

EXPLORATION OF LIP SHAPE MEASURES AND THEIR ASSOCIATION WITH  
TONGUE CONTACT PATTERNS

by

Jessica L. Wagner

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Audiology and Speech-Language Pathology

Brigham Young University

December 2005

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Jessica Wagner

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

\_\_\_\_\_  
Date

\_\_\_\_\_  
Christopher Dromey, Chair

\_\_\_\_\_  
Date

\_\_\_\_\_  
Shawn Nissen

\_\_\_\_\_  
Date

\_\_\_\_\_  
D. J. Lee

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Jessica Wagner in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

---

Date

---

Christopher Dromey  
Chair, Graduate Committee

Accepted for the Department

---

Ron W. Channell  
Graduate Coordinator

Accepted for the College

---

K. Richard Young  
Dean, David O. McKay School of Education

## ABSTRACT

### EXPLORATION OF LIP SHAPE MEASURES AND THEIR ASSOCIATION WITH TONGUE CONTACT PATTERNS

Jessica Wagner

Department of Audiology and Speech-Language Pathology

Master of Science

A variety of tools and techniques have been developed to measure the movements of the vocal tract, specifically of the tongue and lips. In recent years, computer technology has allowed for extensive exploration of these precise movements and for the development of speech recognition systems. However, there has been relatively little work on the combination of visible facial movements and internal articulatory activity. In this study, two different technologies were used to explore the internal and external movements of speech production in eight speakers: palatometry quantified tongue contact patterns and computerized video image analysis was used to derive lip shape parameters. Results showed that the lip measures used here cannot predict the identity of phonemes in all speakers as well as the tongue contact patterns can. Results also indicated that the data from lip measures were strongly influenced by who the speaker was, whereas the palatometric data were not. These results suggest that more variation exists in lip shape

than in tongue contact patterns during speech production. Understanding more about lip measures and vocal tract movement during speech production may potentially benefit the area of speechreading; however, more research is needed to refine the procedures used.

## ACKNOWLEDGMENTS

It is a pleasure to thank the many people who made this thesis possible. I would like to thank Dr. Dromey first of all for his enthusiastic supervision of this study. Throughout the writing of my thesis he was endlessly patient and encouraging and took great efforts to explain things clearly and simply. Completing this thesis from the other side of the country would not have been successful without him. I would like to thank the professors and students in the electrical engineering and statistics departments for their contributions. The secretaries in the ASLP department were invaluable. I would also like to thank my fellow students for their friendship and for providing a stimulating and fun environment in which to learn and to grow.

I wish to thank my entire family for their support. My parents for their absolute confidence in me. My in-laws for all their help and assistance. My sweet husband for his explanations of statistical analyses and constant encouragement. My son for those mornings he slept extra long and allowed me time to write.

Lastly, and most importantly, I wish to thank Heavenly Father for blessing me and for his tender mercies which allowed me to complete this difficult task.

## TABLE OF CONTENTS

	Page
List of Tables .....	viii
List of Figures .....	ix
Introduction.....	1
History of EPG.....	6
Previous Studies with EPG .....	8
Lip Shape Analysis .....	10
Correlation Studies with Lip Shape Analysis .....	15
Other Technologies.....	17
Method .....	19
Participants.....	19
Instrumentation .....	19
EPG instrumentation.....	19
Lip shape instrumentation.....	21
Data Collection .....	21
Data Analysis .....	23
EPG data.....	23
Lip shape data .....	24
Principal Components Analysis.....	27
Results.....	30
Discussion.....	53
References.....	60
Appendix.....	64

## LIST OF TABLES

Table	Page
1. Examples of McGurk Effect .....	3
2. Viseme Groups for English Consonants .....	5
3. Results of Principal Components Analysis of 6 Facial Parameters .....	28

## LIST OF FIGURES

Figure	Page
1. Pseudopalate on Stone Model.....	20
2. Display Screen of LogoMetrix Palatometer .....	22
3. Color Space Conversion .....	25
4. Lip Shape Measurements.....	26
5. Principal Components Analysis of Facial Measures by Subject .....	31
6. Principal Components Analysis of Facial Measures by Sound .....	32
7. Principal Components Analysis of Palatometric Data by Subject.....	33
8. Principal Components Analysis of Palatometric Data by Sound.....	34
9. Principal Components Analysis of Facial Measures by Sound for Subject 0.....	36
10. Principal Components Analysis of Facial Measures by Sound for Subject 1.....	37
11. Principal Components Analysis of Facial Measures by Sound for Subject 2.....	38
12. Principal Components Analysis of Facial Measures by Sound for Subject 5.....	39
13. Principal Components Analysis of Facial Measures by Sound for Subject 6.....	40
14. Principal Components Analysis of Facial Measures by Sound for Subject 7.....	41
15. Principal Components Analysis of Facial Measures by Sound for Subject 8.....	42
16. Principal Components Analysis of Facial Measures by Sound for Subject 9.....	43
17. Principal Components Analysis of Palatometric Data by Sound for Subject 0.....	44
18. Principal Components Analysis of Palatometric Data by Sound for Subject 1.....	45
19. Principal Components Analysis of Palatometric Data by Sound for Subject 2.....	46
20. Principal Components Analysis of Palatometric Data by Sound for Subject 5.....	47
21. Principal Components Analysis of Palatometric Data by Sound for Subject 6.....	48

22. Principal Components Analysis of Palatometric Data by Sound for Subject 7 .....49
23. Principal Components Analysis of Palatometric Data by Sound for Subject 8 .....50
24. Principal Components Analysis of Palatometric Data by Sound for Subject 9 .....51

## Introduction

Language processing is a multi-modal phenomenon; a person perceiving a spoken message will often process visual as well as auditory information. For example under noisy conditions, comprehension is heightened if the speaker's face is entirely visible. The movements of the face supplement the auditory signal and the accuracy of perception is increased (Sumbly & Pollack, 1954).

Many people who are deaf are able to understand an entire message solely by speechreading. Speechreading is the ability to interpret speech by observing the lip and facial expressions of the speaker without any auditory information. A study conducted by Bernstein, Demorest, and Tucker (1998) examined the speechreading ability of participants. The participants, consisting of both people who were deaf and people who were not, identified phonemes in nonsense words. The results showed that, as a whole, people who were deaf performed better in the speechreading tasks than people who had normal hearing. However, the authors also stated that there was a wide range of speechreading proficiency in both groups. The skill of speechreading demonstrates that there is enough information communicated visually when a person speaks to allow another to correctly recognize what is being said. As noted above, even people with normal hearing are able to better comprehend what is being said if they are able to see the face of the person speaking, especially under noisy conditions (Neely, 1956). There is mounting evidence that listeners make use of visual information to increase speech intelligibility (Rosenblum & Saldaña, 1998).

An issue that still remains unclear is what characteristics constitute a good speechreader. This question necessitates further research. Campbell and De Haan (1998)

concluded from an experiment that speechreading was not faster for known or familiar faces than for unfamiliar faces. On the contrary: unfamiliar faces were speechread faster by participants in the study. The authors proposed that speechreading a familiar face activates semantic information about the known person, information that may interfere with the message being spoken. These findings also support the notion that communication is multimodal; it involves visual and auditory as well as cognitive factors.

Chen and Rao (1998) reviewed recent research that examines audio-visual integration in multimodal communication. They stated that human speech is bimodal in nature, involving visual and auditory characteristics. The McGurk effect (McGurk & MacDonald, 1976) also demonstrates the bimodal nature of human speech perception. When people are presented with conflicting auditory and visual stimuli, the perceived sound is a fusion or combination of what was presented in each modality. For example, when a person hears the sound /bɑ/ but sees the speaker saying /gɑ/ on a monitor, the person may not perceive either /gɑ/ or /bɑ/. Instead, what is perceived is a combination or something close to /dɑ/. Some other examples of the McGurk effect or these audio-visual combinations are shown in Table 1. This effect demonstrates that what a person perceives when another speaks depends not only on the acoustics, but also on visual cues such as lip movements (McGurk & MacDonald, 1976).

These observations have motivated some to link speechreading with audio signals in computer speech recognition systems. Eisenberg (2003) reported on the efforts of several researchers who are attempting to boost the accuracy of automatic speech recognition by incorporating a speechreading system. This would be useful in noisy places such as in a car, bus station, or other public place. Eisenberg (2003) reported that

Table 1

*Examples of McGurk Effect (McGurk & MacDonald, 1976)*

---

Auditory Stimuli	Visual Stimuli	Perceived Speech
bɑ	gɑ	dɑ
pɑ	gɑ	tɑ
mɑ	gɑ	nɑ
pɑ	kɑ	tɑ

---

the technology searches for skin-tone pixels and uses statistical models and vision algorithms to estimate the location of facial features, including the corners and center of the lips. The visual and audio features are then combined and analyzed by statistical models to predict what the speaker is saying.

Since the perception of speech can be bimodal, it follows that its production involves both visual and auditory features. Human speech is produced by combining the vibration of the vocal folds and the configuration of the vocal tract. The vocal tract is composed of articulatory organs, which include the lips, soft palate, teeth, and tongue. “Using these articulatory organs together with the muscles that generate facial expressions, a speaker produces speech. Since some of these articulators are visible, there is an inherent relationship between acoustic and visible speech” (Chen & Rao, 1998, p. 838). In other words, there is a relationship between the internal dynamics of the vocal tract and the visible movements of the face.

The basic unit of acoustic speech is called a phoneme. Chen and Rao (1998) proposed that analogous to the phoneme is the ‘viseme’ in the visual domain. The viseme is the basic unit of mouth movement and is therefore the smallest visibly distinguishable unit of speech. Sounds that are similar visually are grouped into the same category that represents a viseme. The visemes for English consonants can be grouped into nine distinct groups, as listed in Table 2 (Chen and Rao, 1998). In the current study a few visemes will be explored in more depth as well as their relationship to the configuration of the vocal tract. This exploration of both the internal and external movements of speech, will help us better understand the influence of vocal tract activity on lip shape and, by extension, will reveal more about the perception of speech from visual

Table 2

*Viseme Groups for English Consonants (Chen & Rao, 1998)*

---

Group	Consonants
1	f, v
2	θ, ð
3	s, z
4	ʃ, tʃ
5	p, b, m
6	w
7	r
8	g, k, n, t, d, y
9	l

---

information. In this research, two different technologies will be used to explore these two aspects of speech production more fully: electropalatography (EPG) will display tongue contact patterns, and computerized video image analysis will be used to quantify lip shapes.

### *History of EPG*

“The technique of electropalatography is designed to record details of the timing and location of tongue contacts with the hard palate during continuous and static speech sounds” (Hardcastle & Gibbon, 1997, p. 149).

Kydd and Belt (1964) stated that Oakly Coles was the first person known to use a form of palatography in 1872. The technique Coles used was static palatography. A powdered mixture was applied to the palate; after the production of a single speech sound, the palate was sketched or photographed. As Kydd and Belt explained, a single speech sound was used because producing a second sound would destroy the pattern of the first sound. Therefore, this method was not representative of what typically occurs in speech. Dynamic observations were not possible using this early procedure.

In the 1960s the single sound limitation of palatography was remedied through the discovery of electronic means to observe the tongue’s contact with the palate (Kydd & Belt, 1964). At the University of Washington, Kydd and Belt mounted silver electrodes on a denture-like plate, which was inserted into the user’s mouth. When the tongue touched these electrodes, tiny voltages were generated. Kydd and Belt used these voltages to detect linguapalatal contact. It was difficult to detect the small electrical charges made from contact with the electrodes, thus defeating this approach. Saliva, which collected between the sensing electrodes, also proved limiting to this method.

Kydd and Belt explained that another problem was the inability to make numerical measures from the linguapalatal contact patterns. However, the potential of dynamic, electronic palatography was becoming a reality.

Throughout the 1960s, experimental systems for measuring dynamic linguapalatal contact based on completing an electrical circuit were developed by laboratories in Russia, Japan, and the United States of America (Fletcher, McCutcheon, & Wolf, 1975). These studies introduced new circuitry to overcome the saliva bridging problems of electronic palatography. Improvements introduced by Fletcher and his team in Alabama (1975) increased the number of electrodes from 24 to 96. Fletcher also helped to standardize the placement of the sensor array which opened the door to numerical measurement and more precise across-speaker comparisons. These advantages made dynamic, electronic palatography more feasible.

Another obstacle that needed to be overcome was the thickness of the pseudopalate. In its early stages, the palate was thick enough to require training to compensate for the articulatory differences caused by the pseudopalate. Baum and McFarland (1997) investigated this barrier and found that, “adaptation to a structural modification of the oral cavity can occur relatively quickly with intensive, target specific practice” (p. 2357). Currently, the pseudopalate is thin enough not to require extensive adaptation training.

Electropalatography has since become a powerful research tool. It has also become effective in clinical settings for treating children with persistent misarticulations and the hearing impaired, as well as other populations. Numerous studies have been conducted with EPG to explore complicated speech movements and patterns as well as to

evaluate EPG as a treatment technique in a variety of disorders.

*Previous Studies with EPG*

Hardcastle and Gibbon (1997) stated, “EPG has become well established in many experimental phonetics laboratories and speech therapy clinics throughout the world as a safe and convenient technique for use in the investigation of an important aspect of tongue activity” (p. 149). The wide applicability of this technique and its usefulness as a tool in phonetic research is reflected in the range of phenomena that has been explored in recent years (Hardcastle & Gibbon, 1997). Areas of research include: the articulatory characterization of lingual fricative production, aspects of lingual articulation and coarticulation in a variety of languages, articulatory dynamics of implosive and egressive stops, alveolar and dental stops in Kimvita Swahili, palatalization and palatal consonants in a variety of languages, symmetry of lingual gestures, emphatic consonants in Arabic, force of articulation in French, vocalized /l/ in English, articulatory correlates of voicing, general articulatory dynamics of the tongue, timing in /kl/ clusters, and rate variation and its articulatory correlates (Hardcastle & Gibbon, 1997).

EPG has proven useful for giving information about tongue shape and placement. Forster and Hardcastle (1998) used the EPG system to look at the speech of two stuttering participants. They compared lingual gestures in stuttering speakers with control speakers in fluent and disfluent speech. The EPG technique was very useful in examining the gestures the tongue was making in these participants. Results showed that compared to control speakers, the stuttering individuals displayed increased tongue contact for alveolar and velar plosives and less contact for fricatives during fluent speech. The authors proposed this may be explained by increased muscular tension that is common in

stutterers, or an inability to produce and maintain delicate neuromuscular control. They also suggested these differences could be a result of a deliberate strategy used by the participant to prevent disfluencies (Forster & Hardcastle, 1998). This study shed more light on the dynamic movements of the tongue in these speakers and proved EPG to be a useful technique in the analysis of articulatory characteristics.

Murdoch, Gardiner, and Theodoros (2000) recently used EPG to document the location and timing of linguapalatal contacts during speech production in a 52-year-old man with dysarthria secondary to multiple sclerosis. EPG gave Murdoch et al. information about tongue contact on the palate as well as the spatial configuration inside the mouth, calculated by the proportion of electrodes which were contacted. For example, the results showed that the participant with multiple sclerosis made contact between the tongue and palate in the correct anatomical location for each phoneme. However, his spatial configurations differed from speakers in the control group in the proportion of electrodes contacted. The speaker with MS consistently made more contact with the palate than did those in the control group. Therefore, Murdoch et al. suggested that rather than concentrating on correcting tongue placement in therapy, treatment should focus on correcting the temporal aspects of the tongue's movement.

Christensen, Fletcher, and McCutcheon (1992) also showed EPG to be helpful in assessing esophageal speakers. Christensen et al. used EPG to assess the linguapalatal contact patterns in a 56-year-old man who had 15 years of experience in using esophageal speech following a laryngectomy. The authors found that this esophageal speaker had a narrower average medial groove width in comparison to normal speakers when producing /s/ and /z/. They hypothesized that this difference in the groove contributed to the

articulatory distortions. Christensen et al. could then recommend that esophageal speakers be taught to form a narrower or tighter linguapalatal groove for /s/ and /z/ to compensate for this. In summary, these previous studies show that EPG can be very useful for both treatment and research.

### *Lip Shape Analysis*

Another field of study that has been explored extensively is lip shape and the visual components of speech information. It is important to understand how vocal tract movement is correlated with the visibly perceptible aspects of speech. There has been relatively little work on the correlation between these two components. Most descriptions of visual speech information have explored the visual component alone and have involved static facial positions or static lip positions. Montgomery and Jackson (1983), for example, described the visual information for vowels in terms of a scaling space representing the degree of lip spreading, rounding, and tongue height. These features were captured in a still photograph and further analyzed.

Another study that examined lip shape and movement was conducted by Ramsay, Munhall, Gracco, and Ostry (1996). They explored the derivation of principal components to represent more complex movement dynamics. They described a method of functional data analysis to examine speech production and further understand the basic principles of articulation. Ramsay et al. monitored lip motion by using OPTOTRAK, which is an optoelectronic tracking system that can transduce the three dimensional position of markers. Eight infrared light-emitting diodes (IREDs) were attached to the border of a single participant's lips using double-sided tape. Another six diodes were attached to the participant's head in order to allow the researchers to correct for head

movement after data collection. The OPTOTRAK system then transduced the information and the researchers analyzed the signals using a waveform editor, filters and algorithms. These data were then analyzed using statistical methods. This paper described these analyses in detail, which allowed other researchers to further explore lip motion.

Le Goff, Guiard-Marigny, and Benoît (1997) also analyzed lip shape. However, their purpose was to create a three dimensional physical model of the lips and the human face. To measure the lip shape, a speaker was video taped from the front and side by two video cameras. The speaker wore blue lip makeup to facilitate the automatic analysis that needs to differentiate the lips from the rest of the face. The blue makeup made the lips darker than any other pixel on the screen. The video frames were then digitized, and the contours of the lips with a reference point on the chin were extracted using software. Finally, geometric measurements were made on each contour to reproduce the lip shape. Le Goff et al. controlled the lip model with five parameters which were easy to measure on a speaker's face. These parameters include width, height, lip contact protrusion, upper lip protrusion, and lower lip protrusion. The authors also created a three dimensional physical model of the entire face using similar techniques. They also evaluated the contribution of the lip model and face model to speech intelligibility. Their results "confirmed the importance of visual information in the perception of speech" (Le Goff et al., 1997, p. 235). They found that the speaker's natural face restored two-thirds of the missing auditory information, the facial model restored half of it and the lip model alone restored one-third of it. This study revealed the importance of visual cues as well as an innovative way to analyze the lip shape alone.

Lucero and Munhall (1999) created a three dimensional physical model of the

human face. They reported using an OPTOTRAK system to analyze facial movements. However, they also incorporated a facial mesh and compared the data to electromyography (EMG) in their study. The facial mesh is a multi-layered model of facial tissues that is deformable. The nodes in the mesh are point masses, which are connected. The nodes are arranged in three layers representing the structure of facial tissues: the epidermis, the fascia, and the skull surface. There is also a layer which represents the dermal-fatty tissues and muscle. The composition of the facial mesh is based on anatomy literature and measures of muscle geometry in cadavers. The virtual mesh is placed over a speaker's face and can then measure the movements. In their study, Lucero and Munhall created a three dimensional physical model of the human face, and set out to model the peripheral speech motor system and thereby increase the understanding of speech motor control and audiovisual speech perception. They created facial movements with the multilayered mesh, collected measurements, and then compared these measurements to EMG data from a real human face. The model represented the soft tissue biomechanics of the human face. Overall, Lucero and Munhall found a good match between the recorded EMG data and the model's movements. Although the model includes many simplifications and is prone to a number of problems, it is a relatively accurate model of the human face and may be useful for further speech perception and production research.

This model of the human face was later refined by Pitermann and Munhall (2001). These authors desired to create a realistic animation of the human face based on three dimensional kinematic recordings. In the process, they analyzed facial shape and movement. As in the previous study, Pitermann and Munhall utilized the OPTOTRAK

system combined with a facial mesh with numerous nodes to create a realistic model of the face. Instead of comparing these data to EMG data of a human speaker, they inverted the data and then compared them. To invert the data, the authors worked backward from the facial movements to predict the characteristics of the underlying EMG data. The inversion technique was dynamic and allowed the researchers to match the dynamics of the face model to a kinematic pattern by continuously updating the muscle activity. The correspondence between the animated face and the human face data was very good and the animation was of high quality. Therefore, the authors refined their method for analyzing facial movement and replicating it accurately. This demonstrates the variety of sophisticated techniques that have contributed to the research field of facial kinematics.

The advantages to this modeling approach were discussed by Munhall and Vatikiotis-Bateson (1998). They suggested that using the human face model for research results in greater computational efficiency and a high level of realism. However, Munhall and Vatikiotis-Bateson (1998) added that the most important outcome of the “synthesis of moving faces from neuromotor and biophysical parameters is that a more rational approach to speech research is brought about by pursuing production and perception [research] together rather than separately” (p. 136). Through creating this human face model, researchers can search for similarities or correlations between speech production and speech perception.

Kroos, Kuratate, and Vatikiotis-Bateson (2002) also analyzed lip shapes as well as the movements of the entire face. Their measures were globally distributed over the whole face but could also be applied to specific regions, such as the lips. The authors described a noninvasive method of measuring face motion during speech production,

which did not interfere with the speaker's articulation. Kross et al. extracted measures from standard video sequences using an image analysis process which concentrated on face structure and motion. They made measures by using an ellipsoidal mesh which was fit to the face of a speaker. An algorithm tracked the position of the mesh nodes relative to the face for an entire video segment. OPTOTRAK and a head mount were used for precise head tracking. The authors compared these measures to three dimensional marker data that were also recorded for the same utterances and the same speaker to ensure reliability.

Recently, this technique has been refined even further. Lucero, Maciel, Johns, and Munhall (2005) measured face motion during speech production by using an ellipsoidal mesh consisting of 38 markers distributed on a speaker's face. Six additional markers were located on a headband to use as a reference and to compute the origin and orientation of a head coordinate system. The 38 markers were grouped into clusters containing one primary marker which drives the movement of all the other secondary markers. An algorithm was used to extract data from the clusters and to further measure the facial movements. This new system allowed the researchers to model facial movement instead of static shapes, by looking at regions of facial movements in the clusters. Doing so provided a more realistic facial animation. Lucero et al. explained that the purpose of this work was to "develop a model of human face biomechanics that may be used as a tool for speech production and perception studies" (p. 405). The researchers refined a technique to create animation systems that realistically reflect the motion patterns across the face, which will further the research of the visible articulators and their movements.

### *Correlation Studies with Lip Shape Analysis*

The present study attempts to examine both the shape of external articulators during speech production and internal vocal tract patterns. There is very little research combining these two aspects in a single study. However, Munhall, Löfqvist, and Kelso (1994) examined the coordination of the lip with the larynx. These authors concentrated on lip motion relative to laryngeal activity and did not analyze lip shape. However, the correlation of these two parameters helped shed light on the linkage between the vocal mechanism and the peripheral speech structures. The focus of their experiment was to investigate the coordination between the larynx and the lips and jaw while a speaker was producing a voiceless consonant. During the speech production, a mechanical perturbation was applied to the lip, also affecting the jaw. The movements of the lip and jaw were analyzed using an optoelectronic technique, while the laryngeal responses were analyzed using transillumination, intraoral pressure, and the acoustic signal. The results of this study revealed that the larynx responded in two ways to this perturbation of the lip. First, there was a delay in onset of glottal abduction. And second, the glottal abduction-adduction cycle increased in duration in the adduction phase. This indicated that the laryngeal articulatory movements can be rapidly altered when a change is applied to the lip during voiceless stop production. This cooperative behavior suggests that “in some very general way the articulators are coupled during speech production” (Munhall et al., 1994, p. 3613) and are therefore coordinated.

It is known that “facial motion during speech is a direct consequence of vocal-tract motion which also shapes the acoustics of speech” (Yehia, Kuratate, & Vatikiotis-Bateson, 2002, p. 555). In other words, “Configuring the vocal tract during speech

simultaneously shapes the acoustics and deforms the face” (Kuratate, Munhall, Rubin, Vatikiotis-Bateson, & Yehia, 1999, p.1). A group of researchers have also explored the motion-acoustics relationship that occurs during the production of speech, namely that which exists between head motion and fundamental frequency. This relationship has been examined in a number of studies (Kuratate et al., 1999, Yehia et al., 2002, Yehia, Rubin, & Vatikiotis-Bateson, 1998). Yehia et al. (2002) proposed that speech acoustics can be used to predict face motion and vice versa. They used this assumption to develop a system that takes speech acoustics as input, and gives as output the coefficients needed to create natural face and head motion in a model. To accomplish this task, Yehia et al. measured face and head motion and speech acoustics. To measure face motion they utilized an OPTOTRAK system with markers placed on the cheeks, chin, and around the lips. To measure head motion, markers were placed around a lightweight frame worn on the head and transduced with the OPTOTRAK system. In previous work (Yehia et al., 1998), linear mappings were used to relate face motion and speech acoustics. The results from this previous study indicated that face motion could be partially determined from some acoustic features of speech, such as the root mean square amplitude and line spectrum frequency pair coefficients. These findings were used to refine the research. Yehia et al. (2002) collected the audio signal, calculated the fundamental frequency, and analyzed the figures using linear predictive coding synthesis and other mathematical analysis procedures. The authors used this information to create realistic talking faces. The results of this study indicated that between 80 and 90% of the variance observed in face motion can be accounted for by the speech acoustics. The authors also found that fundamental frequency estimated from head motion contains natural prosodic

information (Yehia et al., 2002). Further research is essential to create a more accurate system; nevertheless this again demonstrates the strong link between laryngeal activity and head motion.

### *Other Technologies*

A variety of technologies have been used to analyze the movements of the face as well as the vocal tract. Munhall (2001) discussed the use of neural imaging in studying articulation and vocal tract movements. He described the use of X-ray imaging to shed light on vocal tract movements. The use of this technique is limited, however, due to the risks associated with radiation exposure. Other imaging techniques include the use of ultrasound and electromagnetometry. Munhall described these tools in the context of examining the neural control of speech production. This differs from tools previously discussed which were used to make kinematic measures of the articulators. Munhall also explained the use of functional magnetic resonance imaging (fMRI) in speech research. He concluded that fMRI allowed for the first recording of three dimensional images of the complete vocal tract and contributes to the understanding of the vocal tract movements. Computed tomography (CT) has also been a positive addition to research technology. Other techniques include positron emission tomography (PET), magnetoencephalography (MEG), and event-related potentials (ERP), among others, each with its own strengths and weaknesses. These instruments are available to researchers today to aid in expanding the understanding of speech production and vocal tract movement.

There have been many studies which analyzed lip shape or EPG data separately. There have also been numerous studies performed to correlate visual information in

speech perception with audio information, as previously discussed. This current study examines both visual facial information and vocal tract movement to try to understand more about the process of speech production. Palatometrically derived tongue contact patterns were studied as well as visually-based measures of lip shape to accomplish this. The territory of these analyses is relatively uncharted and will increase understanding of the basic processes of human communication. For example, some of the essential characteristics that identify people who speechread well have yet to be determined. Speechreading itself demonstrates that substantial information concerning speech sound formation must be present in visible cues on the face. The current research holds the potential to more fully describe the specific sources and characteristics of such cues.

## Method

### *Participants*

The participants for this study were eight adults associated with Brigham Young University. All participants were native speakers of English, and passed a hearing screening bilaterally at 15 dB for the following frequencies: 4000Hz, 2000Hz, 1000Hz, and 500Hz. All participants passed a speech screening and an oral peripheral exam administered by a speech language pathologist. Each participant agreed to participate in the study by reading and signing a consent form (Appendix), which was approved by the institutional review board. Speakers are numbered 0, 1, 2, 5, 6, 7, 8, and 9. The data from speakers 3 and 4 were excluded because they were not collected simultaneously.

### *Instrumentation*

In the present research tongue contact patterns derived from electropalatography and visually-based measures of lip shape were examined.

*EPG instrumentation.* The EPG system used in this study was the LogoMetrix palatometer. The palatometer has been previously described (Fletcher, 1989). Briefly, the user wears a custom-made artificial palate which is similar to an upper dental retainer. The artificial palate can be made by a speech language pathologist, who makes a dental casting of the user's hard palate and upper teeth. This 'stone model' is used to vacuum-form a thin (0.5mm) plastic sheet to conform to the exact shape of the speaker's upper teeth and hard palate. A flexible printed circuit with 118 gold-plated sensors is glued onto the plastic pseudopalate, allowing the device to interface with a computer to display the contact pattern of the tongue with the sensors on the palate (Figure 1). The surface is



*Figure 1.* Pseudopalate on Stone Model.

smooth because the sensors do not protrude from the artificial palate, thus the tongue can make contact comfortably.

Each sensor is displayed on the computer screen as a black dot. The dots change to filled blue circles as the sensors are contacted by the tongue or lips (Figure 2). The blue circles change back to black dots as the contacts are released.

*Lip shape instrumentation.* Two color digital video cameras were used to capture images of the speaker's mouth from directly in front of the speaker's mouth and also from 90 degrees to the side. Lip shape data were then extracted from these digital video images and recorded as a sequence of x and y coordinates for four sounds: /s, ʃ, i, u/. These sounds were chosen because their articulatory postures can be held constant over time, which facilitates data collection. Therefore, sounds that cannot be held, such as affricates and plosives, were excluded from this study. The /s/ sound was chosen because it is frequently misarticulated in school-age children and is of interest in the field of speech pathology. The vowels /u/ and /i/ were chosen because of their contrast of lip rounding versus lip retraction. The sibilant /ʃ/ was chosen because of its more subtle contrast with /s/. A reference shape was also recorded for each speaker with relaxed and closed lips.

#### *Data Collection*

The data were collected in three separate sessions with each participant. The first session included a routine oral mechanism examination, speech screening, and a hearing screening to determine participant eligibility. During the second session an impression of the speaker's upper teeth and palate was made. Following this, the pseudopalate was made from a stone model of the palate with a thin sheet of orthodontic plastic. The



Figure 2. Display Screen of LogoMetrix Palatometer.

contact electrodes were installed by gluing a flexible printed circuit to the surface of the plastic pseudopalate. During the third session the participant performed a variety of phonetic tasks, including production of the four target sounds using the EPG simultaneously with the cameras capturing lip shapes. Each target phoneme was produced 20 times and was held for approximately 3 seconds. The stimuli were randomized and presented to the speakers by the researcher. Each speaker participated in a 15 minute training period with the pseudopalate in place in order to facilitate speech adaptation. Each participant also wore dark lipstick to facilitate data analysis of the lip shape. These data were then analyzed using custom software applications in Matlab.

#### *Data Analysis*

Because of this study's exploratory nature, the use of new techniques, and the small participant population, conventional group statistics, such as ANOVA models, were not appropriate. However, a principal components analysis was performed on the lip shape and tongue contact data to determine whether the four sounds are separable in their lip or tongue contact patterns.

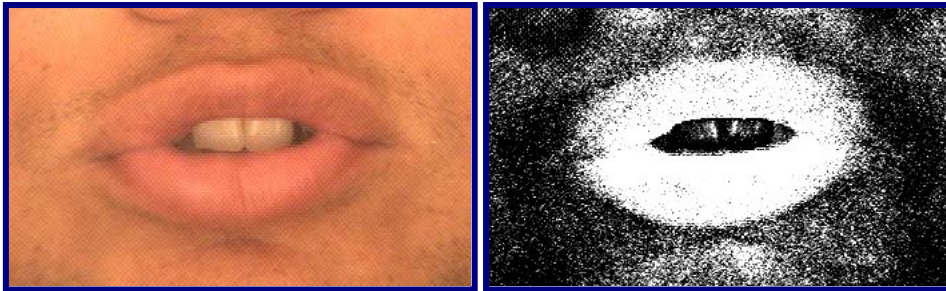
*EPG data.* Traditionally, analysis of palatometric data has been a qualitative examination of the patterns formed by the tongue during speech production. In order to facilitate the correlation of palatometric data with lip shape measurement data, an Excel spreadsheet summarizing the palatometric patterns was created. The spreadsheet displayed all 118 sensors on one axis and time across the other axis. Ones and zeros represented whether the sensor was being touched by the tongue at a particular moment in time. The number one indicated tongue contact, while the number zero indicated no contact. A line of data was taken to represent each sound. These data were later used in

the principal components analysis.

*Lip shape data.* The lip shapes were extracted using a newly developed lip color segmentation algorithm. Because of the lack of contrast between the skin around the mouth and the lip itself, it is difficult to find the edge of the lip. This problem was addressed by using this algorithm. A linear color space conversion technique was developed which makes the lip color high value and the skin color low value to perform lip shape segmentation. This algorithm mapped every point in an image to a value between 0 and 255 depending on the point's color values of red, green, and/or blue.

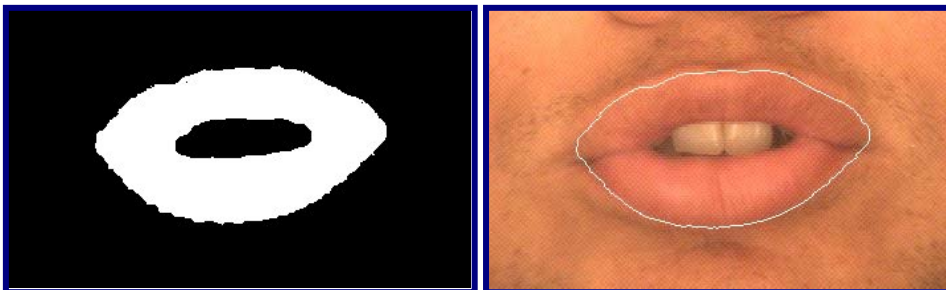
Two sets of data were needed in order to calibrate the lip segmentation algorithm. One set contained points that were part of the lips and the other contained points that were part of the face but not part of the lips. The red, green, and/or blue values of every point in these two data sets were then used to calibrate the lip segmentation algorithm. Once the system was calibrated, points that were most similar to the lips mapped to high values while points that were not similar mapped to low values. Morphological operations were then used to clean up the images. The extracted outlines of the lips were examined and data points were moved manually to correct segmentation errors which removed any sharp edges before analysis (Figure 3).

To analyze the lip shape, a technique was used from Lee, Bates, Dromey, Xu, and Antani (2003). Eight relevant points were identified on each lip shape. Front and side views were both used. From these eight points, four quadrants were used to create quantitative shape data. From these four quadrants, six measurements were made of critical size and angle features. These lip shape parameters include height, width, upper lip protrusion, lower lip protrusion, the upper angle, and the lower angle (Figure 4).



Original Image

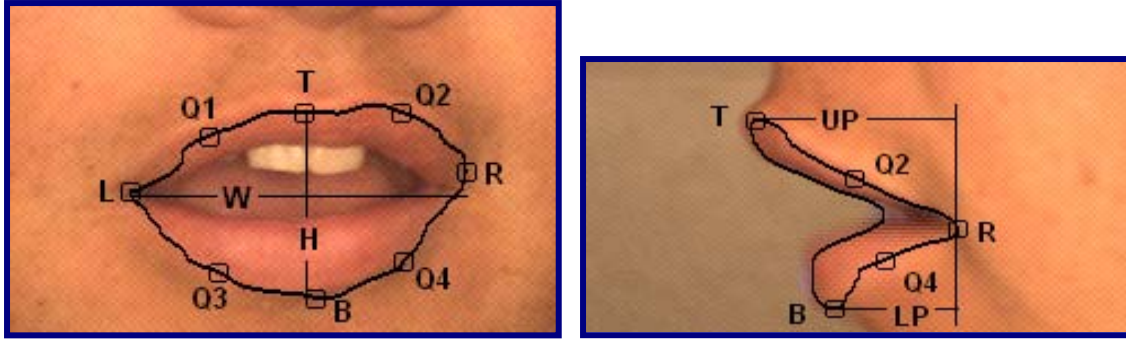
1-D Linear Color Space



Morphological Operation

Superimposed Lip Contour

*Figure 3. Color Space Conversion.*



*Figure 4.* Lip Shape Measurements. T=top, B=base, R=right, L=left, W=width, H=height, UP=upper protrusion, LP=lower protrusion.

*Principal Components Analysis.* A principal components analysis was performed to reduce the overall complexity and aid in the interpretation of the data. The purpose of a principal components analysis is to reduce the number of factors of a data set while retaining as much information as possible. This is accomplished by creating components. For example, to measure basketball skill, instead of using percentage of jump shots, free throws, lay-ups, and 3-pointers, these factors could all be combined into one 'shooting ability' component.

This analysis reduced the 6 facial parameters and 118 data points to fewer components in order to simplify the interpretation of the contributions and interactions of the lip shape data and the EPG data for the 8 speakers. The principal components analysis was also performed to determine the amount of separation between the individual sounds and the speakers. In other words, the analysis helped to determine whether the data could distinguish one sound from another and one speaker from another.

As can be seen in Table 3, two components emerged from the analysis of the 6 facial parameters with eigenvalues greater than 1, accounting for 76.33 percent of the variance. This suggests these two components are the most useful for this study. The first component (eigenvalue = 2.94) was composed of the following four facial parameters: height, width, upper angle, lower angle. These parameters showed loadings of 0.51, -0.45, 0.41, and 0.53 with cross loadings of 0.17, 0.35, -0.19, and 0.13 respectively. The second component (eigenvalue = 1.64) was composed of the remaining two parameters of upper protrusion and lower protrusion. These parameters demonstrated component loadings of 0.58 and 0.67 with cross loadings of 0.27 and -0.12 respectively. The loadings and cross loadings showed how these parameters separate themselves into the

Table 3

*Results of Principal Components Analysis of 6 Facial Parameters*

---

Parameters	Component 1	Component 2
Height	0.510610	0.174499
Width	-.448052	0.347265
Upper Angle	0.412612	-.191839
Lower Angle	0.528208	0.130964
Upper Protrusion	0.271890	0.587063
Lower Protrusion	-.123897	0.671089

---

two components, thus reducing the 6 variables of facial parameters into 2 composite variables. These two components were the only components of the facial measures used for the rest of the analysis. A principal components analysis was also performed for the palatometric data. From this analysis two components were selected to investigate the palatometric data. These components describe the 118 different variables and their weighted combination. These two components are separate and distinct from the two components for the facial measures.

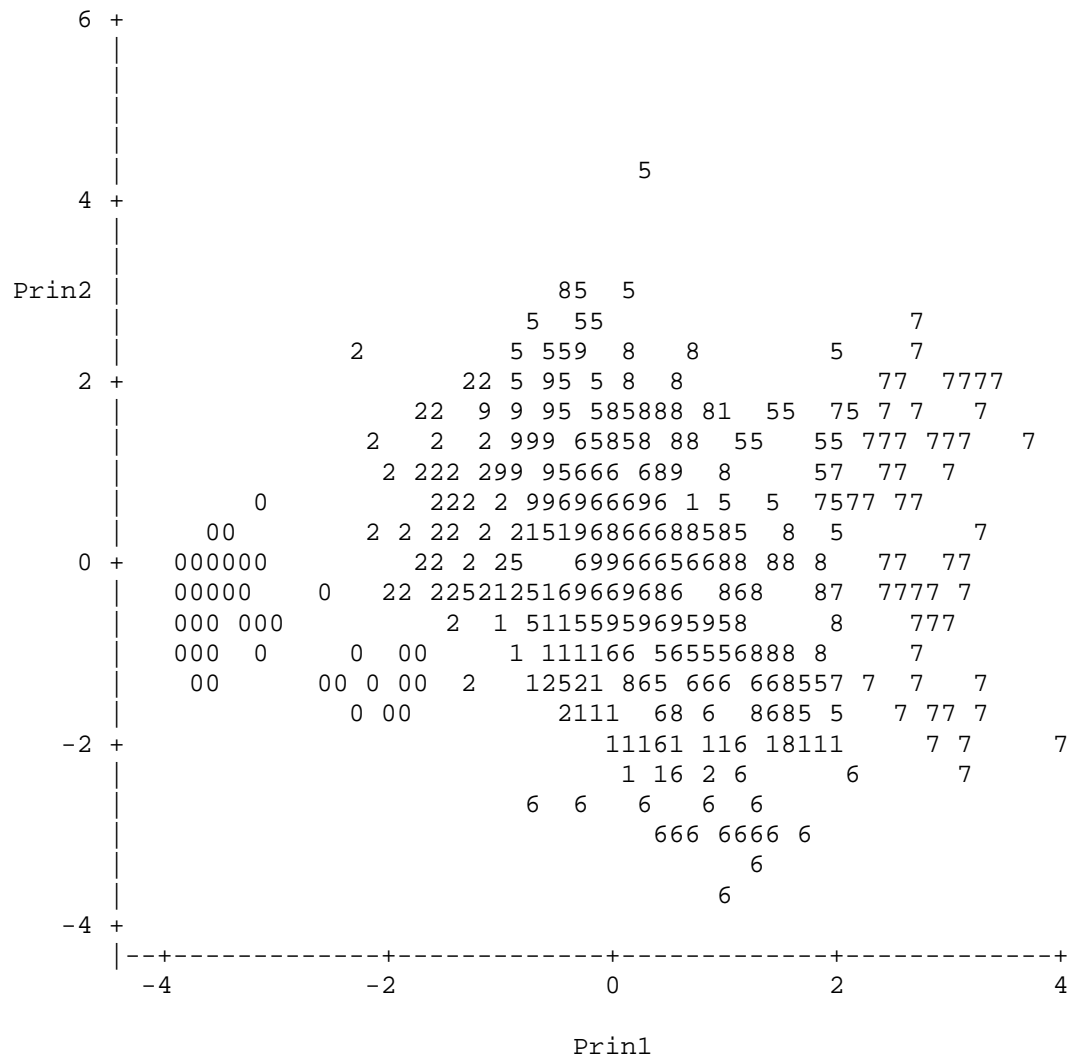
## Results

The two principal components were used to create plots of the facial measures and the palatometric data. The x-axis is principal component two and the y-axis is principal component one for each plot discussed in this section. Figure 5 is a plot of the six facial measures for each individual speaker. The symbols refer to the speaker number. As shown in the plot, speaker 0 and speaker 7 separate themselves out from the rest of the speakers. The other six speakers are not distinguishable. This suggests that the data are influenced by who the speaker is. The facial measures are not generic across the different speakers.

Figure 6 is a plot of the six facial measures by sound across all speakers for principal components one and two. This plot shows the only sound that is distinguishable from the rest is /u/. The other sounds overlap, which suggests it cannot be determined what the sound is from the facial measures alone.

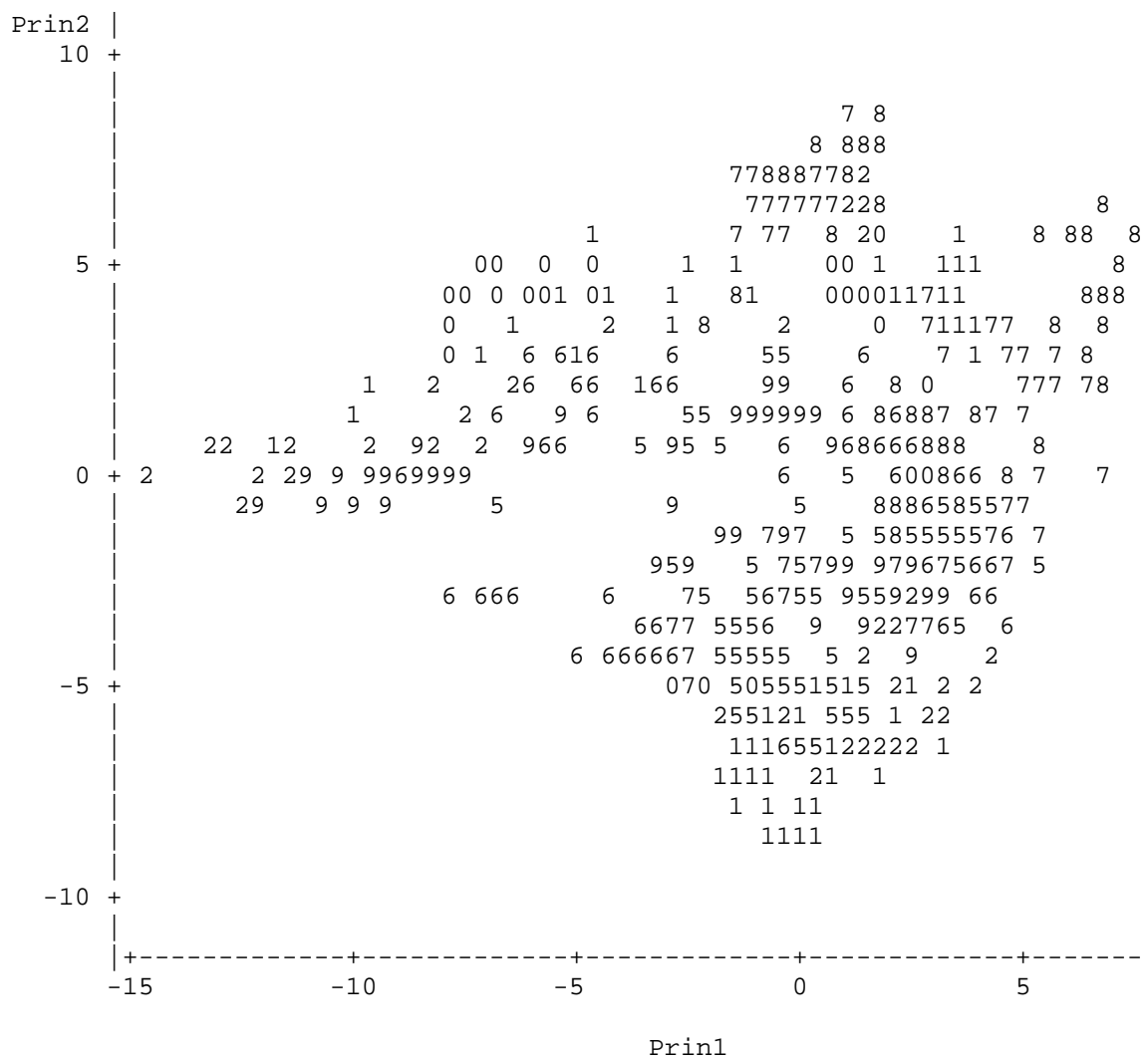
Figure 7 is a principal component plot of the 118 palatometer data points. The x-axis is principal component two and the y-axis is principal component one. The symbols refer to the speaker number. This plot shows that the speakers do not separate from each other for the palatometer data. The speakers overlap in the palatometer data compared to the facial measures. This suggests that the sounds may be generic across the different speakers. In other words the tongue contact data do not distinguish which participant is speaking. Or, in other words, the data are not influenced by the speaker as much as is the case for the facial data.

Figure 8 is a principal component plot of the palatometer data points by sound across all speakers. This plot shows that the sounds do separate well compared to the

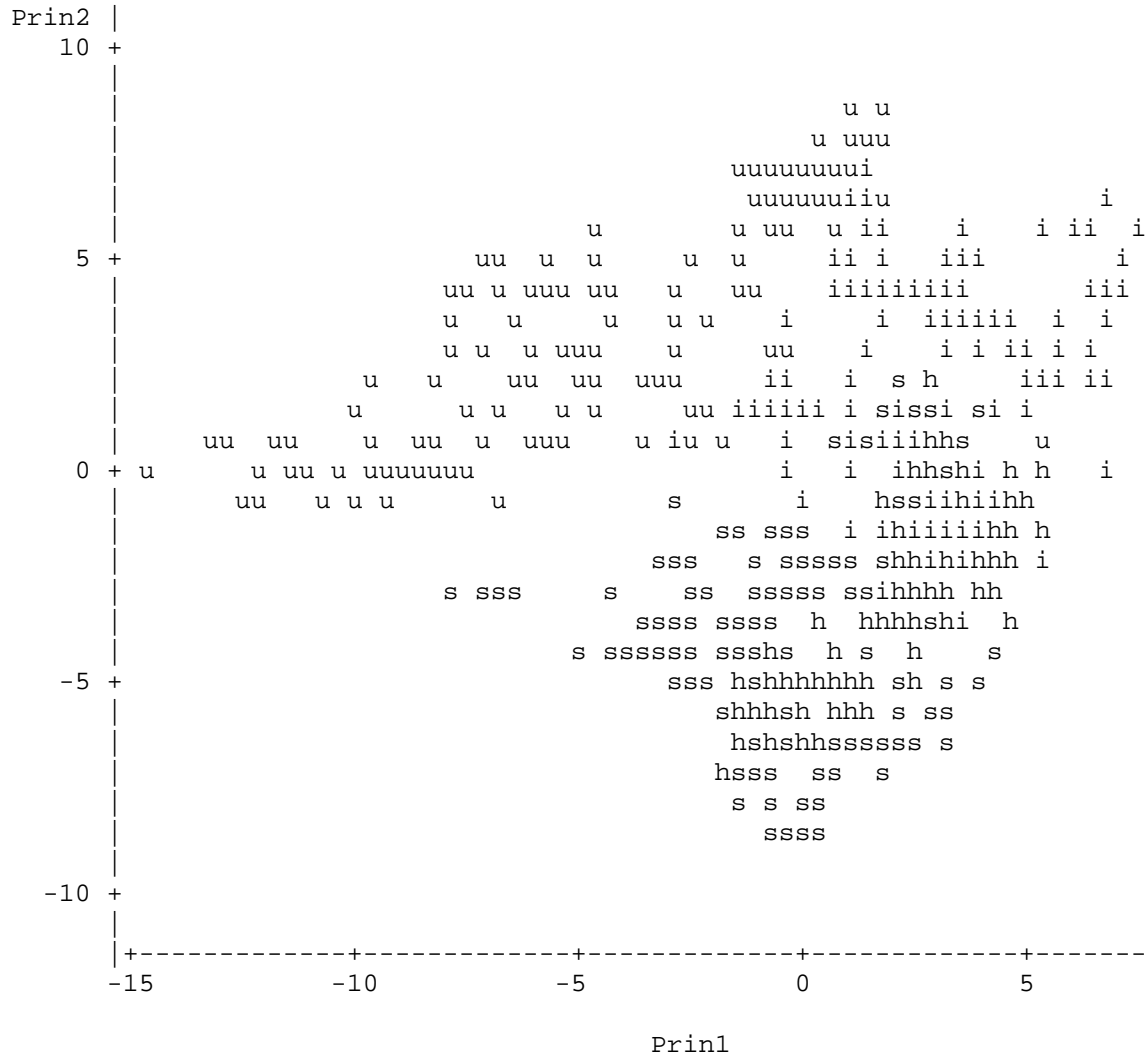


*Figure 5.* Principal Components Analysis of Facial Measures by Subject. The symbols in the plot represent the speaker numbers. All sounds are combined.





*Figure 7.* Principal Components Analysis of Palatometric Data by Subject. The symbols in the plot represent the speaker numbers. All sounds are combined.



*Figure 8.* Principal Components Analysis of Palatometric Data by Sound. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /ʃ/.

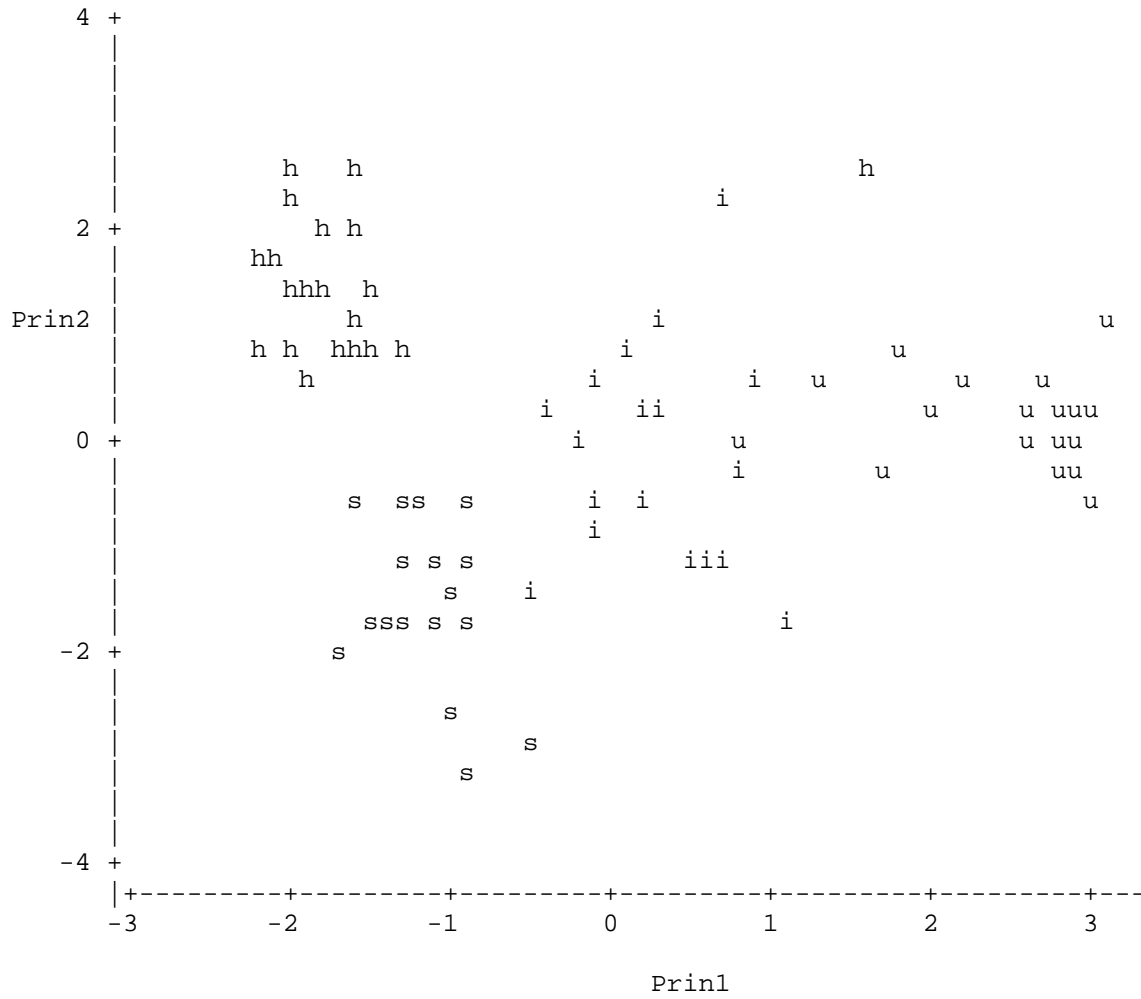
facial measures. This suggests that the specific sound which the speaker is producing can be determined by the palatometer data.

Two plots were also created which combined the facial measures and the palatometer data. One plot displayed these data by sound, the other by speaker. This was done in order to explore the effect of combining the data. The plots produced were very similar to the palatometer plots of sound and speaker. The participants did not clearly separate while the sounds did. This did not add more information than was established by the previous plots. The palatometer data are highly influential in this plot because there are 118 points versus only 6 points for the facial measures. Adding the facial measures to the palatometer data did not give us more information because of the weight of the palatometer data.

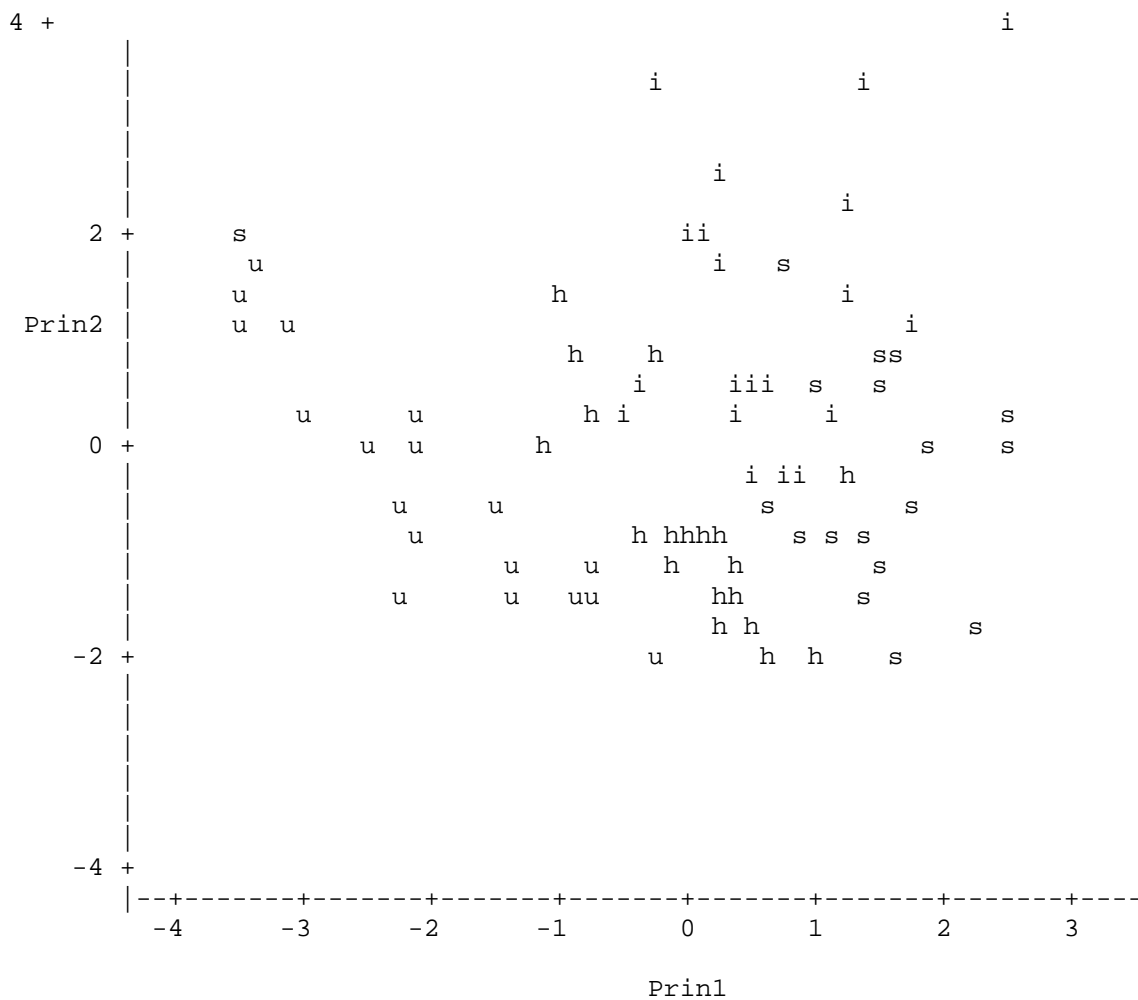
Figure 9 is a plot of the facial measures of sound across principal components one and two for speaker 0 only. This plot shows that the sounds produced by speaker 0 separate out agreeably. This suggests that this speaker's facial data can determine which sound is being produced.

Figures 10 through 16 show the plots of the facial measures of sound for speakers 1, 2, and 5-9 respectively. The plots for these speakers are quite different than the corresponding plot for speaker 0. The sounds do not separate out well, but are overlapping. This suggests that the facial measures cannot determine which sound is being produced by these speakers as readily as they can with speaker 0.

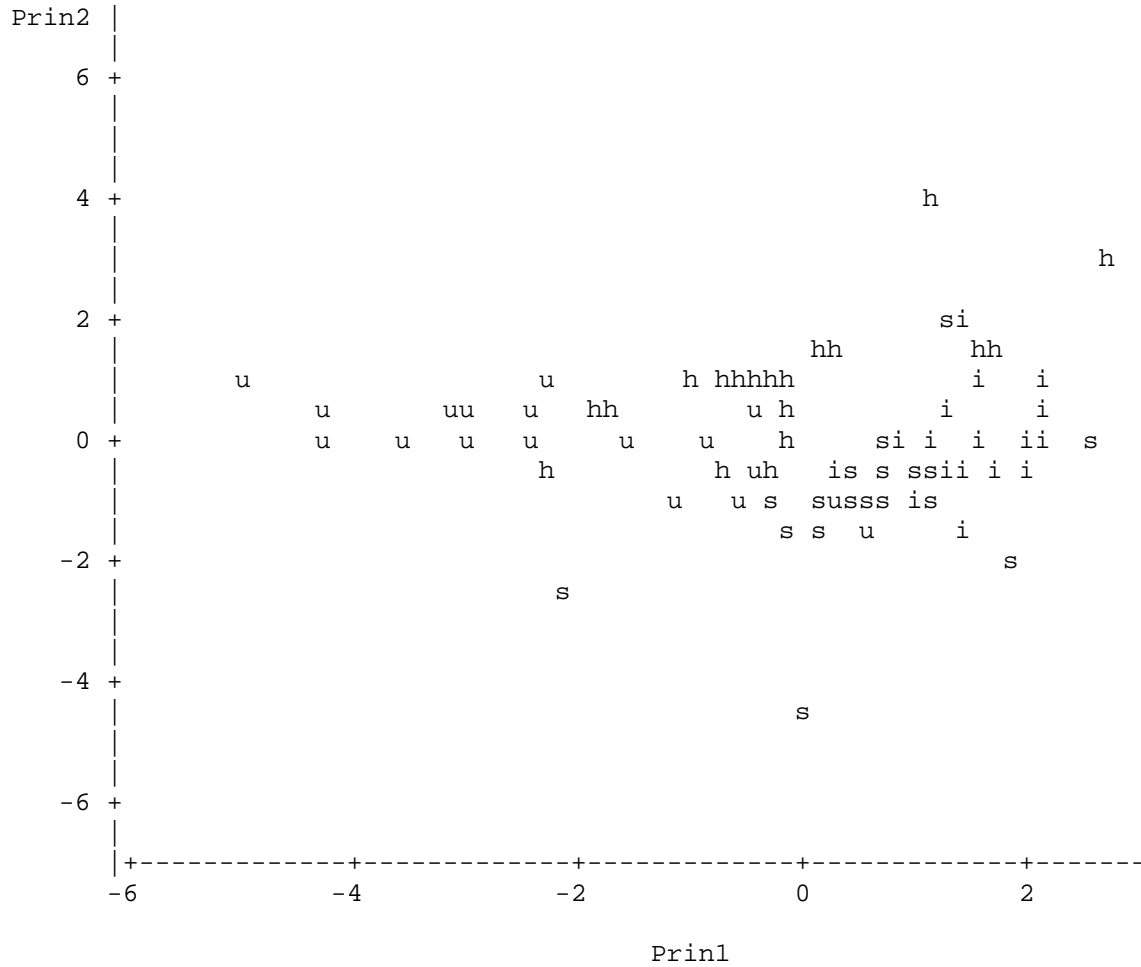
Figures 17-24 are plots for speakers 0-2 and 5-9 respectively. These figures show plots of the palatometer data by sound. The sounds separate distinctly and are not overlapping. This is the case for each speaker. This suggests that the palatometer data can



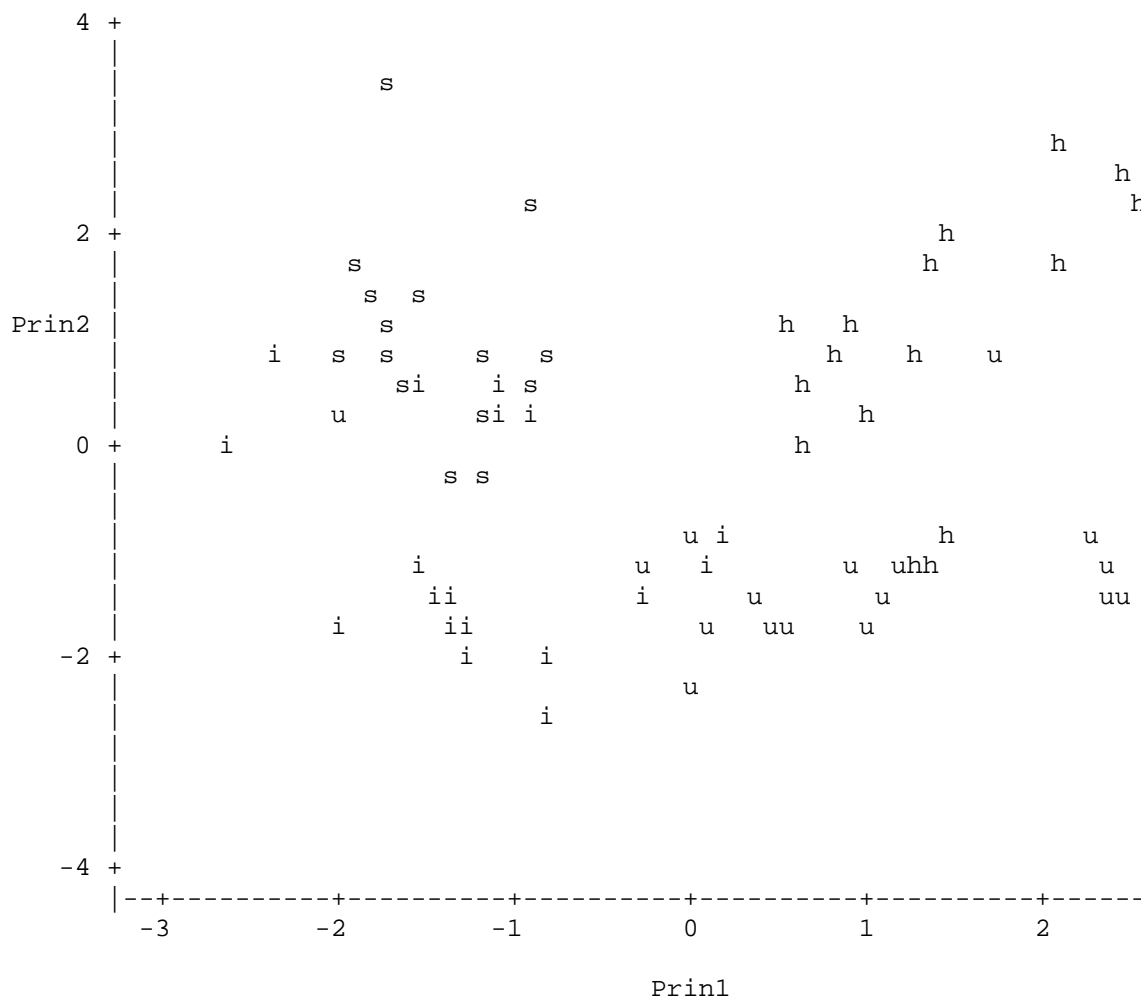
*Figure 9.* Principal Components Analysis of Facial Measures by Sound for Subject 0. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



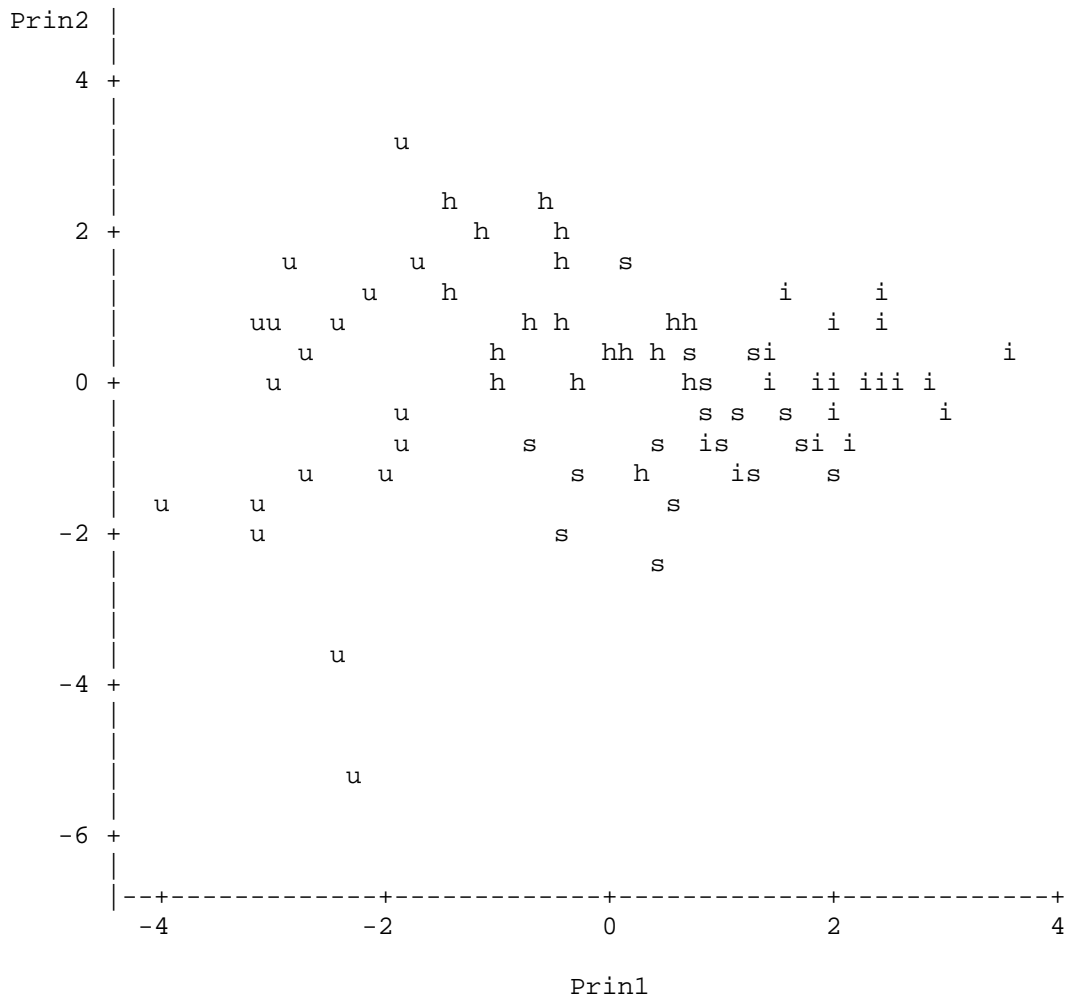
*Figure 10.* Principal Components Analysis of Facial Measures by Sound for Subject 1. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



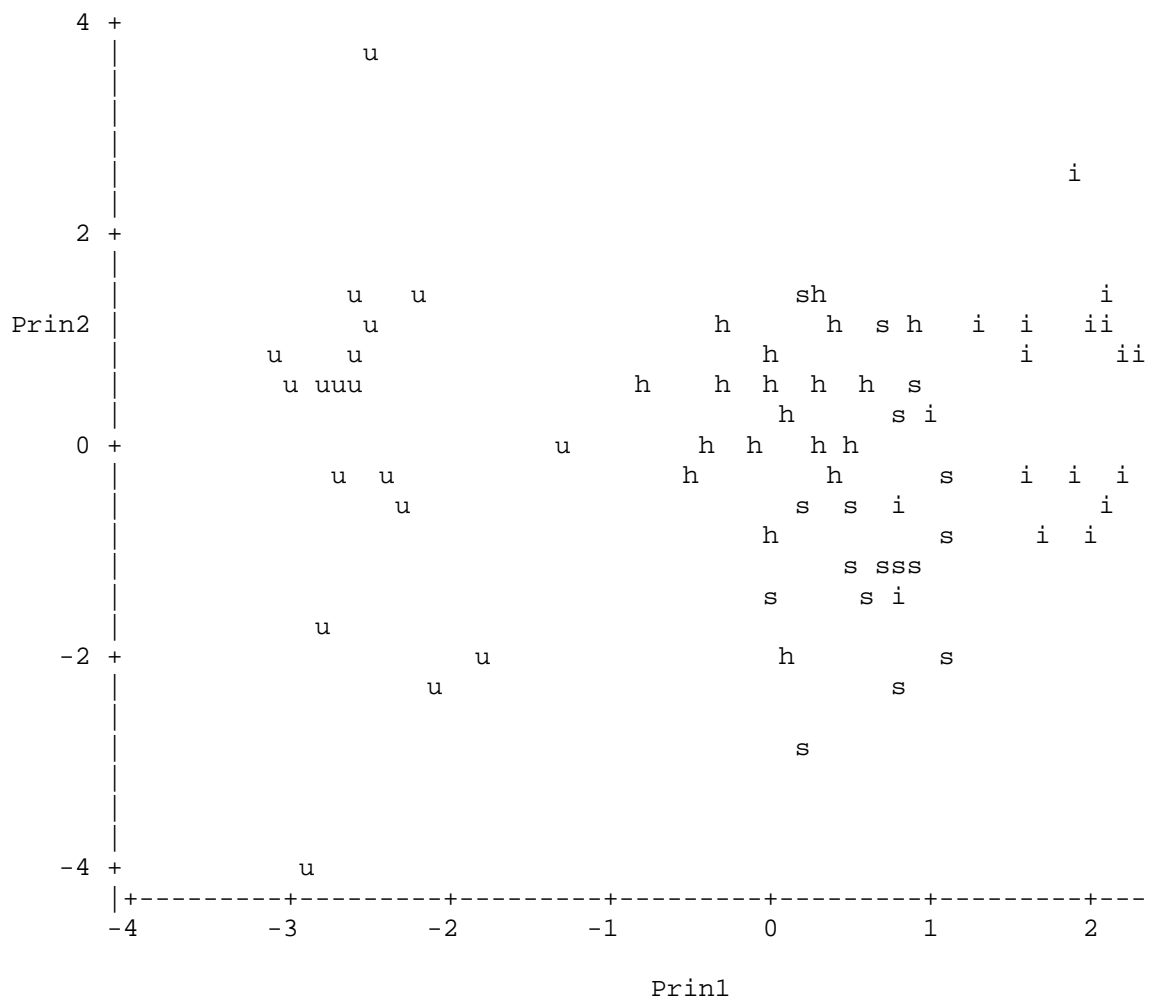
*Figure 11.* Principal Components Analysis of Facial Measures by Sound for Subject 2. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



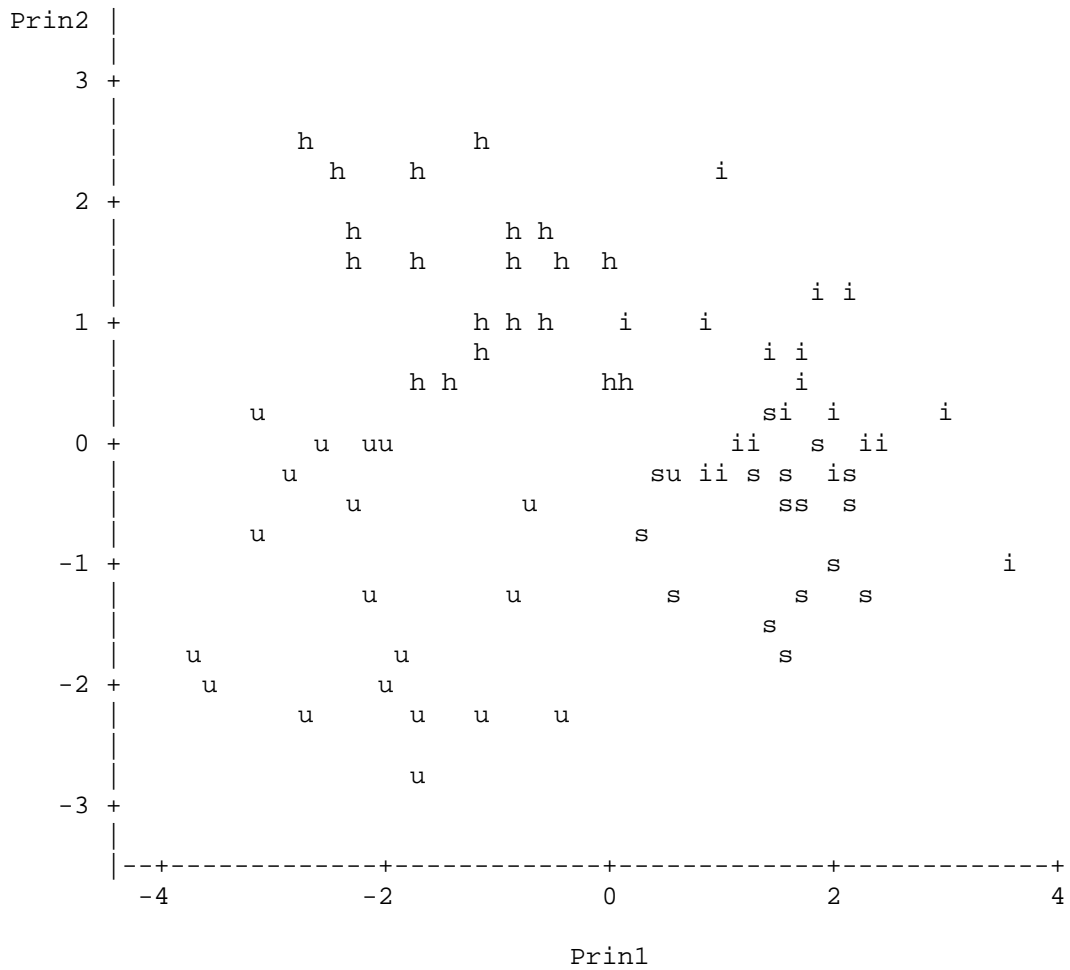
*Figure 12.* Principal Components Analysis of Facial Measures by Sound for Subject 5. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



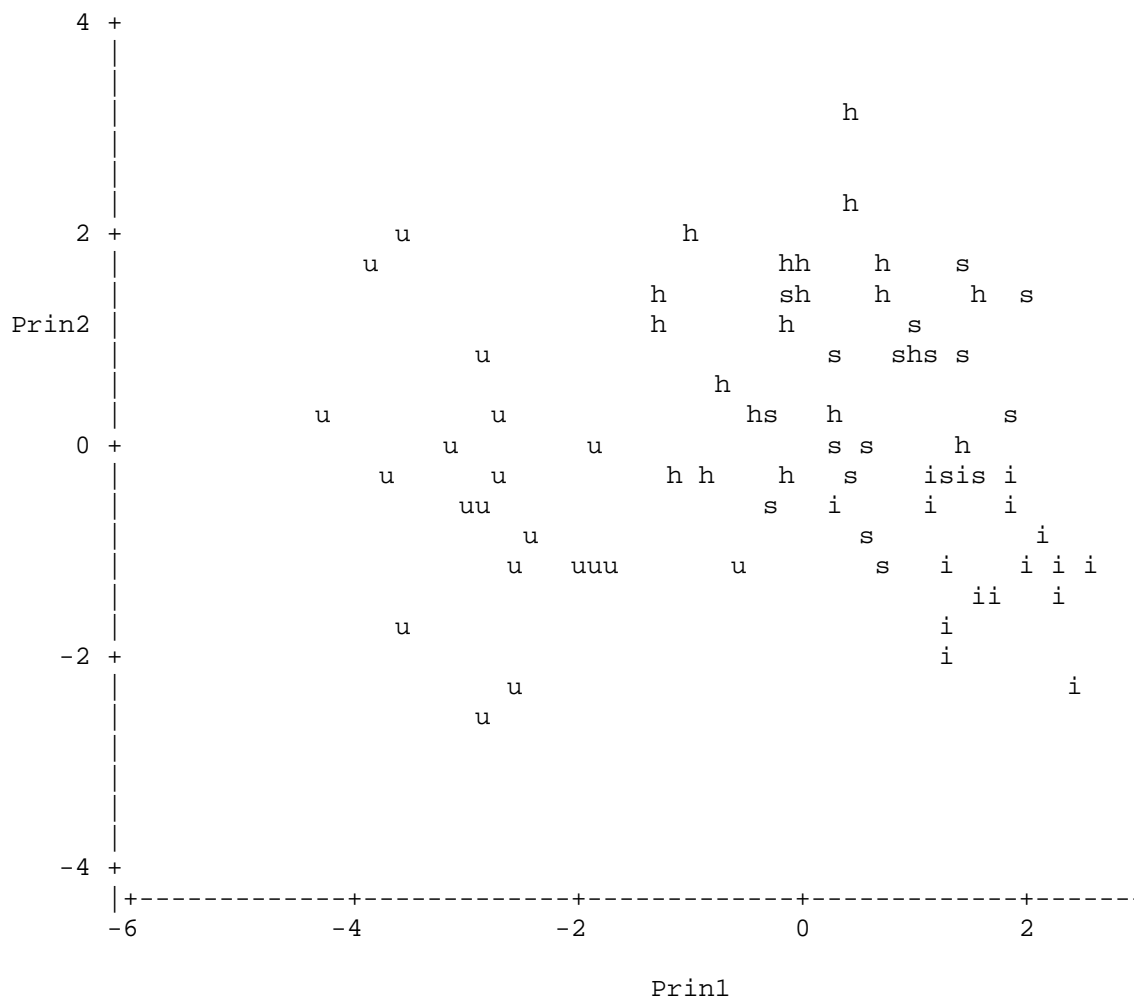
*Figure 13.* Principal Components Analysis of Facial Measures by Sound for Subject 6. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



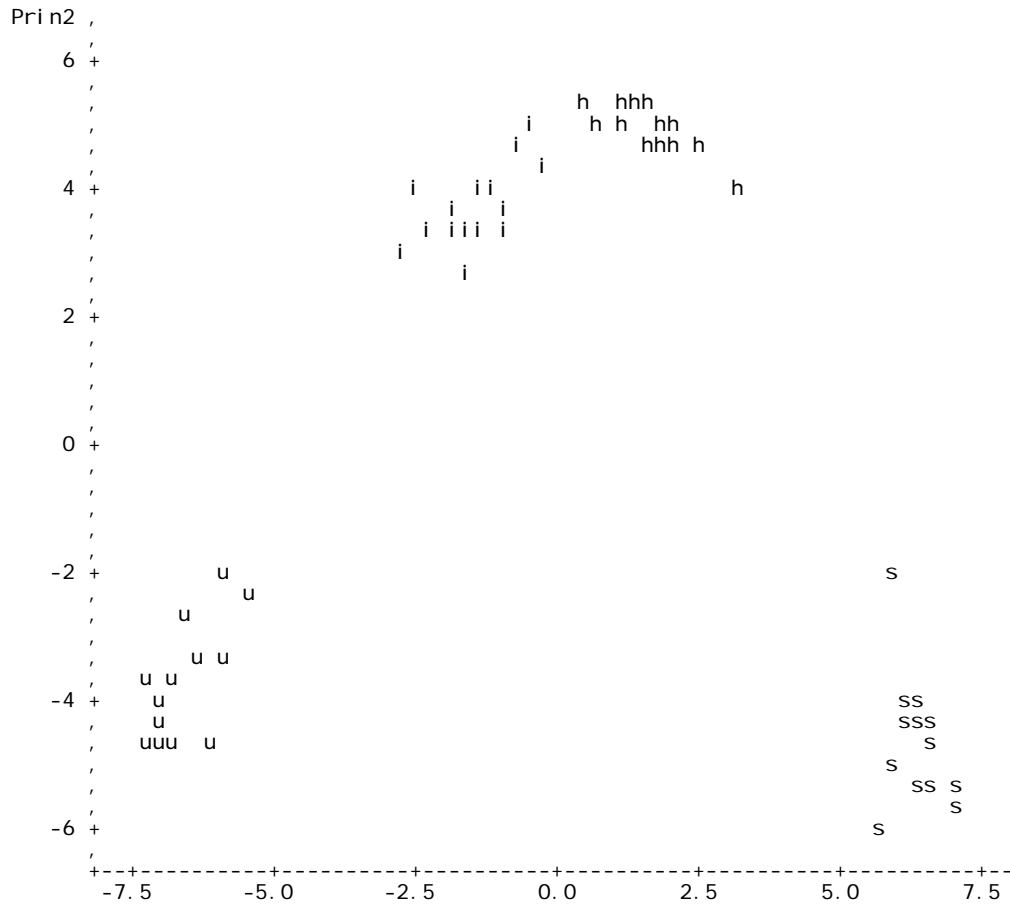
*Figure 14.* Principal Components Analysis of Facial Measures by Sound for Subject 7. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



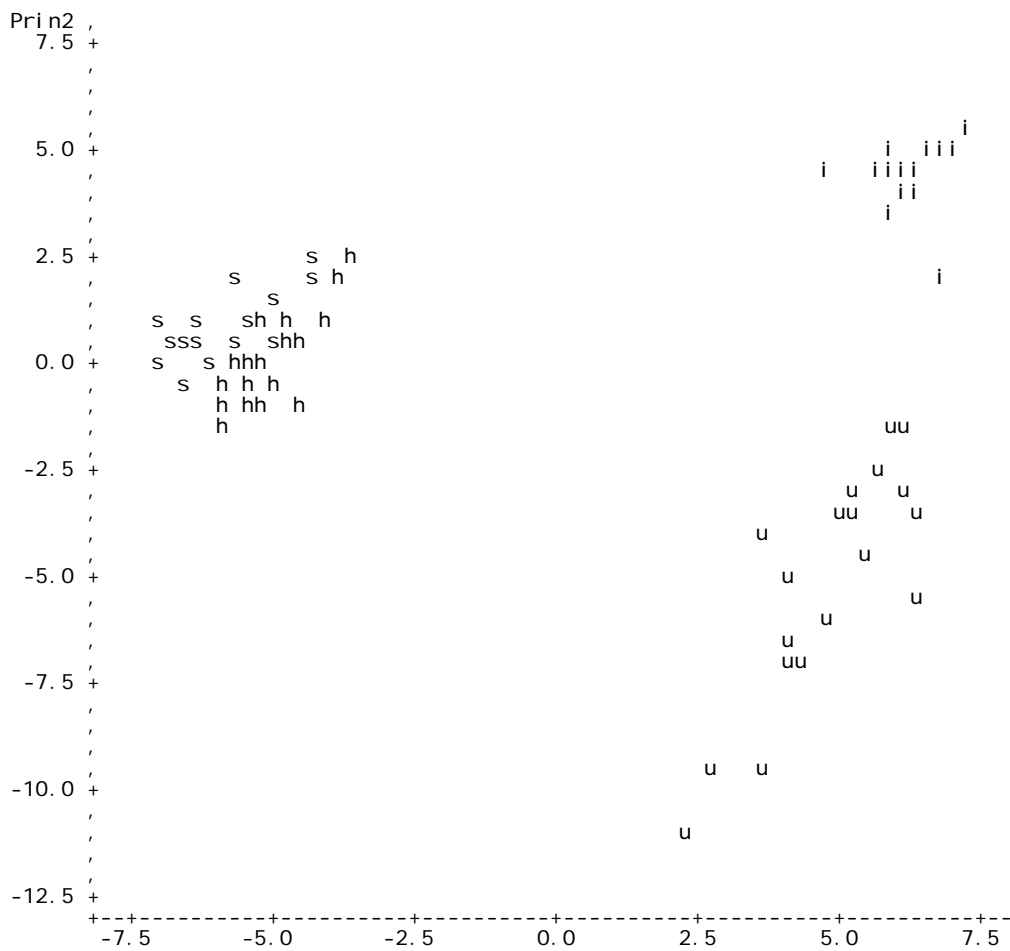
*Figure 15.* Principal Components Analysis of Facial Measures by Sound for Subject 8. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



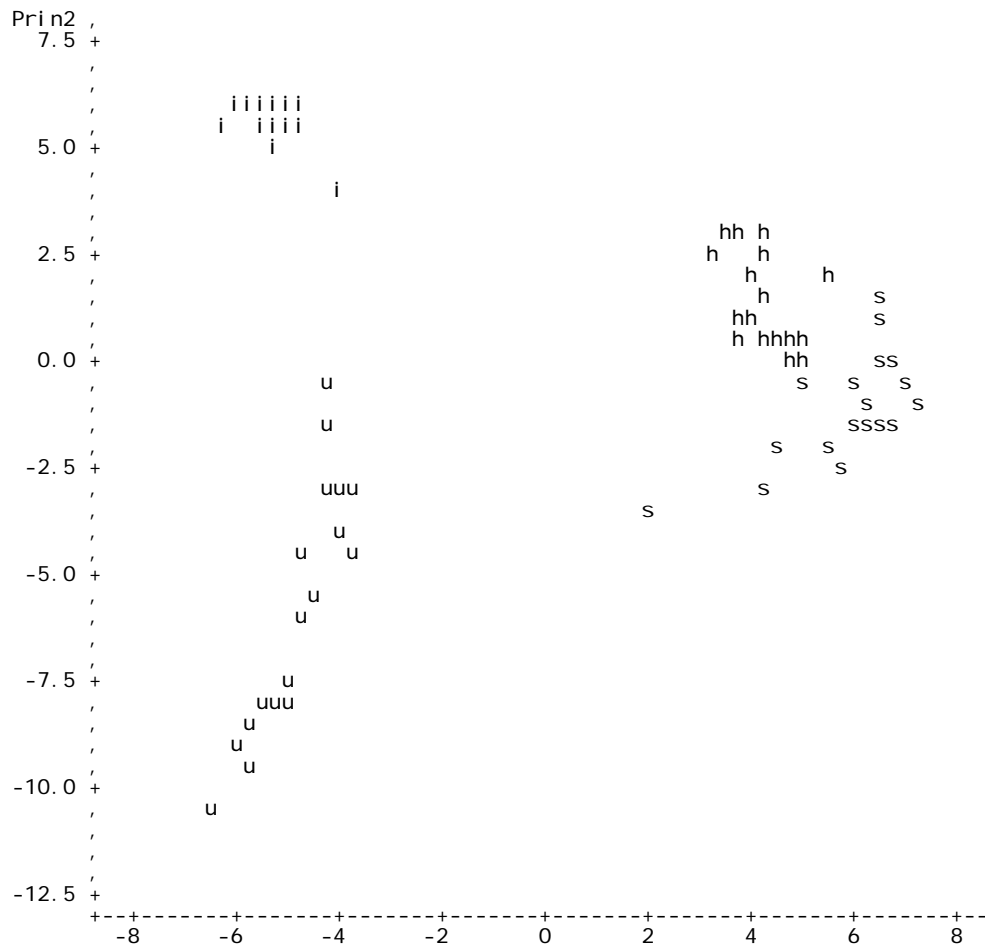
*Figure 16.* Principal Components Analysis of Facial Measures by Sound for Subject 9. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



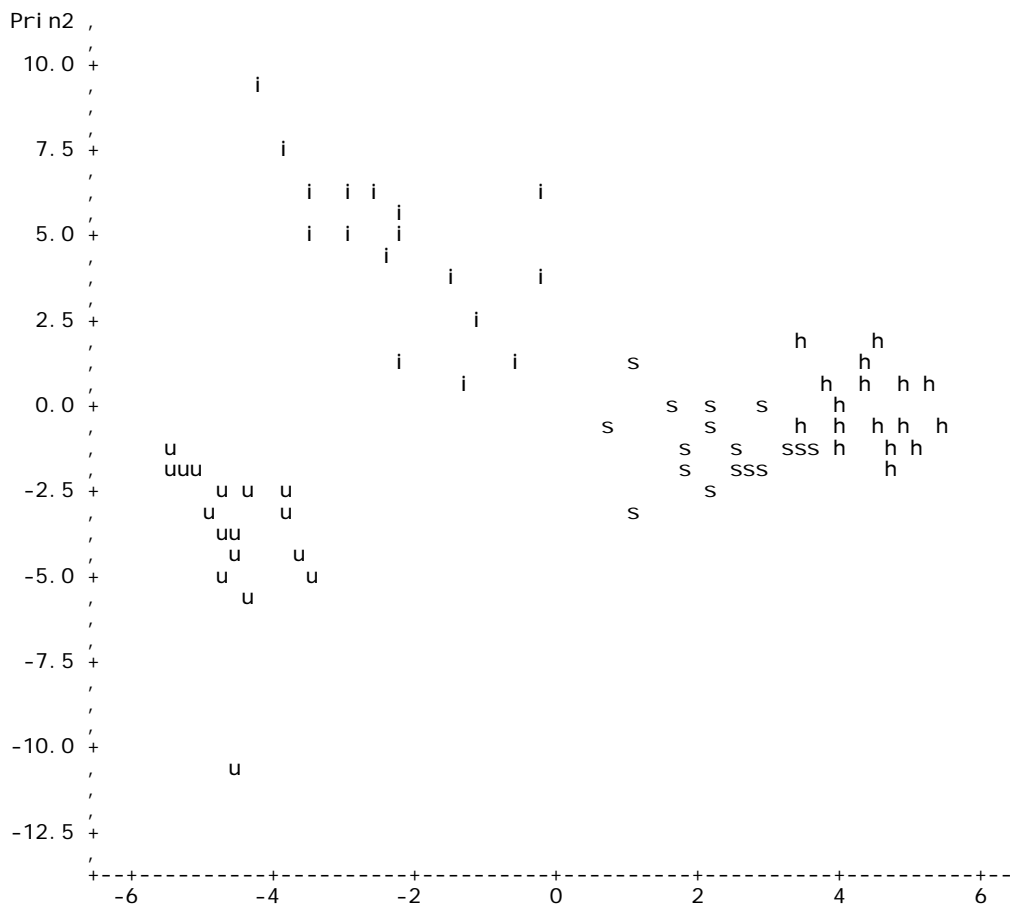
*Figure 17.* Principal Components Analysis of Palatometric Data by Sound for Subject 0. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



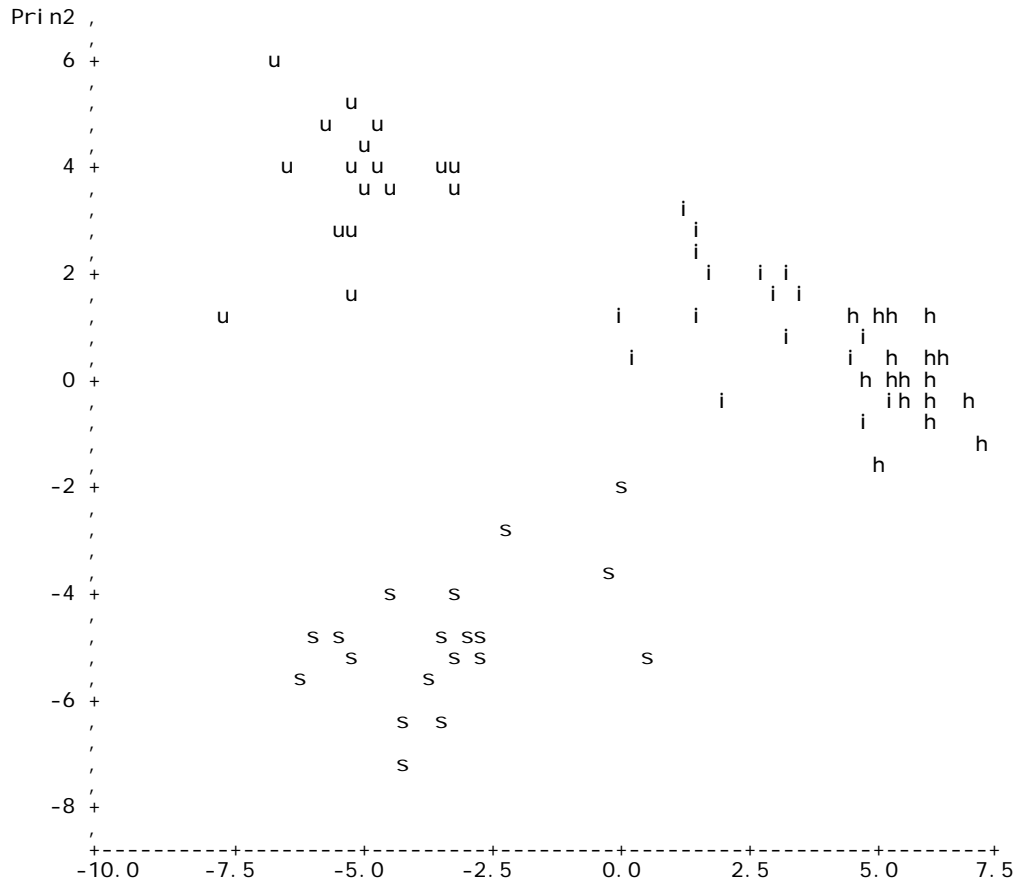
*Figure 18.* Principal Components Analysis of Palatometric Data by Sound for Subject 1. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



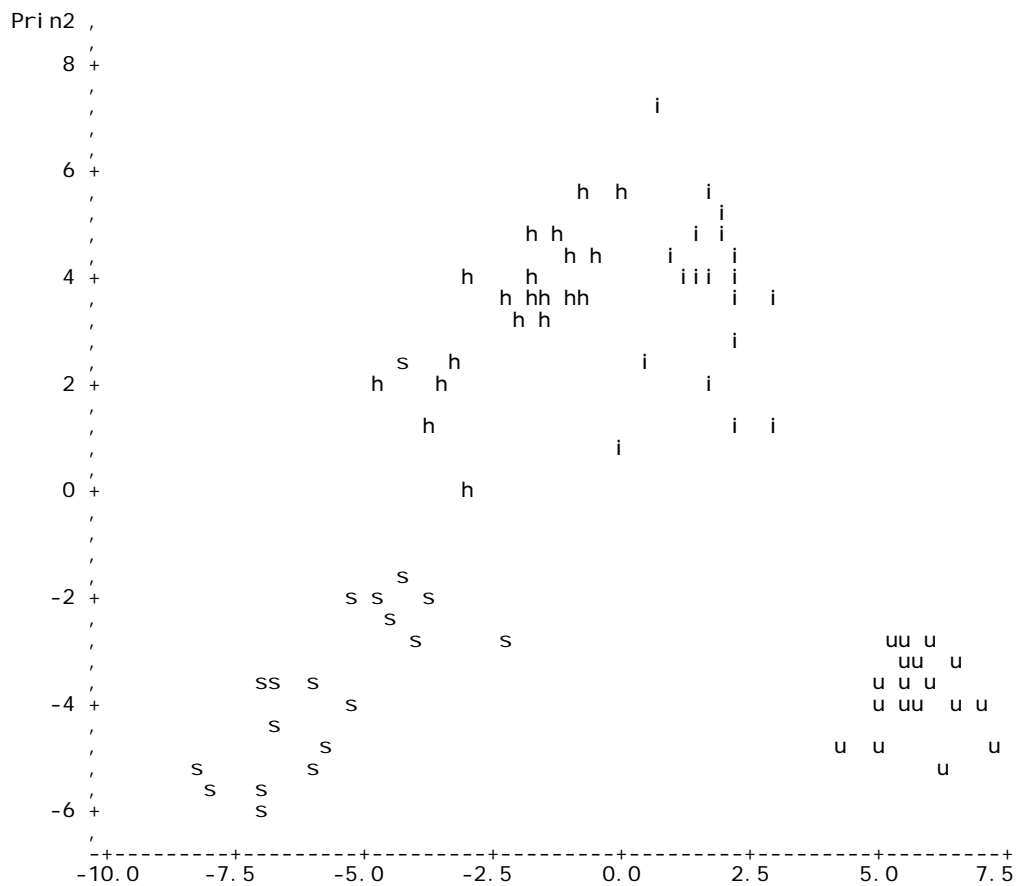
*Figure 19.* Principal Components Analysis of Palatometric Data by Sound for Subject 2. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



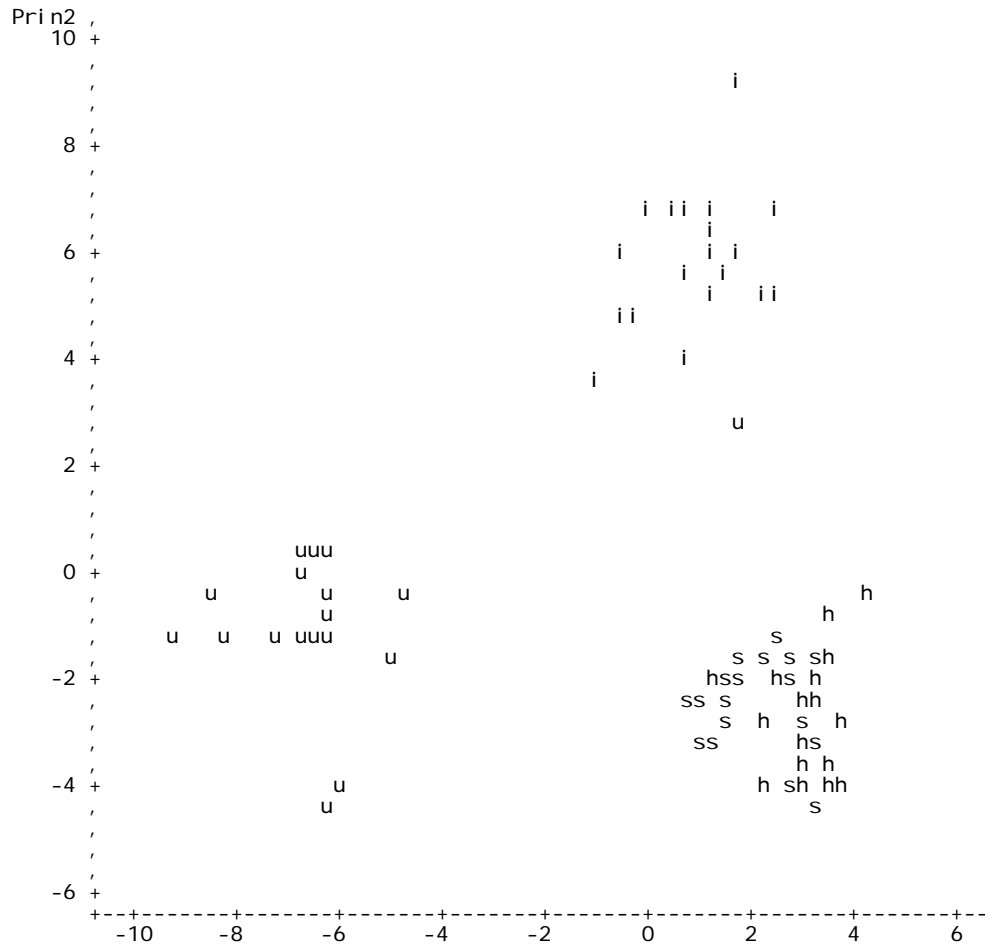
*Figure 20.* Principal Components Analysis of Palatometric Data by Sound for Subject 5. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



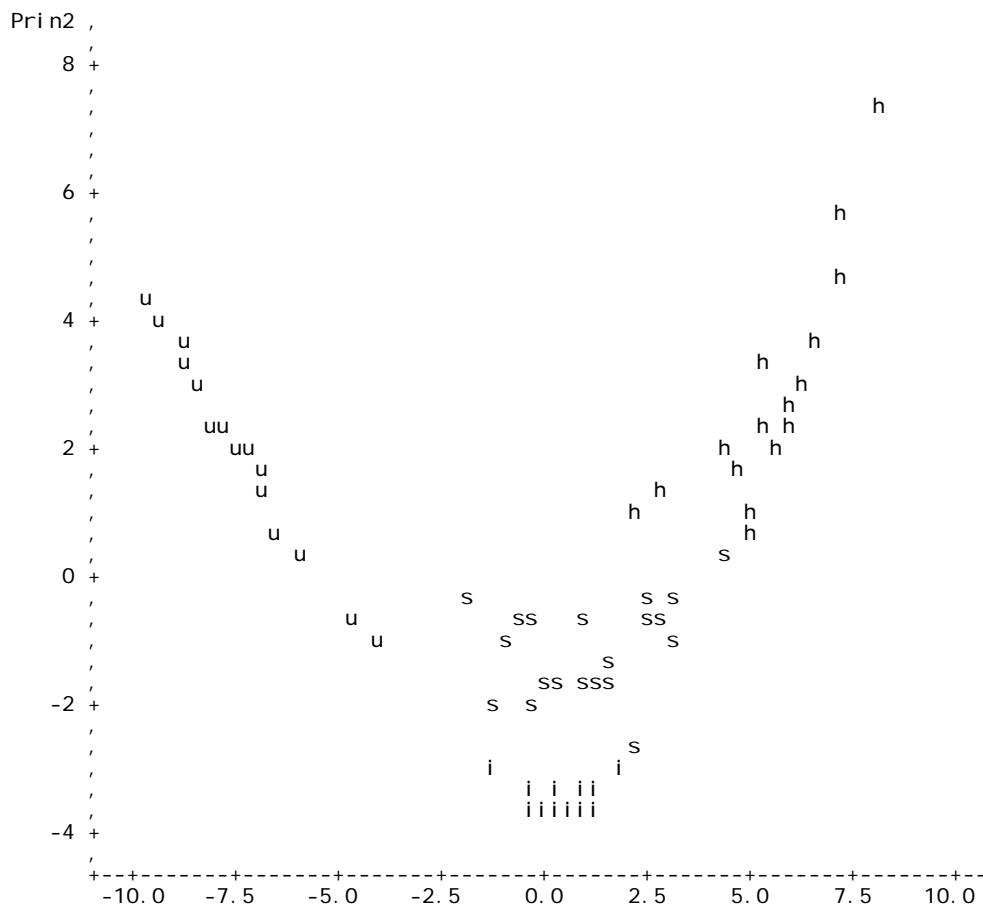
*Figure 21.* Principal Components Analysis of Palatometric Data by Sound for Subject 6. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



*Figure 22.* Principal Components Analysis of Palatometric Data by Sound for Subject 7. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



*Figure 23.* Principal Components Analysis of Palatometric Data by Sound for Subject 8. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /f/.



*Figure 24.* Principal Components Analysis of Palatometric Data by Sound for Subject 9. The symbols in the plot represent the sounds produced. The symbol “h” represents the sound /h/.

determine which sound is being produced by each individual speaker.

With the goal of finding more information about the facial measures, ratios were created for each combination of the six parameters across all speakers. For example, ratios were created for height/lower protrusion, width/upper angle, lower angle/upper protrusion, and so on. A correlation matrix and a principal components analysis were run for these ratios. The results were very similar to the analyses run with the raw data. The plots created showed that the sounds were not better separated by using the ratios variables instead of the raw variables. These analyses were not successful in adding more information about the facial measures data.

Another attempt to gain more information about the facial measures data was made by selecting the data from four participants. The four participants that were chosen had the best separation of sounds for the facial measures data. Speakers 0, 1, 5, and 7 were selected. Ratios were created from the data of these individuals and a principal components analysis was completed. Plots were also created from these data as was done with all the speakers previously. The results of these analyses did not provide more or different information. The plots and outputs were similar to the results completed including all the speakers. The sounds still did not separate out well; there was overlap between the sounds. This again suggests that the facial measures are not good predictors of which sound is being produced by the speaker.

## Discussion

The principal components analyses of the facial measures and the palatometric data allowed the results to be plotted visually in the figures referenced above. These visual representations showed that the sounds could be separated by the palatometric data. This means that it could be concluded which sound was being produced based on the palatometric data alone. This was the case for all speakers included in the study. The analyses also showed that sounds could not be as distinctly separated by the data from the facial measures. It could not be concluded which sound was being produced based on the facial measures alone. Furthermore, the analyses indicated that the data from lip measures were strongly influenced by who the speaker was, whereas the palatometric data were not. Thus, generalizations about sound production can be made from the palatometric data in a way that is not possible from the lip parameters.

The data from the lip measures were influenced by the speaker for all participants with the exception of speaker 0. Speaker 0's sounds separated out well from the facial measures as well as the palatometric data. It was possible to conclude which sound was being produced based on the lip measures alone. This shows the idiosyncratic nature of the way sounds are produced. It could be speculated that speaker 0 articulated phonemes more conspicuously in daily communication compared to the other speakers. It is possible that speaker 0, being the only male, displayed a wider range of motion of the articulators compared to the other speakers.

The discrepancies found among the speakers suggest that people vary in the way they produce phonemes with the visible parts of the vocal tract. This finding is consistent with the speechreading literature. Yakel, Rosenblum, and Fortier (2000) stated that

“speakers vary widely in their visible speech movements, which bears on how difficult they are to speechread (p. 1406).” For the present study, it was reasoned that the lip shapes of the phonemes /u, i, s/ and /f/ would look very different. For example /u/ is produced with lip rounding whereas /i/ is produced with lip retraction. These sounds should be very distinct from one another by looking at the lip shape. However, the present study shows that people vary in the way they produce these phonemes; an /i/ is not always produced with full lip retraction and an /u/ is not always produced with full lip rounding. The shapes did not appear to be as distinct for all speakers as was expected. The concept of motor equivalence helps in explaining this phenomenon. Motor equivalence refers to the use of different movement patterns to perform the same task under different conditions. For example, a word may normally be articulated with a contribution from the jaw, or it can be produced while clenching a pencil between the teeth, which fixes the position of the mandible. The lip and tongue movements required to produce the same sounds are very different in these two cases. Speakers can produce an /i/ with very different lip shapes and a wide range of lip retraction. Therefore, the lips have many degrees of freedom in the different shapes they can form while still producing the same sound. The lip measures used in this study may have been too simplistic to capture these small gradations in shape. Future research could develop more precise measures to explore these nuances.

The palatometric data were good predictors of which sound was being produced. This suggests that there may not be much variation in the way people produce phonemes with the tongue. The tongue contact patterns for the four sounds examined were generally consistent across all speakers. The consistency of the patterns may be explained

by the nature of the sounds used. The phonemes /u/ and /i/ are both vowels which are produced in a very different way from the fricatives /s/ and /ʃ/. By definition, fricatives are produced by forcing air through a narrow channel made by placing two structures close together. To produce /ʃ/, air flow is directed through a groove in the tongue which is positioned between the alveolar ridge and the palate. To produce /s/, the air flow is also directed through a groove in the tongue which is positioned against the alveolar ridge. The production of these two sounds is very similar. A small change in place of articulation could change an /ʃ/ to an /s/; therefore these sounds allow for fewer degrees of freedom in their production. While the lips have more room for variation in their placement in the production of a given sound, the internal vocal tract does not.

The internal vocal tract is also restricted for the vowels /i/ and /u/. Vowels are defined by a number of parameters. Among the most significant are lingual height and the front/back position of the vocal tract narrowing. The vowels /u/ and /i/ are produced with the tongue in a high position relative to the palate. If the tongue were to be lowered, a different vowel would be produced. The vowel /u/ is produced with the tongue positioned towards the back of the mouth, while the vowel /i/ is produced with the tongue further forward. The tongue also makes contact with the sides of the palate during the production of these vowels. The front or back position of these vowels is evident from the palatometric display. During the production of /u/, the sensors on the sides of the back portion of the palate are lit. During the production of /i/, these and also a group of more forward sensors are activated. Small changes in the lingual-palatal contact positions have a substantial impact on which vowel is produced. The sensitivity of the acoustic and perceptual output to the specific place of articulation explains why the tongue contact

patterns are so consistent for these sounds across all speakers.

Because this study was exploratory in nature and involved only a small sample size, there is a clear need for future research. The facial measures were not specific enough to separate the sounds across all speakers. The development and exploration of more precise or subtle facial measures would be beneficial to the prediction of sounds based on visible measures only. The present study only looked at a limited number of lip parameters: height, width, upper angle, lower angle, upper lip protrusion, and lower lip protrusion. More detailed measures could be explored as well as measures of other facial structures besides the lips. Kaplan (1987) explained that speechreading is the “ability to recognize the different sounds of speech by observing movements of the lips, tongue, and jaw (p. 1). This author also explained that “skilled speechreaders depend on much more than the information available from the movements of the lips, tongue, and jaw. They interpret facial expressions, gestures, and body language and use clues available from the situation and what they know about the language (Kaplan, 1987, p.1).” Future research could analyze the movements of facial expression, including movements of the head, eye brows, and forehead as well as eye gaze. These movements are all important in accurate speechreading. Skilled speechreaders also examine information about the *movements* of the articulators, not just their positions (Yakel et al., 2000). The current study focused on a few isolated static sounds by analyzing postures, and this line of research could be expanded by examining dynamic movements as well.

A larger sample size could be used to determine how much variance exists between speakers in the facial measures during the production of a sound. In the current study, both the facial measures and palatometric data from one speaker were predictive of

the sound being produced. More research is needed with a larger population to determine whether this individual was an outlier, especially since he was the only male in the study.

Future research could also focus on comparing instrumental data like those analyzed in this study to the observations made by speechreaders. It would be important to establish that the results from the objective analysis correlate with the subjective judgments about the readability of an individual's speech. Some research has been conducted to determine what makes a speaker easier to speechread. Lesner (1988) stated that desirable traits for a readable speaker include a slightly slower-than-normal rate of speaking, precise articulation, appropriate gestures, and inclusion of appropriate pauses. While all of these traits apply to conversational speech, using precise articulation would make the articulation of isolated speech more readable as well. This could explain why speaker 0 in this present study was more readable by facial measures alone compared to the other speakers: he might have displayed more precise articulation. Interestingly, other studies have shown that female talkers are speechread significantly more accurately than male talkers (Bench, Daly, Doyle, & Lind, 1995; Daly, Bench, & Chappell, 1996). There has also been research on the subjective traits that make a speaker more readable. In selecting talkers for a speechreading test, Bench et al. (1995) used a panel of hearing impaired judges to choose speakers who were "acceptable to speechreaders, by avoiding talkers who might be perceived as having negative personality characteristics or physical peculiarities" (p.174) and who were relatively easy to speechread based on subjective readability scores. In the future, research could include only speakers who were easily speechread to ensure more accurate objective measures and also to determine whether the objective measures correlate with the subjective measures of readability.

Another direction for future research could be the exploration of a better process to correlate or combine the facial measures with the palatometric data. The palatometric data consisted of 118 variables, whereas the facial measures consisted of six. It was difficult to correlate these sets of data together without the palatometric data carrying a heavier weight. Also, using more than two components in the principal components analysis may provide more subtle insights or details which were excluded by using only two components. More elaborate statistical approaches are needed to understand the influence of vocal tract activity on lip shape. Bringing these data together would show how the movements of the internal vocal tract interact with and affect the movements of the external vocal tract.

In summary, the current study examined visual facial information and vocal tract movement to try to understand more about the process of speech production. The results have shown that the visual information used here cannot predict the identity of phonemes in all speakers as well as the tongue contact patterns can. However, only static lip postures were explored, which represent a relatively small portion of the extensive array of visual information available in communication.

Understanding more about lip measures and vocal tract movement during speech production may potentially benefit the area of speechreading and help to explain why some people are more easily speechread than others. This knowledge would also be of substantial benefit to clinicians who work to assess and improve the ability of hearing-impaired individuals to understand speech based on visual information alone. The ability to speechread well is also important for individuals with normal hearing when they are communicating in a noisy environment, listening to speech with a heavy foreign accent,

or interpreting complicated subject matter.

In conclusion, the current study was a preliminary effort in the exploration of the internal and external movements of the vocal tract to understand more about the processes of speech production. While some headway was made in developing analytical measures of speech articulation, more research is needed to refine these procedures. Valuable experience has been gained through this study in the area of speech production, which will aid in unraveling the complexities of speech production as this line of research is continued.

## References

- Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. *Journal of the Acoustical Society of America*, *102*, 2353-3359.
- Bench, J., Daly, N., Doyle, J., & Lind, C. (1995). Choosing talkers for the BKB/A Speechreading Test: A procedure with observations on talker age and gender. *British Journal of Audiology*, *29*, 172-187.
- Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (1998). What makes a good speech reader? First you have to find one. In R. Campbell, B. Dodd, & D. Burnham (Ed.), "Hearing by Eye II" (pp. 211-227). East Sussex, England: Psychology Press.
- Campbell, R., & De Haan, E. H. F. (1998). Repetition priming for face speech images: Speech-reading primes face identification. *British Journal of Psychology*, *89*, 309-323.
- Chen, T., & Rao, R. R. (1998). Audio-visual integration in multimodal communication. *Proceedings of the IEEE, USA*, *86*, 837-852.
- Christensen, J. M., Fletcher, S. G., & McCutcheon, M. J. (1992). Esophageal speaker articulation of /s,z/: A dynamic palatometric assessment. *Journal of Communication Disorders*, *25*, 65-76.
- Daly, N., Bench, J., & Chappell, H. (1996). Gender differences in speechreadability. *Journal of the Academy of Rehabilitative Audiology*, *29*, 27-40.
- Eisenberg, A. (2003, September 11). What's next: Beyond voice recognition to a computer that reads lips. *New York Times*, pp. G8.
- Fletcher, S. G., McCutcheon, M. J., & Wolf, M. B. (1975). Dynamic palatometry.

*Journal of Speech and Hearing Research*, 18, 812-819.

Forster, C., & Hardcastle, W. (1998). An electropalatographic (EPG) study of the speech of two stuttering subjects. *International Journal of Language and Communication Disorders*, 33, 358-363.

Hardcastle, J., & Gibbon, F. (1997). Electropalatography and its clinical applications. In M. J. Ball & C. Code (Eds.), *Instrumental Clinical Phonetics* (pp. 149-193). London: Whurr Publishers.

Kaplan, H. (1987). *Speechreading: A way to improve understanding* (2<sup>nd</sup> ed.). Washington, DC: Gallaudet University Press.

Kroos, C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Video-based face motion measurement. *Journal of Phonetics*, 30, 569-590.

Kuratate, T., Munhall, K. G., Rubin, P. E., Vatikiotis-Bateson, E., & Yehia, H. (1999). Audio-visual synthesis of talking faces from speech production correlates. *Proceedings of EuroSpeech, ESCA, K013*.

Kydd, W. L., & Belt, D. A. (1964). Continuous palatography. *Journal of Speech and Hearing Disorders*, 29, 489-492.

Le Goff, B., Guiard-Marigny, T., & Benoît, C. (1997). Analysis-synthesis and intelligibility of a talking face. In J. P. H van Santen, R. Sproat, J. P. Olive & J. Hirschberg (Eds.), *Progress in speech synthesis* (pp. 235-246). New York: Springer-Verlag.

Lee, D. J., Bates, D., Dromey, C., Xu, X., & Antani, S. (2003). An imaging system correlating lip shapes and tongue contact patterns for speech pathology research. The 16th IEEE International Symposium on Computer-Based Medical Systems,

New York, NY, June 26-27, 2003.

- Lesner, S. (1988). The talker. New reflections of speechreading [Special issue]. *Volta Review*, 90 (5), 89-98.
- Lucero, J. C., Maciel, S. T. R., Johns, D. A., & Munhall, K. G. (2005). Empirical modeling of human face kinematics during speech using motion clustering. *Journal of the Acoustical Society of America*, 118, 405-409.
- Lucero, J. C., & Munhall, K. G. (1999). A model of facial biomechanics for speech production. *Journal of the Acoustical Society of America*, 106, 2834-2842.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 265, 746-748.
- Montgomery, A. A., & Jackson, P. L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. *Journal of the Acoustical Society of America*, 73, 2134-2144.
- Munhall, K. G. (2001). Functional imaging during speech production. *Acta Psychologica*, 107, 95-117.
- Munhall, K. G., Löfqvist, A., & Kelso J. A. S. (1994). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America*, 95, 3605-3616.
- Munhall, K. G., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd, & D. Burnham (Ed.), *Hearing by eye II* (pp. 123-139). East Sussex, England: Psychology Press.
- Murdoch, B. E., Gardiner, F., & Theodoros, D. G. (2000). Electropalatographic assessment of articulatory dysfunction in multiple sclerosis: A case study. *Journal*

*of Medical Speech-Language Pathology*, 8, 359-364.

- Neely, K. K. (1956). Effect of visual factors on the intelligibility of speech. *Journal of the Acoustical Society of America*, 28, 1275-1277.
- Pitermann, M., & Munhall, K. G. (2001). An inverse dynamics approach to face animation. *Journal of the Acoustical Society of America*, 110, 1570-1580.
- Ramsay, J. O., Munhall, K. G., Gracco, V. L., & Ostry, D. J. (1996). Functional data analyses of lip motion. *Journal of the Acoustical Society of America*, 99, 3718-3727.
- Rosenblum, L. D., & Saldaña, H. M. (1998). Time-varying information for visual speech perception. In R. Campbell, B. Dodd, & D. Burnham (Ed.), *Hearing by eye II* (pp. 61-81). East Sussex, England: Psychology Press.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.
- Yakel, D. A., Rosenblum, L. D., & Fortier, M. A. (2000). Effects of talker variability on speechreading. *Perception & Psychophysics*, 62, 1405-1412.
- Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, 30, 555-568.
- Yehia, H. C., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26, 23-43.

## Appendix

## Informed Consent

## Research Participation Form

Participant: \_\_\_\_\_ Age: \_\_\_\_\_

You are invited to participate in a research study being conducted by Jessica Wagner, a graduate student in the Department of Audiology and Speech Language Pathology at Brigham Young University. The faculty director of this research is Christopher Dromey, Ph.D. Other student assistants may help with data collection. You have been invited to participate because you are a native speaker of English, with no history of speech, language or hearing disorders.

This research project is designed to help us learn more about the relationship between measures of lip shape and movements of the tongue. You will be asked to produce certain sounds while using two different types of equipment.

In the first session, which lasts about 15 minutes, you will be administered a routine hearing test to determine that your hearing is normal and that you are qualified for this study. You will also participate in a speech screening and an oral structure examination to ensure your oral mechanism is normal for the purposes of speech. During a second lab visit (lasting about half an hour at a dental lab in Provo), you will have a dental impression made of your upper teeth and hard palate, so that a custom palatometer can be constructed to fit your mouth. The palatometer records the contact your tongue makes with the palate during speech.

In the third, one-hour session, digital pictures will be taken of your lips while you are producing specific sounds. Then you will be asked to produce the same sounds while wearing the palatometer, which is similar to a dental retainer. Electronic sensors on its surface will record the contact your tongue makes with the roof of the mouth. This pseudopalate will be connected to a computer for later data analysis.

Your participation will thus involve three sessions. The first is about 15 minutes long, the second about 30 minutes, and the third about one hour. Please feel free to ask questions at any time before or during data collection.

Numerous studies have been conducted using this equipment, and there are no known risks involved. The equipment should not be uncomfortable or significantly influence your ability to speak during the research. There is the possibility that individuals with a hyperactive gag reflex will experience discomfort when the palatometer is first fitted. However, the device can be trimmed to reduce this discomfort.

Names of all speakers will be kept confidential, and data will be stored on the computer using speaker codes, rather than names. Participation in the study is a voluntary service and you are free to withdraw from the study at any time without any penalty. Your decision to participate or withdraw from the study at any time will have no influence on your grades, standing or relationship with BYU.

If you have any questions regarding this research project you may contact Dr. Christopher Dromey, 133 TLRB, Brigham Young University, Provo, Utah 84602; phone (801) 422-6461. If you have any questions regarding your rights as a participant in a research project you may contact Dr. Renea Beckstrand, Chair of the Institutional Review

Board, 422 SWKT, Brigham Young University, Provo, UT 84602; phone (801) 422-3873.

I agree to participate in this research study. I confirm that I have read the preceding information and that all of my questions have been answered. I hereby give my consent to participate.

---

Signature of Participant

---

Date