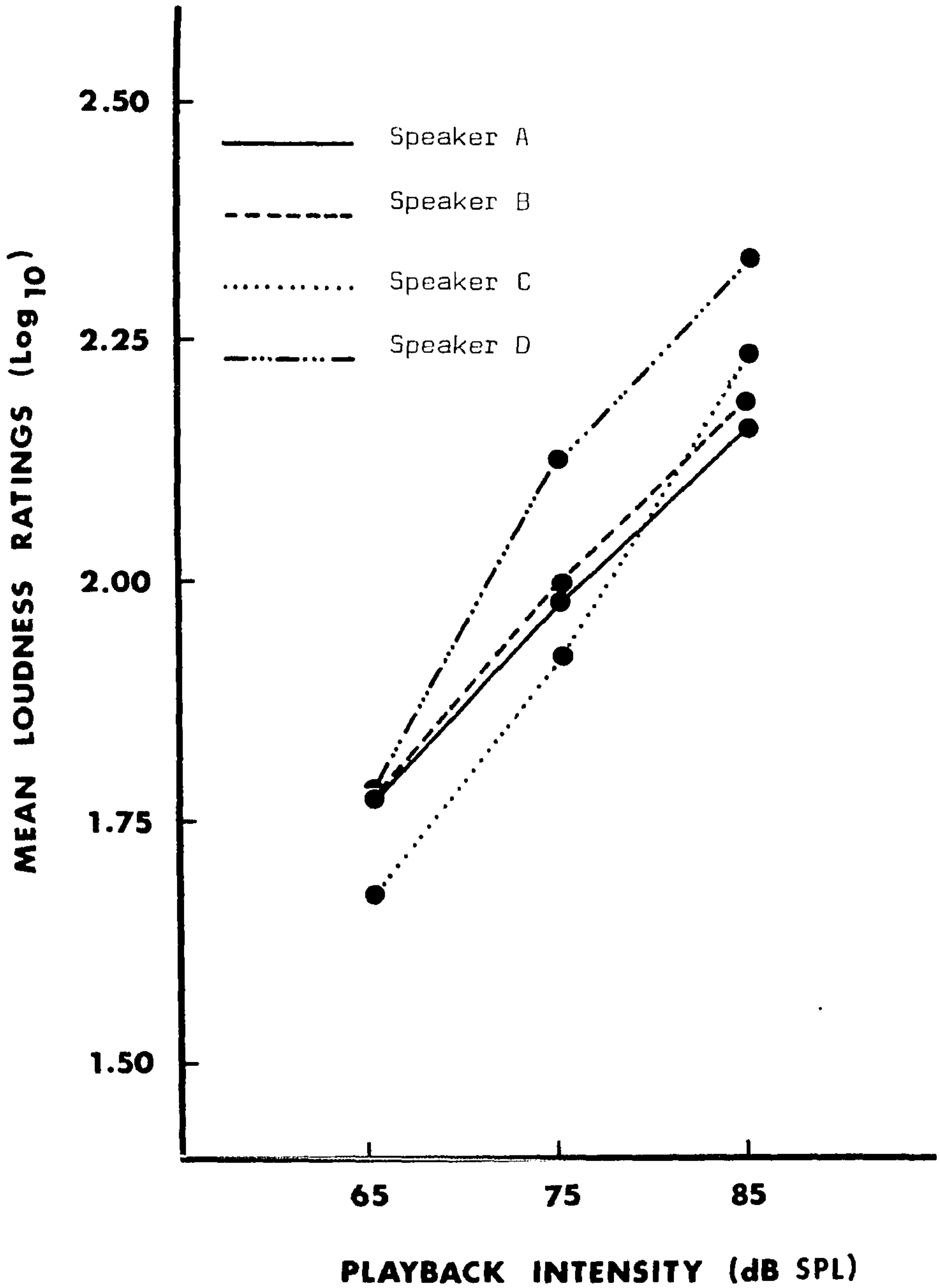

Figure 25.--Mean loudness ratings ($\log_{10}$) for sentences produced by Speakers A, B, C, and D. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

each of the four speakers at each playback level. It is apparent from this plot of means that the rated loudness of the sentence samples increases almost linearly for each speaker as playback intensity is increased. As in the graphic display of production intensity effects in Figure 24, Speaker D is rated louder than the other speakers at all playback levels, with this difference being most apparent at the 75 dB and 85 dB levels. These findings suggest that rated loudness tends to increase constantly as the playback level is increased.

The findings reported in this section suggest that variations in both production intensity and playback level provide cues used by listeners in rating the loudness of sentences. It seems clear, however, that production and playback level cues do not contribute to the perception of loudness in the same manner. Increments in playback level resulted in increases in rated loudness for all four speakers studied. The effects of increased production level on rated loudness, however, varied for individual speakers. In addition, a relatively restricted range of loudness ratings was assigned to samples at the three production levels, compared to the greater range of ratings representing the effects of playback level. These differences are evident in comparing Figure 24 with Figure 25. These findings suggest that variations in playback level provide more definitive and consistent cues to the loudness of sentence samples than do variations in production level.

Each sentence sample provided the listeners with an array of acoustic cues attributable both to the effect of the physiological speech production mechanism as well as to the electronic control of playback intensity. The procedure used to analyze the data obtained in this experiment allowed the investigators to assess, independently, the effects of production intensity and playback level. Presumably, the playback level effects revealed by this analysis reflect the influence upon ratings of the actual SPL at which listeners heard the samples. Production intensity effects appeared to relate more to the amount of effort used during sentence production. If this is so, it can be reasoned that listeners are able to use acoustic cues relating to the intensity at which vowels are heard as well as the effort used by the speaker to produce the sentence.

Effort ratings. Results of the analysis of variance presented in Table 12 reveal that the speaker, production intensity, and playback intensity main effects are significant. None of the interactions involving these factors is significant. As in the previous section, these findings will be presented under two headings: (a) production intensity effects on effort ratings and (b) playback intensity effects on effort ratings.

Production intensity effects on effort ratings. The significant production intensity main effect for effort ratings of sentences shown in Table 12 is plotted for each of the four speakers in Figure 26. In this figure, mean effort ratings, averaged over all playback levels, are presented for each of the four speakers as functions of production intensity. Inspection of this plot of means shows that each of the four speakers displays an increase in rated effort as production intensity is increased from 65 dB to 85 dB. It is also apparent, as shown by the significant speaker main effect, that differences exist among the mean effort ratings for the individual speakers. The mean effort ratings for

TABLE 12

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF SENTENCES PRODUCED BY FOUR SPEAKERS

Source	df	ms	F
Production Intensities (A)	2	0.1896	38.225*
Playback Intensities (B)	2	0.1333	26.877*
AB	4	0.0109	2.214
Speakers (C)	3	0.0376	7.580*
AC	6	0.0108	2.186
BC	6	0.0044	< 1
Residual	12	0.0049	

* $p \leq 0.05$

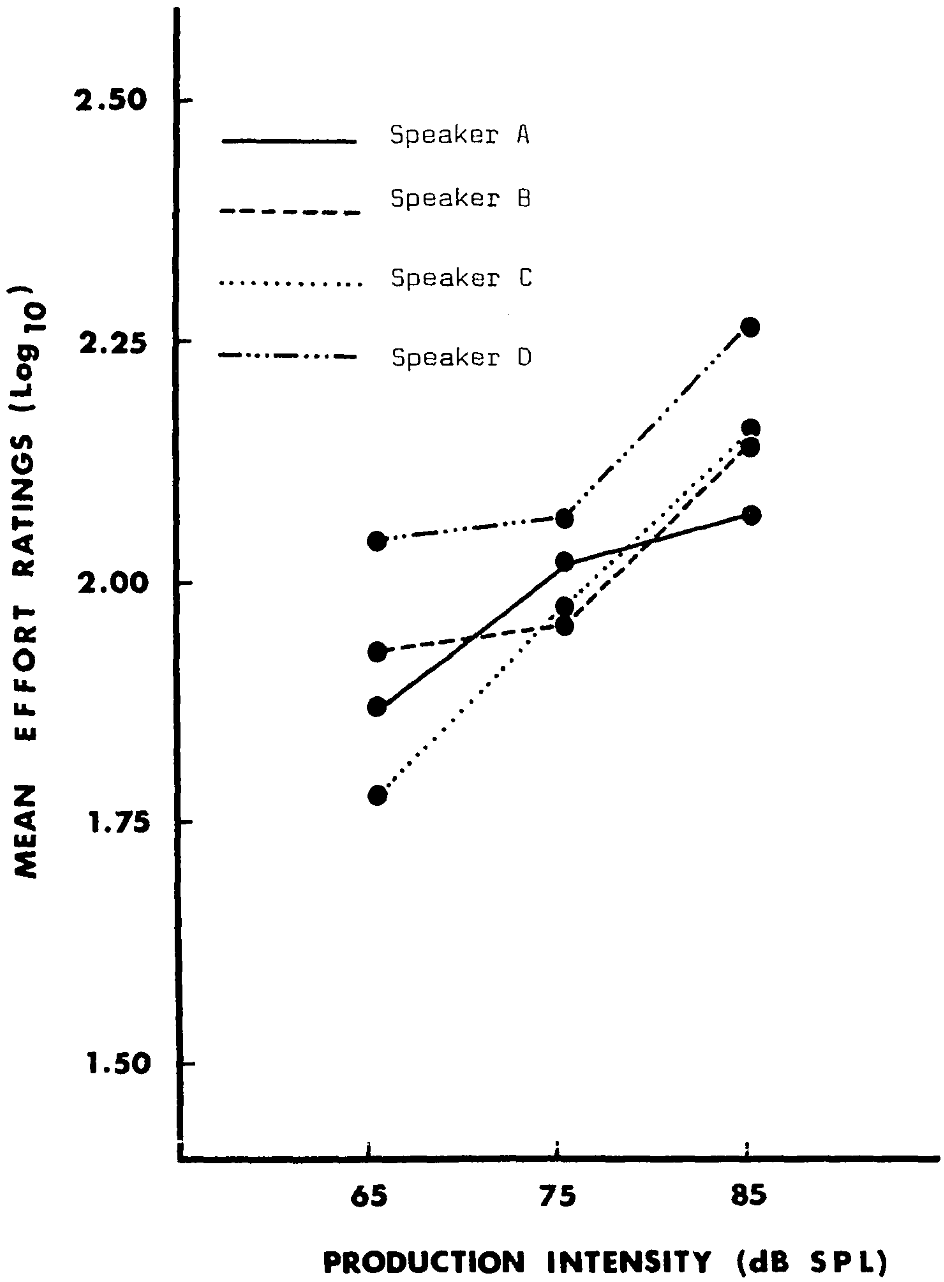

Figure 26.--Mean effort ratings ($\log_{10}$) for sentences produced by Speakers A, B, C, and D. The geometric means of listener ratings at each production intensity have been averaged over all playback levels.

Speaker D, for example, exceed those for the other speakers at all production intensities. Examination of the trends within Figure 26 is of interest. Although the patterns of change in rated effort with changes in production intensity tend to vary for individual speakers, the speaker-by-production intensity interaction is not significant. These findings suggest, in general, that effort ratings increased as production intensity increased. Comparison of Figure 26 with Figure 24 shows that production intensity cues produced greater increments in rated effort than loudness. Thus, it may be concluded that listeners make greater use of acoustic cues which are presumably related to speaker effort when specifically asked to rate effort as opposed to loudness of sentences.

Playback intensity effects on effort ratings. The significant playback intensity main effect shown in Table 12 is presented graphically for each of the four speakers in Figure 27. The means in this figure show that effort ratings increase almost linearly as playback intensity is increased in 10 dB steps from the 65 dB to the 85 dB level. As in the previous analysis of sentences, Speaker D was judged to employ the greatest amount of effort of any of the four speakers and showed the greatest increase in effort from the lowest to the highest playback level. These findings suggest that, as playback level increases, judges tend to rate the sentence samples to be more effortful.

In comparing effort ratings reflecting the effect of production intensity and of playback level (Figures 26 and 27, respectively), it can be seen that the two sets of effort functions are generally similar in direction and extent. That a similar range of effort ratings

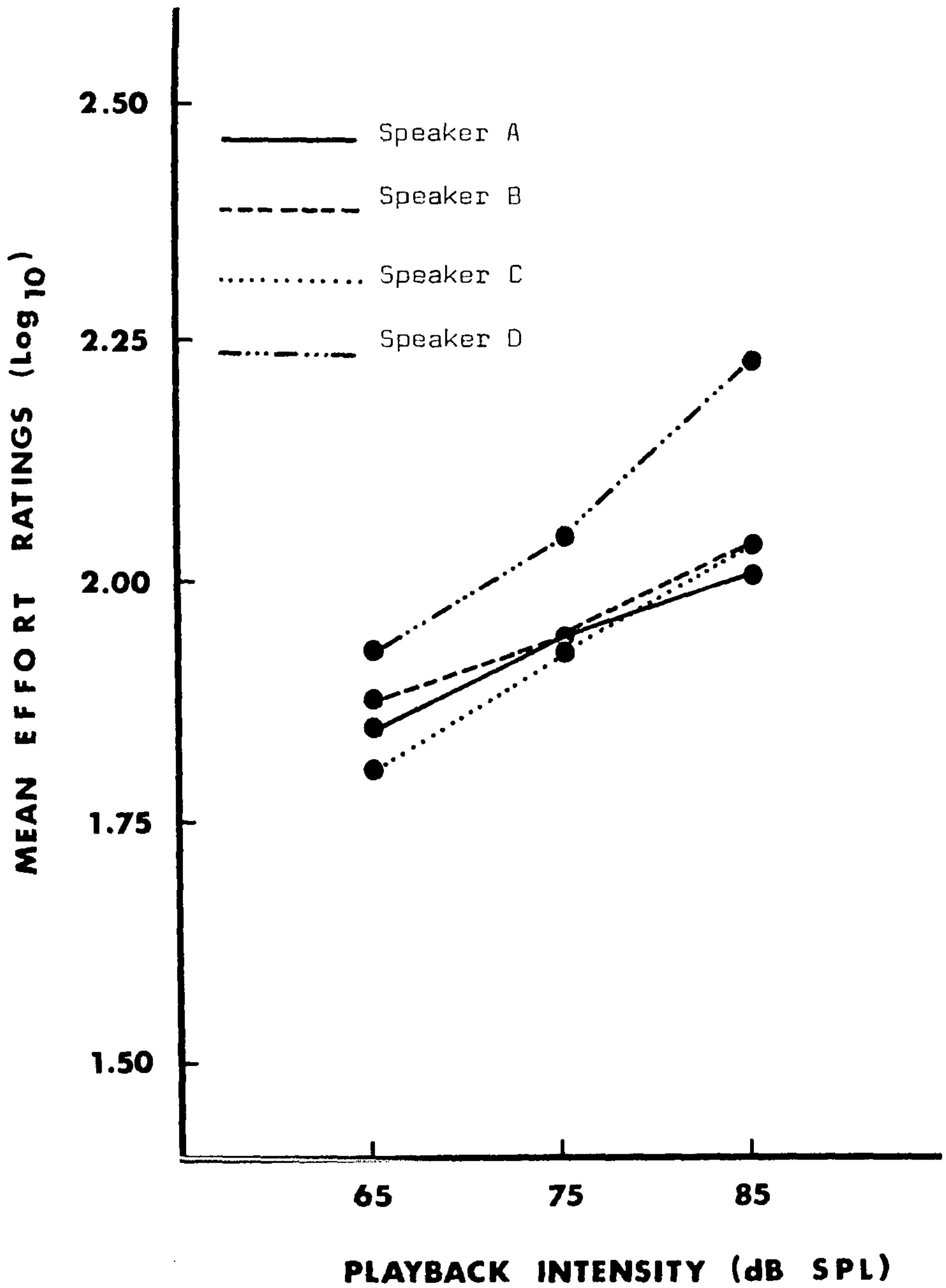

Figure 27.--Mean effort ratings ($\log_{10}$) for sentences produced by Speakers A, B, C, and D. The geometric means of listener ratings at each playback level have been averaged over all production intensities.

characterizes the effects of both production intensity and playback level suggests that judges tend to depend on SPL differences among the samples as well as production intensity cues in estimating the effort used to produce the sentences. Moreover, they appeared to rely on the two cue systems to about the same extent. Thus, cues which are related to differences in speaker production level and the actual intensity of samples at playback seem to contribute about equally to listener ratings of effort for sentences. In contrast, analysis of loudness ratings of the same sentences reflects greater listener dependence on playback level than on production intensity cues.

Comparison of Figures 24 and 26 shows that listeners reported more changes in effort than in loudness as a function of production intensity. This suggests that alterations in the acoustic properties of sentences which are attributable to different speaker production intensities contribute more to listener ratings of effort than loudness.

Conversely, a wider range of loudness ratings than effort ratings is observed when these ratings are plotted as functions of playback level (Figures 25 and 27). Apparently the actual intensity at which the sentences were heard was a more definitive cue to loudness than to effort for these listeners.

Discussion

In the present investigation, effort and loudness rating data obtained from sentences produced by four speakers were analyzed to determine the effects of speakers, production levels, and playback levels upon the ratings. The findings indicate that there is significant speaker-to-speaker variability in loudness-effort ratings of sentences.

This is consistent with findings previously reported for similar ratings of isolated vowels and suggests that results based upon single-subject experiments may not be representative of all speakers.

It is evident that loudness and effort ratings for sentences tend to increase with successive increments in production intensity. However, each 10 dB increment in production level does not result in consistent changes in rated effort or loudness for all speakers, though speaker variability related to production intensity is more characteristic of loudness ratings than effort ratings. As was suggested in the previous analysis of loudness and effort ratings for isolated vowels, there is reason to believe that this variability in ratings may reflect true differences in physiological effort used by different speakers to phonate at the same intensity. Thus, findings that sentences produced by the individual speakers differ in rated effort and loudness are not unexpected.

The findings of the present investigation also indicate that increments in playback level resulted in nearly linear increases in rated loudness and effort for all speakers. Here, there is less of the speaker-to-speaker variability which characterizes the effects of production level. In addition, loudness slopes are seen to rise more steeply as a function of playback level than of production intensity. It appears that cues related to differences in playback level provide a more definitive set of cues for loudness ratings of sentences than do variations in production intensity. In contrast, effort functions plotted over playback levels and production intensities appear similar. Thus, playback level does not appear to be more important than production intensity for effort ratings of sentences. In addition, the effects of playback and production intensity variations on effort and loudness appear to be independent of each other. Thus, listeners tend to respond to playback level variations without regard to the intensity at which the sample was produced and, conversely, to production intensity variations independent of the playback level.

Discussion Summary

The literature cited as background for the present investigation offers a number of divergent hypotheses concerning the perceptual bases of speech loudness. Lane, Catania, and Stevens (16) report that a speaker's judgment of his own vocal level (the autophonic response) was approximately proportional to the 1.1 power of the sound pressure level. An exponent of 0.7, however, was reported for the loudness of isolated vowels judged by listeners, corresponding well to the loudness function reported for pure tones (30, 31). Thus, they concluded, speakers based estimates of their own speaking level on physiological feedback from the degree of vocal effort used as well as SPL, whereas listeners used SPL alone. The findings of Moll and Peterson (22) are consistent with this hypothesis. These investigators report that loudness and effort functions for listener ratings of isolated vowels are nearly identical. They conclude that both effort and loudness ratings of vowels are predicated on the SPL differences among the samples and assert that listeners cannot differentiate between the effort and loudness of vowel samples even when explicitly asked to do so.

In contrast, Ladefoged and McKinney (15), based on the study of CVC syllables, suggest that listeners as well as speakers base their judgments of loudness on estimates of the physiologic effort involved in the production of subglottal pressure. Their conclusion was based on the findings that listeners' loudness judgments were approximately proportional to the 1.2 power of the sound pressure level, which corresponds well to the exponent (1.1) reported by Lane, Catania, and Stevens (16) for the autophonic response.

Somewhat compatible findings are reported by Brandt, Ruder, and Shipp (8). These investigators report that listeners may employ both effort and SPL cues in their assignment of loudness ratings. They report different power functions (0.92 and 0.38) for loudness and for effort, respectively, for sentences produced at equal intensities but heard at different playback intensities. Conversely, when listeners were provided with differences in production levels at uniform playback intensities, they heard slightly more effort differences than loudness differences though the power functions were not high (0.40 for loudness and 0.57 for effort). They assert that listeners are able to differentiate "something called effort" from loudness in rating sentence samples.

The findings of the present study are consistent with those of earlier studies (8, 16, 21, 22) which indicate that the loudness of vowel and sentence samples is an almost linear function of SPL for each of the speakers studied. When mean ratings of vowel and sentence samples are averaged over production intensities and plotted as a function of playback level, nearly a straight line is formed.

The present findings also suggest, however, that loudness judgments can be based on acoustic cues other than SPL differences among vowel and sentence samples. The irregular increments in loudness observed when ratings are averaged over playback levels and plotted as functions of production level suggest that increments in production intensity comprise an additional cue to the loudness of the sample. This cue probably relates to effort-associated alterations of the acoustic spectrum. In this respect, the present findings are consistent with the contention of Ladefoged and McKinney that listeners respond to physiologic effort in assigning loudness ratings. The failure of loudness ratings to increase as a linear function of production intensity, in the present study, suggests that acoustic cues directly related to physiologic effort played a less important role than playback level in determining the magnitude of loudness ratings.

Loudness and effort ratings of sentences appear to differ from each other in the relative use by listeners of cues related to production intensity and playback level. In contrast to loudness ratings, effort ratings appear to be about equally dependent upon production intensity and playback level cues. In addition, listeners appeared to depend upon production level cues more for effort than for loudness estimates. In contrast, they seemed to rely more upon playback level differences in estimating loudness as opposed to effort. This suggests, consistent with findings of Brandt et al., that listeners do, in fact, differentiate "something called effort" from loudness when specifically asked to do so.

The present findings also suggest that the relationship between loudness ratings differs for the isolated vowel and sentence samples. In rating both the effort and loudness of vowel samples, listeners tended to respond to the most obvious difference among the samples, that is, playback level differences or production intensity differences. The similarity of both sets of ratings suggests, as Moll and Peterson (22) contend, that listeners cannot differentiate cues to loudness from cues to effort when specifically asked to do so.

For sentence samples, loudness and effort ratings were different from each other. This suggests that listeners recognize and can label specific cues to effort in sentences. Perhaps these findings are attributable to the presence of consonant-vowel combinations in sentences. Brandt, Ruder, and Shipp (8) have suggested previously that spectral and/or temporal alterations in the relationship of consonants and vowels accompanying alterations in speaking level may signal changes in speaking effort. This cue system, understandably, is unavailable in isolated vowels. Results of the present investigation suggest that while spectral cues relating to speaking effort may be available to listeners within the isolated vowel, this cue system is not sufficiently obvious to listeners to allow them to make true "effort" ratings.

The interrelationships of perceived effort and loudness appear to be complex and somewhat variable from speaker to speaker. The degree of inter-speaker variability found in this study suggests that considerably more information is required before these interrelationships can be defined. It also seems clear that attempts to define the relationship between perceived effort and loudness may be limited if approached solely through perceptual measurements. From the present perceptual data, it can be reasoned that changes in the physiology of speech production do not result in one-to-one changes in perception.

Correlative measures of physiologic events would appear to contribute greatly to our understanding in this area.

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to investigate relationships among two types of intensity-controlled speech samples and listeners' ratings of loudness and of effort. Four normal-speaking adults, two males and two females, served as subjects for this investigation. The subjects individually produced each of the vowels /ɑ/ and /i/, as well as the sentence, "According to the present information the profits are high," at 65-, 75-, and 85-dB SPL. The vowels were produced at each intensity, first with fundamental vocal frequency free to vary with changes in intensity, and then with frequency controlled at a predetermined level. Each vowel production was sustained for seven seconds.

All sample productions were recorded on magnetic tape. Three exact duplicates of each speaker's production of each test sample were then dubbed and spliced in random order into two listening tapes, one of vowels and one of sentences. An attenuator and amplifier were utilized to attenuate the signal 0, 10, and 20 dB. This allowed the experimenter to systematically vary the relationships between the production intensities and playback levels of the test samples. The psychophysical method of magnitude estimation was used to obtain listener ratings of loudness and of effort for the recorded samples.

Using an analysis of variance procedure, the geometric means of listener ratings were compared to provide an index of the effects of production intensities, playback level, vowels (vowel samples only), and fundamental frequency conditions (vowel samples only). Additional graphic displays of mean ratings, presented individually for the speakers, permitted visual inspection of the effects upon listener ratings of the experimental variables.

Within the limitations of this experiment, the following conclusions appear to be warranted:

1. Loudness and effort are more reliably rated for sentences than for vowels. Further, ratings for vowels are more reliable when listeners hear playback level differences among the samples than when they base their ratings on production level cues alone. In contrast, sentence reliability is uniformly high in both of the two rating conditions.

2. Loudness and effort ratings for vowels are affected by vowel identity and by the presence or absence of fundamental vocal frequency change associated with alterations in production intensity. However, the extent to which ratings are affected by these variables is highly speaker-dependent.

3. For sentences as well as for vowels, increased production intensity is associated with irregular increases in loudness and effort ratings.

4. For sentences as well as for vowels, loudness and effort ratings based on production intensity differences show much speaker-to-speaker variability.

5. For sentences and for vowels, increments in playback level are associated with nearly linear increments in rated loudness and effort. Further, loudness and effort increase more rapidly with each 10-dB increment in playback level than for similar increments in production level. This holds especially for loudness and effort ratings of vowels and, to a lesser extent, for loudness ratings of sentences.

6. Effort ratings of sentences increase less rapidly as a function of playback level than do loudness ratings of sentences and loudness and effort ratings of vowels.

7. Effort ratings of sentences increase more rapidly as a function of production intensity than do loudness ratings of sentences and loudness and effort ratings of vowels.

There are several limitations of the present study that merit discussion. Although efforts were made to eliminate from the rating situation the effects of any except the experimental variables, it is possible that the present results were influenced by certain uncontrolled factors. For example, informal listening to the taped samples under conditions which simulated an actual rating session revealed the presence of extraneous background "hum". Two possible sources of this noise appear likely. It is possible that environmental noise present during recording may constitute one source. Another possible source was the electronic equipment used to replay the listening tapes to the listeners. Although the test samples were recorded at a constant tape recorder VU setting, the signal intensities during the final stage of amplification at playback covered a 20 dB range. The higher levels of signal amplification may have amplified low levels of noise which are characteristic of the electronic equipment. Although this background "hum" was noticeable upon listening, it was felt to be of sufficiently low intensity to constitute a minor source of contamination.

The findings of this study may also have been influenced by the method used to present the test samples to the listeners. Evaluation of the effects of the two pitch conditions, in particular, may have been influenced by the fact that vowels produced under both pitch conditions were randomized onto a single listening tape. Thus, listeners may have been confused by the range of pitch-intensity combinations available. In natural speaking situations, pitch change may be a more useable cue to speaking level than is implied by the present findings.

In spite of these limitations, the present study suggests that work already begun toward identifying acoustic correlates of vocal loudness and effort (4, 7, 8) is useful and should be extended. It appears that some of the cues used by listeners are contained within the vowel spectrum. However, since the present results indicate that listeners are able to use the label "effort" when rating the effects of production level in sentences but not vowels, one would suspect that the addition of consonants and/or the imposition of a prosodic pattern upon an utterance may give listeners a more useful set of "effort" cues. Thus, it would be of interest to investigate the relationship of loudness and effort ratings to spectral patterns in consonant-vowel combinations as well as to alterations in pitch and rate patterns within sentences.

Since acoustic cues to speaking effort, aside from intensity, are probably available within isolated vowels, one might hypothesize that this percept might be related to characteristics of the glottal source spectrum. The results of a preliminary investigation (12) suggest that such a relationship may exist. Subsequent investigations might explore further listener perceptions of vocal effort and loudness for speakers presenting vocal fold pathologies.

BIBLIOGRAPHY

1. Allen, G. D. Acoustic level and vocal effort as cues for the loudness of speech. Jour. Acous. Soc. Amer., 49, 1831-1841, 1971.

2. Arkebauer, H. J., Hixon, T. J., and Hardy, J. Peak intra-oral air pressures during speech. JSHR, 10, 2, 195-208, 1967.

3. Austin, D. and Emanuel, F. W. A Further Investigation of Vowel Spectral Noise Levels and Perceived Vowel Quality. A paper presented at the Proceedings of the Oklahoma Speech and Hearing Association Convention, Oklahoma City, Oklahoma, 1973.

4. Bernstein, J. Phonation and Loudness. A paper given at the 84th Annual Meeting of the Acoustical Society of American, Miami Beach, Florida, November, 1972.

5. Black, J. Natural frequency, duration and intensity in reading. JSHD, 14, 216, 1949.

6. Black, J. The pressure component in the production of consonants. JSD, 15, 207-210, 1950.

7. Brandt, J. F. Effects of stimulus bandwidth on listener judgments of vocal loudness and effort. Jour. Acous. Soc. Amer., 52, 705-707, 1972.

8. Brandt, J. F., Ruder, R. F., and Shipp, T. Vocal loudness and effort in continuous speech. Jour. Acous. Soc. Amer., 46, 1543-1548, 1969.

9. Brown, W. S. and Brandt, J. F. Effects of auditory masking on vocal intensity and intraoral air pressure during sentence production. Jour. Acous. Soc. Amer., 49, 1903-1905, 1971.

10. Brown, W. S. and Shearer, S. M. Constancy of intraoral air pressure related to integrated pressure-time measures. Folia Phon., 22, 49-57, 1070.

11. Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 16, 407-424, 1951.

12. Emanuel, F. W. and Counihan, D. T. Some characteristics of oral and nasal air flow during plosive consonant production. Cleft Palate Journal, 7, 249-260, 1970.

13. Fairbanks, G. A physiological correlative of vowel intensity. Speech Monographs, 17, 390-395, 1950.

14. Isshiki, N. Vocal intensity and air flow rate. Folia Phoniatrica, 17, 92-104, 1965.

15. Ladefoged, P. and McKinney, N. P. Loudness, sound pressure and subglottal pressure in speech. Jour. Acous. Soc. Amer., 35, 4, 454-460, 1963.

16. Lane, H. L., Catania, A. C., and Stevens, S. S. Voice level: autophonic scale, perceived loudness and effects of sidetone. Jour. Acous. Soc. Amer., 33, 160-177, 1961.

17. Leeper, H. A. and Noll, J. D. Pressure measurements of articulatory behavior during alterations of vocal effort. Jour. Acous. Soc. Amer., 51, 1291-1295, 1972.

18. Lehiste, I. and Peterson, G. Vowel amplitude and phonemic stress in American English. Jour. Acous. Soc. Amer., 31, 428-435, 1959.

19. Malecot, A. Mechanical pressure as an index of force of articulation. Phonetica, 14, 3, 169-180, 1966.

20. Malecot, A. The lenis-fortis opposition: its physiological parameters. Jour. Acous. Soc. Amer., 47, 1588-1592, 1970.

21. Mendel, M. I., Sussman, H. M., Merson, R. M., Naeser, M. A., and Minifie, F. D. Loudness judgments of speech and nonspeech stimuli. Jour. Acous. Soc. Amer., 46, 1556-1561, 1969.

22. Moll, K. L. and Peterson, G. E. Speaker and listener judgments of vowel levels. Phonetica, 19, 104-117, 1969.

23. Ringel, R., House, A. S., and Montgomery, A. A. Scaling Articulatory Behavior: Intraoral Air Pressure. A paper presented at the 71st Meeting of the Acoustical Society of American, 1967.

24. Sacia, C. F. Speech power and energy. Bell System Tec. Jour., 4, 627-641, 1925.

25. Sacia, D. F. and Beck, C. J. The power of fundamental speech sounds. Bell System Tech. Jour., 5, 393-403, 1926.

26. Scharf, D. S. Variations in vowel intensity measurements. Language and Speech, 9, 250-256, 1966.

27. Stevens, K. N. and House, A. S. Speech perception. In Foundations of Modern Auditory Theory, Vol. 2, J. V. Tobias (Ed.). New York: Academic Press, 1-63, 1972.

28. Steven, S. S. The measurement of loudness. Jour. Acous. Soc. Amer., 27, 815-829, 1955.

29. Stevens, S. S. The direct estimation of sensory magnitudes--loudness. Amer. J. Psych., 60, 1-25, 1956.

30. Subtelny, J., Worth, J., and Sakuda, M. Intraoral pressure and rate of flow during speech. JSHR, 9, 498-518, 1966.

31. Wright, H. N. and Colton, R. H. Some parameters of autophonic level. Presented to the Annual Convention of the American Speech and Hearing Association, November 21, 1972.

APPENDIX A

Instructions to Listeners
Training Tape

Instructions to Listeners
Training Tape

You are going to make ratings of speech samples which I will present to you later. To familiarize you with the procedures which you will use, I want you to practice rating the durations of a number of tones. One tone has been selected as the standard. You are to assign numbers to each of the remaining tones to indicate how long you think each one is in relation to the standard tone. That is, if a tone sounds twice as long as the standard, you will assign the value of 200 to it. You will assign the value of 50 to a tone which sounds half as long as the standard, and so on. You will hear the standard five times before you are to begin making ratings. Thereafter, the standard will be repeated after every fourth sample. Do you have any questions?

APPENDIX B

Instructions to Listeners
Loudness and Effort Ratings of Sentences

Instructions to Listeners
Loudness and Effort Ratings of Sentences

You are now ready to begin making judgments of speech stimuli. You are going to judge the loudness of the sentence, "According to the present information, the profits are high," as it is spoken by two men and two women. One sentence has been selected as the standard. You are to assign to it the value of 100. You are to assign numbers to each of the remaining sentences to indicate how loud each one sounds in relation to the standard. Please rate the loudness of the spoken sentence only and disregard any other feature of the sample you hear. You will assign the value of 50 to a sentence which sounds half as loud, and 200 to one that sounds twice as loud, and so on. As before, you will be allowed to hear the standard several times before you begin judging. It will also be repeated after every fourth sample you judge. The asterisks which appear on your rating sheets before some of the numbers indicate that a standard sample will be played before the number marked.

Now I am going to present additional samples of the same sentences which you have just rated in terms of loudness. During this portion of the listening session, I would like for you to judge the amount of effort which the speaker seems to be using to produce the sentence. All of the conditions will be the same as before. Again, rate only the amount of effort which the speaker seems to be using, and disregard any other feature of the sample. The standard sample, which you will hear several times at the beginning, will be assigned the value of 100. As you listen to the remaining samples, assign the value of 200 to a sample which sounds as if the speaker used twice as much effort to produce it. If the speaker seems to be using half as much effort, assign the value of 50 to that sample.

APPENDIX C

Instructions to Listeners
Loudness and Effort Ratings of Vowels

Instructions to Listeners
Loudness and Effort Ratings of Vowels

Next, we will present samples of the vowels /ɑ/ and /i/, spoken by two women and two men. As before, you are to rate the loudness of these vowels. The vowel which has been selected as the standard will have a value of 100. Assign the value of 200 to all samples which sound twice as loud as the standard, 100 to samples which sound just as loud as the standard, 50 to samples which sound half as loud as the standard, and so on. On your judgment sheets, the asterisks before some of the numbers indicates that the standard will be presented before that sample.

Now I am going to present additional samples of the same vowels. This time you are to rate the amount of effort which the speaker seems to be using to produce them. The standard stimulus has been assigned a value of 100. Assign numbers to the remaining samples which indicate their relation to the standard. Use the value of 200 for vowels which sound as if the speaker used twice as much effort to produce them. If a vowel sounds as if the speaker used half as much effort to produce it, assign the value of 50 to it, and so on.

APPENDIX D

Summaries of Analyses of Variance

TABLE 13

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS
OF VOWEL /u/ PRODUCED UNDER FREQUENCY CONDITION I
BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.0221	12.119*
Playback Intensities (B)	2	0.5942	324.606*
AB	4	0.0054	2.969
Speakers (C)	3	0.0141	7.710*
AC	6	0.0072	3.955*
BC	6	0.0015	< 1
Residual	12	0.0018	

* $p \leq 0.05$

TABLE 14

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS
OF VOWEL /ɑ/ PRODUCED UNDER FREQUENCY CONDITION II
BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.0543	26.581*
Playback Intensities (B)	2	0.6945	339.726*
AB	4	0.0036	1.790
Speakers (C)	3	0.0083	4.097*
AC	6	0.0162	7.940*
BC	6	0.0029	1.433
Residual	12	0.0020	

* $p \leq 0.05$

TABLE 15

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS OF VOWEL /i/ PRODUCED UNDER FREQUENCY CONDITION I BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.0877	30.615*
Playback Intensities (B)	2	0.7464	260.552*
AB	4	0.0070	2.466
Speakers (C)	3	0.0576	20.112*
AC	6	0.0056	1.966
BC	6	0.0010	< 1
Residual	12	0.0028	

* $p \leq 0.05$

TABLE 16

SUMMARY OF THE ANALYSIS OF VARIANCE FOR LOUDNESS RATINGS OF VOWEL /i/ PRODUCED UNDER FREQUENCY CONDITION II BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.0507	14.020*
Playback Intensities (B)	2	0.6678	184.678*
AB	4	0.0122	3.394
Speakers (C)	3	0.0229	6.355*
AC	6	0.0210	5.827*
BC	6	0.0054	1.510
Residual	12	0.0036	

* $p \leq 0.05$

TABLE 17

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWEL /ɑ/ PRODUCED UNDER FREQUENCY CONDITION I BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.0742	6.137*
Playback Intensities (B)	2	0.8463	69.952*
AB	4	0.0011	< 1
Speakers (C)	3	0.0436	3.606*
AC	6	0.0111	< 1
BC	6	0.0076	< 1
Residual	12	0.0120	

* $p \leq 0.05$

TABLE 18

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWEL /ɑ/ PRODUCED UNDER FREQUENCY CONDITION II BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.0855	9.551*
Playback Intensities (B)	2	0.7806	87.206*
AB	4	0.0041	< 1
Speakers (C)	3	0.0169	1.894
AC	6	0.0269	3.009*
BC	6	0.0114	1.264
Residual	12	0.0089	

* $p \leq 0.05$

TABLE 19

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWEL /i/ PRODUCED UNDER FREQUENCY CONDITION I BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.1803	39.136*
Playback Intensities (B)	2	0.9291	201.675
AB	4	0.0044	< 1
Speakers (C)	3	0.0650	14.128*
AC	6	0.0106	2.300
BC	6	0.0090	1.955
Residual	12	0.0046	

* $p \leq 0.05$

TABLE 20

SUMMARY OF THE ANALYSIS OF VARIANCE FOR EFFORT RATINGS OF VOWEL /i/ PRODUCED UNDER FREQUENCY CONDITION II BY SPEAKERS A, B, C, AND D

Source	df	ms	F
Production Intensities (A)	2	0.1003	27.265*
Playback Intensities (B)	2	0.9903	269.076*
AB	4	0.0039	1.083
Speakers (C)	3	0.0638	17.341*
AC	6	0.0320	8.699*
BC	6	0.0086	2.350
Residual	12	0.0036	

* $p \leq 0.05$

APPENDIX E

Spearman Rank Order Correlation Coefficients for Loudness and Effort Ratings of Vowels and Sentences

Spearman Rank Order Correlation Coefficients for Loudness and Effort Ratings of Vowels and Sentences

Correlations between each of four separate groups of listeners and a single "standard" group are presented.

<u>Sentences</u>

Loudness	0.93
Effort	0.86

<u>Vowels</u>

Loudness	0.92
Effort	0.89