6-2004

Acoustic, Kinematic, and Auditory Perceptual Characteristics of Clear Speech

Kristin L. Greilick

Recommended Citation

Greilick, Kristin L., "Acoustic, Kinematic, and Auditory Perceptual Characteristics of Clear Speech" (2004). *Master's Theses*. 4563.
https://scholarworks.wmich.edu/masters_theses/4563

ACOUSTIC, KINEMATIC, AND AUDITORY PERCEPTUAL CHARACTERISTICS OF CLEAR SPEECH

by

Kristin L. Greilick

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Arts
Department of Speech Pathology and Audiology

Western Michigan University
Kalamazoo, Michigan
June 2004

ACKNOWLEDGMENTS

This project has been made possible due to the support and guidance of the following individuals. My greatest debt is owed to my advisor Dr. Stephen Tasko for taking on this challenge and seeing me through. Without his encouragement and tireless explanations I would never have reached any degree of depth on this topic; he is a teacher in the truest sense. I also owe a great debt to Dr. James Hillenbrand for his guidance when this project was still just an idea yet to be explored, as well as his much needed technical support. I would also like to thank Dr. Michael Clark for his encouragement and editorial assistance that has made this paper presentable. Additionally I would like to thank the researchers behind the University of Wisconsin Microbeam Database for their efforts collecting the datasets that made up the foundation of this project. Thanks also to each listener participant for lending his/her time and a keen ear.

A heartfelt thanks to my parents, Paul and Mary and the rest of my family, Carol, Gary, Kayla, Logan, Olivia and Parker for their support in my academic interests. Special thanks go to my parents for instilling curiosity and appreciation for scientific exploration that powered the desire to tackle this project. I am truly grateful for their continued interest and enjoyment in my life experiences.

Kristin L. Greilick

ACOUSTIC, KINEMATIC, AND AUDITORY PERCEPTUAL CHARACTERISTICS OF CLEAR SPEECH

Kristin L. Greilick, M.A.

Western Michigan University, 2004

This study characterized the speech motor transformations that underlie speech clarity changes in a group of 49 healthy young speakers. Clarity judgments based on auditory perception of the speech samples were determined by a panel of 30 undergraduate and graduate students. This study specifically aimed to characterize (1) the auditory-perceptual judgments of clarity, (2) the acoustic measures of clear and causal speech, (3) the temporal and spatial features of articulatory movements of clear and casual speech, and (4) test the hypothesis that auditory perceptual scaling of clarity (perceptual salience) would be associated with kinematic indicators of physical effort.

Results suggested clear speech is accompanied by changes in both the spatial and temporal features of speech motor output. However, temporal measures were more strongly correlated with perceptual ratings of clarity. Auditory perceptual scaling of clarity was not found to be associated with physical effort as defined by a change in peak speed.

TABLE OF CONTENTS

CHAPTER

Table of Contents—continued

Table of Contents—continued

APPENDICES

LIST OF TABLES

List of Tables—continued

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background Information

Variability in Speech Production Processes

The ease with which speakers communicate belies the complexity of the operations underlying speech production. Producing even the simplest of phrases involves transforming a set of abstract linguistic units into continuous bodily movement. These movements, which are bound by the laws of physics and biology, serve to generate and modify acoustic spectra over time to produce sound patterns that listeners can link to meaning. To achieve these movements, the speaker must activate with appropriate onset times, durations, and intensities a diverse set of muscles from respiratory, phonatory, resonatory, and articulatory systems. Further, speakers must readily adjust this complex process to communicate the same linguistic message to meet a variety of environmental and listener characteristics. In spite of this highly efficient transmission of linguistic information, speakers exhibit substantial variation in speech-related movements and resulting sound patterns. The broad goal of much of the research in normal speech production processes is to provide a principled account of the variation observed in a given speech signal. Sources of variation include phonetic context (coarticulation), speech rate, stress patterns, dialect, and adaptation due to environment and listener characteristics. Environment and listener characteristics are relevant, as a speaker often has to alter his/her speech production patterns to increase the clarity of the message.

Speech Clarity as a Source of Speech Production Variation

Clear speech can be defined as speech produced with an attempt to enhance its intelligibility. Intelligibility or clarity of a speech signal may be compromised by environmental noise, physical distance between a speaker and a listener and/or various listener-specific limitations such as hearing impairment. There are numerous motivations for investigating speech production correlates of speech clarity. Clinically, diagnostic and therapeutic procedures benefit from a comprehensive understanding of the features that speakers consider important for message clarity and the motor transforms that underlie it. As a result, there is a growing literature on the acoustic, and to a lesser extent articulatory kinematic, correlates of clear speech. This literature suggests there are a number of measurable differences in the acoustic features of speech produced under clear and casual conditions. Acoustic features of clear speech include a decrease in articulation rate (Picheny, Durlach, & Braida, 1985, 1986), increase in pause time (Picheny, et. al., 1985, 1986), elevated sound pressure level (Picheny, et. al, 1985, 1986), expanded vowel formant spaces (Moon & Lindblom, 1994; Bradlow, 2002), increased frequency of released stop consonants (Picheny, et. al., 1985, 1986) and evidence for larger and more rapid formant transitions (Moon & Lindblom, 1994).

There are also theoretical motivations for studying speech clarity. A comprehensive theory of speech production must account for how speakers communicate the same message or phonetic string under differing conditions and communication demands. For example, Lindblom (1990) has developed a theoretical framework that acknowledges and attempts to account for the wide range in speech clarity observed in everyday communication. Lindblom argues that the principal objective of the speaker is the production of speech that is as perceptually distinctive as the speaker-listener situation demands it to be. This suggests that the speakers possess a degree of flexibility in a number of parameters of production and that a degree of adjustment made in these parameters is determined by what the speaker believes the listener requires. Lindblom suggests that a speech solution for a given situation is the result of a type of cost-benefit competition between a general goal to economize speech gestures and the communicative goal of meeting listener-oriented perceptual demands. Speech production becomes an optimization problem in which speakers make these adjustments along a continuum between "hypospeech", which is highly economical but exhibits limited perceptual salience, and "hyperspeech", which is perceptually salient, but comes at a higher physiologic cost. Casually and clearly produced speech would likely fall in different places along the hypo-hyperspeech continuum. This perspective would predict that as compared to casual speech, clear speech would be associated with greater perceptual salience and greater effort.

Testing such a hypothesis requires methods to quantify both the perceptual salience of a speech signal and physiologic effort involved in producing that speech signal. Evaluating the former can be accomplished by having listeners judge speech intelligibility or perceived signal clarity. Quantifying physiologic effort is somewhat more challenging, in part because effort may be defined in a variety of ways. Based on modeling efforts at finding kinematic solutions that minimize various physiologic costs, Nelson (1983) suggested that peak velocity provides a reasonable index of physiologic effort. Further, Nelson (1983) provided evidence that speech-related jaw movement conforms to a principle of economy of effort.

This approach to studying kinematic effort in clear speech has failed to provide predictable speaker strategies for producing greater articulatory effort in response to the need to vary their speech clarity. This general approach has been the focus of two recent studies. The results of studies done where peak movement speed was used as a relative measure of articulatory effort showed a greater peak speed in the clear condition for some speakers, suggesting that clear speech is produced with greater articulatory effort than casual speech. However, this was not observed for all subjects. More interestingly, what was also noted in these studies was that a greater percentage of the subjects used a variety of other combinations of parameters to produce the clear conditions (Matthies, Perrier, Perkell, & Zandipour, 2001; Perkell, Zandipour, Matthies, & Lane, 2002). Other parameters used to produce the clear conditions included SPL increase, increased vowel duration and an increase in consonant-string duration. Changes in peak speed/duration were highly variable in both studies for the clear condition. While some subjects demonstrated the highest peak velocity in the normal speaking condition, others demonstrated highest peak velocity in the clear condition.

Research Problem and Questions

Current understanding of the acoustic, perceptual and kinematic correlates of clear speech is limited by a number of factors. First, studies thus far have limited sample sizes, with numbers typically ranging between five and ten speakers. Second, minimal information is typically provided about the degree to which speakers achieved clear speech. Auditory perceptual scaling of clarity by listeners would aid in determining the extent of clarity adjustments made by speakers. This type of measurement would be consistent with Lindblom's claim of speech on a continuum of hypo and hyper articulation. Finally, concurrent evaluation at perceptual, acoustic and kinematic levels may provide a clearer view of possible perceptual salience-effort trade-offs that are made by speakers.

Goals of Research

The overall goal of this study was to evaluate auditory perceptual, acoustic and kinematic measures in large group (49) of healthy speakers as they speak under casual and clear conditions. Specifically, this study is designed to

1. Characterize the auditory-perceptual judgments of clarity
2. Characterize acoustic measures of clear and causal speech
3. Characterize temporal and spatial features of articulatory movements of clear and casual speech
4. Test the hypothesis that auditory perceptual scaling of clarity (perceptual salience) will be associated with kinematic indicators of physical effort.

CHAPTER II

LITERATURE REVIEW

Studies of Clear Speech

Kent and Read (2002) define *clear* speech as speech produced in an effort to be highly intelligible which is in contrast to *conversational* speech in which clarity may be compromised. This definition identifies the goal of a speaker to be more acoustically distinctive, thus more clear. Identifying the factors that underlie clear speech has practical and theoretical benefits for understanding normal and disordered communication. Improving hearing aid technology depends on knowing what cues are most relevant for speech understanding. Speech disorders such as dysarthria are often characterized by reduced clarity. Therapy improvements will rely on a better understanding of what speakers attempt to do when they wish to increase speech clarity. From a theoretical perspective, an adjustment in speech clarity demands that a speaker (normal or disordered) transform his/her speech motor output in some way. Evaluating at the kinematic, acoustic and perceptual levels how such transformations occur can provide insights into the organization of the speech motor system.

Acoustic Properties of Clear Speech

The effect of changes in speech clarity on speech acoustics has been reported in a pair of studies by Picheny, Durlach, & Braida (1985, 1986). In the initial study (Picheny, et al. 1985), three speakers produced nonsense sentences spoken in either clear or conversational speech. These sentences were presented to five listeners with sensorineural hearing loss. Intelligibility was scored as the number of correctly identified words.

Intelligibility scores were significantly better for the sentences produced in the clear manner as compared to the conversational manner. Having established the facilitative effect that a clear speech mode can have on intelligibility, the authors followed up with a study comparing acoustic properties of clear versus conversational speech (Picheny, et al., 1985). The sentences used in the earlier intelligibility study were submitted to acoustic analysis. Results revealed that the clear speech condition could be distinguished from the conversational condition along a number of acoustic dimensions. It was observed that clear speech was characterized by a decreased rate of speech. This rate reduction was achieved through the insertion and lengthening of pauses and by increasing the duration of individual speech sounds. The authors also report that stop consonants were more likely to be released and that the intensity of obstruent consonants was greater for the clear condition. The clear speech condition was also characterized by vowel nuclei with formant frequencies that were closer to canonical values. The authors concluded that differences exist in the acoustic characteristics of clear speech and that these features contribute to improved speech intelligibility for hearing impaired listeners.

Moon and Lindblom (1994) evaluated the role of speech clarity on second formant (F2) transition in a set of test words that were designed to vary vowel duration and the degree of F2 displacement. The aim of this study was to look at formant “undershoot” which the authors defined as a reduction in formant values away from hypothetical values and vowel reduction as acoustic centralization. Data was collected on glide-vowel combinations in clear speech versus citation-form speech to identify which variables were the most important determinants of vowel reduction. Five students participated. Results revealed that the clear speech condition had less average undershoot than citation form speech, and that the degree of undershoot was talker specific. Acoustic analysis revealed that vowel formant patterns were affected according to neighboring consonants, as measured by vowel duration and that these dependencies were more limited for clear versus citation form speech. Additionally, clear speech was characterized by increased intensity, longer vowel durations, larger formant displacements and more rapid formant transitions than the citation speech. The authors conclude that clear speech is not simply loud speech, but involves a systematic reorganization of articulatory pattern (Moon & Lindblom, 1994).

Bradlow (2002) studied clear speech to evaluate investigate whether coarticulatory influences are minimized, maintained or exaggerated in clear speech, and whether the vowel space expansion effect of clear speech varies across languages with vastly different vowel inventory sizes. Bradlow defined coarticulation as variations on vowel formant values that depended on the preceding consonant. This study compared the extent of vowel variability in conversational and clear speaking modes for high back vowels that occurred in CV syllables where C varied in place of articulation. English monolingual and Spanish-English bilingual speakers were used in the subject pool. The subjects produced the nonsense words /bu/ and /du/ in a carrier sentence under conversational and clear speaking modes. The monolingual subjects produced the sentences in English while the bilingual subjects produced the sentences in English and Spanish. Specifically, clear speech of both mono and bilingual speakers maintained CV coarticulation for the vowel /u/ when produced following /b/ and /d/. Bradlow suggests that this maintenance of coarticulation shows talker control and serves some listener-oriented purpose rather than purely talker-oriented economy of effort. The second main finding of this study was that clear speech in both English and Spanish involved a similar amount of vowel space expansion. This finding suggests that speakers, regardless of the size of language's vowel inventory, spare no effort in producing clear speech and instead "globally" hyperarticulate, even when perceptual confusion is unlikely. In summary, vowels of clear speech are positioned more peripherally in the vowel space relative to vowels of conversational speech, while still maintaining the coarticulation effect of the preceding consonant.

Another approach to understanding speech clarity is to study the acoustic features of speech disorders that are characterized in part by a reduced intelligibility (i.e. clarity). Kent and colleagues adopted this approach to study the acoustic and intelligibility deficits that occur in dysarthria associated with amyotrophic lateral sclerosis (Kent, Kent, Weismer, Martin, Sufit, Brooks & Rosenbek, 1989). This study evaluated the relationship between the rate of second formant (F2) change during vocalic transitions and speech intelligibility. Kent, et al. (1989) found that for the most unintelligible speakers, the F2 trajectory is typically long in duration and almost flat, indicating little or no articulatory movement during the vocalic transition of a word. Kent, et al. (1989) found that the correlations between the rate of F2 transition and word intelligibility exceeded 0.80. The authors concluded that the second formant (F2) transition slope may be a useful index of intelligibility deficits in dysarthria (Kent, et al., 1989).

Speech Clarity as a Reflection of Different Speech Motor Optimization Strategies

There is substantial evidence that clear speech is produced in fundamentally distinct ways from casual or conversational speech. Clarity-based adjustment in speech motor output is frequently referred to as hyperarticulation or "hyperspeech" (Lindblom, 1990; Bradlow, 2002). Lindblom (1990) argues that the principal objective of the speaker is to produce speech that is as perceptually distinctive as the speaker-listener situation demands it to be. As the speaker-listener situation varies, so do the demands on the speech motor system. This perspective presumes that speakers can exercise substantial flexibility in the control of various speech production parameters and will adjust these parameters along what Lindblom terms a “hyper-hypospeech” continuum. For example, speaking in noisy environments with a stranger would demand a high level of perceptual distinctiveness resulting in hyperspeech. Alternatively, when in a quiet room with a familiar listener, there is less demand for perceptual distinctiveness and the speaker would resort to a form of hypospeech. Lindblom suggests that speakers select a strategy within this hyper-hypospeech continuum that simultaneously meets an output-oriented demand for a perceptually distinct speech signal and a system-oriented demand for biologically efficient speech motor patterns. Therefore, within this framework, speech production becomes an optimization problem in which speakers make these adjustments along a continuum between hypospeech, which is biologically efficient but exhibits limited perceptual salience, and hyperspeech, which is perceptually salient, but carries a higher physiologic cost. This theoretical perspective states that this hypospeech-hyperspeech dimension can account for the range of phonetic variation observed in speech. As output constraints dominate, we expect to see “hyperforms” of phonetic elements. Within this framework, casually and clearly produced speech would likely fall in different places along the hypo-hyperspeech continuum. Specifically, it would predict that as compared to casual speech, clear speech would be associated with greater perceptual salience and greater effort.

One of the challenges of testing such a hypothesis has been defining what is meant by biologic/physiologic effort or cost. Nelson (1983) argued that it is possible to infer physiologic effort by investigating the kinematic features of movement-related velocity histories. A velocity history plots a movement's instantaneous velocity over its time course. Figure 1 shows a stylized velocity history. This velocity history has a single period of acceleration and deceleration. "Unimodal" velocity histories are characteristic of most skilled movements, including speech (Nelson, 1983). From such a plot, a number of kinematic measures may be extracted. The duration of the movement may be determined from the time between the onset of acceleration and the offset of deceleration. The distance moved is equivalent to the time integral of the velocity history (area under the curve). Nelson evaluated a number of optimal control models which sought to minimize some "cost" that could reasonably be considered a measure of movement effort. The parameters used to define cost included peak acceleration (which is proportional to peak force), impulse (time integral of force/acceleration), energy (square of force/acceleration), and jerk (first-order derivative of acceleration). Nelson found that most of the models tested produced optimal velocity histories that were very similar in appearance. The great similarities in the model results regardless of the specific cost used led Nelson to use a minimum-impulse model. The advantage of this model is that the cost is physiologically plausible (time integral of force) and equal to the peak velocity of the movement (Figure 1), which can be observed directly from the kinematic signal. Within such a framework, relatively large peak velocities would characterize "high-cost" or "effortful" movements. Using Lindblom's (1990) theoretical framework, clear speech would be characterized by movements with relatively large peak velocities.

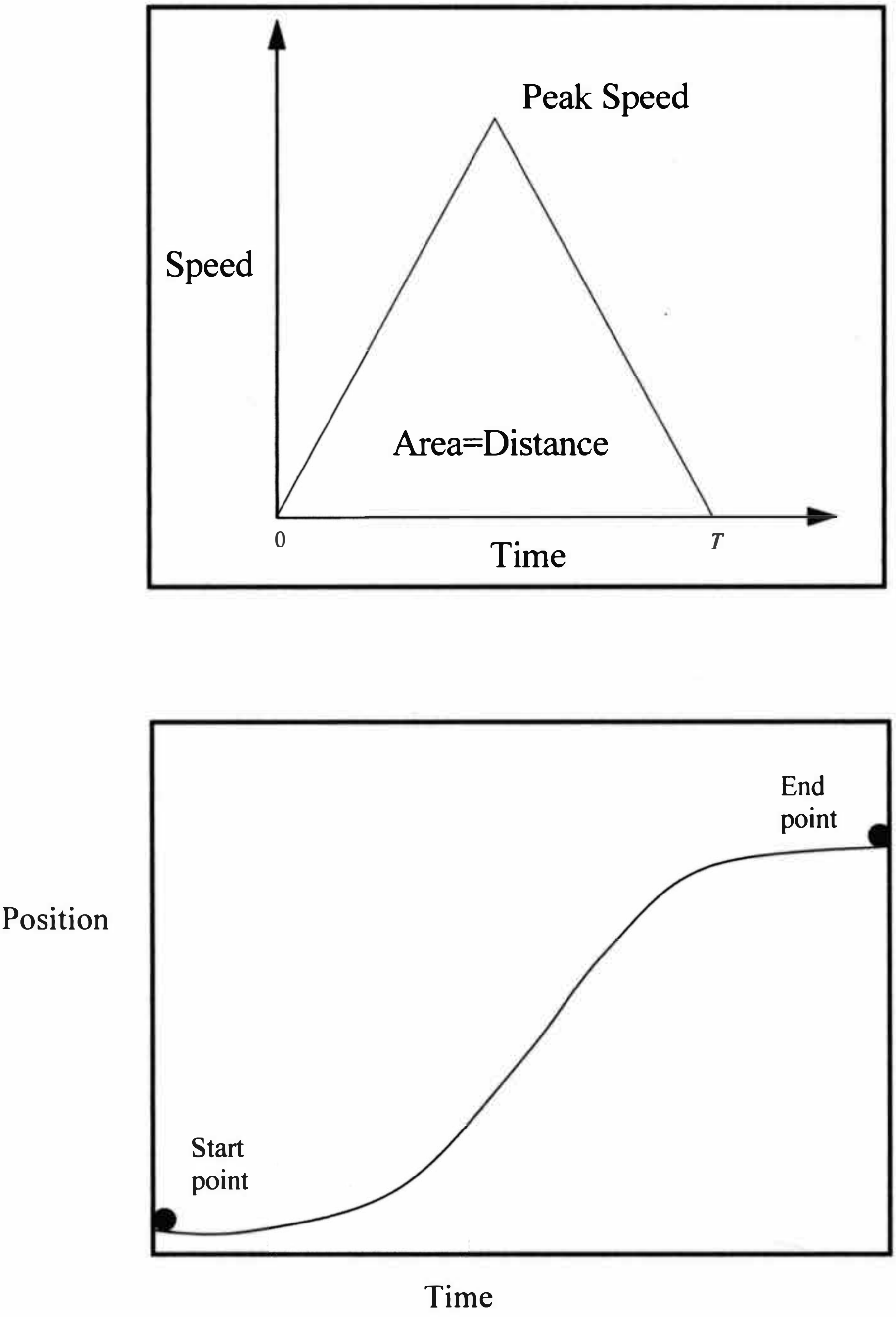

Figure 1. Geometric representations of a one-dimensional movement or velocity history.

Two recent studies have attempted to use the framework developed by Nelson (1983) to evaluate the relationship between "effort" and clear speech. Matthies, Perrier, Perkell, & Zandipour (2001) tested the hypothesis that coarticulation, defined as the cost-conserving overlapping of articulatory movements between neighboring sounds in a sequence, at vowel-consonant and consonant-vowel boundaries, is influenced by the listener's requirement for clarity and the speaker's strategy to economize effort. The main hypothesis was that temporal and spatial aspects of coarticulation would be consistent with a trade-off between clarity requirements and economy of effort and that this tradeoff would be expressed in measures of peak velocity and spatial and temporal measures. One strategy predicted that clear condition utterances would be produced with less coarticulation, or less overlap of articulator movements, and more canonical values of vowel formants, achieved by using more effortful movements. Kinematic and acoustic data were collected on seven subjects in three conditions including normal and clear. Subjects were instructed to read 15 repetitions of a VCV nonsense word embedded in carrier phrases given the conditions of speaking "normally" and speaking " as if someone was in the next room who was checking for clarity and correctness of pronunciation." The principal measures included the value of the second formant of the initial and final vowels, the magnitude of spatial and temporal overlap of consonant-related movement moving into or out of the vowel segment (the "coarticulation" measure), and the peak velocity of the consonant-related opening/closing movement. Results indicated that the clear speech condition was associated with minimal changes in second formant values of vowels and a minimal reduction in coarticulation. Further, peak velocities for the clear conditions were higher in only 7/12 conditions.

In another study, Perkell, Zandipour, Matthies, & Lane (2002) tested the hypothesis that clear speech was associated with greater physiologic effort by collecting kinematic and acoustic data in seven speakers under multiple speaking conditions including 'normal' and 'clear.' They reported differences in peak movement speed, distance and duration among the conditions and among the speakers. However, only three of the seven speakers produced the clear condition utterances with movements that had larger peak speeds. The authors of both studies suggest that while there is support for an increased physiologic effort in producing clear speech by some speakers, there appears to be significant inter-individual variability. Speakers may use other combinations of parameter adjustments to produce the clear condition.

Summary of the Literature

Current research surrounding speech clarity, auditory-perceptual and acoustic data have generated valuable descriptions of how the speech signal varies as a result of a speaker's shift in intelligibility. The acoustic shifts seem to be relatively agreed upon within the research. The limited studies which include kinematic data report inter-individual variability in the way in which these acoustic shifts are achieved as speakers depart from their normal speech modes. Exploring the effects of speech clarity across kinematic, acoustic and auditory perceptual levels will yield a more complete understanding of the phenomenon of speech clarity. Specific goals attempted in this study are to describe differences in acoustic measures of clear and casual speech, characterize temporal and spatial features of articulatory movements of clear and casual speech, and to test the hypothesis that auditory perceptual scaling of clarity will be associated with kinematic indicators of physical effort.

CHAPTER III

METHODS

Speech Task

Speakers

Speakers were selected from the University of Wisconsin X-ray Microbeam Speech Production Database (XRMB-SPD). This publicly available database includes the acoustic signal and synchronous midsagittal-plane motions of eight fleshpoints recorded from 57 neurologically and communicatively healthy, native-English speaking young adults performing a range of speech and non-speech oral activities. Not all speakers could be used in this study. For technical reasons, most speakers in the dataset have some missing data. In some cases a particular record was not recorded. In other cases, a record exists, but a portion of the data was not usable. Each speaker's dataset was inspected to establish that the records containing the casual and clear speech conditions exist and to ensure that the records of interest were relatively complete. Only those speakers whose records of interest were either absent or contained large amounts of missing data were excluded from the study. This process yielded 49 speakers. The subject pool is weighted toward female speakers (29 females and 20 males). The median speaker age is 20.88 years. The speaker age ranges from 18.33 to 36.98 years. Table 1 shows the age, gender, and dialect base (i.e. place of residence during linguistically formative years) for each speaker. It can be seen that the dialect base of the speaker pool is largely from the midwestern states. Therefore, it may be reasonably assumed that the majority of speakers speak an Upper Midwest dialect of American English. Additional details of chosen

Table 1. Brief description of the speakers used in this study. The Dialect Base refers to the speaker's place of residence during his/her linguistically formative years. There were 49 speakers, 20 male and 29 females with a median age of 20.88.

Number	Subject	Gender	Age (yrs)	Dialect Base
1	07	M	27.5	Maquoketa, IA
2	08	M	22.83	Brookfield, WI
3	09	F	30.36	Whitewater, WI
4	11	M	20.04	Brookfield, WI
5	12	M	21.1	Marinette, WI
6	13	F	36.98	Rockford, IL
7	14	F	36.02	Stoughton, WI
8	16	F	20.49	Kiel, WI
9	18	M	19.38	Hudson, WI
10	20	F	25.39	Milford, MA
11	21	F	21.57	Cherry Hill, NJ
12	22	F	20.65	Mankato, MN
13	23	F	25.12	Hong Kong
14	25	F	24.19	Elroy, WI
15	26	F	23.88	Verona, WI
16	27	F	20.85	Blair, WI
17	28	M	21.58	Madison, WI
18	29	F	20.61	Milwaukee, WI
19	30	F	19.45	Edina, MN
20	31	F	19.92	New Holstein, WI
21	33	F	19.41	Minneapolis, MN
22	34	F	20.99	Amery, WI
23	35	F	24.28	Waupon, WI
24	36	F	18.33	Park Forest, IL
25	37	F	20.12	Morgan, CA
26	38	F	20.88	Great Neck, NY
27	39	F	23.77	Rochester, MN
28	40	M	19.21	Green Bay, WI
29	41	M	20.41	Milwaukee, WI
30	42	M	18.56	Greendale, WI
31	43	M	20.05	Waukesha, WI
32	44	M	20.75	Madison, WI
33	45	M	21.24	Mishawaka, IN
34	48	F	21.33	Maywood, IL
35	49	F	19.45	Madison, WI
36	50	F	34.04	Madison, WI
37	51	M	19.19	Madison, WI
38	52	F	26.46	Kewanee, IL
39	53	M	20.72	Waukesha, WI
40	54	F	21.34	Wonewoc, WI
41	55	M	26.71	Denver, CO
42	56	F	22.33	Edina, MN
43	57	M	19.6	Union Grove, WI
44	58	M	23.23	Fair Lawn, NJ
45	59	M	29.28	Sauk City, WI
46	60	F	20.19	Columbus, OH
47	61	M	20.4	Middleton, WI
48	62	F	18.38	Crofton, MD
49	63	M	20.66	Los Angelos, CA

speakers' demographic and physical characteristics can be found in the XRMB-SPD Handbook (Westbury, 1994).

All speakers were paid for their participation. All procedures were approved by the Institutional Review Board of the University of Wisconsin.

Data Acquisition and Processing

Data acquisition was performed according to the procedures described in Westbury (1994). Briefly, articulator motion is recorded by directing a narrow high-energy x-ray beam to track the mid-sagittal position of a number of 2-3 mm diameter gold pellets glued to various structures within and around the oral cavity. Figure 2 shows the approximate location of the pellets. Four pellets (T1-T4) were positioned on the mid-line surface of the tongue. Two mandibular pellets were positioned at the gum line between the mandibular incisors (MI) and between the first and second molar on the left side (mm). The final two pellets were positioned mid-line at the vermilion border of the upper (UL) and lower (LL) lips. Pellets were sampled at rates ranging from 40 to 160 Hz and the simultaneous sound pressure level signal was recorded at 22 KHz. Following acquisition, the position histories for each pellet underwent a series of processing steps. One of the processing steps involved re-expressing the data in an anatomically based Cartesian coordinate system in which the abscissa is located along the maxillary occlusal plane and the ordinate is normal to the abscissa where the central maxillary incisor meets the maxillary occlusal plane. Figure 2 shows the location of the abscissa and ordinate. The position histories of the pellets were low pass filtered at 10Hz and re-sampled at 145 Hz. For a full description of the processing steps see Westbury (1994).

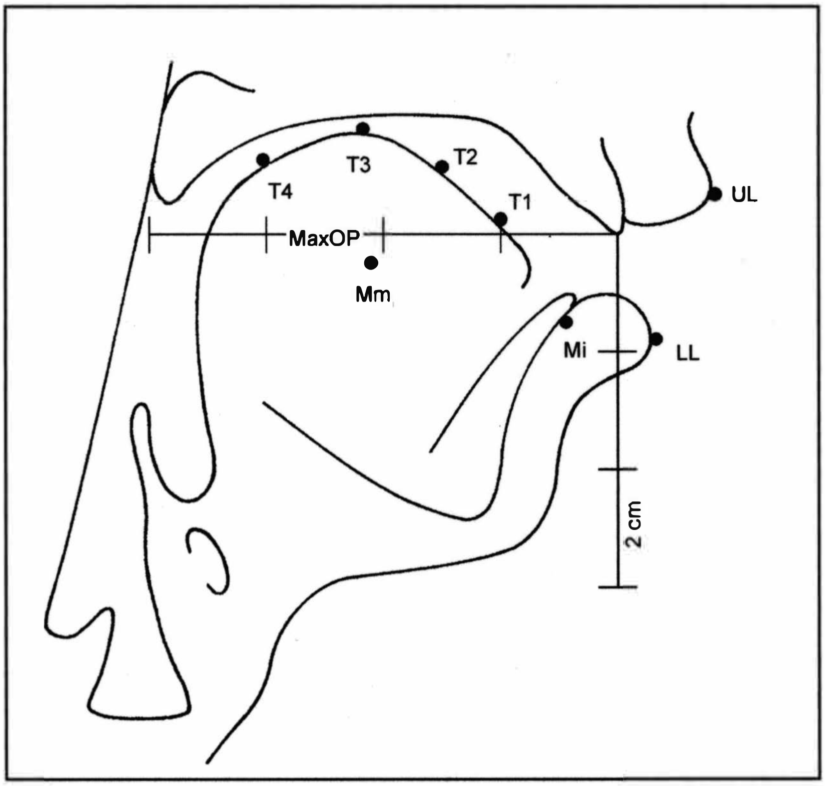

Figure 2. The placement of articulator pellets. For the purposes of this study, the mid-ventral pellet (T2), the mid-dorsal pellet (T3), the mandibular pellet (MI), the upper lip pellet (UL) and the lower lip pellet (LL) were used. This figure also shows the location of the abscissa and ordinate. The abscissa is located along the maxillary occlusal plane (MaxOP) and the ordinate is normal to abscissa where the central maxillary incisor meets the MaxOP.

Movements associated with two vocal tract events are evaluated within this study. First, T2, T3 and MI pellets trace one particular movement of interest in which the tongue and jaw movement is associated with the vocalic transition of the diphthong /ɑɪ/. Second, UL and LL movements associated with vocal tract opening for the phonetic sequence /bɑɪ/ were evaluated. Figure 2 provides a representative coordinate system for the articulatory data. Note that increasing positive 'x' values are associated with anterior movement and increasing negative ‘x’ values indicate posterior movement. Increasingly positive ‘y’ values are associated with a superior movement and increasingly negative ‘y’ values indicate an inferior movement in the vocal tract.

Speech Instructions

Speakers performed an oral reading of the sentence “combine all the ingredients in a large bowl” under clear and casual conditions. The clear condition was collected as a single 17.5 second record. To elicit the clear condition, the specific directions were to “repeat this sentence five times, very distinctly and clearly, as if you are trying to make someone understand you in a noisy environment. Do not pause between words.” Unless technically unfeasible, the first complete production of the test sentence was selected for analysis. Otherwise, the second complete production was used. The casual production of the test sentence was not collected in isolation. Rather, the sentence was part of a long list of sentence level stimuli that comprise the task set. Therefore, for the casual condition, the speaker was prompted to refer to the general instruction to recite the speech stimuli at a comfortable rate and loudness, unless otherwise instructed.

The single word ‘combine’ was extracted for acoustic, kinematic and auditory perceptual analysis in both the casual and clear conditions. This particular word was selected for a number of reasons. First, by choosing to use a single word versus using the whole sentence for this study, a more in-depth analysis was possible. Second, ‘combine’ contains the diphthong /ɑɪ/ which requires a relatively large, phonemically relevant acoustic and articulatory transition. Finally, since this word was the initial word in sentence, it was relatively easy to extract and retain acceptable auditory quality for use in a perceptual study. For each speaker, a single token of the clear and casual productions of the word was selected for inclusion in the analysis.

Auditory Perceptual Evaluation

The general goal of the auditory-perceptual evaluation was to attain a listener panel rating of the clarity difference between a speaker's clear and casual production of the test word. This information evaluates whether speakers actually produce perceptually salient differences between the clear and casual conditions.

Listeners

The listener panel consisted of 30 undergraduate and graduate students from the Western Michigan University Department of Speech Pathology and Audiology. All were native-English speaking with normal hearing.

All members of the listener panel gave informed consent to participate in the study. The consent form was in compliance with the standards set forth by the Human Subjects Institutional Review Board of Western Michigan University. Appendix A shows a sample signed consent form. Some listeners who participated in the study were fulfilling a small extra credit opportunity for a course they were taking within the Department of Speech Pathology and Audiology.

Auditory-Perceptual Evaluation Procedure

Listeners were seated in a quiet room in front of a PC computer which randomly presented the speech stimuli through an amplifier (Realistic MPA-30) and loudspeaker

(Paradigm Titan v.3). The PC computer used a 16-bit sound card to convert the digitized stimuli to analog waveforms. The loudspeaker was located about 100 cm from the listener and signals were presented at a sound pressure level of 70 dB. Stimulus presentation and response recording were controlled by *Alvin* (Gayvert and Hillenbrand, 2003), an open source, Windows-based program for controlling listening experiments. *Alvin* allows the listener to control the rate of stimulus presentation and response.

Figure 3 is an image of the computer screen through which the listener interfaced during the experiment. The following instructions were read to the listeners prior to commencing the experiment. "You will be presented with a series of a paired words read by a number of different speakers. For each presentation, the words will have differing degrees of clarity; the order of the words for each pair is random. Your task is (1) to indicate which of the words is more clear, presentation one or two, and (2) by how much. You will use the mouse to move the marker from the middle of the scale towards the presentation that is clearest. You will indicate the degree of clarity by how close you move the marker to the speech sample, the closer to either end the clearer the word. Try to make each choice match the clarity, as you perceive it." Listeners were allowed to play each stimulus pair as many times as they wished. A total of 196 signal pairs were presented (49 speakers presented four times). The entire listening experiment lasted approximately 20-30 minutes.

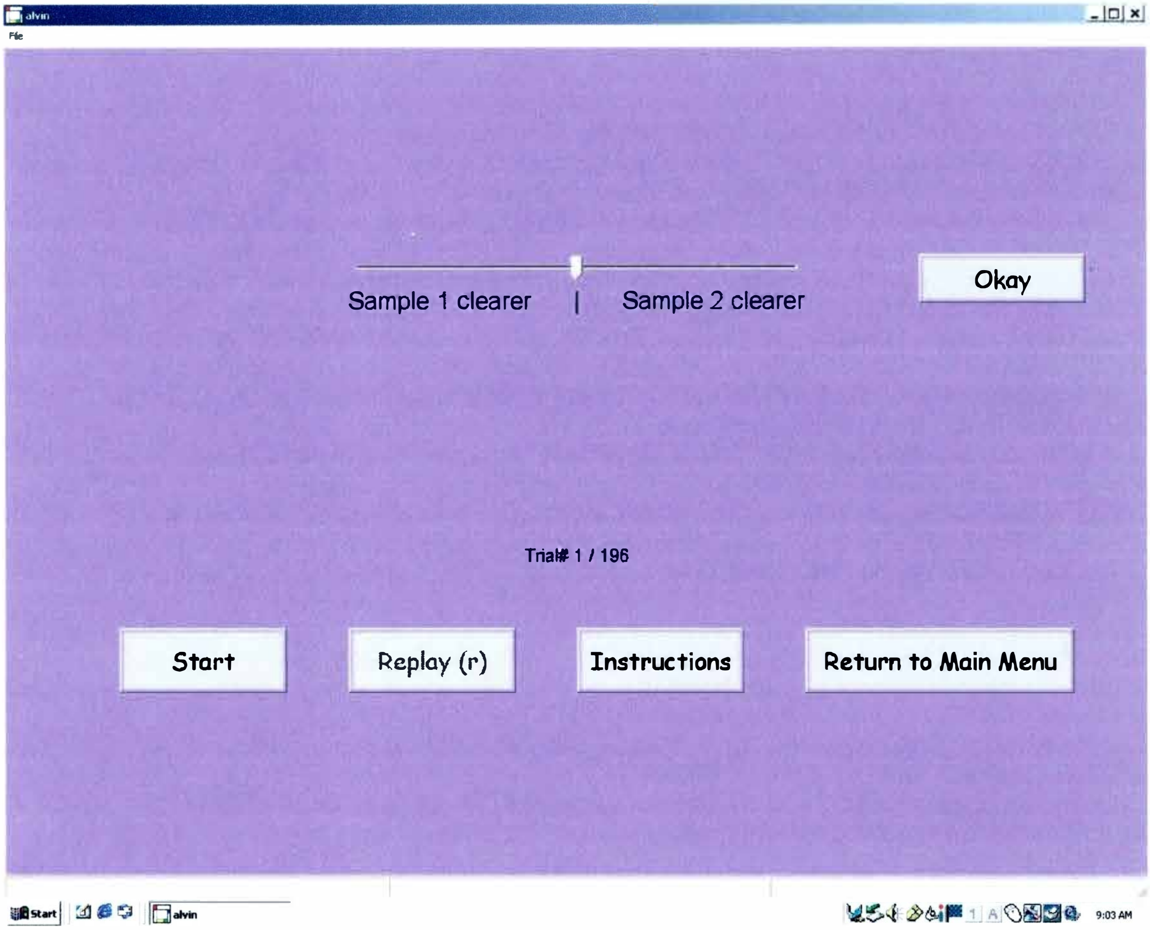

Figure 3. Listener task computer screen. This computer screen image is displayed to the listener throughout the auditory-perceptual experiment. For each presentation, the listener indicates 1. indicates which sample is the clear sample, and 2. the degree of clarity, by placing the marker along the scale in the direction of the clearer sample.

The *Alvin* program recorded listener responses in integers ranging from -500 to + 500. For example, a score of -500 would indicate that the listener perceived sample one to be much clearer than sample 2. A rating of + 500 would indicate that the listener perceived sample two to be much clearer than sample one. A rating of 0 would indicate that the listener could not distinguish the clarity of samples one and two. Given the random stimulus order, the results for each listener were rescaled so that positive ratings would always be associated with the stimulus in the clear conditions. For each stimulus pair (i.e., speaker), a *listener rating* was derived based on the mean of that listener's four ratings. A *panel rating* was then derived based on the mean listener rating across the 30 listeners.

Reliability of Auditory Perceptual Ratings

This study correlated ratings within and across listeners to evaluate the strength of association between a given pair of ratings. Individual listeners rated the stimulus set four times. Therefore, it is possible to evaluate the association between ratings within a listener. In other words, ratings for the first presentation of the stimulus set may be correlated with the second presentation, the second presentation with the third presentation and so forth. This approach evaluates within-listener reliability. Alternatively, listener one's ratings may be correlated with listener two's ratings, listener one's ratings correlated with listener three's ratings and so forth. This evaluates the strength of association of the ratings of two different listeners and addresses reliability across the listener panel.

Figure 4 plots a frequency histogram showing the results of the within listener correlation analysis. This histogram summarizes the Pearson correlation coefficients for all within listener comparisons pooled across the speaker panel (six correlations per listener across 30 listeners). The distribution ranges from a minimum correlation of -0.05 to a maximum correlation of 0.69. The mean correlation is 0.4 and the standard deviation is 0.16.

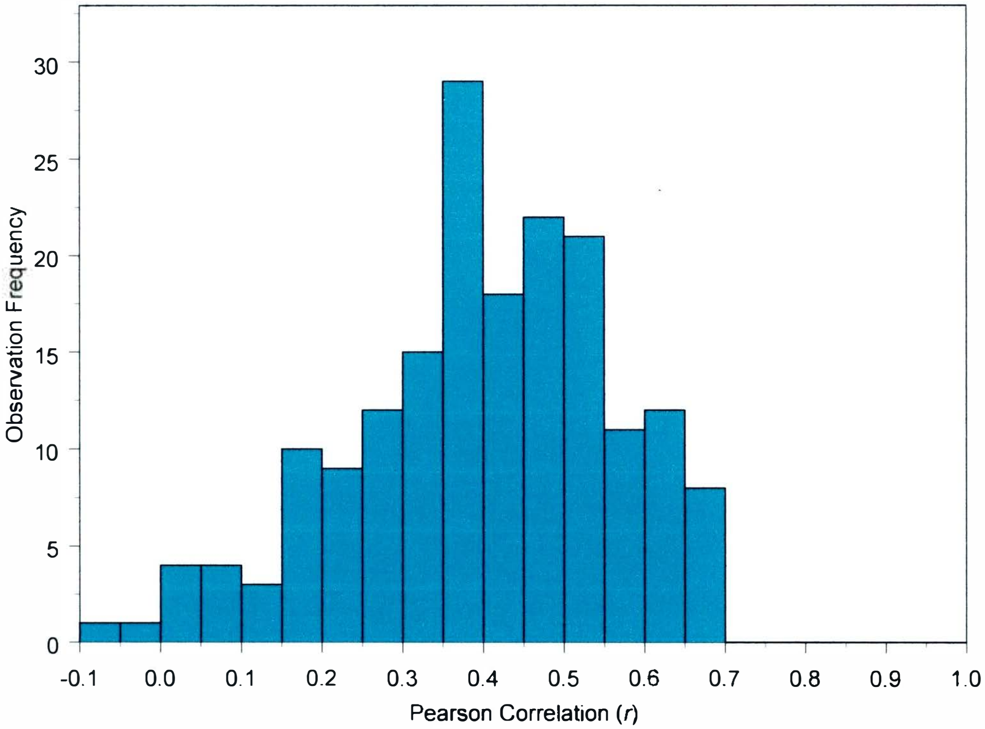

Figure 4. Within listener correlation. Frequency histogram of the Pearson correlation coefficients for all within listener comparisons pooled across the speaker panel (six correlations per listener across 30 listeners).

Figure 5 plots a frequency histogram showing the Pearson correlation coefficients for all possible pairs of listeners that comprise the listener panel. These cross rater correlations range from 0.01 to 0.85. The distributions of correlations have a mean value of 0.57 and a standard deviation of 0.17.

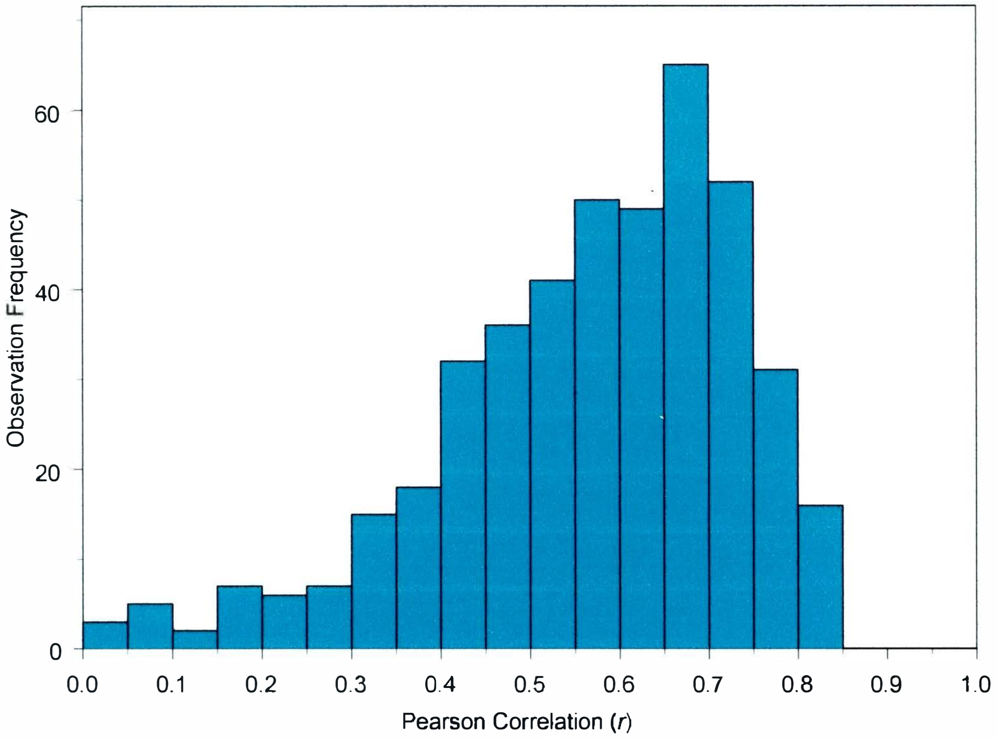

Figure 5. Between listener correlation. Frequency histogram showing Pearson correlation coefficients for all possible pairs of listeners that comprise the listener panel.

Acoustic Analysis of Test Words

Acoustic Phonetic Segment Durations

For each speech token, the durations of the six acoustic phonetic elements that make up the word /kʌmbɑɪn/ were measured. This analysis was performed using a freely available acoustic analysis software package (TF32, Milenkovic). All segmentation was performed using a combined waveform-spectrogram (Bandwidth=300 Hz) display. Figure 6 shows a representative example of the speech token with the acoustic phonetic boundaries in place. The locations of the acoustically defined phonetic boundaries were based on the following segmentation rules.

/k/: The onset of the stop was defined by the sudden increase in mid-frequency energy. The offset was defined by the onset of the following vowel.

/ʌ/: The onset of the neutral vowel was determined by the presence of frequency bands consistent with the first two formants. The offset was defined by the onset of the following bilabial nasal.

/m/: Onset of bilabial nasal was judged as the abrupt loss of energy in the mid to high frequencies. The offset was defined by the onset of the following voiced bilabial stop.

/b/: The onset of the voiced bilabial stop was defined by the sudden increase in noise energy in the low to mid frequencies. The offset was defined by the onset of the following vowel.

/ɑɪ/: The onset of the diphthong was determined by the presence of frequency bands consistent with the first two formants. The offset was defined by the onset of the following alveolar nasal.

/n/: Onset of the alveolar nasal was defined as the point where an abrupt loss of energy in mid to high frequencies occurred. The offset was defined by the cessation of acoustic energy.

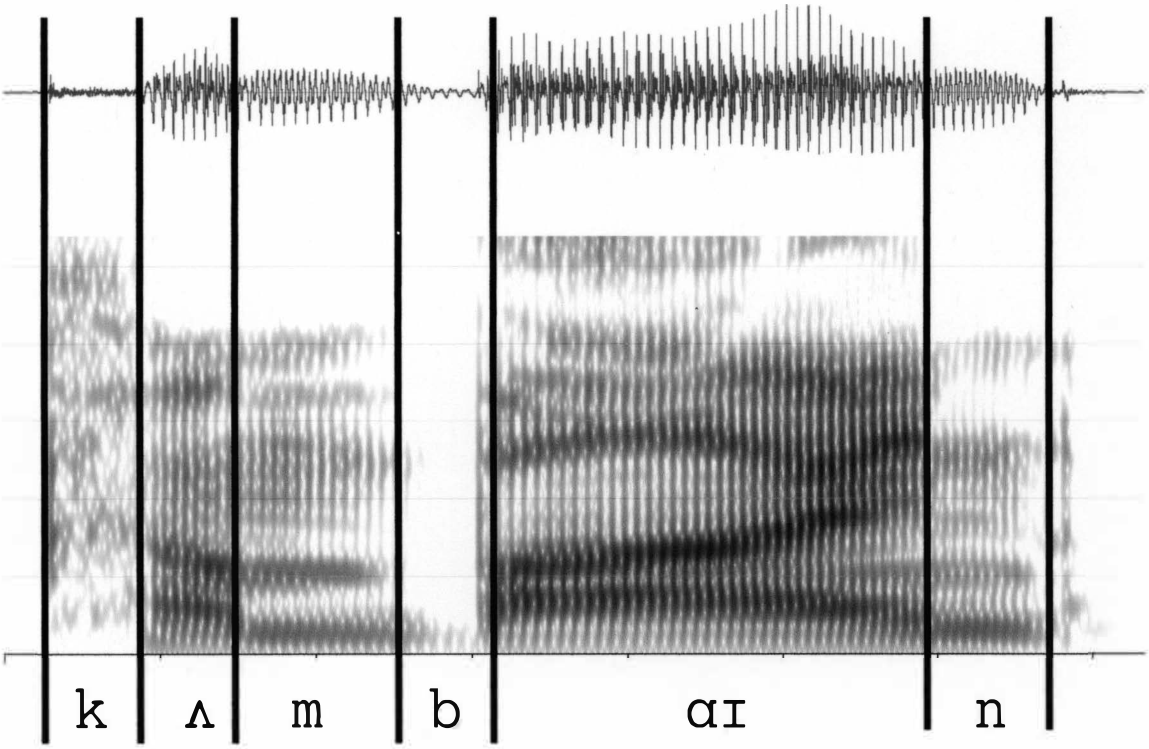

Figure 6. Acoustic segmentation. A plot of the sound pressure level wave form of the word "combine" spoken by an adult male. The phonetic segmentation follows conventionally accepted segmentation rules. The boundaries of the diphthong /aI/ are associated with the appearance of the associated formant frequency bands and F2 transition from a low to a mid range.

Reliability

A repeated analysis on ten randomly selected speakers in the clear and casual conditions was performed by the same experimenter to establish reliability of segmentation of each acoustic token and total word duration. Correlation measures were found to be high. The /k/ was found to have a correlation of 0.83, the /ʌ/ had 0.95, the /m/ had 0.96, the /b/ had a 0.99, the /ɑɪ/ had 0.99, the /n/ had 0.89, and the total word duration had a correlation of 0.99.

First and Second Formant Transition

Measures of the extent and duration of the first (F1) and second (F2) formant transitions for the diphthong /ɑɪ/ were extracted for each token. This analysis was performed using Speech Tool, an acoustic analysis software package (Speech Tool, Hillenbrand and Gayvert). For an individual token, an LPC-based spectrum was derived and the peaks in the spectrum were identified at 10 msec intervals. This automated routine makes no attempt to identify any peak as a particular formant. Following the peak picking algorithm, the experimenter viewed these peaks overlaid on a spectrogram based on the LPC spectrum. From this view the experimenter then hand edited the peaks to ensure they corresponded to the F1 and F2.

The onset and offset of the F1 and F2 transitions were determined using a procedure similar to Kent, Kent, Weismer, Martin, Sufit, Brooks and Rosenbek (1989). The onset of each formant (F1 and F2 were analyzed separately) was determined by a sustained 20 Hz/20 msec rate change. The offset of the formant was determined as the point where the rate change dropped below the 20 Hz/20 msec threshold. This analysis was performed using a custom written Matlab program that allowed the experimenter to select formant onsets and offsets. The F1 and F2 onset and offset frequencies and their time locations were extracted. The following measures were collected or derived for use as dependent measures.

F1 onset (Hz): Frequency when F1 transition onset began

F1 offset (Hz): Frequency when F1 transition ends

F1 range (Hz): F1 offset – F1 onset

F1 duration (msec): Duration of F1 transition

F1 transition rate (Hz/msec): F1 range/F1 duration

F2 onset (Hz): Frequency when F2 transition onset began

F2 offset (Hz): Frequency when F2 transition ends

F2 range (Hz): F2 offset – F2 onset

F2 duration (msec): Duration of F2 transition

F2 transition rate (Hz/msec): F2 range/F2 duration

Fundamental Frequency (F_0)

The fundamental frequency history was derived for the test word using an autocorrelation algorithm described by Milenkovic (TF32, Milenkovic). Analysis was limited to the F_0 contour associated with production of the diphthong /ɑɪ/. The following statistical descriptors of the F_0 contour were derived using a custom Matlab program.

Mean F_0: Mean F_0 value over the contour

Min F_0: Smallest F_0 value over the contour

Max F_0: Largest value over the contour

F_0 Range: Max F_0 - Min F_0

F_0 SD: Standard deviation of the F_0 over the contour

Sound Pressure Level

An estimate of sound pressure level was derived using a freely available acoustic analysis software package (TF32, Milenkovic). The RMS signal was generated from the sound pressure waveform using a 20 msec averaging window. Sound pressure level was expressed in decibels (dB) relative to a 5 Volt reference. Mean and Maximum sound pressure level was derived for the diphthong /ɑɪ/ in the clear and casual conditions.

Articulatory Kinematic Analysis

Kinematic analysis was restricted to two tongue (T2 & T3), one jaw (MI) and two lip (UL & LL) fleshpoints (see Figure 2). Inspection of the records prior to analysis revealed that these fleshpoints exhibited the most prominent movement during production of /ɑɪ/. First order time derivatives of the ‘x’ and ‘y’ positions of each fleshpoint were derived and used to calculate the velocity history of each position history (*dx/dt*, *dy/dt*). The velocity vectors were used to determine fleshpoint speed ($[(dx/dt)^2 + (dy/dt)^2]^{1/2}$). Figure 7 plots the sound pressure waveform, the T2 speed history, and T2 spatial position associated with the production of the test word ‘combine’. Note the large peak in the speed history coincident with the diphthong /ɑɪ/. This kinematic event is associated with the largely superior movement of the fleshpoint (bottom plot). The movement onset/offset was defined as the point in time associated with a change from deceleration

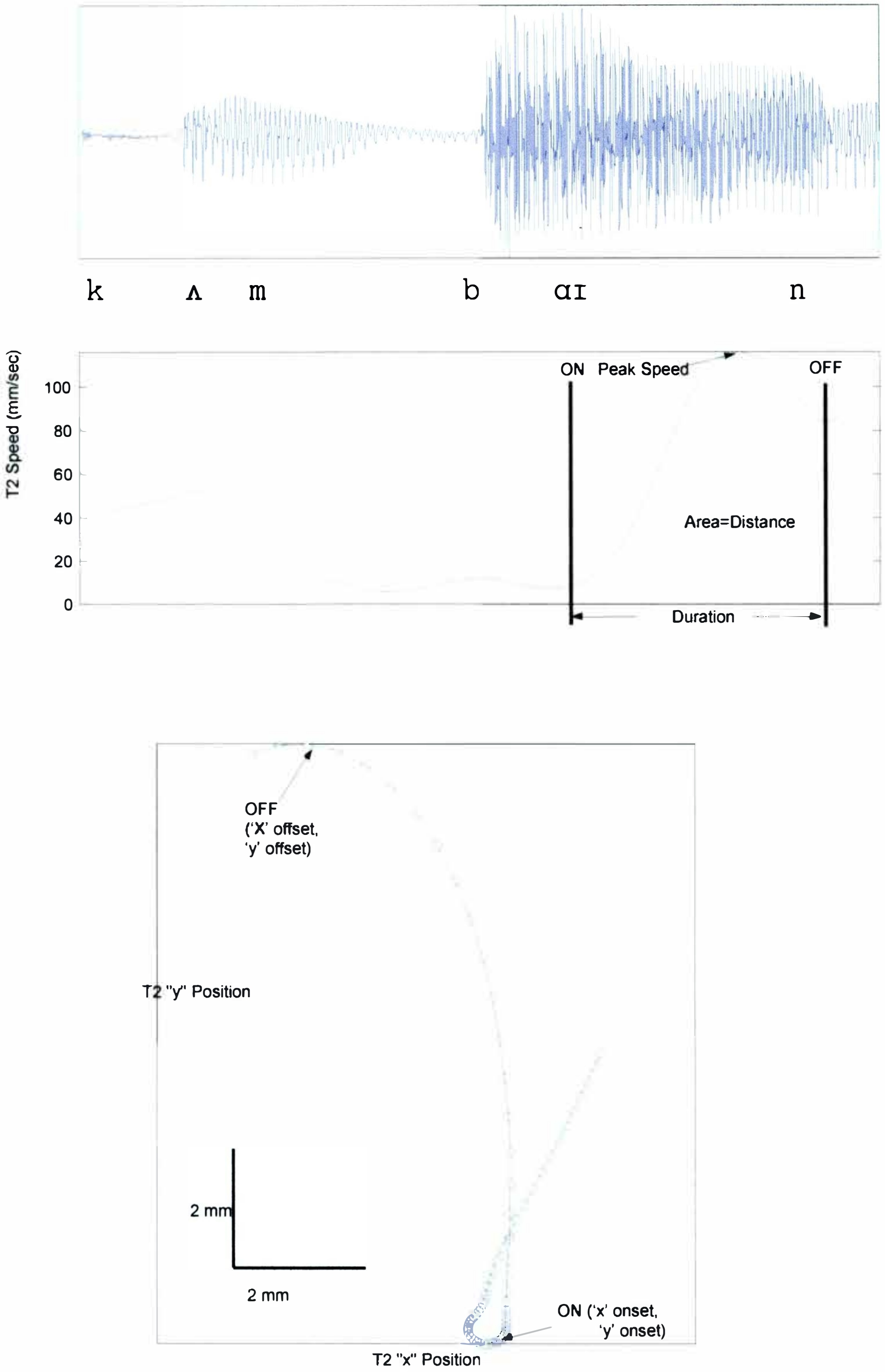

Figure 7. Plots associated with the test word 'combine.' (Top panel) sound pressure waveform, (middle panel) T2 speed history, and (bottom panel) T2 spatial position.

(negative slope in speed history) to acceleration (positive slope in speed history). These time values defined the temporal extent of the movement.

A number of kinematic features were selected from this movement.

Peak Movement Speed: Maximum speed between onset and offset of movement

Movement Duration: Movement offset time – movement onset time

Movement Distance: Distance moved between movement onset and offset time.

Initial Spatial Position: Spatial position associated with the movement onset time.

Final Spatial Position: Spatial position associated with the movement offset time.

Statistical Analysis

The data extracted for this study allow two general statistical questions to be addressed. First, how does speech produced under clear and casual conditions differ along auditory-perceptual, acoustic and articulatory dimensions? Second, are clarity-based differences in acoustic/articulatory kinematic measures associated with auditory-perceptual ratings of clarity differences?

The first question addresses what speakers do when they are asked to speak clearly. For the auditory-perceptual data, a one-sample *t*-test was performed to evaluate the hypothesis that listener ratings of clarity differences were significantly different from zero. A two sample *t*-test was used to evaluate whether the male and female speakers exhibited different auditory-perceptual ratings. For the acoustic and articulatory kinematic measures, clarity based differences were evaluated using a series of individual repeated measures analyses of variance (ANOVA). Speech clarity (Clarity) was considered a within-subjects factor and gender (Gender) was treated as a between-subjects factor. Given the large number of measures analyzed, the *p*-values were adjusted for multiple comparisons ($p < .0008$) in order to hold the Type I error rate to less than 5 %.

CHAPTER IV

RESULTS

Findings

Auditory Perceptual Findings

Recall that the listener's task in the auditory perceptual experiment was to rate the difference in clarity between the casually and clearly produced test words for each of the speakers. Therefore, the listener may judge the clear test word to have greater clarity, equivalent clarity or less clarity than the casually produced test word. Figure 8 plots a frequency histogram of the mean panel ratings for the speaker pool. It may be seen in the figure that the listener panel showed a distribution of mean ratings that are typically greater than zero indicating that the clear productions were judged to have greater clarity than the casual productions. The mean panel rating for the speaker group is 150 with a standard deviation of 85. A one sample t-test was performed to evaluate the hypothesis that the panel ratings of the speaker pool were statistically different from zero (i.e. no judged clarity difference). Results were significant ($t = 12.29, p < .0008$). A two sample t-test comparing males and female speaker ratings suggest that the panel ratings did not differ as a function of gender ($t=-1.36, p < .20$). However, it must be noted that the panel ratings for individual speakers varies widely, spanning from -10 to +299. Panel ratings for many speakers were near zero suggesting minimal differences in perceived clarity of the clear and casual productions.

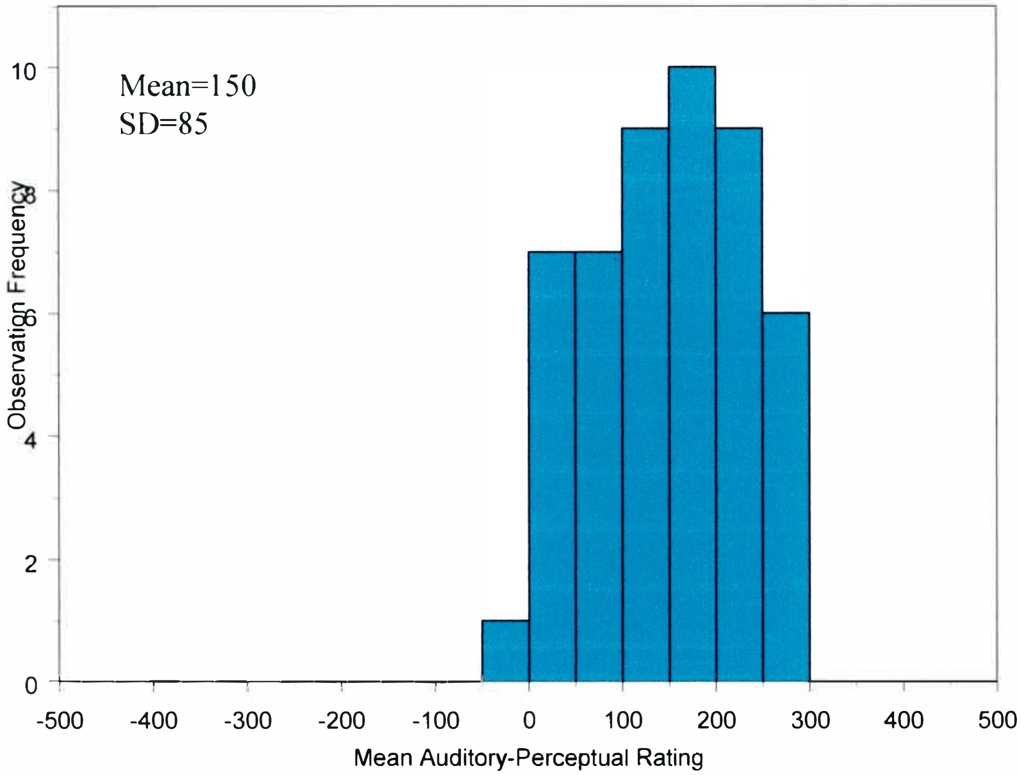

Figure 8. Frequency histogram of mean panel rating for speaker pool.

Figure 9 plots the standard deviation of the mean panel rating as a function of the mean rating. This figure shows the listeners increased their variability as mean ratings increase. In other words as listeners rated higher degrees of clarity, the variability of the rating by the panel also increased. The Pearson correlation relating the mean rating with the standard deviation was 0.74 ($p < .0005$).

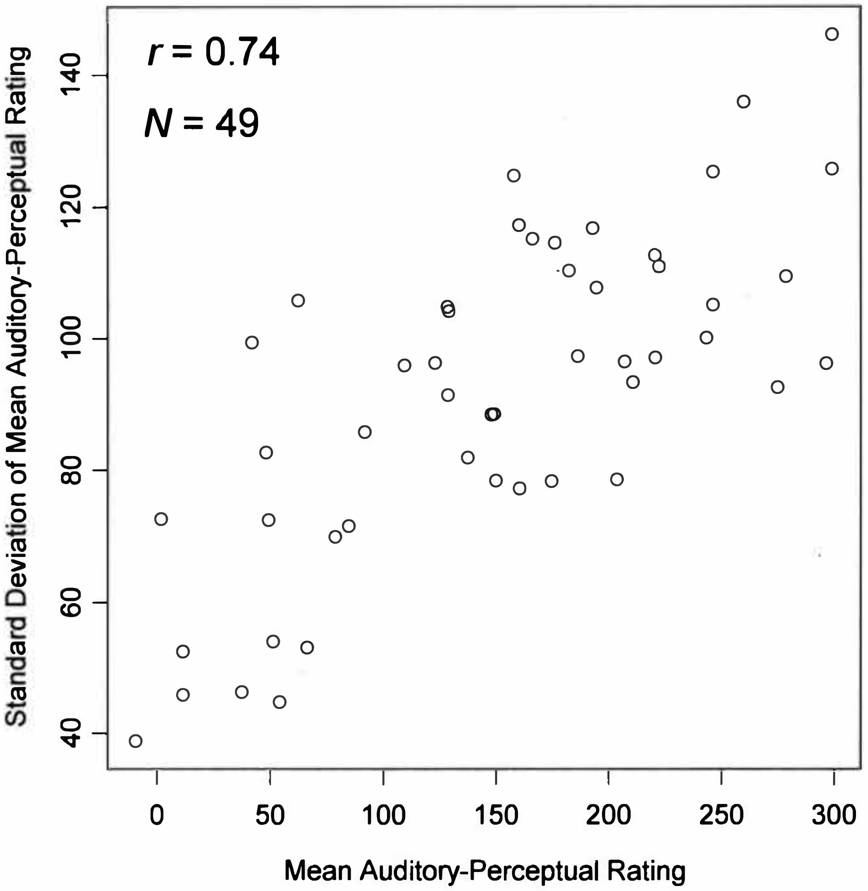

Figure 9. Scatter plot showing the relationship between mean and standard deviation of the auditory-perceptual ratings of clarity. Each data point represents a speaker.

Acoustic Characteristics of Clear Speech

Acoustic Segment Durations:

Table 2 shows the means and standard deviations of acoustic segment durations organized by gender and clarity. Table 3 summarizes the results of a series of repeated measures ANOVAs performed on the duration data. The results indicate that the duration of all six segments and word duration increased during clear speech. Note that there was neither a significant gender effect nor a significant clarity-by-gender interaction.

The data in Table 2 show absolute change in duration. An alternative way to evaluate this data is to express the clarity based changes in duration as a percentage of the duration in the casual condition (i.e. $[\text{duration}_{\text{clear}} - \text{duration}_{\text{casual}}]/\text{duration}_{\text{casual}}*100$). This calculation was performed for each speaker-segment condition.

Table 2. Summary of acoustic duration results. Group means and standard deviations (in parentheses) are organized according to acoustic phonetic segment, gender and condition. M=male, F=female.

	Condition	k	ʌ	m	b	ɑɪ	n	total
M	Clear	61.46 (21.42)	46.29 (16.62)	132.14 (17.06)	25.26 (13.20)	266.29 (53.79)	70.94 (13.58)	602.53 (90.92)
	Casual	48.96 (12.41)	27.93 (9.0)	91.99 (16.31)	15.67 (3.70)	178.92 (39.07)	44.52 (15.1)	408.02 (64.59)
F	Clear	61.72 (16.09)	39.9 (12.82)	117.48 (24.42)	18.07 (8.77)	248.85 (52.59)	66.58 (16.88)	552.62 (96.17)
	Casual	50.74 (11.75)	32.39 (12.28)	88.32 (16.92)	15.09 (6.05)	178.25 (35.10)	46.87 (12.82)	411.63 (58.71)

Table 3. Summary of repeated measures ANOVAs performed on acoustic duration measures.

Factor	*df*	*F* values						
		k	ʌ	m	b	ɑɪ	n	Total
Clarity	1	20.33*	37.51*	94.78*	22.71*	113.13*	66.08*	152.14*
Gender	1	0.08	0.1	4.07	3.5	0.68	0.1	1.54
Clarity -by- Gender	1	0.09	6.61	2.16	6.27	1.27	1.4	3.87
Error	47							

* $p < 0.0008$

First Formant (F1):

Table 4 shows the means and standard deviations of F1 transition measurements in /ɑɪ/ organized by gender and clarity. Table 5 summarizes the results of a series of repeated measures ANOVAs performed on the F1 data. The results indicate the F1 measurements are significantly higher at onset, the transition range is larger, and transition duration is longer during clear speech. Measures at offset show a marginal, however not significant lower position in the clear condition. Note that females have a higher F1 onset, F1 offset and a subsequent greater range than males. There was no significant clarity-by-gender interaction.

Table 4. Summary of first formant (F1) transition results. M=males, F=females.

	Condition	F1 onset (Hz)	F1 offset (Hz)	F1 transition range (Hz)	F1 transition duration (msec)	F1 transition rate (Hz/msec)
M	Clear	760.20 (62.69)	406.20 (88.02)	-354.00 (124.78)	161.10 (61.96)	-2.56 (1.31)
	Casual	716.90 (66.97)	470.50 (98.25)	-246.30 (17.33)	94.43 (41.63)	-3.07 (2.02)
F	Clear	1001.07 (120.64)	528.03 (113.81)	-473.03 (165.45)	134.33 (50.14)	-4.02 (2.09)
	Casual	950.69 (111.01)	566.17 (95.78)	-384.52 (156.79)	92.82 (24.50)	-4.32 (1.78)

Table 5. Summary of repeated measures ANOVAs performed on the first formant (F1) transition measures.

Factor	*df*	*F* values				
		F1 onset	F1 Offset	F1 transition range	F1 transition duration	F1 transition rate
Clarity	1	15.82*	8.05	17.46*	51.43*	1.91
Gender	1	82.73*	22.4*	13.16*	1.73	9.01
Clarity -by- Gender	1	0.09	0.53	0.17	2.78	0.13
Error	47					

* $p < 0.0008$

Second Formant (F2):

Table 6 shows the means and standard deviations of F2 transition measurements organized by gender and clarity. Table 7 summarizes the results of a series of repeated measures ANOVAs performed on the F2 data. The results indicate the F2 measurements are higher at offset, the transition range is larger, and transition duration is longer during clear speech and the F2 measures are marginally, however not significant, higher at onset. The transition rate is faster in the clear condition. Note that females have a higher F2 onset and F2 offset than males. There was no significant clarity-by-gender interaction.

Table 6. Summary of second formant (F2) transition results. M=males, F=females.

	Condition	F2 onset (Hz)	F2 offset (Hz)	F2 transition range (Hz)	F2 transition duration (msec)	F2 transition rate (Hz/msec)
M	Clear	1069.55 (92.28)	1970.50 (212.34)	900.95 (237.89)	205.02 (55.51)	4.50 (1.01)
	Casual	1006.15 (105.85)	1767.90 (150.53)	761.75 (163.00)	139.40 (37.22)	5.81 (2.03)
F	Clear	1333.97 (161.43)	2379.66 (220.87)	1045.69 (246.37)	184.56 (43.28)	5.86 (1.46)
	Casual	1327.55 (124.79)	2180.52 (178.83)	852.97 (161.14)	136.20 (31.49)	6.49 (1.54)

Table 7. Summary of repeated measures ANOVAs performed on the second formant (F2) transition measures.

Factor	*df*	*F* values				
		F2 onset	F2 Offset	F2 transition range	F2 transition duration	F2 transition rate
Clarity	1	5.66	62.45*	38.14*	76.81*	18.18*
Gender	1	73.39*	66.38*	4.84	1.32	6.94
Clarity-by-Gender	1	3.77	0.01	0.99	1.76	2.19
Error	47					

* $p < 0.0008$

Fundamental Frequency (F_o):

Table 8 shows the means and standard deviations of F_o results organized by gender and clarity. Table 9 summarizes the results of a series of repeated measures ANOVAs performed on the F_o data. The results indicate there was no significant clarity effect. F_0 measurements are significantly higher in females than men for mean, minimum, and standard deviation across clarity conditions. There is no significant clarity-by-gender interaction.

Table 8. Summary of fundamental frequency (F_0) results. M=males, F=females.

Gender	Condition	Mean F_0 (Hz)	SD F_0 (Hz)	Min F_0 (Hz)	Max F_0 (Hz)	F_0 Range (Hz)
M	Clear	129.36 (20.83)	6.92 (5.17)	118.79 (18.34)	139.60 (25.04)	20.82 (14.66)
	Casual	130.79 (23.45)	6.98 (5.48)	120.51 (19.35)	139.64 (290.00)	19.13 (14.20)
F	Clear	219.55 (29.18)	8.89 (10.36)	204.85 (35.07)	234.49 (36.84)	29.64 (38.94)
	Casual	217.18 (29.00)	6.80 (6.52)	209.90 (26.58)	229.75 (35.61)	19.85 (17.32)

Table 9. Summary of repeated measures ANOVAs performed on fundamental frequency (F_0) measures.

Factor	*df*	*F* values				
		Mean F_0	SD F_0	Min F_0	Max F_0	F_0 Range
Clarity	1	0.03	0.62	0.49	1.68	0.8
Gender	1	150.68*	181.69*	107.02*	0.69	0.23
Clarity -by- Gender	1	0.47	0.15	0.51	0.84	0.89
Error	47					

* $p < 0.0008$

Sound Pressure Level (SPL):

Table 10 shows the means and standard deviations of SPL measurements organized by gender and clarity. Table 11 summarizes the results of a series of repeated measures ANOVAs performed on the SPL data. The results indicate no significant clarity effect. There is no significant gender effect nor clarity-by-gender interaction.

Table 10. Summary of sound pressure level (SPL) results.

	Mean SPL (dB)		Max SPL (dB)	
Condition	M	F	M	F
Clear	1.66 (0.53)	1.53 (0.54)	2.34 (0.78)	2.16 (0.89)
Casual	1.58 (0.48)	1.60 (0.39)	2.17 (0.7)	2.13 (0.55)

Table 11. Summary of repeated measures ANOVAs performed on sound pressure level (SPL) measures.

Factor	*Df*	*F* values	
		Mean SPL	Max SPL
Clarity	1	0.01	0.66
Gender	1	0.2	0.39
Clarity-by-Gender	1	1.36	0.37
Error	47		

Articulatory Kinematic Features of Clear Speech

Bear in mind the coordinate system used to reference the articulator movements in which the horizontal or x axis represents movement of articulators posterior and anterior and the vertical or y axis traces movement with relative height in the vocal tract. Therefore a relatively positive vertical and positive horizontal position is a higher and more anterior movement, respectively. Alternatively, a relatively negative vertical and a negative horizontal position would reflect a low and posterior position.

Tongue Blade Pellet (T2):

Table 12 shows the means and standard deviations of kinematic results for diphthong-related movement of T2 organized by gender and clarity. Table 13 summarizes the results of a series of repeated measures ANOVAs performed on the T2 data. The results of the T2 position indicate no difference at onset, a significantly higher and more forward position at offset, and the duration and distance are longer in the clear condition. There is no significant gender effect nor clarity-by-gender interaction.

Table 12. Summary of kinematic results for diphthong-related movement of tongue blade (T2). M=males, F=females.

		Tongue Blade (T2)						
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
	Condition	Horizontal (mm)	Vertical (mm)	Horizontal (mm)	Vertical (mm)	Duration (msec)	Distance (mm)	Peak Speed (mm/sec)
M	Clear	-30.02	-1.79	-26.97	14.6	213.21	17.88	143.3
		(4.53)	(4.93)	(3.31)	(3.93)	(49.00)	(4.29)	(32.12)
	Casual	-29.95	-0.79	-29.4	13.2	167.46	15.53	152.27
		(4.33)	(2.5)	(3.96)	(3.52)	(32.56)	(4.14)	(32.03)
F	Clear	-29.37	-2.02	-27.63	14.34	204.02	17.59	150.43
		(3.62)	(3.27)	(3.34)	(2.96)	(44.65)	(3.43)	(29.74)
	Casual	-30.18	-0.31	-28.59	12.65	164.78	14.18	140.11
		(3.37)	(3.16)	(3.18)	(2.9)	(30.53)	(3.54)	(30.35)

Table 13. Summary of repeated measures ANOVAs performed on measures of diphthong-related movement of tongue blade (T2).

T2								
		F values						
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
Factor	*df*	Horizontal	Vertical	HorizoNtal	Vertical	Duration	Distance	Peak Speed
Clarity	1	1.78	7.17	25.28*	31.80*	37.4*	32.39*	0.01
Gender	1	0.001	0.97	0.00	0.13	0.212	0.33	0.02
Clarity-by-Gender	1	0.2	0.29	5.8	0.40	0.046	1.37	3.31
Error	43							

* $p < 0.0008$

Tongue Blade Pellet (T3):

Table 14 shows the means and standard deviations of kinematic results for diphthong-related movement of T3 organized by gender and clarity. Table 15 summarizes the results of a series of repeated measures ANOVAs performed on the T3 data. The results indicate the T3 position is significantly lower at onset, higher and more forward at offset, the duration and distance are longer and the peak speed is greater in the clear condition. There is no significant gender effect nor clarity-by-gender interaction.

Table 14. Summary of kinematic results for diphthong-related movement of tongue blade (T3). M=males, F=females.

		Tongue Blade (T3)						
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
	Condition	Horizontal (mm)	Vertical (mm)	Horizontal (mm)	Vertical (mm)	Duration (msec)	Distance (mm)	Peak Speed (mm/sec)
M	Clear	-45.41	0.85	-40.77	16.14	208.51	17.11	147.10
		(5.73)	(4.90)	(5.23)	(4.21)	(42.03)	(5.47)	(40.14)
	Casual	-45.75	2.90	-42.93	13.90	158.28	12.44	126.80
		(5.69)	(3.07)	(5.03)	(4.12)	(30.14)	(4.33)	(29.41)
F	Clear	-43.52	3.54	-40.61	13.99	190.47	12.13	112.29
		(3.99)	(4.23)	(4.17)	(3.92)	(41.10)	(4.11)	(36.29)
	Casual	-44.05	5.02	-41.24	12.30	144.68	8.75	96.63
		(4.19)	(3.80)	(4.17)	(3.77)	(37.56)	(3.88)	(33.31)

Table 15. Summary of repeated measures ANOVAs performed on measures of diphthong-related movement of tongue blade (T3).

T3								
		F values						
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
Factor	*df*	Horizontal	Vertical	Horizontal	Vertical	Duration	Distance	Peak Speed
Clarity	1	3.06	24.17*	19.15*	40.13*	42.67*	62.92*	16.48*
Gender	1	1.33	4.02	0.44	2.71	2.89	12.1	11.47
Clarity-by-Gender	1	0.15	0.75	6.70	0.22	0.06	1.11	0.14
Error	43							

* $p < 0.0008$

Mandibular Incisor (MI):

Table 16 shows the means and standard deviations of kinematic results for diphthong-related movement of MI organized by gender and clarity. Table 17 summarizes the results of a series of repeated measures ANOVAs performed on the MI data. The results indicate the MI measurements are significantly lower at onset, the duration and distance are longer in the clear condition. There is no significant gender effect nor clarity-by-gender interaction.

Table 16. Summary of kinematic results for diphthong-related movement of mandibular incisor (MI). M=males, F=females.

Mandibular Incisor (MI)								
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
	Condition	Horizontal (mm)	Vertical (mm)	Horizontal (mm)	Vertical (mm)	Duration (msec)	Distance (mm)	Peak Speed (mm/sec)
M	Clear	-1.52 (2.55)	-14.28 (3.96)	-1.33 (1.70)	-9.25 (3.10)	190.34 (43.35)	5.35 (2.64)	41.66 (17.04)
	Casual	-1.71 (2.88)	-12.99 (3.30)	-1.44 (2.12)	-9.48 (2.51)	134.43 (35.77)	3.73 (2.59)	41.43 (25.99)
F	Clear	-3.27 (2.03)	-13.68 (3.18)	-2.12 (1.62)	-8.19 (2.17)	179.53 (46.63)	5.81 (2.76)	49.83 (20.71)
	Casual	-3.22 (1.95)	-12.17 (2.72)	-2.29 (1.64)	-8.64 (1.78)	122.36 (35.06)	3.76 (2.37)	43.26 (21.42)

Table 17. Summary of repeated measures ANOVAs performed on measures of diphthong-related movement of the mandibular incisor (MI).

MI								
		F values						
		Position at Movement Onset		Position at Movement Offset			Movement characteristics	
Factor	*df*	Horizontal	Vertical	Horizontal	Vertical	Duration	Distance	Peak Speed
Clarity	1	0.09	16.42*	0.77	1.12	77.78*	22.36*	1.26
Gender	1	5.8	0.59	2.63	1.59	1.28	0.04	0.58
Clarity-by-Gender	1	0.00	0.00	0.32	0.25	0.04	0.07	1.14
Error	42							

* $p < 0.0008$

Upper Lip (UL):

Table 18 shows the means and standard deviations of a summary of kinematic results for vocal tract opening movement of the UL organized by gender and clarity. Table 19 summarizes the results of a series of repeated measures ANOVAs performed on the UL data. The results indicate the UL measurements are significantly lower at onset in the clear condition. Note that males have a more forward posture at onset and offset. There is no significant clarity-by-gender interaction.

Table 18. Summary of kinematic results for vocal tract opening movement of the upper lip (UL). M=males, F=females.

		Upper Lip (UL)						
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
	Condition	Horizontal (mm)	Vertical (mm)	Horizontal (mm)	Vertical (mm)	Duration (msec)	Distance (mm)	Peak Speed (mm/sec)
M	Clear	15.33 (1.99)	10.42 (2.97)	13.97 (2.23)	13.90 (2.63)	206.38 (50.89)	4.13 (1.83)	33.78 (15.35)
	Casual	15.44 (2.25)	11.77 (3.41)	14.18 (1.94)	13.79 (3.00)	178.52 (27.65)	3.42 (1.47)	31.05 (12.52)
F	Clear	11.34 (1.15)	10.03 (2.82)	9.74 (1.75)	13.23 (3.04)	203.87 (61.13)	4.15 (1.98)	31.21 (11.83)
	Casual	11.53 (1.29)	11.01 (2.21)	10.31 (1.37)	13.29 (2.84)	202.06 (48.82)	3.68 (1.32)	28.65 (8.56)

Table 19. Summary of repeated measures ANOVAs performed on measures for vocal tract opening movement of the upper lip (UL).

UL								
		F values						
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
Factor	*df*	Horizontal	Vertical	Horizontal	Vertical	Duration	Distance	Peak Speed
Clarity	1	0.75	20.31*	3.93	3.40	2.15	6.15	3.33
Gender	1	62.43*	0.47	51.83*	0.42	0.67	0.04	0.89
Clarity-by-Gender	1	0.26	0.38	1.82	0.98	1.03	0.21	0.15
Error	40							

* $p < 0.0008$

Lower Lip (LL):

Table 20 shows the means and standard deviations of kinematic results for vocal tract opening movement of the LL organized by gender and clarity. Table 21 summarizes the results of a series of repeated measures ANOVAs performed on the LL data. The results indicate the LL measurements are significantly lower at offset and the duration and distance are greater in the clear condition. Note that males have a more forward posture at onset and offset. There is no significant clarity-by-gender interaction.

Table 20. Summary of kinematic results for vocal tract opening movement of the lower lip (LL). M=males, F=females.

		Lower Lip (LL)						
		Position at Movement Onset		Position at Movement Offset		Movement characteristics		
	Condition	Horizontal (mm)	Vertical (mm)	Horizontal (mm)	Vertical (mm)	Duration (msec)	Distance (mm)	Peak Speed (mm/sec)
M	Clear	15.04 (2.34)	-6.28 (2.97)	8.29 (3.35)	-22.49 (4.36)	202.17 (29.99)	18.02 (4.08)	176.52 (43.61)
	Casual	14.83 (2.71)	-7.02 (3.50)	9.07 (3.63)	-20.90 (5.27)	157.23 (19.79)	15.29 (5.36)	173.77 (60.32)
F	Clear	11.84 (1.75)	-5.27 (2.58)	5.53 (2.01)	-20.84 (3.77)	181.46 (25.65)	17.08 (3.78)	174.25 (46.91)
	Casual	12.14 (1.39)	-5.14 (2.28)	6.01 (2.05)	-19.16 (3.41)	151.76 (20.46)	15.46 (3.24)	176.14 (38.36)

Table 21. Summary of repeated measures ANOVAs performed on measures for vocal tract opening movement of the lower lip (LL).

LL								
		F values						
		Position at Movement Onset		Position at Movement Offset			Movement characteristics	
Factor	*df*	Horizontal	Vertical	Horizontal	Vertical	Duration	Distance	Peak Speed
Clarity	1	0.18	1.79	9.08	26.21*	115.54*	29.12*	0.50
Gender	1	24.87*	2.86	15.31*	1.46	3.71	0.01	0.04
Clarity-by-Gender	1	2.83	2.16	1.18	0.68	5.66	3.35	0.93
Error	44							

* $p < 0.0008$

Analysis of the kinematic parameters revealed that for all articulators, movements were significantly longer and larger in the clear condition as compared to the casual condition. Within the clear condition, the onset of movement began at a significantly lower tongue and jaw position and ended at a significantly higher tongue position. T3 was the only tongue blade fleshpoints that exhibited a significantly faster movement in the clear condition. The UL and LL were found to be lower at onset and lower at offset respectively.

Correlation between Auditory Perceptual Measures and Articulatory/Acoustic Measures

Pearson correlations were performed on measures of auditory perception of clarity with measures to of acoustic segment durations, F1 and F2 transitions and articulator movement transition and durations.

Auditory Perceptual Measures and Acoustic Measures:

Table 22 shows the Pearson correlation coefficients associating auditory perceptual ratings and clear-casual differences in acoustic segment durations. A positive correlation between listener panel ratings of clarity and increased duration measures of the phonemic segments was found with all segments and overall word duration.

Table 23 shows the Pearson correlation coefficients associating auditory perceptual ratings and clear-casual differences in diphthong-related formant transitions. A significant correlation between panel ratings of clarity and measurements of F1 and F2 transitions at offset position, transition range and transition duration was found.

Specifically, as F1 and F2 increased their difference (F1 offset lower and F2 offset higher) the auditory perceptual ratings of clarity increased.

Figure 10 shows a scatter plot of the correlation each of the 49 speakers word duration increase in the clear condition compared to the casual condition and the mean auditory perceptual rating for each. The correlation is 0.9 suggesting a strong positive association.

Auditory Perceptual Measures and Articulatory Movement data:

Table 24 shows the Pearson correlation coefficients associating auditory perceptual ratings and clear-casual differences in selected kinematic results. A positive correlation was found between all movement measures of duration and panel ratings of clarity. MI and LL measurements were found to be significant for distance as well as duration. Over all, articulator start and end points are more extreme and excursions are longer and larger.

Table 22. Pearson correlation coefficients (r) associating auditory perceptual ratings and clear-casual differences in acoustic segment durations.

	k	ʌ	m	b	ɑɪ	n	total
r	0.35	0.48*	0.63*	0.42	0.86*	0.48*	0.90*

*$p < .005$

Table 23. Pearson correlation coefficients (r) associating auditory perceptual ratings and clear-casual differences in diphthong-related formant transitions.

Formant	Onset frequency	Offset frequency	Transition Frequency Range	Transition duration	Transition rate
F1	0.03	-0.39*	-0.46*	0.74*	0.14
F2	0.22	0.64*	0.47*	0.65*	-0.25

*$p < .005$

Table 24. Pearson correlation coefficients (r) associating auditory perceptual ratings and clear-casual differences in selected kinematic results.

	Position at Movement Onset		Position at Movement Offset		Movement characteristics		
	Horizontal	Vertical	Horizontal	Vertical	Duration	Distance	Peak Speed
T2	0.02	-0.13	0.36	0.36	0.58*	0.38	-0.04
T3	0.05	0.04	0.30	0.18	0.44*	0.23	-0.10
MI	0.01	-0.14	0.05	0.17	0.64*	0.51*	0.19
LL	0.02	0.34	0.01	-0.37	0.54*	0.44*	0.32

*$p < .005$

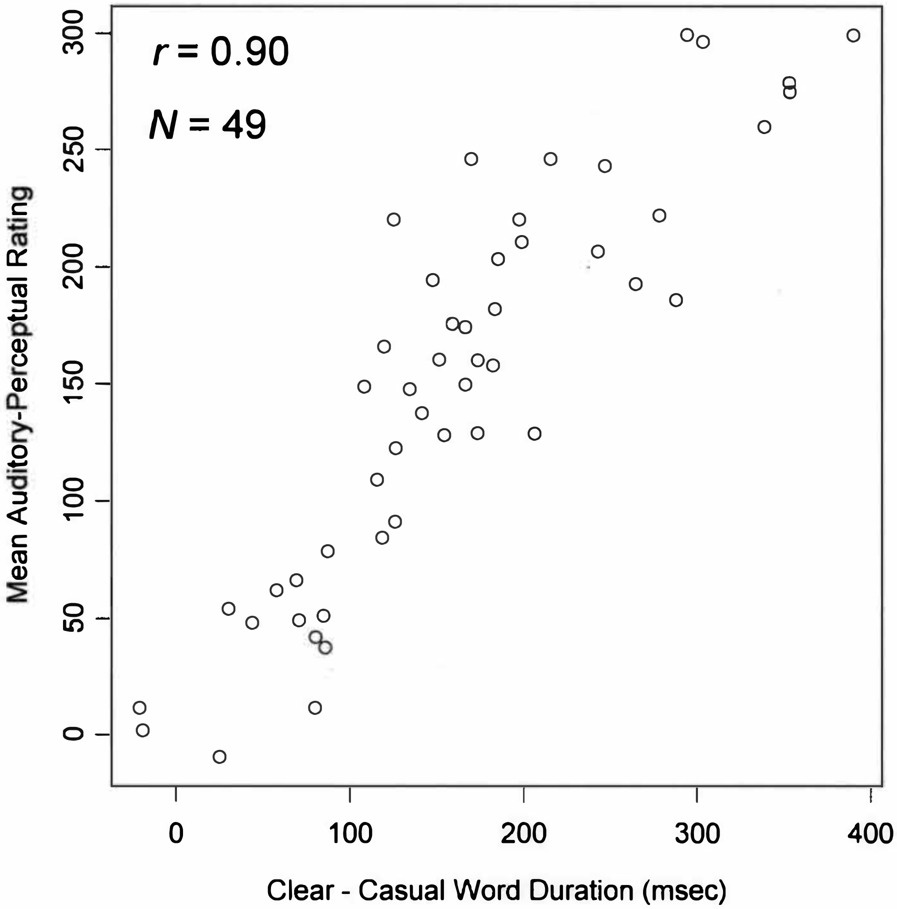

Figure 10. Scatter plot showing the relationship between mean auditory-perceptual ratings of clarity and increase in word duration during clear speech. Each data point represents a speaker.

CHAPTER V

DISCUSSION

Acoustic and Kinematic and Auditory Perceptual Measures of Clear Speech

Previous literature suggests there are a number of measurable differences in the acoustic and kinematic features of speech produced under clear and casual conditions. However, current understanding of clear speech production is limited. The present study was designed to characterize clear speech at the acoustic, articulatory and auditory perceptual levels of analysis. This multilevel approach prompted the hypothesis that auditory perceptual scaling of clarity would be associated with kinematic indicators of physical effort.

A number of factors make this study unique as compared to that of previous work on clear speech. Previous studies have often been limited to a single analysis domain (e.g., acoustic or articulatory kinematic). The present study has concurrent evaluation at perceptual, acoustic, and kinematic levels that has the potential to provide a clearer view of the possible clarity-based adjustment speakers utilize. Another distinctive characteristic of this study is the listener evaluation measures of speakers' clarity. These judgments of relative clarity allowed for correlation analysis between perceptual judgments of clarity change and changes in acoustic and kinematic measures. Finally, this study used a relatively larger sample size than has been previously used in studies on this topic.

Acoustic Features of Clear Speech

Measurements of acoustic segment duration and first and second formant transitions provided significant distinction in the clear condition. The most prominent acoustic finding in the clear condition was the significantly longer word and segment durations. Vowels were more affected by the clarity condition than stop consonants. This suggests that vowels are more susceptible to elongation as compared with stop consonants, a finding which is consistent with previous studies that reported increased duration of individual speech sounds (Picheny, et al., 1985, 1986; Moon & Lindblom, 1994; Perkell, et al., 2002; Bradlow, 2001; Matthies, et al., 2001) in the clear condition. Correlation analysis of auditory perceptual evaluation of clarity and increased duration of acoustic segments and total word duration is significantly high. This correlation will be further discussed later in this chapter.

For the F1 and F2 transition through the diphthong /ɑɪ/ in the clear condition, the relevant features are the position at onset and offset and the duration and range of transition. F1 transition starts at a higher position, and while not significant there was also a general trend for the F1 offset to be lower in the clear condition. Similarly, F2's onset in the clear condition does not meet our criteria for statistical significance, however the number is fairly high and does follow F1 in having a more low and back posture in the initial position. F2 has a significantly higher offset. These positions suggest that at the start of the diphthong speakers are producing a more "/ɑ/-like" position and at the end of the diphthong the position is more /i/-like. An increased duration and range of formant transition in the clear condition is consistent with both F1 and F2. Therefore, during clear speech speakers make larger and longer formant transitions starting and ending at more exaggerated formant frequency positions. Additionally, F2 transition rate was found to be significantly slower in the clear condition, a finding not consistent with Moon and Lindblom (1994) who found clear speech was characterized by increased intensity, longer vowel durations, larger formant displacements and more rapid formant transitions than the citation speech.

This study chose to look at F_0 values given findings of previous literature of trends to be a wider range and generally higher F_0 in clear speech (Picheny, et al., 1985, 1986). Speech clarity did not appear to be influence the fundamental frequency measures. As would be expected, there is significantly higher F_0 measurement in female speakers in the mean, minimum and standard deviation than for males. It is surprising that the F_0 results do not significantly change in the clear condition, however one explanation may be the relatively short segment used for analysis for this study. That is, analysis of a whole sentence produced in the clear condition may show F_0 changes.

Another surprising finding in this study was a lack of clarity-based difference for mean and max sound pressure level (SPL). No systematic changes in SPL as a function of clarity were found. An increase in loudness is expected during clear speech as found in previous literature (Picheny, et al., 1985, 1986). This does not necessarily conclude that SPL variations do not exist. This study was performed using a publicly available database. It is not clear from database documentation how well the experimenters controlled the consistency of the microphone to mouth distance. Such a variation could minimize loudness effects.

Articulatory Features of Clear Speech

Analysis of the kinematic parameters for all articulators revealed that movement excursions were significantly longer and larger as the tongue moved through the diphthong gesture in the clear condition as compared to the casual condition. More specifically, in the clear condition, the onset of movement began at a significantly lower and back tongue and lower jaw position and ended at a significantly higher and forward tongue position. Not surprisingly, these measurements follow the formant transition of increased duration and distance changes in the clear condition.

Analysis of UL measurements indicates little significant change in the clear condition. UL was significant in the clear condition, showing it to be lower at onset. LL measures in the clear condition are significantly lower in the vertical position at offset, with a greater duration and distance traveled. Lip measurements do show a gender difference, as males have a significantly more forward position of the UL and LL at onset and offset, however this does not have a significant effect on clarity. It may be that gender differences could be due to difference in the overall size of vocal tract structures.

Peak speed was not found to be significantly affected in clear speech production. Only one of the tongue blade fleshpoints, T3, exhibited a significantly faster movement in the clear condition. This finding is contrary to an initial hypothesis that peak speed would be affected by the clarity condition. One explanation is that T3 is the only pellet that is capturing speed changes given its location on the tongue, and that an increase in speed is in fact employed during clear speech production. Another explanation is that peak speed is not the employed by speakers and that they produce more effort with larger movements.

Auditory Perceptual Findings

Auditory perceptual analysis generally supports the hypothesis that speakers produced clearer speech in the clear condition. However, the ratings for individual speakers varied from perceptually indistinguishable differences between the two speech conditions to significant consistently perceived differences. This suggests that speakers do not make perceptually uniform clarity adjustments when given the same instruction for producing clear speech.

Correlation between Auditory Perceptual Measures and Articulatory/Acoustic Measures

Correlations between perceptual judgments of clarity and temporal (i.e., acoustic and kinematic durations) measures were generally positive and statistically significant. Spatial/spectral measures (e.g., movement extent, spatial position at onset and offset) were much more variable in their association with clarity ratings. A positive correlation between listener panel ratings of clarity and increased duration measures of the acoustic segments was found with all segments and overall word duration. Likewise, a significant correlation was found between panel ratings of clarity and measurements of F1 and F2 transitions. Specifically, as F1 and F2 increased their range (F1 offset lower and F2 offset higher) the auditory perceptual ratings of clarity increased.

The association between auditory perceptual ratings and clear-casual differences in the kinematic results are positive for all the movement measures of duration. This may suggest that listeners are tuning into duration changes more than spatial/spectral measure changes when identifying clear speech.

Kinematic Indicators of Effort

One of the hypotheses of this study was that auditory perceptual scaling of clarity (perceptual salience) would be associated with kinematic indicators of physical effort. However this study only found one articulator (T3) that had an increase in speed. In general this study found peak speed to not be significant or correlated with clear speech production. This does not necessarily mean that speakers are not employing greater effort. In fact there is good reason to speculate that speakers may in fact be increasing effort given the significant extent of excursion through increase in duration and distance. The question of effort may be better realized through adopting the theory of effort as indicated by increases in other measures such as a combination of distance and duration as is seen in this study. For example, to travel from point A to point B a person may have a fixed distance to travel but increase their speed, or they may keep their speed the same but increase the distance traveled. In either case more effort is required but is expressed in different ways depending on the situation. The later example may be a more likely scenario for clear speech. As we see in this study, clear speech allows for more extreme articulation posture, thus an increase in distance traveled with no change in speed. Using distance and duration measurements to indicate effort we can conclude that an increase in effort is present in clear speech.

Limitation of Present Study and Suggestions for Future Directions

Experimental Limitations

Design of this study was limited by the pre existing data set of speaker subjects. This did not allow for experimenter control over instruction given to speakers to elicit clear and casual speech. Additionally, the experimenter could not control for microphone placement to ensure consistent distance between the speaker and the microphone during recordings.

Auditory perceptual data was collected using two presentations of the same word, both intelligible, that required a clarity scaling by the listener. It may be interesting for a future study to present the stimulus embedded in noise to create an intelligibility distinction to define clarity. In this suggested method listeners may be forced to rely on the characteristics that matter most for clear versus unclear speech as opposed to casual speech.

Another alternative for characterization of clarity may be to use synthetic speech in order gain more control over specific acoustical measurements. Presentations to listeners could then be altered in systematic ways to better characterize what a listener may be tuning into when perceiving clarity.

Conclusion

In summary, the principle findings of this study show that when producing clear speech speakers increased durations of sound segments, make larger, longer formant transitions which start and end at more exaggerated formant frequency positions, articulators, typically have start and end movement in more extreme positions that increase movement distances and durations, and do not appear to systematically vary speak speed with speech clarity. Speech clarity is more strongly associated with duration measures.

The original hypothesis that auditory perceptual scaling of clarity would be associated with movement indicators of physical effort was not supported. However as previously discussed, this hypothesis may be better tested in future work using a broader or alternative definition of effort.

Appendix A Listener Consent Form

WESTERN MICHIGAN UNIVERSITY
H. S. I. R. B.
Approved for use for one year from this date:

OCT 0 7 2003

x
HSIRB Chair

Consent Form: Listening Experiment

Western Michigan University
Department of Speech Pathology and Audiology

Principal Investigator: James M. Hillenbrand

I have been invited to participate in a research project entitled "Acoustic Correlates of Phonetic Quality." The purpose of this research is to gain a better understanding of how the ear and brain process and recognize speech signals. If I choose to participate, I will be listening to samples of either naturally produced or computer-generated speech. My job will be to make certain judgments about the speech. For example, I might be asked to identify the word or speech sound that was spoken, or I might be asked to judge whether two sounds are the same or different. The precise task will be explained to me before the listening session begins. I will also be asked to participate in a screening procedure consisting of: (1) a brief hearing test, (2) a short interview, and (optionally) an assessment of dialect. The hearing screening consists of listening to three soft tones presented over headphones to each ear. The interview consists of answering two questions about my language background and any history of speech disorders. The listening experiment and screening task together will last approximately 30-60 minutes.

There are no known risks associated with these procedures. However, as in all research, there may be unforeseen risks to the participant. If an accidental injury occurs, appropriate emergency measures will be taken; however, no compensation or additional treatment will be made available to me except as otherwise stated in this consent form.

All the information that is gathered as part of this listening session is confidential. That means that my name will not appear on any papers describing this research. My responses on the listening task will be stored in a computer file under my subject number only There will be no way to associate this subject number with my name.

WESTERN MICHIGAN UNIVERSITY
H. S. I. R. B.
Approved for use for one year from this date:

OCT 0 7 2003

x
HSIRB Chair

I will not directly benefit from this research. However, the knowledge that is gained could be used to aid in the treatment of speech or voice disorders, and to improve methods of voice synthesis. If I should consent to participate in this study, I may withdraw my consent at any time without effect on grades or my relationship with Western Michigan University. If I have any questions or concerns about this study, I may contact James Hillenbrand at 387-8066. I may also contact the Chair of the Human Subjects Institutional Review Board at 387-8293 or the Vice President for Research at 387-8298 with any concerns that I have.

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner of both pages. Subjects should not sign this document if the corner does not show a stamped date and signature.

Signature of Research Subject

02/24/04
Date

Signature of Experimenter or Research Assistant

2/24/04
Date

Appendix B HSIRB Approval Letter

WESTERN MICHIGAN UNIVERSITY

Centennial
1903-2003 Celebration

Human Subjects Institutional Review Board

Date: October 7, 2003

To: James Hillenbrand, Principal Investigator

Cc: David Ross, Grants and Contracts
Proposal # 5 RO1 DC01661-11

From: Mary Lagerwey, Chair

Re: Final Extension of Approval, HSIRB Project Number 01-11-06

This letter will serve as confirmation that an extension to your research project entitled "Acoustic Correlates of Phonetic Perception" has been granted by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now continue to implement the research as described in the original application.

Please note that you may **only** conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. **You must submit a new protocol if the project extends beyond the termination date noted below.** In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the continued pursuit of your research goals.

Approval Termination: October 7, 2004

Walwood Hall, Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276

Human Subjects Institutional Review Board

Kalamazoo, Michigan 49008-5456
616 387-8293

WESTERN MICHIGAN UNIVERSITY

Date: November 26, 2001

To: James Hillenbrand, Principal Investigator

From: Mary Lagerwey, Chair Mary Lagerwey

Re: HSIRB Project Number: 01-11-06

This letter will serve as confirmation that your research project entitled "Acoustic Correlates of Phonetic Perception" has been **approved** under the **expedited** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may **only** conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: November 26, 2002

BIBLIOGRAPHY

Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. *Laboratory Phonology 7.* New York: Mouton de Gruyter.

Bradlow, A. R. and Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112(1),* 272-284.

Kent, R. D. (1997). *The Speech Sciences.* San Diego, CA: Singular Publishing Group, Inc.

Kent, R. D. and Read, C. (2002). *Acoustic Analysis of Speech.* Albany NY: Delmar.

Kent, R. D., Kent, J. F., Weismer, G., Sufit, R. L., Brooks and Rosenbek, J. C. (1989). Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects. *Clinical Linguistics and Phonetics, 3(4),* 347-358.

Lieberman, P. and Blumstein, S. E. (1998). *Speech Physiology, Speech Perception, and Acoustic Phonetics.* Irthlingborough, Northants: Woolnough Bookbinders Ltd.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-439). Netherlands: Kluwer Academic Publishers.

Matthies, M., Perrier, P., Perkell, J. S. and Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research, 44*, 340-353.

Moon, S-J. and Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96,* 40-55.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics, 46,* 135-147.

Perkell, J. S. (1999). Articulatory Processes. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 333-370). Malden, MA: Blackwell Publishers Ltd.

Perkell, J. S., Zandipour, M., Matthies, M. L. and Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America, 112(4),* 1627-1641.

Picheny, M. A., Durlach, N. I. and Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96-103.

Picheny, M. A., Durlach, N. I. and Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434-446.

Weismer, G., Kent, R. D., Hodge, M. and Martin, R. (1988). The acoustic signature for intelligibility test words. *Journal of the Acoustical Society of America,* 84(4), 1281-1291.

Westbury, J. R. (1994). *X-ray microbeam speech production database user's handbook.* Madison: University of Wisconsin at Madison, Waisman Center.