5-1-2002

Differences in Acoustic Measures of Vowels in Ventriloquial and Normal Speech

Corrie S. M. Wiesner

Recommended Citation

Wiesner, Corrie S. M., "Differences in Acoustic Measures of Vowels in Ventriloquial and Normal Speech" (2002). *Theses and Dissertations*. 3230.
https://commons.und.edu/theses/3230

DIFFERENCES IN ACOUSTIC MEASURES OF VOWELS IN VENTRILOQUIAL AND NORMAL SPEECH

by

Corrie S. M. Wiesner
Bachelor of Science, Minot State University, 2000

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science

Grand Forks, North Dakota
May
2002

This thesis, submitted by Corrie M. Wiesner in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

(Chairperson)

John P. Madden

This thesis meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Joseph N. Benoit

Dean of the Graduate School

April 29, 2002

Date

PERMISSION

Title Differences in Acoustic Measures of Vowels in Ventriloquial and Normal Speech

Department Communication Sciences and Disorders

Degree Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University to North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University to North Dakota in any scholarly use which may be made of any material in my thesis.

Signature Corrie Wiesner

Date April 15, 2002

TABLE OF CONTENTS

CHAPTER

LIST OF ILLUSTRATIONS

LIST OF TABLES

Table Page

ACKNOWLEDGMENTS

The editor expresses sincere appreciation to Graduate School staff members Dr. John Madden and Dr. Kanishka Marasinghe, who graciously proofread the manuscript, offered suggestions, and participated in this study's development. A special thanks to Dr. Manish Rami who willingly dedicated time, energy, and ideas to benefit the development of this manuscript.

Additional thanks to my father who gladly and willingly agreed to be part of this study and to my loving husband who kindly offered his support and constant encouragement.

ABSTRACT

This study investigated and compared the acoustic properties of vowels in ventriloquial and normal speech. Voice recordings of a 51 year-old male participant producing 10 words containing target vowels, three times each were made in both normal and ventriloquial speech. Standard acoustic measures for frequency were gathered using Tiger Electronics Inc., Dr. Speech Science, Ver.2.0. Fundamental frequency, first, and second formant frequencies were analyzed as compared between the two types of speech. Although the results revealed no statistically significant differences in first and second formant frequencies, slight variations do exist. Statistically significant differences were found for two words in fundamental frequency. The results would seem to suggest that as long as vowel production is within a "range" of variability, vowels may be accurately perceived.

CHAPTER I

VENTRILOQUISM

The word ventriloquism is derived from two Latin words, ‘ventor’, meaning the belly, and ‘liquor’ which means to speak (Bergen, 1983). Contrary to the meaning of these words, ventriloquial speech does not come from the belly but instead comes from the larynx, where all speech originates. Ritchard & Maloney (1987) define ventriloquism as “the art of speaking without moving your lips and jaw, in which additional tongue movements replace all visible lip and jaw movements. Using this system, the ventriloquist creates sounds and voices that, by means of acting and illusion, seem to come from a different source” (p. 4).

Visual Cues Used in Ventriloquism

Good ventriloquism depends on two major factors, visual cues and more importantly, auditory cues. The skill of making a voice sound as if it is coming from a puppet is called the “ventriloquism” effect and relies upon the presence of visual cues. According to Jack & Thurlow (1973), “the ‘ventriloquism’ effect refers to the perception of speech sounds as coming from the same direction as the visually observed speaker (even though they are not actually coming from the same direction)” (p. 967). This effect has been a focus of study for many years and findings indicate that the strength of the ventriloquism effect is heavily dependant upon the movement of a puppet’s mouth. This movement helps to orient spatial attention. (Jack & Thurlow, 1973, Warren, Welch, & McCarthy, 1982, and Bertelson, Vroomen, DE Gelder, & Driver, 2000).

Auditory Cues Used in Ventriloquism

In addition to the visual stimulus, auditory cues are essential to "capturing" the audience. Many acoustic and articulation changes need to be made so that the voice is perceived to come from the puppet rather than the ventriloquist. The ventriloquist not only has to act as if the puppet is talking, but he/she must also keep his/her mouth from moving. The difficulty in developing this skill arises because there are a number of English sounds that are produced with mouth and lip movement. There are ten English phonemes that are known as "problem letters" to the ventriloquist. These letters are the three bilabials /p, b/ and /m/, the four fricatives /f, v, θ, ð/, the liquid /w/, and the two lip rounding vowels /o/ and /u/ (Ritchard & Moloney, 1987). Before attempting to understand precisely what occurs during the ventriloquial production of these phonemes, it is important to understand how they are produced in normal speech.

CHAPTER II

PHONATION

The understanding of a speech signal is based on a knowledge of how sound is transmitted. Sound is produced when a disturbance causes changes in pressure in a gas, liquid, or solid medium. Air is the primary medium through which sound is transmitted and is composed of molecules that have a tendency to equalize in the atmosphere, meaning they spread themselves in equal distances in relation to each another. The movement of the vocal folds set into motion by air coming from the lungs is the disturbance that causes a change in pressure (Ferrand, 2001).

Air is primarily made up of oxygen, nitrogen, and hydrogen. These molecules possess inherent energy that causes them to move around at extremely high speeds in seemingly random patterns. This is known as the Brownian motion (Ferrand, 2001). Air possesses the properties of mass, elasticity, and inertia (Fucci & Lass, 1999), which allow the air particles to vibrate.

Mass, Elasticity, and Inertia

Mass refers to any form of matter that is capable of vibrating. The property of elasticity is a force that restores and is described by Ferrand (2001) as "the property of an object to be able to spring back to its original size, form, location, and shape after being stretched, displaced, or deformed" (p.15). Inertia is used to describe the fact that "a body in motion will remain in motion, while a body at rest will remain at rest (unless acted upon by an external force)" (Fucci & Lass, 1999, p. 72). For example, when pushed on a swing, a person (who has mass) remains in the forward motion (the property of inertia) until the direction is reversed and the person returns to the original position (the property of elasticity). Inertia once again becomes a factor when the person swings past the point of origin.

This same process is true for the movement of air particles that are displaced in the transmission of sound. Once the particles of air are set into motion, inertia takes over and they continue to move until they encounter other molecules. They then set these other molecules into motion and the property of elasticity returns them to their original position, where inertia once again takes over (Borden, Harris, & Raphael, 1994). Therefore, sound is transmitted from one location to another because each molecule in motion disturbs neighboring molecules that move the sound further and further away from the original source. Ambient pressure is the term used to refer to the constant pressure in the atmosphere. For a sound to be generated, this ambient pressure must be disturbed in a systematic manner, such as the vibration of the vocal folds. When molecules are displaced by vocal fold vibration and approach one another to collide, an area of positive pressure results. This is called compression. Rarefaction, on the other hand, is used to describe the area of low pressure caused when molecules of air move away from each other following a collision. Sound is then heard by the listener when the tympanic membrane is moved inward by compression, and outward by rarefaction (Ferrand, 2001).

CHAPTER III

ACOUSTIC PROPERTIES OF SOUND

Waveforms

The movement of air particles during the transmission of sound, can be graphically represented in the form of a wave called a waveform (Ferrand, 2001). Plotting the characteristic of time on the horizontal axis and the characteristic of amplitude on the vertical axis creates a waveform. A wave is described by Ferrand (2001) as, "a disturbance that moves through a medium" (p.15). It is characterized by changes in a medium caused by the movement of individual particles of that medium.

A waveform also demonstrates the linear and temporal concepts of wavelengths, cycles and periods. A cycle is the term used to describe the vibratory movement of a particle from its resting position to its point of maximum displacement in one direction, then back to its resting position, and then to its point of maximum displacement in the opposite direction. The time it takes for one cycle to be completed is generally expressed in seconds and is termed the period of the cycle. The distance between two points of a duplicate phase in two contiguous cycles of a wave is the wavelength (Fucci & Lass, 1999).

Intensity

Intensity, measured in decibels (dB) is defined by amplitude. Amplitude is referred to as the "maximum displacement of the particles of a medium" (Fucci & Lass, 1999, p. 78). Amplitude is used to show the intensity or the energy of the sound. It is generally measured from the baseline to the point of maximum displacement of the particles. Intensity is a measure of the energy found in the sound. Kaplan, Gladstone, & Lloyd (1993) further define intensity as the "amount of sound energy; the greater the sound energy, the higher the intensity. Intensity, in a general way, corresponds to perceived loudness" (p. 391). Therefore, the further the particles of air are displaced from their place of rest, the greater the amplitude. The greater the amplitude, the higher the intensity. The higher the intensity, the louder the sound is perceived to be.

Frequency

Frequency is another acoustic measure used to characterize sound. It is related to the temporal measurement of period. Frequency is "the number of vibratory cycles per second" (Borden et al., 1994), and its unit of measurement is Hertz (Hz) (Ferrand, 2001). A wave in which every cycle takes the same amount of time is labeled as a periodic wave and is characteristic of a pure tone (Ferrand, 2001). A pure tone is defined as, "a sound wave that has only one frequency" (Ferrand, 2001, p. 24). A pure tone is seldom heard in everyday life. Sounds are more often complex tones rather than pure tones. Waves that are composed of more than one frequency are known as complex waves. Complex waves are due to the combination of and interference between different frequencies. Speech sounds comprise complex tones consisting of a number of frequencies (Borden et al., 1994).

Fundamental Frequency and Harmonics

In speech, frequency is perceived as the pitch of an individual's voice (Ryalls & Behren, 2000). For example, it is generally found that a male's voice is lower in pitch than that of a female. Ferrand (2001) offers this explanation of pitch: "the higher the frequency (the more cycles per second), the shorter the duration is the period and the shorter the wavelength. The lower the frequency (the fewer cycles per second), the longer in duration is the period and the longer is the wavelength" (p. 21). The most basic measure of frequency of voice is that of fundamental frequency (F_0), or the lowest rate of vocal fold vibration (Ryalls & Behrens, 2000, Martin, 1997). Multiple frequencies actually are produced by the vocal folds from this fundamental frequency. These multiple frequencies are called harmonics, and they bear a direct relationship to the fundamental frequency. They can be predicted based on a mathematical relationship to the fundamental frequency. Each harmonic frequency is a whole number multiple of the fundamental frequency, or the first harmonic (Borden et al., 1994, Ryalls & Behrens, 2000). For example, if the first harmonic is 100 Hz (100 x 1), the second harmonic would be 200 Hz (100 x 2), the third harmonic, 300 Hz (100 x 3), and so on.

Formant Frequencies

The harmonics (or the harmonic spectrum) are "fine tuned" or filtered by the vocal tract to produce peaks called the *formant frequencies* (Ryalls & Behrens, 2000). These formant frequencies are the result of the filtering of sound (fundamental frequency and harmonics) through the resonance of the vocal tract. Shipley & McAfee (1998) describe resonance as "the vibration of one or more structures related to the source of the sound; vibration above or below the sound source (the larynx for speech)" (p. 461). A sound is created elsewhere and the resonator vibrates in conjunction with it if the sound is at or near the resonant frequencies of the resonator" (Borden et al., 1994).

Ferrand (2001) outlines three characteristics of the vocal tract resonator. First, it is a quarter-wave resonator. The vocal tract is thought of as such because it is open at one end (the lips) and closed at the other (the glottis). Second, the vocal tract can also be thought of as several air pockets linked and joined together. Each air container acts as a filter to transmit those frequencies within the bandwidth of its resonances and attenuate frequencies outside its bandwidth. Each container has its own resonating frequency. These resonate frequencies are what are known as the formant frequencies. The third characteristic of the vocal tract resonator is that it is a variable resonator. Its resonant frequencies are dependent upon the shape of its cavities. Thus, the formants of the vocal tract change whenever the articulators are moved to produce a sound. The type and amount of resonance is dependent upon the shape and configuration of the vocal tract.

The formant frequencies important in the analysis of speech are labeled F_1, F_2, and F_3 (Ryalls & Behrens, 2000). F_1 is the result of the size and shape of the back cavity of the vocal tract, that is, from the vibrating vocal folds to the point of first constriction of the articulators. F_2 is dependent upon the size and shape of the front cavity, or from the point of first constriction to the teeth. The size and shape of the cavity created between the teeth and the lips is associated with F_3.

The size and shape of the resonating cavity directly influences the frequency of sound produced (Ryalls & Behrens, 2000). To illustrate this concept, imagine blowing air across the top of a pop bottle. When the bottle is full of liquid, the sound made when air is blown across the top is high in frequency. This is due to the small chamber in which the air resonates. However, a low frequency sound is produced when there is very little liquid in the bottle. The area in which the air resonates is much larger. The same holds true for speech sounds. The larger the resonating cavity, the lower the sound frequency. The smaller the resonating cavity, the higher the sound frequency.

Measuring Fundamental Frequency and Formant Frequencies

These acoustic attributes of sound can all be measured with the aid of a spectrographic analysis. Spectrographic analysis of speech is "a dynamic analysis that reveals spectral features in a nearly continuous fashion" (Kent, 1997, p. 344). These spectral features include measures of frequency, amplitude, and duration. "Frequency is displayed on the vertical axis, time is represented on the horizontal axis, and intensity of acoustic energy is represented by the darkness of the trace on the screen" (Ferrand, 2001, p.200). From a spectrographic analysis, fundamental frequency, formant frequencies can be examined.

To obtain more exact estimates of the formant frequencies of a vowel, amplitude spectra can be used (Borden et al., 1994). An amplitude spectrum displays the amplitude of the signal harmonics as a function of frequency. As sound, originating from the vibration of the vocal folds, is filtered through the vocal cavities, some frequencies are intensified and others are attenuated. The frequencies that become intensified depend upon the size and shape of the resonating cavities. These frequencies are the formant frequencies and are represented by peaks on an amplitude spectra.

CHAPTER IV

VOWEL PRODUCTION AND ANALYSIS

When vowels are analyzed spectrographically, several acoustical characteristics are revealed. The formant frequencies are greatly dependent upon the manner in which the articulators, most importantly the tongue, are configured. Shriberg & Kent (1995) state that vowels produced when the tongue is close to the roof of the mouth are called high vowels. Contrarily, vowels produced when the tongue is depressed in the mouth are the low vowels. Vowels for which the tongue is in intermediate positions are described using the terms mid-high, mid, or mid-low. For example, the /i/ vowel can be described as a high front vowel because to produce it, the front part of the tongue is drawn up to the top of the mouth. The position of the tongue is responsible for the spectrographic features for each vowel. Kent & Read (1992) report that a general rule of thumb is that F1 varies according to the tongue height and that F2 varies depending on tongue advancement (the anterior-posterior orientation of the tongue). Based on spectrographic analysis, it is found that the low vowels have a high F1 frequency and high vowels have a low F1 frequency.

The ventriloquial "problem vowels" of /u/ and /o/, are "rounded" vowels. They are made by rounding the lips in order to elongate the oral cavity. They are also called "back" vowels, meaning that they are "produced in the lowest position, with the tongue depressed in the mouth" (Shriberg & Kent, 1995, p. 26). According to Kent (1997) and Kent & Read (1992), the rounding of the lips lengthens the vocal tract causing the formant frequencies to become lower, and more specifically, "the back vowels have a small F1-F2 separation, a large F2-F3 separation, and a variable F1-F0 difference reflecting tongue height" (p. 350). The term separation can be described as the difference between the values of the two formant frequencies (Kent & Read, 1992).

In ventriloquial speech the lips must not move in order to maintain the illusion that the voice is coming from someplace other than from the individual producing the voice. Producing the back vowels /u/ and /o/ in normal speech requires the rounding and elongation of the lips. However, in ventriloquial speech, the lips must be kept relaxed and still. Therefore, the formants of these vowels could be significantly different than vowels produced in normal speech. The rounded vowels, /u/ and /o/ would not have the elongated vocal tract thus causing the formant frequencies to be higher.

The vowels /a, ɑ, ɛ, æ, ʊ, e, i, ə/ would require a comparatively smaller, more constricted oral cavity in ventriloquial speech as opposed to normal speech (Ritchard & Moloney, 1987). This effect would make the formant frequencies higher (Ryalls & Behrens, 2000). In addition, ventriloquial speech demands a tense vocal tract, thus potentially heightening the vocal intensity and pitch (Kent, 1997).

In order for an audience to perceive that a puppet is actually speaking, the issue of sound recognition is involved. The way in which speech sounds are perceived greatly relies on the "formant specification of vowels" (Kent & Read, 1992, p. 92). Syrdal & Gopal (1986) studied the perception of vowels according to their auditory representation. They report that regardless of speaker differences, individuals possess the inherent ability to normalize sounds they hear. No matter what the stimulus, as long as it bears resemblance to a known sound, the listener can make strong conclusions as to the nature of that sound. In addition to the acoustic signal, Rosch (1975), reported that native vowels may be organized based on prototypes, meaning that sounds are learned by assimilating nearby members of the same phonetic category.

The purpose of this study was to examine the acoustic differences in ventriloquial speech as compared to "normal" speech production. The variations in fundamental frequency and formant frequencies were examined and compared. It was expected that because ventriloquial speech is created and produced differently as compared to normal speech (through variations made to the vocal tract), the standard acoustical measures would be statistically different. The findings reported in this study may have significant implications for those individuals who need to acquire compensatory articulation skills due to anatomic and/or physiologic difficulties.

CHAPTER V

METHOD

Ventriloquist

The ventriloquist was a 51 year-old adult male with 15 years of experience in performing as a clown and ventriloquist for birthday parties, seminars, and skits. He has no history of speech, language, or neurological disorders. His hearing was screened according to ASHA's guidelines for hearing screening and found to be normal (American Speech-Language-Hearing Association Audiologic Assessment Panel 1996, 1997).

Materials

A laptop computer (Dell, model LC600), a lapel microphone (Lavalier microphone audiotechnica, model ATR35s), and a stationary color video camera (RCA Camcorder, model CC415) were used to record the participant's speech productions. A list of words used to elicit the vowels and diphthongs of English (heat, hid, head, hat, hot, hood, hoot, hut, hurt, hard) was used to make the audio recordings. Dr. Speech Science, Ver. 2.0 (Tiger Electronics Inc.) program was used to conduct the acoustic analysis of the speech samples recorded.

Procedure

The ventriloquist was seated in a comfortable chair in a quiet room setting. A microphone was attached to a boom placed approximately 30 cm from the ventriloquist's lips with an orientation of 0° azimuth and -30° altitude. The ventriloquist was asked to produce a series of words that contained the target vowels. Each word was recorded three times. The ventriloquist produced the words first in normal speech, and then in ventriloquial speech. These speech samples were digitally recorded as well as audio-video taped. Standard acoustical measures for frequency (fundamental frequency and formant frequencies) were obtained.

CHAPTER VI

RESULTS

Three utterances of each word in the two speaking conditions formed the 60 tokens. The fundamental frequency as well as the first and second formant frequencies were obtained for each utterance in the normal and ventriloquial speaking conditions. The means and standard deviations for the fundamental frequency and the first and second formant frequencies in both the normal and ventriloquial speaking conditions were calculated and are presented in Tables 1 and 2. A boxplot comparing the formant frequencies for each word in the normal speaking condition and another for the ventriloquial speaking condition is presented in Figures 1 and 2.

A *t* test for paired samples was used to identify statistical differences in each measure of individual words for normal and ventriloquial speech. Since multiple comparisons were conducted, the alpha level was adjusted to $p = 0.0025$ using the Bonferroni correction to avoid a family-wise error. A statistically significant change in fundamental frequency was seen between normal and ventriloquial speech ($p = 0.0025$). A similar paired comparison for the first and second formant frequencies was not found to be significant ($p > 0.0025$). The t values for all comparisons are presented in Table 3.

The formant frequencies obtained in this study were compared to the values for F1 and F2 obtained by Peterson & Barney (1950) and are represented in Figures 3 and 4. These data suggest vocal tract configuration during production of ventriloquial speech does not vary significantly from that of normal speech.

Table 1

Means and standard deviations (SD) for fundamental (F0), first formant (F1), and second formant (F2) frequencies for Normal Speech.

	F0 (Hz)		F1 (Hz)		F2 (Hz)	
Word	Mean	SD	Mean	SD	Mean	SD
Heat	108.14	24.43	230.22	20.22	2463.91	51.07
Hid	100.81	1.39	319.89	39.95	2071.40	114.93
Head	99.23	3.62	477.35	27.27	2008.76	147.14
Hat	95.17	10.23	688.24	68.08	1699.11	41.59
Hot	101.01	14.79	495.09	128.99	1245.88	161.95
Hood	103.34	5.64	373.23	29.09	1249.92	41.06
Hoot	113.85	6.66	275.06	10.09	1023.96	56.28
Hut	111.79	29.07	478.69	56.06	1400.15	94.96
Hurt	120.13	19.60	356.51	20.13	1326.84	88.19
Hard	96.85	15.49	389.91	81.42	1310.48	206.14

Table 2

Means and standard deviations (SD) for fundamental (F0), first formant (F1), and second formant (F2) frequencies for Ventriloquial Speech.

	F0 (Hz)		F1 (Hz)		F2 (Hz)	
Word	Mean	SD	Mean	SD	Mean	SD
Heat	257.08	7.29	253.97	38.52	2344.20	160.06
Hid	200.83	14.91	372.53	12.38	2107.77	121.59
Head	182.75	10.06	505.58	30.24	2176.71	113.02
Hat	177.43	19.77	407.63	162.92	1587.43	394.81
Hot	208.45	16.01	486.62	135.14	1256.27	114.00
Hood	220.50	9.83	416.33	17.94	1452.29	50.58
Hoot	247.68	1.43	334.19	3.02	1305.29	49.18
Hut	236.68	11.50	449.37	37.27	1324.87	74.21
Hurt	239.32	9.31	417.45	7.79	1461.17	125.36
Hard	222.81	22.49	467.39	20.48	1129.64	30.46

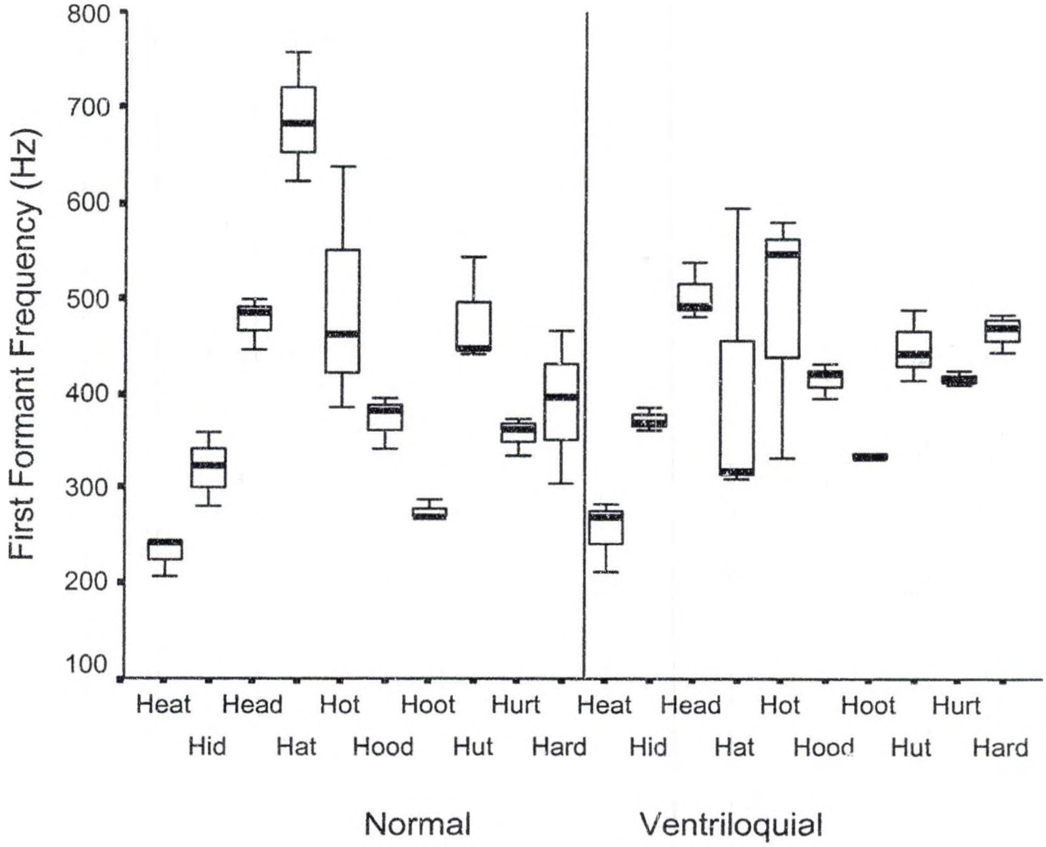

Figure 1. Boxplots of first formant frequencies for each word in Normal and Ventriloquial speech.

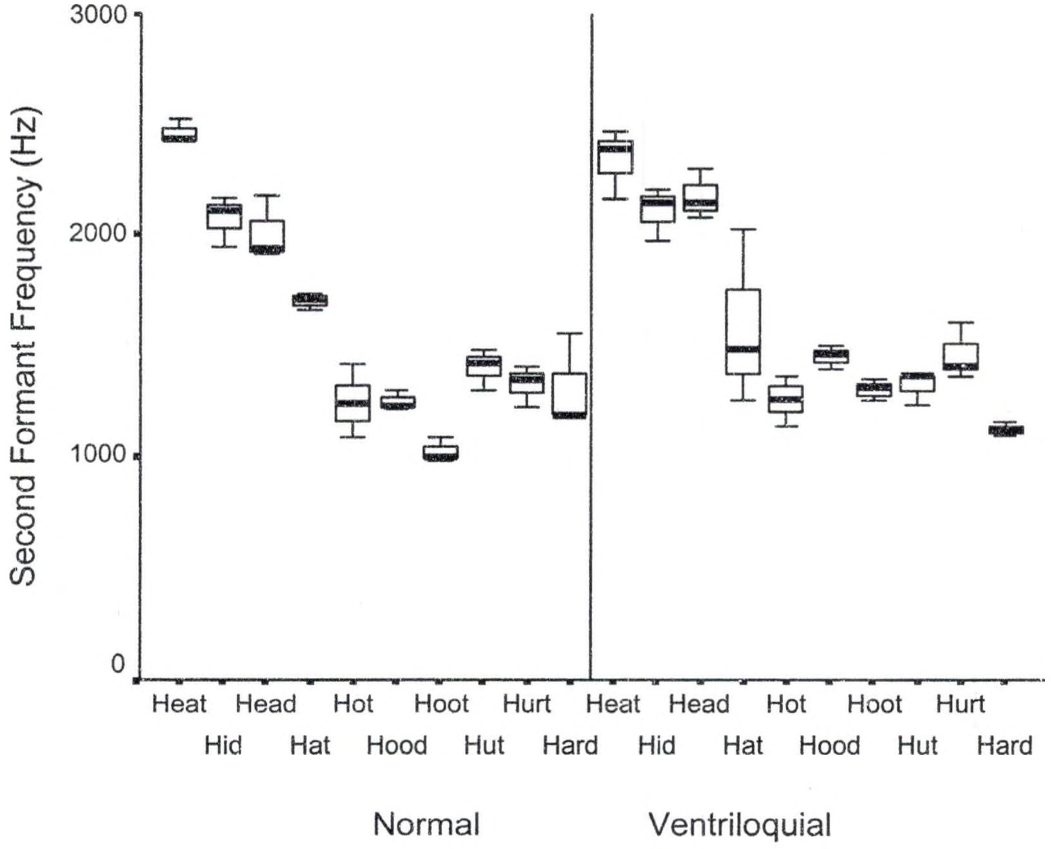

Figure 2. Boxplots of second formant frequencies for each word in Normal and Ventriloquial speech.

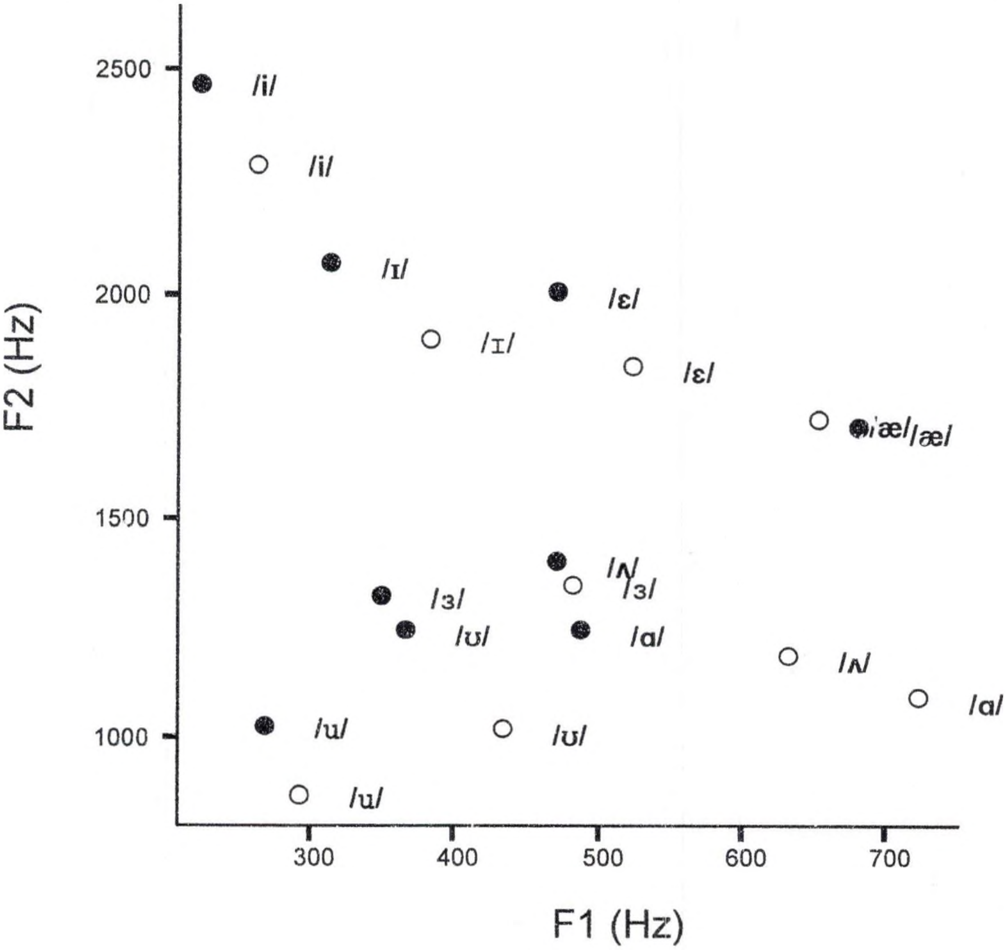

Figure 3. Scatterplot of first and second formant frequencies of Normal speech (solid circles) compared to first and second formant frequencies from Peterson & Barney (1950) data (outlined circles).

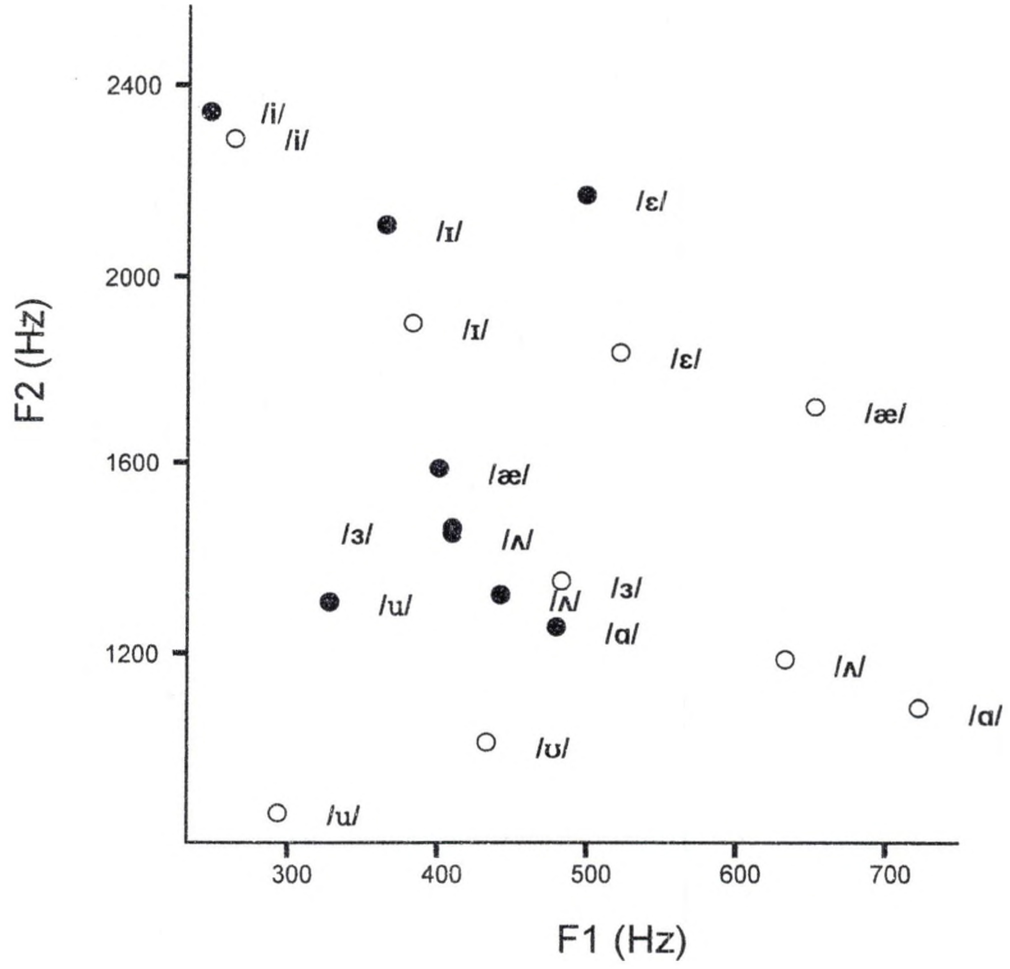

Figure 4. Scatterplot of first and second formant frequencies of Ventriloquial speech (solid circles) compared to first and second formant frequencies from Peterson & Barney (1950) data (outlined circles).

Table 3

t values for comparisons of fundamental frequency (F0), first formant frequency (F1) and second formant frequency (F2) between Normal and Ventriloquial speech for each word.

Word	F0	F1	F2
Heat	-9.85	-0.84	1.66
Hid	-10.78	-2.41	-0.37
Head	-16.43	-1.64	-1.22
Hat	-7.88	2.68	0.48
Hot	-9.67	0.06	-0.78
Hood	-20.07*	-4.36	-3.83
Hoot	-22.55*	-8.12	-10.91
Hut	-5.33	0.72	0.99
Hurt	-7.15	-3.79	-3.54
Hard	-5.88	-1.32	1.50

*$p < 0.0025$

CHAPTER VII

DISCUSSION

The data from this experiment do not strongly support the hypothesis that vowels produced in ventriloquial speech are the result of a different vocal tract configuration as compared to that required for normal vowel production. Specifically, there were no statistically significant differences in formant frequencies between vowel production in normal and ventriloquial speech for all vowels ($p > 0.0025$). However, a statistically significant difference was found between the fundamental frequencies for normal and ventriloquial speech for /u/ and /ʊ/ ($p < 0.0025$).

Based on the values obtained for the first and second formant frequencies, there does not seem to be a significantly large change in the size and shape of the vocal tract cavities between the two types of speech. On a perceptual level, it may be assumed that slight variations in F1 and F2 (size, shape of resonating cavities) are still understood by the listener. It seems logical then to assume that the articulators are being used in a relatively similar manner during the two speaking conditions. If the vocal tract cavities are being configured similarly, similar values for F1 and F2 would be obtained, as is shown in the case of this study (see Figures 1 and 2).

The findings of the present study support those found by Syrdal & Gopal (1986) whose study involved a quantitative perceptual model of vowel recognition. This model was based on the idea that a pattern of auditory excitation occurs when vowels are produced. They transformed the auditory distance between formants into bark differences. A "bark" is a unit of measure that represents acoustic energy falling within a "critical band," otherwise defined as approximately 1.3 mm along the basilar membrane and about 1300 cochlear neurons. Using the values obtained in Peterson & Barney (1952), Syrdal & Gopal examined the variability in formant and bark values in closely and widely spaced vowels. Closely spaced vowels are those in which the first and second formants are close in value such as /u/ and /o/. Vowels in which the first and second formants have a wider range of value are widely spaced vowels (e.g., /a/ and /ɔ/). Their results suggest that there is less variability in formant and bark difference for closely spaced formants than for widely spaced formants. Traunmüller (1981) suggests that variability in formant distance (distance between F1 and F2) in widely spaced vowels can still be identified by the listener. This means that slight variations in formant distance are perceived by the listener as the same vowel. This would indicate that variations in the resonating cavity would not affect intelligibility. Likewise, in the current study, it is shown that although statistically significant changes in the formant frequencies between the two types of speech were not found, slight variations do exist as evidenced in figures 3 and 4. These slight variations seem to be "ignored" by the listener and are therefore understood.

Liu, Tsao, & Kuhl (2002) studied the vowel space produced by subjects with Cerebral Palsy. Results indicate that the more extreme acoustic measures for a vowel are perceptually perceived as better exemplars and aid in the identification of vowels. Therefore, it is suggested that overall speech intelligibility is improved by increasing articulatory space. Likewise, Turner, Tjaden, & Weismer (1995) found that vowel space accounts for 45% of the variance in speech intelligibility. The current study presents evidence that ventriloquial and normal speech are produced in a similar manner. There is no significant difference between formant values and yet ventriloquial speech seems to be understood by the listener.

“Motherese” or “parentese” is used to define the exaggerated speech parents/caregivers use when talking with their infants. Research on infant development shows that adult speakers, when engaged in speaking to infants, use a significantly larger vowel space (Kuhl et al., 1997, Liu et al., 2000). This suggests that adult speakers not only speak more intelligibly but also provide better exemplars of vowels when addressing infants. It is suggested, in the context of this study, that this overexaggeration of vowel production is not necessarily needed as ventriloquial and normal speech are produced in relatively the same manner and intelligibility is not affected.

The differences found in measures of fundamental frequency (lower frequencies for normal speech, higher frequencies for ventriloquial speech) were significant in this study ($p < 0.0025$). This difference could be attributed to the fact the ventriloquist gives his puppet a character. The perceptually higher pitched voice is part of the puppet’s character. Yet, despite the fact that there is a change in pitch between the two types of speech, the ventriloquial speech is still intelligible.

There were several limitations to this study. One is the number of samples collected. Since only three utterances of each word containing the target vowels were obtained and analyzed, there may not be adequate statistical power to detect the presence of a small differences. If this study is replicated, it is suggested that the number of utterances recorded be increased. In addition, analyzing the productions of several participants would increase the reliability of the manner in which vowels are produced during ventriloquial speech. Also, perhaps more detailed and accurate data could be gathered if bark differences (Syrdal & Gopal, 1986) were measured as opposed to formant frequencies.

An extension of this project would be to collect data concerning the accuracy with which subjects are able to identify vowels produced in both types of speech. It would certainly appear that there is a one-to-one correspondance between the changes in F0, F1, and F2 and speech perception. For example, Ryalls & Lieberman (1982) conducted a study that examined fundamental frequency and vowel perception. They found that an increase in pitch leads to a decrease in intelligibility. Taking the frequency measures from Peterson & Barney (1952), they presented synthesized vowels to their subjects using three conditions for the F0: 1) average: 135 Hz, 2) low: 100 Hz, and 3) high: 250 Hz. Their participants identified vowels in the average and low condition of the F0 with greater accuracy than the high F0 vowels. Further evidence demonstrating that the F0 of a vowel influences vowel perception is presented by Fant et al. (1974), who found that the same formant "organization can result in the perception of two different vowels." Given this information, would a listener be able to correctly identify vowels if visual and sentence/context cues were eliminated in ventriloquial speech?

It would greatly add to this project if the effect of F3 on perception/intelligibility were examined. Although F3 was not closely examined in this study, it would appear that in the case of ventriloquial speech (where F3 is significantly reduced due to limited lip movement), F3 would have a significant impact on vowel production and hence perception.

In conclusion, this study does have significant implications for those individuals who need to acquire compensatory articulation skills due to anatomic and/or physiologic difficulties. A vowel can be identified by a listener even if there is variability in formant values. As such, if an individual could use his/her functional articulators to compensate for his/her anatomical and/or physiological difficulties, intelligibility can still be maintained. The integrity of vowel perception can be retained if the functional articulators are used in such a manner that only slight variations in formant values are produced.

References

American Speech-Language-Hearing Association Audiologic Assessment Panel 1996. (1997). *Guidelines for audiologic screening.* Rockville, MD: Author.

Bergen, E. (1938). How to become a ventriloquist. New York: Presto Books.

Bertelson, P., Vroomen, J., DE Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. Perception & Psychophysics, 62, 321-332.

Borden, G. J., Harris, K. S., & Raphael, L. J. (1994). Speech science primer: physiology, acoustics, and perception of speech (3rd ed.). Baltimore: Williams & Wilkins.

Fant, G, Carlson, R., & Granström, B. (1974). The [e] – [ø] ambiguity. Proc. Speech Communication Seminar. Stockholm 3, 117-121.

Ferrand, C. T. (2001). Speech science: an integrated approach to theory and clinical practice. Boston: Allyn & Bacon.

Frieda E. M., Walley, A. C., Flege, J. E., Sloane, M. E. (2000). Adults' perception and production of the English vowel /i/. Journal of Speech, Language, and Hearing Research, 42, 129-143.

Fucci, D. J. & Lass, N. J. (1999), Fundamentals of speech science. Boston: Allyn & Bacon.

Jack, C. E. & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. Perceptual and Motor Skills, 37, 967-979.

Kent, R. (1997). The speech sciences. San Diego, CA: Singular Publishing Group, Inc.

Kent, R. D. & Read, C., (1992). The acoustic analysis of speech. San Diego, CA: Singular Publishing Group, Inc.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. S., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. Science, 277, 684-686.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. Journal of the Acoustical Society of America, 100, 2425-2438.

Liu, H-. M., Tsao, F-. M., & Kuhl, P. K. (2000). Support for expanded vowel triangle in mandarin motherese. International Journal of Psychology, 35 (3-4), 337.

Liu, H-. M., Tsao, F-. M., & Kuhl, P. K. (December, 2002). The correlation of vowel space and speech intelligibility for individuals with cerebral palsy. Poster session presented at the annual meeting of the Acoustical Society of America, Fort Lauderdale, Florida.

Marin, F. N. (1997). Introduction to audiology. (6th ed.). Boston: Allyn & Bacon.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. The Journal of the Acoustical Society of America, 24, 175-184.

Ritchard, D. & Moloney, K. (1987). Ventriloquism for the total dummy. New York: Villard Books.

Rosch, E. (1975). Cognitive reference points. Cognitive Psychology, 7, 532-547.

Ryalls, J. & Behrens, S. (2000). Introduction to speech science: From basic theories to clinical applications. Boston: Allyn & Bacon.

Ryalls, J. H. & Lieberman, P. (1982). Fundamental frequency and vowel perception. Journal of the Acoustical Society of America, 72, 1631-1634.

Shipley, K. G. & McAfee, J. G. (1998). Assessment in speech-language pathology: A resource manual. San Diego, CA: Singular Publishing Group, Inc.

Shriberg, L. & Kent, R. (1995). Clinical phonetics (2nd ed.). Boston: Allyn & Bacon.

Syrdal, A. K. & Gopal H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America, 79, 1086-1100.

Thurlow, W. R. & Jack, C. E. (1973). Certain determinants of the "ventriloquism" effect. Perceptual and Motor Skills, 36, 1171-1184.

Traunmuller, H. (1981). Perceptual dimension of openness in vowels. Acoustical Society of America, 69, 1465-1475.

Warren, D. H., Welsch, R. B., McCarthy, T. J. (1982). The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. Perception & Psychophysics, 30, 557-564.