Identifying the Acoustic Onset for English Semivowels

by

Christine Yin-Ling Joe

B.A., The University of British Columbia, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Audiology and Speech Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2010

Abstract

Today's hearing aids are sophisticated devices that use complex signal processing to alter the acoustic signal. As digital circuit complexity and power efficiency evolve, even more advanced processing algorithms will be possible and will need to be evaluated. Most existing measures of hearing aid processing involve global acoustic (e.g., Articulation Index) or global behavioural (e.g., Hearing in Noise Test) analyses. Such measures have not been shown sensitive enough to detect local acoustic or behavioural changes to individual speech segments that result from complex processing algorithms. The purpose of this study was to provide information to help in the development of a standardized test that can be used for phoneme by phoneme acoustic analysis and speech recognition for the purpose of evaluating the effects of complex hearing aid processing. Such a test would require clear acoustic boundaries for the onset and offset of each phoneme, which to date, have not been determined for semivowel sounds. Using items from the University of Western Ontario Distinctive Features Differences (UWODFD) test, I evaluated the acoustic boundaries at which the English intervocalic semivowels were just perceived by Canadian English listeners. This study aimed to 1) establish the acoustic onset of semivowel identification within the UWODFD items, and 2) evaluate whether that point could be predicted by magnitude of spectral change, formant pattern, and/or formant transition duration. Eight listeners were presented time-sliced UWODFD test tokens and were asked to identify the sound out of a list of 21. A multivariate regression was performed to determine the amount of variance accounted for by each predictor variable. The acoustic boundary for phoneme recognition was determined for each semivowel, using an operational definition of 75% correct recognition. This study successfully established the acoustic boundary for each semivowel. Different combinations of acoustic variables were needed to predict the

recognition of different semivowel sounds, however formant ratios and transition duration consistently stood out to be important. No absolute ratio values or transition durations were found to identify the acoustic onsets, although a reduced range of ratio values was observed to separate perception and non-perception.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I am grateful for the NSERC Discovery Grant awarded to Dr. Lorienne Jenstad, that provided the funds to be able to carry out this research.

I am thankful for Dr. Lorienne Jenstad, who encouraged me to take-up this thesis project in the first place. Thank you for keeping me grounded yet encouraged through the many ups and downs in the process of completing this project.

Thank you to Dr. Valter Ciocca and Ms. Sharon Adelman, committee members, for your important advice on this project along the way.

Many thanks to my family, who have been understanding, encouraging, supportive, and kind.

Thanks to my friends for their encouraging words and prayers.

Dedications

To my family and friends

1 Introduction

Hearing aids have progressed dramatically in the last few years due the introduction of digital hearing aids in 1996 (Strom, 2006). Digital hearing aids have allowed the implementation of advanced signal processing and transformed hearing aids into devices that can employ complex algorithms to enhance speech, such as adaptive directional microphones, feedback cancellation, and adaptive noise suppression. As digital circuit complexity and power efficiency evolve, even more advanced processing algorithms will be possible and will need to be evaluated. Currently, effects of hearing aid processing on speech may be quantified with acoustic or behavioural measures, which include average measures of acoustic effects (e.g., Articulation Index and Speech Transmission Index) and overall speech recognition tests (e.g., Hearing in Noise Test, Speech in Noise Test, Synthetic Sentence Identification, and CUNY Nonsense Syllable Test). These global, or average, measures have not been shown sensitive enough to detect local, or small scale, changes to individual speech segments that result from complex processing algorithms, such as alterations to a phoneme's spectral or temporal composition, or phoneme-specific speech intelligibility (Boothroyd & Medwetsky, 1991; Cox, Alexander, Gilmore, & Pusakulich, 1988; Nilsson, Soli, Sullivan, 1994; Stelmachowicz, Kopun, Mace, Lewis, & Nittrouer, 1995; Turner & Robb, 1987). In addition, the relationship between acoustics and speech intelligibility is complex; some studies have shown that the same processing effect may be beneficial, detrimental, or negligible to speech intelligibility, depending on the phoneme being altered (Balakrishnan, Freyman, Chiang, Nerbonne, & Shea, 1996; Freyman, Nerbonne, & Cote, 1991; Jenstad & Souza, 2005). For those reasons, it is important to develop an evaluation measure that is capable of identifying the local changes to the speech

signal as a result of complex hearing aid processing and that is appropriate for both acoustic and behavioural measures.

Many types of advanced hearing aid algorithms are designed to enhance or reduce very short segments of speech. These local changes, defined by a timecourse of a few milliseconds, will manifest differently according to a phoneme's characteristics, such as inherent energy and spectral composition (Bor, Souza, & Wright, 2008; Hedrick & Rice, 2000; Jenstad & Souza, 2005). For example, compression is a processing strategy in which gain is automatically adjusted according to the intensity of the input level. One inherent feature of compression is release time, which is the time required for the processor to react to a decrease in input level. Shorter release times from compression result in a smoothing of the temporal envelope and a decrease in intensity difference between adjacent vowels and consonants. This difference in amplitude between adjacent vowels and consonants is called the "consonant to vowel ratio". The smoothing effect has been found to be greater for inputs with larger intensity differences, so voiceless consonants are more affected than voiced consonants, and stops, fricatives, and affricates are more affected than approximants and nasals (Jenstad & Souza, 2005; Souza, Jenstad, & Folino, 2005). The change in consonant-vowel ratio resulting from compression has also been shown to alter the spectral composition of stop release bursts; particularly, compression has been shown to increase high frequency energy within labial stop bursts more than alveolar stop bursts, which inherently have high frequency energy within the burst (Hedrick & Rice, 2000). Increasing channels of compression, a processing option in which the audible frequency range is split to allow compression processing to be applied independently to different frequency regions, results in a significant decrease in F1 and F2 spectral contrast (which is the difference between the formant's spectral peak and immediate adjacent trough) for high front,

high back, low back, low front vowels but not for mid front vowels /e, ɛ/ (Bor, Souza, & Wright, 2008).

The ability to detect local acoustic changes to the signal should therefore be an important feature in hearing aid evaluation tools. However, acoustic information on its own is an incomplete evaluation of hearing aid processing because it does not give information on how the acoustic alterations impact speech intelligibility (e.g., Valente, Hosford-Dunn, & Roeser, 2008; Van Tasell, 1993). Acoustic changes in the speech signal may have varying effects on performance. It has been shown that some large acoustic alterations to the speech signal may be neither beneficial nor detrimental to speech recognition (Jenstad & Souza, 2005). In addition, intended acoustic alterations to the signal, such as increased consonant-vowel ratio, may lead to decreased recognition for one sound but increased recognition in another (Balakrishnan, Freyman, Chiang, Nerbonne, & Shea, 1996; Freyman, Nerbonne, & Cote, 1991). The complex relationship between acoustic alterations and speech intelligibility is likely due to the fact that different acoustic cues are important for the perception of different phonemes (e.g., Dubno & Levitt, 1981). Therefore, in order to properly evaluate the effects of processing algorithms, it is necessary to be able to relate acoustic alterations to performance on speech recognition tasks.

The Articulation Index (AI) and Speech Transmission Index (STI) are methods based on the long-term average speech spectrum (LTASS) that predict speech recognition based on the proportion of each frequency band audible to the listener. In both methods, frequency bands are weighted according to their importance for speech intelligibility, and the sum of the weighted bands corresponds to predicted speech intelligibility. If the sum of those weighted bands is 0, the signal is not predicted to be intelligible, but if the sum is 1.0, speech is predicted to be maximally intelligible (Amlani, Punch, & Ching, 2002). Though the AI and STI relate acoustics to speech

perception, these measures are limited because they are global measures (Stelmachowicz, Kopun, Mace, Lewis, & Nittrouer, 1995); that is, the acoustic measurements in AI and STI are based on the speech signal's long-term average gain per one-third octave frequency band, over approximately a 64 second time period, and predicted speech intelligibility is an average score based on intelligibility of overall stimuli lists, such as sentence tests, monosyllabic tests, closed set tests, and nonsense syllable tests (Amlani, Punch, & Ching, 2002). These long-term averages of acoustics and behaviour do not provide phoneme-specific information, and have been reported to be not suitable to detect the local acoustic and intelligibility changes that result from complex processing (Dubno, Dirks, & Schaefer, 1989; Stelmachowicz, Kopun, Mace, Lewis, & Nittrouer, 1995).

Previous studies have looked at phoneme by phoneme error analyses, showing that it is possible to relate errors in phoneme perception to acoustics (e.g., Dubno & Levitt, 1981; Krull, 1990; Miller & Nicely, 1955). For example, Dubno and Levitt (1981) conducted a phoneme by phoneme analysis and found that specific acoustic parameters may explain consonant confusions among some syllables, but no one set of acoustic parameters could predict confusions among all syllables. Some studies have shown that this type of analysis is possible and informative for quantifying hearing aid processing (e.g., Bor, Souza, & Wright, 2008; Davies-Venn, Souza, Brennan, & Stecker, 2009; Jenstad & Souza, 2005); however, to date, this detailed analysis is done neither often nor routinely.

The purpose of this study is to help develop a standardized test that is a phoneme by phoneme analysis of acoustics and behaviour, and that can be used routinely to evaluate the effects of complex hearing aid processing. The speech recognition test chosen for this evaluation tool is the University of Western Ontario Distinctive Features Difference test (UWODFD),

which has already been shown to be a successful measure of acoustics and behaviour, though not necessarily in the same study together (e.g., Jamieson, Brennan, & Cornelisse, 1995; Jenstad, Barnes, Hayes, 2008; Jenstad, Seewald, Cornelisse, & Shantz, 1999). The UWODFD test is a phoneme-based speech test consisting of two-syllable nonsense sounds with the target phoneme placed intervocalically (e.g., aDil, aBil, aJil), with tokens from multiple talkers (Cheesman & Jamieson, 1996).

The ideal speech stimulus for this purpose would be one in which phoneme identification is generalizable across talkers, is based solely on the acoustic information available in the signal rather than context, and is representative of information extracted from running speech (Van Tasell, 1993). The UWODFD stimuli address these issues. First, the UWODFD stimuli were spoken by 4 talkers, two females and two males, making results more generalizable compared to speech tests using stimuli spoken by one speaker. Second, unlike real word speech tests, which have been shown to improve speech recognition scores due to lexicon (word frequency) and semantic context effects (Boothroyd & Nittrouer, 1988), the UWODFD stimuli are nonsense words and are not subject to such contextual confounds. Third, although shorter stimuli can restrict the acoustic content to information solely available for the phoneme of interest, speech tests consisting of short stimuli may introduce experimental error due to listener opportunities to learn artifactual aspects of the signal (e.g. idiosyncrasies in target token related to talker or processing artifacts) (Van Tasell & Trine, 1996). The UWODFD words are two syllables, which may decrease the listener's ability to learn artifactual information for identification compared to one syllable sounds. Finally, intervocalic consonants have been shown to be much more common in conversational speech than initial consonants (Pickett, Bunnell, & Revoile, 1995). Because the UWODFD target sound occurs intervocalically it serves to approximate the

contextual acoustic cues available in "running speech."

In addition, for this purpose, acoustic measures should correspond to behavioural measures. Within the speech stimuli used, it is necessary that each phoneme's acoustic boundaries (onsets and offsets) be established and based on listeners' perceived onset and offset. This is important because any changes observed in the waveform within the acoustic boundaries could then be related to behaviour, and vice versa, so that a direct comparison between acoustics and behaviour can be made (e.g., Dubno & Levitt, 1981). Within the UWODFD speech set, the initial perception of the stops, affricates, fricatives, and nasals can be reliably recognized by patterns in the waveform based on many previous studies (e.g., Jenstad & Souza, 2005; Kennedy, Levitt, Neurman, & Weiss, 1998; Smits, 2000), and these patterns have already been used for segmentation within the waveform (Jenstad, Barnes, & Hayes, 2008). On the other hand, no reliable acoustic markers have been identified to signal the initial perception of the four English semivowels /w, j, r, l/, therefore no such boundaries have been established for the UWODFD semivowels.

In light of this, my paper proposes a study to evaluate the temporal boundary at which the English intervocalic spoken semivowels from the UWODFD test are consistently perceived by Canadian English listeners. For sounds like semivowels, that do not have clear acoustic markers to signal their perceptual onsets, auditory analysis has been used for their segmentation in the waveform (Balakrishnan, Freyman, Chiang, Nerbonne, & Shea, 1996; Kennedy, Levitt, Neuman, & Weiss, 1998; Jenstad & Souza, 2005). Auditory analysis is the subjective parsing of sound segments using listeners' auditory perception of their onset and offset. I aim to use objective behavioural measures to determine the perceptual boundary associated with UWODFD semivowel perception, and acoustic analysis to determine whether acoustic predictors may be

useful to signal the onset of perception. I aim to provide information that will ultimately be used to combine both acoustic and behavioural results obtained with stimuli from the UWODFD test, as part of the development of a new standard hearing aid evaluation tool sensitive to local acoustic changes resulting from complex processing,

The Problem of Segmenting Semivowels

Speech is not a string of discrete sound segments and vocal tract movements. Rather, during speech, sounds are produced uninterrupted and the vocal tract is in continuous motion. As a result, adjacent articulatory gestures overlap; this is named co-articulation (Ladefoged, 2000). Sound segments can be subjected to backward co-articulation, which occurs when the articulatory characteristics of a previous segment are seen in a later segment, and forward co-articulation, which occurs when articulatory characteristics of a later segment influence preceding segments (Daniloff & Moll, 1968).

Stops, nasals, and fricatives are articulated distinctly from vowels so that even in the midst of co-articulation, their waveforms possess defining characteristics that can be consistently recognized as the onset of their acoustic boundaries. Stops are formed by a complete closure of two articulators; the onset of a stop is characterized by the obvious cessation of energy in the spectrogram (Ladefoged, 2000). In nasals, two articulators are completely constricted and the soft palate is lowered so that air flows through the nose. The addition of the nasal branch to the vocal tract creates a characteristic low frequency murmur (due to the elongated vocal tract) and a sudden decrease in energy (due to the anti-resonances in the vocal tract), both marking the transition from vowel to nasal. Fricatives are described as the narrowing of two articulators; their initial boundary can be recognized by a sharp increase in noise in the waveform (Borden, Harris, & Raphael, 2003).

Semivowels are difficult to parse because they are articulated similarly to vowels and, except of the /rl/ cluster as in "snarl", must occur adjacent to a vowel (Espy-Wilson, 1987). Both vowels and semivowels are articulated orally without complete closure of the vocal tract, so air flows freely through the vocal tract with no audible frication and no inhibition of voicing. Also like vowels, semivowels are distinguished by tongue position and the pattern of resonances those positions create. Although semivowels are acoustically more similar to vowels than to consonants, they are considered to be consonants because English phonotactic constraints require them to occur at the periphery of syllables, like other consonants. The result of their articulatory properties is a gradual acoustic transition from vowel to semivowel, and a waveform that does not possess clear acoustic landmarks to mark the onset or offset of the consonant (Borden, Harris, & Raphael, 2003).

Formant Frequencies

Resonances within the vocal tract are called formant frequencies. Formant information is one of several acoustic cues used to identify voicing, place, and manner of articulation of stops, fricatives, affricates, and nasals. For vowels and semivowels, formant information is the crucial cue for their identification and discrimination from one another (Dalston, 1975; Delattre, Liberman, Cooper, & Gerstman, 1952; Espy-Wilson, 1992; Fry, Abramson, Eimas, Liberman, 1962; Lehiste, 1964; Lisker, 1957; O'Connor, Gerstman, Liberman, Delattre, Cooper, 1957).

Source-Filter Model of Speech Production. Acoustic output of speech is commonly considered the result of the combination of a sound source of energy (i.e. vocal fold vibrations in the larynx) and a filter (i.e., the resonant response of the supraglottal vocal tract). This is called the source-filter model of speech production (Fant, 1960).

Vocal fold vibrations are achieved from aerodynamic forces, muscular tension, and tissue elasticity. The vocal folds are first drawn together by activities of various laryngeal adducter muscles, such as the transverse interarytenoid muscle (which adducts the arytenoids cartilages at the back of the larynx) and the lateral cricoarytenoid muscles (which rock the muscular processes of the arytenoids). When the vocal folds come together, the velocity of the air particles passing through the glottis increases, resulting in a pressure drop between the medial edges that pulls the vocal folds together. Subglottal pressure then builds up, forcing the folds to break apart, and the cycle repeats again (Borden, Harris, Raphael, 2003; Titze, 1988). The activity of the vocal folds causes the airstream flowing from the lungs to be parsed into rapid puffs, eliciting an acoustic shock wave that propagates to the outside. The number of glottal openings per second is called the fundamental frequency.

A spectrum is a graphical representation of a sound source's component frequencies. The spectrum of the sound source consists of the fundamental frequency and its many harmonics. Harmonics are multiples of the fundamental frequency. These harmonics decrease in intensity by about 12 dB per octave as frequency increases (Fant, 1960). The frequency response of the supraglottal vocal tract filters the output of the glottal source. The harmonics of the sound source that are close to the resonant frequencies of the tract increase in energy and those that are far from the resonant frequencies decrease in energy. The result is a sound wave that contains the same fundamental frequency and harmonics of the glottal sound source, except the amplitudes of the harmonics have been modified, altering the quality of sound and the perceived sound segment (Borden, Harris, Raphael, 2003).

Formant Frequencies: Resonances of the Human Vocal Tract. Every tube will resonate naturally at specific frequencies, depending on its length and configuration. In neutral

position, the human vocal tract is an acoustic resonator that acts similarly to a tube open at one end and closed at the other end. In such a tube, resonances occur when the velocity of air molecules is at maximum at the open end. The first, or lowest, resonant frequency has a frequency whose wavelength is four times the length of the tube. When 1/4 of the wave is in the tube or tract, the velocity of air molecules will reach a maximum at the opening of the tube (or in the human resonator, the lips). At the closed end of the tube (or the glottis in the human vocal tract) the velocity of air molecules is at a minimum but pressure is high. The vocal folds act as a "dead end" for the air molecules where they have little room to move. The second frequency at which such a tube resonates is three times the lowest resonant frequency; its vocal tract includes two points of maximum velocity and two points of maximum pressure. The third resonant frequency in such a tube has a frequency that is five times the lowest resonant frequency; it includes three points of maximum velocity and three points of maximum pressure (Chiba and Kajiyama, 1941).

Points of maximum velocity and pressure are important because resonant frequencies change when constrictions or perturbations occur close to points of maximum velocity and pressure. Researchers have examined the changes in resonant frequency as a function of perturbations in some region along the vocal tract, indicating the effects on formant frequencies when moving from one vocal tract configuration to another, and how a vocal tract should be manipulated to shift formant frequencies a certain way (Chiba & Kajiyama, 1941; Fant, 1980; Schroeder, 1967). In general, they show that in the vocal tract, constrictions at points of maximum velocity reduce resonant frequency, and constrictions at points of maximum pressure increase resonant frequency.

Semivowel Articulation and Acoustic Consequences

Semivowels can be sub-divided into glides /w, j/ and liquids /r, l/. Glides are produced with articulators that are in continuous motion. Glides are produced with a narrowing of the oral cavity that is more constricted than high vowels, but not constricted enough to produce turbulent noise. The acoustic consequence of this vocal tract configuration is a relatively low F1 (around 250 to 300 Hz), resulting in a reduction in F1 spectral peak amplitude. The narrowed vocal tract constriction also reduces the amplitude of the glottal pulses. As air flows through the constriction during changing glottal volume flow, a pressure drop across the constriction occurs, resulting in a reduction in trans-glottal pressure during the rise phase of the glottal flow. The low F1 and source modification reduce the amplitude of the glide during the constricted interval and serve to enhance the contrast between the glide and its adjacent vowel (Stevens, 1998).

Like glides, liquids /r, l/ are also produced with a constriction that is not sufficiently narrow to cause turbulent noise at the vicinity of the constriction. In liquids /r, l/ the tongue blade is raised so that the edges of the blade near the tongue tip are close to the hard palate causing a bifurcation of the airway. This point of constriction is much smaller than that of glides, resulting in a relatively higher F1 frequency. The small constriction also introduces acoustic impedances due to vocal tract walls and kinetic resistance, which tends to broaden the low-frequency spectral peak and reduces glottal source amplitude. Like in glides, the reduced glottal amplitude helps to contrast the liquid and the syllabic vowel (Stevens, 1998).

Semivowel Perception Studies

Second and Third Formant Pattern. Studies using synthetic and natural speech stimuli have revealed that semivowels can be distinguished from one another using the 2nd and 3rd formant transition patterns (Lisker, 1957; O'Connor, Gerstman, Liberman, Delattre, & Cooper,

1957). Formant transitions are shifts in formant frequencies. These transitions reflect the change in vocal tract resonances as the vocal tract shape changes from the position of the consonant to that for the proceeding vowel, and vice versa. O'Connor, Gerstman, Liberman, Delattre, and Cooper (1957) attempted to synthesize syllable-initial (CV) semivowels by varying the starting point of second formant transition before the vowels /e, a, o/ in three conditions: no third formant, a straight third formant, or a rising third formant. They found that second formant transition starting point was crucial for identifying /w/, /r, l/, and /j/; this formant transition must start low for /w/ (around 600 Hz), mid-range for /r, l/, and high for /j/ (around 2760 Hz). The third formant distinguishes the liquids /r/ and /l/, with a rising third formant important for identifying /r/, and a straight third formant for identifying /l/. Lisker (1957) found the same patterns when he tried to synthesize intervocalic semivowels.

Spectrographic analysis of adult real speech (Dalston, 1975; Espy-Wilson, 1992; Lehiste, 1964) demonstrates the same F2 and F3 frequency transition pattern for correct semivowel perception. Dalston (1975) investigated the spectral and temporal acoustic characteristics of correct and incorrect /w, r, l/ phonemes produced in word-initial position by children and adults and Espy-Wilson (1992) examined the spectral characteristics of all four semivowels in prevocalic, intervocalic, and postvocalic position with varying points of stress. As in previous synthetic speech studies, Dalston (1975) and Espy-Wilson (1992) observed that a low F2 reliably distinguishes /w/ from /r, l/ and a low starting F3 distinguishes between /r/ and /w, l/.

Transition Duration. Formant transition duration is a crucial cue for distinguishing semivowels from other classes of speech sounds. Liberman, Delattre, Gerstman, and Cooper (1956) found they could synthesize syllable-initial stop consonants, semivowels, and vowels simply by varying the rate of formant transition before the vowel /ɛ/ (as in "bet"). When they

initiated the vowel with a very brief formant transition of 15 to 30 ms, listeners reported hearing /bɛ/ and /gɛ/. When they increased the duration of the transition to 40 or 50 ms listeners reported hearing semivowel glides /w/ and /j/ at the start of the syllable, and when they increased the duration to 150 ms or more, listeners reported hearing vowels /i/ and /u/ at the beginning of the syllables. They, and others (e.g., Cooper, Ebert, & Cole, 1976; Schwab, Sawusch, & Nusbaum, 1981), reported syllables with short initial transitions are perceived as beginning with stop consonants, and those with longer transitions are perceived as beginning with semivowels.

Further research shows that the required formant transition durations for stop or semivowel perception are influenced by overall syllable duration. Miller and Liberman (1979) found that the stop/glide distinction was influenced by the duration of the following vowel; particularly that a longer adjacent vowel duration shifted the stop/glide boundary toward a longer transition so that more stops are perceived. The required change in transition duration related to vowel duration has been reported to be non-linear (Miller & Baer, 1983).

Identifying the Visual Correlates of the Auditory Perception of /w j r l/

Although the acoustic properties and perceptual cues of /w, j, r, l/ have been studied extensively, it remains difficult to identify from waveforms and spectrograms the point at which the listener has sufficient acoustic information to just perceive the presence and the identity of naturally spoken /w, j, r, l/. In natural speech, where semivowels occur most frequently inter-vocalically (Pickett, Bunnell, & Revoile, 1995), the smooth transition from vowel to semivowel and co-articulation of neighboring sounds makes it difficult to determine when the formant transitions of the vowel end and the transitions of the semivowel begin. Several researchers have used a gating paradigm to evaluate the distribution of perceptually relevant information along the temporal dimension (e.g. Furui, 1986; Kurowski & Blumstein, 1987; Ohman, 1966; Smits,

2000). In this paradigm, a spoken stimulus is time-sliced relative to a landmark (known as the gating landmark) and is presented to a listener in segments of varying duration. Participants are then asked to identify the speech sound presented (Grosjean, 1996). By varying the cut-off points in the stimulus, the perceptual relevance of various parts of the signal can be measured. Gating studies reveal that across listeners, correct identification of specific phonemes will occur at a common time point in relation to the gating landmark (e.g., Grimm, 1966; Ohman, 1966; Smits, 2000). Many researchers have used the categorical data to determine the temporal distribution of perceptually important information, such as manner, place, and voicing, per phoneme (e.g., Smits, 2000). Others have used the data to investigate perceptual significance of aspects of the waveform (e.g., Furui, 1986).

The gating paradigm has been shown to be valid and useful in spoken word recognition research. Researchers have used the gating paradigm to replicate a number of effects found with other paradigms such as context, word length, and word frequency (Craig & Kim, 1990; Craig, Kim, Ryner, & Chirillo, 1993; Walley, Michela, & Wood, 1995), and subjects reportedly show the same results when they are under time pressure (Tyler & Wessels, 1985). In addition, the gating paradigm allows precise control over the acoustic and phonetic information presented to subjects, and may indicate how much acoustic-phonetic information is needed to identify the stimulus (Grosjean, 1996). Grosjean (1996) asserts a main issue regarding the gating paradigm is that it may not be considered a real "on-line" paradigm because the task does not directly reflect the online activation of phonemes in perception. In the gating task, listeners engage in a decision process that may have no part in speech perception since listeners normally incorporate additional processing time and acoustic information. However, this paradigm "offers the best

window into the listener's resolution of ambiguity as the speech signal unfolds" (Smits, Warner, McQueen & Cutler, 2003, p. 563).

Selected Studies on Acoustic Correlates to Semivowel Perception. It has been shown that maximum spectral transition points on the waveform quantify consonant and vowel perception. Furui (1986) recorded all 100 phonotactically-possible short syllables of Japanese spoken by two trained female speakers. These syllables included Japanese /w, j, r/. Initial and final truncated versions of the syllables were presented to listeners for identification of both the consonant and vowel. Correct identification curves for each syllable revealed "perceptual critical points," defined as the point where 80% correct was exceeded for the first time. At these points, percent correct identification for the truncated syllable as a function of truncation point changed abruptly. By calculating the spectral transition change in 5 ms increments, Furui (1986) found these perceptual critical points to be related to maximum spectral change. He found a speech wave of approximately 10 ms in duration that includes the point of maximum spectral transition bears the most important information for consonant and syllable perception.

Relative energy difference between adjacent vowels and semivowels has also been reported as useful for their speech class identification (Espy-Wilson, 1992; Espy-Wilson, 1994). Espy-Wilson (1992) quantified and analyzed the distinctive linguistic features that characterize the semivowels in American English. Based on the distinctive feature theory, this study investigated the acoustic correlates for the linguistic features sonorant, syllabic, consonantal, high, back, front, and retroflex. The final acoustic correlates were then tested on a group of 233 polysyllabic words, each spoken once by two females and two males. Within the "defined" semivowel boundaries, Espy-Wilson (1992) found the quantified acoustic properties generally successful at separating the semivowels from other sounds. Particularly, the acoustic measure

for the feature [-syllabic], defined as a significant dip in the mid frequency range relative to energy in an adjacent vowel, was generally successful at separating vowels and semivowels.

The above studies give insight to how the onset of /w, j, r, l/ may be quantified on the waveform. Furui's (1986) Japanese study of CV syllables revealed the existence of perceptual critical points, and found that maximum spectral transition points may predict consonant identification, including /w, j, r/. Espy-Wilson (1992) showed that relative measures of the difference in acoustic energy between vowels and semivowels are useful to separate the two speech classes. Furui's (1986) study is limited because, within the purpose of the current study, it is unclear whether this cue occurs for intervocalic consonants as well. In addition, although Espy-Wilson (1992) successfully showed that within a "defined" region relative acoustic measures can be useful for identifying the presence or absence of semivowels, it is unclear whether the same types of measures are useful for defining exact onset of the perceptual boundary between semivowels and vowels.

Purpose

This study attempts to answer two questions. First, for VC segments extracted from the UWODFD speech stimuli, is there a common time point along the acoustic waveform at which all listeners reliably perceive the semivowels? If so, are there consistent acoustic cues to predict those perceptual boundaries? Within the formant transitions (to be defined in detail below), the acoustic cues of interest are a) formant patterning, described as ratios between the centre frequency of each pair of the first, second, and third formants, b) spectral change of formants 1, 2, and 3 (measured as formant slope over a 4 ms time frame), and c) duration within the formant transition required for correct identification (measured in absolute time and percentage of transition duration).

The gating paradigm was deemed an appropriate measure to identify the point at which listeners have sufficient information to just perceive the semivowels. The initial gating landmarks were chosen at points thought to be within the transition from non-perception to perception of the semivowel, but were not hypothesized to be perceptual critical points; they were simply starting points for creating the stimuli.

Formant transitions are operationally defined as the region beginning at the point of greatest energy within the initial vowel and ending at the point of lowest energy within the adjacent semivowel. Energy was calculated as the rms total across a wide bandwidth; that is, from 0 to 10,000 Hz. Figure 1.1 is an illustrative example of the operationally-defined formant transition of talker F2 sound W, as the sound changes from a vowel neutral /ʌ/ to the /w/ sound. Compared to vowels, semivowels are more constricted and therefore usually have less energy in the low-to mid frequency range (Espy-Wilson, 1992; Stevens, 1998).

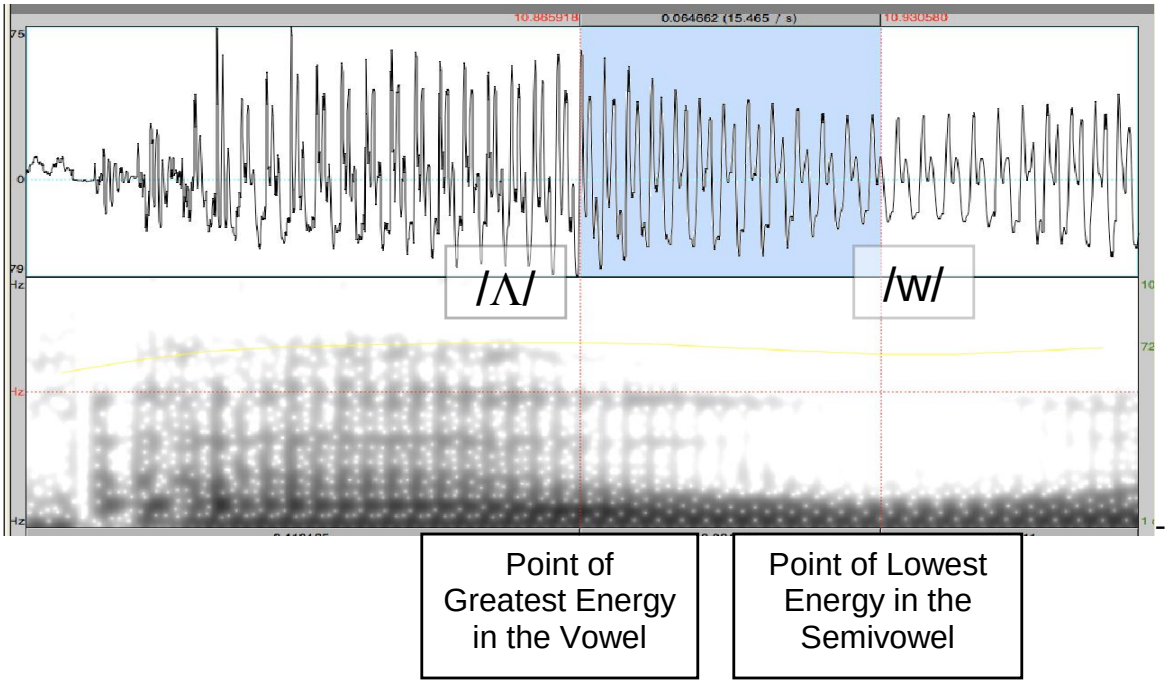


Figure 1.1 F2 W Formant Transition Region between /ʌ/ and /w/.

The first acoustic cue of interest is formant patterning. Past studies have identified the formant frequency trends, namely magnitude and direction of centre frequency F1, F2, and F3 related to semivowel identifcation in synthetic and real speech (Dalston, 1975; Lisker, 1957; O'Connor, Gerstman, Liberman, Delattre, Cooper, 1957). Since formant frequency values will vary due to speaker differences (Peterson & Barney, 1952), I wanted to explore whether defining the distance between the formant frequencies in terms of ratios would normalize the values and whether these values could be used to identify the onset of perception. The second acoustic parameter of interest is formant slope. Furui (1986) showed that Japanese listeners were able to identify word initial consonants based on maximal spectral change; it remains uncertain whether this cue is useful for consonant identification in intervocalic consonants. The third acoustic parameter of interest is formant transition duration. Several studies have shown formant transition duration to be important for semivowel identification (Cooper, Ebert, & Cole, 1976; Liberman, Delattre, Gerstman, & Cooper, 1956; Schwab, Sawusch, & Nusbaum, 1981), and the duration required for correct identification is related to following adjacent vowel length (Diehl & Walsh, 1989; Miller & Baer, 1983; Miller & Liberman,1979). This suggests that absolute transition duration may not be a suitable for semivowel perception because it varies according to the adjacent vowel. Instead, it may be more suitable to define semivowel perception in terms of relative transition duration length. Since Miller and Baer (1983) showed that the relationship between vowel duration and formant transitions for semivowel identification is not linear, it will be of interest to examine percentage of duration, compared to absolute duration, of the formant transition needed for semivowel identification. Such an examination would normalize the transition duration and eliminate the complex influence of the following vowel length.

Hypotheses

Based on previous gating studies, it was predicted that across listeners, individual target semivowels /w, j, r, l/ will be just recognized at a common time point along the target waveforms (Furui, 1986; Ohman, 1966). Also, based on classic studies of semivowel perception, Furui's 1986 study, and studies on semivowel transition durations, these time points were hypothesized to be predicted by one or more of the following acoustic cues: a) formant patterning for formants 1,2, and 3, b) magnitude of spectral change for formants 1, 2, and 3, and c) percent or absolute transition duration (Dalston, 1975; Diehl & Walsh, 1989; Lisker, 1957; Miller & Baer, 1983; Miller & Liberman,1979; O'Connor, Gerstman, Liberman, Delattre, Cooper, 1957).

2 Method

Participants

Ten participants took part in this study. All participants were fluent speakers of Canadian English and had normal hearing (i.e., 20 dB HL or better for octave frequencies between 250 Hz to 8 kHz) in the test ear. One participant, male (age 24, subject 1), took part in the pilot study. One female participant failed to attend the required number of study sessions and was therefore disqualified (subject 5). The remaining eight participants, 4 males and 4 females (ages 22 to 38) completed the experimental study.

Participants were screened for cognitive function using the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) and for short-term memory deficits using the digit span subtest of the Wechsler Scale Form I. The minimum scores required were 26/30 for the MMSE and 11/15 for the digit test. All participants passed both tests.

Stimuli

Using Praat Version 5.018, test tokens were created by time slicing the set of 21 consonant target test items, spoken by 4 talkers, from the University of Western Ontario Distinctive Feature Differences (UWODFD) Test (Cheesman & Jamieson, 1996 adapted from Feeney and Franks, 1982). The original 21 test items from the UWODFD Test were 2-syllable nonsense words in the form of /Λ**C**Il/, in which **C** was one of 21 target consonants /b, tʃ, d, f, g, h, j, k, l, m, n, p, r, s, ʃ, t, δ, v, w, y, z/ (e.g. afil). The test items were spoken by four talkers, two males (M1, M2) and two females (F1, F2), who ranged in voice and speaking style. All talkers were native speakers of central Canadian English.

The parameters used in Praat for analysis and processing are listed in Appendix A. For stimulus creation (details of each condition described below), speech segments were extracted by placing the cursor at the nearest zero crossing and were ramped at the onset/offset using a 10 ms linear ramp to avoid a broadband transient caused by an abrupt onset of offset, which is distracting to listeners and introduces auditory masking effects (Pols & Schouten, 1978).

Ungated Tokens. To obtain the ungated tokens, all the UWODFD Test tokens /b, tʃ, d, f, g, h, j, k, l, m, n, p, r, s, ʃ, t, δ, v, w, y, z/ were time sliced and reduced from two syllable test tokens /Λ**C**Il/ to one-syllable test tokens containing only the first syllable /Λ**C**/. These test tokens were time sliced so that when presented, listeners would unambiguously hear the correct consonant. In total, 84 ungated test tokens were created (21 sounds x 4 talkers).

Practice Tokens. To obtain the gated practice tokens, the consonants /b, t, m, s/ were chosen from the set of sounds that were not semivowels, to include a variety of manner, place, and voicing characteristics. All UWODFD test tokens containing these sounds were time-sliced in 10-ms intervals, at seven time points before and after a chosen landmark (i.e. -70, -60, -50, -40, -30, -20, -10, 0 (landmark), 10, 20, 30, 40, 50, 60, 70 ms).These time points were always shifted to the nearest zero-crossing. The landmarks for the stops /b, t/ were chosen to be the mid point in the silent gap. The landmark for /s/ was the point where the waveform resembled the steady-state aperiododicity of the fricative with no obvious sinusoidal pattern visible; that is, the frication noise showed no over-lay onto a sinusoidal pattern and simply centred over the zero-crossing (Figure 2.1). The landmark for /m/ was the nearest zero-crossing where the waveform just began to resemble the steady-state portion of the nasal sound, in both morphology and amplitude (Figure 2.2). The specific point of the landmark chosen was not critical, because it

was simply used as a reference point for labeling the time slices. In total, 240 of these gated tokens were created (15 time slices x 4 sounds x 4 talkers).

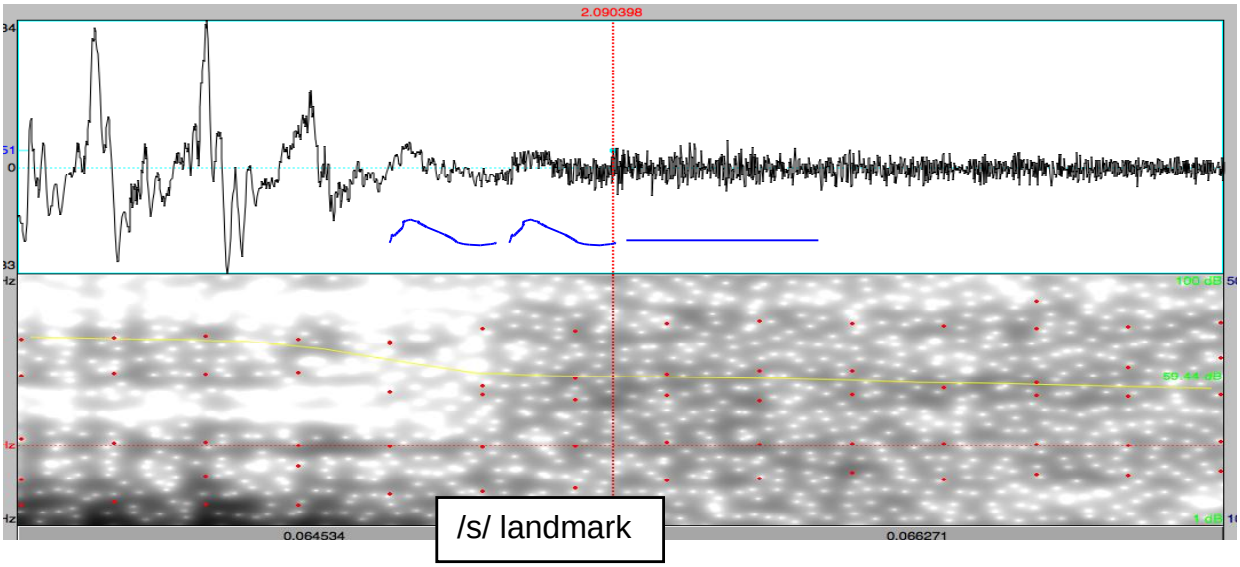


Figure 2.1 Talker M2 "aSil" Landmark (13 ms window).

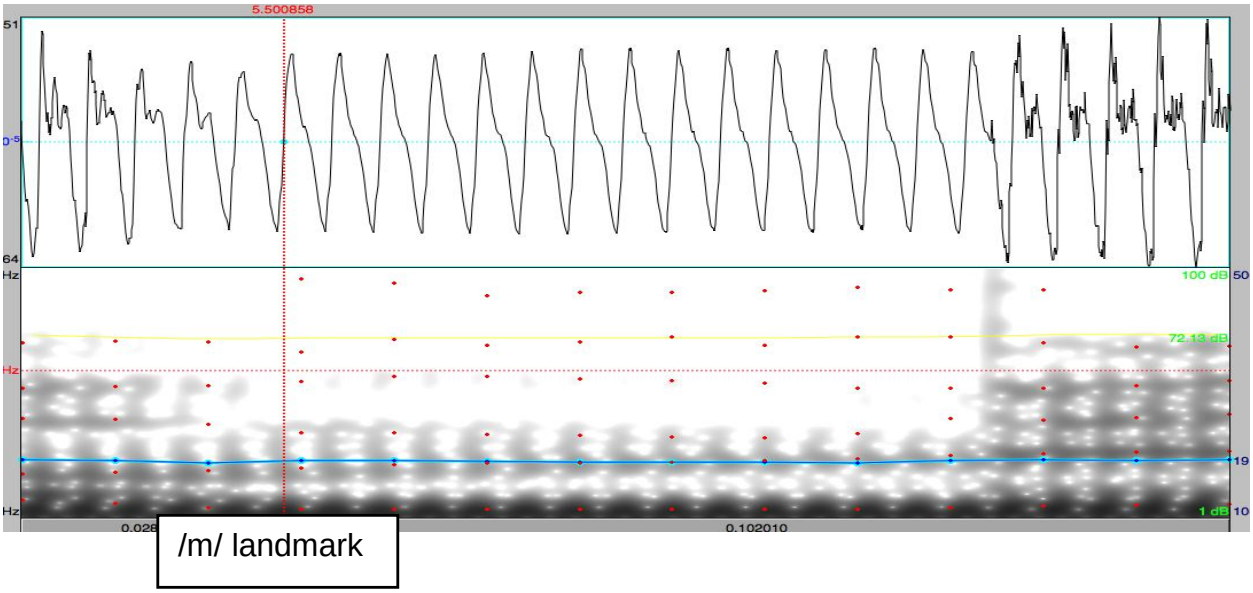


Figure 2.2 Talker F2 "aMil' Landmark (13ms window).

Target Tokens. To obtain target tokens, all UWODFD Test tokens containing semivowels /w, j, r, l/ were time-sliced in 5 ms steps, 14 times before and after the landmark (i.e.

-70, -65, -60, -55, -50, -45, -40, -35, -30, -25, -20, -15, -10, -5, 0 (landmark), 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 ms). These time points were always chosen at the nearest zero-crossing to the nominal gating point. The landmarks for /w, j, r, l/ were chosen to be the nearest zero-crossing where the waveform first resembled the morphology and amplitude of the steady-state portion of the semivowel (Figures 2.3 and 2.4). As before, the exact location of the landmark was not considered to be critical. In total, 464 gated target tokens were presented (29 time slices x 4 sounds x 4 talkers).

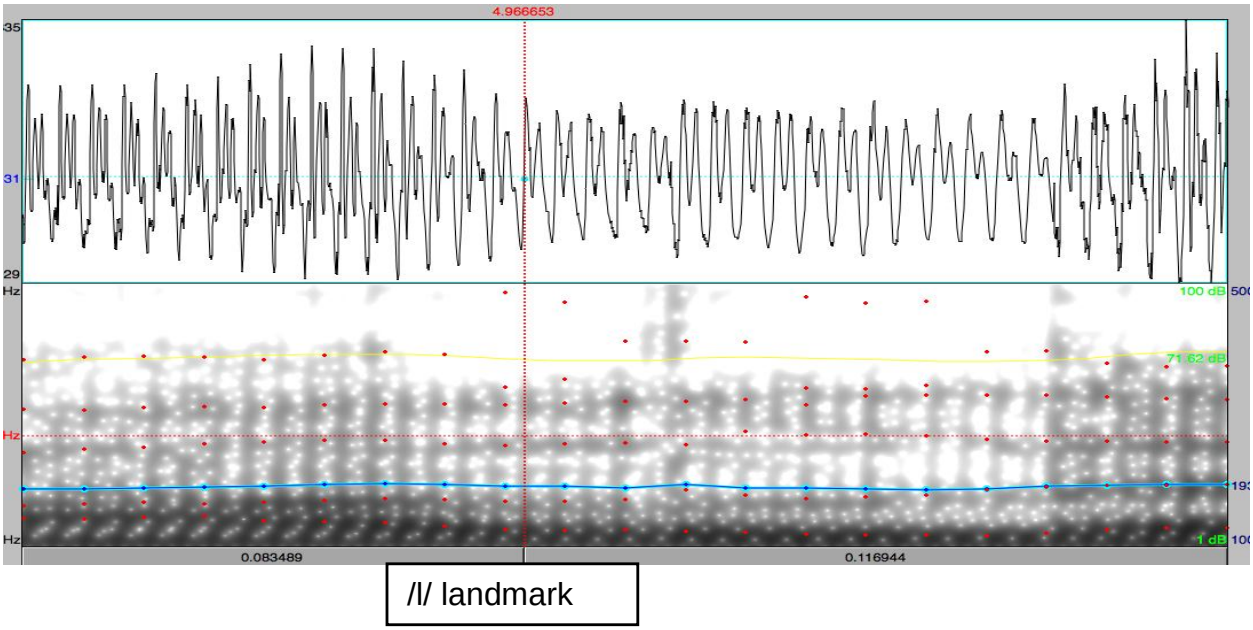


Figure 2.3 Talker F2 "aLil" Landmark (20 ms window).

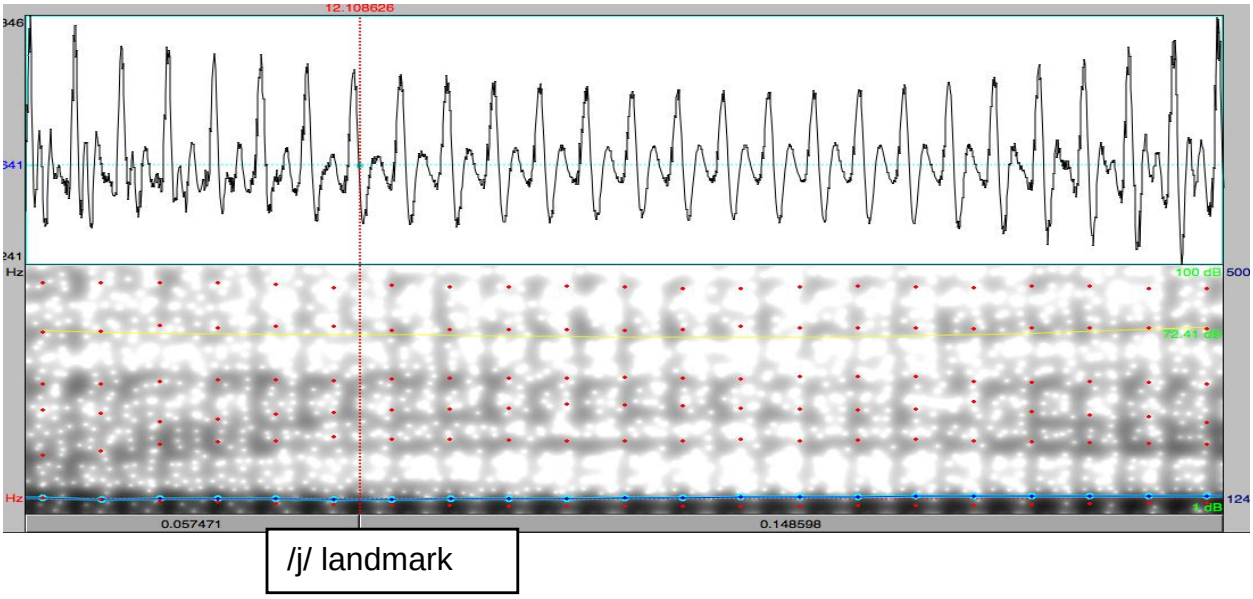


Figure 2.4 Talker M1 "aYil" Landmark (20 ms window).

Equipment

Normal hearing thresholds for all subjects were confirmed using a Grason Stadler GSI-61 audiometer.

The test stimuli for speech perception were stored on a local PC. On playback, the digital stimuli were routed to the Tucker-Davis Technologies (TDT) System 3. The stimulus was then processed by an RP2.1 Enhanced Real-Time Processor and the level was controlled by a programmable attenuator (TDT PA5) under computer control. The output was delivered to an HC7 headphone driver to be converted from digital to analog and then delivered to a Sennheiser HD250 linear headphone only to the right phone. All stimuli were calibrated once at the beginning of the project, using a Larson-Davis 824 Sound Level meter, to present 70 dB SPL through the right phone. Daily listening checks were subsequently conducted.

Speech perception was measured in a double-walled audiometric sound booth. Participants were seated in a height adjustable chair facing a touch screen computer monitor. SykofizX 2.0 software was used to present the stimuli and collect responses.

Procedure

Each participant took part in three study sessions of approximately 2 hours each. All screening measures were performed at the beginning of the first session. For data collection, participants were seated in a sound-proof booth in front of a touch screen monitor that presented a set of 21 possible responses. The response alternatives were presented on the screen as a**B**, a**CH**, a**D**, a**F**, a**G**, a**H**, a**J**, a**K**, a**L**, a**M**, a**N**, a**P**, a**R**, a**S**, a**SH**, a**T**, a**TH**, a**V**, a**W**, a**Y**, a**Z**. The listener's task was to choose the consonant heard by pressing the corresponding button on the screen. Listeners were instructed that sometimes it would be easy to identify the consonant and sometimes it may be more difficult. Even if the consonants were hard to identify at times, listeners were encouraged to make a guess.

In the first study session, listeners began with a practice test (Practice A). The goals of Practice A were to 1) screen the listeners for participation by evaluating whether listeners were able to recognize /w, j, r, l/ in their ungated form, and to 2) teach the listeners to extract meaningful phoneme information from gated stimuli. Practice A included all 84 ungated stimuli (one syllable versions /Λ**C**/ of all 21 consonants x 4 talkers) and all 240 gated practice tokens (i.e. /b, t, m, s/). The stimuli were presented in random order in a single block. Passing criterion was 75% correct identification for the ungated semivowel sounds /w, j, r, l/ across the 4 talkers. After passing Practice A, listeners were presented three blocks of the experimental stimuli, with the option of breaks between blocks. Each block included all 84 ungated stimuli (21 consonants x 4 talkers) and all 464 gated target tokens /w, j, r, l/ for the 4 talkers (548 tokens total), which were presented in random order to the listener without replacement. The number of tokens presented in this experimental session was 1644 ([3 blocks x 548 tokens]). Each gated target token was classified three times.

In the second study session, listeners began with a second practice test (Practice B). The purpose of Practice B was to 1) remind listeners of the gated sounds they would be presented and 2) re-familiarize listeners with the task. Practice B included one-third of the tokens from Practice A, for a total of 108 tokens. After Practice B, listeners were presented four blocks of the experimental stimuli, again, with the option of a break in between. The number of tokens presented in this experimental session was 2192 ([4 blocks x 548 tokens]). Each gated target token was classified four times.

Like the second study session, the third study session began with Practice B. After practice B, listeners were presented three blocks of the experimental stimuli, with the option of a break in between. Each gated target token was classified three times. The number of tokens presented in this experimental session was 1644 ([3 blocks x 548 tokens]). After the third and final study session, listeners had classified each gated target token 10 times in total for each of the 4 talkers ([3 times in the first session] + [4 times in the second session] + [3 times in the final session]). In total, each listener participated in about seven hours of study sessions.

Data Analysis

Treatment of the Data. The scores on the UWODFD speech task were calculated at each gating point in percent correct for each talker and each semivowel. A rationalized arcsine unit (RAU) was used to convert percent correct scores into a more suitable scale for statistical analyses. RAU transforms data from a proportion or percentage scale into a scale in which variance is independent of observed mean. Percent correct scores are mapped onto a range of -23 to +123 and units correspond closely to percent correct values from 15% to 85% (Studebaker, 1985; Studebaker, McDaniel, & Sherbecoe, 1995).

The UWODFD task, with its 21 response items, generally can be treated as an open-set task. However, in the current study, there was a potential for response biases as the task was heavily weighted on only four responses. That is, although listeners responded from a list of 21 sounds, the speech task was weighted, presenting the target semivowel tokens approximately five times more often than the ungated distractor tokens. Close examination of false alarm rates for each token showed that false alarm rates were fairly low and were unlikely to affect the analyses; despite this, false alarm rates are provided for each of the conditions for reference. False alarm rates were calculated across subjects for each semivowel, by dividing the number of incorrect responses by the total number of presentations that were not the targeted response.

In this study, 75 RAU for at least seven out of eight listeners was chosen as the "perceptual critical point". Similar values have been used in other studies to describe the critical point (Furui, 1986; Smits, 2000). A RAU of 25 or less for seven out of eight listeners was chosen as the boundary signifying no usable acoustic information.

Figures 2.5 and 2.6 illustrate speech recognition as a function of gating points for all eight listeners (listed by subject number), with the 25% and 75% correct points, and false alarm rates marked for two tokens. Because RAUs correspond closely to percent correct values of 15% to 85% (Studebaker, 1985; Studebaker, McDaniel, & Sherbecoe, 1995), all scores shown on the figures are in percent correct, for ease of representation. Graphs of the complete list of 16 target sounds are presented in Appendix B.

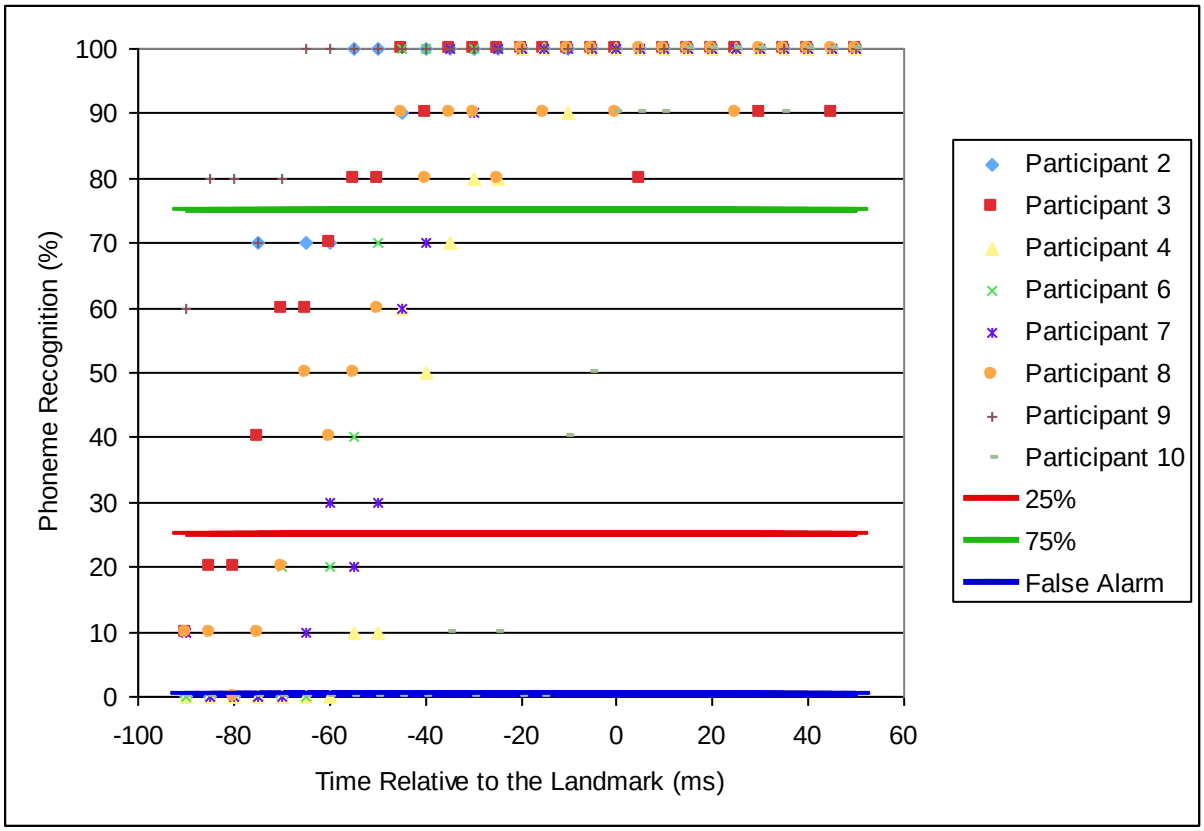


Figure 2.5 Phoneme Recognition for Individual Participants as a Function of Gating Time for Talker M2 Sound "aY".

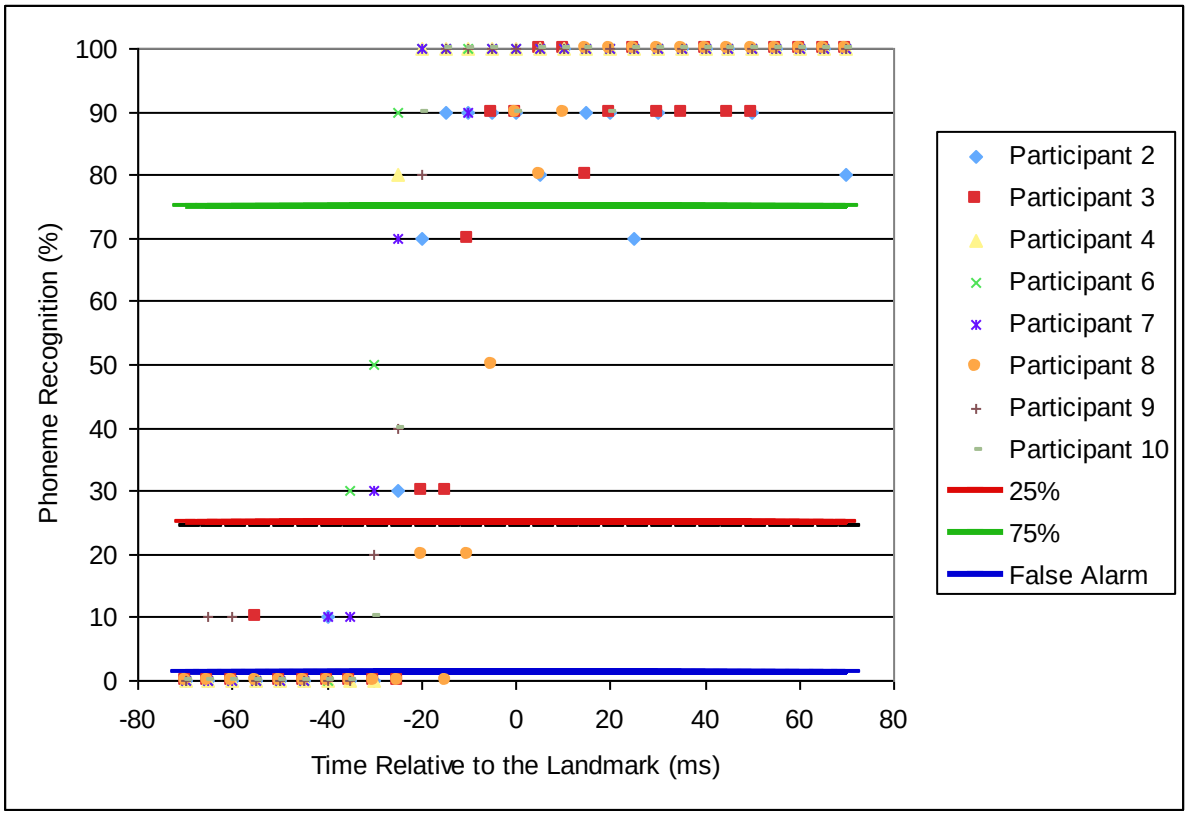


Figure 2.6 Phoneme Recognition for Individual Participants as a Function of Gating Time for Talker F2 Sound "aW".

A multiple regression analysis was performed for each semivowel to determine the amount of variance in semivowel perception accounted for by the acoustic variables. All four talkers were included in the analysis of each semivowel. The variables examined as possible predictors for perception of each semivowel were formant slopes (F1, F2, and F3 slope), formant ratios (F3:F1 ratio, F3:F2 ratio), and percent duration of transition and absolute duration of transition required within the operationally defined formant transition boundaries. F2:F1 ratio was not included in the regression analysis because an initial correlation analysis between variables revealed it was very highly correlated with F3:F1 ($r = +.834$, $p < 0.001$) and F3:F2 ($r = -.608$, $p < 0.001$,). The regression model was built by examining the contribution of the main-effect variables in a forward regression procedure, using alpha-level criterion of .05. Higher

order models were considered until adding a new variable accounted for less than 1% additional variance. Separate regression analyses were run for the percent of transition and absolute duration variables due to their high correlation value (r= +.940, p=0.01).

Acoustic Analyses

Formant Tracking. Spectrographic analysis of the 16 target speech tokens (a**L**, a**R**, a**W**, a**Y** spoken by four talkers) was conducted using Praat Version 5.018 to provide the formant values for the regression analysis. Formant tracking was based on linear predictive coding (LPC), and measurement parameters for each talker and semivowel were set according to the best representation of formant pattern based on visual inspection (see Appendix A and C for parameter values). The LPC method was unable to generate reliable formant patterns for any of the talker M2 sound R slices and some sound slices for other talkers; therefore, those time slices were not included in the regression analyses. Table 2.1 lists the time slices omitted from the regression analyses for each talker and sound, described relative to the landmark and the number of useable time slices for each target sound, out of a maximum possible 116 time slices.

Table 2.1 *Time Slices Omitted from Analysis*

Sound	Speaker	Missing Time Slices (relative to the gating landmark)	Total Time Slices Missing	Total Time Slices Included out of a Maximum 116
L	F1	-15 to +70	18	98
R	F1 M2	+50 to +70 All	5 29	82
W	F1 M2	+5 to +70 +60 to +70	14 3	99
Y	none	none	none	116

Calculating Formant Slope and Formant Ratio. Formant slopes and ratios were calculated for all useable time slices for every semivowel sound. In this study, spectral change for each formant was calculated as formant frequency slope, and the degree of separation between the formant frequencies at a particular time point was calculated as a ratio between formant frequencies.

Formant slope is the magnitude of formant frequency change over a 4 ms time frame. For a given time slice (i.e., 0, -5, -10, etc.), formant slope was calculated by finding the difference between formant frequency 2 ms before the time sliced point and 2 ms after the time sliced point, and dividing the frequency difference by 4 ms. Formant ratio is the magnitude of formant frequency separation at a given time point. Formant ratio was calculated by dividing the larger formant value by the smaller formant value. For example, to determine F3:F1 ratio, the F3

value for a given time point was divided by the F1 value at the same time point. The same calculation was used to determine F3:F2 and F2:F1 ratios for each time slice.

Calculating Percent Duration of Transition and Absolute Duration of Transition. Percent duration and absolute duration of the formant transitions were both calculated by first defining the formant transition boundaries within the vowel-consonant (VC) cluster. The formant transition onset was defined as the point of greatest overall energy within the vowel and the offset was defined as the point of lowest overall energy within the consonant (semivowel). Percent duration and absolute duration of transition were calculated for all the time slices in the study. Percent duration of transition was calculated by subtracting the time sliced point from the point where maximum energy occurred in the initial vowel, and dividing that value over the entire time region required for the greatest energy in the initial vowel to fall to the lowest energy in the following semivowel. Absolute duration of transition was calculated by subtracting a time sliced point from the point of maximum energy in the initial vowel.

3 Results

Acoustic Predictors of /l/ Identification

The correlations among all the possible predictor variables for /l/ are presented in Table 3.1, with significant correlations starred.

Table 3.1 *Pearson Correlations Among All Possible Predictor Variables for /l/*

	F3:F2	F3:F1	F1 Slope	F2 Slope	F3 Slope
F3:F2	1.000				
F3:F1	.091	1.000			
F1 Slope	.052	.038	1.000		
F2 Slope	.013	.187*	-.096	1.000	
F3 Slope	-.037	-.056	-.134	-.186*	1.000
Percent Duration of Transition	.327*	.752*	.054	.147	-.165
Absolute Duration of Transition	.143	.935*	.039	.163	-.109

Note. *Correlation is significant at the 0.05 level (1-tailed)

The regression model for /l/ analyzed with percent duration of transition is given in Table 3.2. The only significant predictor was the F3:F1 ratio, which accounted for 85.8% of the variance in RAU. The results show that percent correct scores improved when F3:F1 ratio increased. The regression model for /l/ analyzed with absolute transition of duration is given in

Table 3.3; this model accounted for 88.2% of the variance in RAU and included two significant predictors: F3:F1 ratio and absolute duration of transition. These results show that percent correct scores improved when F3:F1 ratio and absolute duration of transition increased.

Table 3.2 *Final /l/ Model: Regression Analysis including Percent Duration of Transition*

	R Square	R Square Change	Sig. F Change
F3:F1 ratio	.858	.858	.000

Table 3.3 *Final /l/ Model: Regression Analysis including Absolute Duration of Transition*

	R Square	R Square Change	Sig. F Change
F3:F1 ratio	.858	.858	.000
Absolute Duration of Transition	.882	.024	.000

Acoustic Predictors of /r/ Identification

The correlations among all the possible predictor variables for /r/ are presented in Table 3.4, with significant correlations starred.

Table 3.4 *Pearson Correlations Among All Possible Predictor Variables for /r/*

	F3:F2	F3:F1	F1 Slope	F2 Slope	F3 Slope
F3:F2	1.000				
F3:F1	-.174	1.000			
F1 Slope	.145	.082	1.000		
F2 Slope	.047	.095	.139	1.000	
F3 Slope	.040	.095	.060	-.050	1.000
Percent Duration of Transition	-.484*	.836*	-.045	.063	-.005
Absolute Duration of Transition	-.494*	.838*	-.039	.065	-.001

Note. *Correlation is significant at the 0.05 level (1-tailed)

The regression model for /r/ analyzed with percent duration of transition is given in Table 3.5. It accounted for 87.7% of the variance in RAU and included three significant predictors: percent duration of transition, F3:F2 ratio, and F1 slope. The data showed that increases in percent duration of transition and F1 slope, and decreases in F3:F2 ratio were related to improved percent correct scores. The regression model for /r/ analyzed with absolute duration of transition is given in Table 3.6; it accounted for 88.3% of RAU variance and included three significant predictors: absolute duration of transition, F3:F2 ratio, and F1 slope. In this model, increases in absolute duration of the transition and F1 slope, and decreases in F3:F2 ratio were related to improved percent correct scores.

Table 3.5 *Final /r/ Model: Regression Analysis including Percent Duration of Transition*

	R Square	R Square Change	Sig. F Change
Percent Duration of Transition	.833	.833	.000
F3:F2 ratio	.860	.027	.000
F1 Slope	.877	.017	.002

Table 3.6 *Final /r/ Model: Regression Analysis including Absolute Duration of Transition*

	R Square	R Square Change	Sig. F Change
Absolute Duration of Transition	.845	.845	.000
F3:F2 ratio	.868	.023	.000
F1 Slope	.883	.015	.002

Acoustic Predictors of /w/ Identification

The correlations among all the possible predictor variables for /w/ are presented in Table 3.7, with significant correlations starred.

Table 3.7 *Pearson Correlations Among All Possible Predictor Variables for /w/*

	F3:F2	F3:F1	F1 Slope	F2 Slope	F3 Slope
F3:F2	1.000				
F3:F1	.740*	1.000			
F1 Slope	.029	.108	1.000		
F2 Slope	.197*	.325*	.170*	1.000	
F3 Slope	.123	-.020	.125	.059	1.000
Percent Duration of Transition	.283*	.382*	.296*	.567*	.044
Absolute Duration of Transition	.353*	.563*	.291*	.547*	.002

Note. *Correlation is significant at the 0.05 level (1-tailed)

The regression model for /w/ analyzed with percent duration of transition is given in Table 3.8. It accounted for 89.3% of the variance in RAU and included three significant predictors: percent duration of transition, F3:F1 ratio, and F2 slope. Percent duration of transition, F3:F1 ratio, and F1 slope all increased as percent correct scores improved. The regression model for /w/ analyzed with absolute duration of transition is given in Table 3.9. It accounted for 87.0% of RAU variance and included two significant predictors: absolute duration of transition and F3:F2 ratio. The data showed that increases in absolute duration of transition and F3:F2 ratio led to improved percent correct scores.

Table 3.8 *Final /w/ Model: Regression Analysis including Percent Duration of Transition*

	R Square	R Square Change	Sig. F Change
Percent Duration of Transition	.805	.805	.000
F3:F1 ratio	.882	.078	.000
F2 slope	.893	.010	.003

Table 3.9 *Final /w/ Model: Regression Analysis including Absolute Duration of Transition*

	R Square	R Square Change	Sig. F Change
Absolute Duration of Transition	.836	.836	.000
F3:F2 ratio	.870	.035	.000

Acoustic Predictors of /j/ Identification

The correlations among all the possible predictor variables for /j/ are presented in Table 3.10, with significant correlations starred.

Table 3.10 *Pearson Correlations Among All Possible Predictor Variables for /j/*

	F3:F2	F3:F1	F1 Slope	F2 Slope	F3 Slope
F3:F2	1.000				
F3:F1	.164*	1.000			
F1 Slope	.038	.214*	1.000		
F2 Slope	-.083	-.510*	.094	1.000	
F3 Slope	-.037	-.006	.137	.217*	1.000
Percent Duration of Transition	-.026	.749*	-.373	-.100*	1.000
Absolute Duration of Transition	-.051	.710*	.085	-.349*	-.103

Note. * Correlation is significant at the 0.05 level (1-tailed)

The regression model for /j/ analyzed with percent duration of transition is given in Table 3.11. It accounted for 82.9% of the variance in RAU and included three significant predictors: percent duration of transition, F3:F1 ratio, and F3:F2 ratio. The data showed that increases in percent duration of transition and F3:F1 ratio, and decreases in F3:F2 ratio led to improved percent correct scores. The regression model for /j/ analyzed with absolute duration within transition is given in Table 3.12. It accounted for 83.1% of RAU variance and included three significant predictors: F3:F1 ratio, absolute duration of transition, and F3:F2 ratio. Increases in F3:F1 ratio and absolute duration of transition, and decreases in F3:F2 ratio led to improved percent correct scores.

Table 3.11 *Final /j/ Model: Regression Analysis including Percent Duration of Transition*

	R Square	R Square Change	Sig. F Change
Percent Duration of Transition	.674	.674	.000
F3:F1 ratio	.769	.088	.000
F3:F2 ratio	.829	.068	.000

Table 3.12 *Final /j/ Model: Regression Analysis including Absolute Duration of Transition*

	R Square	R Square Change	Sig. F Change
F3:F1 ratio	.658	.658	.000
Absolute Duration of Transition	.765	.108	.000
F3:F2 ratio	.831	.065	.000

Descriptive Analyses at the 25% and 75% Correct Points

Recall, 75 RAU (or 75% correct responses) for at least seven out of eight listeners represented the perceptual critical point for semivowel identification, and a RAU of +25 (or 25% correct responses) or less for seven out of eight listeners indicated no usable acoustic information for the listener. The significant acoustic predictors for each semivowel, indicated by the regression analyses, were examined across talkers for each semivowel at each of these two points, with the goal of describing the differences between the acoustic measures at these two points. The 25% and 75% correct points for each of the 16 UWODFD target sounds are given in Table 3.13.

Table 3.13 *The 25% and 75% Correct Points for Each Semivowel*

Talker	Sound	25% Phoneme Recognition Boundary Relative to the Landmark (ms)	75% Phoneme Recognition Boundary Relative to the Landmark (ms)
F1	L /l/	-55	-15
	R /r/	-35	11
	W /w/	-75	-10
	Y /j/	-99	-60
F2	L /l/	-39	27
	R /r/	16	40
	W /w/	-35	-5
	Y /j/	-89	-45
M1	L /l/	-60	0
	R /r/	-65	-35
	W /w/	-45	15
	Y /j/	-70	-19
M2	L /l/	-30	55
	R /r/	-110	-59
	W /w/	-54	36
	Y /j/	-79	-30

An illustrative example of this analysis is shown in Figure 3.1. The beginning of the shaded area represents the point at which percent correct is 25% and listeners were not able to recognize talker M1's /l/ sound, and the end of the shaded area represents the point at which percent correct is 75% and listeners were consistently able to correctly perceive it. F3:F1 ratio and absolute duration of transitions (the significant acoustic predictors determined by the regression analyses) were then compared at these two points. The values of each acoustic predictor at the two points of interest are listed in Table 3.14. The values presented are values

across the 4 talkers, showing averages values for formant slope and a range of values for formant ratios and percent/absolute transition duration. The values are discussed more fully in the following sections.

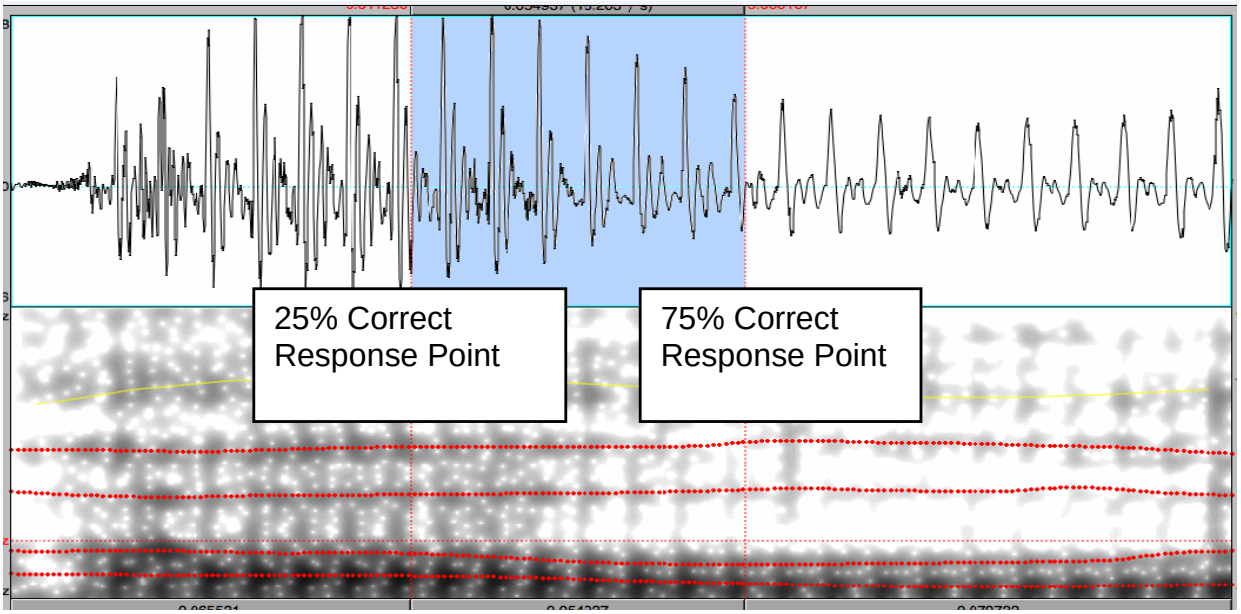


Figure 3.1 Comparing the 25% Correct and 75% Correct Points in the Waveform and Spectrogram of /l/ as Spoken by Talker M1. The beginning of the shaded area represents the point where responses were 25% correct for seven out of eight listeners and the end of the shaded area represents the point where responses were 75% correct for seven out of eight listeners.

L Sounds. The acoustic predictors for the perception of /l/ were F3:F1 ratio and absolute duration within the transitions; their values are given in Table 3.14. Across the four talkers, F3:F1 ratios ranged from 4.08 to 5.11 at the 25% correct point, and 6.39 to 7.46 at the 75% correct point. On average, F3:F1 ratio was larger at the 75% correct point compared to the 25% correct point, indicating better recognition scores when F3 and F1 formants are more separated. The absolute duration of transitions across the 4 talkers varied from -17 to 15 ms at the 25% correct point and 49 to 88 ms at the 75% correct point.

R sounds. The acoustic predictors for the perception of /r/ were the percent and duration within the transitions, F3:F2 ratio, and F1 slope; their values at the 25% and 75% correct points are given in Table 3.14. At the 25% and 75% correct points, the percent duration of transitions across talkers varied from -25% to 14% and 12% to 47% respectively, and the absolute duration of transition varied from -20 to 12 ms and 10 to 44 ms respectively. Across the four talkers, F3:F2 ratio ranged from 1.76 to 1.81 at the 25% correct point and 1.46 to 1.71 at the 75% correct point. On average, F3:F2 ratio was smaller at the 75% correct point compared to the 25% point, indicating better /r/ recognition when F3 and F2 frequencies were closer together. Finally, on average, F1 slope was greater at the 75% correct point compared to the 25% point, indicating that a greater magnitude of formant change is related to improved /r/ recognition.

W sounds. The acoustic predictors for the perception of /w/ were the percent and duration within the transitions, F3:F2 ratio, F3:F1 ratio, and F2 slope; their values are presented in Table 3.14. At the 25% and 75% correct points, the percent duration of transitions across talkers varied from -5% to 25% and 60% to 83% respectively, and the absolute duration of transitions varied from -5 to 20 ms and 50 to 86 ms respectively. For formant ratios, values at the 25% and 75% correct points ranged from 2.07 to 3.80 and 3.16 to 3.44 respectively for F3:F2, and 4.44 to 5.43 and 6.46 to 9.07 respectively for F3:F1. Both F3:F2 and F3:F1 ratios showed larger values at the 75% correct point compared to the 25% point, indicating that /w/ recognition scores improved when F3 was more separated from F2 frequency, and F3 was more separated from F1 frequency. Finally, on average, F2 slope magnitude was smaller at the 75% point compared to the 25% point, indicating that a decreased magnitude of formant change is related to improved /w/ recognition.

Y sounds. The acoustic predictors for the perception of /j/ were the percent and duration within the transitions, and F3:F1 and F3:F2 ratios; their values are given in Table 3.14. At the 25% and 75% correct points, the percent duration of transitions across talkers varied from -32% to 9% and 6% to 42% respectively, and the absolute duration of transitions varied from -41 to 13 ms and 7 to 57 ms respectively. For formant ratios, values at the 25% and 75% correct points ranged from 4.41 to 6.27 and 6.18 to 8.19 respectively for F3:F1, and 1.40 to 1.65 and 1.25 to 1.36 respectively for F3:F2. On average, at the 75% correct point F3:F1 ratio was larger and F3:F2 ratio was smaller than at the 25% point. This indicated better /j/ recognition when the F3 and F1 frequencies were farther separated, and when F3 and F2 frequencies were closer together.

Table 3.14 *Predictor Variables Identified as Significant: Comparison of Values at the 25% and 75% Correct Response Points*

Sound	Variables Included in the Model	Value at 25% Correct Point (no usable information)	Value at 75% Correct Point (usable information)
L /l/	Absolute Duration of Transition	-17ms – 15 ms	49 ms – 88 ms
	F3:F1 Ratio	4.08 to 5.11 M = 4.54	6.39 to 7.46 M = 6.98
R /r/	Percent Duration of Transition	-25% to +14%	12% to 47%
	Absolute Duration of Transition	-20 ms – 12 ms	10 ms – 44 ms
	F3:F2 Ratio	1.76 to 1.81 M = 1.79	1.46 to 1.71 M = 1.60
	F1 Slope	-2755.92 Hz/sec	-3709.24 Hz/sec
W /w/	Percent Duration of Transition	-5% to +25%	60% to 83%
	Absolute Duration of Transition	-5 ms – 20 ms	50 ms – 86 ms
	F3:F1 Ratio	4.44 to 5.45 M = 5.09	6.49 to 9.07 M = 7.51
	F3:F2 Ratio	2.07 to 3.80 M = 2.72	3.16 to 3.44 M = 3.34
	F2 Slope	-8417.77 Hz/sec	1816.02 Hz/sec

Sound	Variables Included in the Model	Value at 25% Correct Point (no usable information)	Value at 75% Correct Point (usable information)
Y /j/	Percent Duration of Transition	-32% to +9%	6% to 42%
	Absolute Duration of Transition	-41 ms – 13 ms	7 ms – 57 ms
	F3:F1 Ratio	4.41 to 6.24 M = 5.02	6.13 to 8.19 M = 7.53
	F3:F2 Ratio	1.40 to 1.65 M = 1.51	1.25 to 1.36 M = 1.31

4 Discussion

This study established perceptual critical points (the point at which 7 of 8 listeners achieved 75% correct) for all 16 semivowel sounds of the UWODFD speech test. Formant slope, formant ratio, and percent and absolute duration of transition were then examined as possible acoustic predictors to identify the perceptual critical points.

Spectral Change

The results of my study were somewhat consistent with Furui's (1986) consonant recognition study, which included three semivowels: /r, w, j/. His study showed that consonant identification scores for gated syllables were related to maximum spectral transition position. The regression analyses in the present study identified formant slope as an important acoustic predictor for two semivowel sounds, /r, w/. For /r/, F1 slope was positively correlated with percent correct score, so that as F1 slope magnitude increased, so did percent correct scores. The same trend was observed for /w/ recognition and F2 slope. For the other semivowels /l, j/, spectral change was not a significant predictor for semivowel identification. Though somewhat consistent with Furui's (1986) study, overall, if spectral change was an extremely important predictor of semivowel identification, I would have expected it to be reflected as a top predictor for all the regression analyses for each semivowel sound.

Two reasons may explain some of the differences observed in this study compared to Furui's (1986) study. First, my measure of spectral change was different from Furui's. Furui measured the change in overall frequency content, calculating cepstrum coefficients and 30 ms Hamming windows to create a linear approximation to represent the log-spectral envelope of the phonemes (Furui, 1986). On the other hand, I measured the change in individual formant frequency content, calculating the slope in each formant frequency over a 4 ms window. As

such, it is possible that my measures were identifying different aspects of spectral change, and therefore produced different results.

Second, Furui (1986) used consonant initial (CV) utterances and measured spectral change in the vocalic transitions following the consonant, while my study used consonant final (VC) utterances extracted from VCVC utterances and measured spectral change in the vocalic transitions preceeding the consonant. Co-articulation studies have shown that vocalic transitions preceding and following consonant sounds exhibit different spectral characteristics (Lee, 1997; Modarresi, Sussman, Lindblom, & Burlingame, 2004). It is possible that the trend observed in Furui's study is associated only in vocalic transitions following the consonant, rather than the vocalic transitions preceding the consonant, which is of interest in this study.

Formant Patterning

Formant patterning is the degree of separation between formant frequencies at a given time point. In this study, formant patterning was calculated as formant ratios F3:F1 and F3:F2. The relationship between improved speech identification and the formant patterning observed in this study seems to agree with past literature on semivowel production and perception (e.g., Espy-Wilson, 1992; Lisker, 1957; Stevens, 1998).

F3:F1 Patterning. This study showed that as the separation between F3 and F1 frequency increased, so did semivowel identification scores for /w, j, l/. This was expected, because semivowel articulation is more constricted than vowels and therefore tends to show a decrease in F1 frequency, and intervocalic /w, j, l/ has been shown, on average, to have higher F3 values than the preceding vowel (Espy-Wilson, 1992). The results indicate that when semivowel articulation is made more complete so that F1 and F3 separation increases, identification improves.

F3:F2 Patterning. For /w/, identification increased when the separation between F3 frequency and F2 frequency increased. This agrees with past perception studies (e.g., Lisker, 1957), that perception of /w/ requires the F2 frequency to be low and the F3 frequency to be slightly high. On the other hand, for /j/ and /r/, identification seemed to increase when the separation between F3 and F2 frequency decreased. These results were also expected, considering that for /j/, F2 frequency is predicted to be significantly higher than its preceding vowel, closing the distance between it and F3, and for /r/, F3 frequency has been shown to be significantly lower than the preceding vowel, closing the gap between it and F2 (Lisker, 1957).

For each regression model in this study, formant ratio was identified as a significant predictor of recognition, indicating that the separation between formant frequencies is a salient acoustic cue for semivowel identification. In this study a range of F3:F1 or F3:F2 ratio values were observed to signal the point at which listeners were able consistently to perceive semivowel sounds. For example, the F3:F1 ratios extracted from the points where listeners were just able to recognize the /j/ sound varied depending on the talker, ranging from 6.13 for talker F1 to 8.19 for talker F2. No absolute formant ratio values were observed to indicate the onset of any semivowel sound, which was expected based on the results of Dalston (1975) and Espy-Wilson (1992), indicating that formant ratio may not be suitable for semivowel segmentation.

Interestingly, for the most part, the range of ratio values observed at the point where listeners responded at chance (25%) compared to the point where listeners consistently perceived the sound (75%) showed no overlap (refer to tables 3.14 and 3.15), lending support to its importance as a cue. It may be that distinct ranges of ratio values are associated with perception and non-perception, and these boundaries do not overlap.

Percent and Absolute Duration of Transition

Miller and Baer (1983) showed that the transition duration required for correct semivowel identification differed depending on the length of the adjacent following vowel, and the relationship was not linear. To address the issue of varying transition durations, I examined normalized transition lengths (percent duration of transition) and then absolute transition length (absolute duration of transition) for comparison. Percent and absolute duration of transition were both significant predictors of semivowel perception; the regression analyses of each sound listed at least one of them as a top predictor of semivowel identification. Overall, regression analyses measured with either variable resulted in regression models that explained similar amounts of variance for each sound, with neither of the variables resulting in significantly better (i.e., more explained variance) regression models than the other. Despite this general trend, the amount of variance explained by percent and absolute duration was not identical. For example, percent duration of transition was not included in the regression model for /l/ whereas absolute duration of transition was, explaining an additional 2.4% of the variance. Overall, both regression models calculated with percent duration of transition and absolute duration of transition gave similar results, indicating that the definition of the transition was not critical for improved percent correct scores; rather, it is simply the amount of information available that is. This study did not give evidence for perceptual normalization of transition duration length.

No consistent amount of duration needed within the semivowel transitions was identified in this study. The percent and absolute duration of transition required for each sound to reach the perceptual critical point varied widely, with the percent duration of transition ranging from 12%-47% for /r/, for 60%-83% /w/, and 6%-42% for /j/, and the absolute duration of transition ranging from 49ms-88 ms for /l/, 10ms-44 ms for /r/, 50ms-86 ms for /w/, and 7ms-57 ms for /j/.

Percent required for /l/ is not presented here because it was not a significant predictor of /l/ perception.

The inconsistencies observed in the amount of percent and absolute duration of transition required may be due to an inaccurate definition of formant transition. In this study, formant transitions were operationally defined as the region between the greatest energy within the initial vowel to lowest energy within the following semivowel in a bandlimited range of 0-10,000 Hz. However, semivowels have been reported to usually have less energy specifically in the low-to mid frequency range (Espy-Wilson, 1992). As such, it may have been more suitable to define the formant transitions as the region between the greatest and lowest energy using a bandlimited range of 640-3000 Hz (based on Espy-Wilson, 1992) rather than the broader frequency range used in the current study.

Conclusion and Implications

The Acoustic Onset of Semivowel Identification. In this study no one set of acoustic variables could accurately predict the identification of each semivowel. This finding is similar to Dubno and Levitt's (1981) observation that acoustic variables are useful to predict consonant confusions, but different combinations of acoustic variables are needed to predict the results of different types of syllables. Some acoustic variables consistently did stand out to be important; namely, formant ratios and transition duration (in percent and/or absolute). At the perceptual critical points, the magnitude of F3:F1 and F3:F2 ratios were consistent with previous data on formant patterning and semivowel perception. However, the value of the ratios depended on speaker characteristics; therefore, no consistent F3:F1 or F3:F2 ratio was observed that could be used for semivowel identification. Noteworthy, a distinct range of ratio values was observed to separate the perception and non-perception of the semivowels, which may be useful for

semivowel identification. Percent and duration within formant transitions were also found to be important predictors of recognition, but again, no consistent percent or duration required within the transitions was observed to predict semivowel recognition. The current study shows that percent/duration within the transition and formant ratios cannot be used as acoustic correlates of the initial auditory perception of /w, j, r, l/.

Implications. The purpose of this study was to help establish a standardized test that is a phoneme by phoneme analysis of acoustics and behavioural speech recognition, and that can be used routinely to evaluate the effects of complex hearing aid processing. The UWODFD speech test was considered the speech test of choice for such a measure. Accurate acoustic boundaries related to phoneme perception are imperative for hearing aid researchers to correctly measure the acoustic effects of processing on phoneme recognition. The systematically evaluated perceptual boundaries established in this study provide useful information for determining the acoustic onsets of the UWODFD semivowel tokens, which to date, had not been reliably established (Jenstad, Barnes, & Hayes, 2008).

Because no acoustic properties were found to absolutely identify the perceptual onset of semivowel identification, my study re-affirms the need to use auditory analysis for segmentation of semivowels. The smooth acoustic transition between vowel and semivowel, and the variability of speaker rate, style, and pitch makes it difficult to determine clear acoustic rules for segmentation. Auditory analysis has been used for semivowel segmentation in the past (e.g., Balakrishnan, Freyman, Chiang, Nerbonne, & Shea, 1996; Kennedy, Levitt, Neuman, & Weiss, 1998; Jenstad & Souza, 2005), and unless evidenced otherwise, still seems to be the most appropriate method of semivowel segmentation to date.

References

Amlani, A., Punch, J., & Ching, T. (2002). Methods and applications of the audibility index in hearing aid selection and fitting. *Trends in Amplification, 6*, 81-129.

Balakrishnan, U., Freyman, R. L., Chiang, Y. C., Nerbonne, G. P., & Shea, K. J. (1996). Consonant recognition for spectrally degraded speech as a function of consonant-vowel ratio. Journal of the Acoustical Society of America, 99, 3758-3769.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84*, 101-114.

Boothroyd, A., & Medwetsky, L. (1991). Spectral distribution of /s/ and the frequency response of hearing aids. *Ear and Hearing, 13*, 150-157.

Bor, S., Souza, P., & Wright, R. (2008). Multichannel compression: Effects of reduced spectral contrast on vowel identification. *Journal of Speech, Language, and Hearing Research, 51*, 1315-1327.

Borden, G. J., Harris, K. S., & Raphael, L. J. (2003). *Speech science primer: Physiology, acoustics, and perception of speech.* Philadelphia: Lippincott Williams & Wilkins.

Cheesman, M F., & Jamieson, D. G. (1996). Development, evaluation and scoring of a nonsense word test suitable for use with speakers of Canadian English. *Canadian Acoustics, 24*, 3-11.

Chiba, T., & Kajiyama, M. (1941). *The Vowel: Its Nature and Structure*. Tokyo: Kaiseikan.

Cooper, W. E., Ebert, R. R., & Cole, R. A. (1976). Perceptual analysis of stop

consonants and glides. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 92-104.

Cox, R. M., Alexander, G., Gilmore C., & Pusakulich, K. (1988). Use of the connected speech test (CST) with hearing impaired listeners. *Ear and Hearing, 9*, 198-267.

Craig, C., & Kim, B. (1990). Effects of time gating and word length on isolated word recognition performance. *Journal of Speech and Hearing Research, 33*, 808-815.

Craig, C., Kim, B., Rhyner, P., & Chirillo, T. (1993). Effects of word predictability, child development and aging on time-gated speech recognition performance. *Journal of Speech and Hearing Research, 36*, 832-841.

Dalston, R. M. (1975). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America, 57*, 462-469.

Daniloff, R. G., & Moll, K. (1968). Coarticulation of liprounding. *J. Speech Hear. Res. 11*, 707-721.

Davies-Venn, E., Souza, P., Brennan, M., & Stecker, G. (2009). Effects of audibility and multichannel wide dynamic range compression on consonant recognition for listeners with severe hearing loss. *Ear and Hearing, 30*, 494-504

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color: observations on one and two formant vowels synthesized from spectrographic patterns. *Word, 8*, 195-210.

Diehl, R.L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America, 85*, 2154–2164

Dubno, J., Dirks, D., & Schaefer, A. (1989). Stop-consonant recognition for normal hearing listeners and listeners with high-frequency hearing loss. II: Articulation index predictions. *Journal of the Acoustical Society of America, 85*, 355-364.

Dubno, J., & Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *Journal of the Acoustical Society of America, 43*, 249-261.

Espy-Wilson, C. Y. (1987). An acoustic-phonetic approach to speech recognition: Application to the semivowels. Technical Report No. 531, Research Laboratory of Electronics, MIT.

Espy-Wilson, C. Y. (1992). Acoustic measures of linguistic features distinguishing the semivowels /w r j l/ in American English. *Journal of the Acoustical Society of America, 92*, 736-757.

Espy-Wilson, C. (1994). A feature-based semivowel recognition system. *Journal of the Acoustical Society of America, 95*, 65-72.

Fant, G. (1980). The relations between area functions and the acoustic signal. *Phoetica*, 55-86.

Fant, G. (1960) *Acoustic Theory of Speech Production*. Mouton, The Hague.

Feeney, M., & Franks, J. (1982). Test-retest reliability of a distinctive feature difference test for hearing aid evaluation. *Ear and Hearing, 3*, 59-65.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical method for grading the cognitive state for patients for the clinician. *Journal of Psychiatric Research, 12*, 93-99.

Freyman, R. L, Nerbonne, G. P., & Cote, H. A. (1991). Effect of consonant ratio

modification on the amplitude envelope cues for consonant recognition. *Journal of Speech and Hearing Research, 34*, 415-426.

Fry, D. B., Abramson, A. S., Eimas, P.D., & Liberman A. M. (1962). The identification and discrimination of synthetic vowels. *Lang. Speech, 5*, 171-189.

Furui, S. (1986). On the role of spectral transition for speech perception. *Journal of the Acoustical Society of America, 80*, 1016-1025.

Grimm, W. A. (1966). Perception of segments of English-spoken consonant-vowel syllables. *Journal of the Acoustical Society of America, 40*, 1454-1461.

Grosjean, F. (1996). Gating. *Language and Cognitive Processes, 11*, 597-604.

Hedrick, M. S., & Rice, T. (2000). Effect of a single-channel wide dynamic range compression circuit on perception of stop consonant place of articulation. *Journal of Speech, Language, & Hearing Research, 43*, 1174-118

Jamieson, D., Brennan, R., & Cornelisse, L. (1995). Evaluation of a speech enhancement strategy for normal and hearing impaired listeners. *Ear and Hearing, 16*, 274-286.

Jenstad, L.M., Barnes, S., & Hayes, D. (2008). Properties of the Distinctive Features Differences test for hearing aid research. International Hearing Aid Research Conference, Lake Tahoe, CA, 2008.

Jenstad, L. M., Seewald, R., Cornelisse, L., & Shantz, J. (1999). Comparison of linear gain and wide dynamic range compression hearing aid circuits: Aided speech perception measures. *Ear and Hearing, 20, 117-126.*

Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language,*

and Hearing Research, 48, 651-667.

Kennedy, E., Levitt, H., Neuman, A., & Weiss, M. (1998). Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. *Journal of the Acoustical Society of America, 103,* 1098-1114.

Kurowski, K., & Bluemstein, S. E. (1987). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America, 81,* 1917-1927.

Krull, D. (1990). Relating acoustic properties of perceptual responses: A study of Swedish voiced stops. *Journal of the Acoustical Society of America, 8,* 2557-2570.

Ladefoged, P. (2000). *A course in phonetics* (4th ed.). Boston: Thomson Wadsworth.

Lee. J (1997). The asymmetry of C/V coarticulation in CV and VC structures and its implications on phonology. *Studies in Linguistics, 27, 1997,* 139-152.

Lehiste, I. (1964). Acoustical characteristics of selected English consonants. Int. *J. Am. Ling., 30,* 1-197.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology, 52,* 127-137.

Lisker, L. (1957). Minimal cues for separating /w, r, l, y/ in intervocalic position. *Word, 13,* 256-67.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Percept. Psychophys., 25,* 457-465.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America, 73,* 1751-1755.

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some

English consonants. *Journal of the Acoustical Society of America, 27,* 301-315.

Modarresi, G., Sussman, H., Lindblom, B., & Burlingame, E., (2004). An acoustic analysis of the bidirectionality of coarticulation in VCV utterances. *Journal of Phonetics, 32*, 291-312.

Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95,* 1085-1099.

O'Connor, J., Gerstman L., Liberman, M., Delattre, P. & Cooper, F. (1957). Acoustic cues for the perception of initial /w, j, r, l/ in English. *Word, 13,* 24-43.

Ohman, S. E. G. (1966). Perception of segments of VCCV utterances. *J. Acoust. Soc. Am., 40,* 979-988.

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

Pickett, J. M., Bunnell, H. T., & Revoile, S. G. (1995). Phonetics of intervocalic consonant perception: retrospect and prospect. *Phonetica, 52,* 1-40.

Pols, L. & Schouten, M. (1978). Identification of deleted consonants. *Journal of the Acoustical Society of America, 64,* 1333-1337.

Schroeder, M. R. (1967). Determination of the geometry of the human vocal tract by acoustic measurements. *Journal of the Acoustical Society of America, 41*, 1002-1010.

Schwab, E. C., Sawusch, J. R., & Nusbaum, H. C. (1981). The role of second-formant transitions in the stop-semivowel distinction. *Percept. Psychophys. 29,* 121-128.

Smits, R., Warner, N., McQueen, J., & Cutler, A. (2003). Unfolding of phonetic

information over time: A database of Dutch diphone perception. *Journal of the Acoustical Society of America, 113,* 563-574.

Smits, R. (2000). Temporal distribution of information for human consonant recognition in VCV utterances. *Journal of Phonetics, 28,* 111-135.

Souza, P., Jenstad, L. M., & Folino, R. (2005). Using multichannel wide-dynamic range compression in severely hearing-impaired listeners: effects on speech recognition and quality. *Ear and Hearing, 26,* 120-131.

Stelmachowicz, P., Kopun, J., Mace, A., Lewis, D. E., & Nittrouer, S. (1995). The perception of amplified speech by listeners with hearing loss: Acoustic correlates. Journal of the Acoustical Society of America, 98, 1388-1399.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Strom, K. E. (2006). The HR 2006 dispenser survey. *Hearing Reviews, 13,* 16-38.

Studebaker, G. A. (1985). Research note: A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28,* 455-462.

Studebaker, G. A, McDaniel, D. M., & Sherbecoe, R. L. (1995). Evaluating relative speech recognition performance using the proficiency factor and rationalized arcsine differences. *Journal of the American Academy of Audiology, 6,* 173-182.

Titze, I. R. (1988). Regulation of vocal power and efficiency by subglottal pressure and glottal width. In O. Fujimura (Ed.), *Vocal fold physiology: Voice production, mechanisms, end functions* (pp. 227-238). New York: Raven.

Turner, C., & Robb, M. P. (1987). Audibility and recognition of stop consonants in normal and hearing-impaired subjects. *Journal of the Acoustical Society of America, 81,* 1566-1573.

Tyler, L., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception and Psychophysics, 38*, 409-420.

Valente, M., Hosford-Dunn, H., & Roeser, R. (2008). *Audiology treatment*. New York: Thieme.

Van Tasell, D. (1993). Hearing loss, speech, and hearing aids. *Journal of Speech and Hearing Research, 36*, 228-224.

Van Tasell, D., & Trine, T. (1996). Effects of single-band syllabic amplitude compression on temporal speech information in nonsense syllables and in sentences. *Journal of Speech and Hearing Research, 39*, 912-922.

Walley, A., Michela, V., & Wood, D. (1995). The gating paradigm: Effects of presentation format on spoken word recognition by children and adults. *Perception and Psychophysics*, *57*, 343-351.

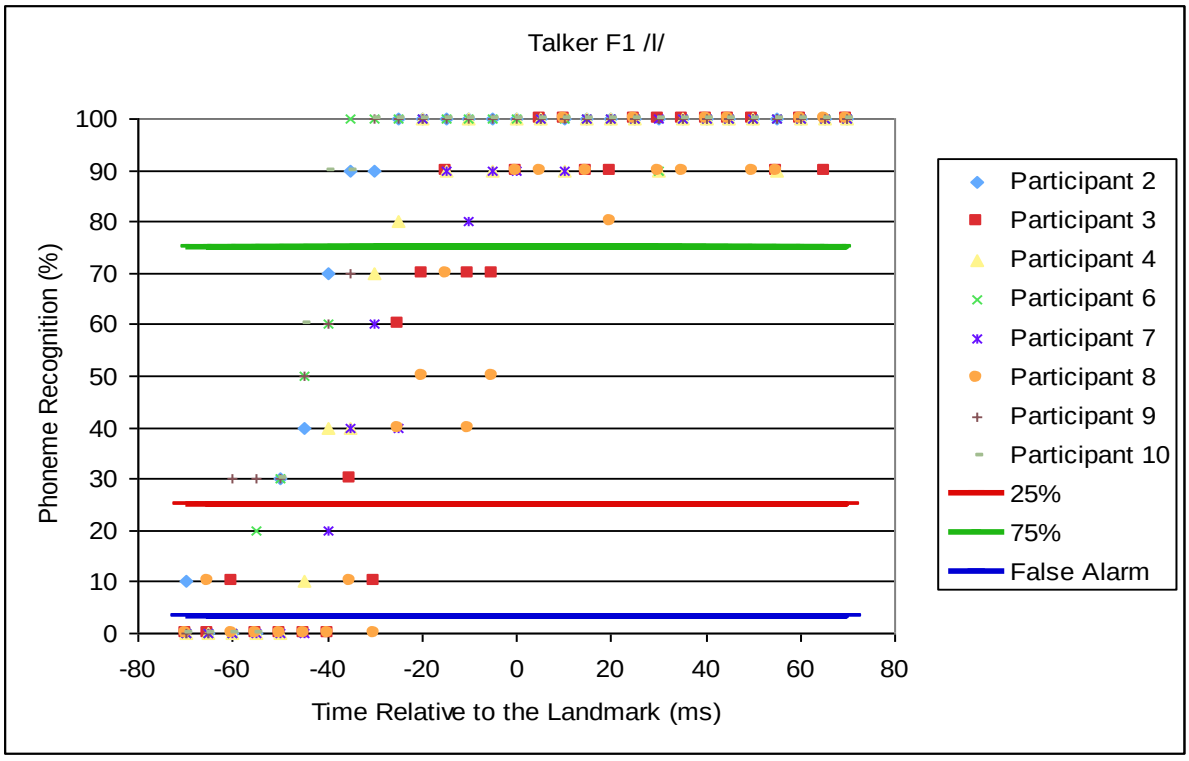
Appendix A

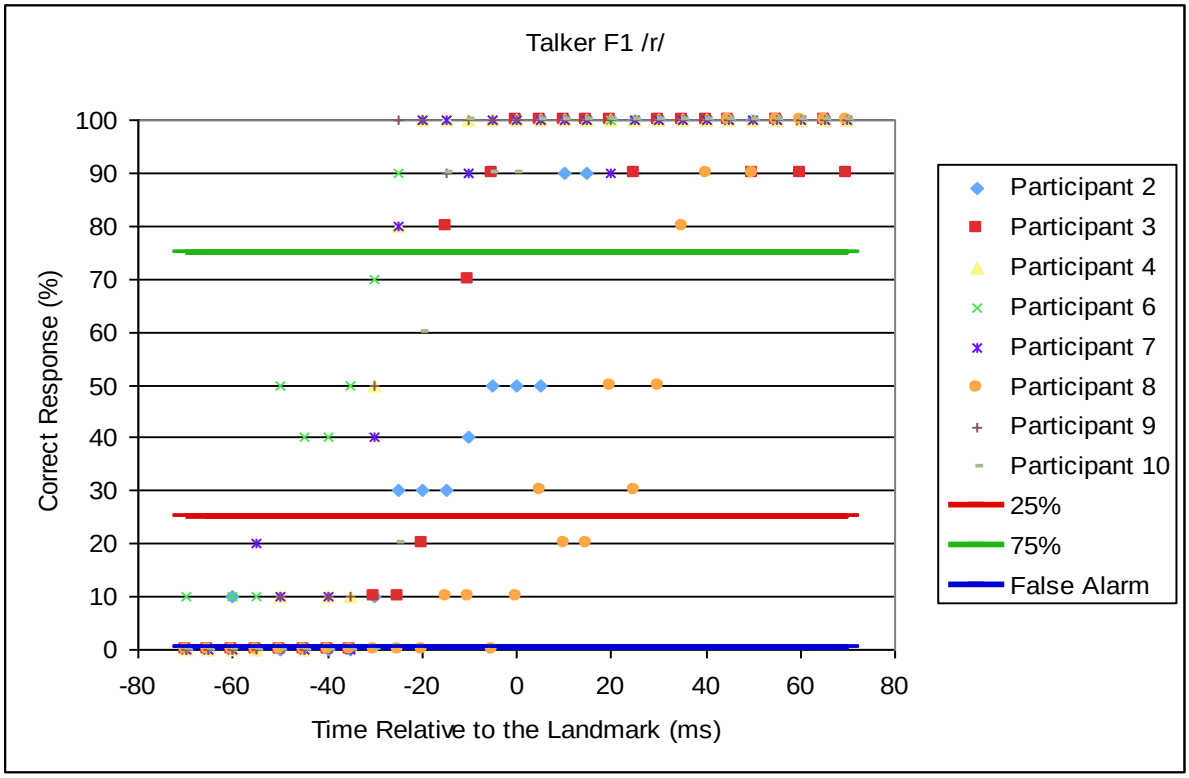
Praat Version 5.018 Parameters

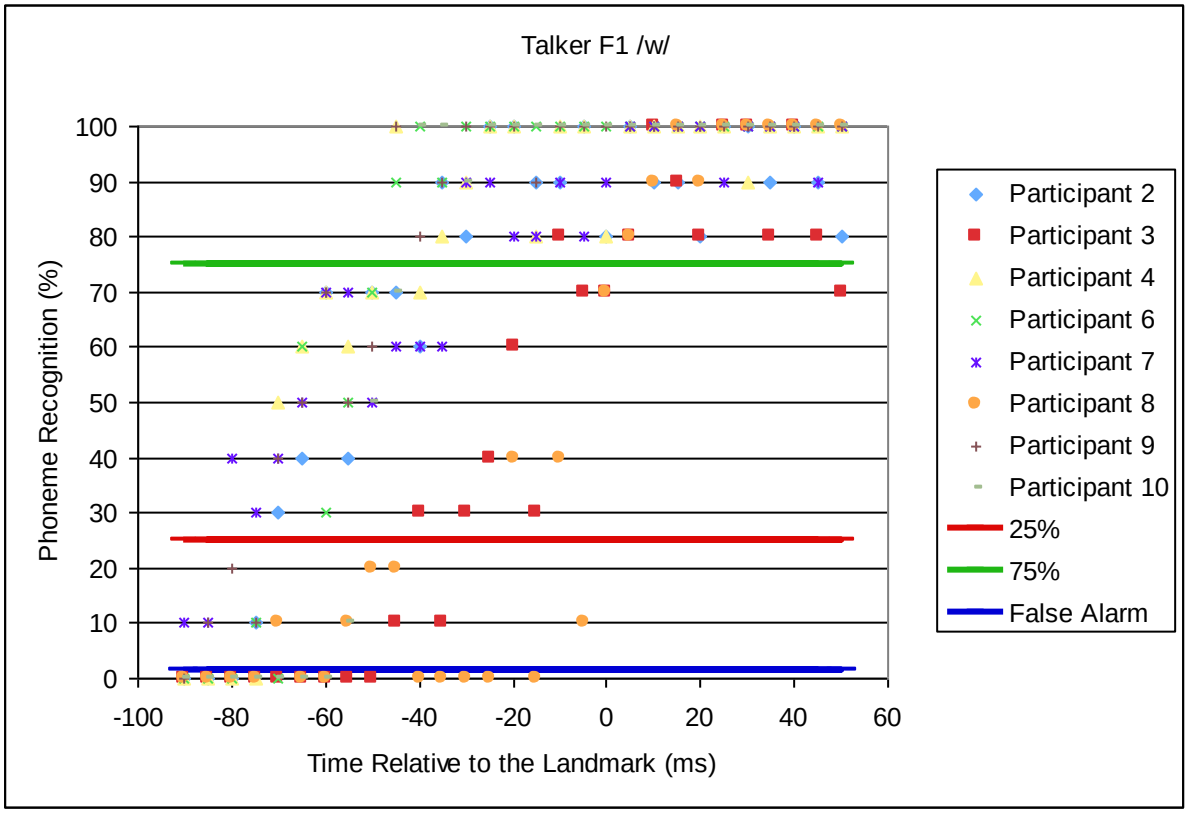
View	Sound autoscaling: yes Time step setting → time step strategy: fixed → fixed time step: 0.01
Spectrum	Spectrogram show: yes Spectrogram view from: 0 Hertz Spectrogram view to: 8000Hertz Spectrogram window length: 0.005 seconds Spectrogram dynamic range: 60 dB Spectrogram number of time steps: 1000 Spectrogram number of frequency steps: 250 Spectrogram method: Fourier Spectrogram window shape: Gaussian Spectrogram autoscaling: no Spectrogram maximum: 100 dB/Hz Spectrogram pre-emphasis: 0 dB/octave Spectrogram dynamic compression: 0
Pitch	Pitch show: Yes Pitch floor: 100 Hertz Pitch ceiling: 500 Hertz Pitch units: Hertz Pitch view from:0 Hertz Pitch view to: 0 Hertz Pitch method: Autocorrelation Pitch very accurate: no Pitch max. number of candidates: 15 Pitch silence threshold: 0.03 of global peak Pitch voicing threshold: 0.45 (periodic power/total power) Pitch octave costs: 0.01 per octave Pitch octave jump cost: 0.35 per octave Pitch voiced/unvoiced cost: 0.14 Hertz
Intensity	Intensity show: yes Intensity view from: 1 dB Intensity view to: 100 dB
Pulses	Pulses: no

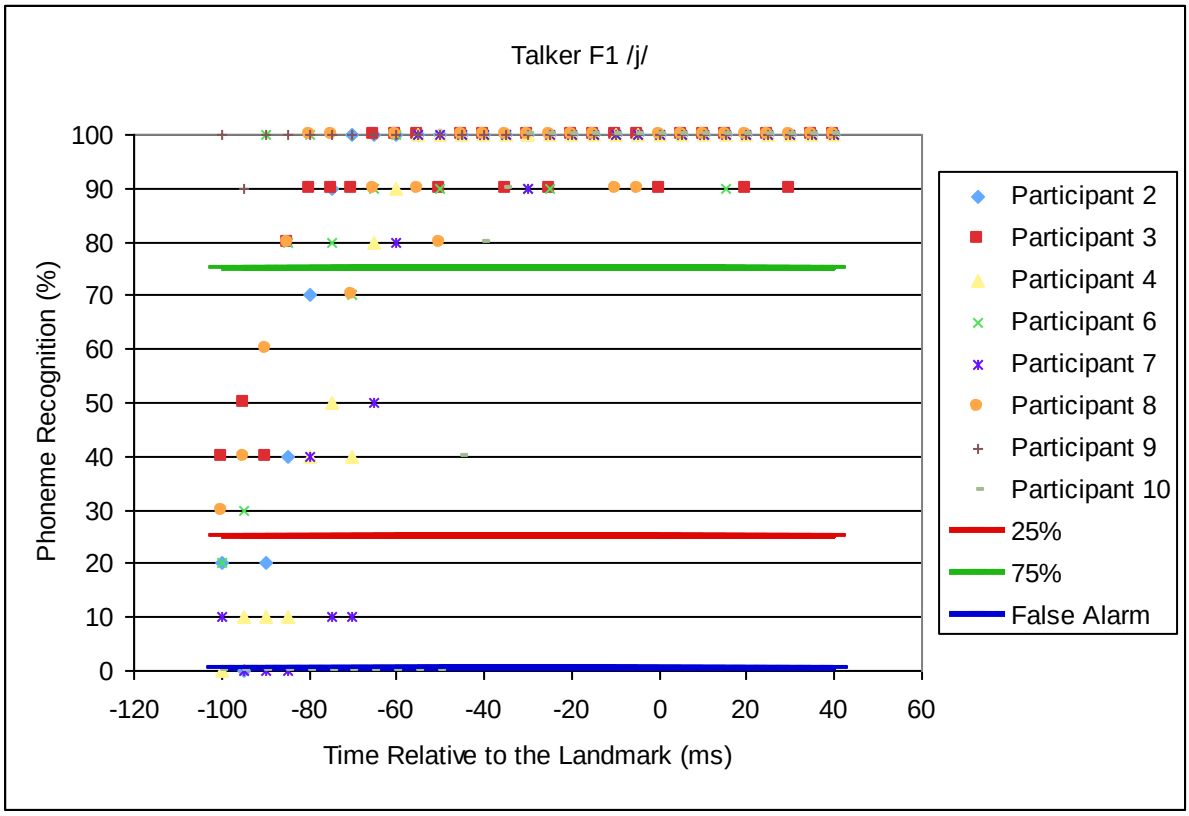
Appendix B

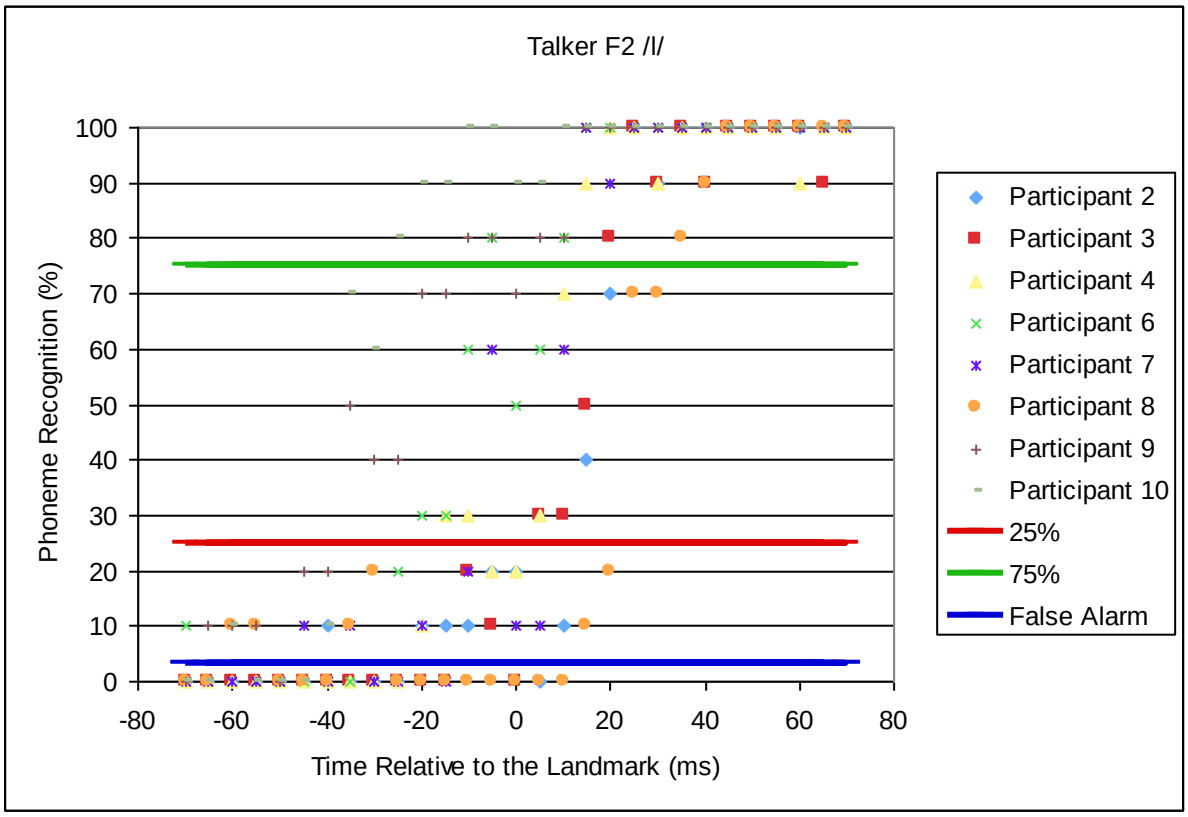
Phoneme Recognition for Individual Participants as a Function of Gating Time for UWODFD Semivowel Tokens

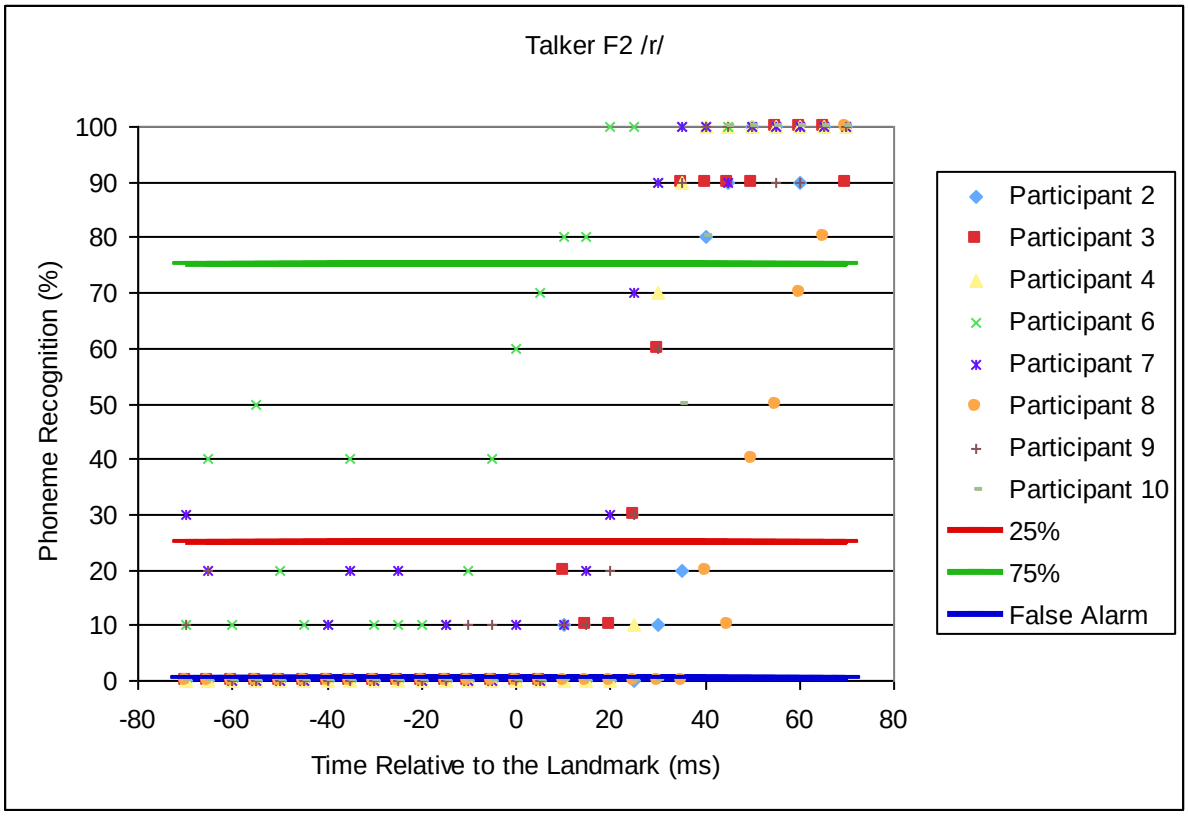


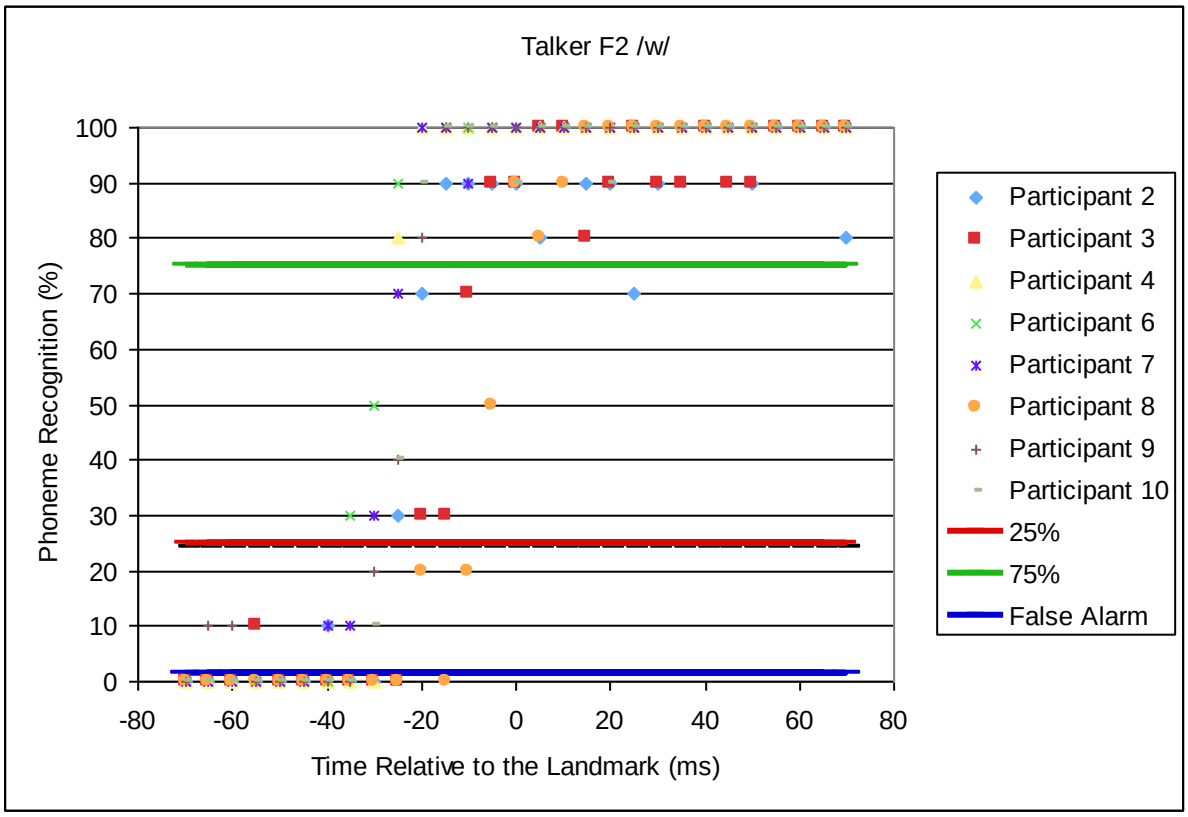

Talker F1 /r/
Correct Response (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

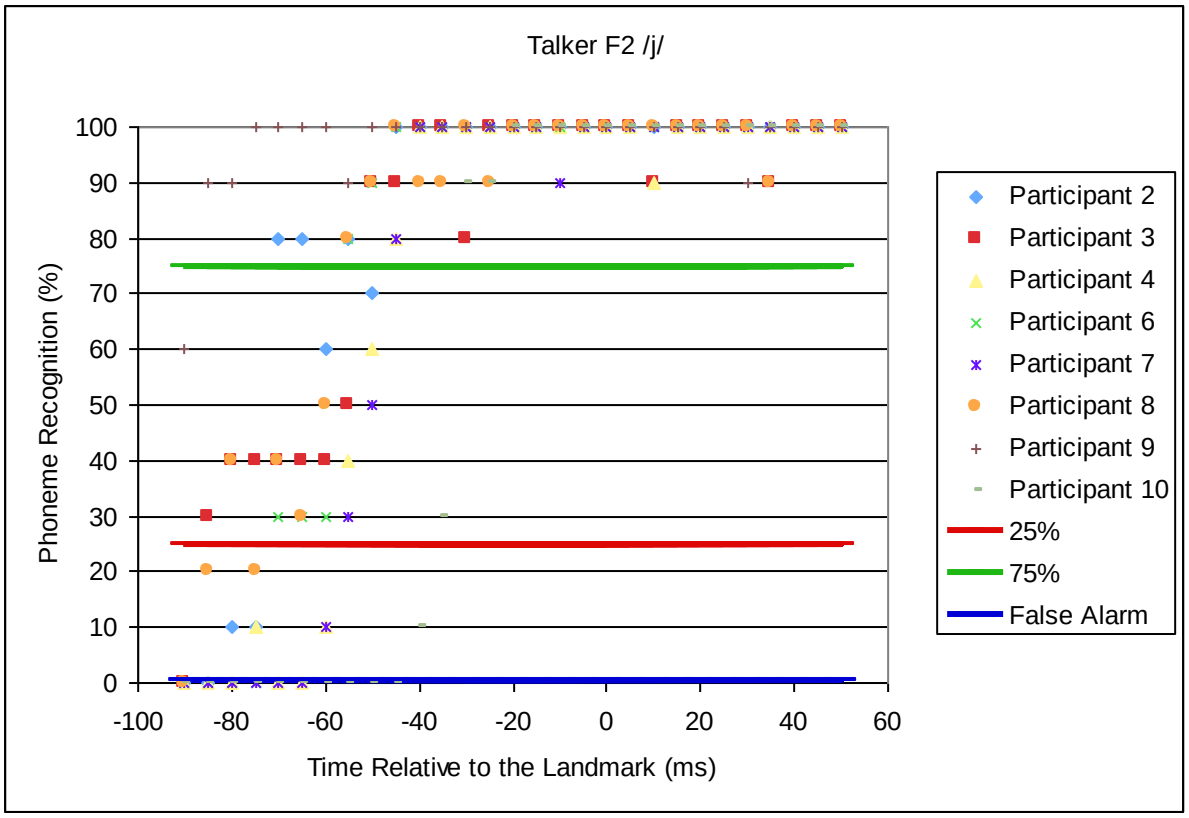

Talker F1 /w/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-100
-80
-60
-40
-20
0
20
40
60
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

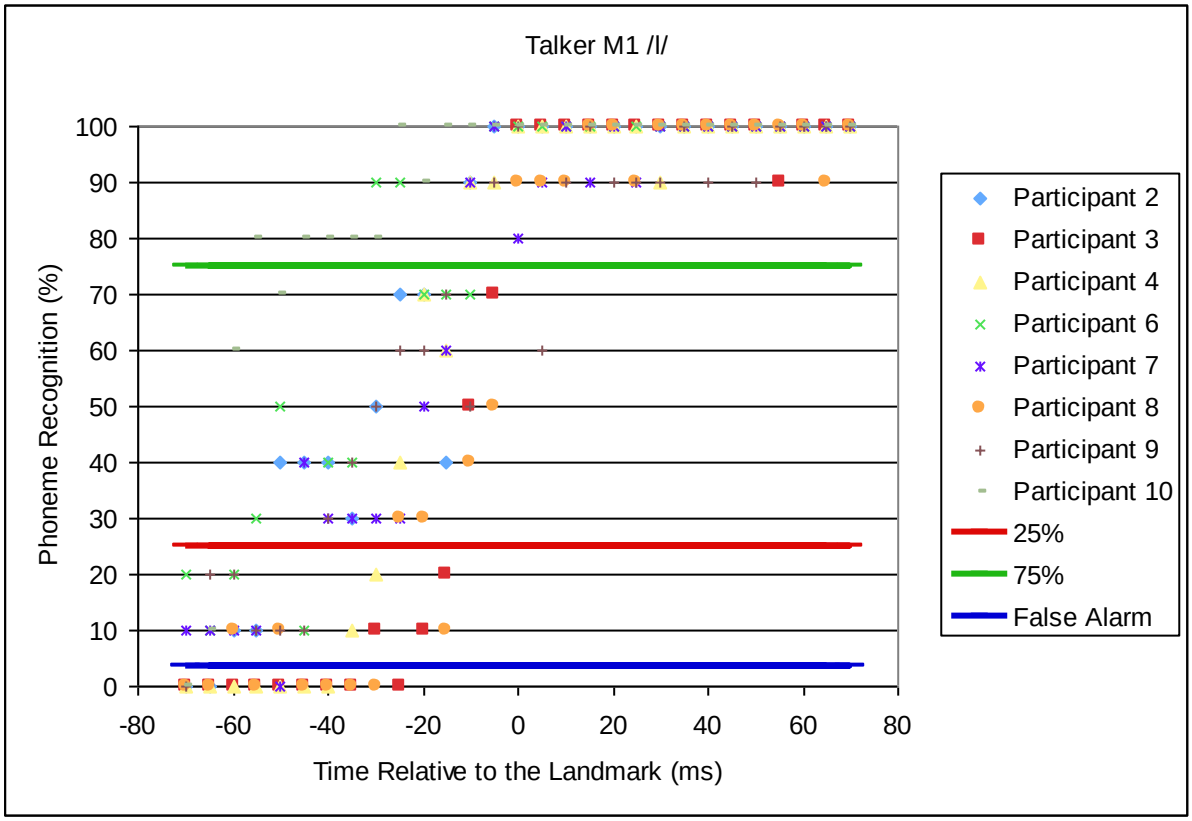

Talker F1 /j/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-120
-100
-80
-60
-40
-20
0
20
40
60
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

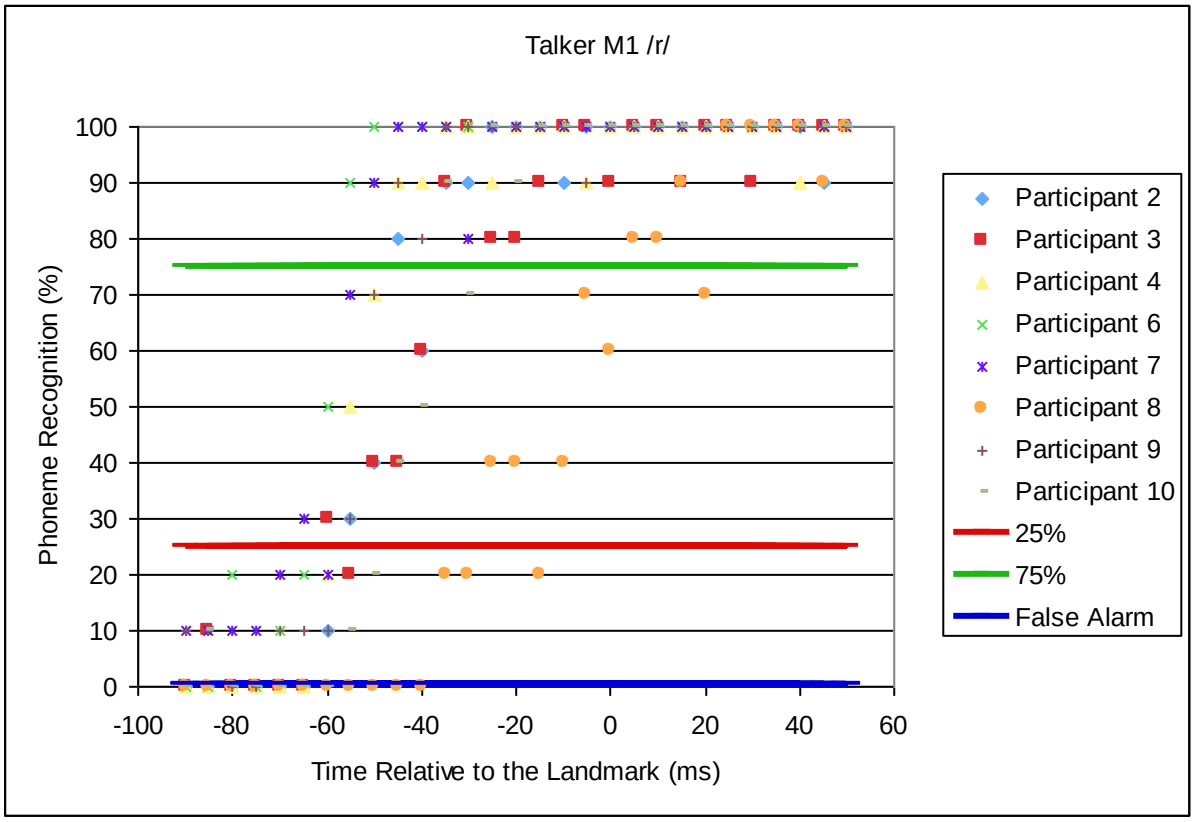

Talker F2 /l/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

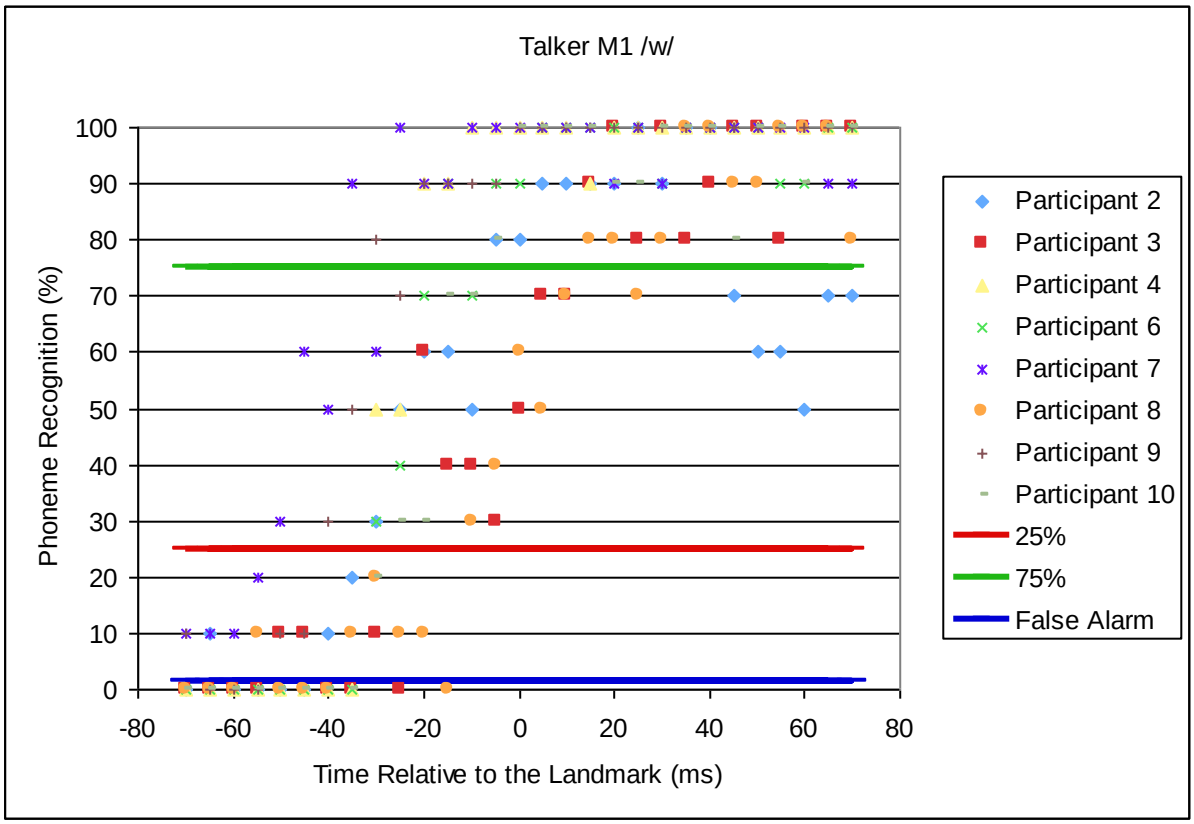

Talker F2 /r/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

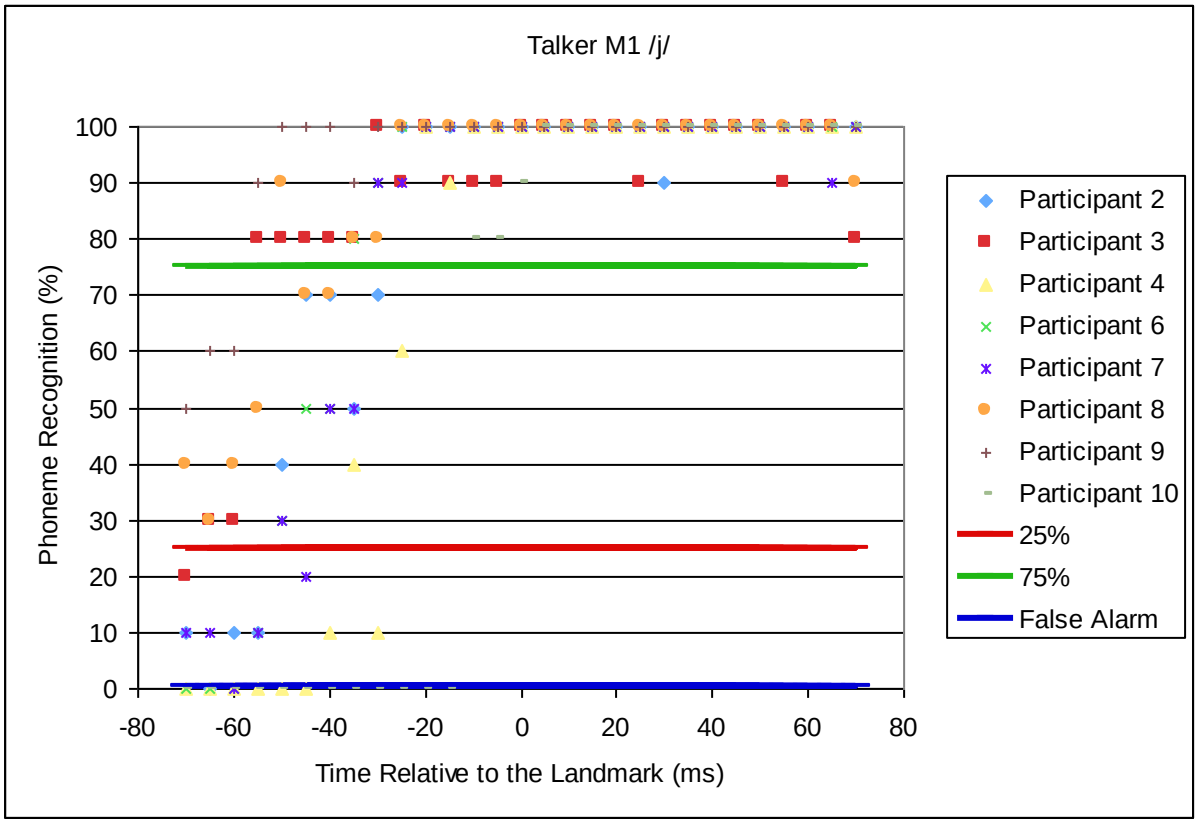

Talker F2 /w/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

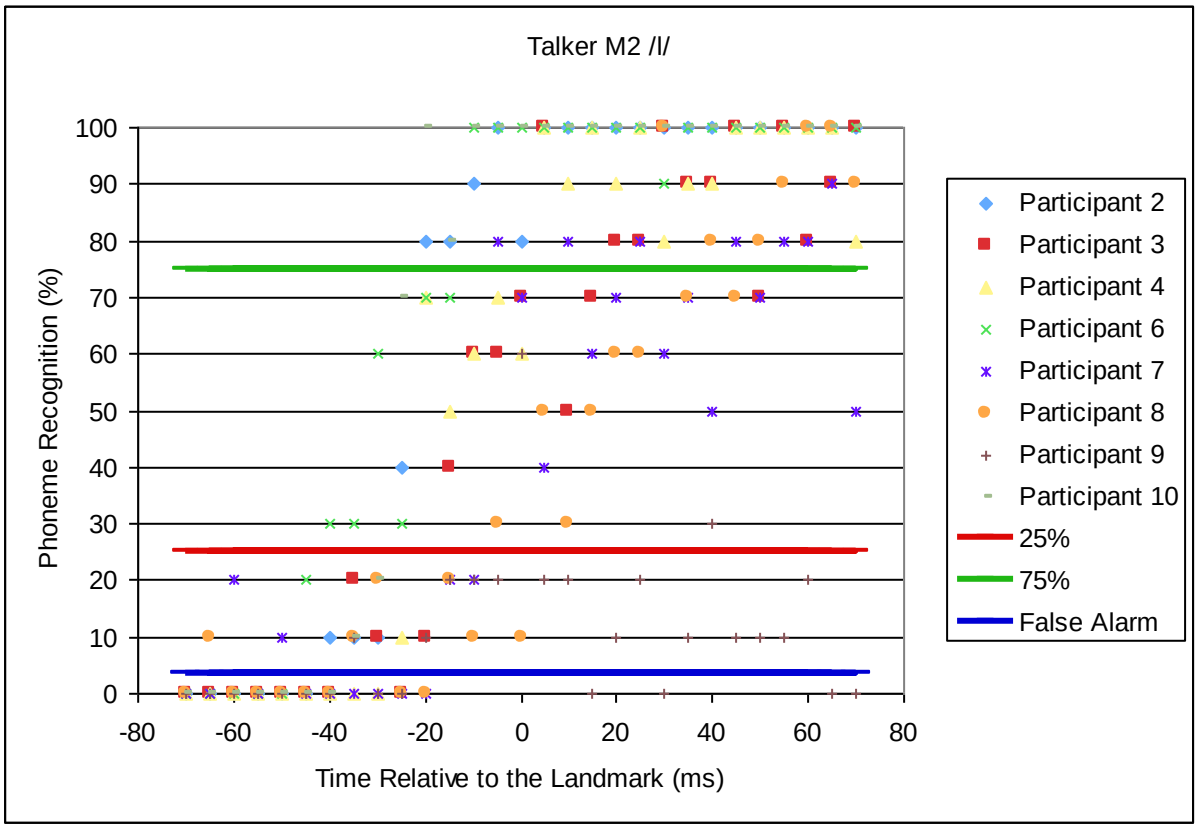

Talker F2 /j/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-100
-80
-60
-40
-20
0
20
40
60
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

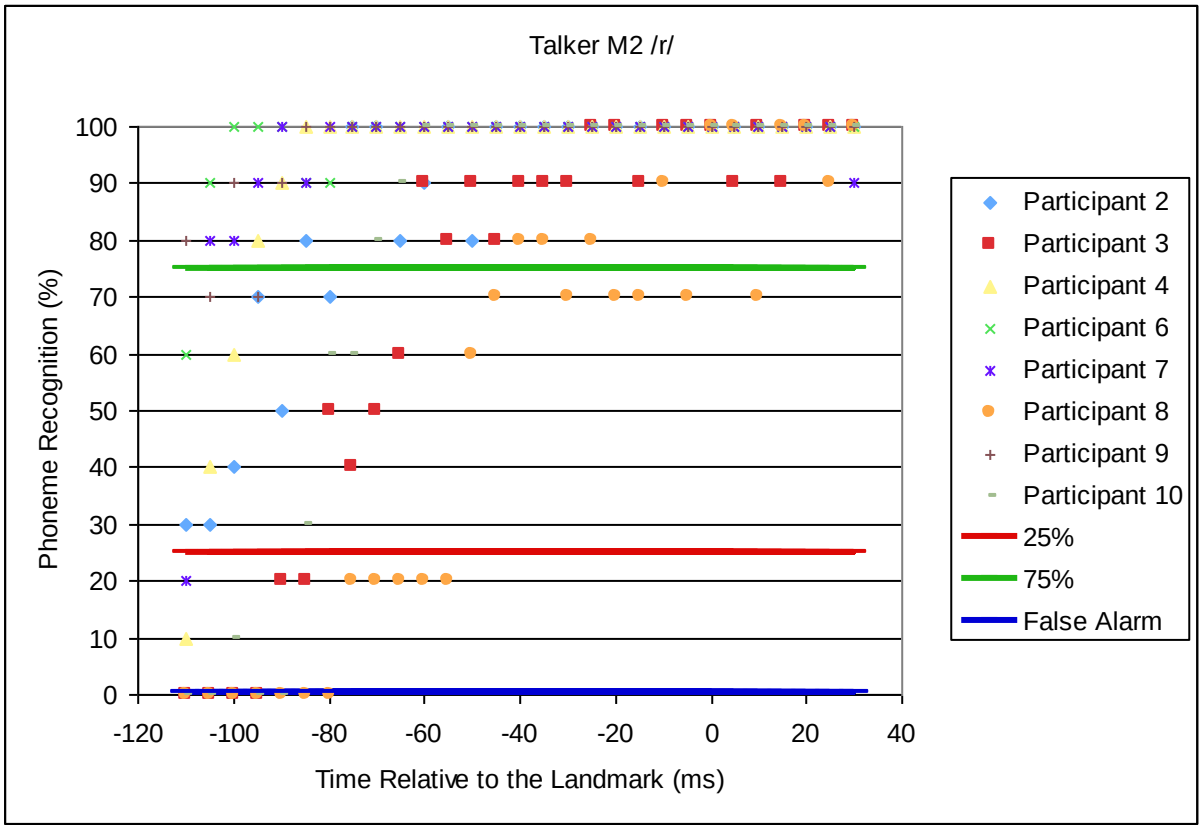

Talker M1 /l/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

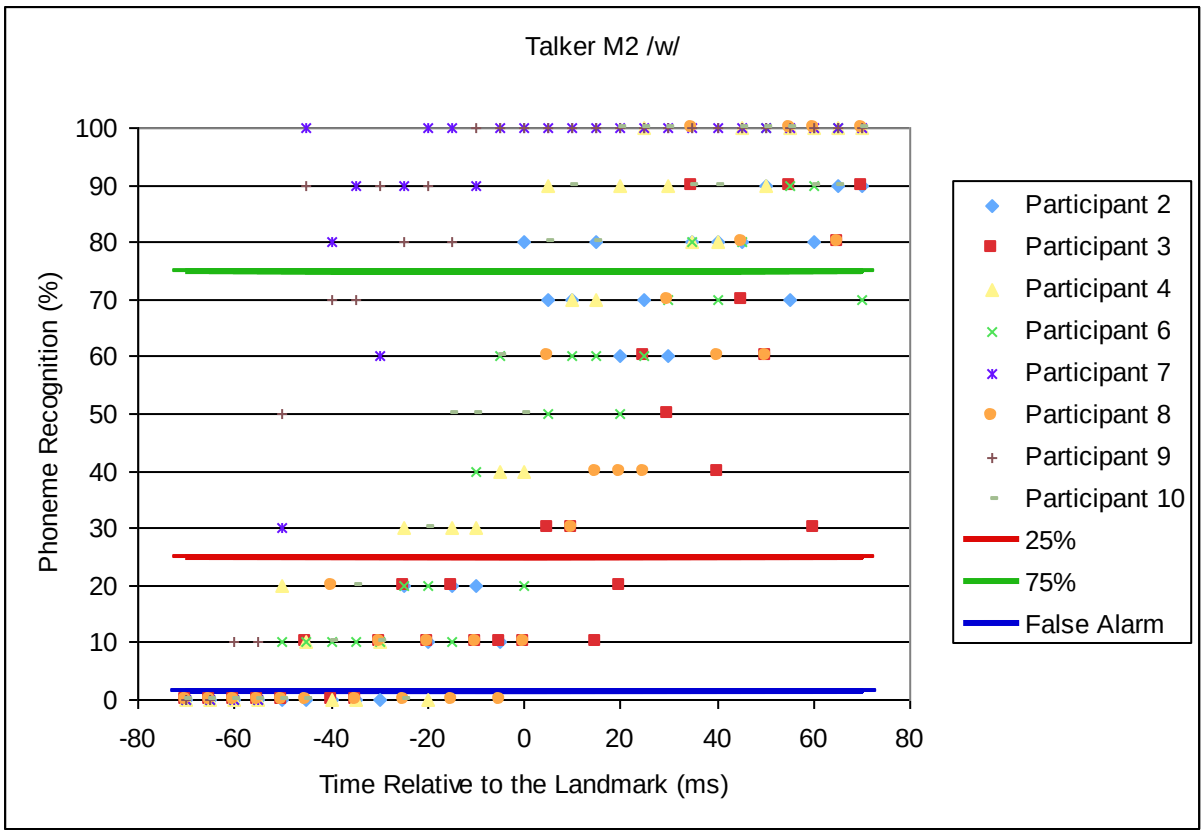

Talker M1 /r/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-100
-80
-60
-40
-20
0
20
40
60
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

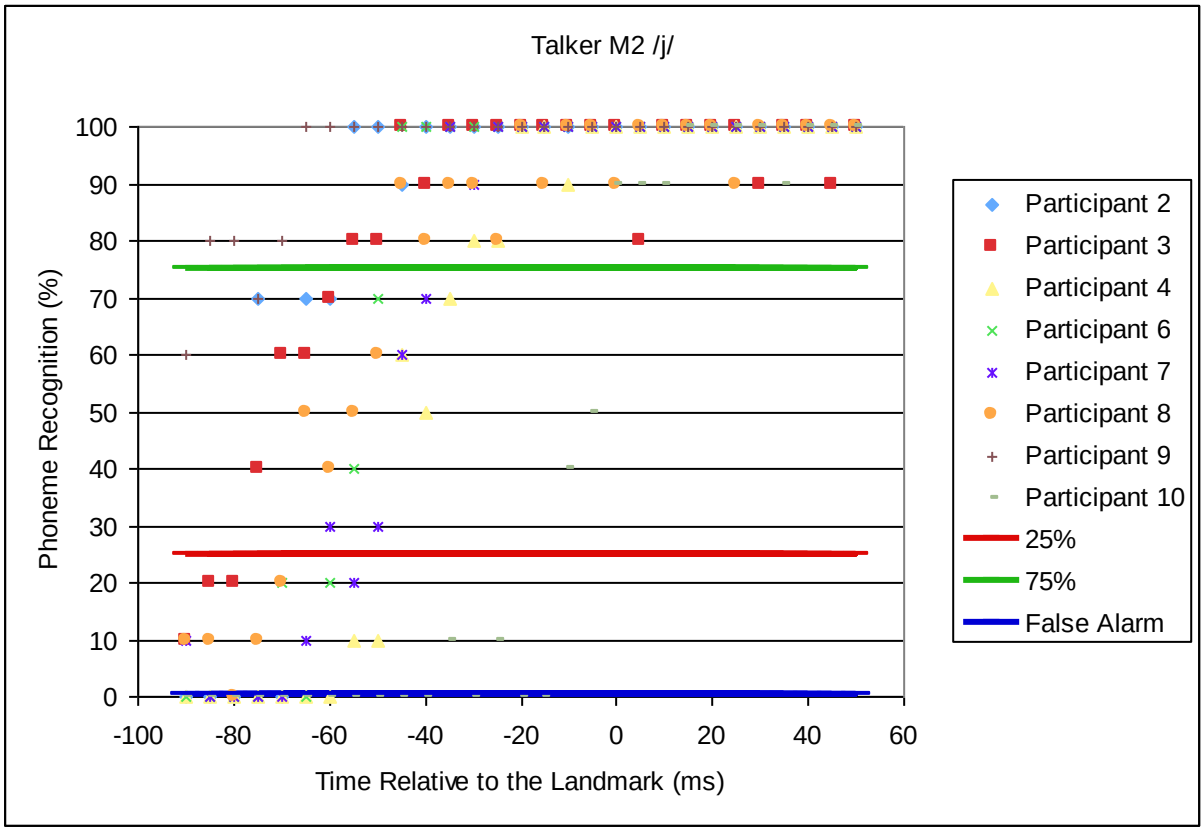

Talker M1 /w/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm


Talker M1 /j/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm


Talker M2 /l/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm


Talker M2 /r/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-120
-100
-80
-60
-40
-20
0
20
40
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm


Talker M2 /w/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-80
-60
-40
-20
0
20
40
60
80
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm


Talker M2 /j/
Phoneme Recognition (%)
100
90
80
70
60
50
40
30
20
10
0
-100
-80
-60
-40
-20
0
20
40
60
Time Relative to the Landmark (ms)
Participant 2
Participant 3
Participant 4
Participant 6
Participant 7
Participant 8
Participant 9
Participant 10
25%
75%
False Alarm

Appendix C

Formant Measurement Parameters

Talker	Sound	Formant Settings	Pitch Setting	Spectrogram Settings
F1	L	Max Formant: 5700 Hz # Formants: 5 Window Length: 0.025 sec Dynamic Range: 80 dB	Pitch Range: 100-300 Hz	View Range: 0-10,000 Window Length: 0.004 Dynamic Range: 80 dB
	R	Max Formant: 5700 Hz # Formants: 5 Window Length: 0.025 sec Dynamic Range: 80 dB		
	W	Max Formant: 5500 Hz # Formants: 5 Window Length: 0.025 sec Dynamic Range: 80 dB		
	Y	Max Formant: 5200 Hz # Formants: 5 Window Length: 0.025 sec Dynamic Range: 80 dB		
F2	L	Max Formant: 4900 Hz # Formants: 4 Window Length: 0.025 sec Dynamic Range: 80 dB	Pitch Range: 100-300 Hz	View Range: 0-10,000 Window Length: 0.004 Dynamic Range: 80 dB
	R	Max Formant: 3900 Hz # Formants: 4 Window Length: 0.025 sec Dynamic Range: 80 dB		
	W	Max Formant: 4600 Hz # Formants: 4 Window Length: 0.025 sec Dynamic Range: 80 dB		
	Y	Max Formant: 4600 Hz # Formants: 4 Window Length: 0.025 sec Dynamic Range: 80 dB		
M1	L	Max Formant: 5500 Hz # Formants: 5 Window Length: 0.025 sec Dynamic Range: 80 dB	Pitch Range: 50-200 Hz	View Range: 0-10,000 Window Length: 0.006 Dynamic Range 80 dB
	R	Max Formant: 4000 Hz # Formant: 4 Window Length: 0.025 sec Dynamic Range: 80 dB		

Talker	Sound	Formant Settings	Pitch Setting	Spectrogram Settings
M1	W	Max Formant: 4000 Hz # Formants 4 Window Length: 0.025 sec Dynamic Range: 80 dB	Pitch Range: 50-200 Hz	View Range: 0-10,000 Window Length: 0.006 Dynamic Range 80 dB
	Y	Max Formant: 4000 Hz # Formants: 4 Window Length: 0.025 sec Dynamic Range: 80 dB		
M2	L	Max Formant: 5200 Hz # Formants: 5 Window Length: 0.025 sec Dynamic Range: 80 dB	Range: 50-200 Hz	View Range: 0-10,000 Window Length: 0.006 Dynamic Range: 70 dB
	R	n/a		
	W	Max Formant: 5800 Hz # Formants: 6 Window Length: 0.025 sec Dynamic Range: 80 dB		
	Y	Max Formant: 5800 Hz # Formants: 6 Window Length: 0.025 sec Dynamic Range: 80 dB		

Appendix D

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - FULL BOARD

PRINCIPAL INVESTIGATOR:	INSTITUTION / DEPARTMENT:	UBC BREB NUMBER:
Lorienne Jenstad	UBC/Medicine, Faculty of/Audiology & Speech Sciences	H09-00144

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):

Christine Y.L. Joe

SPONSORING AGENCIES:

Natural Sciences and Engineering Research Council of Canada (NSERC)
UBC Faculty of Medicine

PROJECT TITLE:

Acoustic and Behavioral Effects of Hearing Aid Processing

REB MEETING DATE:	CERTIFICATE EXPIRY DATE:
May 14, 2009	**May 14, 2010**

DOCUMENTS INCLUDED IN THIS APPROVAL: — **DATE APPROVED: June 9, 2009**

Document Name	Version	Date
Consent Forms:		
Consent - Community	1	May 28, 2009
Consent - Student	2	May 28, 2009
Advertisements:		
Recruitment Flyer	2	May 28, 2009
Questionnaire, Questionnaire Cover Letter, Tests:		
Questionnaire	1	December 9, 2005
Letter of Initial Contact:		
Initial Contact	1	April 17, 2009

The application for ethical review and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board and signed electronically by one of the following:

Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Anita Ho, Associate Chair