EXAMINING LISTENERS' ABILITY TO PERCEIVE VOWEL-INHERENT SPECTRAL CHANGES

by

Kathleen L Chiddenton

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
March 2013

DALHOUSIE UNIVERSITY

SCHOOL OF HUMAN COMMUNICATION DISORDERS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "EXAMINING LISTENERS' ABILITY TO PERCEIVE VOWEL-INHERENT SPECTRAL CHANGES" by Kathleen L Chiddenton in partial fulfilment of the requirements for the degree of Master of Science.

Dated: March 22nd 2013

Supervisor: _____

Readers: _____

_____

_____

DALHOUSIE UNIVERSITY

DATE: March 22nd, 2013

AUTHOR: Kathleen L Chiddenton

TITLE: EXAMINING LISTENERS' ABILITY TO PERCEIVE VOWEL-INHERENT SPECTRAL CHANGES

DEPARTMENT OR SCHOOL: School of Human Communication Disorders

DEGREE: M.Sc. CONVOCATION: May YEAR: 2013

_____

Signature of Author

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

One family of theories regarding vowel perception suggests that onset and offset formant-frequencies are important for identification and that the shape of the transitions themselves are not otherwise perceptually important. The present study determined just-noticeable-differences in deviations from linear formant trajectories. Diphthong-like stimuli were manipulated by inserting a point of inflection into the otherwise linear transition. Several parameters were manipulated including vowel duration, location of the inflection point in time, direction of formant change, and fundamental frequency. Data from the first experiment indicate that listeners are largely insensitive to deviations from linearity of formant trajectory but that large enough deviations could eventually be detected. The size of these deviations seems dependent on the range of onset-offset formant frequencies. However, a second experiment in which only the first half of stimuli was presented thereby affecting the frequency range of the stimuli, gave different results. Results from these experiments along with several hypotheses are presented.

LIST OF ABBREVIATIONS USED

dB	decibel
f_0	fundamental frequency
F_1	first formant
F_2	second formant
Hz	hertz
ms	millisecond
VISC	vowel-inherent spectral change
CVC	consonant-vowel-consonant (eg. Bid)

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Dr. Michael Kiefte because without his support, knowledge, and guidance this research would not have been possible. Dr. Kiefte kept me motivated and invested himself fully in the outcome of these results. The invested time he put into this research is greatly appreciated.

In addition to a supportive supervisor, I was fortunate enough to have an equally enthusiastic committee consisting of Dr. Steven Aiken and Dr. Terrance Nearey. I thank them for their knowledge and support, as well as their overall interest in my results. They provided many helpful suggestions throughout the entire process that helped shape my thesis.

These results would not have been possible without the help of my participants. I cannot thank them enough for taking time out of their busy schedule to help out with my study.

I would like to thank my family and friends for their continued support and patience throughout these few years. This has been a stressful and time consuming process and I would never have made it through without so much support in my personal life. You all mean the world to me and I can't thank you enough.

This work was supported by a fellowship from the Social Sciences and Humanities Research Council of Canada.

CHAPTER 1 INTRODUCTION

The way in which listeners perceive vowels has been a debated topic since the earliest investigations of speech perception and production. Vocal-tract resonances, or formant frequencies, have been identified as the most important characteristic of a vowel in the identification process (see Kiefte et al., 2012; Rosner and Pickering, 1994, for reviews). Many studies have shown that the first two or three formant centre frequencies (F_1 and F_2) are the main spectral properties used for vowel identification. However, it remains unclear how human listeners track these formants in the ever changing environment of natural speech production.

In addition, there are many other acoustic properties involved in vowel production and researchers have looked into the effects these other properties may have on enhancing vowel identification when paired with formant frequencies. Some of these properties include global spectral shape and formant amplitude (Aaltonen, 1985; Bladon and Lindblom, 1981; Zahorian & Jagharghi, 1993). The extent to which any of these properties may affect vowel identification has been studied extensively (e.g., Bladon and Lindblom, 1981; Ito, et al., 2001; Kiefte and Klunder 2005; Kiefte et al., 2010). The difficulty in determining the importance of each property is that results are dependent on the specific environment and/or circumstances under which they were tested during the experiment. Nearey (1989) reminds us that none of these factors can be fully ignored when trying to create a complete model for vowel perception. Although the relative

importance of each factor is dependent on the circumstances in which it was tested, none of the factors can be discounted.

1.1 THE SIMPLE TARGET MODEL OF VOWEL PERCEPTION

The "simple target" model of vowel perception characterizes vowels as discrete points represented by the first two or three formants (e.g., Delattre et al., 1952). This implies that each vowel can be differentiated from one another based on the vocal tract configuration at a single point in time when the formant frequencies are particularly stable (Peterson, 1952, 1961). However, many pattern-recognition studies report better identification of vowels in models that incorporate the spectral change of the vowel rather than a measurement sampled at a single time (e.g. Nearey and Assmann, 1986). In addition, Hillenbrand and Nearey (1999) completed a study in which a pattern classifier was significantly more accurate when trained on two points drawn from the formant pattern of the vowel rather than a single point.

Hillenbrand et al. (1995) provided support for a formant theory of vowel perception in which the formant trajectories serve to better discriminate English vowels in a crowded acoustic space. This is shown similarly for both the traditionally dynamic diphthongs such as /eɪ / as well as monophthongs such as /ɪ /. Given the crowded vowel space in English, if vowels are plotted by their center frequency many vowels overlap or are found to be close to one another. Hillenbrand et al. (1995) were able to show that the formant transitions of these close vowels were unique and therefore helped to distinguish them

from one another. Additionally they concluded that if listeners were to attend only to a single, relatively stable point of the vowel they would have difficulty identifying silent-center vowels which requires listeners to attend to brief portions of both the onglide and offglide of a vowel. With only brief onglides and offglides, Nearey and Assmann (1986) were able to obtain high vowel identification rates among listeners, showing the importance of the formant endpoints.

Hillenbrand et al (2001) looked into the dynamic changes associated with speech to determine if vowel identity was maintained in a varying consonant environment. Using both one and two-point classification models, they found similar results to their research in 1995, showing that more accurate identification occurs when focused on two points (incorporating spectral change) even when the vowel may be influenced by the consonant environment. They did not claim that consonant-vowel coarticulation does not play a role in identifying vowels but they do suggest that a simple two-point model (similar to what has been used in isolated vowel identification) may be adequate to classify them.

1.2 VOWEL IDENTIFICATION BASED ON BROAD SPECTRAL SHAPE

Despite the extensive work on identifying the important aspects of formants in speech, many researchers still feel that formant theories of vowel perception have not been conclusively shown to be fully adequate. Several alternative models have been proposed based on the justification that predicting vowel identification based on formant

frequencies alone has not been shown (see Kiefte, et al., 2012; Rosner and Pickering 1994 for reviews).

As opposed to assessing the specific properties of a vowel that are most important in perception, some researchers have argued that we should assess the broadly defined global spectrum of the vowel (Bladon and Lindbolm, 1981). The auditory spectrum shape contains a more complete vowel description and may therefore be a more important feature in vowel perception. Bladon and Lindbolm (1981) indicate that changes in formant frequencies or spectral peaks lead to changes in spectral shape and that this can in turn explain why experimental results may appear to favour spectral peaks. Zahorian and Jagharghi (1993) hypothesized that the complete spectral shape may provide more information and therefore lead to more accurate vowel discrimination when compared to formants in an isolated CVC word. Their results showed that both the formants and the global spectral shape were important and adequate features for vowel identification but that using both is redundant. The advantage the authors noted in using formants for identification was that a significant amount of information can be carried in relatively few features. Zahorian and Jagharghi did find that the global spectral shape was able to provide a more complete description of the vowel, but when focus is on identification either was sufficient. Their results indicated that static spectral cues are more important than temporal cues but that trajectories are an important secondary cue.

1.2.1 SECONDARY SPECTRAL FEATURES

Overall, speech-perception research has assumed that the first three formant frequencies are most important in vowel identification, whereas secondary spectral features such as formant bandwidth, spectral tilt, and formant amplitude are seen as non-essential pieces of information (e.g., Klatt, 1982; Kiefte et al., 2012). However, Ito et al. (2001) found that when either the first or second formant is removed from the vowel spectra, vowel identification rates for the five Japanese vowel categories are still high, which led them to believe that formants are not essential features in vowel perception. In 2005, Kiefte and Klunder published a paper to further this research because, as they pointed out, Ito et al. had removed the formant after the stimuli had been synthesized normally. This meant that the spectral tilt, or relative balance between high and low frequency, which is highly dependent on the missing formant (F_1 or F_2), remained unchanged and continued to provide information about the missing formant. In order to explain results by Ito et al., Kiefte and Klunder (2005) suggested that in the absence of F_2, the spectral tilt is used by listeners to identify the vowel. Therefore, secondary spectral features appear to have an effect on vowel perception when the primary features such as formants are either removed or distorted.

Researchers continue to question whether or not other spectral features such as formant amplitude can provide additional information necessary for accurate vowel perception. Formant amplitudes and bandwidth can both be largely deduced by the formant frequencies due to a predictable relationship that has been observed between them (Fant, 1956). Given this, it seems that formant amplitude would provide no additional

information needed for vowel identification that was not already provided by the formant frequencies alone. Jacewicz (2005) tested the effects that formant amplitude variation could have on overall vowel identification and/or perception. The main conclusions from the study were that formant amplitude impacts vowel naturalness which could in turn affect proper identification. Those vowels that sounded less natural were more likely to be mis-identified than vowels sounding more natural to the listener.

Kiefte et al. (2010) found further evidence to suggest that formant amplitude does play a role of vowel perception. By manipulating formant amplitudes and introducing spurious formants, researchers were able to asses when listeners heard the vowel /i/ and when they heard /u/. Based on their results, the researchers suggested three related factors of the formant amplitude that contribute to vowel identification. The first is simultaneous masking of spectral peaks in which formant amplitude manipulation can lead to stronger peaks masking the effects of the formants at lower amplitudes. This can lead to misperceiving higher formants as a lower one. For example, if F_2 peak is inaudible because of the masking effect from F_1, F_3 may mistakenly be perceived as F_2 resulting in misidentification of the vowel. The second factor is local spectral contrast which was described above as the overall shape of spectral amplitude peaks as a degree of prominence in relation to the nearby spectral energy. The final factor described was auditory segregation. Authors suggested that amplitudes falling out of range of some predetermined threshold, which was based on experience with natural speech and the variations between formant frequency and amplitude, are segregated from the vowel itself and essentially ignored (Kiefte et al., 2010). Although a significant body of evidence

supports the claims that formants are the primary feature necessary for vowel identification, many researchers have been able to show that other spectral properties can also play an important secondary role.

1.3 VOWEL INHERENT SPECTRAL CHANGE (VISC)

In the English language, vowels have been traditionally categorized as either monopthongs or diphthongs. It has long been established that diphthongs are characterized by changing formant frequencies over time, separating them from the more stable monophthongs. However, there is a growing volume of research that has observed some traditional monophthongs in English to resemble diphthongs in that they too show vowel inherent spectral changes (VISC; Nearey and Assmann, 1986; Hillenbrand et al., 1995). This change can be best described as a formant trajectory or frequency change in the F_1-F_2 space. The VISC noted in monophthongs such as /ɪ / can be as large in magnitude as that found in diphthongs such as /e/and it appears to play an important role in vowel identification and discrimination (e.g., Assmann and Katz, 2005; Hillenbrand et al., 2001). Many researchers have found that listeners' vowel identification rates change when presented with stimuli that have typical formant trajectories as opposed to flat or reversed trajectories (Nearey and Assmann, 1986; Hillenbrand and Nearey, 1999).

Three hypotheses on how VISC is perceived by listeners have been proposed; the onset-offset hypothesis, the slope hypothesis, and the direction hypothesis (Nearey and Assmann, 1986; Gottfriend et al., 1993; Morrison and Nearey, 2007). The onset-offset

hypothesis suggests that the formant frequencies of two vowel targets, one near the onset and one near the offset of the vowel, are necessary spectral properties in vowel perception (Nearey and Assmann, 1986). The slope hypothesis claims that perceptual cues are based on the initial target frequency and the rate of change over time for the formant frequencies, irrespective of the offset frequencies. Interestingly, this hypothesis suggests that the exact point of the vowel offset is not important and that it is the speed of the vowel transition rather than the duration of the vowel which influences perception. Similar to the slope hypothesis, the direction hypothesis suggests that the initial vowel frequency and the direction of formant movement is the relevant piece of information, irrespective of vowel duration, speed, or the exact point of vowel offset. These hypotheses differ in which aspects of VISC are relevant for perceiving vowels. All three are in agreement that the initial formant frequency is important in identifying vowels but they differ on which additional information is necessary. Several studies have been conducted to evaluate which of these hypotheses can better represent how listeners perceive VISC. Results in favour of the onset-offset hypothesis were presented first by Strange et al. (1983) and again by Nearey and Assmann (1986) using silent-centre vowel stimuli to determine how identifiable the vowels were. With only brief onglides and offglides, high identification rates were still obtained. Additionally, Strange et al. (1983) varied the duration of the silent centre and they argued that with this change, as well as speaker variability, static features of the vowel become ambiguous and the essential features for identification lie within the dynamic spectral features impacted by coarticulation.

1.3.1 LISTENERS SENSITIVITY TO VISC

The onset-offset hypothesis has been shown by many researchers to be the most likely explanation for how we identify vowels and this suggests that we are not attending to the centre of the vowel and are therefore not perceiving the entire transition (e.g., Nearey and Assmann, 1986; Morrison and Nearey, 2007). Kewley-Port and Goodman (2005) felt it was reasonable to hypothesize that these dynamic formant frequencies cannot be used as cues for vowel identification if listeners are unable to perceive them. Even though every vowel in English has been shown to be dynamic, monophthongs are often characterized as "quasi-steady state" because they possess more limited formant change than diphthongs. This limited change in frequency could be too small for auditory perceptual abilities making it seem as though we are not attending to the changes.

Kewley-Port and Goodman studied the threshold for discrimination of formant change in the vowel nucleus (2005). They compared the threshold of detection with the amount of formant change observed in natural speech and concluded that the vowel transitions in speech measured at least four times larger than the threshold. Researchers also found that perception thresholds were smaller for vowels that showed more dynamic change. A possible explanation for these results presented by the researchers was that when vowels can be identified more easily by their dynamically changing formants, we become more perceptive to the changes present.

Liu et al (2012) compared English with Chinese listeners, given that Mandarin Chinese has a much less crowded vowel space and therefore more distance between individual

vowels than English does. They predicted that the sensitivity of listeners in these two languages to formant transitions would differ based on the size of the vowel space they are accustomed to listening to. In order to measure sensitivity, the researchers were looking for thresholds of vowel-formant discrimination defining this as the smallest change of formant frequency that is detectable. Results indicated that English listeners were more sensitive to formant changes in both English and Mandarin vowels in that they had smaller thresholds. Liu et al. concluded that this was in agreement with their hypothesis indicating that English speakers are more sensitive to formant changes because of the small space in which all of the English vowels must be contained.

1.3.2 FLEXIBILITY TO PERCEIVE INCOMPLETE/IMPRECISE FORMANT TRANSITIONS

With such an acute sensitivity to changes in formant frequency, it seems difficult to fathom that listeners can recognize vowels consistently in everyday speech given the high variability among speakers. Formant transitions rarely reach their target in connected speech, and yet the perceptual system seems to somehow compensate for these vowel reductions by perceptually overshooting the offset frequency of the vowel (e.g., Divenyi, 2009). Within natural speech, a single vowel can vary based on age and gender of the speaker, speaking rate, and coarticulation factors. As described above, English vowels reside in a closely packed vowel space and therefore a variation in formant frequency of one vowel can easily push it into the frequency domain of another, and yet listeners perceive them accurately (Divenyi, 2009). This ability shows the true flexibility of the perceptual system and further complicates the understanding of how we can perceive

vowels as different from one another. In trying to understand how listeners perceive VISC, we must establish conditions under which vowel perception becomes altered and under which conditions the change is not significant enough to be detected.

1.3.3 PERCEPTUAL TRACKING OF VISC OVER TIME

There is a large amount of evidence suggesting that the dynamic properties of the vowel are necessary for identification (Nearey and Assmann, 1986; Gottfriend et al., 1993; Morrison and Nearey, 2007). Formant frequencies change over the duration of the vowel and these are thought to be the most important pieces of information involved in vowel perception. However, it has never been conclusively shown that listeners perceptually follow formant frequencies across time. Most results have been in favour of the onset-offset hypothesis for vowel identification such that formant tracking over time is not actually required.

Morrison and Nearey (2007) reported a study to determine which aspects of the formant change in vowels were relevant in perception. They used three types of synthetic vowels to differentiate among the three different VISC hypotheses: Formant trajectories were either straight transitions across the duration of the vowel, elbowed with an initial steady state for the first quarter of the vowel duration followed by gliding formant transitions, or constant with no formant-frequency changes throughout the vowel duration. Their results provided support for the onset-offset hypothesis indicating the perceptual importance of the initial and final formant frequency with little attention being given to the changes in between by listeners – i.e., listeners identified the straight and elbowed formant

trajectories as the same vowel. Interestingly, the introduction of a bend relates closely to the perception of silent centered stimuli shown by Nearey and Assman such that the bend and the silent center had no effect on the perception of the vowel or its identification (1986). Further, anecdotal evidence suggested that listeners were not even able to detect differences between these stimuli suggesting that they were not at all sensitive to the formant manipulations used to test vowel identification.

1.4 JUSTIFICATION FOR THIS STUDY

With a significant body of research pointing to the importance of the dynamic properties of formant transitions in vowels for identification but no supporting evidence that listeners track these changes over time, it is important to test the threshold at which listeners can perceive formant changes over time. By extending the research done by Morrison and Nearey, we tested the limits of the onset-offset theory. One can assume that there is a point at which listeners must be able to detect a change in formant frequency in the middle of the vowel if it deviates from the straight trajectory sufficiently. By examining the threshold at which listeners detect difference in the stimuli in which the deviation of the elbow differs, we may gain some insight into the perceptual processes involved in formant perception and tracking. A better understanding of how listeners perceive important characteristics of vowels can also offer a better understanding of speech perception in general.

In addition, although listeners may not have been sensitive to the manipulations used in the study by Morrison and Nearey (2007), it is useful to know the smallest deviation from a straight formant transition that is detectible by listeners. This information would help to evaluate the practical limits of the theories that have been proposed for vowel perception.

In this thesis, Chapter 2 will outline the procedures and results obtained from the initial experiment, following which, a brief discussion of these results will be provided. As shown, these results brought forth more questions, requiring a second round of experimentation. This second experiment is outlined in chapter 3 with procedures, results, and a brief discussion being outlined. Finally, chapter 4 will provide a more in depth discussion in which the results of both experiments are combined to produce an overall conclusion to this study. Suggestions for future research will also be discussed at this time.

CHAPTER 2 EXPERIMENT 1

2.1 METHODS

2.1.1 PARTICIPANTS

Fifteen participants were recruited from the School of Human Communication Disorders at Dalhousie University. All subjects were 18 or older and had normal hearing as determined by an audiometric hearing screening performed on the day of participation. As is customary with hearing screenings, participants were tested at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz in both ears. Normal hearing was defined as having thresholds of 25 dB or less at all frequencies tested.

2.1.2 APPARATUS

Participants were seated in an IAC double-walled sound-treated booth and were fit with Beyer-Dynamic DT 150 circumaural headphones. All stimuli were presented via an Edirol UE-25EX external AD/DA attached to an iMac. Presentation of the stimuli was controlled by PsychoPy.

2.1.3 STIMULI

The experiment was a three-alternative forced-choice 1-up 3-down adaptive staircase design which converges on the 0.841 probability of the psychometric function (Levitt, 1971). Stimuli were presented in a series of 3 vowel-like sounds in succession. The base stimulus, from which all others were derived, was 100 ms long and the vowel was

modeled after the diphthong /ɛɪ / as in 'hay'. All stimuli were generated using a MATLAB implementation of the KLATT80 speech synthesizer (Kiefte et al, 2002; Klatt, 1980). All synthesis parameters were maintained at their software defaults as described by Klatt with the exceptions described below.

In the base condition, F_1 onset frequency was set at 425 Hz and offset frequency was 394 Hz. Second formant, F_2 transitioned from 1900 Hz at onset to 1971 Hz at offset. For the deviant stimuli, an inflection point or bend was introduced at 25% of the vowel duration or 25 ms (see Figure 1). The deviation of this frequency from the linear interpolation between onset and offset was varied until responses converged on a detection threshold. Fundamental frequency (f_0) was set at 125 Hz at onset and fell to 88 Hz at offset. However, several other series were also tested (see Table 1 for testing parameters).

Bends were inserted into the stimuli at either 25% of the vowel duration or at 50% depending on the condition. Many stimuli were synthesized such that this deviation became larger or smaller from the original trajectory based on the participant's response in an effort to determine their threshold (point at which they are just able to perceive a difference in the stimuli). These deviations brought the second formant down below its original position at that point in time (see Figure 1). The step size, whether it was moving the formant deviation up or down, was 5 Hz.

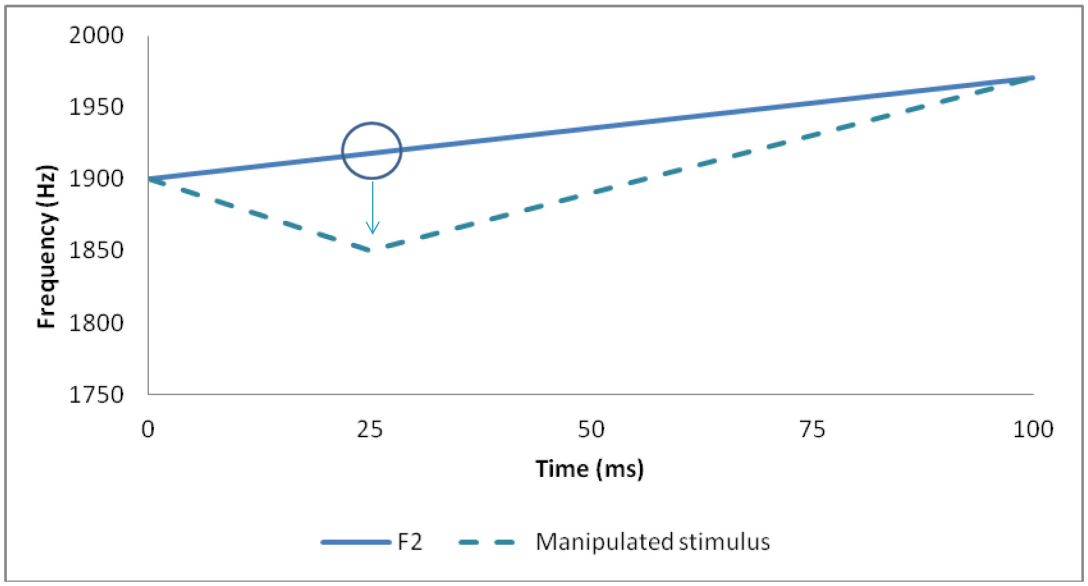

Figure 1 The base condition is shown such that the solid blue line depicts the original F_2 trajectory for the vowel (ranging from 1900 to 1971 Hz). The dotted line shows how the formant was manipulated at 25% of the vowel duration in order to find each participants threshold for detection. Note the arrow indicating the direction the deviation will go until the participant can perceive a change.

Another series of stimuli was generated in which the bend occurred at 50 ms or 50% of stimulus duration. In another, the duration of the vowel was extended to 200 ms with the inflection point still at 25% of the duration (50 ms). In a fourth series, the formant transitions for both F_1 and F_2 were reversed in time (offset and onset values switched) such that the stimuli sound like /ɪ / instead of /ɛɪ/. Another series was identical to the first (100 ms duration and inflection point at 25 ms) except the fundamental frequency was dropped to 110 Hz at onset and 77 Hz at offset. In an additional condition, the first formant was varied to examine the changes this would have over changing F_2. Additionally, a condition in which the deflection direction was altered showed the bend in F_2 moving above the original trajectory as opposed to below it, as in the base condition. This caused the offglide of the vowel to slope downward. Finally, in order to

test the effects on changing the range of F_2, /aɪ/ as in 'hi' was also used. This vowel has a steeper formant change. F_1 ranges from 800 Hz at onset to 394 Hz at offset and F_2 ranges from 1215 Hz to 1971 Hz. Note that this is the same offset F_2 value as the base condition.

Table 1 Research parameters for experiment one are outlined.

Condition	Vowel	Duration	Bend	Different F_0	Formant being manipulated	Deflection direction
1	eɪ	100 ms	25%	No	F_2	Down
2	eɪ	100 ms	50%	No	F_2	Down
3	eɪ	200 ms	25%	No	F_2	Down
4	ɪ	100 ms	25%	No	F_2	Down
5	eɪ	100 ms	25%	Yes	F_2	Down
6	eɪ	100 ms	25%	No	F_1	Down
7	eɪ	100 ms	25%	No	F_2	Up
8	aɪ	100 ms	25%	No	F_2	Down

The primary purpose of these variations was to examine the effects of changing the slope of F_2 transition (100 versus 200 ms duration) and the direction (/ɪ / versus /ɛɪ/). By placing the inflection point at 50% of the duration, we can also examine the effects of changing transition slope independent of duration. Altering the fundamental frequency allows us to examine the interaction between formant and fundamental frequencies. The first formant was altered to determine if listeners are more sensitive the changes in F_2 or F_1. The additional vowel /aɪ / was tested because of the larger formant transitions. Comparing this to the base condition vowel allows us to assess how more significant formant transitions affect a listener's perception to change.

2.1.4 PROCEDURE

Each participant was tested one at a time in the sound booth. Stimuli were played at a comfortable listening level (approximately 70 dB). For each trial the listener heard a set of three vowel-like stimuli. Two of these stimuli were identical and they were smooth formant trajectories. The third stimuli featured the inserted bend. Presentation of this stimulus was random such that the odd stimulus could be presented first, second, or third. Using a handheld numeric keypad, participants were instructed to choose either 1, 2, or 3, corresponding to the stimulus they believed sounded different from the other two. Once their response was chosen, a new stimulus set was automatically played. If the participant was correct three times in a row the deviation from straight formant transition became smaller in 5 Hz increments. If the participant was wrong once, the deviation became larger. The various conditions were presented in randomized order to each participant. Total testing time ranged from 40 min to an hour depending on the participant. A threshold was determined based on the average of the last 6 reversals.

2.2 RESULTS

The results for each condition are represented graphically based on the average threshold across participants (Figures 2-9). Error bars represent that standard error of the mean. In each figure, the comparitor stimulus is represented by a dotted line. The angled lines represent thresholds, or the smallest deviation necessary from the straight transition for participants to detect a stimulus change. The absolute threshold in each condition represents the frequency at which participants could hear a difference. The deviation

threshold will refer to the frequency change from the original trajectory to the absolute threshold.

Figure 2 represents the threshold results for the base condition. The threshold for detection of a bend at 25% of the vowel was 1830 Hz. This is a deviation threshold of 87 Hz from the original stimulus. When the bend was inserted at 50% of the vowel (Figure 3) the absolute threshold was 1830 Hz with a larger deviation threshold of 105 Hz from the original formant. Figure 4 shows results for condition 3 in which the vowel was extended to 200 ms duration, effectively changing the slope of the formant. The figure shows that this manipulation does little to change the absolute frequency of the threshold for the deviation. The threshold for the 200 ms condition was 1817 Hz with a deviation threshold of 101 Hz.

Figure 5 shows that reversing the direction of F_2 had little effect on the absolute threshold of the inflection point (1839 Hz). Nonetheless, the deviation threshold was much larger (114 Hz versus 87 Hz from the base condition). This suggests that the relative magnitude of the deviation in F_2 needed to be much larger for the /ɪ/ stimulus but that the absolute threshold of the frequency deviation was approximately the same as for /ɛɪ/. Additionally, when the fundamental frequency of the vowel was altered, it had little effect on the threshold when compared to the base condition (Figure 6).

Results show that when F_1 is being altered the deviation threshold change necessary for detection is 48 Hz (Figure 7), giving an absolute threshold of 465 Hz. In Figure 8, the

deflection direction condition is shown compared to the base condition. When the bend was inserted above the original stimulus, the deviation threshold was 159 Hz, giving an absolute threshold of 2077Hz. This is shown compared to the base condition in which a deviation threshold of 87 Hz was found.

Finally, an additional vowel was tested in order to evaluate listener perception with more dramatic formant transitions. Figure 9 shows the results from testing the vowel /aɪ/ compared to /eɪ/. The absolute threshold for this new vowel was 1193 Hz with a deviation threshold of 211 Hz (compared to a deviation threshold of 87 Hz for the original vowel tested in condition 1).

A within subject analysis of variance was performed on results from conditions 1-5. There were no significant effects for any of these conditions ($F_{4,56}$=0.288, n.s.). The remaining conditions were not included in that analysis as they were not expected to give comparable results.

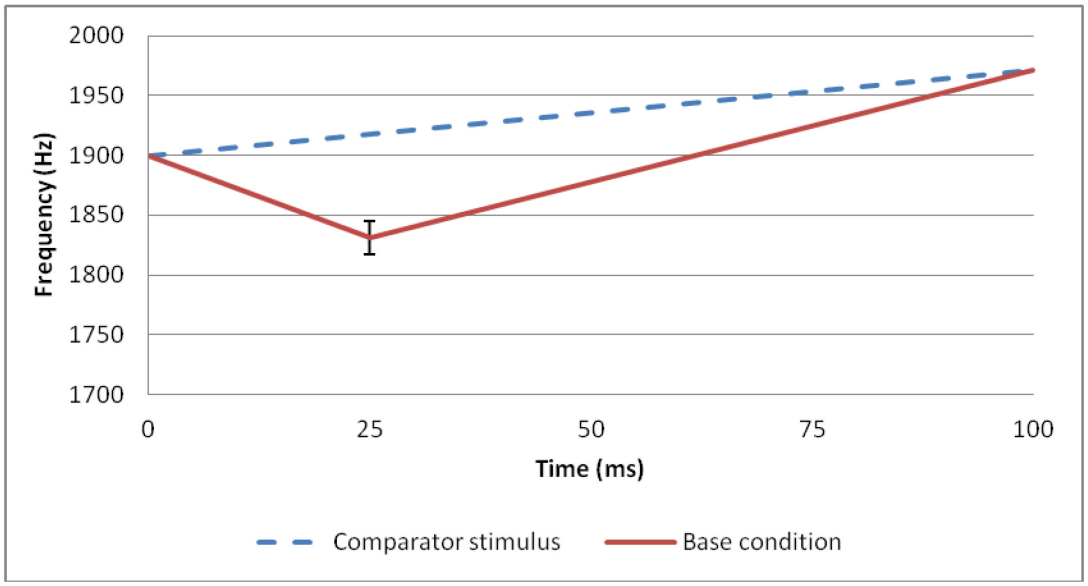

Figure 2 The base condition threshold is shown in red (condition 1). The dotted blue line outlines the original stimulus. Results are averaged over all of the participants and error bars are shown. Absolute threshold = 1830 Hz; deviation threshold = 87 Hz.

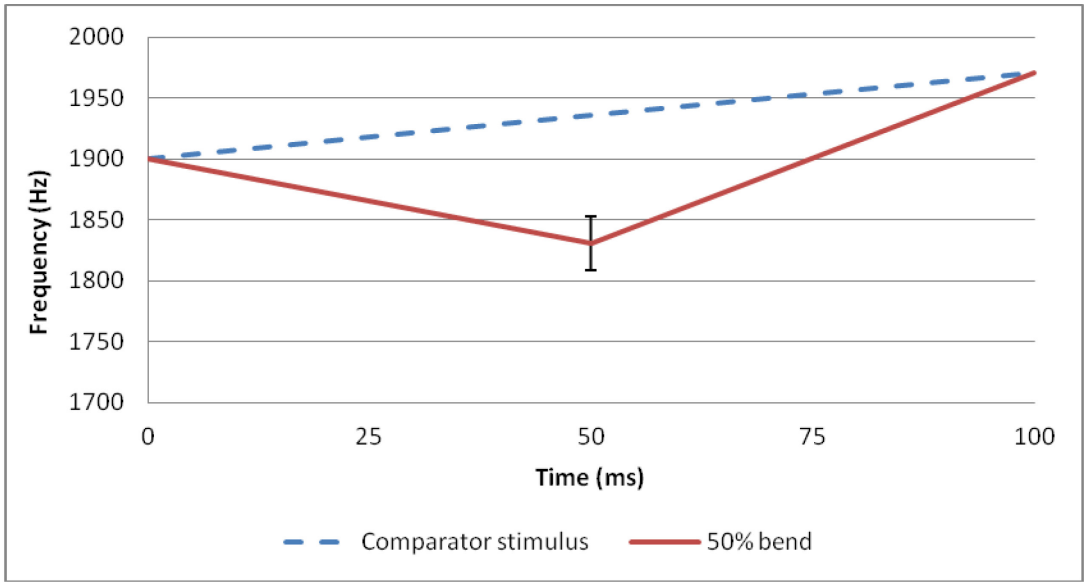

Figure 3 Condition 2 used the midpoint, 50% of the vowel duration, as the deviation point. Absolute threshold = 1830 Hz; deviation threshold = 105 Hz.

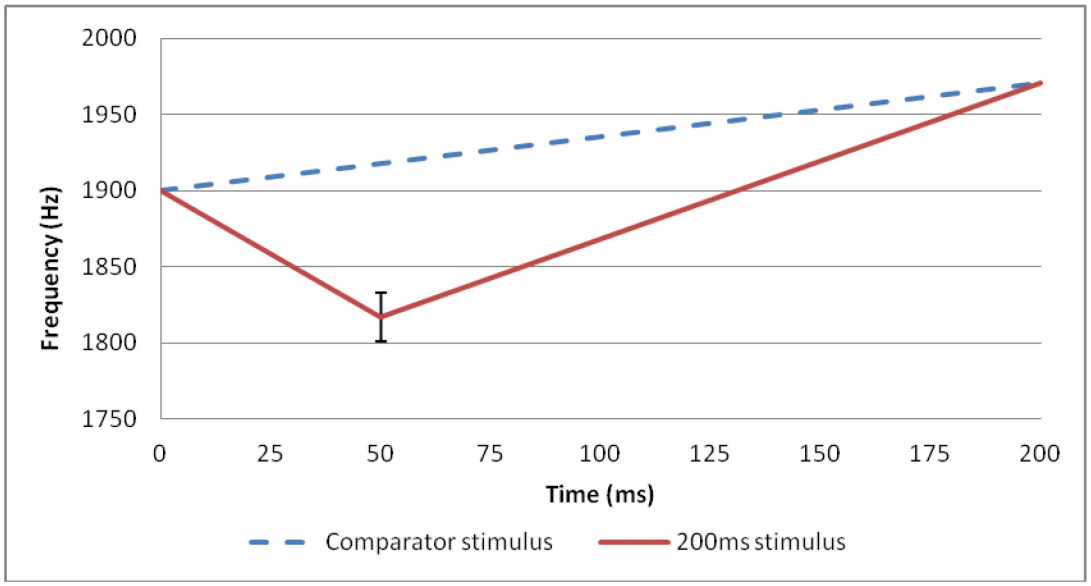

Figure 4 In condition 3 the vowel stimulus was doubled to 200 ms and the bend was maintained at 25% of the vowel duration which is now 50 ms. Absolute threshold = 1817 Hz; deviation threshold = 101 Hz.

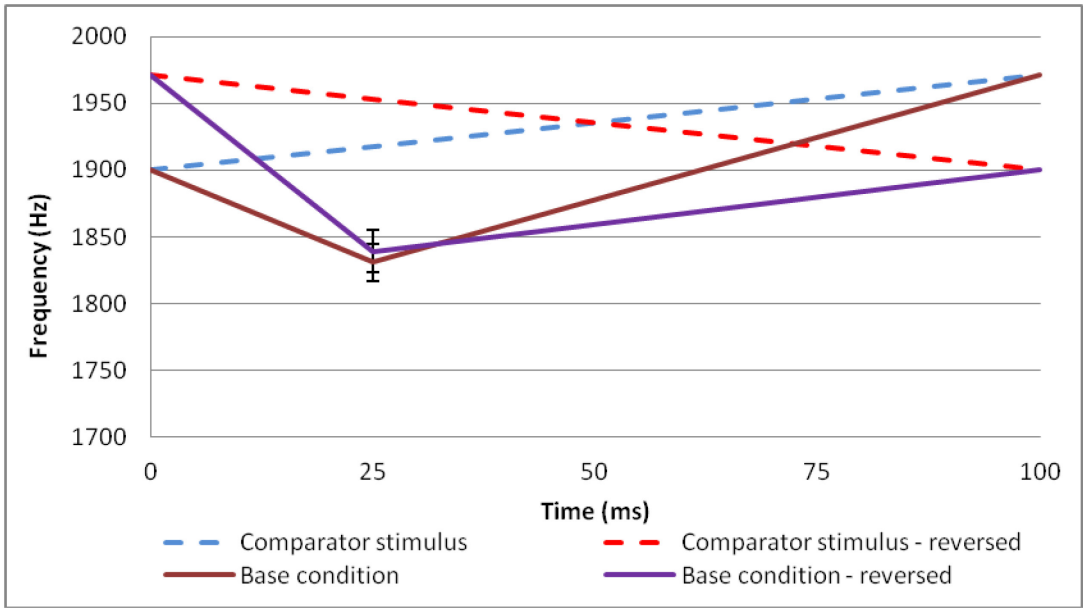

Figure 5 Condition 1 (base) is shown overlapped by condition 4 (reverse condition). In the reverse condition all formants (F_0, F_1, and F_2) were reversed such that the onset and offset switched. The original stimulus, once reversed, is shown by the red dotted line. Condition 1 absolute threshold = 1830 Hz; deviation threshold = 87 Hz. Condition 4 absolute threshold = 1839 Hz; deviation threshold = 114.

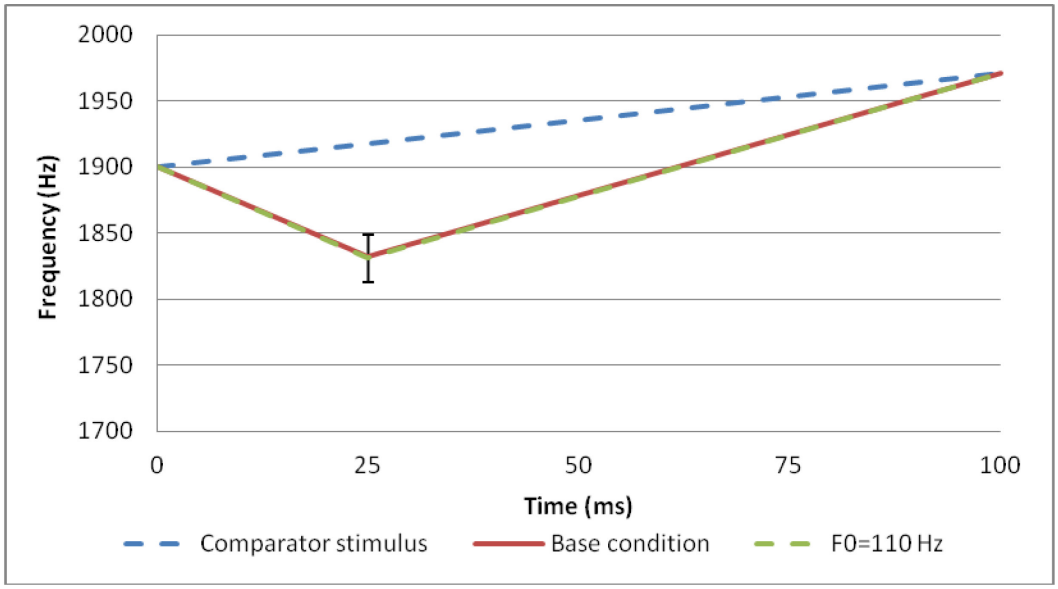

Figure 6 Condition 1 (base) is shown overlapping with condition 5 in which the fundamental frequency contour was scaled down to 110 Hz. This change was done to ensure the fundamental frequency was not affecting listeners' perception. Condition 5 absolute threshold = 1830 Hz; deviation threshold = 87 Hz.

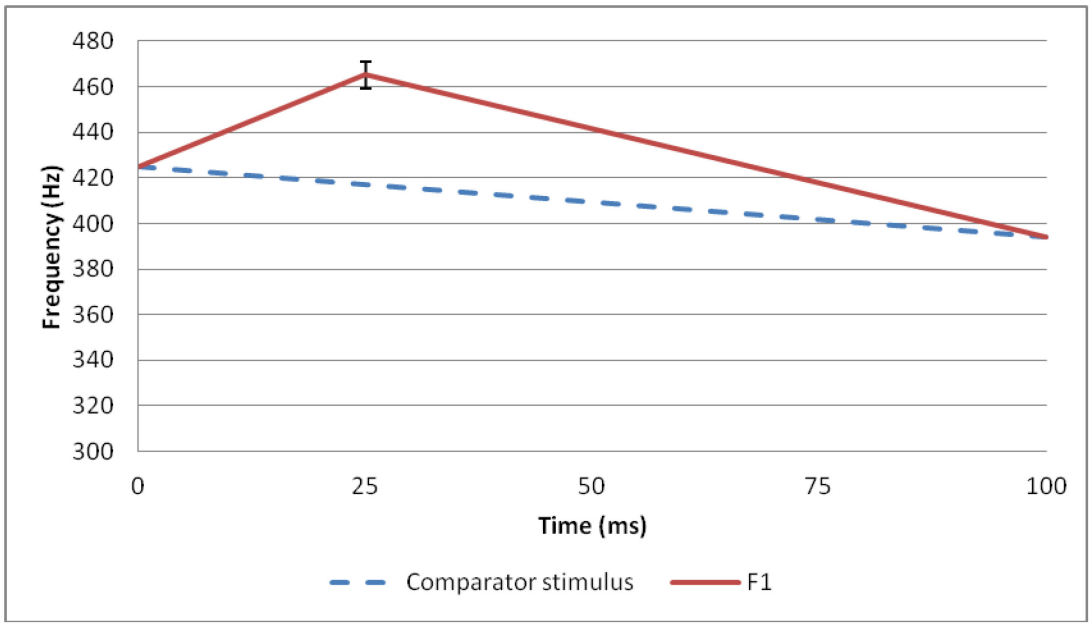

Figure 7 In condition 6 F_1 is varied at 25% of the 100 ms stimulus. F_2 is not depicted however it remains unchanged. Absolute threshold = 465 Hz; deviation threshold = 48 Hz.

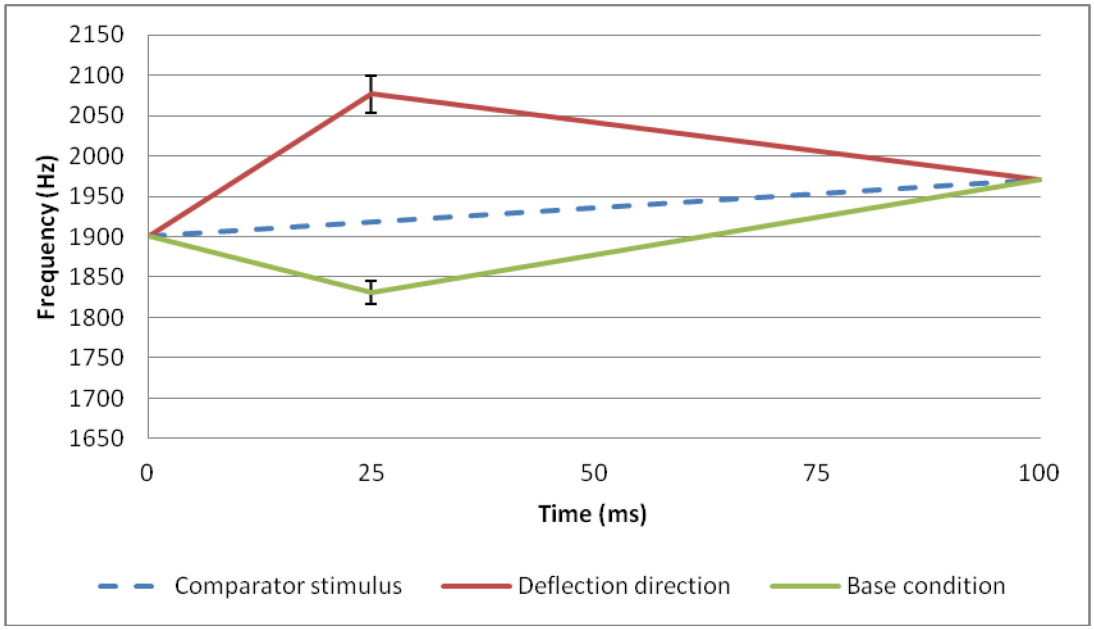

Figure 8 Condition 1 (base) is shown in green and compared to condition 7 (deflection direction). For this condition the bend was moved positively, above the original stimulus causing the offset of the vowel to descend in frequency. Condition 7 absolute threshold = 2077 Hz; deviation threshold = 159 Hz.

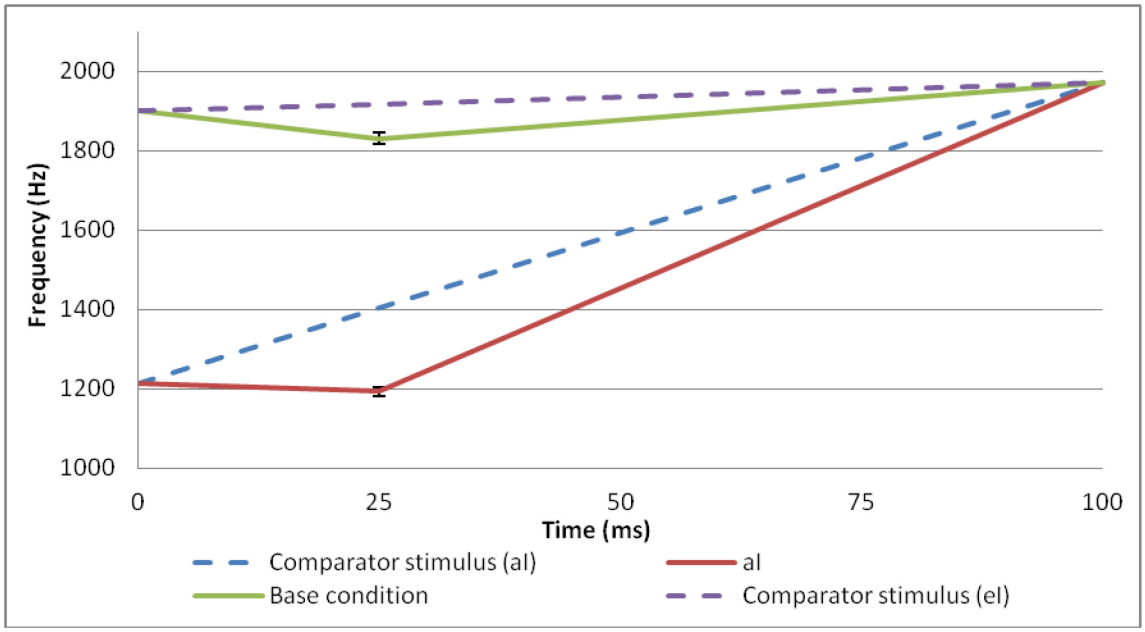

Figure 9 Condition 1 (base) is shown in green and compared to the vowel /aɪ/ , condition 8, which is shown in red. This graph shows the more dynamic formant transitions that occur in this vowel and thresholds are shown for each. Condition 8 absolute threshold = 1193 Hz; deviation threshold = 211 Hz.

2.3 DISCUSSION

The purpose of this study was to test the limits of the onset-offset theory of vowel perception by extending the work done by Morrison and Nearey (2007). Their results provided support for this theory in that listeners' identification patterns for straight and elbowed trajectories with the same endpoints were statistically indistinguishable. The most interesting result from their study was anecdotal evidence that suggested listeners not only believed the two stimuli to be the same vowel, but that they could not even hear a difference between them. This indicated that a small bend in the formant trajectory was not perceptible to listeners in this study. However, by introducing a sufficiently large deviation, we were able to show that listeners can perceive differences in formant frequency over time. Some participants reported hearing /eɪwi/ when the deviation was large enough to detect. One participant reported hearing /eɪli/.

Many conditions were tested in this research and some provided more interesting results than others. Briefly, conditions 6-8 shown in Figures 7-9 provided less relevant results than the other conditions as they cannot be compared directly with the base condition. When F_1 is manipulated in condition 6, listeners appear to be more sensitive to changes when compared to F_2 manipulations. However, this is not unexpected because the range of formant transition is much smaller for the first formant than for the second. It is also known that listener sensitivity to lower frequencies is greater than for higher frequencies (Gelfand, 2010). When the bend was brought above the original stimulus, as shown in figure 8 (condition 7), it seems to have a larger deviation threshold. This result proved to be unimportant as it was not directly comparable to other conditions. Lastly, when

another vowel, aɪ, was assessed in condition 8, listeners appeared to be less sensitive to changes when the formant transitions were larger. This vowel has a larger range in formant transition (especially for F_2 which was being tested). When the formant transition is larger, it appears that the deviation threshold is also larger. That being said, it appears as though the threshold falls tighter to the original trajectory in that the onset now appears to be flat until threshold and then there is a steady change for the offset whereas in the base condition, eɪ, the onset of the vowel because a negative slope down towards the threshold followed by a positive slope towards the offset. The pattern followed by the threshold found in aɪ is very similar to the stimuli used to test VISC theories in Morrison and Nearey (2007).

As shown in the Figures 1-6, the absolute threshold reached for each condition was similar despite several differences among the conditions. The approximate value of 1830 Hz appears to be the frequency at which F_2 must be lowered to in order for listeners to detect a change irrespective of the experimental condition. The F_2 in these stimulus conditions varied in slope, direction, onset/offset frequency. Despite the fact that the *absolute* threshold appears to be constant across these conditions, the *deviation* threshold does not. When comparing the base condition with condition 2 in which the bend was inserted at 50%, the deviation threshold is not the same (Figures 2 and 3). In other words, for the bend to be perceived, a larger change was necessary in the second condition than was needed for the base condition.

This absolute threshold for detection of a deviation in F_2 seems to be consistent despite the slope of the formant. This is demonstrated by condition 3 in which the stimulus was extended to 200 ms in duration (Figure 4). With the same onset and offset values as the base condition but double the duration, the slope of this condition is much shallower than the base condition. However, with this change, the absolute threshold still falls within the same approximate frequency of 1830 Hz. Similarly, when the vowel was reversed in condition 4 the slope became negative and this had little to no effect on the absolute threshold value.

Finally, when the fundamental frequency was altered in condition 5 the threshold value was unaffected. The frequency harmonics nearest the threshold formant frequency of 1830 Hz at the instant of the elbow for the experiment in which f_0 is reversed in time are 1734 Hz, 1850 Hz, 1966 Hz, etc with f_0=116 Hz. The harmonics at the instant of the elbow for the base condition were 1730 Hz, 1832 Hz, 1934 Hz, etc for f_0=102 Hz. Given that there does not appear to be a relationship between the frequencies of harmonics and the threshold for perception of the formant deviation, the threshold we are seeing throughout all of the conditions cannot be easily explained by harmonic frequencies.

In order to understand these results, it is necessary to consider the variables that are held constant across these conditions. With slope, duration, direction, and deviation threshold all varying, there seems to be only one similarity across the stimuli. The range of F_2 frequencies across the duration of the stimulus remains the same, irrespective of other parameters. It is possible that this is the primary factor that influences the magnitude of

the deviation necessary for listeners to perceive a change. If this is true, it implies that listeners' ability to detect these deviations can only be determined after the stimulus is complete as listeners must hear the offset of the stimuli before the range of F_2 frequencies is determined.

This phenomenon of retrospective listening has been shown when using click stimuli (Shore et al., 1998). Sensory saltation is an illusion where a stimulus occurring later in time affects the perception of a stimulus occurring earlier (Geldard, 1982). It has been shown to occur when multiple stimuli occur in close temporal proximity at various locations. Shore and his research team showed this mislocalization to occur when presenting "four clicks on the left followed by four clicks on the right" (1998). When the timing between clicks was reduced, listeners perceived the clicks to be moving continuously from left to right as opposed to discontinuously as they were presented. When the stimulus was cut in half, presenting only four clicks on the left, listeners perceived the sound as coming only from that side. Similar results were found when four clicks were presented on the right. When these stimuli were combined into eight clicks total the listener then perceived sound movement. When the clicks are presented in short succession the result is a complex stimulus that the brain must make sense of. Shore and colleagues suggested that this phenomenon is the brain's way of providing the simplest explanation to this complex set of sounds and that it is only after the listener hears all of the sounds that they are able to then perceive the earlier sounds as moving continuously (1998).

It is unknown if the results presented here could represent a similar phenomenon of retrospective listening. In order to test this theory, a second experiment was conducted using stimuli similar to these, but which were truncated such that the range of F_2 frequency is reduced while preserving the slope and onset frequency of the formants. Similarly to Shore et al., we would expect to see a change in perception when only half of the stimulus is presented. If the threshold is dependent on the onset and offset frequency value, changing the offset value should in turn change the absolute threshold.

CHAPTER 3 EXPERIMENT 2

3.1 METHODS

3.1.1 PARTICIPANTS

Fifteen additional participants were recruited to participate in the second experiment. Similarly to experiment one, all subjects were 18 years or older and had normal hearing as determined by an audiometric hearing screening performed on the day of participation.

3.1.2 APPARATUS

There was no change in the apparatus used from experiment 1.

3.1.3 STIMULI

As in the first experiment, a three-alternative force-choice 1-up 3-down adaptive staircase design was used, converging on the 0.841 probability of the psychometric function (Levitt, 1971). Stimuli were presented in a series of 3 vowel-like sounds in succession. All stimuli were derived from the previous experiment as described below.

The base condition was retested in order to evaluate the results from a new set of participants. The stimulus was 100 ms long and the vowel was modeled after the diphthong /ɛɪ / as in 'hay'. As a reminder, F_1 onset frequency was set at 425 Hz and offset frequency was 394 Hz. Second formant, F_2 transitioned from 1900 Hz at onset to 1971 Hz at offset. For the deviant stimuli, an inflection point or bend was introduced at

25% of the vowel duration or 25 ms. The deviation of this frequency from the linear interpolation between onset and offset was varied until responses converged on a detection threshold. Fundamental frequency (f_0) was set at 125 Hz at onset and fell to 88 Hz at offset. In condition 2, this stimulus truncated at 50% of its duration. The bend was introduced at 50% of the new 50 ms duration (ie. 25 ms). For this condition the new F_2 ranged from the original onset value of 1900 Hz to a new offset value that varied depending on the condition – i.e., each new truncated stimulus was always exactly the same as the first half of the corresponding 100 ms stimulus. The offset value was located at half of the duration and it always pointed towards but never reached the original offset value. For the base condition, this new offset value was 1876 Hz. Additionally, the fundamental frequency for the truncated condition was halved.

In order to test the effects of slope in this experiment, condition 3 was created such that the formant values were identical to the base condition however the vowel was extended over 400 ms duration. In condition 4 this stimulus was cut in half such that the duration was 200 ms, the bend was at 100 ms, and the offset frequency for F_2 was 1868 Hz.

For condition 5 the reverse of the base condition was retested (similarly to condition 4 in experiment 1). The formant transitions for both F_1 and F_2 were reversed in time (offset and onset values switched) such that the stimuli sound like /ɪ / instead of /ɛɪ/. In condition 6 this stimulus was also cut in half such that F_2 ranged from the original onset value of 1971 Hz to a new offset value of 1869 Hz.

In condition 7 the fundamental frequency was reversed in time, turning the vowel from a more natural pitch declination to a pitch contour resembling an interrogative. In this case, the f_0 rose from an 88 Hz onset to 125 Hz at offset. This alteration was done to further examine the interaction between the formants and the fundamental frequency.

Table 2 Research parameters for experiment two are outlined.

Condition	Vowel	Duration	Bend	Reverse F_0	Original stimulus halved?
1	eɪ	100 ms	25%	No	No
2	eɪ	50 ms	50%	No	Yes
3	eɪ	400 ms	25%	No	No
4	eɪ	200 ms	50%	No	Yes
5	ɪ	100 ms	25%	No	No
6	ɪ	50 ms	50%	No	Yes
7	eɪ	100 ms	25%	Yes	No

3.1.4 PROCEDURE

There were no changes to the procedure from experiment 1. Participants were instructed in the same way and performed the same tasks. The only variation was in the stimuli being presented.

3.2 RESULTS

As was shown in experiment 1, the results for each condition in this experiement are again represented graphically based on the average threshold across participants (Figures 10-13). Error bars represent the standard error of the mean and the comparitor stimulus is

represented by a dotted line. This comparitor was always presented twice, and the different stimulus was presented once. The order of this presentation was random such that the different stimulus could be in the first, second, or third position. The angled lines represent thresholds, or the smallest deviation necessary from the straight transition for participants to detect a stimulus change. Both absolute and deviation threshold values are presented for each condition in figure captions.

When comparing conditions 1-6 (represented in figures 10, 11, and 12), results show that when these conditions are truncated to half their duration, the threshold is largely unaffected. In Figure 10, the base condition is shown compared to half of the same stimulus. The absolute thresholds are 1830 Hz and 1829 Hz respectively. The results of the base condition found in experiment 2 confirm the results obtained from the first experiment in which the threshold of the base condition was also found to be 1830 Hz.

In condition 3, the extended vowel with a duration of 400 ms, the absolute threshold was 1820 Hz (as shown in figure 11). When this stimulus was halved as in condition 4, the absolute threshold was 1818 Hz. The results of the halved condition, at 200 ms duration, is comparable to the results obtained in experiment 1 in which the 200 ms duration stimulus had a bend inserted at 25% of the vowel. The threshold of this condition, which is shown above in Figure 4, was 1817 Hz. For the condition shown in Figure 11, the bend is inserted into the halved 200 ms duration stimulus at 50% of the vowel duration, or 100 ms into the stimulus.

Figure 12 shows similar results when condition 5 (the reverse condition) is compared to condition 6 which is halved. Absolute thresholds obtained for these conditions are 1850 Hz and 1853 Hz respectively showing little to no change when the stimulus is halved. Again, the reverse threshold obtained in the second experiment is comparable to that found in experiment 1 (1839 Hz).

An additional condition was added to experiment 2 in order to evaluate the interaction between the formants and the fundamental frequency. Figure 13 shows condition 7 in which the fundamental frequency is reversed, changing the vowel from a more natural pitch declination to a pitch contour resembling an interrogative. This condition is compared to the base condition and the absolute threshold values obtained are 1796 Hz and 1830 Hz respectively. Deviation thresholds are 122 Hz for condtion 5 and 87 Hz for condition 1.

A within-subject analysis of variance was performed on these results and significance was found ($F_{6,84}$=2.666, p=0.02). A Bonferonni post-hoc comparison revealed that the significance was between conditions 6 and 7.

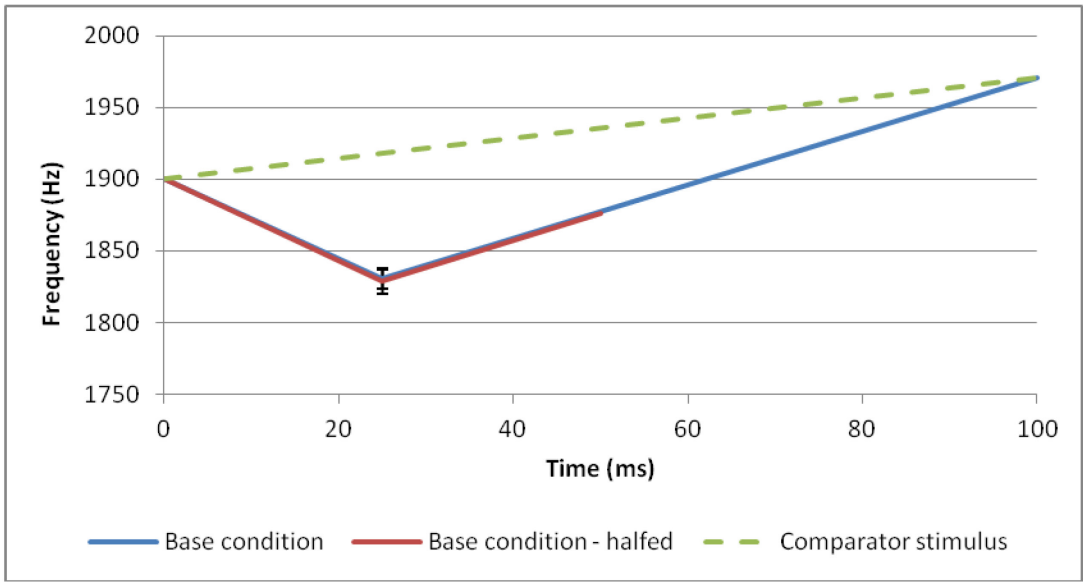

Figure 10 Condition 1 (base) is shown in blue and compared to condition 2 (halved stimulus) shown in red. Condition 1 absolute threshold = 1830 Hz; deviation threshold = 87 Hz. Condition 2 absolute threshold = 1829 Hz; deviation threshold = 88 Hz.

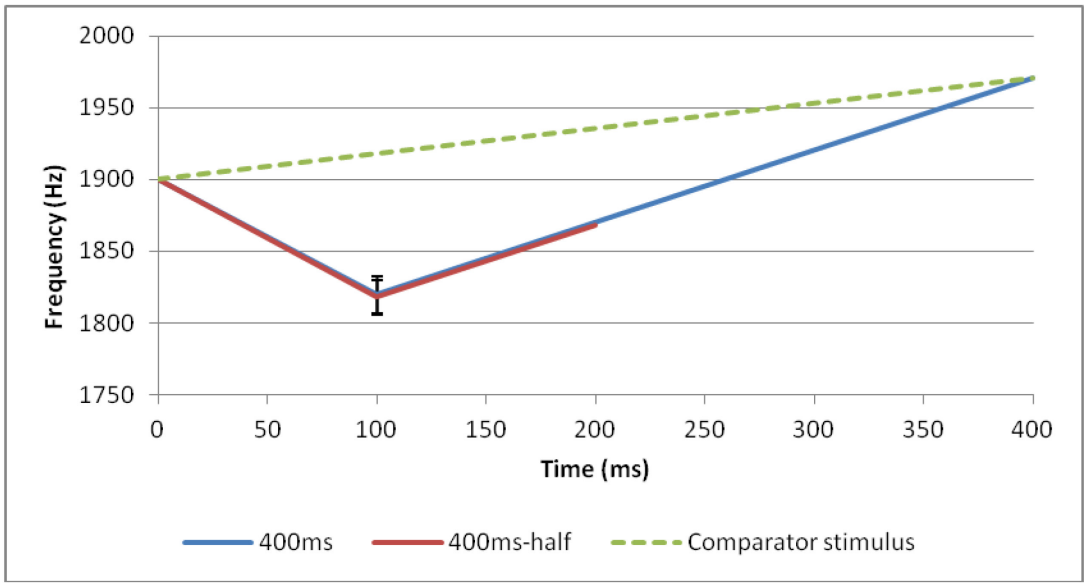

Figure 11 Condition 3 shows the base condition extended over a 400 ms duration (shown in blue). This threshold is shown compared to condition 4 in where the 400 ms stimulus was cut in half (shown in red). Condition 3 absolute threshold = 1820 Hz; deviation threshold = 98 Hz. Condition 4 absolute threshold = 1818 Hz; deviation threshold = 100Hz.

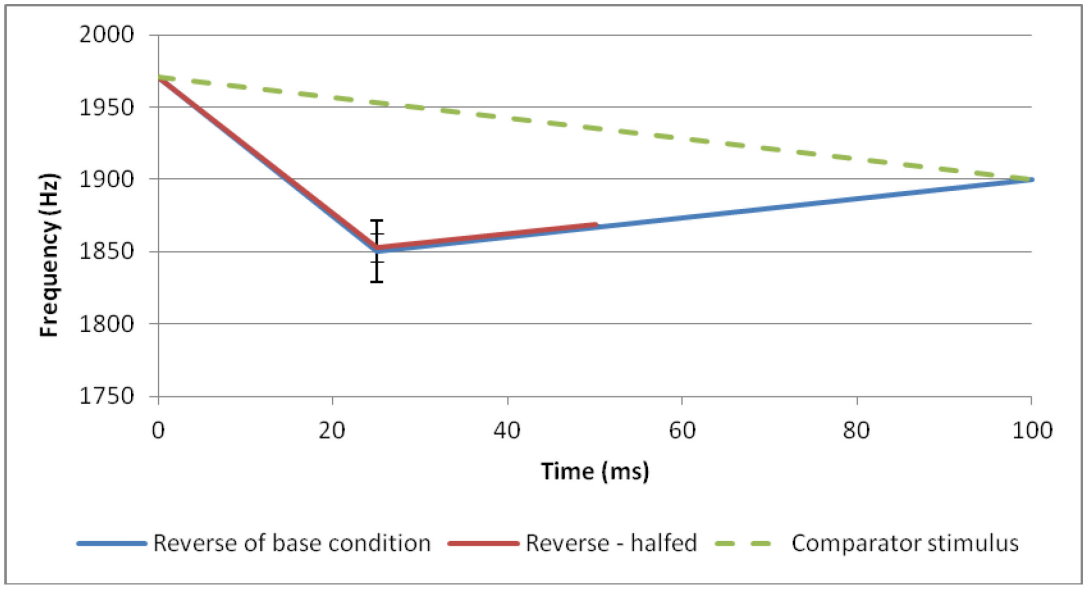

Figure 12 Condition 5 (reverse) is shown in blue and is being compared condition 6 in which the reverse stimulus was cut in half (shown in red). Condition 5 absolute threshold = 1850 Hz; deviation threshold = 103 Hz. Condition 6 absolute threshold = 1853 Hz; deviation threshold = 100 Hz.

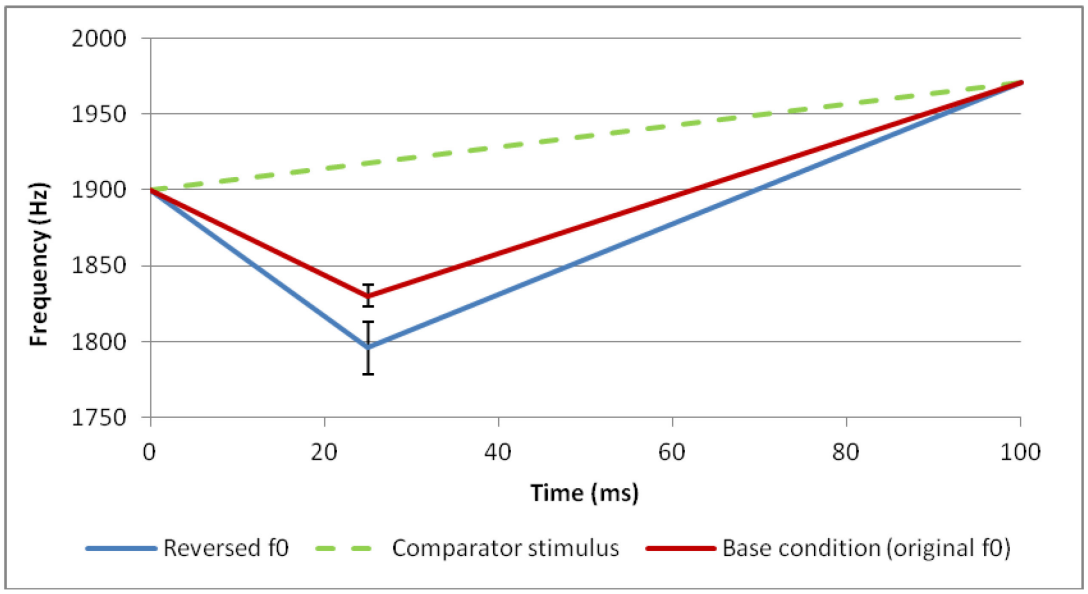

Figure 13 Condition 1 (base) is shown in red and compared to condition 7 in which the fundamental frequency of the stimulus was reversed (show in blue). Condition 7 absolute threshold = 1796 Hz; deviation threshold = 122 Hz.

3.3 DISCUSSION

The purpose of this second experiment was to further assess whether the auditory saltation phenomenon was a plausible explanation for the results noted in experiment one. As a reminder, the first experiment showed that the threshold for detection of a bend embedded within the vowel was constant for multiple stimuli. The only obvious constant between these conditions was the range of frequencies spanned by the stimulus (i.e., the absolute difference between the onset and offset frequency). From this, the auditory saltation phenomenon was introduced as a possible explanation for these results. If listeners are making a perceptual decision retrospectively, then we would need to change the range of frequencies in order to impact the threshold. By halving the stimuli in duration, we would expect to see a change in the threshold frequency, and it would likely be a smaller absolute threshold. However, when the range of the formant endpoints was altered, reducing the offset frequency, similar thresholds were still noted.

In Figure 10, a comparison is made between conditions 1 and 2. As shown, the thresholds for each are very similar. In figure 11 a new condition was introduced such that the base stimulus was extended to 400 ms duration over the same frequency range (condition 3). This condition was also halved (condition 4) and again the thresholds remained unchanged. In Figure 12 a comparison is made between conditions 5 and 6 (reverse and reverse half). Again, the thresholds are the same for each condition. This is contradictory to the initial conclusion that the threshold would be dependent upon the range of the formant endpoints. When comparing these conditions, there are very few similarities between them. They vary in their duration, slope, offset frequency, direction

of the formant transition, and in the range of frequencies spanned by the vowel. It has become more difficult to determine what variables are being maintained in order to explain why the threshold for detection seems to stay the same throughout conditions.

Condition 7 (shown in figure 13) has f_0 reversed in order to assess the relationship between perceived formant frequency and fundamental frequency. This condition is similar to that shown in Figure 6 of experiment one (condition 5 for that experiment), in which the f_0 was scaled down. This was the only condition during the second experiment in which a new threshold value was obtained. Unfortunately, this is contradictory to what we would expect to see. From the first experiment, it seemed as though a new f_0 had no effect on the threshold and therefore it was assumed that a relationship between perception of the bend and the fundamental frequency was not at play. Based on results shown here, there may be an interaction. Harmonics nearest the threshold frequency of 1796 Hz at the instant of the bend for the condition in which the f_0 is reversed in time are 1791 Hz, 1876 Hz, 1961 Hz, etc for f_0=85 Hz. Similar to the first experiment, there is no pattern between the threshold of the bend and the harmonic frequencies at that point in time. With such a limited knowledge on how we perceive formants and the impact formant peaks, or harmonics, have on this perception, more research is warranted to further investigate these contradicting results. Conclusions cannot be reached based on the minimal results obtained in these two conditions.

In Chapter 4 a general discussion will outline possible conclusions that can incorporate the results from both experiments. In addition, suggestions will be made for future research to further investigate the role formant transitions play in vowel perception.

CHAPTER 4 GENERAL DISCUSSION

The results of our experiments indicate that listeners can perceptually track formants across time. Results also show that the threshold for detection of these stimuli is nearly the same despite many variable manipulations. Our preliminary conclusions following the first experiment suggested that the absolute range of frequencies spanned by F_2 was responsible for the threshold we obtained. If this were true, then the listeners would be performing a phenomenon referred to as retrospective listening (Shore, et al., 1998). This would indicate that listeners must first hear the entire stimulus, including the offset frequency, before they can determine if a change somewhere in the middle was detected. In order to provide further support for this claim, our second experiment was conducted in which the stimulus was truncated and the offset frequency was changed. Based on these results, our initial conclusion was disproven. Therefore, the threshold appears to be unrelated to the absolute range of the formant transition nor did our first experiment show an example of the retrospective listening phenomenon.

Different threshold frequencies were noted when another vowel was tested (see Figure 9). This was expected because the new vowel spanned an entirely different frequency range. These results lead us to wonder if there is something about the frequency range used in this experiment for the base condition, and all of the modifications of it, which leads to the similar threshold frequency results obtained. Additional testing would need to be done using other vowels that span other frequency ranges. For example, we tested the vowel /aɪ/ but other manipulations could be made to this vowel in order to extend the

duration, insert the bend at 50% of the vowel, etc. These changes may not result in the same threshold of detection. This would show that the frequency range we tested was impacting the thresholds we obtained.

Further investigation is necessary into possible interactions between the fundamental frequency and the formants of a vowel. In experiment one, f_0 was modified and this had no effect on the threshold; however, in experiment two when the f_0 was modified such that it was reversed, a new threshold was obtained. With only two examples of f_0 modification, it is impossible to conclude whether or not an interaction is occurring between f_0 and the formant frequencies. Additionally, we were unable to identify what this interaction would be if it did exist. It would be interesting to further explore a possible interaction between fundamental frequencies and the formant peaks.

In 2000, Dissard and Darwin performed an experiment asking listeners to adjust the formant frequency of one sound such that it matched another. They showed that matches were more reliable when the fundamental frequencies of the sounds matched as opposed to when they did not (Dissard and Darwin, 2000). They also showed that the difference between the matches was larger for sounds with high f_0 than sounds with low f_0. The introduction of a new fundamental frequency requires listeners to make a more abstract level of matching. This task is more difficult and more abstract when the harmonics are resolved as opposed to unresolved (Dissard and Darwin, 2001). These authors admit that more testing is required to support their evidence. It could be a case that when the fundamental frequency of the stimulus was altered, the interaction between it and the

formants was changed such that a new threshold for detection was found. This is only a preliminary finding based on limited testing. Nonetheless, continued research is warranted in this area to further assess any interactions that may be at play here.

A final explanation for these consistent results that should be investigated is the phenomenon of categorical perception. It may be the case that when the second formant is lowered to approximately 1830 Hz part way through its transition, listeners are able to hear a new phoneme or phonemes and it is not until this perceptual change occurs that they notice any change in the formant trajectory. In study done by Liberman, et al. (1957), researchers were able to show that graded changes in F_2 transitions of synthetic, two formant CV stimuli resulted in the perception of 3 different phonemes: /b/, /d/, and /g/. When the second formant trajectory changed from sharply rising, to more gradual, and to a negative slope listeners perception changed as to which phoneme they heard. We are suggesting that it may be possible that only when the second formant was lowered to approximately 1830 Hz midway through the trajectory were participants able to here an entirely new phoneme and at this point recognized the stimulus as different. Additional testing would be necessary in order to confirm this hypothesis. Testing should involve phoneme labeling of the stimuli in order to determine whether the "different" stimulus sounded like a new phoneme or if listeners were able to hear a change within the same phonemic category. This would allow us to determine if listeners are able to discriminate the stimuli better than they can label them or if these skills are parallel such that perception is categorical in nature.

4.1 CONCLUSIONS

In conclusion, listeners can perceptually track formants over time but the mechanism for this is largely unknown. This experiment tested the acuity of listeners by assessing their threshold for detection when a formant trajectory is modified. This threshold appeared to be a constant that was held despite many variable modifications to the original stimulus. Initial thoughts were that listeners performed retrospective listening and the threshold would be dependent on the absolute range of formant frequencies. When this was found to be untrue in experiment two we could only speculate about how to explain the results. With little to compare between the stimuli, it is possible that the frequency range being tested was too small to notice a threshold change. It's also possible that an interaction between the fundamental frequency and the formant peaks of a vowel exist. However, this interaction was not proven nor could we suggest what it would be. Finally, a categorical perception explanation could suffice. The threshold frequency of many of our conditions could be a point at which listeners detect an entirely different phoneme. Prior to this point, a slight change is undetectable because listeners still perceive the original phoneme. Again, this conclusion is speculative and additional testing will be necessary in order to convincingly explain these results.

4.2 FUTURE DIRECTIONS

Further investigation into how listeners track formants over time is warranted. Based on the results obtained here, it will be essential to focus on any possible interactions between

formants and harmonics, and how these impact our perception of vowels. Theories of vowel perception still have gaps in them and more research is necessary to fill these in. Based on the results obtained, we believe that if F_2 was moved up by 50 Hz we would see a threshold shift of this magnitude as well.

In the future, it would be interesting to see if, instead of halving the base condition as was done in experiment two, the base condition was extended to be 150% of the original length. While doing this, it would be important to maintain the onset value and simply extend the formant trajectory with a continuous slope. It would be helpful to determine if the threshold frequency of a bend would change with these variables or if it would remain the same, as it did when the stimulus was halved. When the stimulus was halved, a narrow frequency range was made even narrower and therefore, it's possible that there is not enough of a difference to see it statistically. By extending the base stimulus, the range of the second formant would be increased and a change in the threshold may be more noticeable statistically.

REFERENCES

Aaltonen, O. (1985). The effects of relative amplitude levels of F2 and F3 on the categorization of synthetic vowels. *The Journal of Phonetics*. **13**: 1-9.

Assmann, P. F., and Katz, W. F. (2005). Synthesis fidelity and time-varying spectral change in vowels. *The Journal of the Acoustical Society of America*. **117**: 886-895.

Bladon, R. A. W., and Lindblom, B. (1981). Modeling the judgement of vowel quality differences. *The Journal of the Acoustical Society of America.* **69** (5): 1414-1422.

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*(3), 195-210.

Dissard, P., and Darwin, C. J. (2000). Exacting spectral envelopes: Formant frequency matching between sounds on different and modulated fundamental frequencies. *The Journal of the Acoustical Society of America*. **107**: 960-969.

Dissard, P., and Darwin, C. J. (2001). Formant-frequency matching between sounds with different bandwidths and on different fundamental frequencies. *The Journal of the Acoustical Society of America*. **110** (1): 409-415.

Divenyi, P. (2009). Perception of complete and incomplete formant transitions in vowels. *The Journal of the Acoustical Society of America.* **126** (3): 1427-1439.

Fant, C. G. M. (1956). On the predictability of formant levels and spectrum envelopes from formant frequencies. In *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday*, edited by M. Halle, (Mouton, The Hague), 109–120.

Fox, R. A., Jacewicz, E., and Chang, C. Y. (2010). Auditory spectral integration in the perception of diphthongal vowels. *The Journal of the Acoustical Society of America*. **128** (4): 2070-2074.

Geldard, F. A. (1982). Saltation is somethesis. *Psychology Bulletin*. **92**: 136-175.

Gelfand, S. A. (2010). Auditory sensitivity. In: *Hearing: An Introduction to Psychological and Physiological Acousticals 5th edition*. Colchester, UK: Informa Healthcare, 166-185.

Gottfried, M., Miller, J. D., and Meyer, D. J. (1993). Three approaches to the classification of American English vowels. *Journal of Phonetics*. **21**: 205-229.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. N. (2001). Effect on consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*. **109**: 748-763.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America.* **97**: 3099-3111.

Hillenbrand, J. M. and Nearey, T. M. (1999). Identification of resynthesized /hvd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*. **105**, 3509-3523.

Ito, M., Tsuchida, J., & Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception. *The Journal of the Acoustical Society of America.* **110** (2), 1141-1149.

Jacewicz, E. (2005). Listener sensitivity to variations in the relative amplitude of vowel formants. *Acoustics Research Letters Online*. **6** (3): 118-124.

Kewley-Port, D. and Goodman, S. S. (2005). Thresholds for second formant transitions in front vowels. *The Journal of the Acoustical Society of America*. **188** (5): 3252-3260.

Kiefte, M., Enright, T., and Marshall, L. (2010). The role of formant amplitude in the perception of /i/ and /u/. *The Journal of the Acoustical Society of America*. **127** (4). 2611-2621.

Kiefte, M. and Klunder, K. R. (2005). The relative importance of spectral tilt in monophthongs and diphthongs. *The Journal of the Acoustical Society of America*. **117** (3): 1395-1404.

Kiefte, M., Klunder, K. R., and Rhode, W. S. (2002). Synthetic speech stimuli spectrally normalized for nonhuman cochlear dimensions. *Acoustic Research Letters Online*, **3**: 41-46.

Kiefte, M., Nearey, T. M., and Assmann, P. F. (2012). Vowel perception in normal speakers. In *Handbook of Vowels and Vowel Disorders*, edited by M. J. Ball and F. Gibbon, 160-185 (Psychology Press, New York).

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*. **67** (3): 971-995.

Klatt, D. H. (1982). Prediction of perceived phonetic distance from critical-band spectra: A first step. *Acoustics, Speech, and Signal Processing, IEEE International Conference.* **7**: 1278-1281.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*. **49**: 467-477.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychology Review*. **74**: 431-461.

Liberman, A. M., Harris, K. S., Hoffmann, H.S., and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *The Journal of Experimental Psychology*. **54**: 358-368.

Liu, C., Tao, S., Wang, W., and Dong, Q. (2012). Formant discrimination of speech and non-speech sounds for English and Chinese listeners. *The Journal of the Acoustical Society of America*. **132** (3): EL 189-195.

Morrison, G. S., and Nearey, T. M. (2007). Testing theories of vowel inherent spectral change. *The Journal of the Acoustical Society of America*. **122** (1): EL 15-22.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America.* **85** (5), 2088-2113.

Nearey, T. M., and Assmann, P. F. (1986). Modeling the role of vowel inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*. **80**: 1297-1308.

Peterson, G. E. (1952). The information-bearing elements of speech. *The Journal of the Acoustical Society of America*. **24**: 629-637.

Peterson, G. E. (1961). Parameters of vowel quality. *The Journal of Speech and Hearing Research*. **4**: 10-29.

Rosner, B. S., and Pickering, J. B. (1994). *Vowel Perception and Production* (Oxford University Press).

Shore, D. I., Hall, S. E., and Klein, R. M. (1998). Auditory saltation: A new measure for an old illusion. *The Journal of the Acoustical Society of America*. **103** (6): 3730-3733.

Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America*. **74** (3): 695-705.

Strange, W. (1989). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America*. **85**: 2081-2087.

Zahorian, S. A., and Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*. **94** (4): 1966-1982.

APPENDIX A

INDIVIDUAL PARTICIPANT THRESHOLD FREQUENCIES FOR EACH CONDITION

Table A.1 Threshold frequencies for each of the fifteen participants in experiment one's conditions. Condition numbers are across the top and participant numbers are down the first column.

	1	2	3	4	5	6	7	8
1	1702	1888	1811	1852	1881	449	2106	1180
2	1884	1896	1877	1890	1867	435	1994	1261
3	1835	1837	1839	1835	1891	475	2078	1162
4	1814	1864	1730	1707	1861	498	2197	1175
5	1889	1907	1857	1872	1840	471	2002	1194
6	1731	1774	1839	1897	1863	468	2015	1128
7	1785	1772	1637	1837	1702	513	2011	1140
8	1818	1860	1835	1837	1886	442	2079	1182
9	1830	1732	1861	1677	1658	456	2183	1173
10	1842	1838	1779	1851	1793	503	2291	1188
11	1875	1894	1867	1883	1852	456	1997	1296
12	1872	1883	1838	1859	1881	460	2029	1248
13	1862	1584	1827	1871	1779	451	2114	1242
14	1846	1852	1834	1859	1859	476	2081	1132
15	1882	1890	1823	1859	1867	427	1982	1191

Table A.2 Threshold frequencies for each of the fifteen participants in experiment two's conditions. Condition numbers are across the top and participant numbers are down the first column.

	1	2	3	4	5	6	7
1	1862	1841	1761	1784	1863	1824	1788
2	1844	1789	1861	1771	1913	1769	1712
3	1870	1856	1774	1812	1877	1862	1847
4	1836	1856	1807	1846	1825	1849	1846
5	1842	1798	1842	1716	1860	1866	1797
6	1815	1820	1806	1846	1901	1895	1847
7	1831	1760	1814	1894	1813	1817	1782
8	1781	1841	1864	1854	1889	1816	1772
9	1821	1854	1720	1781	1580	1829	1607
10	1841	1776	1859	1834	1842	1887	1837
11	1771	1857	1868	1869	1822	1907	1842
12	1800	1811	1743	1771	1874	1841	1812
13	1851	1862	1854	1810	1891	1887	1844
14	1844	1849	1846	1831	1892	1878	1837
15	1851	1867	1881	1854	1909	1864	1771