Acoustic Characterization of the Glides /j/ and /w/ in American English

by

Elisabeth Hon Hunt

B.S.E. Electrical Engineering
Princeton University, 2003

S.M. Electrical Engineering
Massachusetts Institute of Technology, 2005

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2009

Signature of Author: ____

Department of Electrical Engineering and Computer Science
May 19, 2009

Certified by: ____

Kenneth N. Stevens
Emeritus Professor of Electrical Engineering and Health Sciences and Technology
Thesis Supervisor

Accepted by: ____

Terry P. Orlando
Professor of Electrical Engineering
Chair, Department Committee on Graduate Students

Acoustic Characterization of the Glides /j/ and /w/ in American English

by

Elisabeth Hon Hunt

Submitted to the Department of Electrical Engineering and Computer Science
on May 19, 2009 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Electrical Engineering

ABSTRACT

Acoustic analyses were conducted to identify the characteristics that differentiate the glides /j,w/ from adjacent vowels. These analyses were performed on a recorded database of intervocalic glides, produced naturally by two male and two female speakers in controlled vocalic and prosodic contexts. Glides were found to differ significantly from adjacent vowels through RMS amplitude reduction, first formant frequency reduction, open quotient increase, harmonics-to-noise ratio reduction, and fundamental frequency reduction. The acoustic data suggest that glides differ from their cognate high vowels /i,u/ in that the glides are produced with a greater degree of constriction in the vocal tract. The narrower constriction causes an increase in oral pressure, which produces aerodynamic effects on the glottal voicing source. This interaction between the vocal tract filter and its excitation source results in skewing of the glottal waveform, increasing its open quotient and decreasing the amplitude of voicing.

A listening experiment with synthetic tokens was performed to isolate and compare the perceptual salience of acoustic cues to the glottal source effects of glides and to the vocal tract configuration itself. Voicing amplitude (representing source effects) and first formant frequency (representing filter configuration) were manipulated in cooperating and conflicting patterns to create percepts of /V#V/ or /V#GV/ sequences, where Vs were high vowels and Gs were their cognate glides. In the responses of ten naïve subjects, voicing amplitude had a greater effect on the detection of glides than first formant frequency, suggesting that glottal source effects are more important to the distinction between glides and high vowels.

The results of the acoustic and perceptual studies provide evidence for an articulatory-acoustic mapping defining the glide category. It is suggested that glides are differentiated from high vowels and fricatives by articulatory-acoustic boundaries related to the aerodynamic consequences of different degrees of vocal tract constriction. The supraglottal constriction target for glides is sufficiently narrow to produce a non-vocalic oral pressure drop, but not sufficiently narrow to produce a significant frication noise source. This mapping is consistent with the theory that articulator-free features are defined by aero-mechanical interactions. Implications for phonological classification systems and speech technology applications are discussed.

Thesis Supervisor: Kenneth N. Stevens
Title: Emeritus Professor of Electrical Engineering and Health Sciences and Technology

Acknowledgments

It is customary for students in their PhD theses to thank their advisors first and foremost out of a long list of essential contributors to their graduate experience. Let this implicit obligation not detract, however, from the sincerity of the gratitude and admiration that I hold for my own advisor, Professor Kenneth Noble Stevens. I consider myself extremely fortunate and honored to have had the opportunity to study with Ken during my time at MIT. On top of the example of his illustrious career and brilliant mind in speech science, Ken is the most sincere, supportive, and truly caring advisor that a student could ever ask for. His excitement about speech research, his honest and open scientific perspective, and his *joie de vivre* are truly infectious, and have made him beloved by all of his students and colleagues. It is with great pride that I now join the teeming ranks of Ken Stevens alumni!

I am also very grateful to my committee members for their guidance and support throughout my thesis research. Dr. Stefanie Shattuck-Hufnagel has become almost a co-advisor to me while Ken has been navigating his transition into well-deserved retirement, and she has been a wonderful mentor through the later years of my graduate study and my planning for the future. The final presentation of this thesis owes a great deal to Stefanie, and my work was helped immeasurably by her tremendously positive encouragement, as well as the lovely flowers with which she adorns our offices and lifts all of our spirits. Prof. Louis Braida also gave me great encouragement during my research, as well as many valuable suggestions, honest evaluation, and fresh perspective on my work.

Research in speech science at MIT has taught me the great value and satisfaction that is gained from close collaboration with other laboratories and departments that may be tackling the same problems from various different angles. In particular, I would like to thank Daryush Mehta, who provided me with software he coded for noise analysis of speech signals, as well as with numerous enlightening conversations about speech technology and analysis methods. I would also like to thank the members of Daryush's Voice Quality Study Group, from whose meetings I gained great insight informing the aerodynamic theories of this thesis. From the Linguistics department, I would especially like to thank Prof. Adam Albright, who provided me with essential background for the phonological issues treated in this thesis. For insight into recent advances in quantal/enhancement theory, I would like to thank Prof. Helen Hanson, whose collaboration with Ken Stevens continuously produces new fascinating topics for discussion.

Thanks are certainly due to the four speaking subjects and ten listening subjects who generously participated in my research experiments. Over the years, I have been attempting to hone my experimental techniques to become progressively less tedious and sleep inducing, but I am sure there is still much room for improvement on that front. I am very grateful to each of my subjects, as none of this work would have been possible without them.

Many people gave me important technical and logistical support in carrying out this research, and in navigating the complex institution that is MIT. I would like to thank Seth Hall, Joe Perkell, Satra Ghosh, Janet Slifka, and Miwako Hisagi for all of their assistance in learning the ins and outs of our laboratory hardware and software. Above all, I cannot give enough thanks to Arlene Wint, who is the Oracle of our lab group. Every random question I ever had I took to Arlene, and she was always able to help me with a swift solution or a sympathetic ear.

Warm thanks to all of the members of the Speech Communication Group, especially Xuemin Chi, Nancy Chen, Tony Okobi, and Yoko Saikachi, for their friendship and support during my five years in this wonderful lab. The level of collegiality, camaraderie, and true sense of family in the Speech Group is a rare gift, and I have been very glad to have this as my home at MIT. Special thanks also to Bill Cutter and the Music department at MIT, which was a joyful and fulfilling second home to me during grad school.

Love and thanks to my parents, Henry and Michele Hon, for their never-ending love, support, and guidance, and for making me the person that I am today. To my brother Gregory, and my sister Stephanie, for all the blessings of family. And finally, to my husband Kevin, for being "the best".

Table of contents

Table of figures

Table of tables

1 Introduction

1.1 Glides: /j/ and /w/

The research described in this thesis has as its focus a group of speech sounds that occupies a somewhat ambiguous classification between vowels and consonants. This thesis accepts the common practice of specifying this class of sounds using the term “glide”, a name which evokes the smooth transitions between these sounds and adjacent sound segments, in terms of both production and acoustics. Glides are always found in onset position within a syllable, and they directly precede the vowel nucleus of the syllable to which they belong. In the time region between a glide and a vowel, acoustic parameters such as formant frequencies and amplitudes change in a smooth and continuous transition between the two segments (Sun, 1996). These smooth formant movements are directly observable since the glides generally exhibit steady periodic voicing, exciting all of the visible/perceptible vocal tract formant resonances.

In English, the generally accepted set of glide segments is composed of /j/ and /w/. Although other glides besides these two have been reported cross-linguistically, they are not considered in this study, since they do not occur in English and are relatively rare in other languages. Ladefoged & Maddieson (1996) report that glides other than /j, w/ occur in less than 2% of the world’s languages; by contrast, 85% of languages use the palatal glide /j/, and 76% use the labial glide /w/. Thus, the glides /j, w/ are an important and ubiquitous component of human language, and the fact that they have not received the same depth of acoustic analysis as other types of sound segments is an inequity that deserves to be rectified.

The glides are produced by raising the tongue dorsum in order to produce a constriction with the palate (and also at the lips, for /w/). This constriction is narrow enough to weaken the spectrum amplitude of the glide segment relative to the adjacent vowel, but not narrow enough to cause the type of acoustic discontinuity that is present in a consonant (Stevens, 2002). The articulatory difference between glides and vowels or consonants is thus thought to be a matter of the degree of constriction in the vocal tract; the constriction should be narrower than that of a vowel, but not as narrow as that of a consonant.

The idea that glides occupy a phonological category of their own, especially as distinct from the related high vowels /i/ and /u/, has not been accepted by all researchers. Some prefer the term “semivowel” to that of “glide”, labeling them as vowel-like segments that only *function* like consonants (Ladefoged & Maddieson, 1996). Others deny *any* difference between glides and vowels, other than their relative positions in the syllable (e.g. Selkirk, 1984a). This thesis addresses the question of glide characterization from an acoustic standpoint, highlighting cues in the speech signal that code for the distinctive features of glides vs. other segments.

It is assumed in this study that sound segments are represented in the lexicon as bundles of binary distinctive features (Jakobson, Fant, & Halle, 1952). A distinctive feature is the smallest categorical unit which is capable of creating a contrastive distinction between words in a language. A change in the value of a single distinctive feature within a single sound segment in a word has the potential to create a different word; for example, the minimal pair “pat”/”bat” differs only in value of the feature [stiff vocal folds], which labels the first consonant of the word as being either voiced ([-stiff vocal folds]) or voiceless ([+stiff vocal folds]). The minimal pair “bat”/”bait” differs only in the value of the feature [low] for the vowel, and the minimal pair “pat”/”pass” differs only in the value of the feature [continuant] for the final consonant. Features such as [low] have been termed *articulator-bound* features, since they are tied to the actions of particular articulators in the vocal tract, in this case the tongue body. Features such as [continuant], on the other hand, are *articulator-free* features, since they specify the manner of articulation, but not which articulator is being used (Halle, 1992). Articulator-free features divide the total inventory of sound segments into major phonological classes, such as vowels, fricatives, stops, affricates, etc. If the glides in fact constitute a separate feature class from vowels, there should be an articulator-free feature related to their manner of articulation that makes this distinction.

This thesis aims to investigate the case for a distinctive feature specification for glides, through evidence gained from a study of the acoustics of glides in American English. The remainder of Chapter 1 provides a summary of current knowledge about the phonology, production, and acoustics of glides. Chapter 2 describes a new set of acoustic analyses of glides produced naturally by American English speakers, and Chapter 3 describes a perceptual study of the acoustic cues used by listeners to detect glides in speech signals. A discussion of the conclusions drawn from this work is given in Chapter 4, and Chapter 5 suggests avenues for future work and applications for the knowledge gained.

1.2 Phonology of glides

The vowel and consonant sound segments are widely accepted to inhabit very separate spaces in the distinctive feature inventory. Each has its own defining feature: [+vocalic] (or some variation thereupon) for vowels, and [+consonantal] for consonants. The glide class, however, is relegated to some ambiguous space between the other two classes, defined not by its own feature but by the negation of one or both of the others ([-vocalic, -consonantal], perhaps) (Jakobson & Halle, 1956; Chomsky & Halle, 1968, Kenstowicz & Kisseberth, 1979). The exact space it occupies is a matter of some debate among phonologists, since some characteristics of the glide class seem to overlap with the other two classes. Although the glides occupy syllable boundary positions like consonants, they are not normally considered to exhibit the [+consonantal] feature, for various reasons. For example, they are not produced with a “radical” obstruction in the vocal tract (Chomsky & Halle, 1968), they do not produce an abrupt discontinuity in the acoustic signal (Stevens, 1998), and they do not have any zeros in their acoustic spectra (Jakobson, Fant, & Halle, 1952). Rather, the debate generally centers on whether glides are different enough from vowels to merit their own feature class, as evidenced by the fact that glides are often termed "semivowels". Some generative phonologists have argued that there is no need for a feature distinction between the glides /j, w/ and the high vowels /i, u/, since they can be differentiated instead by syllable theory alone (Selkirk, 1984a).

Jakobson *et al*. (1952), for example, do not postulate a separate feature class for the glides /j, w/. Instead, they assume a rule in English, that unstressed /i, u/ become non-syllabic when adjacent to another vowel, as in "ye" (phonemically /iii/) or "woo" (phonemically /uuu/). Catford (1988) describes /j, w/ as vowels, but very short ones; in his view, a semivowel is formed when a vowel is not held, but is merely approached and then immediately released. By these accounts, the glides may differ from vowels in terms of syllable position or duration, but this does not constitute a difference of distinctive features. Selkirk (1984a) advocates replacing all of the major class features with a "sonority index" that ranks the high vowels /i, u/ as less sonorous than other vowels. In this system, a high vowel is perceived as a glide if it is adjacent to a more sonorous vowel, and "glidehood and vowelhood are defined with respect to context". Again, according to this view there is no phonological 'need' for a glide class as opposed to vowels.[1]

Despite some phonologists' desire to eradicate the glide class in search of a simpler inventory of features, several arguments have been raised in favor of categorizing glides separately from vowels. Parker (2002) points out that the movement to differentiate between glides and high vowels based solely on syllable position is circular reasoning, since syllabicity is currently only well defined based on the prior classification of vowels and non-vowels. The claim that glides are differentiated from vowels merely by their short duration is contradicted by the fact that geminate glides have been attested in a

[1] Selkirk's system's handling of two adjacent high vowels is somewhat awkward, however. In the word "you" (phonemically /iu/), for example, both high vowels are equally sonorous, and the choice of which to perceive as a glide must be lexically specified.

fair number of languages (Ladefoged & Maddieson, 1996). A recent survey of languages with gemination found that over half of them permit glides to be geminated, and that these languages are widespread geographically and linguistically (Maddieson, 2008). In Maddieson's review of acoustic studies, geminate glides display longer durations, steadier formant frequencies, and slower formant transitions than corresponding single glides, with which they contrast phonologically. Aoyama & Reid (2006) measured geminate glide durations in Guinaang Bontok averaging over 120 ms – well above the duration of many vowels (Stevens, 1998). Thus, neither brevity nor rapid movement can be considered an inherent property of the universal class of glides.

Chitoran (2002) showed that glide-vowel pairs and vowel-vowel diphthongs are perceptually different in Romanian, in addition to patterning differently phonologically. She demonstrated differences in duration and transition time between the glides and vowels, and also showed that glides were produced using more tongue contact with the palate than vowels. Padgett (2008) argues that glides are featurally different from vowels cross-linguistically, citing cases in which glides and vowels contrast, and phonological processes that treat glides differently from high vowels. He notes that this different treatment stems from differences in vocal tract constriction degree between glides and high vowels, pointing to narrower constriction as the root of a featural distinction between glides and all vowels. This is in agreement with Chomsky & Halle's (1968) identification of the feature [-vocalic] with any constriction greater than that in a high vowel. The [vocalic] feature has fallen into disuse since the advent of modern theories of syllabicity; however, its reintroduction to distinguish between glides and vowels has been advocated by phonologists such as Nevins & Chitoran (2008), who note that glides do not pattern as vowels in some phonological processes because of their different degree of constriction.

This proposal finds support in an acoustic study of the glides /j, w/ and the vowels /i, u/ in Amharic, Yoruba, and Zuni by Maddieson & Emmorey (1985), who found that glides are indeed produced with greater constriction degree than their corresponding vowels cross-linguistically. The narrower constriction for the glides was evident acoustically from a lower frequency of the first formant (F1) for both glides, a lower frequency of the second formant (F2) for /w/, and a higher frequency of the third formant (F3) for /j/. The acoustic measurements in Maddieson & Emmorey's study are somewhat problematic, however, in that they were visually estimated from spectrograms, which present challenges for the accuracy and repeatability of estimated formant values. The acoustic analyses pursued in this thesis address this accuracy problem by using more objective spectral measurement techniques, as described in the next chapter. A larger selection of acoustic measurements was also carried out, in addition to formant frequency measurements, in order to evaluate as many potential cues to the presence of glides as possible.

This thesis begins to remedy the dearth of detailed acoustic studies in the current literature on glides. Acoustic analysis is relatively rare in phonological descriptions of the relationship between glides and vowels, and where it does appear it is often limited to rough formant measurements from spectrograms (e.g. Maddieson & Emmorey, 1985; Chitoran, 2002). A certain amount of "eyeballing" is often required in this measurement method, since formant peaks may appear quite broad in spectrograms, and the gray-scale representation of amplitude has inherent visual limitations. Any attempt at spectral analysis is welcome, however; the majority of production and perception studies on glides have focused only on duration and transition rate measurements (e.g. Lehiste & Peterson, 1961; Miller & Liberman, 1979; Miller & Baer, 1983; Mack & Blumstein, 1983; Chitoran, 2002), and more of these compare glides to stop consonants (i.e. /w/-/b/) than compare glides to vowels.

In addition to measurements of formant frequencies in glides and vowels, there is a need for spectral studies of the acoustic cues generated by the interaction between the glottal source and the vocal tract filter in these sound segments. As mentioned in Section 1.1, the distinctive feature differentiating glides from vowels (as well as that differentiating glides from other consonants) should be an articulator-free feature, since it specifies the manner of vocal tract constriction without being limited to the use of a particular articulator. According to a new aspect of feature theory proposed by Stevens & Hanson (in press), articulator-free features arise from *aero-mechanical interactions*, the aerodynamic consequences of airflows and pressure drops in a vocal tract with various constrictions along its length. (Articulator-bound features, on the other hand, arise from *acoustic resonator coupling*, including interactions among formant frequencies, according to this new aspect of the theory.) The glide class feature should therefore be acoustically related to the aerodynamic effect of the narrow oral constriction on the airflow from the glottis, as described in the next section.

In addition, the labels given to the distinctive features that distinguish between major classes of sound segments should capture generalities about which classes are alike in terms of acoustics, production, and phonology. For example, [-sonorant] groups stop consonants and fricatives together; they are alike in their obstruent production, and they also often pattern together phonologically and in opposition to [+sonorant] consonants. For glides, there is more than one option available, assuming that a featural distinction from vowels is warranted. Stevens & Hanson (in press) include in their inventory a feature [glide], which sets the glides completely apart from both vowels and consonants. In their hierarchical feature system, vowels are [-glide], glides are [+glide], and all other consonants are unspecified for that feature. Chomsky & Halle (1968) use the feature [vocalic] instead of [glide], creating the possibility for glides to be grouped with other consonants in opposition to vowels. That is, the feature [-vocalic] is shared by glides and all other consonants, while the feature [-consonantal] is shared by glides and vowels. In Chomsky & Halle's system, the double specification [-vocalic, -consonantal] is shared only by glides and the laryngeal consonants /h, ʔ/. Glides have been known to pattern phonologically with the laryngeal consonants in some languages (Parker, 2002); they have also been described as patterning with other sonorant consonants such as liquids (Levi, 2008). Ultimately, the feature inventory chosen has most value when the available feature categories can be used to describe the sets of sounds that participate together in various phonological processes. The acoustic data gathered in this thesis may be brought to bear on the decision of which feature label is most appropriate for glides, by comparison with the acoustics and inferred articulation for other sound segments with which the glides might potentially be grouped.

1.3 Production and acoustics

The speech production stage of this research focuses on identifying and measuring potential acoustic correlates of the glide feature class, as distinct from the vowel and consonant classes. Stevens (1998) defines glides as “a class of consonants produced with a constriction that is not sufficiently narrow to cause a significant average pressure drop across the constriction during normal voicing”. The lack of a vocal tract closure producing a *significant* pressure drop is a clear distinction between glides and other consonants; its acoustic correlate is a lack of abrupt discontinuity in the acoustic signal (Stevens, 2002). The dividing line between glides and vowels (especially the closely related high vowels), however, has heretofore been less clearly defined. Chomsky & Halle (1968) suggest that the [-vocalic] feature that differentiates glides from vowels is defined by a constriction that is greater in degree than that for a high vowel; but the threshold boundary required to create this category distinction along the continuum of constriction degrees has not been established in terms of articulation and acoustics. The span of constriction degrees that are slightly less narrow than those of full consonants may encompass additional aerodynamic effects, short of significant pressure build-up with turbulence noise, whose acoustic consequences should be investigated.

Both of the standard American English glides /j/ and /w/ are produced with relatively narrow constrictions in the oral cavity which makes up the front part of the vocal tract. Figure 1 shows x-ray

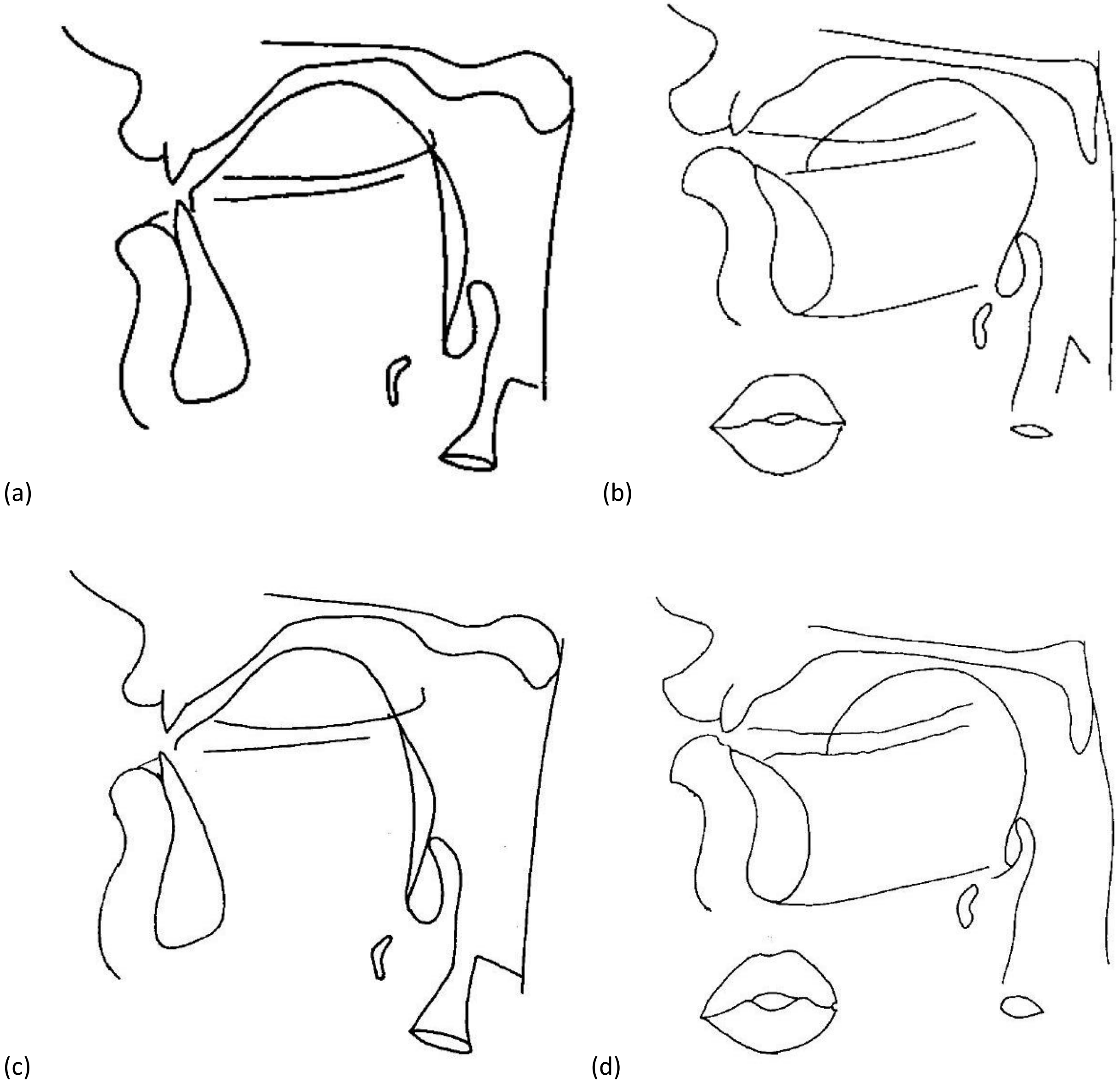


Figure 1: Midsagittal sections of the vocal tract for the glides /j/ (a) and /w/ (b), as well as the corresponding high vowels /i/ (c) and /u/ (d). For /u/ and /w/, the frontal lip contour is also shown, to illustrate the lip rounding. Note that, in these examples, the tongue body appears higher and closer to the hard palate in /j/ (a) than in /i/ (c), and the lip opening appears smaller in /w/ (b) than in /u/ (d). From Bothorel *et al.* (1986).

tracings of the vocal tract configurations for both of these glides, as given in Bothorel *et al.* (1986). For the palatal glide /j/, the tongue body is raised and fronted to create a long constriction with the hard palate. The configuration is similar to that for the high front vowel /i/, but with a narrower constriction in the palatal region. For the labial glide /w/, the lips are rounded to create a narrow and extended opening, and the tongue body is raised and backed to create a secondary constriction in the velar region. The configuration is similar to that for the high back vowel /u/, but again with a narrower primary constriction at the lips. Cross-sectional areas for constrictions in glides are expected to be in the range of 0.2-0.4 cm^2, while areas for vowels are larger (Stevens, 1998).

For both vowels and glides, there is a steady voicing source at the glottis, and it can be assumed that there is no significant acoustic coupling to the subglottal or nasal cavities; therefore the sound output at the mouth is the result of the filtering of the glottal source by an all-pole vocal tract transfer function. The lowest resonant frequency (F1) of this transfer function for the glides /j, w/ and high vowels /i, u/ can be modeled as a Helmholtz frequency:

$$F1^{'} = \frac{c}{2\pi\sqrt{\frac{V\mathrm{l}_\mathrm{c}}{A_c}}}$$

where:

F1’ = the first formant frequency neglecting the effect of yielding vocal tract walls

c = the speed of sound

V = the volume of the air cavity behind the constriction

l_c = the length of the constriction

A_c = the cross-sectional area of the constriction

For both the glides and the high vowels, the first formant frequency is made relatively low by creating a large cavity behind a long constriction, causing the terms *V* and l_c to be large. (In /w/ and /u/, both the labial and the velar constrictions contribute to the lowering of F1 (Stevens, 1998).) Low first formant frequency is a correlate of the articulator-bound feature [+high], which is shared by the high vowels and glides. For the glides, the constriction is narrower than for the high vowels, making the area term A_c smaller, and thus making *F1′* smaller. The frequency of the first formant is therefore expected to be somewhat lower for glides than for high vowels. However, vocal tract wall effects may limit the degree to which F1 can be lowered in this range.

When the first formant frequency is approximated using the assumption that the vocal tract walls are hard, it is possible for F1 to decrease to zero when a complete closure is made. This is not possible in reality, however, since the vocal tract in fact has yielding walls. The first formant frequency with a closed vocal tract is given by:

$$F1_c = \frac{\sqrt{A_w}}{2\pi\sqrt{M_{sw}\,C_A}}$$

where:

A_w = the surface area of the vocal tract walls

M_{sw} = the mass of the walls per unit area

C_A = the acoustic compliance of the closed cavity

Fant (1972) and Fant *et al.* (1977) report typical values of $F1_c$ to be around 190 Hz for males and 220 Hz for females. This represents the lower limit of the first formant frequency due to constrictions in the vocal tract. It cannot reach zero because the mass term M_{sw} of the vocal tract walls cannot be infinite.

When a constriction is made in the vocal tract, but not a complete closure, the actual frequency of the first formant is calculated as:

$$F1 = \sqrt{(F1')^2 + (F1_c)^2}$$

where *F1′* is calculated as above, assuming hard walls. Figure 2 shows a graph of F1 vs. F1′, from Stevens (1998, p. 159). The actual F1 approaches the limit $F1_c$ as the constriction is narrowed to bring F1′ close to zero. Note that the yielding walls have greater effect as F1 becomes very low. In this low

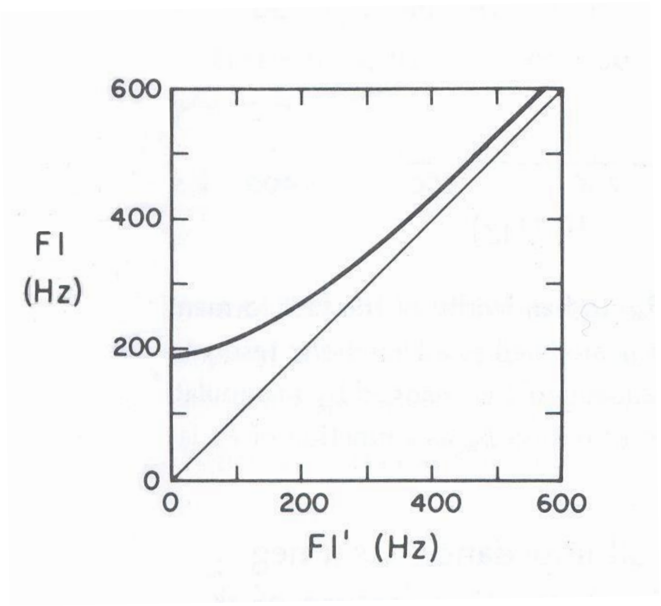


Figure 2: Natural frequency F1 for a constricted vocal tract with yielding walls, as a function of natural frequency F1' computed on the assumption of hard walls (i.e., M_{sw} = ∞). Deviation of the curve from the diagonal line is a measure of the effect of the walls. From Stevens (1998).

region, F1 is less sensitive to changes in the constriction or vocal tract characteristics, and the curve in Figure 2 becomes relatively flat. Although the absolute lower limit of F1 may be around 190 Hz (for males), such a low frequency is only achieved by constricting the vocal tract to such a degree that pressure is built up behind the constriction, and turbulence noise may be generated. Stevens (1998) estimates that F1 for glides may only lower to about 260 Hz before generating significant pressure build-up. If F1 is made lower, the intraoral pressure may become so large as to generate turbulence noise at the constriction.

The effects of the yielding vocal tract walls become more pronounced as a constriction is narrowed and F1 becomes lower, suggesting that wall effects will be more apparent in glides than in vowels. There may be a range of constriction degrees for which F1 remains relatively unchanged, as the oral constriction is narrowed slightly more than in a high vowel. Another effect of the walls is their contribution to the bandwidth (B1) of the first formant peak. Since the walls do not have infinite impedance, some acoustic energy is lost through them, leading to increased formant bandwidths. The bandwidth contribution of the walls of a constriction tube can be calculated as (Stevens, 1998, p. 157):

$$B_w = \frac{G_{sw} S \rho c^2}{2\pi A}$$

where:

G_{sw} = the specific acoustic conductance of the walls

(This term is larger around F1 than at higher frequencies.)

S = the cross-sectional perimeter of the tube

ρ = the density of air

c = the speed of sound in air

A = the cross-sectional area of the tube

Another cause of increased first formant bandwidth in glides is the kinetic pressure drop due to the acoustic resistance of the constriction. As F1 for glides is modeled as a Helmholtz resonance, this bandwidth contribution can be calculated as (Stevens, 1998, p. 163):

$$B_c = \frac{U}{2\pi l_c A_c}$$

where:

U = the volume velocity of air traversing the constriction

l_c = the length of the constriction

A_c = the cross-sectional area of the constriction

This bandwidth contribution is comparable to those from other sources, such as the vocal tract walls, and increases as the constriction is narrowed, decreasing the area term A_c. Adding together the bandwidth contributions from the yielding walls and the pressure drop across the narrow constriction,

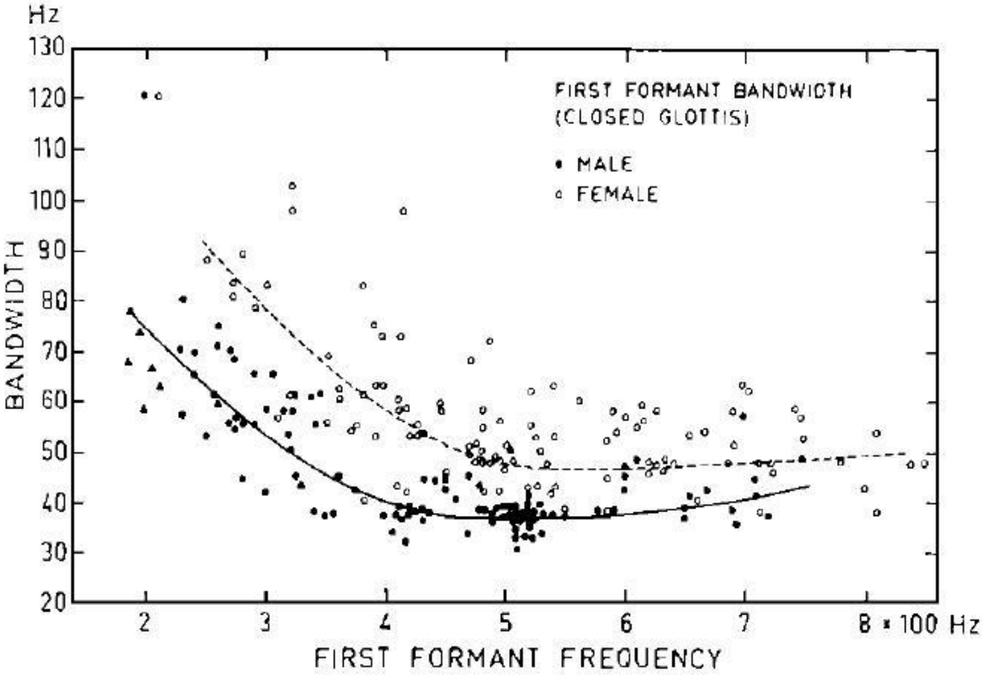


Figure 3: "Bandwidth values for the first formant plotted against the formant frequency. Each closed circle represents a vowel sample of one of three male subjects, and an open circle represents a sample of one of three female subjects. Representative values are estimated by visual inspection of the plots, and curves are drawn for male and female subjects separately. Bandwidth values for articulations with bilabial closures by a male subject are also added in this graph (closed triangles)." (Fujimura & Lindqvist, 1971)

the overall bandwidth (B1) of the first formant for glides is expected to be around 100-150 Hz, which is larger than that for most vowels (Stevens, 1998).

Fujimura & Lindqvist (1971) used an external sweep-tone signal and an analysis-by-synthesis procedure to collect data on first formant frequencies and bandwidths in vowels for male and female speakers. Their data are given in Figure 3. Note that B1 increases in both curves as F1 is made lower for more constricted vowels. Fujimura & Lindqvist also collected data from one male speaker for the stop consonant /b/, with full closure in the vocal tract. These data points lie at the upper left end of the F1-

B1 curve for male subjects. The F1-B1 relationship for glides is expected to fall along the curve between the /b/ data points and the adjacent points for high vowels, since the constriction for glides is intermediate between that of a vowel and that of an obstruent consonant.

An additional effect of the pressure drop across the narrow constriction in a glide is to change the shape of the waveform of the acoustic source at the glottis. Since the total pressure drop from the lungs to the output at the mouth must be equal to the sum of the pressure drops across individual vocal tract elements, the following equation can be written (Stevens, 1998, p. 93):

$$\Delta P = R_g U_g + M_A \, dU_g/dt + R_c U_g$$

where:

ΔP = the pulmonary or alveolar pressure

R_g = the resistance of the voicing constriction at the glottis

U_g = the volume velocity of air traversing the glottis

M_A = the acoustic mass of the air in the vocal tract

R_c = the resistance of the vocal tract constriction

The increased narrowing of the constriction in a glide (relative to a vowel) causes increases in both the mass term M_A and the resistance term R_c. If ΔP is assumed to be constant, then these increases must be balanced by changes in the glottal volume velocity U_g. These changes are illustrated in Figure 4, from Stevens (1998, p. 519). The narrowing of the vocal tract constriction causes a decrease in the peak amplitude of each glottal pulse, as well as an airflow delay in the open phase which skews the waveform to the right (Fant, 1983). In addition, the high airway impedance causes pressure fluctuations immediately above the glottis, which influence the mechanical motion of the vocal folds. The result is an increase in open time of about 10% during the glottal pulse (Bickley & Stevens, 1986). The combined effects on the glottal waveform reduce the overall spectrum amplitude during the constricted interval.

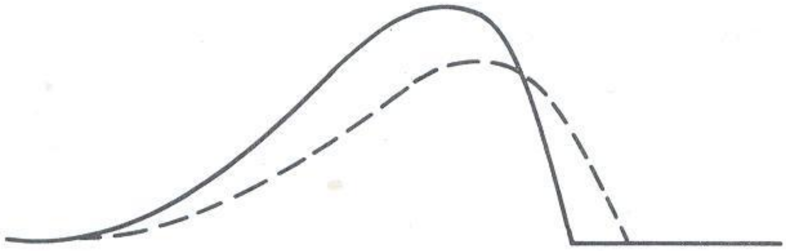


Figure 4: The solid line shows the typical shape of a pulse of glottal volume velocity for an open vowel. The dashed line indicates schematically the modification of this pulse for a glide which is produced with a relatively narrow constriction in the vocal tract. From Stevens (1998).

Bickley & Stevens (1986) found an additional effect on the glottal source, from the increased vocal tract impedance due to a narrow constriction in the vocal tract, in a decrease in the fundamental frequency of phonation (F0). They found that the increase in glottal open time was not completely offset by an equal decrease in closed time, resulting in an increase in the glottal period. This effect has been modeled in recent studies of acoustic loading on the vocal folds (Zañartu *et al*., 2007; Titze, 2008), which predict that F0 becomes more decreased as the reactance of the vocal tract becomes more inertive. Figure 5, from Titze (2008), shows that the greatest F0 decrease occurs when the intended phonation frequency corresponds with the frequency of the peak in the vocal tract inertance, which occurs just below the frequency of the first formant. Titze *et al*. (2008) observed this effect in natural speech, which demonstrated F0 perturbations when F0 crossed F1. This supports Bickley & Stevens' (1986) finding that F0 decreases more for females than for males with the same vocal tract constriction area, since females' baseline F0 is higher and closer to F1 than males'. Such effects on F0 may be expected in glides produced with narrow constrictions and low F1 that may come close to F0.

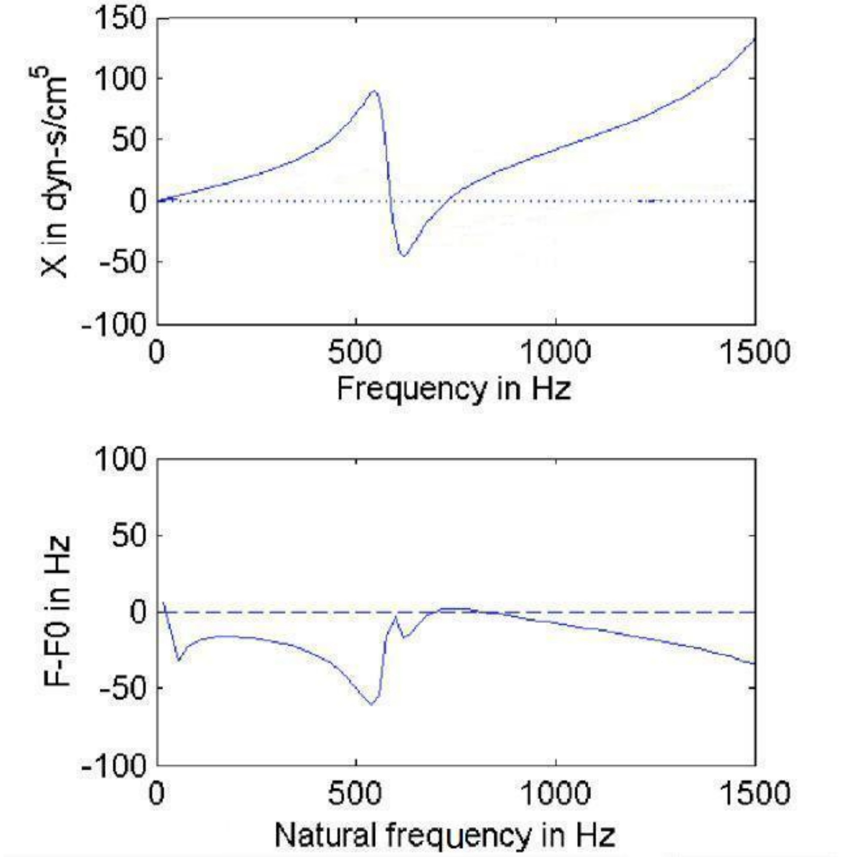


Figure 5: (top) Vocal tract reactance curve. (bottom) The difference between the fundamental frequency F0 and its frequency in the absence of vocal tract loading. From Titze (2008).

Some of the disparate acoustic consequences of forming glides with constrictions narrower than those in vowels may have cooperating contributions toward a combined acoustic cue for glides. For instance, reduction of the first formant frequency (F1), increase of the bandwidth of the first formant peak (B1), and reduction of the amplitude of the glottal pulse could all contribute to the reduction of the overall amplitude of the acoustic signal during the glide segment. Decreasing F1 reduces the amplitude of all higher formants, since they ride on the “skirt” of the frequency response of the lower pole. Increasing B1 reduces the amplitude of the first formant peak, since the amplitude of any formant peak is inversely proportional to its bandwidth. Reducing the amplitude of the glottal waveform causes a direct reduction of the overall signal amplitude, and Stevens (1998) suggests that this may be the principal cause.

1.4 Summary of potential acoustic cues

The acoustic analyses in this research investigate several potential acoustic correlates of the [-vocalic] feature that differentiates glides from related vowels, as motivated by the acoustic theory described in Section 1.3. These acoustic correlates are expected to relate to the production mechanism of a narrow constriction, which is of a greater degree than that found in a high vowel but not great enough to produce a consonantal pressure drop. The following are examples of potential acoustic cues for glides:

$\underline{A_{RMS}}$: Proponents of the idea that the glide-vowel distinction is based on a lack of syllabicity or sonority tend to agree that the most likely perceptual correlate is a lack of intensity or loudness as compared to the adjacent vowel (Parker, 2002; Padgett, 2008; Selkirk, 1984). Glides, inhabiting the syllable boundaries, should have a weaker intensity than the vowels at the syllable nuclei; RMS amplitude (A_{RMS}) provides a quantitative measure of this intensity relationship. From a production standpoint, the decrease in amplitude of the acoustic signal during the glide segment may have more than one contributing factor. Constricting the oral cavity causes a decrease in the first formant frequency (F1), which contributes to the overall amplitude reduction due to the transfer function characteristics of the formants (Stevens, 1998; Fant, 1962). Producing a narrow constriction also causes a reduction in the transglottal pressure during the rising phase of the glottal pulse, modifying its shape to one of reduced amplitude (Fant, 1997; Stevens, 1998). Losses from the yielding vocal tract walls also cause an increase in the bandwidth of the first formant (B1), contributing to the reduction in overall amplitude of the signal. Stevens (1998) suggests that reduced low-frequency amplitude before a vowel is the principal acoustic requirement for a glide segment.

F1: The narrow constriction in the front part of the vocal tract for a glide has the effect of decreasing the frequency of the first formant peak (F1) relative to that of a vowel. This decrease in F1 contributes to the decrease in overall amplitude of the glide segment, accentuating the loudness contrast with the following vowel. The vocal tract configurations for the palatal glide /j/ and the rounded labial glide /w/ are such that F1 may be made as low as possible, since both create a Helmholtz resonance with a large cavity volume behind a long narrow constriction. For the labial glide /w/, a velar constriction is also formed, contributing to the lowering of F1. Because of the finite acoustic mass of the vocal tract walls, there is a limit to the lowest frequency that F1 can achieve in a glide configuration; Stevens (1998) estimates this frequency to be about 260 Hz for an adult male. The value may be slightly higher for females, but the wall loss effects cause reduced sensitivity to differences in vocal tract dimensions for F1.

B1: The bandwidth of the first formant (B1) is expected to be larger for glides than for vowels, again because of the narrower constriction in the vocal tract for the glide segments. One contributing factor is the loss caused by the acoustic resistance due to the kinetic pressure drop across the narrow constriction. Another is the vocal tract wall losses, which come into play when F1 is low. Losses at the glottis will also contribute, resulting in an expected bandwidth of 100-150 Hz (as compared to about 80 Hz for high vowels) (Stevens, 1998).

OQ: Open quotient (OQ) is the percentage of the glottal vibratory cycle during which the vocal folds are open and do not touch each other. It is expected to be larger for glides than for vowels, due to the skewing of the glottal waveform that is caused by the increased airway impedance from the narrow constriction in the oral cavity (Bickley & Stevens, 1986; Stevens, 1998). The acoustic consequence of an increase in OQ is an increase in the magnitude difference between the first two harmonics (H1 – H2) of the speech spectrum (Klatt & Klatt, 1990; Sundberg *et al.*, 2005).

HNR: Several researchers have identified frication noise or turbulence as a negative correlate of sonority (Parker, 2002), suggesting that a less sonorous sound segment (such as a glide) might exhibit more of a noise component in its acoustic signal than a more sonorous sound segment (such as a vowel). Padgett (2008) claims that the tendency toward frication noise due to a narrow vocal tract constriction is a key featural distinction between glides and vowels. Palatalizing mutations (similar to "that you" becoming "thatchoo" in English) partly arise because "a stop release through a narrow constriction (such as that of a glide) is turbulent, and can be perceptually reanalyzed as affrication." Turbulence noise may be allowed into the acoustic signal of a glide as a side effect of narrowing the vocal tract constriction to achieve a first formant frequency lower than the minimum vocalic threshold. Stevens (1998) estimates that F1 cannot be lowered beyond about 260 Hz if pressure build-up is to be avoided, but the allowance of some turbulence noise may permit further lowering of F1, enhancing the low-frequency contrast between the glide and the following vowel. In addition, the increased open quotient of the glottal waveform, skewed by the aerodynamic effects of the narrow oral constriction, could possibly manifest acoustically in additional aspiration noise during the glide segment. A measure targeting both of these turbulence contributions is the harmonics-to-noise ratio (HNR), the ratio of the power in the voicing component of the sound signal to that of the noise component.

F0: The aerodynamic effects of the oral constriction on the glottal source may also have the effect of decreasing the fundamental frequency of phonation (F0) in a glide relative to that in an adjacent vowel (Bickley & Stevens, 1986). Acoustic modeling suggests that this effect will be most pronounced when F0 and F1 are close together. Because F0 is strongly affected by prosodic considerations, this source effect may be variably present in different prosodic environments.

1.5 Approaches to study

The main objective of this thesis is to complete a detailed and comprehensive acoustic analysis of canonically produced glides in American English, focusing specifically on those characteristics which separate them from the closely related high vowels. The study is novel in the breadth of potential acoustic cues to glidehood covered, and the detailed analysis which each cue is given. Continuous spectral measurements are taken over the entire duration of each glide, such that the glide landmarks and their acoustic targets can be located with the maximum degree of temporal precision. In addition, the database of natural recordings created (to be described in Chapter 2) contains glides in combination with all of the English tense vowels, in a balanced set of controlled prosodic contexts. The perceptual experiment presented in Chapter 3 provides a ranking for the glides' acoustic cues in terms of perceptual salience, as well as insights into the relationships between glides and other sound segments. Finally, a large set of acoustic evidence can be brought to bear on the issue of glide identity, with applications for phonology, recognition, synthesis, and general understanding of speech production and perception.

2 Acoustic analyses of intervocalic glides

2.1 Database of recordings

The acoustic analyses presented in this chapter were performed on a database of recordings of natural speech created specifically for this study. The database contains tokens of intervocalic glides produced by four native speakers of American English, two female (labeled 'F1' and 'F2') and two male ('M1' and 'M2'). Each target nonsense token consisted of one of the two glides /j, w/ flanked on both sides by one of the six English tense vowels /i, u, e, o, æ, ɑ/. These six vowels represent all possible combinations of the vowel features [high], [low], and [back], as illustrated in Table 1. The same vowel context was used on either side of the glide in each target token, resulting in vowel-glide-vowel (VGV) tokens such as /ojo/, /ɑwɑ/, etc. The inclusion of all six vowel contexts in this study represents a significant increase in completeness and complexity over previous analyses of glide production. For example, Chitoran (2002) was only able to investigate glides preceding the vowel /ɑ/, due to the desire to compare minimal pairs with corresponding vowel-vowel diphthongs in real Romanian words. Maddieson & Emmorey (1985) included three vowel contexts /i, ɑ, u/ in their study of glide production in three different languages, but they were essentially only able to draw cross-linguistic conclusions from the tokens of glides flanked by their "cognate vowels" (i.e., from the /iji/ and /uwu/ tokens), due to the inability to control for the effects of coarticulation between glides and vowels of differing height and backness. The present study investigates the complete set of vowel place distinctions in an effort to catalogue the coarticulatory effects of different vowel contexts on adjacent glides, as well as to identify acoustic cues which may be less sensitive to such coarticulation.

Table 1: Distinctive feature specification of the six tense vowels in English. The horizontal categories correspond with the high/mid/low height distinction, and the vertical categories correspond with the front/back distinction among vowels.

	[+ high, - low] (high)	**[- high, - low] (mid)**	**[- high, + low] (low)**
[- back] (front)	/i/	/e/	/æ/
[+ back] (back)	/u/	/o/	/ɑ/

An additional level of complexity is added and controlled in this study by specifying the prosodic contour in which the target glide occurs. Five different prosodic contexts, varying based on the location and type of pitch accent with respect to the target glide, were created by embedding the target tokens within specific carrier phrases. The statement carrier phrase, “He said /VGV/ again,” was used to elicit a high pitch accent within the target token. The location of the pitch accent was controlled by directing the subject to emphasize either the first or second vowel in the target token. The question carrier phrase, “He said /VGV/ again?” was used to elicit a low pitch accent within the target token, again with the pitch accent location controlled by directing the subject to emphasize either the first or second vowel. A fifth prosodic context, in which no pitch accent occurred during the target token, was elicited using the carrier phrase, “PLEASE don’t say /VGV/ again.” Instructions to the subject to emphasize the word “PLEASE” ensured that the single pitch accent in this phrase was located three syllables earlier than the target token. This distance is required for the target token not to demonstrate the effects of any pitch accent, since accent-related acoustic effects have been found to spread across two neighboring syllables (Okobi, 2006). Throughout this thesis, the five prosodic contexts produced in this study will be abbreviated as follows:

H1: High pitch accent located on the vowel preceding the target glide, e.g., “He said /ɑ̲jɑ/ again.”

H2: High pitch accent located on the vowel following the target glide, e.g., “He said /ɑjɑ̲/ again.”

L1: Low pitch accent located on the vowel preceding the target glide, e.g., “He said /ɑ̲jɑ/ again?”

L2: Low pitch accent located on the vowel following the target glide, e.g., “He said /ɑjɑ̲/ again?”

NP: No pitch accent during the target token, e.g., “PLEASE don’t say /ɑjɑ/ again.”

To the author’s knowledge, this is the first acoustic study of glides to consider multiple controlled prosodic contexts and their effects on glide production.

All combinations of the 2 glides /j, w/ x 6 vowel contexts /i, u, e, o, æ, ɑ/ x 5 prosodic contexts (H1, H2, L1, L2, NP) were elicited from each of 4 speakers (F1, F2, M1, M2), for a total of 240 tokens in the database. Recordings were made in a sound-attenuating chamber in the laboratory of the Speech Communication Group at MIT, with the subject seated with a fixed microphone approximately six inches from the lips. The subject was prompted by text appearing on a computer monitor to read the phrases that were displayed on the screen. Incorrect vowel or prosodic productions were corrected through verbal instructions from the experimenter and immediately re-recorded. Recordings were made in five consecutive blocks, corresponding to the five prosodic contexts elicited. The decision was made to record each prosodic context in a separate block, because preliminary experiments showed that subjects had difficulty correctly producing the desired vowel and pitch accent combinations when both were varied simultaneously and cued textually on the computer monitor. Within each prosodic block, the glide and vowel combinations were elicited in random order. The recorded speech was low-pass filtered and digitized at a sampling rate of 10 kHz, and the target /VGV/ tokens were excised from their carrier phrases for the acoustic analyses to be described in the following sections.

2.2 RMS amplitude (A_{RMS})

2.2.1 Method

Most of the acoustic measurements presented in this chapter were made using the xkl software developed by Dennis Klatt for the Speech Communication Group at MIT. The software provides for simultaneous viewing of the full-scale and zoomed-in waveform, spectrogram, and spectral slices from Hamming-windowed segments of a speech recording. For most measurements, the length of the analysis window was set to be equal to the length of a single pitch period, since this has been shown to be the preferred window length for quasistationary analysis of speech signals with rapidly varying spectral characteristics (Smits, 1994). Spectral analysis using short-time Fourier transforms makes the assumption that the speech segment within the analysis window is essentially stationary; for sound segments with rapid transitions such as glides, the window must be relatively short in order to provide accurate measurements. In addition, we are interested in charting the exact time-course of the movements of various acoustic characteristics through the transitions into and out of glides; therefore very fine time resolution is desirable. In order to pinpoint and describe the target of maximum articulatory excursion in a glide, ideally the window length should be as short as possible to maximize temporal precision. However, the length of one glottal cycle is the minimum period of interest, since the acoustic variations between the closed and open phases within each cycle add confounding effects to measurements using shorter windows.

The analysis window was therefore maintained at an optimal length equal to the length of the pitch period under analysis. For each /VGV/ token, repeated measurements were taken over a continuous interval spanning from the midpoint of the first vowel, through the glide segment, to the midpoint of the second vowel. One measurement was made at each pitch period during that interval, and the analysis window was updated before each measurement to equal the length of that pitch period. The

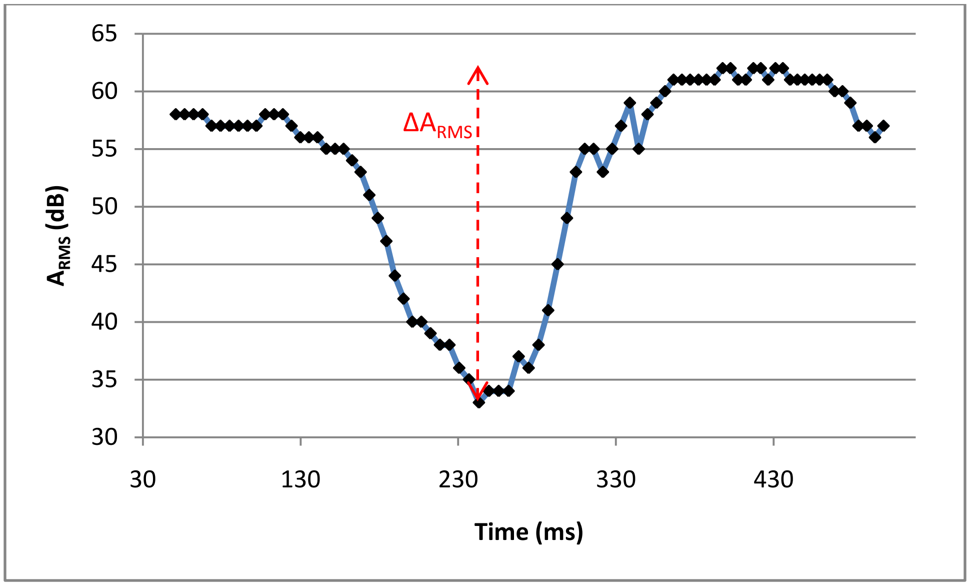


Figure 6: Measured A_{RMS} for a token of /æjæ/ with high pitch accent on the second vowel, produced by speaker F1. Measurement points from individual pitch periods (black diamonds) are connected by straight lines to form a continuous contour. The measurement interval extends from the midpoint of the first vowel /æ/, through the midpoint of the glide /j/ (around time 240ms), to the midpoint of the second vowel /æ/. The red line represents the ΔA_{RMS} measurement, calculated as the difference between the minimum amplitude of the glide segment and the maximum amplitude of the adjacent vowel.

pitch period lengths were calculated as the distance between the waveform zero-crossings just before the maximum amplitude excursions of the pitch periods, corresponding to the time between the beginnings of the closed phases of the glottal cycles. For each pitch period, the analysis window was centered on the closed phase of the cycle, in order to achieve maximum accuracy in capturing the excitation of the vocal tract by the glottal source (Smits, 1994). The RMS amplitude (A_{RMS}) of the signal segment under the analysis window was calculated automatically by the xkl software.

Figure 6 shows an example of A_{RMS} measurements made from a single token of /æjæ/ produced by female speaker F1. The individual measurements from consecutive pitch periods are connected with straight lines to form a continuous contour. Note the decrease in amplitude that occurs from the first vowel segment to the glide at around 240 ms, and the matching increase in amplitude returning to the following vowel. The magnitude of this amplitude excursion was quantified for each token in a single measurement ΔA_{RMS}, calculated as the difference between the minimum amplitude of the glide segment and the maximum amplitude of the adjacent vowel.

2.2.2 Results

The results of A_{RMS} measurements from all of the tokens in the /VGV/ recordings database show that a decrease in amplitude is a very consistent acoustic characteristic of glides in relation to adjacent vowels. The average amplitude decrease from vowel to glide across all four speakers and all contexts is 14.6 dB. 95% of the glides in the database were produced with an amplitude decrease of 5 dB or more. 90% were produced with an amplitude decrease of 7 dB or more. In fact, two of the speakers (F1 and F2) produced all their glides with an amplitude decrease of 7 dB or more.

Figure 7 shows the average ΔA_{RMS} for each speaker, averaged across both glides and all vowel and prosodic contexts. Significant inter-speaker differences are clear from these data, with speaker F1 producing the largest average ΔA_{RMS}, and speaker M2 producing the smallest average ΔA_{RMS}. Since the minimum ΔA_{RMS} was at least 3 dB for speakers F1, F2, and M2, it is clear that their typical glide production involved significant amplitude decrease from the vowel to the glide. Speaker M2 produced a few glides without visible amplitude decrease, but a t-test confirms that his average ΔA_{RMS} was significantly above zero (t(59)=14.12; p =.000). Glides thus exhibit significantly decreased amplitude compared to their adjacent vowels for all four speakers.

Analyses of variance were conducted to determine the effects on ΔA_{RMS} of the various factors controlled in this study: glide segment /j, w/, vowel height context (high, mid, low), and prosodic context (H1, H2, L1, L2, NP). Since significant interactions were found between speaker and all other factors, separate ANOVAs were conducted for each speaker's data. To compensate for the artificially increased

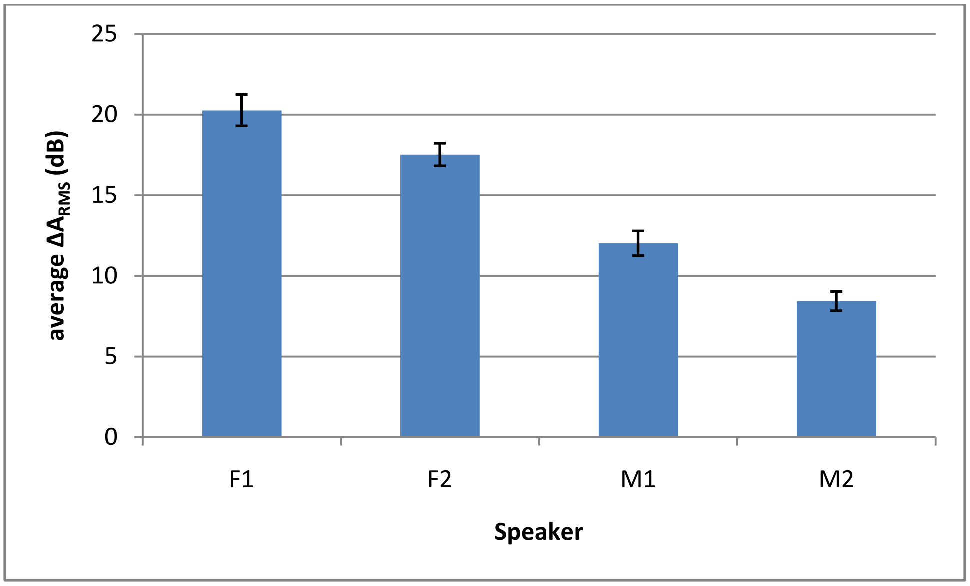


Figure 7: Measured ΔA_{RMS} for each speaker in the /VGV/ database, averaged across glide segments, vowel and prosodic contexts. Error bars represent standard error of the mean.

likelihood of Type I errors arising from the use of multiple statistical tests in this study, a conservative threshold level ($\alpha = .01$) was selected *a priori* to determine significance throughout this thesis. No significant interactions between factors other than speaker were observed for the ΔA_{RMS} measure.

Average ΔA_{RMS} was larger for the glide /w/ than for the glide /j/ for all four speakers, although this difference was statistically significant only for speakers F1 (F(1,30)=22.579; p=.000) and M1 (F(2,30)=15.48; p=.000). Figure 8 shows the average ΔA_{RMS} for each speaker, separated by glide segment /j, w/. The difference in amplitude reduction between /j/ and /w/ tokens is likely due to the difference in spectral tilt between the two different glide segments. In /w/, the backed tongue body has the effect of lowering the second formant frequency (F2), which reduces the overall spectrum amplitude in the same way that lowering F1 does (the higher formants riding on its "skirt" are reduced in amplitude by the frequency reduction of F2). The lowering of F2 for /w/ adds yet another source of amplitude reduction to those already mentioned in Section 1.3 for glides in general. In /j/, by contrast, the fronted

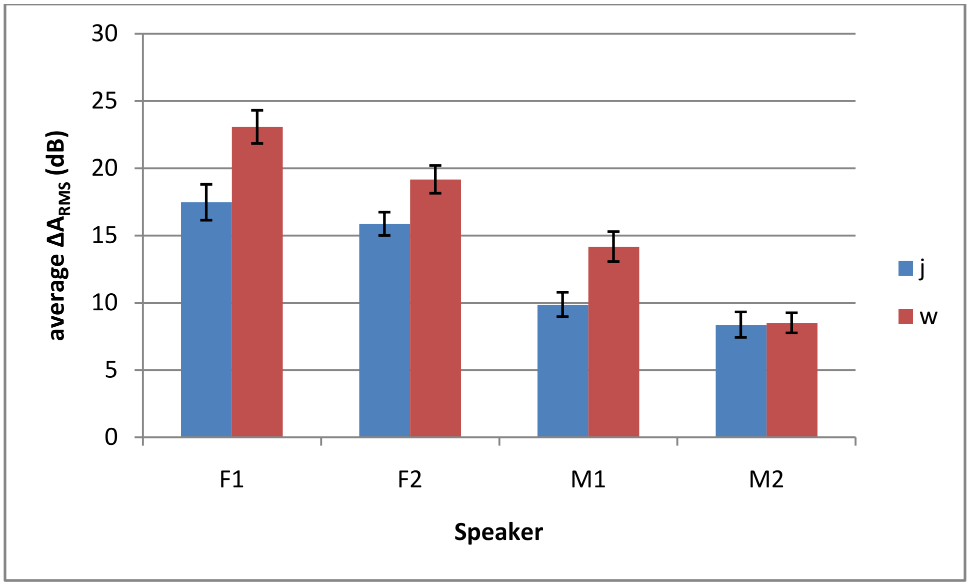


Figure 8: Measured ΔA_{RMS} for each speaker in the /VGV/ database, separated by glide segment /j, w/, and averaged across vowel and prosodic contexts. Error bars represent standard error of the mean.

tongue body has the effect of raising F2, which boosts the amplitude of the higher formants and decreases spectral tilt, causing /j/ to experience less amplitude reduction than /w/. Overall, however, the movement of F2 is outweighed by several other factors, and both types of glide segments demonstrate some degree of amplitude reduction in comparison with the adjacent vowel.

The main effect of vowel height context on ΔA_{RMS} was significant only for speaker M1 (F(2,30)=5.652; p=.008), indicating that the degree of amplitude reduction in glides is in general unaffected by the surrounding vowel context. This is also clear from Figure 9, which shows the average ΔA_{RMS} for each speaker, separated by vowel height context. Note that the differences between vowel height contexts are not significant for most of the speakers, and the differences between the means are not in the same direction across speakers. Tukey pairwise comparisons show that the only significant difference is between the mean of the high vowel context /i, u/ and the mid vowel context /e, o/ for speaker M1 (p=.009). This isolated pairwise difference is unlikely to be related to generalizable aspects

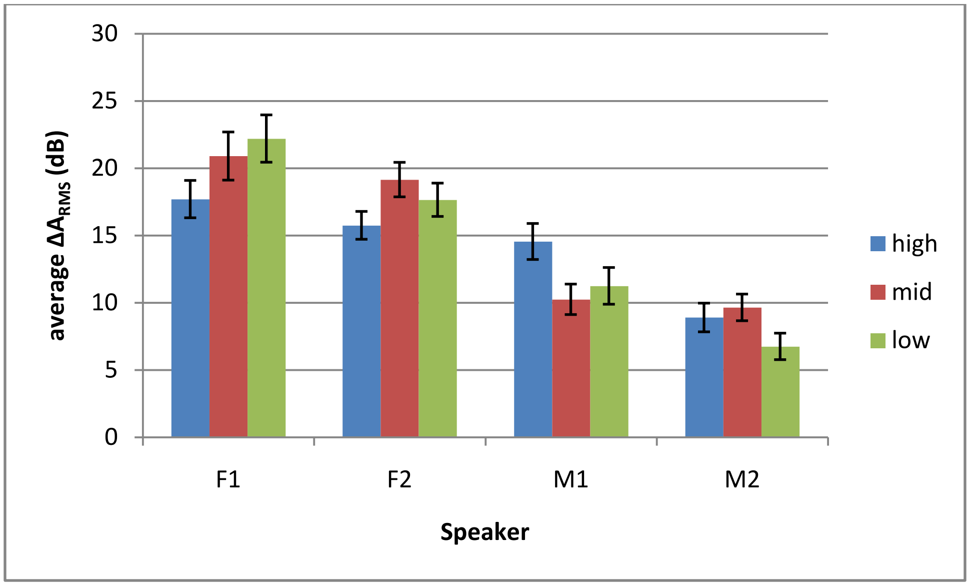


Figure 9: Measured ΔA_{RMS} for each speaker in the /VGV/ database, separated by vowel height context, and averaged across glide segments and prosodic context. Error bars represent standard error of the mean.

of the production of glides, especially since a comparison of the high and low vowel contexts, which are the most dissimilar in terms of articulation, does not show a significant difference (p=.05).

There was a significant main effect of prosodic context on ΔA_{RMS} for all four speakers (F(4,30)=19.423,4.318,10.329,12.047; p≤.007). Figure 10 shows the average ΔA_{RMS} for each speaker, separated by prosodic context, and Table 2 shows the results of Tukey pairwise comparisons of the factor level means. For most of the speakers, glides preceding a low-pitch-accented vowel (L2) are produced with significantly greater amplitude reduction than glides following either type of pitch-accented vowel (L1 or H1). In addition, glides preceding a high-pitch-accented vowel (H2) are produced with significantly greater amplitude reduction than one or both types of post-pitch-accent glides (L1 or H1) for speakers F1 and M2. The non-pitch-accented glides (NP) do not follow a consistent pattern across speakers; for speakers M1 and M2, they pattern with the post-pitch-accent glides, and are produced with significantly less amplitude reduction than one or both types of pre-pitch-accent glides

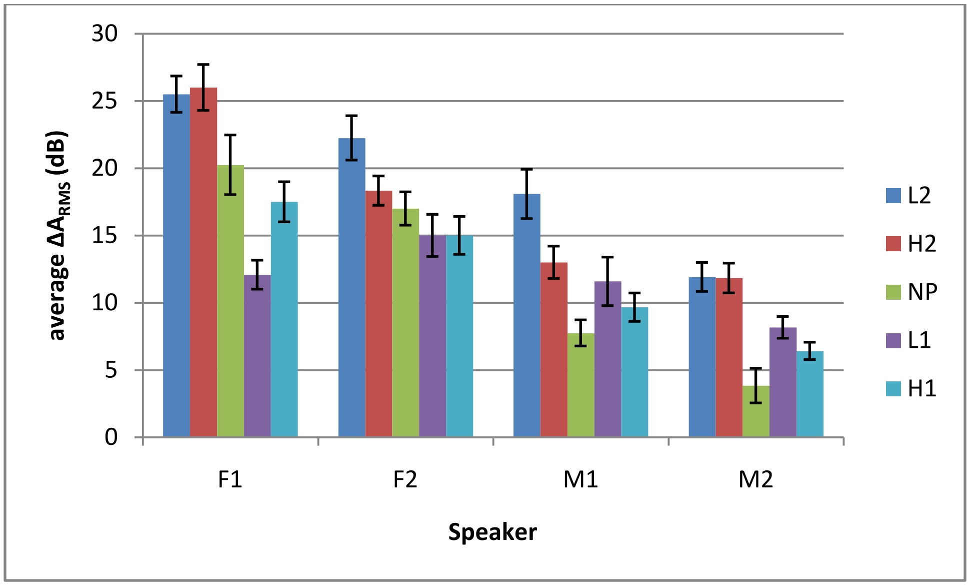


Figure 10: Measured ΔA_{RMS} for each speaker in the /VGV/ database, separated by prosodic context, and averaged across glide segments and vowel contexts. Error bars represent standard error of the mean.

(L2 or H2); for speaker F1, they pattern with the pre-pitch-accent glides, and are produced with significantly greater amplitude reduction than glides following low-pitch-accented vowels (L1). Overall, the general conclusion that can be drawn from these data is that glides tend to be produced with greater amplitude reduction when preceding pitch-accented vowels than when following pitch-accented vowels.

Table 2: Results of Tukey pairwise comparisons of ΔA_{RMS} means by prosodic context. Only significant differences (α = .01) are shown.

	Speaker F1	**Speaker F2**	**Speaker M1**	**Speaker M2**
L2 – H1	p = .001	p = .01	p = .000	p = .005
L2 – L1	p = .000	p = .01	p = .006	
L2 – NP			p = .000	p = .000
H2 – H1	p = .001			p = .006
H2 – L1	p = .000			
H2 – NP				p = .000
NP – L1	p = .001			

2.2.3 Discussion

The results of RMS amplitude measurements on /VGV/ tokens in natural speech show that glides are very consistently produced with a significant reduction in amplitude compared to the adjacent vowel. This amplitude reduction has been described as arising from a number of acoustic effects of the articulation of glides, involving a narrow constriction target in the oral region of the vocal tract. The effects of such a narrowing may include reduced first formant frequency (F1), increased first formant bandwidth (B1), decreased transglottal pressure and concomitant weakening and skewing of the glottal source waveform, all of which contribute to reduction of the overall amplitude of the speech signal during the glide segment (see Section 1.3).

Analyses of variance show little effect of the surrounding vowel context on the magnitude of the amplitude reduction in glide segments. If the ΔA_{RMS} measure is regarded as an indication of the relative strength of a particular glide segment, as reflected in the degree to which the oral tract is constricted during its production, then this result indicates that glide segments are produced with relatively equal strength in different vowel contexts. This makes amplitude characteristics a good possible candidate for a relatively invariant acoustic cue to the presence of glides in differing contexts.

By contrast, the effect of prosodic context was highly significant in the analysis of ΔA_{RMS}. Pairwise comparisons show that the amplitude reduction is generally greater in pre-pitch-accent glides than in post-pitch-accent glides. Since intervocalic glides are known to syllabify with the following vowel, especially when they occur word-medially (Gick, 2003), these data suggest that the glide constriction gesture is stronger in pitch-accented syllables than in non-pitch-accented syllables. This is in agreement with similar studies of other sound segments; for example, Pierrehumbert & Talkin (1992) found that the articulatory gesture for laryngeal consonants /h, ʔ/ is also strengthened when beginning a pitch-accented syllable. Interestingly, their study also observed the strengthening effect through the magnitude of amplitude reduction from the adjacent vowel into the consonant; this is indicative of an articulatory and acoustic similarity between glides and laryngeals which will be discussed in greater detail in later chapters.

2.3 First formant frequency (F1)

2.3.1 Method

Accurate measurement of the first formant frequency (F1) is quite difficult for speech signals with varying fundamental frequency (F0). Since the vocal tract filter is excited by a periodic glottal source, only the energy at harmonic multiples of F0 can be seen in the resulting acoustic spectrum. If the first formant does not exactly line up with one of the harmonics, its frequency and amplitude will not be accurately represented in a long-window DFT. The perceptual system's recognition of formant quality is known not to be affected by changing F0 and harmonic location, however. Thus, accurate measurement of F1 is desirable for our understanding of the perception of speech. (See Klatt (1986) for a discussion of the difficulties of F1 measurement and various solutions.)

Pitch-synchronous short-window analysis was chosen for this study, since it is capable of producing more accurate F1 results in the face of varying F0 than other analysis methods such as linear prediction. Limiting the analysis window to the length of a single pitch period allows the natural response of the vocal tract transfer function to be observed without the harmonic glottal excitation. However, great care must be taken in determining the correct placement of the window during the closed phase of each glottal cycle, as misplacement within the pitch period can result in very irregular spectra.

In order to determine the optimal placement of the analysis window for this study, a short experiment with synthetic vowel formant frequencies was undertaken. The position of a pitch-period-length Hamming window was systematically varied along a glottal cycle of the synthetic waveform, and the measured formant frequencies were recorded at each position. The position was measured with respect to the zero-crossing just before the maximum amplitude excursion of the waveform (the assumed moment of glottal closure), as a percentage of the length of the pitch period. The optimal match with the specified formant frequencies used in the synthesis was found when the center of the analysis window was placed 30%-40% of the pitch period length later than the identified zero-crossing. This window placement was used in all of the pitch-synchronous acoustic analyses undertaken in this study.

Measurements of the first formant frequency (F1) were taken pitch-synchronously from each of the /VGV/ tokens in the recordings database, using the xkl software. The software's peak-picking function was used on pre-emphasized DFT spectra, with placement of the pitch-period-length analysis window as described above. One measurement was taken from each pitch period over an interval starting at the midpoint of the first vowel, continuing through the glide segment, and ending at the midpoint of the second vowel, as in Section 2.2.1. The length of the analysis window was updated before each measurement, to be equal to the length of the current pitch period.

2.3.2 Results

Figure 11 shows an example plot of measured F1 contours, in which the individual measurement points are connected by straight lines to form continuous curves. The plot combines the F1 contours for all six vowel contexts, for a single glide /j/ produced in a single non-pitch-accented (NP) prosodic context by a single speaker F1. The curves are temporally aligned by setting the time of minimum amplitude during the glide segment (measured in Section 2.2) to be time zero, since this is the likely temporal location of the glide landmark in perception and lexical access (for an overview of landmark theory, see Stevens (2002) and Slifka *et al.* (2004).) Note that in general the time of the minimum F1 during the glide lines up with the time of the minimum A_{RMS}. This is to be expected, since both are acoustic effects

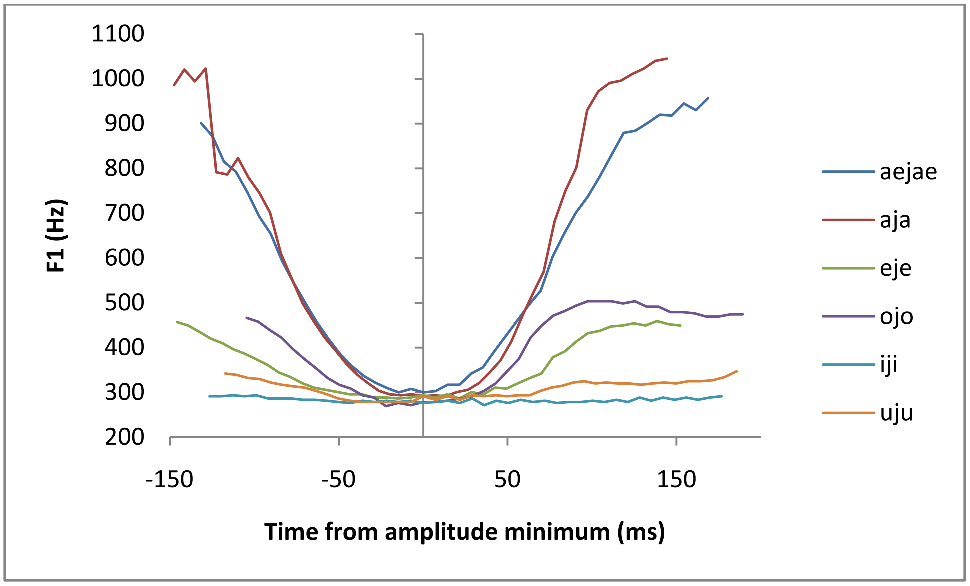


Figure 11: Time contours of the first formant frequency (F1) during the non-pitch-accented (NP) glide /j/, produced by speaker F1. The different curves correspond to the six vowel contexts /i, u, e, o, æ, ɑ/. Time is shown in relation to the point of minimum amplitude during the glide segment (time zero on the x-axis).

of the same vocal tract constriction gesture essential to the production of the glide, and in fact the F1 decrease contributes to the A_{RMS} decrease.

In Figure 11 it can be seen that the reduction in F1 is greatest between adjacent low vowels /æ, ɑ/ and the glide, and less large between the mid vowels /e, o/ and the glide. In order for the same minimum F1 target to be reached during the glide, the first formant must travel a greater distance from its higher frequency position in low vowels. In the high vowel contexts /i, u/, F1 hardly changes at all from the vowel to the glide; this suggests that the minimum F1 target for a glide is not in fact lower than the high vowel F1 for this speaker.

The minimum F1 target is not always the same for glides in different vowel contexts, however. There is often significant coarticulation between the F1 of an adjacent vowel and the minimum F1 of a

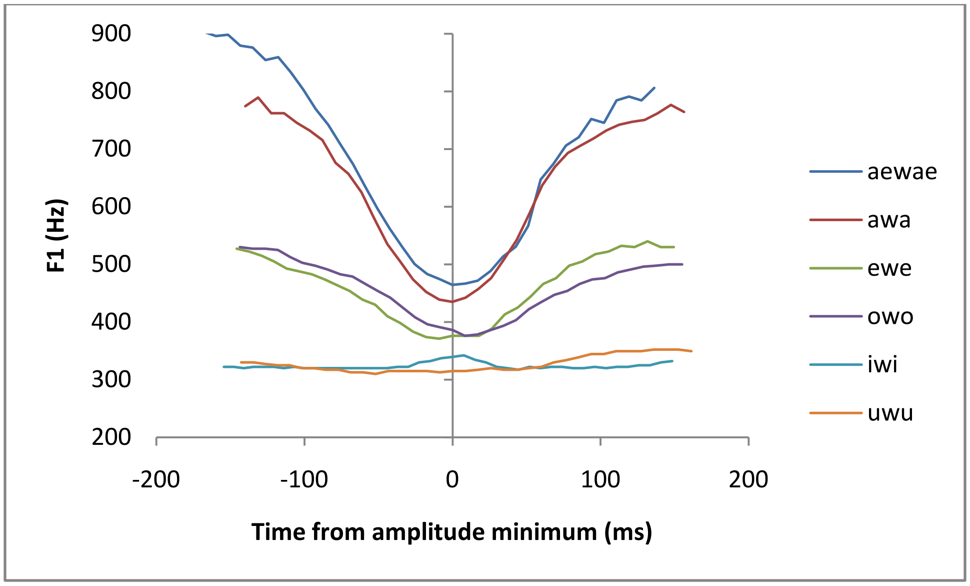


Figure 12: Time contours of the first formant frequency (F1) during the non-pitch-accented (NP) glide /w/, produced by speaker M1. The different curves correspond to the six vowel contexts /i, u, e, o, æ, ɑ/. Time is shown in relation to the point of minimum amplitude during the glide segment (time zero on the x-axis).

glide segment, such that the glide's F1 target appears to migrate in the direction of the vowel's F1. An example can be seen in Figure 12, which shows the F1 contours for the glide /w/ produced in the non-pitch-accented (NP) prosodic context by speaker M1. The effect of vowel height coarticulation is clear in this example, as the minimum F1 reached during the glides in low vowel contexts /æ, ɑ/ is higher than that of mid vowel contexts /e, o/, which in turn is higher than that of high vowel contexts /i, u/.

The coarticulatory effect on F1 was confirmed across the /VGV/ database through analyses of variance, which found a significant main effect of vowel height context for all speakers (F(2,30)=8.652,17.71,44.422,117.52; p≤.001). (Again, due to the presence of significant interactions between speaker and other factors, separate ANOVAs were run for each speaker. No interactions between factors other than speaker were significant.) Figure 13 shows the average F1 minimum during glides ($F1_{min}$) for each speaker, separated by vowel height context, and Table 3 gives the results of Tukey

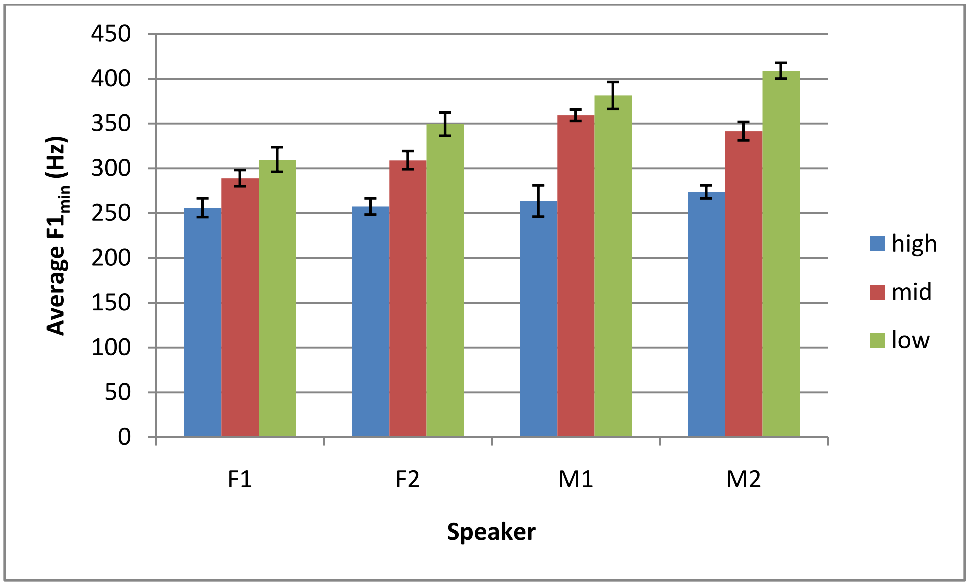


Figure 13: Measured F1 minimum ($F1_{min}$) during glides for each speaker in the /VGV/ database, separated by vowel height context, and averaged across glide segments and prosodic context. Error bars represent standard error of the mean.

pairwise comparisons of the factor level means. $F1_{min}$ is significantly higher in low vowel contexts than in high vowel contexts for all speakers, and is significantly higher in mid vowel contexts than in high vowel contexts for three of the four speakers. The difference between low vowel contexts and mid vowel contexts is significant only for speaker M2. From Figure 13, it is clear that the coarticulatory spread is largest for speaker M2 and smallest for speaker F1; however, all four speakers show spreading of the $F1_{min}$ target according to vowel height context to some degree.

Table 3: Results of Tukey pairwise comparisons of $F1_{min}$ means by vowel height context. Only significant differences (α = .01) are shown.

	Speaker F1	**Speaker F2**	**Speaker M1**	**Speaker M2**
low – high	p = .001	p = .000	p = .000	p = .000
mid – high		p = .006	p = .000	p = .000
low – mid				p = .000

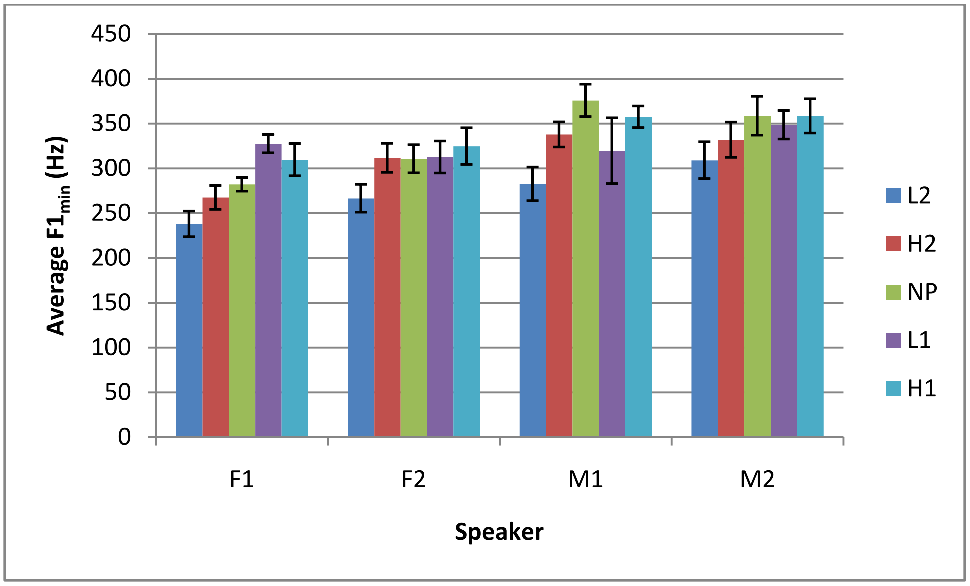


Figure 14: Measured F1$_{min}$ for each speaker in the /VGV/ database, separated by prosodic context, and averaged across glide segments and vowel contexts. Error bars represent standard error of the mean.

The analysis of variance also investigated the effects of the type of glide segment and prosodic context factors. The main effect of glide segment was significant only for speaker M2 (F(1,30)=21.505, p=.000), with /j/ having a lower average F1$_{min}$ than /w/. This appears to correspond with this speaker's production of the "cognate" high vowels; his F1 for /i/ was consistently lower than his F1 for /u/. It seems that this speaker produces narrower palatal constrictions than labiovelar constrictions, whether in vowels or glides. Other studies support this difference between /i/ and /u/ across speakers (e.g. Stevens, 1998, p. 288; Maddieson & Emmorey, 1985, p. 167); however, the difference is too slight to carry over into glides for any of the other speakers in this study.

The main effect of prosodic context was significant for speakers F1, M1, and M2. Figure 14 shows the average F1$_{min}$ for each speaker, separated by prosodic context, and Table 4 gives the results of Tukey pairwise comparisons of the factor level means. For three of the speakers, glides preceding a low-pitch-accented vowel (L2) are produced with significantly lower minimum F1 than glides following one or both

Table 4: Results of Tukey pairwise comparisons of $F1_{min}$ means by prosodic context. Only significant differences ($\alpha = .01$) are shown.

	Speaker F1	Speaker F2	Speaker M1	Speaker M2
H1 – L2	p = .002		p = .000	p = .001
L1 – L2	p = .000		p = .002	
NP – L2			p = .000	p = .001
H2 – L2			p = .003	
L1 – H2	p = .01			

types of pitch-accented vowel (L1 or H1). In addition, $F1_{min}$ for pre-low-pitch-accent glides (L2) is significantly lower than for non-pitch-accented glides (NP) for speakers M1 and M2, and it is significantly lower than for pre-high-pitch-accent glides (H2) for speaker M1. $F1_{min}$ for pre-high-pitch-accent glides (H2) is significantly lower than for post-low-pitch-accented glides (L1) only for speaker F1. Overall, the general conclusion that can be drawn from these data is that glides are sometimes produced with lower target F1 when preceding low-pitch-accented vowels than in other prosodic contexts.

2.3.3 Discussion

The effect of prosodic context on the minimum first formant frequency ($F1_{min}$) reached during a glide segment is similar to the effect observed in Section 2.2 on the magnitude of amplitude reduction. Glides beginning pitch-accented syllables, especially low-pitch-accented syllables, seem to have a tendency to be produced with a strengthened articulatory gesture. The strengthened oral constriction gesture has the effect of lowering both F1 and A_{RMS} during the glide segment, as compared to the adjacent vowel.

F1 and A_{RMS} measurements differ, however, in terms of the effect of the surrounding vowel height context. No consistent effect of vowel height context was found for the ΔA_{RMS} measure; however, the effect of vowel height context on the minimum F1 reached during the glide segment ($F1_{min}$) was highly significant. Coarticulation with the adjacent vowel caused $F1_{min}$ to be significantly higher in glides next to low vowels than in glides next to high vowels. The extent of this coarticulation was such that $F1_{min}$ in glides next to low vowels was often higher than the F1 of many high vowels. This indicates that glides are not always produced with a vocal tract constriction that is narrower than that of a high vowel. When adjacent to a low vowel with a very open oral cavity and high F1, a glide may be produced with a constriction that is only about as narrow as that of a mid vowel.

Although the constriction degree and resulting F1 may vary widely between glides in different vowel height contexts, the fact that ΔA_{RMS} does not vary significantly across these different contexts shows that at least one acoustic characteristic of the glide-vowel distinction remains invariant. In low vowel contexts, the movement between the high F1 of the vowel and the coarticulated $F1_{min}$ of the glide is already large enough to reduce the overall spectral amplitude by the same amount that it is reduced in other contexts, without constricting the oral cavity to a greater degree than in any vowel. On the other hand, in high vowel contexts, a similar degree of amplitude reduction is obtained, even though it was seen in Section 2.3.2 that F1 does not change appreciably from the high vowel to the glide. Since the amplitude is reduced in the absence of F1 change, it must be hypothesized that the oral constriction is made narrower than the point at which the wall effects begin to inhibit the further lowering of F1, as schematized in Figure 2 on page 24. In this scenario, the amplitude reduction arises from sources other than lowered F1, such as increased first formant bandwidth or decreased transglottal pressure. The latter possible sources will be tested in other acoustic measures to follow in this chapter. It is important to note, though, that although the specific source of the effect may differ between vowel contexts, the amplitude reduction characteristic of glides does not differ significantly between them. Thus, ΔA_{RMS} may provide a measure of acoustic invariance that can be exploited in the perception/recognition of glides in spite of variation in $F1_{min}$ across contexts.

2.4 Open quotient (OQ)

2.4.1 Method

Open quotient (OQ) measurements are of interest in order to determine whether the narrow oral constriction in a glide produces aerodynamic effects on the glottal source. As described in Section 1.3, if the glide constriction is sufficiently narrow, some pressure will be built up in the oral cavity behind it, and the transglottal pressure drop will decrease if the subglottal pressure remains constant. The increased oral pressure and decreased transglottal pressure have the effect of skewing the glottal cycle toward a larger percentage of time with the vocal folds open (increased OQ). As this study did not have access to direct physical measures of OQ, the correlated acoustic measure H1* - H2* was investigated in the /VGV/ recordings database.

H1 and H2 are the amplitudes of the first and second harmonics, respectively, that can be measured from the spectrum of an acoustic speech signal using an analysis window covering multiple pitch periods. When the spectrum is taken from the glottal volume velocity source without vocal tract filtering, the difference between the two harmonics (H1 – H2) is proportional to the value of OQ, as shown in panels (a) and (c) of Figure 15 from Hanson (1995). Greater positive differences H1 – H2 correspond to larger percentage values of OQ. However, the acoustic spectrum accessible from the radiated speech has the glottal source filtered by the vocal tract transfer function, resulting in changes to the harmonic amplitudes H1 and H2, as shown in panels (b) and (d) of Figure 15. Inverse filtering calculations must be performed in order to accurately uncover the OQ relationships between speech segments with different formant frequencies. For inverse filtering of H1 – H2, it is sufficient to subtract the effect of the first formant peak of the vocal tract transfer function, in order to arrive at the corrected measure H1* - H2* (Hanson, 1995).

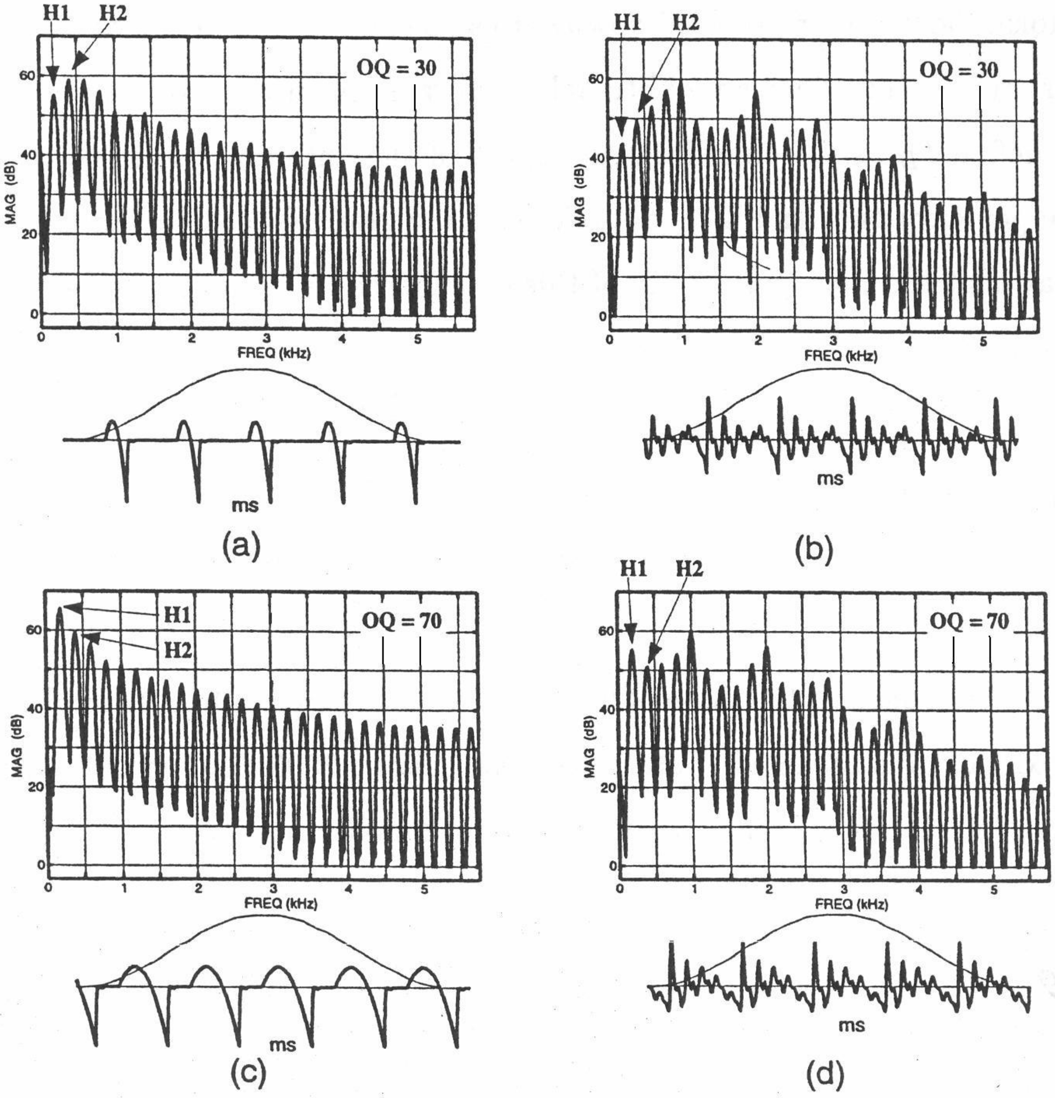


Figure 15: Waveforms and spectra of a synthetic glottal volume-velocity source corresponding to different manipulations of open quotient (OQ). The fundamental frequency is in the range for an adult female speaker. Panels (a) and (c) show spectra and derivatives of the volume-velocity sources, while panels (b) and (d) show the spectra of the vowel /æ/ synthesized using those volume-velocity sources. (a)-(b) OQ is 30%; (c)-(d) OQ is 70%. From Hanson (1995).

For glides and vowels, the vocal tract can be modeled as an all-pole transfer function of the following form:

$$T(\omega) = \left(\frac{s_1 {s_1}^*}{(s - s_1)(s - {s_1}^*)}\right)\left(\frac{s_2 {s_2}^*}{(s - s_2)(s - {s_2}^*)}\right) \cdots \left(\frac{s_n {s_n}^*}{(s - s_n)(s - {s_n}^*)}\right)$$

where s = jω, s_n = (α_n + jω_n), s_n^* = (α_n – jω_n), and n is the number of vocal tract resonant frequencies (formants) under consideration. The transfer function for the isolated first formant is thus given by:

$$T_1(\omega) = \frac{(\alpha_1 + j\omega_1)(\alpha_1 - j\omega_1)}{\big(j\omega - (\alpha_1 + j\omega_1)\big)\big(jw - (\alpha_1 - j\omega_1)\big)}$$

where ω = 2πf, and ω_1 = 2π(F1). Substituting α_1 = π(B1) and simplifying gives:

$$T_1(f) = \frac{(F1)^2 + (\frac{B1}{2})^2}{(F1)^2 - (f + j\frac{B1}{2})^2}$$

The corrections to the amplitudes of the first two harmonics can then be made by converting the magnitude of the "boost" of the first formant transfer function to dB and subtracting it from the measured harmonic amplitude, thus:

$$H1^* = H1 - 20\log_{10}|T_1(F0)|$$

$$H2^* = H2 - 20\log_{10}|T_1(2F0)|$$

As can be seen from the above equations, the calculations to derive H1* - H2* from the measured H1 and H2 require estimates of the first formant frequency (F1) and the first formant bandwidth (B1). For this study, F1 measurements for all of the tokens in the /VGV/ database have already been presented in Section 2.3, leaving only B1 to be estimated. The most direct method of measuring B1 is to calculate it from the time domain representation of the waveform, as in Fant (1997) and Hanson (1995). Figure 16 illustrates the method of determining the bandwidth from the exponential decay factor of the formant oscillation envelope, once the signal has been band-pass filtered around the formant frequency. Unfortunately, the glide recordings collected in this study are not suitable for this time-domain analysis, because the low F1 of the glide (and high vowel) tokens is too close to the fundamental

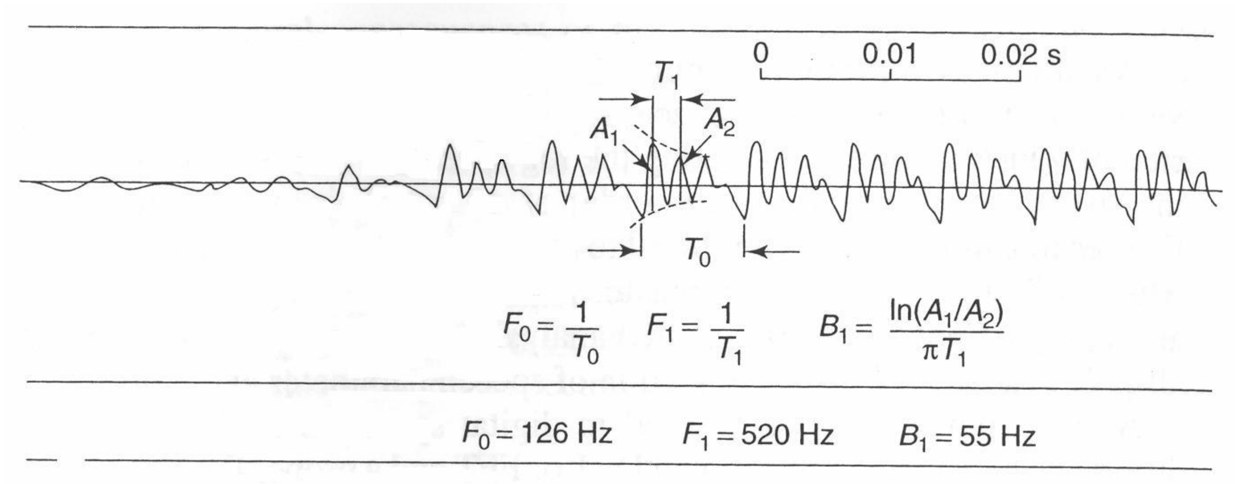


Figure 16: Illustration of the time-domain extraction of the first formant bandwidth (B_1), after low-pass filtering to remove the energy above the first formant frequency (F_1). If the F_1 oscillation is assumed to be a damped sinusoid of the form $e^{-\pi B_1 t} \cos(2\pi F_1 t)$, then B_1 can be estimated by measuring the decay rate of the first formant waveform, according to the formulae above. From Fant (1997).

frequency (F0). The measurement strategy illustrated in Figure 16 requires at least two full oscillations at the first formant frequency to be visible within each pitch period; this requirement is not met when F1 is close in frequency to F0, as can be seen in the example in Figure 17. In this token of /iji/, produced with high pitch accent on the first vowel (H1), F1 and F0 are so close that not even one full oscillation at F1 can be observed within a pitch period, making time-domain extraction of B1 impossible. Hanson (1995) avoids this problem by restricting her analysis to productions of the low vowel /æ/, whose F1 is much higher than typical F0 for males or females. However, this strategy is not useful in a study of glides, which characteristically have low F1. In addition, Hanson's corrections for H1* - H2* neglect the contribution of B1; however, this strategy has been shown to result in significant error unless F1 is more than its bandwidth (B1) away from the harmonic frequencies (Iseli & Alwan, 2004).

Since direct measurement of B1 could not easily or accurately be carried out in this study, estimates were made based on the data reported by Fujimura & Lindqvist (1971). Other studies have also

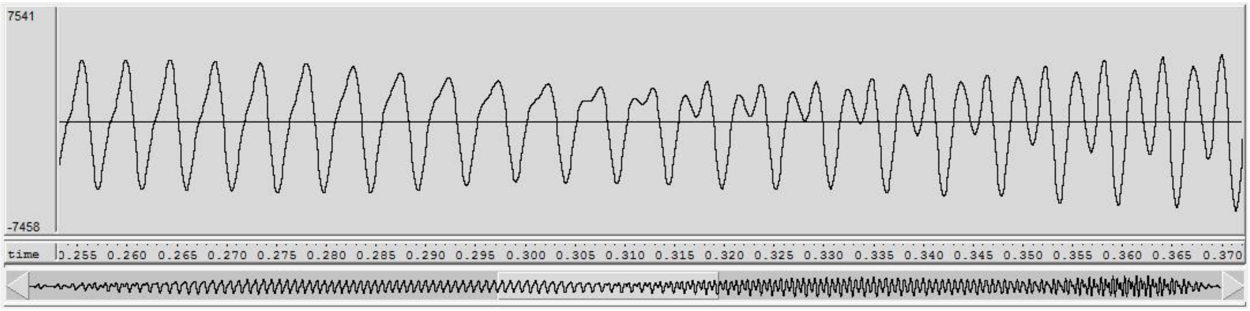


Figure 17: Time-domain waveform section of the glide segment in /iji/, produced with high pitch accent on the first vowel (H1) by speaker F1, with the first formant isolated through low-pass filtering.

provided data on formant bandwidths with relation to frequency (e.g. Fant, 1972; Okobi, 2006), but the Fujimura & Lindqvist study has the advantage of providing data from female speakers as well as male. Fujimura & Lindqvist were able to bypass the issue of proximity between F1 and F0 by exciting the vocal tract using a sweep-tone external signal, eliminating the harmonic nature of the acoustic spectrum. They then used an analysis-by-synthesis procedure to determine the synthetic formant frequencies and bandwidths that matched the observed spectra.

Fujimura & Lindqvist's (1971) plot of B1 vs. F1 for male and female speakers was given in Figure 3 of Chapter 1; it is reprinted here as Figure 18. Fujimura & Lindqvist's representative curves were drawn by visual inspection of the plotted data points; for the estimation performed in this study, the curves were fitted with 2nd-order polynomials, as follows:

$$B1_M = (3.2941 \times 10^{-4})(F1)^2 - \ 0.3609(F1) + \ 133.736$$

$$B1_F = (3.4783 \times 10^{-4})(F1)^2 - \ 0.4341(F1) + \ 178.1354$$

where B1$_M$ is the first formant bandwidth for a male speaker, and B1$_F$ is the first formant bandwidth for a female speaker. Although Fujimura & Lindqvist's data include productions of all of the vowels used in this study (as well as many more), it cannot be claimed with certainty that extrapolation of the curves fitted to their data will accurately represent the F1-B1 relationship in glides, since those sound segments

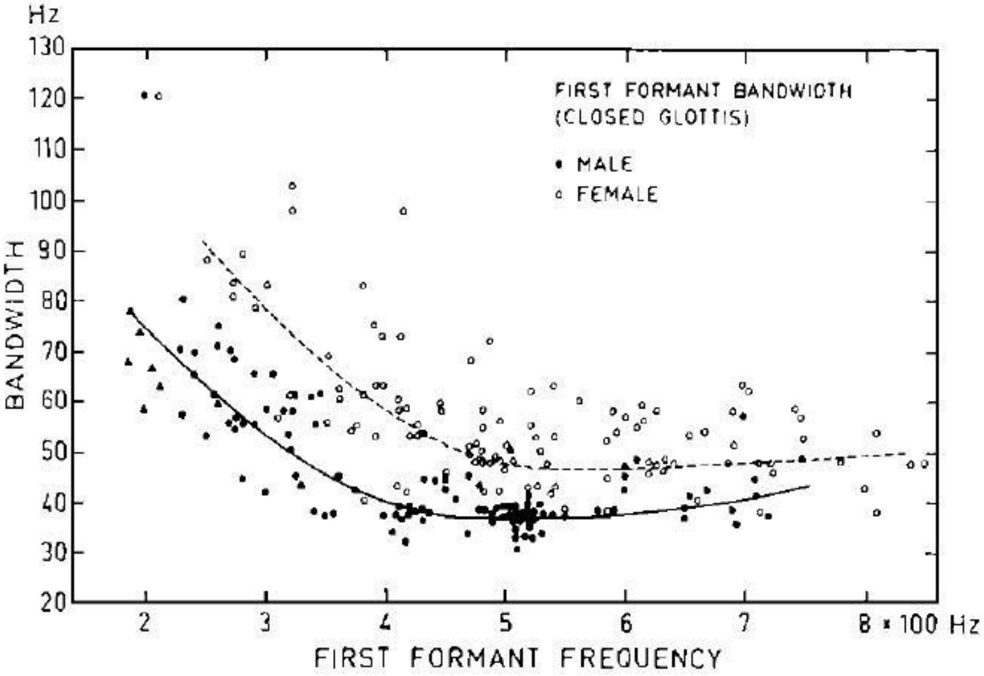


Figure 18: "Bandwidth values for the first formant plotted against the formant frequency. Each closed circle represents a vowel sample of one of three male subjects, and an open circle represents a sample of one of three female subjects. Representative values are estimated by visual inspection of the plots, and curves are drawn for male and female subjects separately. Bandwidth values for articulations with bilabial closures by a male subject are also added in this graph (closed triangles)." (Fujimura & Lindqvist, 1971)

were not included in their study. However, since Fujimura & Lindqvist did include some productions of the stop consonant /b/, and since those data points do agree with the low-F1 end of the curve trajectory for the vowels, it is reasonable to assume that glide data points would tend to interpolate between those of the stops and the vowels along the same curves. The left endpoints of Fujimura & Lindqvist's curves are also consistent with the closed-mouth B1 data given by Fant *et al.* (1977), which averaged 76 Hz for males and 94 Hz for females.

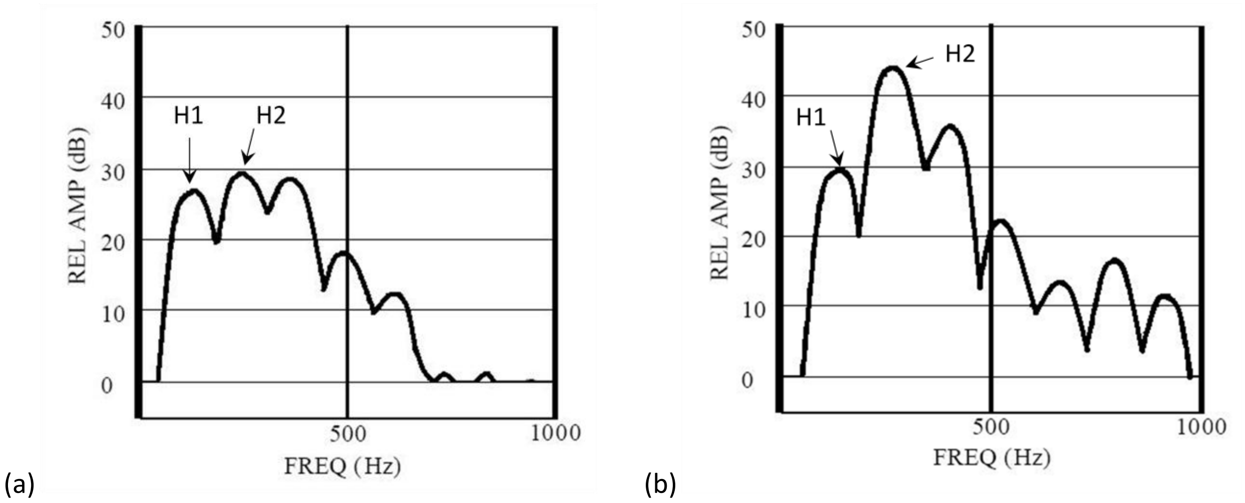


Figure 19: Low-frequency spectra illustrating the amplitude of the first two harmonics during the token /iji/, produced with high pitch accent on the second vowel (H2) by speaker M1. Inverse filtering was applied to the measured values H1 and H2 to remove the effect of the first formant and arrive at the source estimates H1* and H2* (see text). At the glide landmark in (a), H1*-H2* is larger than at the vowel landmark in (b), indicating that the open quotient is increased during the glide.

Two measurements of H1 and H2 were made from each /VGV/ token in the recordings database; one measurement was made at the glide landmark, and one measurement was made at the adjacent vowel landmark. The glide landmark was defined as the time at which the amplitude (A_{RMS}) reached a minimum during the glide segment, and the vowel landmark was defined as the time at which A_{RMS} reached a maximum during the vowel segment, following Stevens (2002). If the global maximum A_{RMS} did not occur at a time at which the formant frequencies had reached steady state values for the vowel (for instance, an amplitude peak might occur when resonant frequencies crossed during the transition between the glide and the vowel), the time at which F1 reached a maximum was used for the vowel landmark instead. H1 and H2 were measured from a DFT spectrum with 22.3 ms analysis window centered on the glide or vowel landmark. This window length assured that at least two pitch periods were included under the window for all speakers, allowing the individual harmonics to be observed and measured, as in Figure 19. F1 measurements corresponding to the time of the glide or vowel landmark

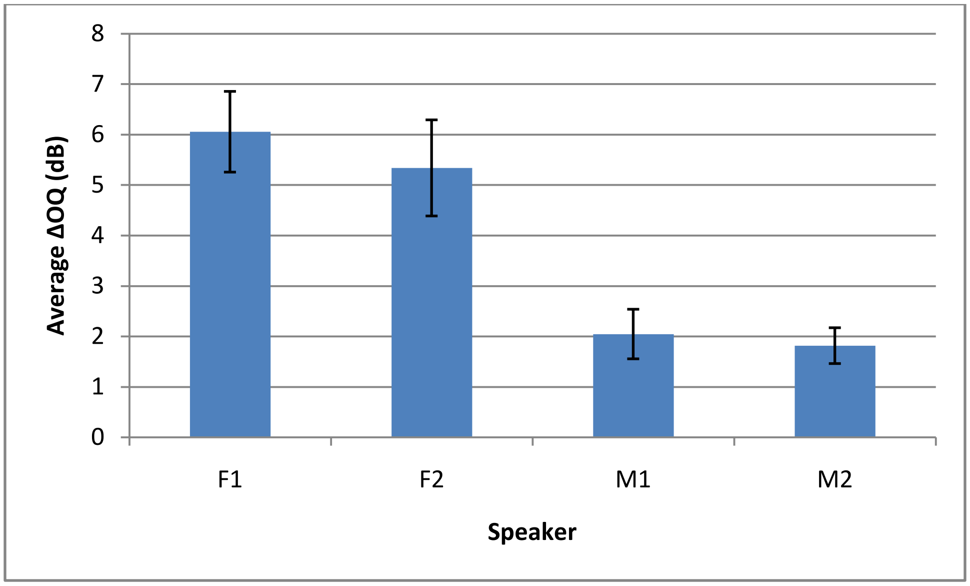


Figure 20: Measured ΔOQ for each speaker in the /VGV/ database, averaged across glide segments, vowel and prosodic contexts. Error bars represent standard error of the mean.

were taken from the data presented in Section 2.3, and these measurements were used to estimate B1 according to the fitted curves listed above. The inverse filtering calculations described in this section were then performed to arrive at H1* and H2*, and the H1* - H2* of the vowel was subtracted from the H1* - H2* of the glide for each token. The resulting measure, $(H1^*\text{-}H2^*)_{glide} – (H1^*\text{-}H2^*)_{vowel}$ will hereafter be designated as ΔOQ. (It should be remembered, however, that this is not a direct measurement of percent open quotient of the glottal cycle, but a correlated acoustic measure of harmonic amplitude relations, expressed in dB.) A positive number for ΔOQ indicates that the glide has a larger open quotient than its adjacent vowel; a negative number indicates that the glide has a smaller open quotient than the adjacent vowel.

2.4.2 Results

Figure 20 shows the average results of ΔOQ measurements for each speaker in the /VGV/ database. Paired t-tests with Bonferroni adjustments for multiple comparisons confirm that the average ΔOQ is

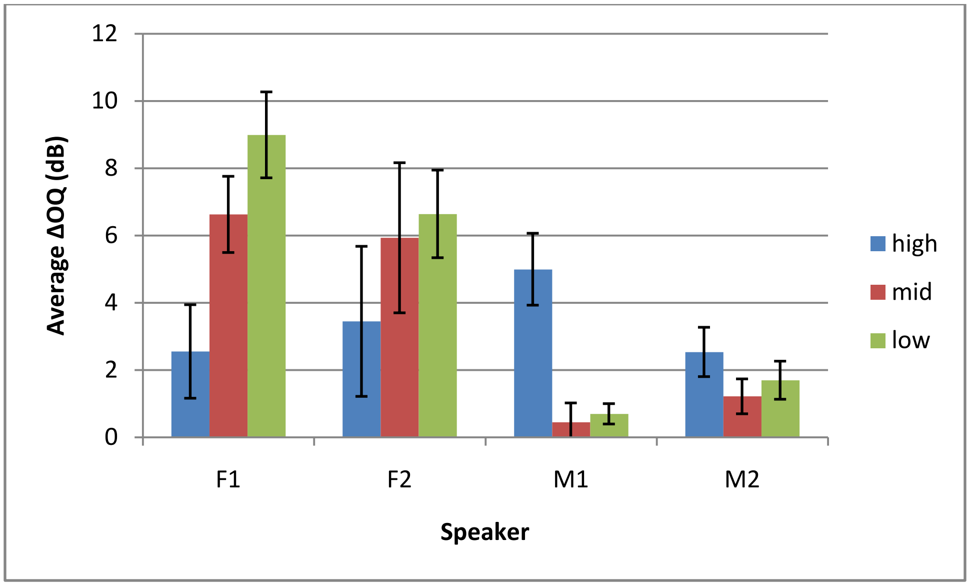


Figure 21: Measured ΔOQ for each speaker in the /VGV/ database, separated by vowel height context, and averaged across glide segments and prosodic context. Error bars represent standard error of the mean.

significantly greater than zero for all four speakers (t(59=7.566, p=.000 for speaker F1; t(59)=5.605, p=.000 for speaker F2; t(59)=4.159, p=.000 for speaker M1; t(59)=5.111, p=.000 for speaker M2). Thus, a common characteristic of glides across speakers is that their open quotient is increased relative to that of the adjacent vowel.

Separate analyses of variance were conducted for each speaker to assess the effect of the glide segment, vowel height context, and prosodic context factors controlled in the database. No interactions between these factors were significant. No main effects of glide segment /j, w/ were significant, and no main effects of prosodic context were significant.

The main effect of vowel height context (high, mid, low) was significant for speakers F1 and M1 (F(2,30)=7.493,16.395; p≤.002). Figure 21 shows the average ΔOQ for each speaker, separated by vowel height context, and Table 5 shows the results of Tukey pairwise comparisons of the factor level means.

Table 5: Results of Tukey pairwise comparisons of ΔOQ means by vowel height context. Only significant differences (α = .01) are shown.

	Speaker F1	**Speaker F1**	**Speaker M1**	**Speaker M2**
high – low	p = .002 (diff. = -6.44)		p = .000 (diff. = 4.3)	
high – mid			p = .000 (diff. = 4.55)	

The difference between the mean ΔOQ for high vowel contexts /i, u/ and that for low vowel contexts /æ, ɑ/ is significant for both speaker F1 and speaker M1; however, the difference is in different directions for the two speakers. For speaker F1, ΔOQ is greater in low vowel contexts than in high vowel contexts; while for speaker M1, ΔOQ is greater in high vowel contexts than in low vowel contexts. This may be explained by the difference in coarticulatory effects observed for the two speakers in Section 2.3. Recall that the coarticulatory spread of $F1_{min}$ with vowel height context was smallest for speaker F1 and larger for speaker M1. For speaker F1, glide targets in all vowel height environments were produced with sufficient narrowing of the oral constriction to make the first formant frequency (F1) quite low. If this narrowing were of such degree as to affect the transglottal pressure and skew the glottal waveform, one would expect an appreciable increase in OQ during glides in all vowel height contexts. The magnitude of this increase, ΔOQ, would be greater when coming from a very open, unconstricted vowel such as /æ, ɑ/ than when coming from a narrower vowel such as /i, u/. A similar scenario may be true for speaker F2, although there was more variation in her productions, and her differences in ΔOQ means did not reach statistical significance. By contrast, speaker M1 produced quite low $F1_{min}$ in glides in high vowel contexts, but his $F1_{min}$ was significantly higher in mid and low vowel contexts due to coarticulation. Thus, the oral constriction narrowing may not have been of a sufficient degree to affect his glottal waveform as much in mid and low vowel contexts, and his ΔOQ would be smaller (close to zero) in mid and low vowel contexts than in high vowel contexts. The situation may be similar for speaker M2, although the larger variation in his productions explains a lack of statistical significance.

2.4.3 Discussion

Of the acoustic characteristics of glides investigated in this study, ΔOQ is the most directly related to the aerodynamic effects of the narrow oral tract constriction on the waveform of the glottal source. While the amplitude of the acoustic output receives contributions from the first formant frequency and bandwidth as well as the glottal source amplitude, H1* - H2* after inverse filtering provides specific information about the shape of the source waveform, with the vocal tract filter effects theoretically removed. Significant increases in H1* - H2* in glides compared to the H1* - H2* in the adjacent vowels indicate that some oral pressure does in fact build up behind the glide constriction, causing the transglottal pressure drop to decrease and the glottal waveform shape to be skewed rightward. This results in amplitude reduction as well as open quotient increase during the glide segment.

Statistical analyses of the acoustic data in this study reveal that average OQ is significantly larger in glides than in adjacent vowels, across speakers. The magnitude of the increase, ΔOQ, is not significantly different for different glide segments /j, w/, and is not significantly affected by prosodic context. Some inter-speaker variation was observed with respect to the effect of vowel height context, but the manifestation of these variations is consistent with the aerodynamic interaction hypothesis in the face of coarticulatory effects, as described in the previous section.

Aerodynamic effects of the vocal tract shape on the glottal excitation source have not been studied before as they pertain to the production and acoustics of glides. It is hoped that future research will develop an increased focus on nonlinear source-coupling phenomena in these and other similar sounds, as has only recently begun to be suggested for consonants in general (Zañartu, Mongeau, & Wodicka, 2007). Because this thesis focuses on acoustic analysis of glides, the changes in the glottal waveform shape can only be inferred from amplitude changes in the harmonic spectrum. Future research using direct methods such as electroglottograph measurements should be able to better quantify the increase in open quotient percentage that occurs during glides. At present, the acoustic investigation of glottal source effects during glides is already a significant departure from and expansion on previous studies focusing exclusively on formant frequencies and duration characteristics (see Section 1.2).

2.5 Harmonics-to-noise ratio (HNR)

2.5.1 Method

The harmonics-to-noise ratio (HNR) is defined as the power of the periodic voicing component of the speech signal divided by the power of the aperiodic noise component. In effect, it is a comparison of the strength of the glottal excitation source to that of any turbulence noise sources that may be active at some point along the vocal tract filter. If the speech signal contains frication noise, the noise source is located at a constriction in the oral or pharyngeal tract; if there is aspiration noise, the source is located near the glottis, just in front of the periodic source (Stevens, 1998). If an increased salience of either of these types of turbulence noise is a characteristic that distinguishes glides from vowels, as has been suggested (Padgett, 2008), the increased noise power would be acoustically observable through a decrease in the HNR. Measurements were made on this study's /VGV/ recordings database to investigate this possibility.

HNR measurements were made using a pitch-scaled harmonic filter (PSHF) algorithm modified from Jackson & Shadle (2001), as implemented in MATLAB by Mehta (2006). The PSHF uses a spectral comb-filtering technique to separate the harmonic peaks from the speech spectrum, leaving behind the noise floor that appears between the harmonics. An interpolation method is then used to fill in the "holes" in the noise component at the harmonic frequency bins, and to adjust the amplitude of the separated harmonic component accordingly. Implementation of the comb filter requires estimates of the fundamental frequency (F0) for successive windowed segments of the speech signal; these estimates are provided by the Praat speech processing tool. Mehta's code creates new .wav files of the separated harmonic and noise components, calculates their power, and outputs the original speech file's HNR automatically.

For the measurement of HNR in glides and vowels, 100 ms segments were excised from the /VGV/ recordings to be processed by the Mehta code. For each token, one segment centered on the amplitude-minimum landmark was excised from the glide, and one segment centered on the amplitude-maximum landmark was excised from the vowel. (The glide and vowel landmarks were located as described in Section 2.4.1.) For each segment, the HNR was calculated and recorded, and the separated harmonic and noise components were retained for further analysis. ΔHNR was calculated by subtracting the HNR of the vowel segment from the HNR of the glide segment for each token. A negative number for ΔHNR would indicate that a glide has lower HNR than its adjacent vowel. This could be caused by a decrease in the power of the harmonic component or an increase in the power of the noise component; either phenomenon would result in a decrease in HNR.

2.5.2 Results

ΔHNR results for each speaker, averaged across glide segment, vowel height contexts, and prosodic contexts, are shown in Figure 22. Paired t-tests with Bonferroni adjustments for multiple comparisons confirm that the mean ΔHNR is significantly below zero for all four speakers (t(59)=-6.359, p=.000 for speaker F1; t(59)=-6.602, p=.000 for speaker F2; t(59)=-3.037, p=.002 for speaker M1; t(59)=-3.678, p=.000 for speaker M2). This indicates that glides are produced with significantly lower HNR than their adjacent vowels, across speakers.

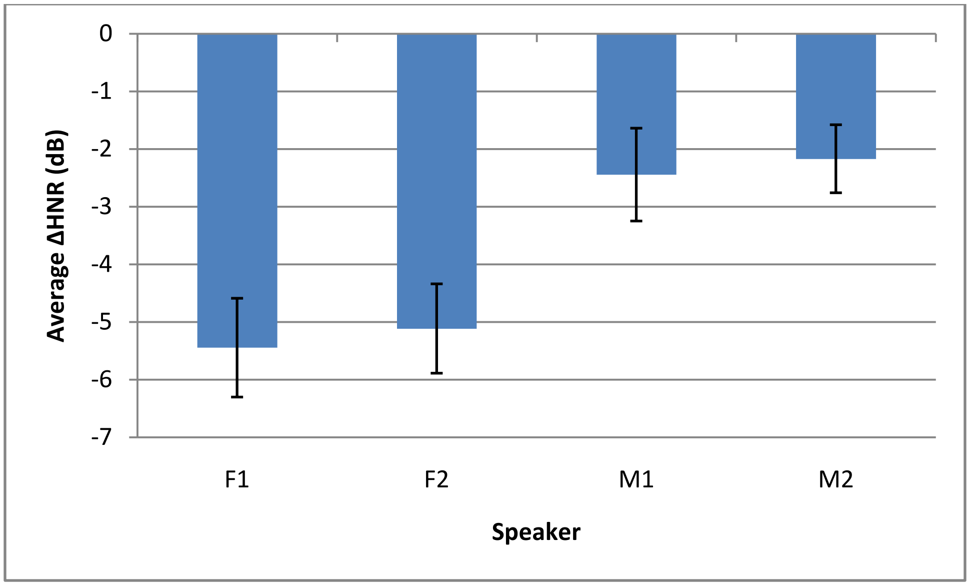


Figure 22: Measured ΔHNR for each speaker in the /VGV/ database, averaged across glide segments, vowel and prosodic contexts. Error bars represent standard error of the mean.

Analyses of variance showed no significant differences in ΔHNR with respect to vowel height context or prosodic context. The main effect of glide segment /j, w/ was significant for speakers F1 and M1 (F(1,30=7.745,23.261; p≤.009). Figure 23 shows the average ΔHNR for each speaker, separated according to glide segment /j, w/; the direction of the difference is the same for all speakers. The average ΔHNR is more negative for productions of the glide /w/ than for productions of the glide /j/, indicating that the HNR decreases more for /w/. This is to be expected given that ΔA_{RMS} was found in Section 2.2.2 to be larger for /w/ than for /j/, since a larger harmonic amplitude reduction will produce a larger HNR decrease even if the noise power is the same.

Although it has been shown that HNR is significantly reduced in glides compared to their adjacent vowels, this does not necessarily confirm that the contribution of noise is increased in glides, since HNR can also be reduced simply through the reduction of the harmonic power in the absence of any change in noise power. That the harmonic power is reduced can be inferred from the consistent reduction in

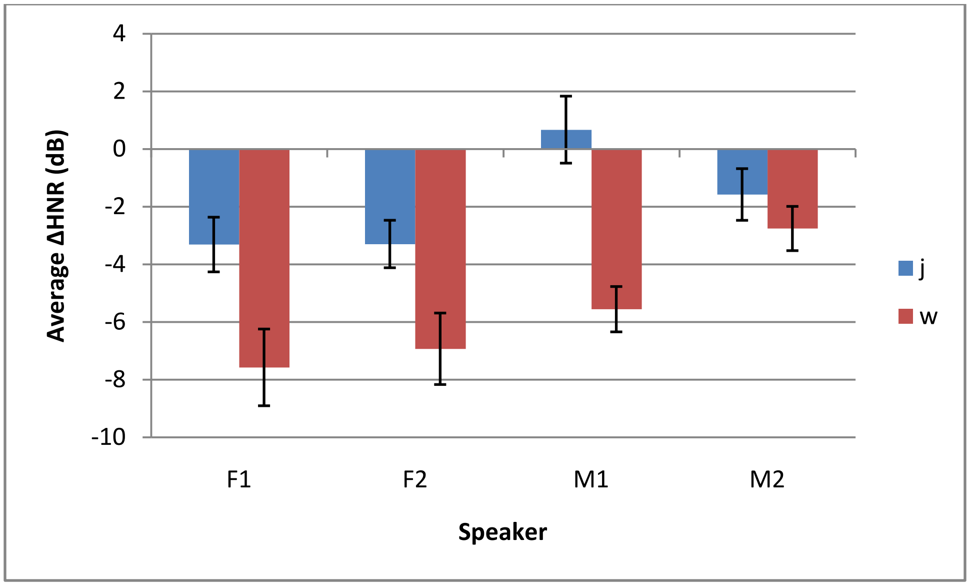


Figure 23: Measured ΔHNR for each speaker in the /VGV/ database, separated by glide segment /j, w/, and averaged across vowel and prosodic contexts. Error bars represent standard error of the mean.

A_{RMS} that was reported in Section 2.2; determining whether the noise power is also increased requires further investigation. For this purpose, it is convenient that Jackson & Shadle's (2001) pitch-scaled harmonic filter allows for the reconstruction of separate speech signals representing the isolated harmonic and noise components of the original signal. The power of the isolated noise component of the glide can be calculated and compared to the power of the isolated noise component in the adjacent vowel. The results of such analysis for this study's /VGV/ database, however, do not find that the noise power is at all increased in the glide segment. It appears that the glide's HNR reduction can be attributed mostly to the reduction in harmonic amplitude already noted, rather than to any increase in noise power.

However, the fact that the calculated noise power in the glide is not greater than the calculated noise power in the adjacent vowel does not necessarily contradict the addition of a new noise source in the glide, as can be seen from Figure 24, from Jackson & Shadle (2000). This figure plots the separated

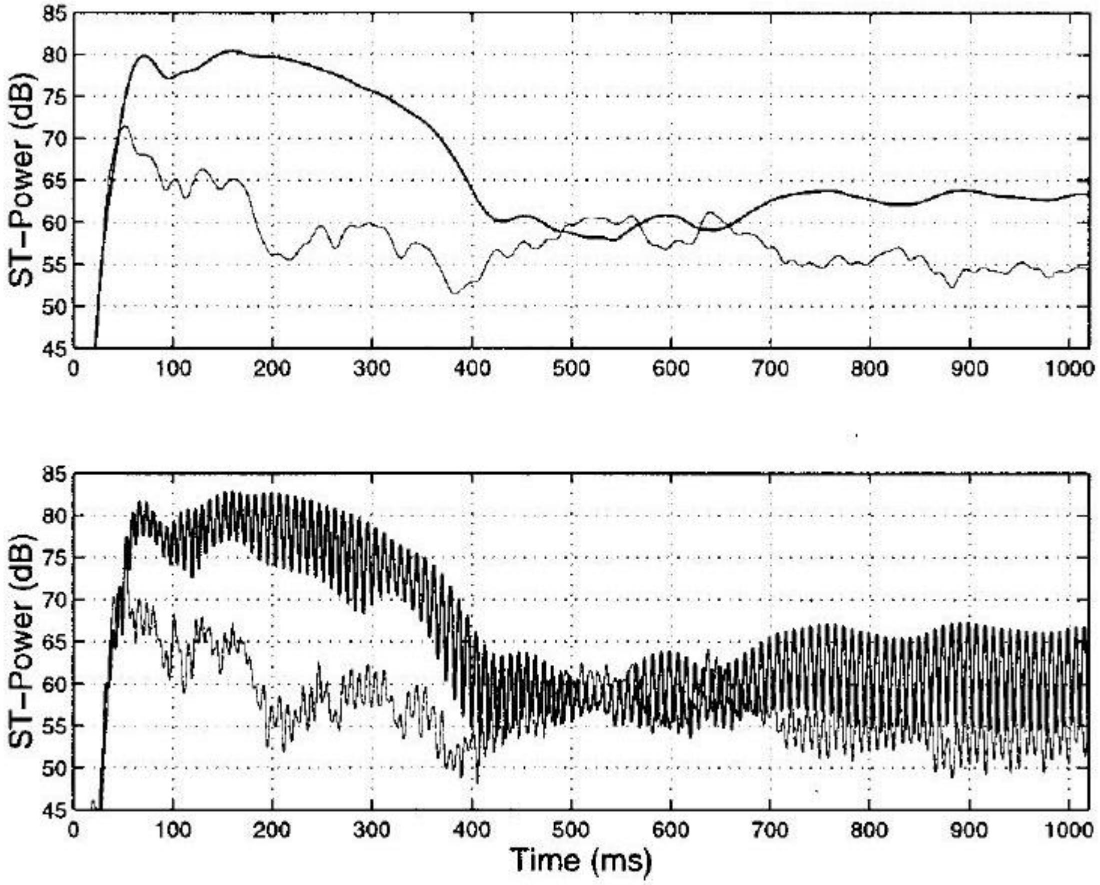


Figure 24: The short-time power calculated over the medium term (top, 32 ms analysis window) and the short term (bottom, 8 ms analysis window) for the decomposed components from a production of /ɑz:/ by a male speaker: (thick) harmonic component, and (thin) noise component. From Jackson & Shadle (2000).

harmonic and noise components simultaneously for a production of the syllable /ɑz:/, in which the vowel transitions to a voiced strident fricative at around time 400 ms. Note that the power of the noise component (thin line) is not significantly higher in the fricative portion than in the vowel portion, even though such a strident fricative is certainly produced with an added noise source in the vocal tract. The fact that the new noise source does not significantly increase the overall noise power is attributed to a

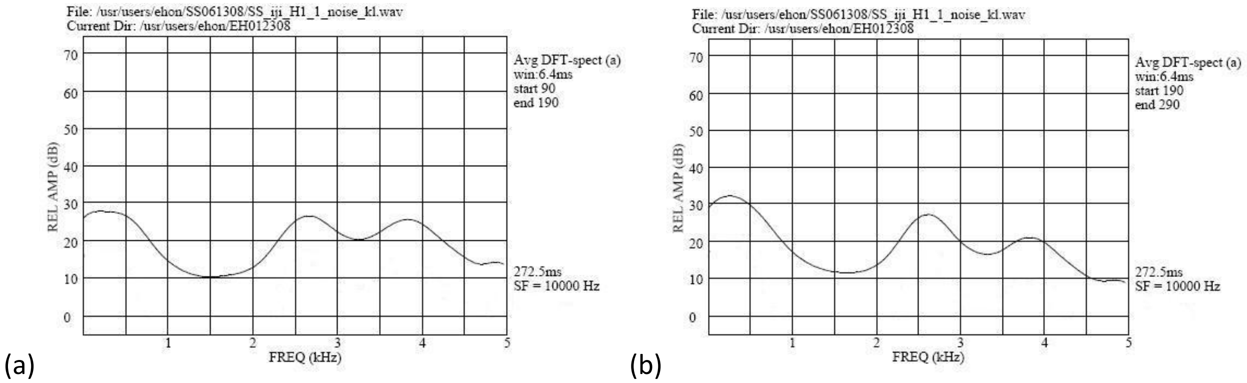


Figure 25: Comparison of the smoothed average spectra of the noise components of the vowel /i/ and the glide /j/, from the token /iji/ produced with high pitch accent on the first vowel (H1) by speaker F2. (a) average noise spectrum of the vowel /i/, (b) average noise spectrum of the glide /j/.

change in noise source location; the pre-existing noise power in the vowel region arises from an aspiration source, which is replaced in the fricative by a frication source, without significantly altering the total noise power (Jackson & Shadle, 2000). If such a change in noise source type and location also occurs for glides, the lack of total noise power increase could still be consistent with the possible addition of a frication noise source at the glide oral constriction.

A cursory spectral analysis of the glide and vowel noise components, however, does not seem to support the idea that a change in noise source occurs between the two segment types. For example, Figure 25 compares the average spectrum of the vowel /i/ with that of the glide /j/, from the token /iji/ produced with high pitch accent on the first vowel (H1) by speaker F2 (one of the tokens which showed the greatest decrease in HNR in the glide compared to the vowel). Note that the overall shapes of both noise spectra are very similar, with peaks corresponding to the formant frequencies of /i-j/. If a frication noise source were present in the glide and not in the vowel, one would expect to see a boost to the spectrum in the high frequency region for /j/ (Stevens, 1971), but this does not appear to be the case. It seems more likely that an aspiration noise source (exciting all of the vocal tract formant frequencies) is active in both the vowel and the glide. Its power does not decrease in the transition between the vowel and the glide, but the harmonic source amplitude does decrease significantly, resulting in decreased HNR in the glide.

2.5.3 Discussion

ΔHNR measurements on the /VGV/ recordings database show that glides have significantly reduced HNR compared to their adjacent vowels. However, the acoustic evidence points to reduced harmonic amplitude as the primary cause of the reduced HNR, rather than any increase in noise power. In particular, there does not appear to be a change in the location of the active noise source during the transition between a vowel and a glide. Rather, it appears that aspiration noise is present in both segments; its average power is not greater in glide segments than in adjacent vowels, even though the average open quotient is significantly larger. The lack of acoustic evidence for frication noise activity in glide segments may contradict assertions that turbulence associated with the narrow constriction distinguishes glides from vowels and motivates phonological processes such as palatalizing mutations (Padgett, 2008). However, it is important to remember that the present acoustic study is restricted to glide productions in intervocalic contexts in American English; it is possible that glides are produced with more evidence of frication in other segmental contexts and/or in other languages.

The fact that noise power is not increased in glides relative to their adjacent vowels does not preclude the perceptual interpretation of decreased HNR as increased noisiness, as is evident from Figure 24. The fricative /z/ can be assumed to be perceived as a noisier segment than the adjacent vowel /ɑ/, even though the level of the noise power is not much increased in the fricative. The important factor would appear to be the relation between the harmonic component and the noise, i.e., the HNR, rather than the absolute level of the noise itself. Given this comparison, it is reasonable to ask whether the decreased HNR of glides might be perceived by listeners as increased noise salience, regardless of the absolute levels of the isolated harmonic and noise components. Since it was found in the previous section that the magnitude of the HNR reduction (ΔHNR) was not significantly affected by vowel height context or prosodic context, this noise salience pattern might provide a relatively invariant acoustic cue for the detection of glides in speech.

If the decreased HNR of glides is indeed perceived as increased noise salience, this may also provide a convenient means of enhancing or strengthening the glide-vowel distinction through the deliberate addition of noise. Although frication noise does not appear to be present in the non-emphatic recordings of glides collected in this study, it is possible that speakers could add an extra frication noise source in certain circumstances to heighten the HNR contrast between an important glide and the adjacent vowel. This enhancement strategy would be more likely in a language like English than in one with contrasting palatal or bilabial voiced fricatives, since noisy glides can be freely produced in English without impinging on the acoustic territory of another phoneme. (Phenomena of language-specific enhancement have been discussed in depth by Keyser & Stevens (2006).) Future studies, perhaps of emphatic speech, should investigate this possibility.

2.6 Fundamental frequency (F0)

2.6.1 Method

As mentioned in Section 2.2.1, pitch period lengths were calculated in all of the /VGV/ tokens in the recordings database, so that pitch-synchronous spectral measurements could be performed with appropriate analysis window lengths. The pitch period lengths were calculated as the distance between the waveform zero-crossings just before the maximum amplitude excursions of the pitch periods, corresponding to the time between the beginnings of the closed phases of successive glottal cycles. The periods were calculated for every glottal cycle over a time interval starting from the midpoint of the first vowel, continuing through the glide segment, and ending at the midpoint of the second vowel. An

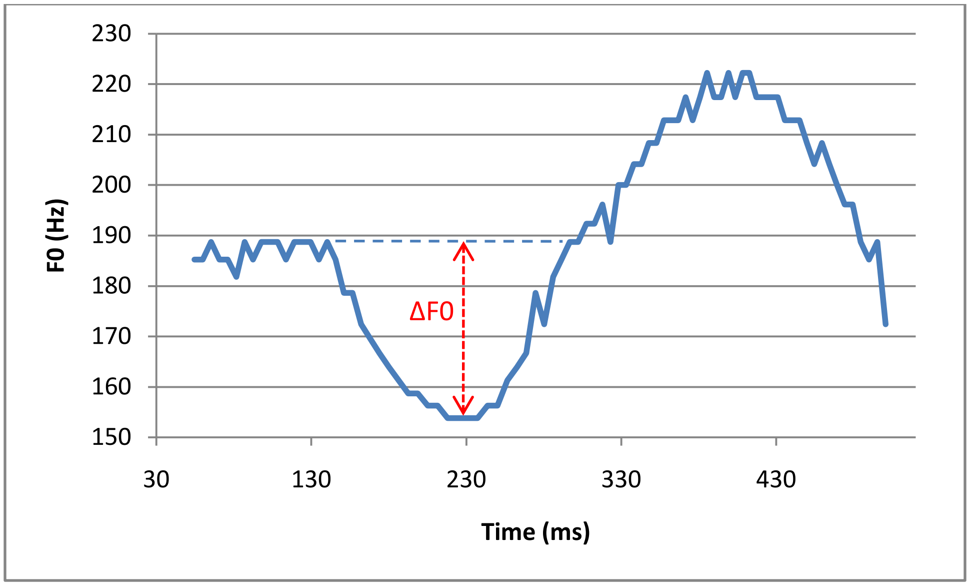


Figure 26: Measured F0 contour from a token of /iwi/ produced with high pitch accent on the second vowel (H2) by speaker F1. The midpoint of the glide /w/ occurs around time 230 ms. The dashed blue line is a conservative projection of the F0 contour without deviation caused by vocal tract loading on the glottal source. The red arrow illustrates the measurement ΔF0, calculated by subtracting the minimum F0 reached during the glide from the projected non-interaction F0.

additional acoustic measurement, fundamental frequency (F0), could then be determined as a function of time by inverting the period of each glottal cycle. Dynamic excursions within the F0 contours could signal the presence of glides with narrow vocal tract constriction, since the resulting loading on the glottis may have the effect of lowering F0 (Bickley & Stevens, 1986; Titze, 2008; Zañartu *et al.*, 2007).

Figure 26 shows an example F0 contour measured from the token /iwi/ produced with high pitch accent on the second vowel (H2) by speaker F1, with the individual measurement points connected by straight lines to form a continuous curve. The glide landmark is located around time 230 ms; note that there is a pronounced local minimum in the F0 contour at this point. This is not to be expected from the prosodic contour alone; the only prosodic target in this token is the high pitch accent on the second vowel, accounting for the maximum peak in F0 around time 400 ms (see Selkirk (1984b) for an overview of the placement of pitch accents relative to syllables). There is no prosodic reason for the extra F0 minimum to occur during the glide segment, but a possible cause is the involuntary effect on F0 of the vocal tract constriction loading on the glottal source. The brief narrowing of the constriction during the glide causes F0 to decrease momentarily from its assumed prosodic contour if no loading were present, which is represented by the dashed line drawn in Figure 26. (It should be noted that this horizontal line is the most conservative estimate of the non-interaction F0 contour; it is possible that the normal prosodic contour would have F0 increase somewhat steadily between the first vowel and the second vowel, rather than remaining flat until the peak.) In tokens in which this F0 deviation occurred around the time of the A_{RMS} and F1 minima during the glide segment, its magnitude (ΔF0) was quantified by subtracting the minimum F0 from the projected estimate designated by the dashed line, as illustrated by the red arrow in Figure 26.

Figure 27 shows a similar plot of the F0 measurements for an example token with a falling prosodic contour. The token is /uju/, produced with a low pitch accent on the second vowel (L2) by speaker F2. The glide landmark occurs around time 230 ms; again, there is an extra local F0 minimum during the glide segment, in addition to the global minimum reached at the low pitch accent target of the second vowel around time 400 ms. Since an acoustic study does not have access to the exact shape that the F0 contour would have taken in the absence of source-filter interaction during the glide segment, the magnitude of the F0 deviation (ΔF0) can only be consistently measured with respect to the conservative horizontal projection designated by the dashed line. The true ΔF0 is likely to be somewhat underestimated if the non-interaction contour would have been steeply sloped, and the degree of underestimation may be systematically larger in some prosodic contexts than in others.

According to Titze (2008), ΔF0 is predicted to be larger the closer F0 and F1 are brought together. Since this happens to a varying degree in glides of varying constriction strength (as measured by degree

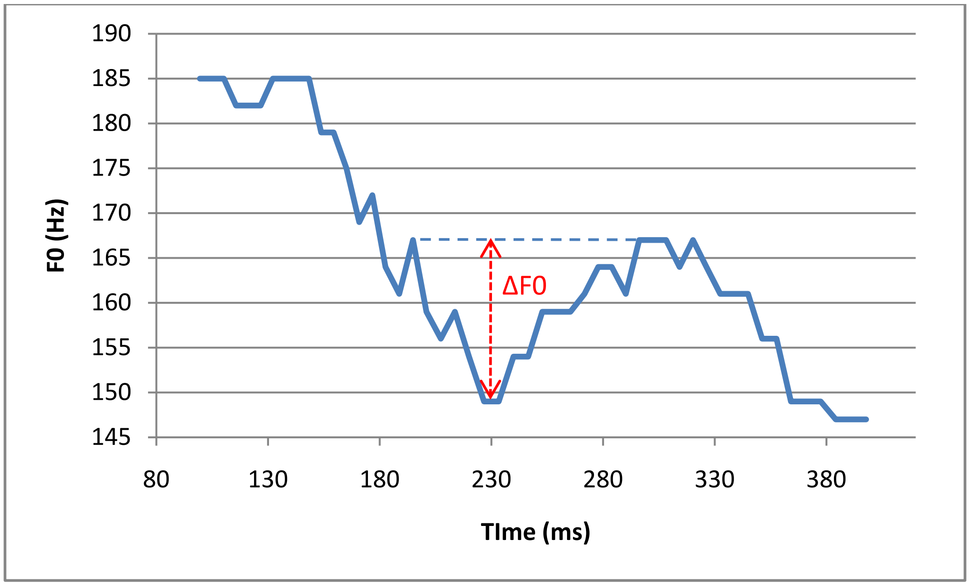


Figure 27: Measured F0 contour from a token of /uju/ produced with low pitch accent on the second vowel (L2) by speaker F2. The midpoint of the glide /j/ occurs around time 230 ms. The dashed blue line is a conservative projection of the F0 contour without deviation caused by vocal tract loading on the glottal source. The red arrow illustrates the measurement ΔF0, calculated by subtracting the minimum F0 reached during the glide from the projected non-interaction F0.

of F1 lowering) and with speakers who have different F0 producing various prosodic contours, this study's database of glide recordings provides very relevant data with which to test this prediction. Accordingly, a measure of the proximity between F1 and F0 was also made for each glide token, as illustrated in Figure 28. This measure was calculated by subtracting the projected estimate of the non-interaction F0 from the minimum F1 reached during the glide.

2.6.2 Results

A measurable downward deviation from the surrounding F0 contour was present in many, but not all, of the glide segments in this study's /VGV/ database. Table 6 lists the percentage of glides that do exhibit an F0 deviation (i.e., ΔF0 > 0), as conditioned by speaker and surrounding vowel height context. From this table, it can be seen that the rate of occurrence of F0 deviation is higher for the female

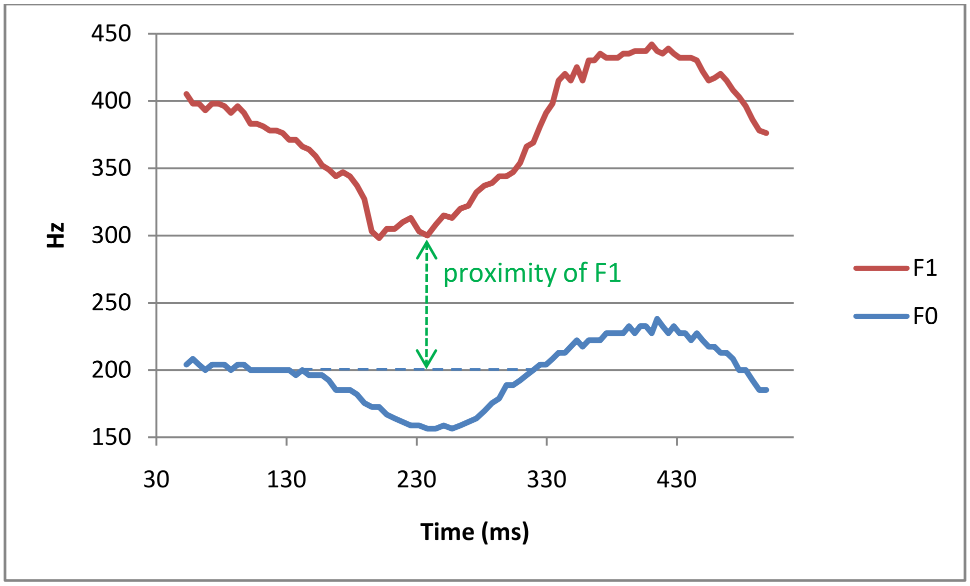


Figure 28: Measured F0 and F1 contours from a token of /uju/ produced with high pitch accent on the second vowel (H2) by speaker F1. The midpoint of the glide /j/ occurs around time 240 ms. The dashed blue line is a conservative projection of the F0 contour without deviation caused by vocal tract loading on the glottal source. The green arrow illustrates the measurement of the proximity between F1 and F0, calculated by subtracting the projected non-interaction F0 from the minimum F1 reached during the glide.

speakers than for the male speakers, and is higher in high vowel contexts than in low vowel contexts. These trends are in agreement with the source-filter interaction hypothesis, which predicts that greater loading is placed on the glottal source when F0 and F1 are in close proximity. This condition may be met to a greater extent when F0 is higher, as it typically is in female speech, or when F1 is lower, as it was shown to be in glides in high vowel contexts in Section 2.3.

Table 6: Percentage of tokens for each combination of speaker and vowel context that showed measurable downward deviation from the normal prosodic F0 contour during the glide segment (out of 20 total tokens in each speaker-vowel category in the /VGV/ database).

	Speaker F1	**Speaker F2**	**Speaker M1**	**Speaker M2**
high vowel context	85%	90%	85%	75%
mid vowel context	85%	75%	70%	40%
low vowel context	70%	60%	50%	45%

The rate of occurrence of F0 deviation is also likely to depend on prosodic context, since prosodic contrasts are expressed through F0 movements, but the dependence is not expected to be in simple relation to the five prosodic categories elicited in this database. The specific F0 contour of a particular glide segment arises from a combination the speaker's average F0, the F0 range of the particular utterance, and the overlaid prosodic targets. However, the true variable of concern here, regardless of contextual category, is the degree of proximity between F0 and F1 during the glide segment. If the hypothesis of acoustic loading on the glottal source is correct, then the magnitude of the deviation, ΔF0, should be correlated with the proximity between F0 and F1 across all tokens.

This condition was tested on this study's database by plotting the ΔF0 measure against the F1-F0 proximity measure for all tokens, as shown in Figure 29(a). A Spearman rank correlation analysis indicates that the two measures are highly correlated (ρ=-.558; t(238)=-10.374; p=.000), confirming that closer proximity between F1 and F0 coincides with greater ΔF0 in glide segments. A downward deviation in F0 can be conditioned by the narrow (low F1) vocal tract constriction of the glide segment causing loading on the glottal source, but the degree of deviation is evidently dependent on how close the non-interaction F0 and F1 were to begin with. Thus, this acoustic characteristic of glides is likely to be more commonly present in speakers with high average F0, such as females.

2.6.3 Discussion

A local minimum in fundamental frequency (F0) has been found to be a common characteristic of glide segments in predictable contexts. The magnitude (and thus the observability) of the F0 deviation (ΔF0) is correlated with the proximity between F0 and the first formant frequency (F1). Thus, greater ΔF0 is expected to occur for glides produced with narrower vocal tract constrictions, i.e., greater strength of articulation, or less coarticulation with neighboring segments. The F0 deviation in glides is also expected to be more common for speakers with higher average F0, such as females.

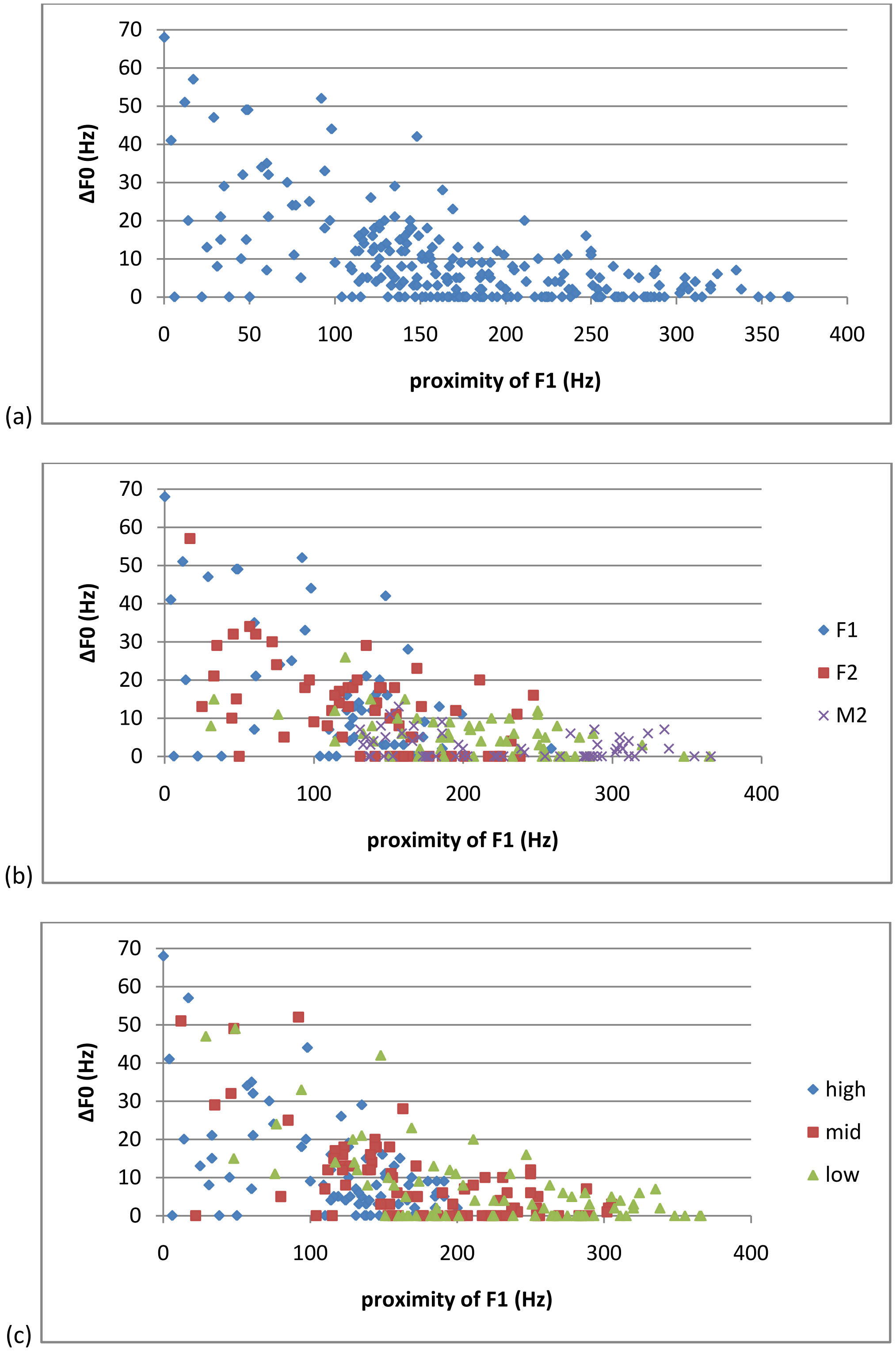


Figure 29: Scatterplots of measured ΔF0 vs. F1-F0 proximity for all 240 tokens in the /VGV/ database. (a) all tokens plotted together; (b) tokens plotted separately by speaker; (c) tokens plotted separately by vowel height context.

If the correlation between ΔF0 and the relative articulatory strength of glide segments is evident to listeners, this acoustic cue may provide another possible avenue for enhancement of the glide-vowel contrast. In the non-emphatic recordings collected in this study, it is assumed that the F0 effects in glides are involuntary and conditioned by the aerodynamic interaction between the glottal source and the vocal tract filter. However, F0 can also be independently controlled through deliberate laryngeal adjustments, and speakers may find it desirable to voluntarily add further F0 decrease to their production of glides to enhance them in certain circumstances. As with the idea of noise enhancement discussed in Section 2.5.3, the possibility of enhancement through F0 adjustments could be investigated through future studies of glides in emphatic or hyper-articulated speech.

It is interesting to note that the effect of the glide articulation on F0 seems to run counter to the direction of the intrinsic pitch of vowels. Studies of many languages have consistently shown that high vowels, which also have relatively low F1, are produced with higher F0 than low vowels. In this study, however, the low F1 of glides has been shown to be correlated with decreased F0. It would seem that the two phenomena, vowel intrinsic pitch on one hand and glide F0 effects on the other, arise from different mechanical and aerodynamic interactions in the speech production system. If vowel intrinsic pitch is indeed an involuntary consequence of articulation, the most likely cause seems to be the pull of the tongue on the laryngeal system (Whalen & Levitt, 1995). On the other hand, the theoretical explanation for decreased F0 in glides is the acoustic loading and decreased transglottal pressure produced by the narrower constriction in the vocal tract. If both theories are correct, it would appear that intrinsic pitch and source-filter interactions could both be active simultaneously, in antagonistic contribution to the same acoustic parameter of F0.

A further question of interest is whether other voiced consonants might exhibit the same F0 characteristics as glides have shown in this study. For example, it is widely known that obstruent consonants have their own form of intrinsic pitch that manifests in the initial F0 of the following vowel; F0 is lower following voiced obstruents than following unvoiced obstruents. The aerodynamic effect of reduced transglottal pressure has been suggested as one possible source of lowered F0 in voiced stops (Hombert, Ohala, & Ewan, 1979), as it has been applied to the F0 of glides in this thesis. However, the aerodynamic hypothesis for stops is rejected by some in favor of the vocal-fold tension hypothesis, which suggests that F0 is raised for voiceless stops by stiff vocal folds, not lowered for voiced stops. The fact that nasal consonants have the same intrinsic pitch as voiced stops has been used as an argument against the aerodynamic F0-lowering hypothesis, since nasals are regarded as neutral segments with respect to pitch (Hanson, 2009). However, in light of the results of the current study on glides, this stance should perhaps be reconsidered. It may be that nasals and stops both experience the same F0-lowering effects of narrow vocal tract constrictions that glides do; this would not preclude the simultaneous and opposite contribution of the vocal-fold stiffness parameter for the obstruent intrinsic pitch distinction.

2.7 Summary

This chapter has described the analysis of various potential acoustic cues to the distinction between glides and vowels. Some of these cues arise from the vocal tract cavity configurations that are formed when the tongue is placed at its characteristic height for the glides /j, w/. These types of cues have been described in previous research, although their dependence on vowel and prosodic context has not been systematically explored before now. Other cues investigated in this study are expected to arise from aerodynamic effects on the glottal voicing source, which are caused by a pressure drop across the narrow constriction formed in the oral cavity for glides. These latter types of cues have been relatively unstudied in previous research. In the literature reviewed for this thesis, Stevens's *Acoustic Phonetics* (1998) is the only work to suggest the possibility of glottal source effects in direct reference to glides;

however, even that discussion states confidently that glides are “produced with little or no pressure drop in the airways above the glottis” (Stevens, 1998, p. 513)

Of the first type of acoustic cues for glides, relating to vocal tract cavity configuration, the most commonly studied are the formant frequencies, including F1. The degree of F1 lowering during a glide reflects the narrowness of the constriction formed by the tongue in the oral part of the vocal tract, up to a certain point. Section 2.3 of this thesis has shown that, for glides in high vowel contexts, F1 does not decrease much further than its already low frequency during the vowel. If the vocal tract constriction is made narrower in the glide than it is for a high vowel, F1 begins to become less sensitive to the constriction area due to wall effects. It was also found in Section 2.3 that the $F1_{min}$ target of glides is highly sensitive to coarticulation with the F1 of adjacent vowels. $F1_{min}$ in low vowel contexts is significantly higher than $F1_{min}$ in high vowel contexts, and it is sometimes even higher than typical F1 for high vowels. Thus, coarticulatory effects can potentially blur the acoustic distinction between glides and some vowels, if the focus remains solely on formant frequencies.

Although the F1 characteristics of glides were found to vary through coarticulation with neighboring vowels, Section 2.2 showed that the amplitude characteristics are not sensitive to vowel height context. Glides were produced with significant amplitude reduction compared to their adjacent vowels for all four speakers in the database, and 90% of all tokens were produced with $\Delta A_{RMS} \geq 7$ dB. Since amplitude reduction is automatically produced by reduction in F1, A_{RMS} can be partly considered an acoustic cue related to vocal tract cavity configuration. The F1 contribution to A_{RMS} is greatest in glides in low vowel contexts, since the F1 movement in frequency is quite large despite the effects of coarticulation. However, the fact that similar ΔA_{RMS} is also found in high vowel contexts, despite the relative lack of F1 movement there, indicates that the A_{RMS} characteristics of glides also receive contributions from glottal source effects. For glides in high vowel contexts, A_{RMS} can be considered to belong completely to the category of acoustic cues related to aerodynamic source-filter interaction, since the contribution of F1 appears to be somewhat negligible in that context.

The presence of aerodynamic effects on the glottal source in glides was confirmed in Section 2.4 through open quotient analyses. Glides were produced with significantly greater OQ than their adjacent vowels for all four speakers, indicating that the glottal waveform is skewed to some degree during glides. This can arise from a pressure drop formed across the vocal tract constriction, building pressure in the cavity above the glottis and causing the transglottal pressure drop to decrease. ΔOQ was found to vary with vowel height context in a manner predictable from the aerodynamic source-filter interaction hypothesis, although its variation was not as great as that for $F1_{min}$. Another acoustic cue related to glottal source effects is the local minimum in F0 found in glides in Section 2.6. Since this cue is expected to arise from acoustic loading by the vocal tract filter on the glottal source, its magnitude and observability depend on the narrowness of the vocal tract constriction as well as the proximity of the intended F0 to F1. Thus, ΔF0 exhibits significant contextual variation, and is likely best used as a contrast-enhancing cue to the glide-vowel distinction.

In Section 2.5, it was found that the harmonics-to-noise ratio of glides was significantly reduced compared to their adjacent vowels for all four speakers. Analysis of the isolated noise components in the glides and vowels suggested that this HNR decrease was attributable more to the voicing amplitude reduction already noted than to any increase in noise power or addition of a new noise source during the glide. It seems that the canonical glide productions do not include acoustic evidence of frication noise, although it is possible that speakers might choose to deliberately add frication noise to enhance the HNR contrast in some situations. Added noise and deliberately decreased F0 are two potential avenues for enhancement that were identified in this chapter. It is suggested that future studies investigate their potential use in emphatic or hyper-articulated glides. The acoustic analyses of prosodic contexts in this study found some evidence of strengthening of glide articulations in pitch-accented syllables, as specifically reflected in effects on ΔA_{RMS} and $F1_{min}$; these types of prosodic environments could possibly be utilized to elicit enhanced productions.

Each of the acoustic cues investigated in this chapter (A_{RMS}, F1, OQ, HNR, F0) has been found to provide information that characterizes glides and their distinction from vowels. All (except perhaps F0) can be defined as acoustic relations between the articulatory target of the glide segment and that of the adjacent vowel segment. This target occurs at a specific moment in time, corresponding with the glide landmark, and the acoustic characteristics need not be described with reference to durational relationships or transition rate. This argues for the use of a distinctive feature to classify glides separately from vowels, rather than merely an appeal to syllabicity (for a discussion of features defined by target acoustics, see Stevens & Hanson (in press).)

Stevens & Hanson (in press) also observe that articulator-free features, such as the potential feature for glidehood, are usually defined by aero-mechanical interaction rather than acoustic resonator coupling. This suggests that the acoustic cues related to aerodynamic source-filter effects may be more central to defining the glide feature than those related to vocal-tract cavity configuration. This ranking is supported in the acoustic data by the fact that A_{RMS}, OQ, and HNR, but not F1, are active in the glide vs. high vowel distinction, which is the articulatory category boundary of interest in categorizing glides as non-vowels. In the following chapter, further evidence will be brought to bear on this ranking through a comparison of the perceptual salience of different cues distinguishing between glides and vowels.

3 *Perceptual study of glide detection*

This chapter presents a perceptual study that was undertaken to test the relative salience to listeners of some of the potential acoustic cues discussed in Chapter 2 to the distinction between glides and vowels. Arguments against a separate feature class for glides hold that the glides /j, w/ are lexically high vowels /i, u/ occurring outside of the syllabic nucleus. Thus, the issue of interest for this study is whether certain acoustic cues can be used by listeners to differentiate glides specifically from their "cognate" high vowels (i.e., to differentiate /j/ from /i/, and /w/ from /u/). In order to provide evidence supporting a featural distinction rather than a mere difference in syllabicity, the cues to be investigated should characterize a time-specific acoustic target that is unrelated to durational cues, and no information about whether glides or high vowels are perceived should be provided by lexical syllabicity constraints.

In addition, there is a larger purpose to this study than to simply rank the acoustic cues to glidehood in terms of their relative perceptual salience (although that in itself would provide new knowledge to advance models of human perception, speech recognition, and lexical access). It is perhaps of even greater value to relate these acoustic cues to articulatory properties, in order to understand which articulatory-acoustic relations are most important in defining a feature related to the category boundary between glides and vowels. In Chapter 2, the acoustic characteristics identified for glides were theoretically associated with two different physical sources. One is the shape of the vocal tract filter, and the other is the shape of the glottal excitation as conditioned by its coupling to that filter. It was mentioned in Section 2.7 that the glottal source effects are more consistent with other articulatory-acoustic relations defining articulator-free features, since they arise from aero-mechanical interactions. In addition, the analyses of Chapter 2 suggest that the acoustic correlates of the glottal source effects, especially the characteristic amplitude reduction occurring in glides, are more invariant across segmental contexts than the isolated filter effects are. In this chapter, the two types of acoustic cues will be compared from an additional angle, that of their relative importance in the perceptual distinction between glides and high vowels.

Of the acoustic cues related to the vocal tract filter configuration (primarily formant frequencies), the one most specifically related to the distinction between glides and high vowels is the first formant frequency (F1). F1 is known to decrease in frequency whenever a constriction is narrowed at any location in the oral cavity, and its direction of change is the same for both glides /j/ and /w/. (By contrast, the other formant frequencies, especially F2, are expected to move in different directions for /j/ and /w/ because their constrictions are located at different places along the vocal tract.) The F1 data collected from the /VGV/ database indicate that F1 often does decrease by a small amount between high vowels and glides, although it was shown in Section 2.3 that the amount of F1 decrease in high vowel contexts is quite small compared to that in other vowel height contexts. The question of interest for this study is whether the F1 decrease that does occur in glides in high vowel contexts is enough to cause the perceptual detection of the glide, in isolation from other acoustic cues. If natural amounts of F1 decrease are not perceptible, or are not utilized in listeners' perception of glides, this would cast doubt on the previously offered hypothesis that the vocal tract filter shape by itself can differentiate glides from high vowels.

Of the acoustic cues to glidehood related to glottal source effects, the most consistent across all tokens was found in Chapter 2 to be the decrease in voicing amplitude, ΔA_{RMS}. This acoustic parameter also has the advantage of being relatively simple to manipulate and verify in speech synthesis. Other acoustic effects of source-filter coupling were associated with the glide-vowel distinction in Chapter 2, including open quotient (OQ) increase and fundamental frequency (F0) decrease; however these were found to be more variable with context and less strongly different between glides and vowels than

ΔA_{RMS}. In order to limit the complexity of the synthesis required for the experiment and the time commitment required of the experimental subjects, the voicing amplitude cue was therefore selected as the single representative of glottal source effects for this perceptual study. If the amplitude decrease is found to cue the presence of glides without contribution from the F1 cue, this would support the hypothesis developed earlier in this section, that effects on the glottal source are central to the distinction between glides and high vowels.

It should be noted that the results of this perceptual study need not necessarily point to the exclusive use of one acoustic cue or the other in the perception of glides. Listeners may use both cues to varying degrees in their detection of glides, especially since they often occur together in natural speech signals. When the cues are isolated from each other in synthetic speech stimuli, listeners may trade the information provided by one cue off that of the other cue; such trading relations have been studied extensively in the speech perception literature for other sound segments. Repp (1982) provides a review of phonetic trading relation studies for many different phonetic category distinctions; it is worth noting, however, that the only category boundary study mentioned involving glides tested only the durational and transition rate cues to the distinction between glides and stop consonants. To the author's knowledge, the perceptual category boundary between glides and vowels has not been investigated with synthetic control of isolated acoustic cues before now, and the movement beyond durational characteristics to target-related parameters is quite novel for glides.

3.1 Synthesis

Artificial speech synthesis was used to create the perceptual stimuli for this study, so that the F1 and amplitude parameters could be isolated and manipulated independently of other acoustic cues and of each other. Since the category boundary of interest is between glides and high vowels, real-word tokens were chosen to create minimal pairs, in which one member contained a glide and the other contained only the corresponding high vowel. The minimal pairs selected were: “see yeast” vs. “see east” (/si#jist/ vs. /si#ist/) and “Sue woos” vs. “Sue oohs” (/su#wuz/ vs. /su#uz/). The intervocalic placement of the glides allowed for direct comparison with the acoustic data gathered from the recorded /iji/ and /uwu/ tokens in Chapter 2. The placement of a word boundary within the stimulus tokens was required since /VV/ sequences are not allowed word-medially in English if the two vowels are the same. The /V#GV/ and /V#V/ sequences were flanked with consonants in order to create pairs of real one-syllable words; alveolar fricatives were chosen for the flanking consonants because they are relatively easy to synthesize with natural-sounding results.

The KLSYN cascade formant synthesizer (Klatt D. H., 1980) was used to create the stimulus tokens for this experiment. Natural productions of all four of the selected word pairs were recorded in the author’s voice for use as copy-synthesis templates. The word pairs were recorded within the carrier sentence, “PLEASE don’t say ___ ___ again.” Emphasis was placed on the word “PLEASE” so that the target words would not receive any pitch accent. This in turn allowed the author to easily produce “see east” and “Sue oohs” without glottalizing the vowel beginning the second word, so that the minimal pairs were truly /V#GV/ vs. /V#V/, not /V#GV/ vs. /V#ʔV/. (Glottalization of word-initial vowels is common, but not required, in English, and is less expected in unaccented syllables (Pierrehumbert & Talkin, 1992; Dilley *et al.*, 1996).) The author produced multiple recordings of each word pair, and the durations of the /V#GV/ and /V#V/ tokens were compared. Within each minimal pair, the average duration of the token containing the glide was longer than that of the token without the glide, which is to be expected since the former contains one more sound segment than the latter. In order to eliminate the durational cue to the distinction in the perceptual experiment, the naturally produced /V#GV/ and /V#V/ tokens with the most similar durations were chosen from each minimal pair for copy-synthesis, and the duration of the synthesized tokens was set at a constant value intermediate between the two.

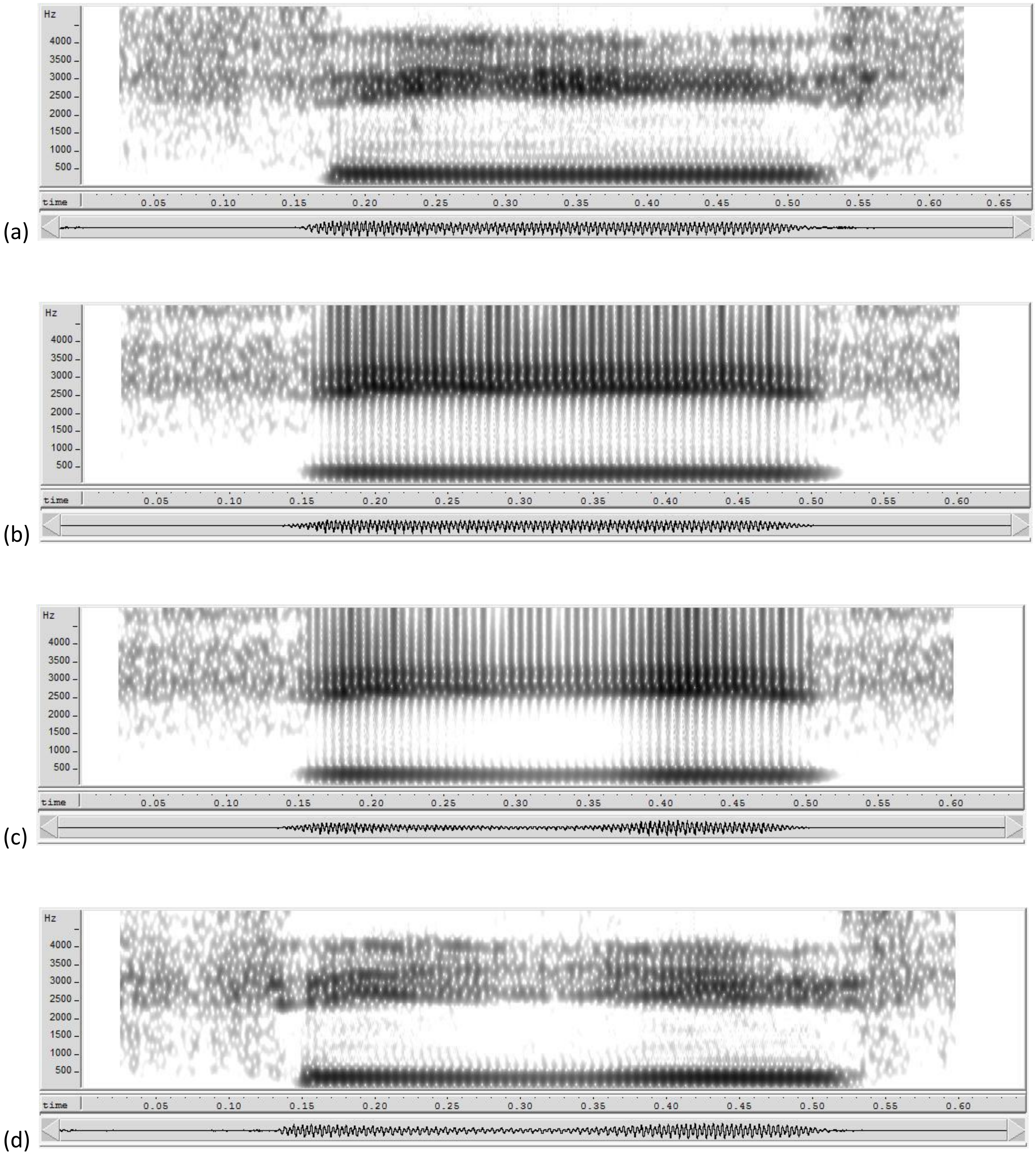


Figure 30: Spectrograms of naturally produced and copy-synthesized "see eas(t)" and "see yeas(t)".
(a) Naturally produced “see eas(t)”. (b) Synthetic “see eas(t)”, copy synthesized from naturally produced “see eas(t)” in (a). (c) Synthetic “see yeas(t)”, created by altering the AV and F1 contours of synthetic “see eas(t)” from (b). (d) Naturally produced “see yeas(t)”, used as a reference for the AV and F1 contours for synthetic “see yeas(t)” in (c).

From the natural speech recording of “PLEASE don’t say ‘see east’ again,” a portion extending from the beginning of the /s/ in “see” to the end of the /s/ in “east” was excised and used as a template for copy-synthesis of an artificial stimulus token. Figure 30 (a) and (b) show spectrograms of the naturally produced and copy-synthesized tokens of “see eas(t)”, respectively. The sampling rate was 10 kHz for both the recorded and the synthesized tokens. A synthetic token of “see yeas(t)” was then created from the synthetic “see eas(t)” token, by manipulating only the amplitude of voicing (AV) and first formant frequency (F1) parameters of the synthesizer. The naturally recorded token of “see yeas(t)” was used as a guide in the manipulation of the AV and F1 contours to create synthetic “see yeas(t)” out of the synthetic “see eas(t)” baseline. Figure 30 (c) and (d) show spectrograms of the synthetic and naturally produced “see yeas(t)”, respectively.

Synthetic “Sue oohs” and “Sue woos” tokens were created following the same procedure used for “see east” and “see yeast” above. Spectrograms of the naturally produced and copy-synthesized tokens can be seen in Figure 31. Note that the only differences between synthetic “Sue oohs” and “Sue woos” (and between synthetic “see eas(t)” and “see yeas(t)”) are in the AV and F1 contours used in the synthesis. All other parameters, including total duration, F0, and other formant frequencies, are identical between the two members of the synthetic minimal pair. This guarantees that any perceptual difference between the tokens is cued by one of the two acoustic parameters under investigation, namely voicing amplitude (AV) and F1. In particular, the fact that the durations of the two members of each minimal pair were kept equal satisfies the conditions given in the introduction to this chapter for the experiment’s targeting of featural acoustic cues to glidehood, rather than cues to syllabicity. Durational cues cannot be recruited to determine whether two vocoids are present or three; thus the target acoustic cues alone must either suffice or fail to provide the perceptual distinction between the glide and simple high vowel tokens.

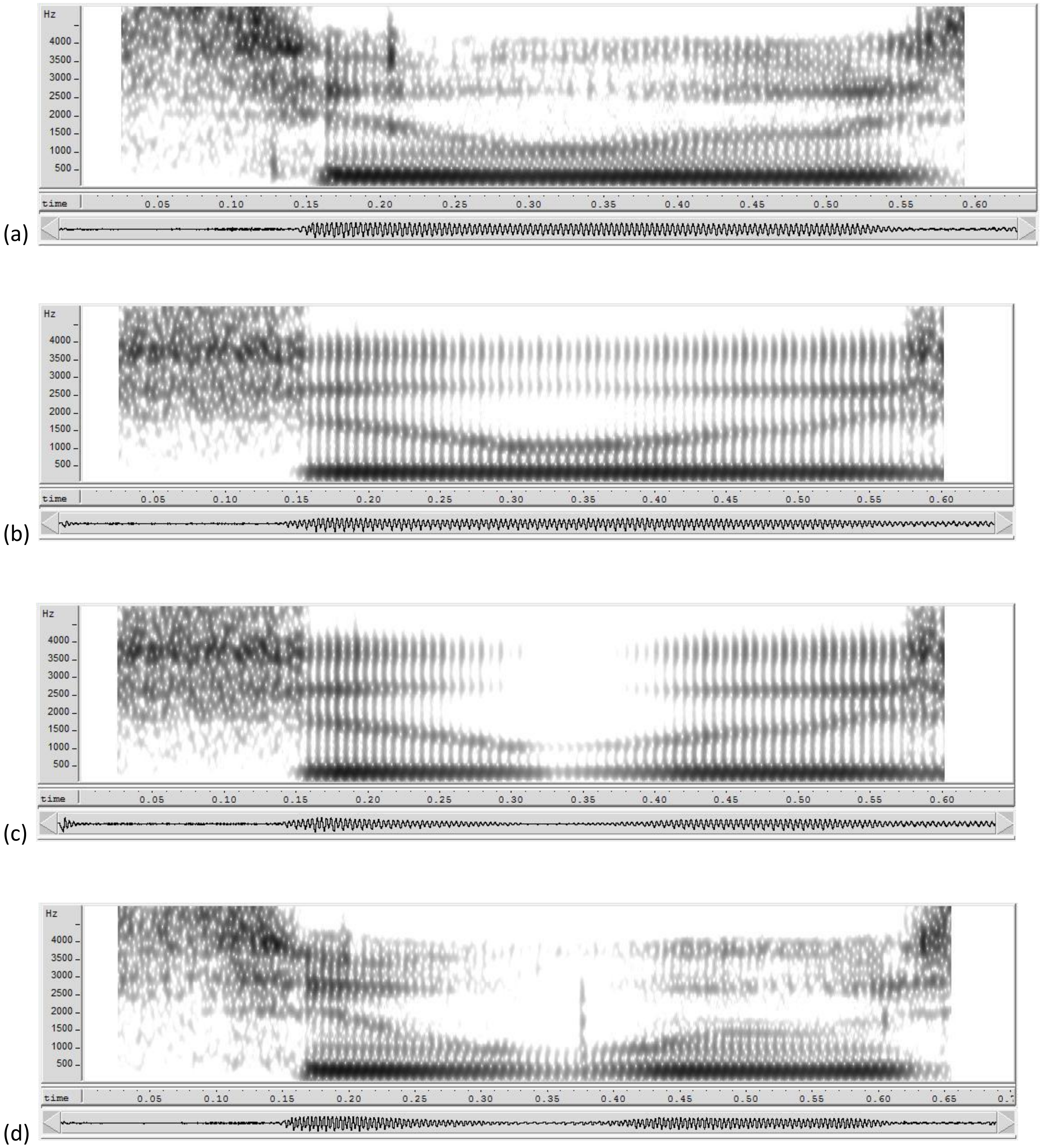


Figure 31: Spectrograms of naturally produced and copy-synthesized "Sue oohs" and "Sue woos". (a) Naturally produced “Sue oohs”. (b) Synthetic “Sue oohs”, copy synthesized from naturally produced “Sue oohs” in (a). (c) Synthetic “Sue woos”, created by altering the AV and F1 contours of synthetic “Sue oohs” from (b). (d) Naturally produced “Sue woos”, used as a guide for the AV and F1 contours for synthetic “Sue woos” in (c).

A set of continua of synthetic tokens, with acoustics intermediate between those of /V#V/ and /V#GV/, was created for each minimal pair by varying the amount to which the AV and F1 parameters were manipulated in equally-spaced steps. For the most /V#V/-like token, the AV and F1 contours were completely flat in the time between the two vowels; for the most /V#GV/-like token, large dips were placed in the AV and F1 contours to create the amplitude and F1 decrease for a glide between the two vowels. The maximum amount of decrease used to create a glide was 16 dB in the AV contour and 80 Hz in the F1 contour. These maximum ΔAV and ΔF1 values were selected following inspection of the range of ΔA_{RMS} and ΔF1 values in the /VGV/ database from Chapter 2 for glides adjacent to high vowels. Tokens with intermediate ΔAV and ΔF1 values were synthesized to create continua of sizes of AV and F1 dips, in steps of 2 dB for AV and 10 Hz for F1. All of the durational values, including the time and duration of the AV and F1 minima, were held constant at the values determined from the copy-synthesis of the naturally produced “see yeast” and “Sue woos”, as described above.

The AV and F1 contours used to generate synthetic “see east”/”see yeast” and “Sue oohs”/”Sue woos” continua are shown in Figure 32 and Figure 33, respectively. Nine contours were synthesized for each continuum, differing from each other in equally spaced steps of the size of the glide-like dip in the AV or F1 parameter. For the AV parameter, ΔAV was varied in 2 dB steps, from ΔAV = 0 dB (flat AV contour) to ΔAV = 16 dB (maximum AV dip). For the F1 parameter, ΔF1 was varied in 10 Hz steps, from ΔF1 = 0 Hz (flat F1 contour) to ΔF1 = 80 Hz (maximum F1 dip). Since the durational values were determined using the naturally produced “see yeast” and “Sue woos” as copy-synthesis guides, the durations of the AV and F1 dips differ between the synthetic “see east”/”see yeast” and “Sue oohs”/”Sue woos” continua. The amount of AV or F1 overshoot in the vowel following the potential glide was also determined through this copy-synthesis procedure, and therefore differs between the two continua.

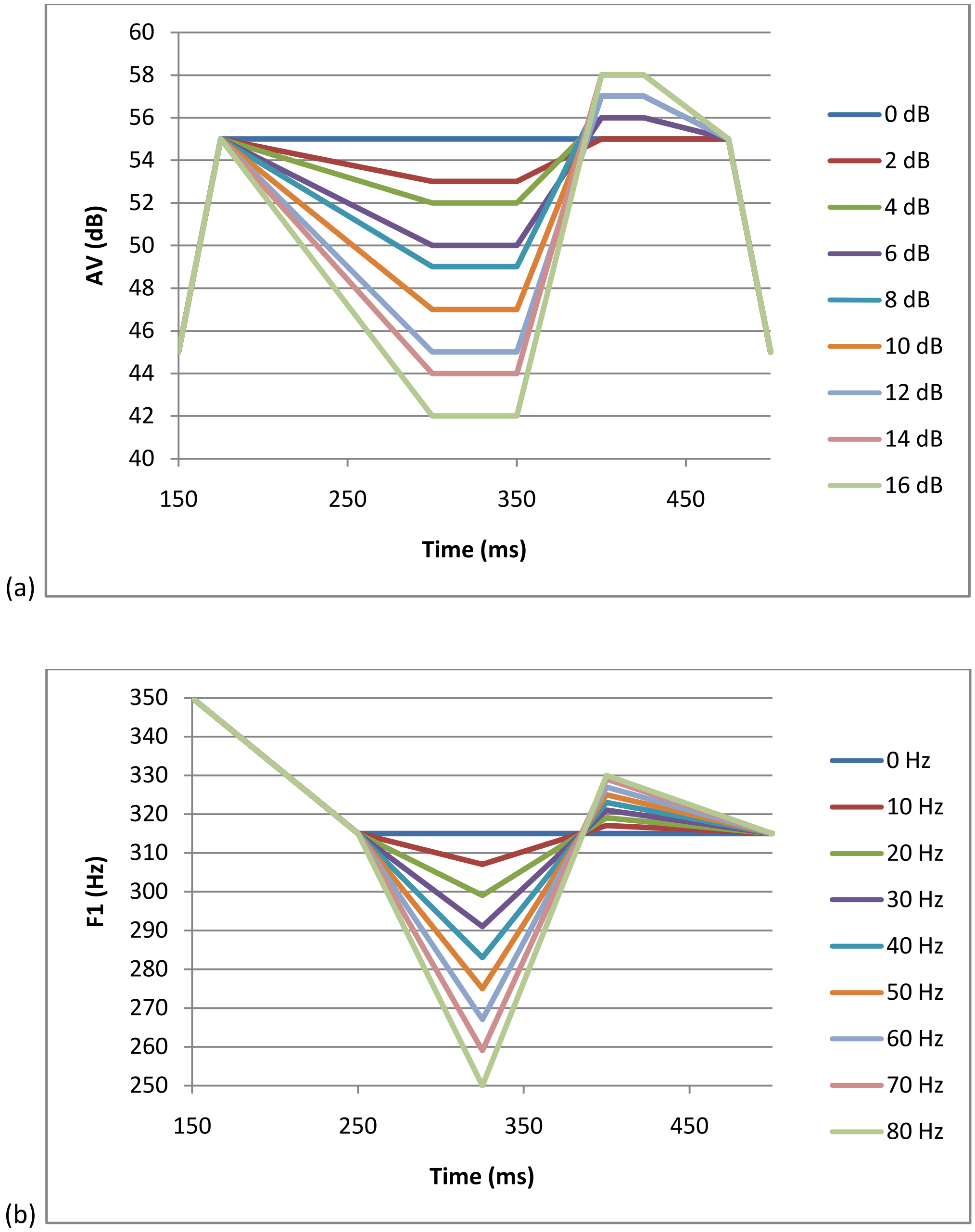


Figure 32: Amplitude of voicing (AV) and first formant frequency (F1) contours used to synthesize "see east"/"see yeast" continua. (a) AV contours, with ΔAV varied in steps of 2 dB, from 0 dB (flat AV contour) to 16 dB (maximum AV dip). (b) F1 contours, with ΔF1 varied in steps of 10 Hz, from 0 Hz (flat F1 contour) to 80 Hz (maximum F1 dip).

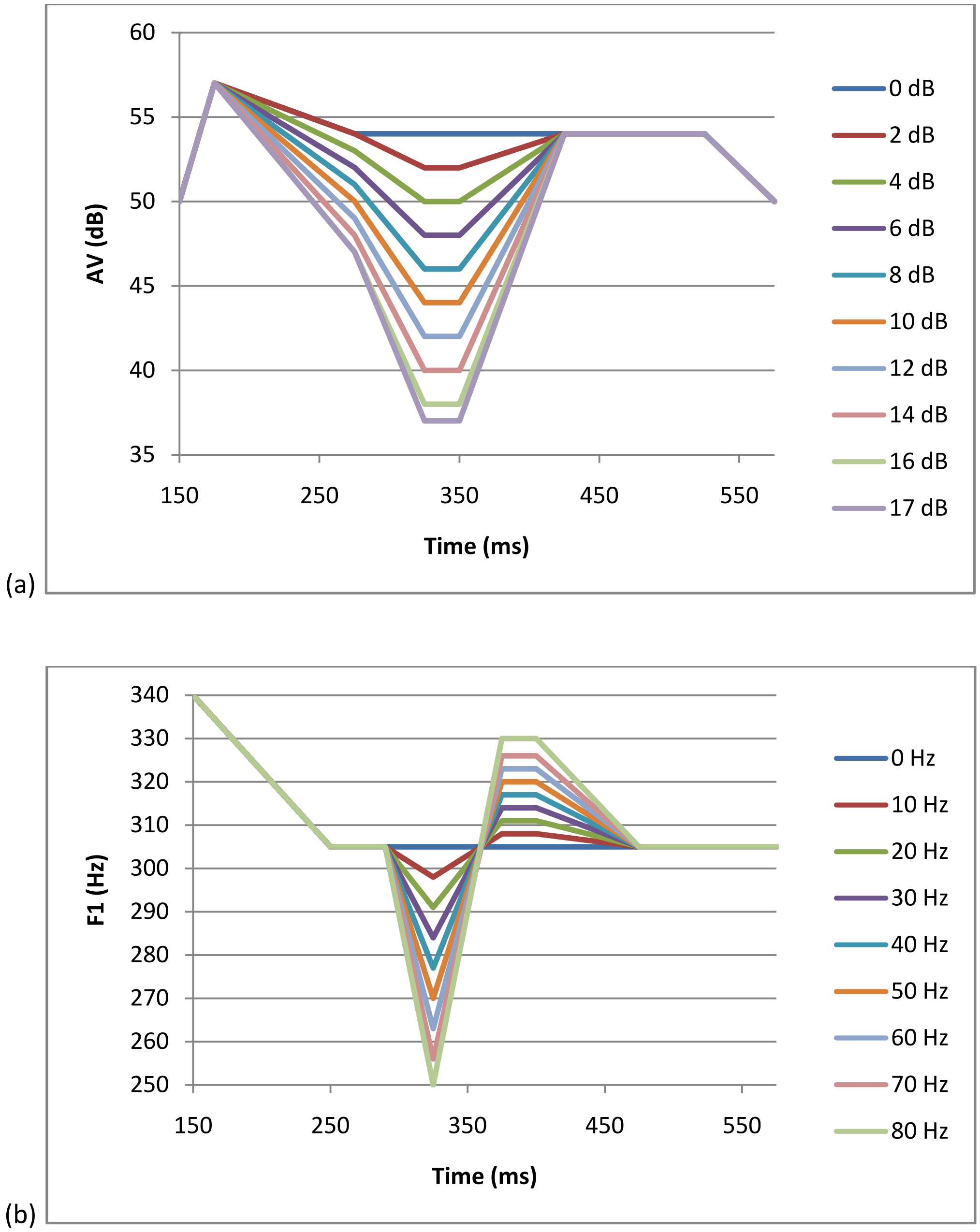


Figure 33: Amplitude of voicing (AV) and first formant frequency (F1) contours used to synthesize "Sue oohs"/"Sue woos" continua. (a) AV contours, with ΔAV varied in steps of 2 dB, from 0 dB (flat AV contour) to 16 dB (maximum AV dip). (b) F1 contours, with ΔF1 varied in steps of 10 Hz, from 0 Hz (flat F1 contour) to 80 Hz (maximum F1 dip).

Each synthetic token was reinserted into the naturally recorded “PLEASE don’t say ___ ___ again” carrier phrase to create the stimulus tokens for the perceptual experiments. For “see east”/”see yeast”, the naturally produced “PLEASE don’t say ‘see east’ again” was used as the carrier phrase. The synthetic “see eas(t)”/”see yeas(t)” was spliced into the carrier phrase to replace the naturally produced target words. The naturally produced /t/ burst at the end of “east” was retained in the carrier phrase, in order to preserve the formant transitions into the vowel beginning “again”. For “Sue oohs”/”Sue woos”, the naturally produced “PLEASE don’t say ‘Sue oohs’ again” was used as the carrier phrase, with the synthetic “Sue oohs”/”Sue woos” spliced in to replace the naturally produced target words. Again, care was taken to preserve the formant transitions by ending the splicing before visible formants replaced the frication noise in /z/. Figure 34 shows spectrograms of the naturally produced carrier phrases, and the stimulus tokens formed by splicing the synthetic target words into those carrier phrases, for the example of the synthetic tokens with ΔAV = 0 dB and ΔF1 = 0 Hz.

If all combinations of nine AV contours and nine F1 contours were presented to listeners in the perceptual experiment, this would result in 81 different stimulus tokens for each of the potential glides /j/ and /w/. Since it was desirable to present at least three repetitions of each test stimulus in order to average the responses, it was thought that this number of test items would be too large to accommodate the listening subjects’ likely level of patience and attention span. In addition, it was possible that some levels of ΔAV or ΔF1 would not be as informative as others in identifying the potential perceptual category boundary between glides and high vowels. Therefore, a pilot experiment, described in the following section, was conducted to identify the best subset of AV and F1 contours to be used in the main experiment.

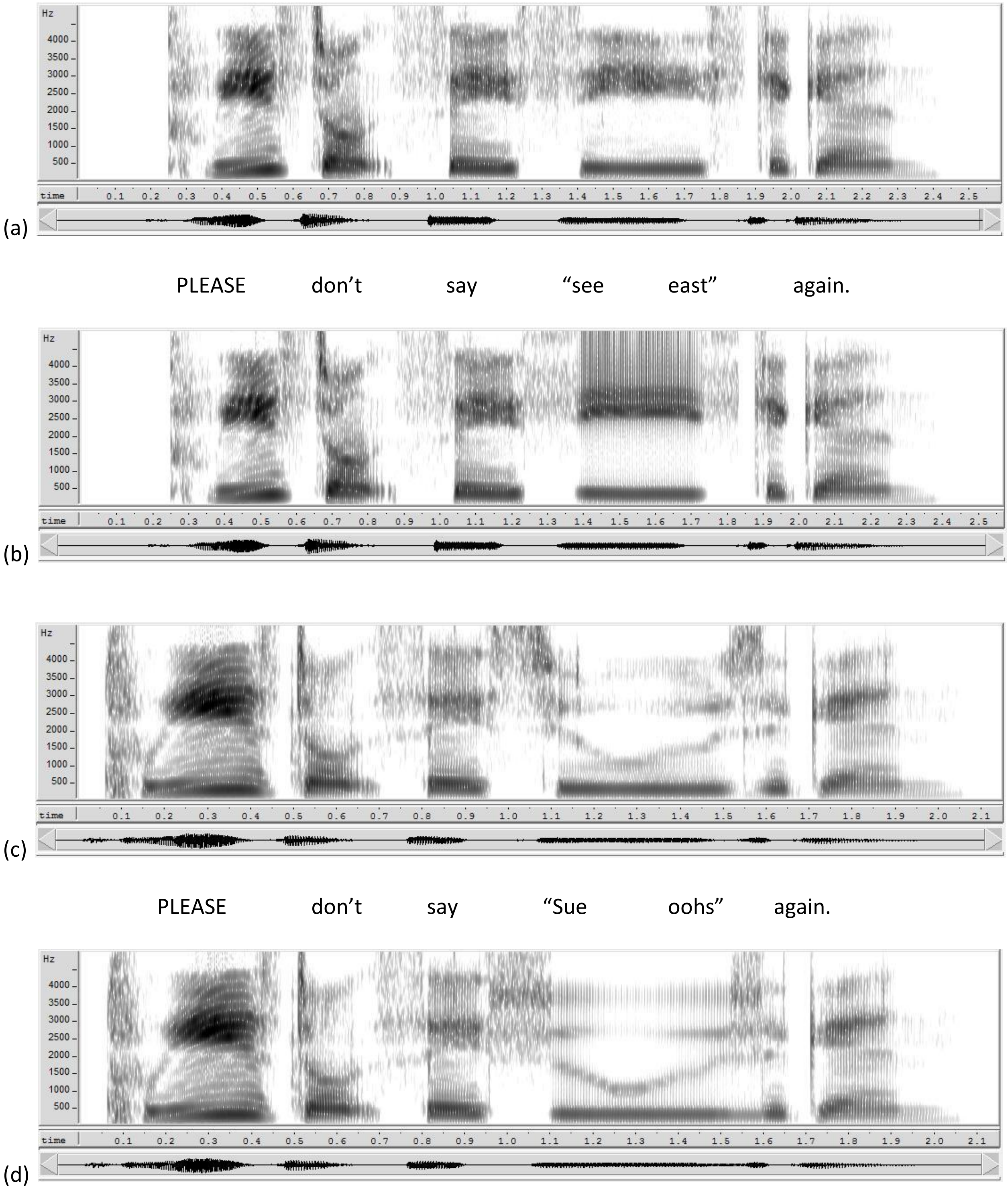


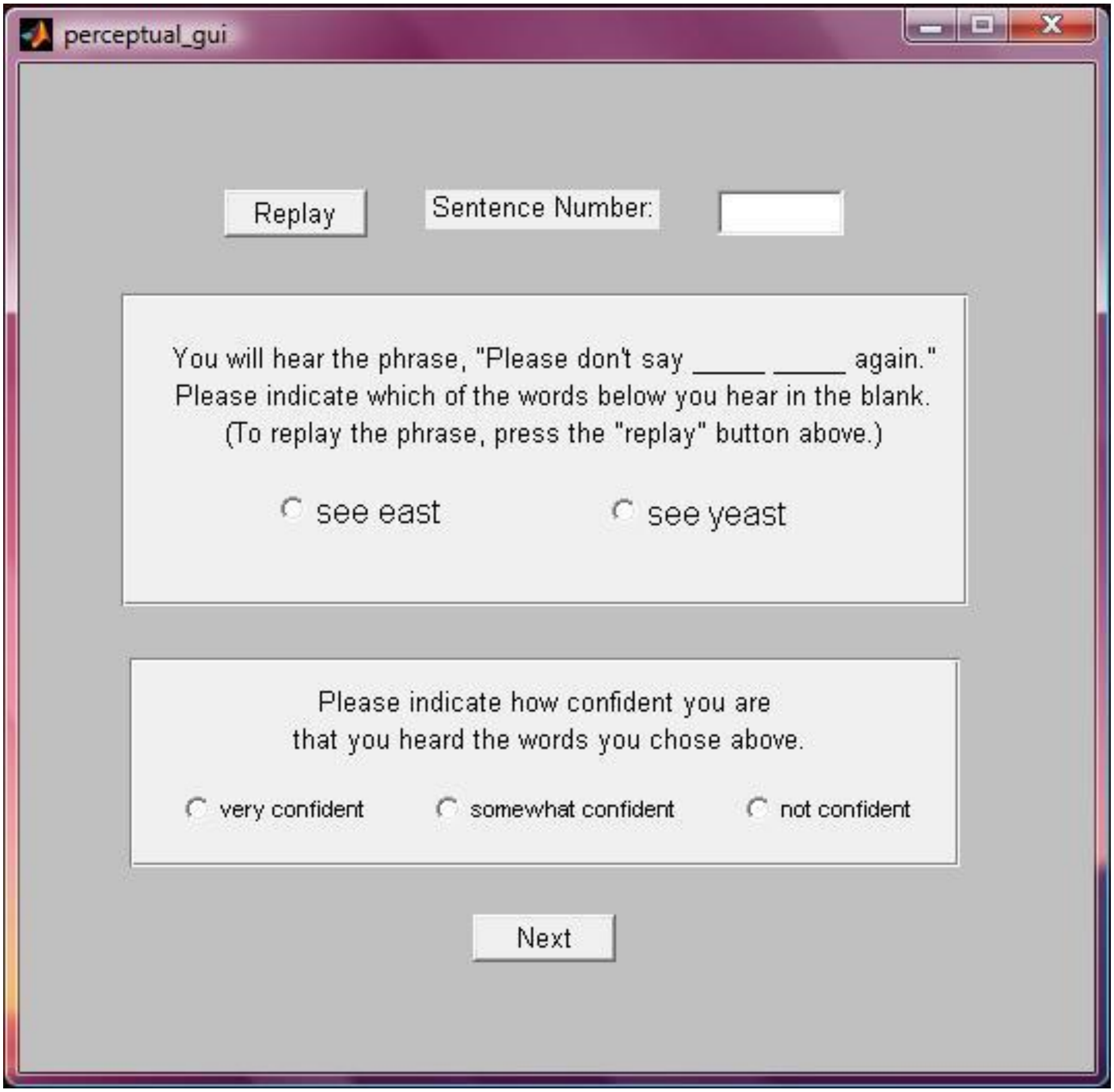
Figure 34: Spectrograms of example perceptual stimulus tokens in carrier phrases. (a) Naturally produced "PLEASE don't say 'see east' again." (b) Synthetic "see eas(t)" with flat AV and F1 contours, spliced into naturally produced carrier phrase from (a). (c) Naturally produced “PLEASE don’t say ‘Sue oohs’ again.” (d) Synthetic “Sue oohs” with flat AV and F1 contours, spliced into naturally produced carrier phrase from (c).

3.2 Pilot experiment

3.2.1 Method

All nine AV contours and all nine F1 contours for the synthetic “see east”/”see yeast” and “Sue oohs”/”Sue woos” continua were tested in the pilot perceptual experiment. However, in order to limit the time required of subjects and potential tiring and frustration, not all combinations of each of the AV and F1 contours were used together. Instead, a representative subset consisting of four continua was formed. One AV continuum combined the flat ΔF1 = 0 Hz contour with all nine AV contour levels, and another AV continuum combined the maximum-dip ΔF1 = 80 Hz contour with all nine AV contour levels. In addition, one F1 continuum combined the flat ΔAV = 0 dB contour with all nine F1 contour levels, and the other F1 continuum combined the maximum-dip ΔAV = 16 dB contour with all nine F1 contour levels. These four continua included a total of 32 contour combinations. Three repetitions of each combination were tested for each of the two minimal pairs (for /j/ and /w/), for a total of 32 contour combinations X 3 repetitions X 2 minimal pairs = 192 test tokens.

Figure 35 shows the MATLAB graphical user interface (GUI) created to conduct the perceptual experiments for this study. Stimulus tokens were presented to the subject through headphones, and the subject could replay the current stimulus token as many times as desired by pressing the “replay” button at the top of the GUI. In the first panel, the subject was asked to indicate whether they heard the member of the minimal pair containing no glide (“see east” or “Sue oohs”) or the member containing the glide (“see yeast” or “Sue woos”). In the second panel, the subject was asked to indicate their confidence that they had heard or not heard a glide, from one of three choices: “very confident”, “somewhat confident”, or “not confident”. The experiment moved on to a new stimulus token when the subject pressed the “next” button at the bottom of the GUI.



You will hear the phrase, "Please don't say _____ _____ again."
Please indicate which of the words below you hear in the blank.
(To replay the phrase, press the "replay" button above.)

○ see east ○ see yeast

Please indicate how confident you are
that you heard the words you chose above.

○ very confident ○ somewhat confident ○ not confident

Figure 35: MATLAB GUI used for perceptual experiments.

The GUI automatically recorded the subject's responses using a 6-point scale, as follows:

1 = glide absent, very confident

2 = glide absent, somewhat confident

3 = glide absent, not confident

4 = glide present, not confident

5 = glide present, somewhat confident

6 = glide present, very confident

The responses, as coded by this 6-point numeric scale, will hereafter be referred to as the subject's "glide ratings".

In order to make the instructions to the subjects and the GUI as clear and simple as possible, the /j/ and /w/ stimuli were presented in two separate blocks during the experiments. At the beginning of each block, an exposure phase (described to the subjects as a practice phase) presented the subjects with the maximum variation in AV and F1 contours that they would hear during the experiment. The exposure phase consisted of 10 tokens with combinations of ΔAV = 0 dB, ΔAV = 16 dB, ΔF1 = 0 Hz, and ΔF1 = 80 Hz, presented in random order. The test phase then began with 10 tokens whose responses were not included in the analysis of results; the exclusion of these first test tokens was intended to remove any learning curve effects from the study's results. The 96 true test tokens followed in random order, with the last 10 tokens the same as the initial 10 excluded learning-curve tokens. The total number of items in the pilot experiment was thus: (10 exposure tokens + 10 learning-curve tokens + 96 test tokens) x 2 minimal pair blocks = 232 total items.

3.2.2 Results

The pilot experiment was taken by two subjects (P1 and P2), both of whom were speech science researchers. For each of the 32 test stimuli, the subjects' responses on the 6-point scale were averaged

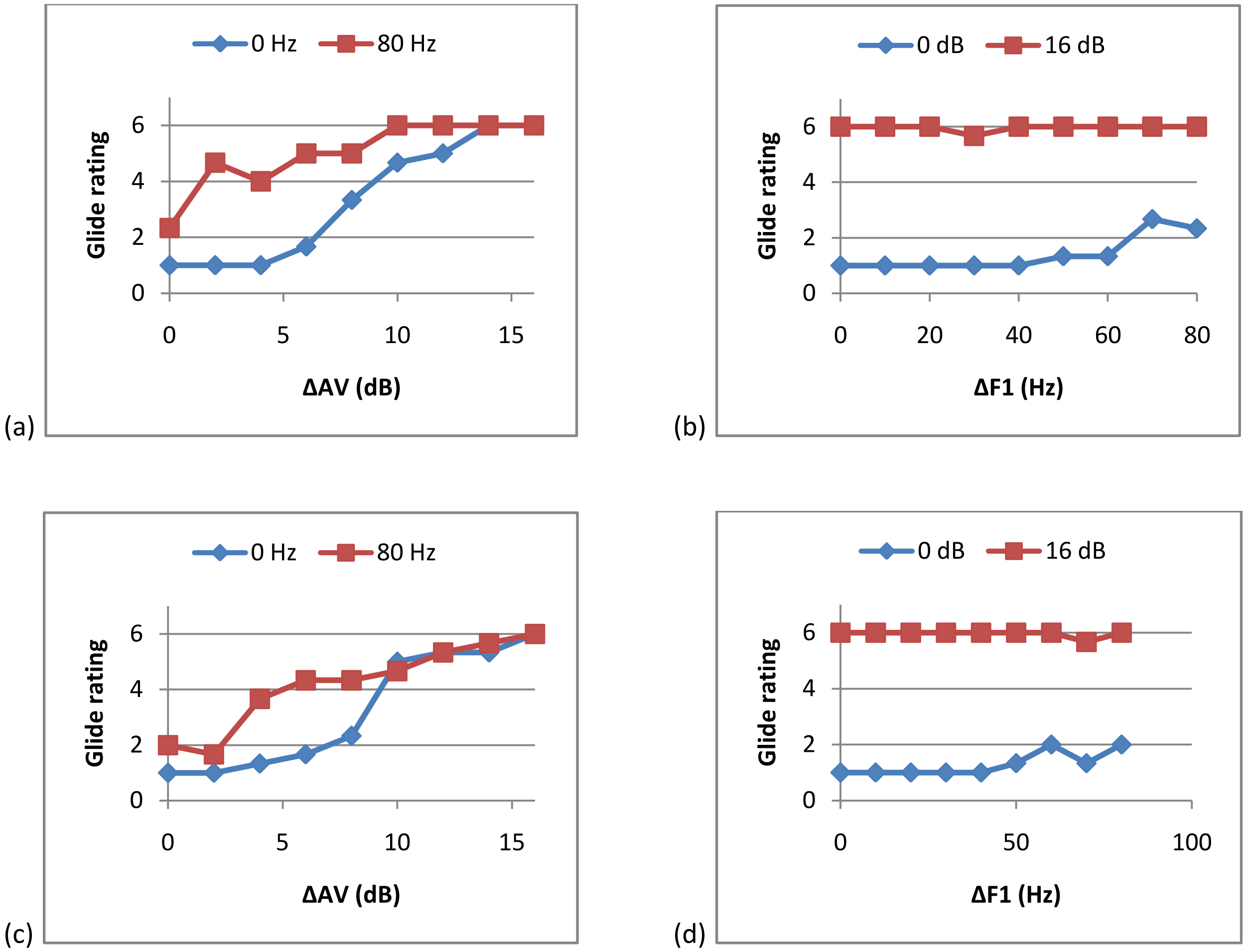


Figure 36: Average glide ratings from pilot perceptual subject P1. (a) Glide ratings for "see east"/"see yeast" as a function of ΔAV, for two values of ΔF1 (0 Hz and 80 Hz). (b) Glide ratings for "see east"/"see yeast" as a function of ΔF1, for two values of ΔAV (0 dB and 16 dB). (c) Glide ratings for "Sue oohs"/"Sue woos" as a function of ΔAV, for two values of ΔF1 (0 Hz and 80 Hz). (d) Glide ratings for "Sue oohs"/"Sue woos" as a function of ΔF1, for two values of ΔAV (0 dB and 16 dB).

across the three repetitions. Those averaged responses are plotted in Figure 36 for subject P1, and in Figure 37 for subject P2. In each figure, the top panels ((a) and (b)) plot the responses for "see east"/"see yeast", and the bottom panels ((c) and (d)) plot the responses for "Sue oohs"/"Sue woos". The left panels ((a) and (c)) plot the responses from the two AV continua, and the right panels ((b) and (d)) plot the responses from the two F1 continua.

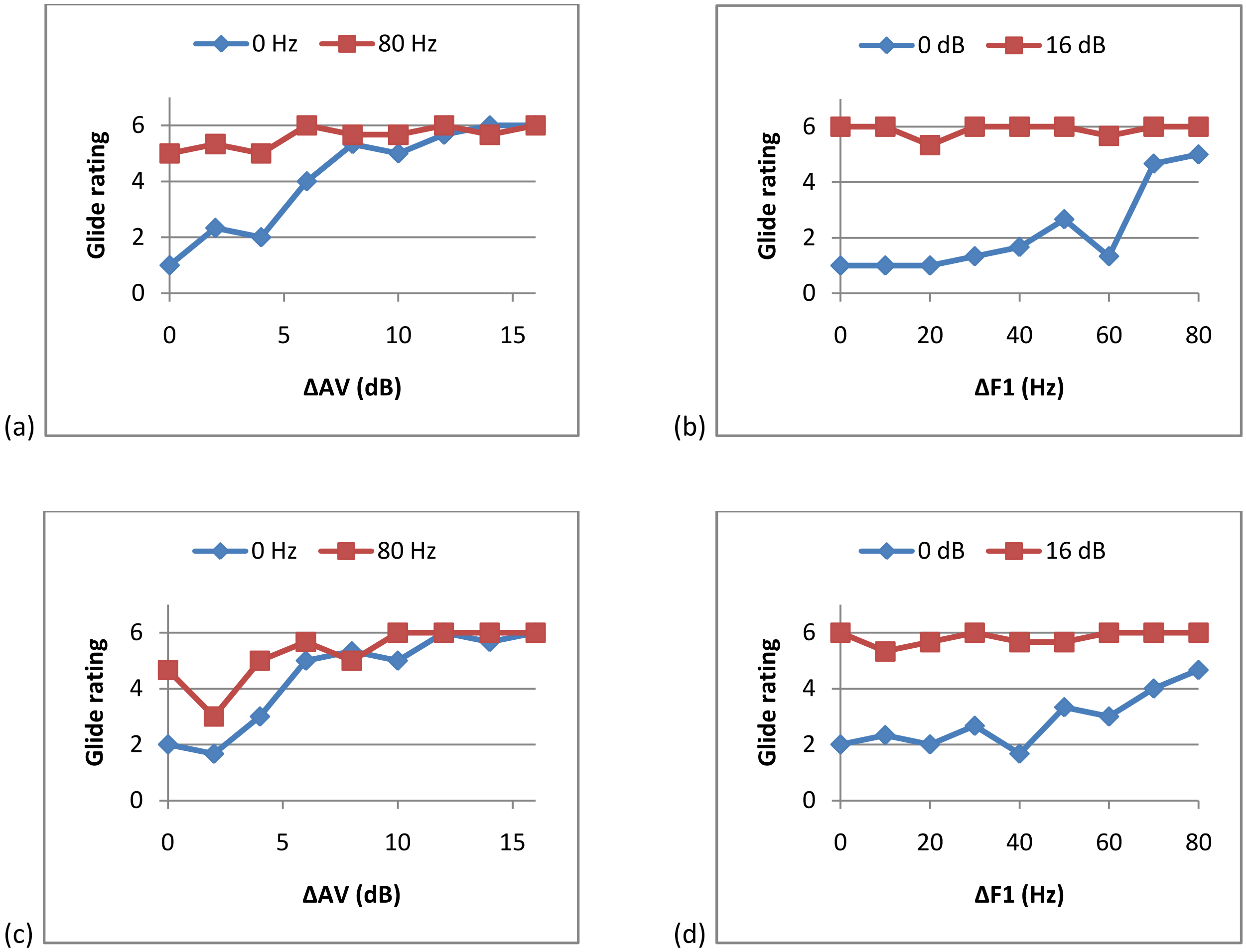


Figure 37: Average glide ratings from pilot perceptual subject P2. (a) Glide ratings for "see east"/"see yeast" as a function of ΔAV, for two values of ΔF1 (0 Hz and 80 Hz). (b) Glide ratings for "see east"/"see yeast" as a function of ΔF1, for two values of ΔAV (0 dB and 16 dB). (c) Glide ratings for "Sue oohs"/"Sue woos" as a function of ΔAV, for two values of ΔF1 (0 Hz and 80 Hz). (d) Glide ratings for "Sue oohs"/"Sue woos" as a function of ΔF1, for two values of ΔAV (0 dB and 16 dB).

From the responses plotted in Figure 36 and Figure 37, it was judged that most, if not all, of the ΔAV levels were of interest in testing the perception of glides and high vowels, since all of the ΔAV values elicited somewhat different glide ratings. However, it was decided that some of the ΔF1 values could be removed from the experiment, since the glide ratings hardly changed for some of the ΔF1 levels, especially those closer to ΔF1 = 0 Hz. Therefore for the main perceptual experiment, to be presented in the next section, all nine ΔAV levels were used in combination with five selected ΔF1 levels: ΔF1 = 0 Hz, 40 Hz, 60 Hz, 70 Hz, and 80 Hz.

The pilot subjects' responses to the synthetic stimuli also provide some preliminary information about the question of interest to this study, namely which of the acoustic parameters tested, AV or F1, is most important to the perceptual distinction of glides from high vowels. The response curves for the AV continua show that the dependence of the glide/vowel percept on the size of the amplitude dip is strong, whether the F1 contour is flat or maximally perturbed. In general, the response curves for the AV continua start at low average glide ratings (indicating that the subject did not hear a glide) at small values of ΔAV, and rise to high average glide ratings (indicating that the subject did hear a glide) at large values of ΔAV. Between the flat F1 contour (ΔF1 = 0 Hz) and the maximally perturbed F1 contour (ΔF1 = 80 Hz), the entire response curve is shifted toward higher glide ratings, but the rightward rising shape of the curve does not change much.

On the other hand, the response curves for the F1 continua appear much flatter than those for the AV continua. This is especially true for subject P1, for whom the response curve with ΔAV = 0 dB never crosses higher than an average glide rating of 3, indicating that this subject heard no glide if an amplitude dip was not present, regardless of the size of the F1 dip. For both subjects, the response curve for the F1 continuum with ΔAV = 16 dB is almost completely saturated at a glide rating of 6, indicating that both subjects heard glides when the amplitude dip was of maximum size, regardless of the size of the F1 dip.

The relative lack of dependence of the glide ratings on ΔF1, compared to the strong and consistent dependence on ΔAV, provides preliminary evidence that the amplitude (source) cue is more important to these listeners than the F1 (filter) cue in distinguishing glides from high vowels. The fact that the stimulus tokens with ΔAV = 16 dB and ΔF1 = 0 Hz were given an average glide rating of 6 by both subjects indicates that the amplitude dip is sufficient to cue the presence of a glide in the absence of any perturbation to F1. By contrast, the stimulus token with ΔAV = 0 dB and ΔF1 = 80 Hz was heard (unconfidently) as containing a glide only by subject P2; the maximum F1 dip was not sufficient to cue the presence of a glide for subject P1 in the absence of source amplitude perturbation. It should be cautioned, however, that both of the subjects for this pilot experiment were phonetically trained researchers, and their responses might not accurately reflect those of the untrained general population. The behavior of a larger sample of naïve and untrained listeners was investigated in the main experiment, presented in the following section.

3.3 Main experiment

3.3.1 Method

The main perceptual experiment used the stimulus tokens synthesized as described in Section 3.1, forming continua including nine levels of ΔAV (0 dB, 2 dB, 4 dB, 6 dB, 8 dB, 10 dB, 12 dB, 14 dB, 16 dB) and five levels of ΔF1 (0 Hz, 40 Hz, 60 Hz, 70 Hz, 80 Hz). All combinations of each AV contour with each F1 contour were generated, for a total of 9 X 5 = 45 test stimuli for each minimal pair (“see east”/”see yeast” and “Sue oohs”/”Sue woos”). The experimental procedure was identical to that of the pilot experiment, described in Section 3.2, with separate blocks for the /j/ and /w/ minimal pairs. As in the pilot experiment, each block was preceded by a ten-token exposure phase and ten learning-curve tokens that were not used in the analysis. Each test stimulus was repeated three times within the random order of the test block, for a total of (10 exposure tokens + 10 learning-curve tokens + 135 test tokens) x 2 minimal pair blocks = 310 total items.

The main experiment was completed by ten naïve subjects drawn from the MIT community, all of whom were paid for their participation. Since the experimental interface allowed for each test item to be replayed as many times as the subjects desired before entering their response for that item, the total length of time for completion of the experiment varied widely between subjects, and they were compensated accordingly. Each subject's responses were recorded automatically by the MATLAB GUI on the 6-point glide rating scale, and their ratings were averaged across the three repetitions of each test stimulus.

3.3.2 Results

The majority of the subjects who participated in this experiment were able to follow the instructions without confusion and interpret the synthetic stimuli as the /V#V/ or /V#GV/ target sequences they were intended to represent. A few subjects, however, gave some indication that their expectation of glottalization on the word-initial vowel in "see east" or "Sue oohs" unduly influenced their performance in the experiment. Judging from feedback spontaneously provided by these few subjects, they were reluctant to accept /V#V/ as a possible sequence with the vowel combinations used here, preferring instead a realization closer to /V#ʔV/.[2] This confusion was not anticipated in the planning of this experiment, since glottalization is in all cases optional in American English, and has been found to be significantly less prevalent in words that are not pitch accented and not at the beginning of an intonational phrase (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996). However, it seems that the expectations of individual listeners may vary with respect to /V#V/-conditioned glottalization, and the stronger expectations of a few subjects produced anomalous results for this study, which will be discussed below. Since the glottalization expectation seemed to operate differently for a couple of

[2] The use of the symbol /ʔ/ may not strictly be appropriate to represent the glottalization of a word-initial vowel in American English, since it is an optional allophonic variation rather than a contrastive phoneme in this language. Although such word-initial glottalization is often referred to as the insertion of a "glottal stop", its articulatory production is rarely completely occlusive as in a full stop consonant (Pierrehumbert & Talkin, 1992). However, similar productions and acoustics to American English glottalization have been attested cross-linguistically for /ʔ/, including in languages in which it is phonemically contrastive. Thus, in the interest of brevity and generalizability, the symbol /ʔ/ will be used here to refer to glottalization of word-initial vowels.

subjects between the /j/ and /w/ minimal pairs, the results from the two blocks will be presented separately in this section.

For the “see east”/”see yeast” minimal pair, two subjects (subjects S4 and S6) were excluded from the analysis of the results, due to the apparent randomness of their responses. These two subjects completed the experiment in far less time than the other subjects required, and it is possible that they were more interested in being paid than in providing carefully considered responses. In addition, one of the two subjects (S4) reported after the experiment that he “couldn’t tell the difference between a glottal stop and a glide,” and that he “tended to pick whichever choice [he] was already looking at” on the GUI when the stimulus was played. (Such feedback was not requested from any subject, but many felt the desire to offer spontaneous feedback after completing the experiment.) Apparently, this subject’s confusion at the lack of glottalization cues was so great that he could not make decisions based on the other acoustic cues available, and instead submitted responses based on some form of visual hysteresis. Alternatively, it is possible that S4 heard the ΔAV and/or ΔF1 cues interchangeably as cues to glides or glottalization, and therefore had trouble choosing between the /V#GV/ and /V#ʔV/ tokens he believed he heard. Further discussion of whether similar acoustic cues might sometimes be shared by glides and glottalization is provided later in this section, and in Section 3.3.3.

Since the two subjects mentioned above provided responses to this experiment that were not consistently dependent on either ΔAV or ΔF1, whether from lack of interest, expectations of glottalization, or other reasons, their results were considered uninformative on the subject of ranking amplitude and F1 cues to glidehood, and were therefore excluded from further analysis. The results from the remaining eight subjects, averaged together for the “see east”/”see yeast” minimal pair, are plotted in Figure 38. Panel (a) shows the average glide ratings, on the 6-point scale, plotted with respect to ΔAV; each separate curve represents a constant ΔF1 value. Note that the leftmost point of all five

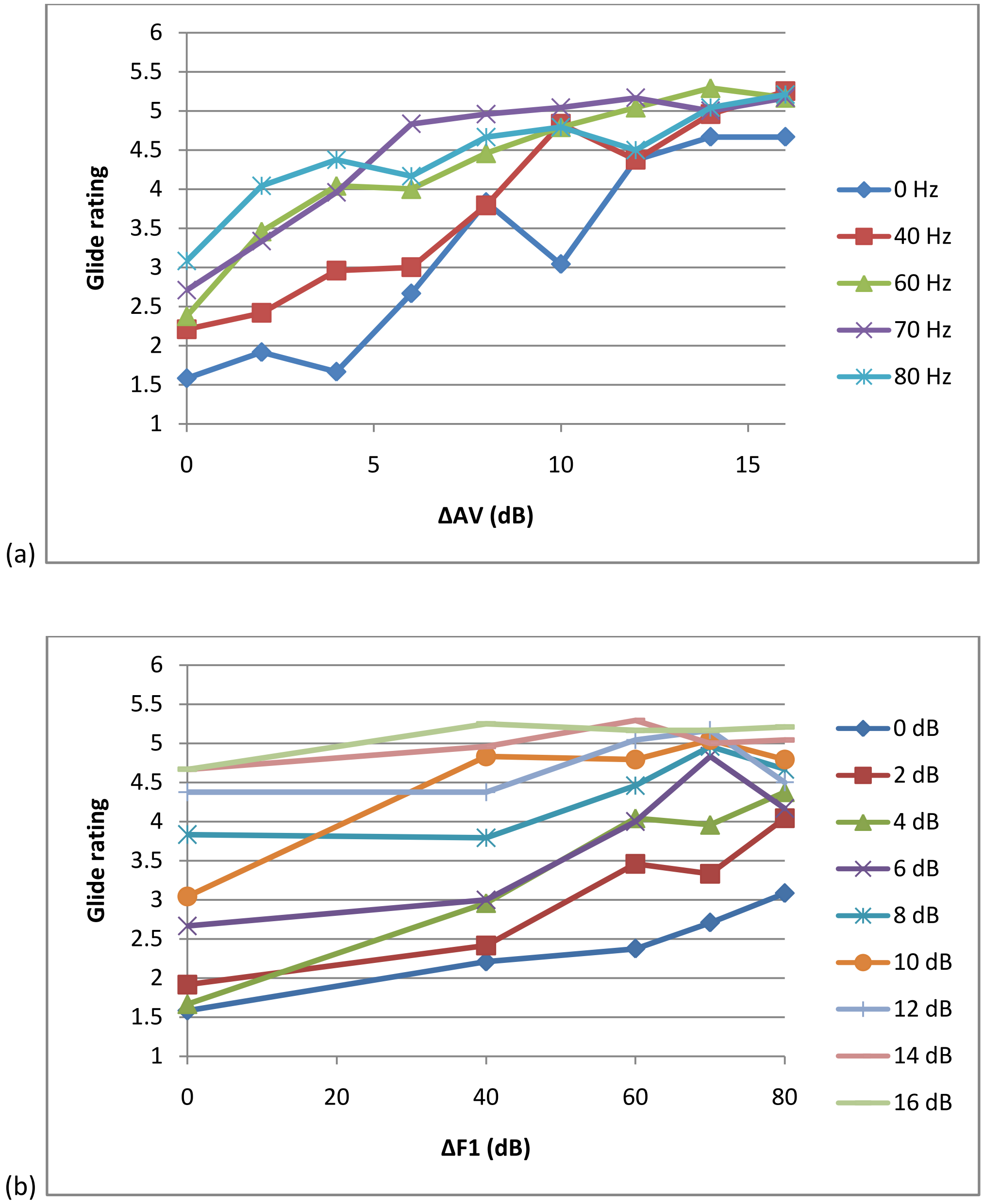


Figure 38: Average glide ratings across eight subjects for the "see east"/"see yeast" minimal pair. (a) Average glide ratings as a function of ΔAV; ΔF1 is constant at one of five levels for each colored curve. (b) Average glide ratings as a function of ΔF1; ΔAV is constant at one of nine levels for each colored curve.

curves, corresponding to ΔAV = 0 dB, is below the glide rating midpoint of 3.5, indicating that tokens with flat amplitude contours were perceived as containing no glide, regardless of the size of the F1 dip. The rightmost point of all five curves, corresponding to ΔAV = 16 dB, is above the glide rating midpoint of 3.5, indicating that tokens with maximum amplitude dip were perceived as glides, again regardless of the size of the F1 dip. By contrast, the dependence of the glide ratings on ΔF1, plotted in panel (b), is appreciably more flat. Note that the endpoints of the curves, corresponding to ΔF1 = 0 Hz and ΔF1 = 80 Hz, fall on either side of the glide rating midpoint of 3.5, in order of their ΔAV.

The strong positive slope of the ΔAV continua responses in Figure 38(a), compared with the relatively flatter ΔF1 continua responses in Figure 38(b), indicate that the perceptual detection of the /j/ in “see yeast” is more dependent on its amplitude contour than on its F1 contour. This perceptual judgment is not completely independent of F1, however, since the average glide ratings do increase somewhat for large values of ΔF1, especially when ΔAV is small. These observations were confirmed through a three-factor analysis of variance, with subject, ΔAV, and ΔF1 as factors (the three-way interaction was not significant, and was removed from the analysis). The main effects of both ΔAV and ΔF1 were significant (F(8,224)=72.994, p=.000 for ΔAV; F(4,224)=47.248, p=.000 for ΔF1); however, the significance level of the ΔAV effect is much higher than that of the ΔF1 effect. This supports the observation that the average glide ratings are more dependent on ΔAV than on ΔF1 for /j/. The interaction between the ΔAV and ΔF1 factors was also significant (F(32,224)=2.984, p=.000), as is clear from the differences in shape and slope between the curves in each individual panel. Also significant were the main effect of subject (F(7,224)=11.343, p=.000) and the interaction between subject and ΔAV (F(56,224)=2.059, p=.000); the interaction between subject and ΔF1 was not significant. The subject*ΔAV interaction reflects the fact that the slope of the ΔAV continua responses differs for each subject; all of the eight subjects’ response curves have positive slopes, however.

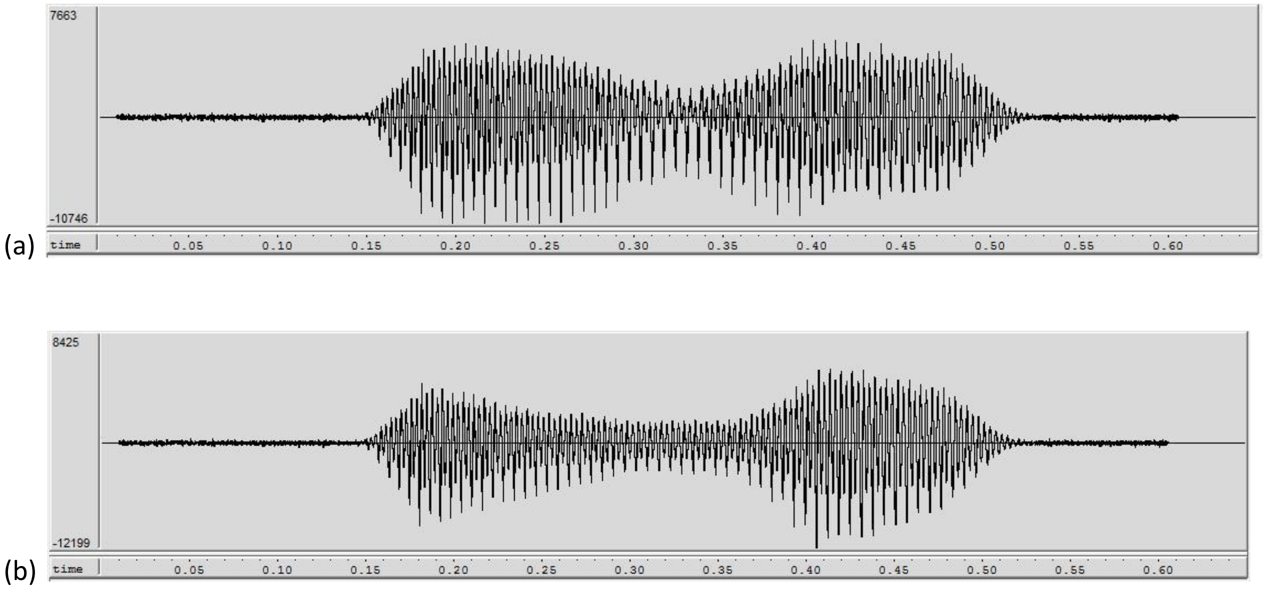


Figure 39: Waveforms of two synthetic "see east"/"see yeast" tokens that were given similar average glide ratings in the main perceptual experiment. (a) ΔAV = 0 dB, ΔF1 = 80 Hz. (b) ΔAV = 10 dB, ΔF1 = 0 Hz.

There is some evidence of a trading relation between ΔAV and ΔF1 in the "see east"/"see yeast" perceptual distinction shown in Figure 38. That is, if a certain glide rating can be achieved through a certain value of ΔAV, that same glide rating might be elicited through some smaller value of ΔAV together with some larger value of ΔF1. For instance, a similar average glide rating of about 3 is given to the token with ΔAV = 10 dB and ΔF1 = 0 Hz, and also to the token with ΔAV = 0 dB and ΔF1 = 80 Hz. However, it is possible that some of the effect of ΔF1 is not due to perception of the decrease in the actual formant frequency, but to perception of its concomitant decrease in spectral amplitude. (Since the KLSYN cascade formant synthesizer mimics the natural production of speech, any decrease in the F1 parameter is accompanied by the amplitude decrease warranted by the acoustics of the typical vocal tract.) Figure 39 shows the waveforms of the two tokens that both elicited an average glide rating of about 3, the one in panel (a) through an 80 Hz dip in F1 with flat AV, and the one in panel (b) through a 10 dB dip in AV with flat F1. The amplitude effect of the F1 dip in panel (a) is clear, although it is not as large as the amplitude dip in panel (b) that elicited the equivalent percept. Probably the effect of ΔF1 on the glide ratings is significant in its own right as a spectral frequency change, but the quantitative dependence of the glide ratings on ΔF1 may in fact be overestimated due to the fact that F1 movement partially contributes to the amplitude effect.

The relative dependence of the glide ratings on ΔF1 seems to be if anything even smaller in the "Sue oohs"/"Sue woos" minimal pair, whose results are plotted in Figure 40. These graphs represent the average of the responses from six subjects, rather than the eight subjects that were included in the "see east"/"see yeast" analysis. For "Sue oohs"/"Sue woos", the same two subjects (S4 and S6) were excluded as for "see east"/"see yeast", for the same reasons described above. In addition, one more subject (S3) provided responses that did not depend on either ΔAV or ΔF1 for "Sue oohs"/"Sue woos", again apparently due to confusion about the possible requirement of glottalization in "Sue oohs". This subject spontaneously reported that he would have chosen "Sue oohs" if he had "heard a break between [the two words]". Since no glottal stop was included in the synthesis, S3 therefore almost always chose "Sue woos", regardless of the other acoustic cues presented (his average glide rating was about 4 for all stimuli). Since S3 did attend to ΔAV along with the majority of the other subjects for the /j/ minimal pair, it is unclear what aspect of the /w/ minimal pair made him change his paradigm. Possibly the difference in behavior was a result of durational differences between the two sets from the copy-synthesis procedure. Since S3's responses were uninformative in terms of comparing the effects of ΔAV and ΔF1, they were not included in the analysis below for "Sue oohs"/"Sue woos".

One further subject (S7) was excluded from the main analysis for "Sue oohs"/"Sue woos" because, although he did attend to the acoustic cues provided, his response trends were not in the same direction as the majority of the subjects. S7's anomalous responses demonstrate a very consistent and intriguing pattern, providing further information about the perception of glottalization, and as such will

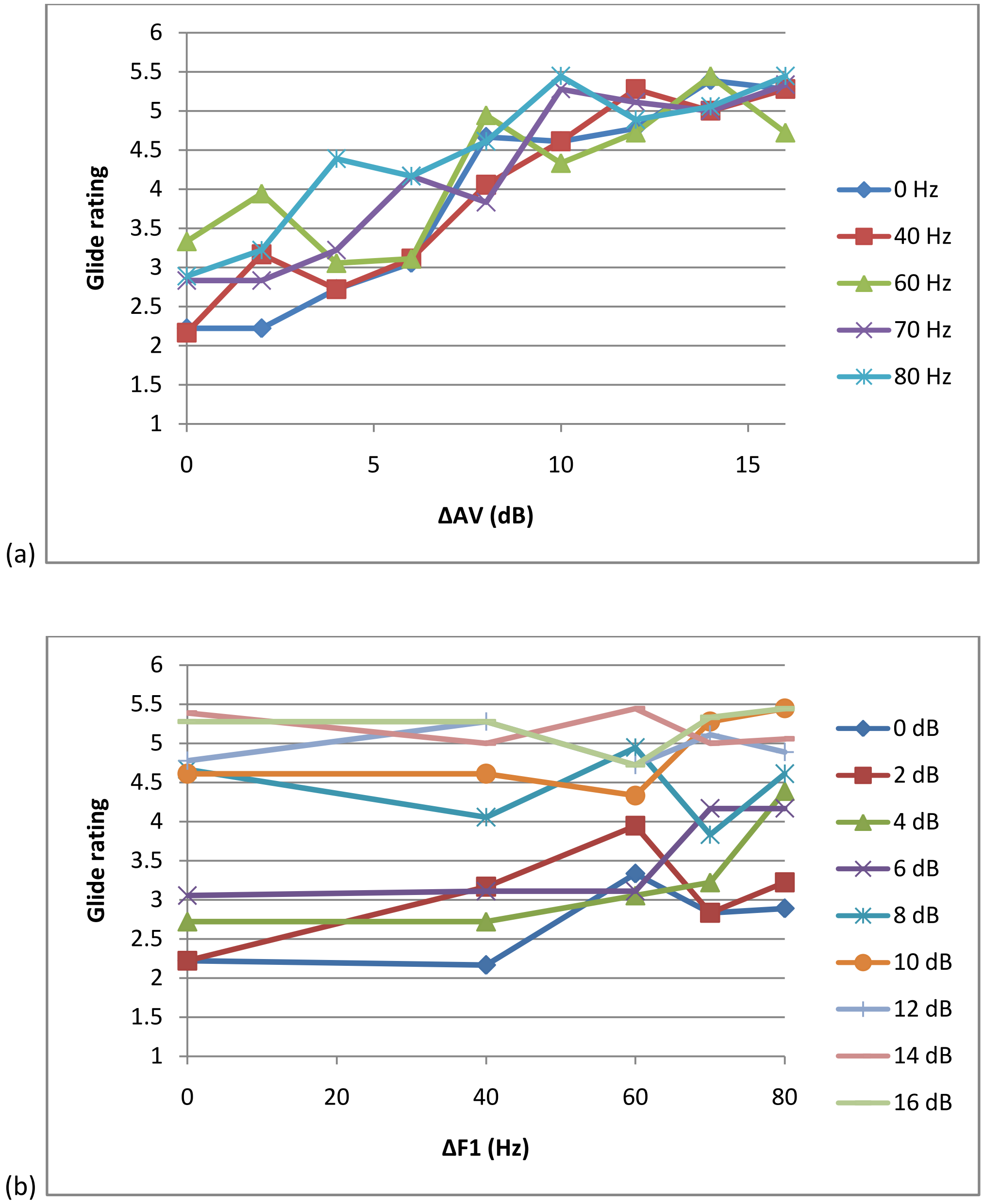


Figure 40: Average glide ratings across six subjects for the "Sue oohs"/"Sue woos" minimal pair.
(a) Average glide ratings as a function of ΔAV; ΔF1 is constant at one of five levels for each colored curve.
(b) Average glide ratings as a function of ΔF1; ΔAV is constant at one of nine levels for each colored curve.

be presented in detail below. However, the more conventional results from the remaining six subjects are presented first.

It can be seen in Figure 40 that the average response curves for “Sue oohs”/”Sue woos” are similar in shape to those for “see east”/”see yeast”. The dependence of the glide ratings on ΔAV is strong, with all ΔAV = 0 dB tokens clustering below the glide rating midpoint of 3.5 (indicating that no glide was heard when there was no dip in amplitude), and all ΔAV = 16 dB tokens clustering above the glide rating midpoint of 3.5 (indicating that a glide was heard when the amplitude dip was of maximum size), regardless of the value of ΔF1. The ΔAV continua curves in panel (a) are less spread from each other than in Figure 38, and the ΔF1 continua curves in panel (b) are less upwardly sloped, indicating that the listeners’ dependence on ΔF1 is even smaller for the perception of /w/ than it was for /j/. An analysis of variance confirms this observation, with subject, ΔAV, and ΔF1 as factors (again, the three-way interaction was not significant). The main effects of both ΔAV and ΔF1 were significant (F(8,160)=57.186, p=.000 for ΔAV; F(4,160)=5.494, p=.000 for ΔF1); however, the significance level of the ΔAV effect is much higher than that of the ΔF1 effect. The interaction between the ΔAV and ΔF1 factors was also significant (F(32,160)=2.307, p=.000), as were the main effect of subject (F(5,160)=23.160, p=.000) and the interaction between subject and ΔAV (F(40,160)=2.011, p=.001); the interaction between subject and ΔF1 was not significant. The subject*ΔAV interaction reflects the fact that the slope of the ΔAV continua responses differs for each subject; all of the six subjects’ response curves have positive slopes, however.

A brief digression is now in order to address the anomalous but intriguing responses of subject S7 to the main perceptual experiment. An ANOVA performed on S7’s average glide responses without interactions showed that his responses were more dependent on the ΔAV parameter than the ΔF1 parameter, in similar fashion to the subjects included in the main analysis. For the /j/ minimal pair, S7’s main effect of ΔAV was more significant than his main effect of ΔF1 (F(8,32)=8.988, p=.000 for ΔAV; F(4,32)=4.505, p=.005 for ΔF1). For the /j/ minimal pair, S7's main effect of ΔAV was significant, and his main effect of ΔF1 was not (F(8,32)=8.116, p=.000 for ΔAV; F(4,32)=1.583, p=.203 for ΔF1). However, the dependence of S7's glide ratings on ΔAV differs from that of the other subjects in that it is in the opposite direction for /w/ than for /j/, as can be seen in Figure 41.

In panel (a) of Figure 41, it can be seen that S7's response curves for the ΔAV continua of "see east"/"see yeast" slope upward to the right, indicating that he heard larger amplitude dips as cues to the presence of the glide /j/, as did the other subjects. By contrast, S7's response curves for the ΔAV continua of "Sue oohs"/"Sue woos" in panel (b) of Figure 41 slope downward to the right, indicating that he head larger amplitude dips as cues to the glide-absent "Sue oohs" rather than the glide-present "Sue woos". In fact, it is clear from unsolicited feedback given by S7 after the experiment that he heard the amplitude dips as cues for /ʔ/ rather than /w/ in the "Sue oohs"/"Sue woos" block, prompting him to choose "Sue oohs" with confidence when the amplitude dip was large. His comments, sent in an e-mail after the experiment, included the following:

> *"To my ear, the words "see yeast" were sometimes clearly distinguishable when the two words were joined by a sufficiently closed and prolonged "y" sound, but the words "see east" were always somewhat ambiguous due to the lack of a clear space and glottal attack. In contrast, the words "Sue oohs" were sometimes clearly distinguishable when separated by a sufficient space and glottal attack, but the words "Sue woos" were always somewhat ambiguous due to the lack of a sufficiently closed and prolonged "w" sound."*

S7's comments, combined with the trends in his glide ratings, clearly indicate that he heard the amplitude dips as cues to the glide /j/ in the "see east"/"see yeast" pair, but as cues to the glottalized /ʔ/ in the "Sue oohs"/"Sue woos" pair. That the same acoustic parameter was interpreted in opposite ways between the two blocks may be explained by durational relationships, as implied by S7's comments' emphasis on the "prolonged" nature of the potential glide sound. Recall from Section 3.1

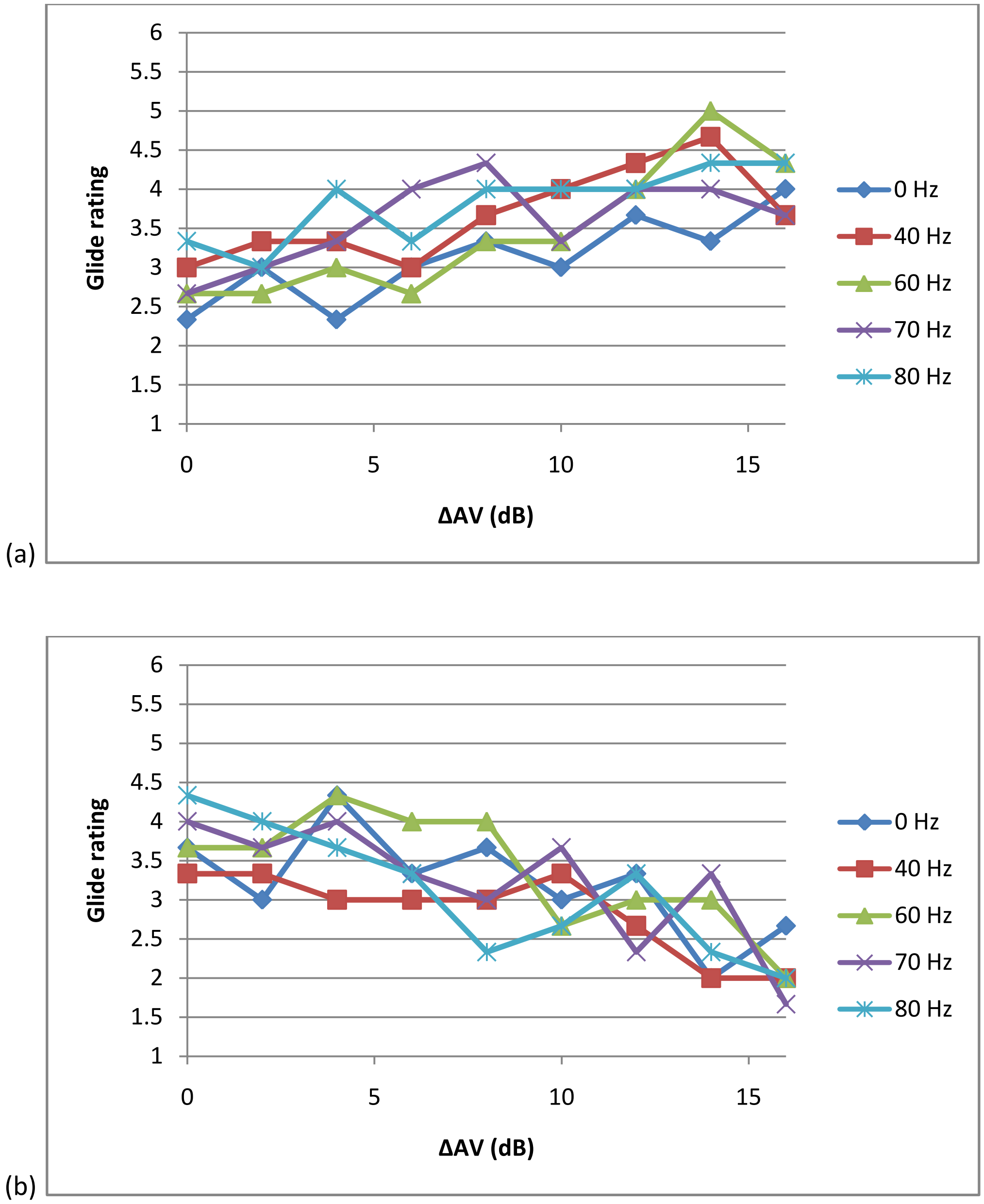


Figure 41: Average glide ratings for subject S7 plotted as a function of ΔAV; ΔF1 is constant at one of five levels for each colored curve. While the pattern of this subject's responses for detection of the glide/j/ agrees with the majority of the other subjects, his response pattern for detection of the glide /w/ differs due to perceptions of glottalization (see text). (a) Average glide ratings for the "see east"/"see yeast" minimal pair. (b) Average glide ratings for the "Sue oohs"/"Sue woos" minimal pair.

that the synthetic “see yeast” and “Sue woos” were created using different naturally produced tokens as guides for the placement and durations of the amplitude dips, therefore they differed between the two glides /j/ and /w/. By comparing Figure 32(a) on page 91 with Figure 33(a) on page 92, it can be seen that the minimum amplitude was held longer in the /j/ tokens than in the /w/ tokens. It seems that the longer amplitude minimum may have caused the percept of the glide /j/ in “see yeast” for S7, while the shortening of the amplitude minimum caused his percept to change to /ʔ/ in “Sue oohs”.

S7’s use of the intervocalic amplitude reduction as an acoustic cue alternatively to the presence of a glide or glottalization, possibly depending on its duration, brings up an interesting acoustic and possibly productional similarity between these two types of sounds. In a perceptual experiment with copy-synthesized speech very similar in procedure to the experiment reported here, Hillenbrand & Houde (1996) found that a reduction in amplitude does cue the presence of intervocalic /ʔ/ for multiple listeners. Although this type of glottalization realized with continuous voicing (i.e., lack of complete stop-like occlusion) is best approximated with reductions in both amplitude and fundamental frequency (F0), their study found that the amplitude reduction by itself was often sufficient to elicit the percept of /ʔ/ in sequences like /o#ʔo/ vs. /o#o/. Hillenbrand & Houde’s findings provide a plausible explanation for the anomalous responses of subject S7 in the current study: Glottal source amplitude reduction is a shared acoustic characteristic of both glides and /ʔ/. /ʔ/ is usually cued by a concomitant reduction in F0, which was not present to a large degree in this study of glides (ΔF0 for Hillenbrand & Houde’s /ʔ/ was on the order of 50 Hz, while ΔF0 for glides here was closer to 10 Hz). The fact that ΔAV was not accompanied by a sufficiently large ΔF0 may explain why most subjects, unlike S7, did not hear ΔAV consistently as a cue to /ʔ/ in the current study on glides.

Hillenbrand & Houde's (1996) finding that amplitude reduction was sufficient to cue the presence of /ʔ/ in the absence of F0 reduction was called into question, however, by the findings of a similar contemporary study by Pierrehumbert & Frisch (1997). Pierrehumbert & Frisch found that synthetic /ʔ/ could be cued by F0 reduction in the absence of other cues, but that amplitude reduction by itself was not sufficient to elicit the percept. Hillenbrand & Houde guessed that durational differences may have brought about the discrepancy between the two studies; however, the results of the current study suggest that the discrepancy may actually arise from the segmental contexts that were chosen to flank the target sounds in the two studies on /ʔ/. Pierrehumbert & Frisch (1997) synthesized /ʔ/ in word pairs such as "heavy oak"/"heavy yoke"; i.e., they compared sequences of /V#ʔV/ to /V#GV/. Since the current study has shown that source amplitude reduction is a very consistent acoustic characteristic and perceptually salient cue for glides, it is plausible that /ʔ/ would be indistinguishable from /j/ on the basis of the amplitude reduction alone. However, amplitude reduction may still be a sufficient cue to distinguish /ʔ/ from surrounding vowels. Although a dip in amplitude would not be expected to distinguish /V#ʔV/ from /V#GV/, since both /ʔ/ and /G/ would cause amplitude reduction, it could distinguish /V#ʔV/ from /V#V/, since no amplitude reduction would be present in the latter sequence. This is precisely the type of minimal pair that was constructed by Hillenbrand & Houde (1996), whose synthetic /o#ʔo/ and /o#o/ could be distinguished by ΔAV in the absence of any ΔF0.

Thus, current evidence seems to favor a shared acoustic cue between glides and glottalization, i.e., amplitude reduction of the glottal source. Although experiments with synthetic speech may allow ΔAV to operate in isolation from other cues, thus possibly producing the type of confusion between glides and /ʔ/ demonstrated by S7 and a couple other subjects in this study, it is likely that natural speech tokens would exhibit other cues, such as changes in F0 and formant frequencies, that could serve to disambiguate between the two types of sounds. However, the acoustic similarities between glides and laryngeal consonants such as /ʔ/, as well as the similarities in production that these acoustic congruencies may imply, may provide a compelling reason to consider both categories and their relationship when choosing a possible distinctive feature class for glides. This will be discussed further in later sections of this thesis.

3.3.3 Discussion

The main perceptual experiment results from Figure 38 on page 104 and Figure 40 on page 108 are reproduced in Figure 42, with only the continua on the edges displayed. That is, the plots are reduced to the AV continua with ΔF1 at its minimum and maximum values, and vice versa. The fact that the AV continua in panels (a) and (c) slope strongly upward to the right, compared with the flatter F1 continua in panels (b) and (d), indicates that the ΔAV cues are more important to the perceptual categorization of glides and high vowels than the ΔF1 cues. This ranking was confirmed through analyses of variance. The main effect of ΔAV was highly significant for both of the glides /j/ and /w/. By contrast, the main effect of ΔF1 was not significant for /w/, and was less significant than the ΔAV effect for /j/.

In particular, the right end of the AV continuum with ΔF1 = 0 Hz in panels (a) and (c) of Figure 42 lies well above the glide rating midpoint of 3.5, indicating that amplitude reduction is sufficient to cue the presence of a glide in the absence of any ΔF1 cue. By contrast, the right end of the F1 continuum with ΔAV = 0 dB in panels (b) and (d) of Figure 42 lies below the glide rating midpoint of 3.5, indicating that F1 reduction is not sufficient to cue the presence of a glide in the absence of amplitude reduction. These results rank amplitude cues as primary over F1 cues in the perceptual detection of glides. This does not,

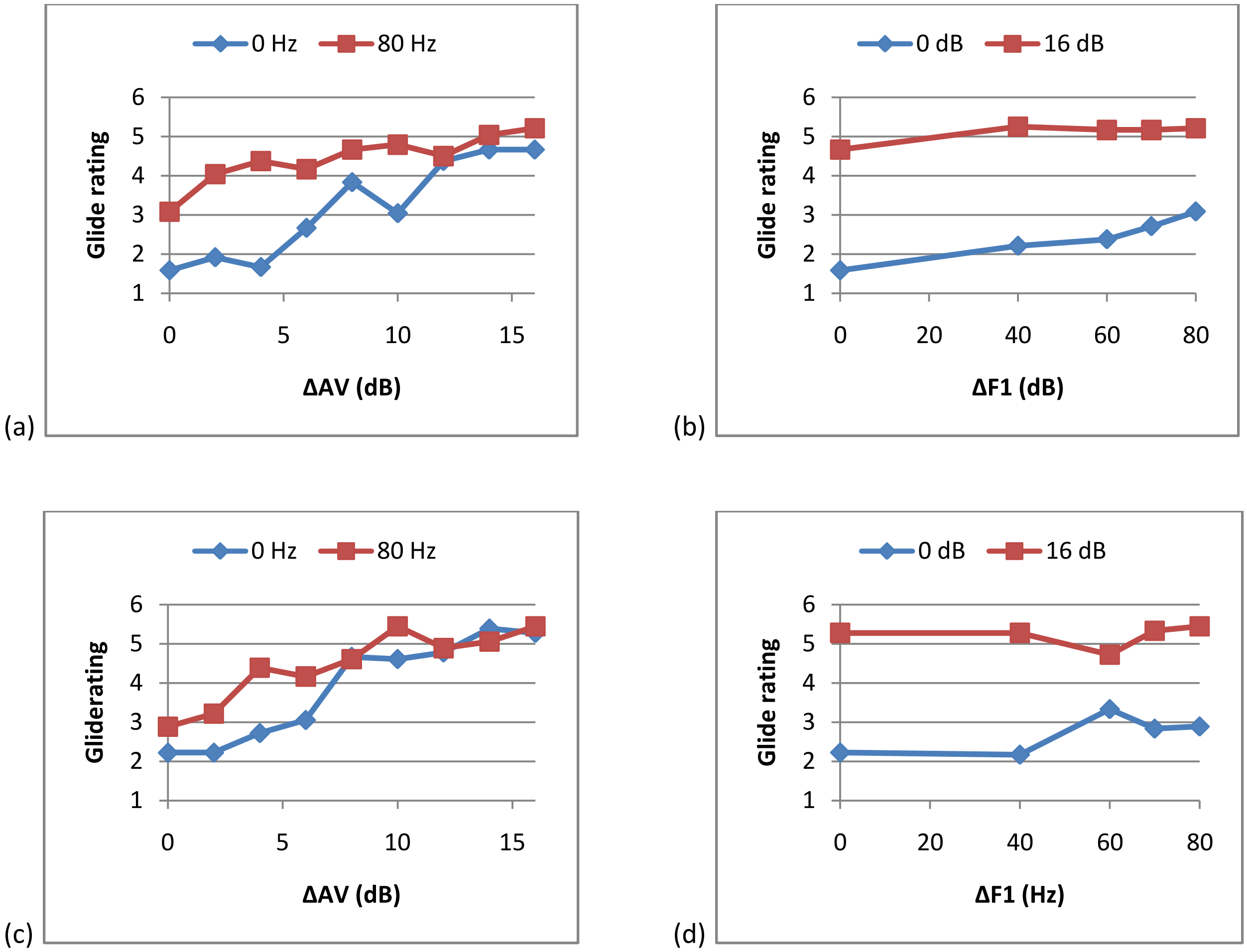


Figure 42: Glide ratings averaged across subjects included in analysis of main perceptual experiment.
(a) Average glide ratings for "see east"/"see yeast" as a function of ΔAV, for two values of ΔF1 (0 Hz and 80 Hz).
(b) Average glide ratings for "see east"/"see yeast" as a function of ΔF1, for two values of ΔAV (0 dB and 16 dB).
(c) Average glide ratings for "Sue oohs"/"Sue woos" as a function of ΔAV, for two values of ΔF1 (0 Hz and 80 Hz).
(d) Average glide ratings for "Sue oohs"/"Sue woos" as a function of ΔF1, for two values of ΔAV (0 dB and 16 dB).

however, constitute evidence that F1 cues are ignored in perceptual judgments between glides and high vowels. Redundant acoustic cues have been found to influence the reaction time required for categorical perception decisions, even when the decision conditioned by the primary cue is unambiguous (Whalen, Abramson, Lisker, & Mody, 1993). However, it would appear from the current experiment that the glide percept indicated by the ΔAV cue generally outweighs that of the ΔF1 cue, whether or not both are perceived and considered.

Because of the type of synthesis used in this perceptual study, it can be confidently concluded that the perceptual weighting given to ΔAV over ΔF1 reflects the greater importance of acoustic cues to the characteristics of the glottal source over those of the vocal tract filter shape. Isolated reductions in the AV parameter can only be perceived as arising from reductions in amplitude of the excitation source itself, since the formant frequencies do not change. On the other hand, an isolated reduction in the F1 parameter brings about its own natural reduction in the amplitude of the signal, since the cascade formant synthesizer approximates the acoustics of natural formants. That the glide rating responses of subjects are not affected by the amplitude reduction arising from the F1 parameter decrease to the same degree as an equivalent reduction in the AV parameter suggests that the perceptual system may be able to separate the amplitude contributions from the glottal source and vocal tract filter effects. It therefore appears to be specifically the contribution to ΔA_{RMS} of the weakening and skewing of the glottal source pulse that is most important to the perceptual distinction between glides and high vowels.

By observing where the response curves to the AV continua in panels (a) and (c) of Figure 42 cross the glide rating midpoint of 3.5, it can be concluded that the average category boundary between the perception of glides and high vowels is at about ΔAV = 8 dB in the absence of F1 decrease. Measurements of the actual ΔA_{RMS} of the synthetic glide tokens, using the method described in Section 2.2.1, indicate that the correspondence between ΔAV and ΔA_{RMS} is accurate to within ±1 dB (the level of accuracy of the xkl analysis tool). This phonetic boundary is well reflected in the production data from the /VGV/ natural recordings presented in Chapter 2, for which 90% of the glides in high vowel contexts were produced with $\Delta A_{RMS} \geq 8$ dB. (The average amplitude decrease for naturally produced glides in high vowel contexts was 14.2 dB.) Interestingly, the same value of ΔAV = 8 dB in the absence of F0 decrease was found to be the phonetic boundary for perception of /ʔ/ in Hillenbrand & Houde (1996).

This correspondence further strengthens the evidence for acoustic similarity between glides and glottalization, possibly reflecting a common effect on glottal vibration between these two types of sound segments.

3.4 Summary

The perceptual experiments undertaken in this chapter have isolated and varied two acoustic parameters that were identified in Chapter 2 as potential cues to the distinction between glides and high vowels. Naïve subjects were asked to choose whether they heard “see east” or “see yeast”, or alternatively “Sue oohs” or “Sue woos”, given only differences in intervocalic dips in voicing amplitude (AV) or first formant frequency (F1). Results indicate that the subjects’ perception of the glide /j/ or /w/ was conditioned more by the size of the amplitude dip (ΔAV) than by the size of the F1 dip (ΔF1). It was also found that an AV reduction was sufficient by itself to cue the presence of a glide, while an F1 reduction by itself was not. Since the AV reduction represents the weakening of the glottal source due to pressure build-up behind the narrow vocal tract constriction in a glide, and the F1 reduction represents the spectral characteristics of the narrow constriction in the vocal tract filter itself, the results of this study indicate that the filter’s loading effects on the glottal source are more central to the acoustic perception of glides than the shape of the filter alone.

The main experiment presented in Section 3.3 determined that the perceptual boundary between glides and high vowels occurs when the amplitude reduction is about 8 dB in the absence of F1 movement. A future question of interest is how this approximate category boundary of ΔA_{RMS} = 8 dB is specifically related to the articulatory configuration of glides. Several researchers have found that the intensity of speech sound in general is increased by 8-9 dB when the transglottal pressure is doubled (Isshiki, 1964; Tanaka & Gould, 1983; Sundberg, Titze, & Scherer, 1993). This would suggest that the

8 dB amplitude decrease marking the category boundary for glides could arise if the transglottal pressure drop were reduced to half of its value in the adjacent vowel. Such a decrease in the transglottal pressure drop would be brought about by the pressure built up behind the narrow oral constriction formed during the glide. If the subglottal pressure is assumed to remain constant through the sequence, as it has been shown to in normal, non-emphatic speech (Ohala, 1990; Sundberg, Elliot, Gramming, & Nord, 1993), then the halving of the transglottal pressure drop would be balanced by an equivalent increase in the oral pressure drop, such that the pressure drops across the glottal and supraglottal constrictions would be about equal in magnitude. This would occur if the glottal and supraglottal constrictions were adjusted to be about equal in cross-sectional area. On the other hand, the data indicating that an 8 dB amplitude decrease would require the transglottal pressure to be halved may underestimate the amplitude effects of such a pressure reduction. The available data on the relation between transglottal pressure and sound intensity show a certain degree of spread; some studies' data suggest that a doubling of transglottal pressure may be associated with larger intensity differences, up to about 13 dB (Isshiki, 1964; Holmberg *et al.*, 1988; Stathopoulos & Sapienza, 1993). In this case, the glide category boundary could be associated with an oral pressure drop that is not quite as large as the weakened transglottal pressure drop. Future articulatory experiments could investigate the actual constriction areas and airflows produced during glides, in order to better quantify the pressure drops across the glottal and oral constrictions. However, the current study has been limited to acoustic measures, and must leave the precise articulatory modeling to future work.

This chapter has also provided some perceptual evidence that the primary acoustic cue to glides, the reduction in glottal source amplitude in relation to the adjacent vowel, is also shared by the laryngeal consonant /ʔ/. This may be informative since the laryngeal consonants have been a source of uncertainty in distinctive feature systems (Parker, 2002), and some have suggested that they might share a class with the glides. In fact, Jakobson *et al*. (1952) call the laryngeal consonants /h, ʔ/ “glides” while maintaining that /j, w/ are types of vowels. The argument has been made, though, that /j, w/ and /h, ʔ/ should actually be classified together as [-consonantal, -vocalic] (e.g., Chomsky & Halle, 1968; Nevins & Chitoran, 2008), since they are both produced with a narrow constriction that does not produce turbulence noise (the constriction for glides is located within the oral tract, while the constriction for laryngeals is at the glottis). In the following chapter, it is suggested that the feature [-vocalic] might apply to both glides and laryngeals because of their common acoustic manifestation of defining effects on the glottal source, rather than an unspecific reference to “constriction degree”.

4 Summary and conclusions

4.1 Acoustic characteristics of glides

This thesis has presented a detailed study of the acoustic characteristics of glides in American English. At issue is what acoustic characteristics distinguish the glides /j/ and /w/ from the related high vowels /i/ and /u/, what information such acoustic characteristics provide about the articulation of glides, and how this information can inform the classification of glides in terms of distinctive features. Through acoustic measurements carried out on a database of intervocalic glides produced naturally by four speakers, it was shown in Chapter 2 that glides are significantly different from adjacent high vowels in terms of several acoustic characteristics. These acoustic differences are consistent with the production hypothesis that glides differ from adjacent vowels (and critically, from adjacent high vowels) in terms of constriction degree.

The glides /j/ and /w/ are known to be produced with constrictions in the oral cavity at the same locations as for /i/ and /u/, respectively. However, the acoustic data presented in this thesis support the hypothesis that the glides are produced with greater degree of constriction than adjacent high vowels. The further narrowing of the constriction changes the shape of the vocal tract filter between the high vowel and the glide, causing perturbations in the formant frequencies. These formant frequency movements are acoustic characteristics of the vocal tract filter shape alone, and would not be affected by any change in the excitation source that is coupled to the filter. For both /j/ and /w/, the narrowing of the vocal tract constriction causes the first formant frequency (F1) to be lowered relative to its value in the adjacent high vowel. However, it was shown in Section 2.3 that the amount of F1 lowering in high vowel contexts is rather small, probably due to the fact that the effects of the vocal tract walls limit the range of possible F1 movement at such low frequencies.

Yet, despite the fact that F1 is relatively insensitive to the movement between high vowel and glide, the few previous studies of the target acoustics of glides (those that did not focus solely on durational measurements) report only formant frequency measurements, suggesting that those researchers considered the vocal tract filter shape to be the only important (or perhaps the only available) source of production distinction between glides and high vowels. For instance, Maddieson & Emmorey (1985) reported that F1 is significantly lower in glides than in the corresponding high vowels across three different languages. However, the differences in their reported means are never greater than about 40 Hz, and it was shown in Chapter 3 of the current study that an F1 decrease of 40 Hz or less is hardly ever enough in isolation to cause the percept of a glide in a high vowel context. This raises the question of whether the vocal tract filter shape itself may not be the real key to the distinction between glides and high vowels.

In fact, the data presented in Chapter 2 suggest that additional acoustic phenomena, other than formant frequency movements, arise from the formation of the narrow vocal tract constriction in glides. In glides flanked by high vowels, there is a significant reduction in the overall amplitude of the signal (A_{RMS}), far in excess of the size of amplitude reduction that could result directly from the small F1 movement observed in these environments. In addition, there is a significant increase in open quotient (OQ), a significant decrease in harmonics-to-noise ratio (HNR), and often a decrease in fundamental frequency (F0). The combination of these acoustic effects suggests that the glottal excitation source is not independent of the vocal tract shape through which it is filtered during a glide, but is itself affected by the aerodynamic effects of the glide's narrow vocal tract constriction.

Although the vocal tract constriction for glides has hitherto been assumed to be wide enough that airflow through it is unhindered, as it is for vowels, the glide's constriction may in fact be narrow enough to cause pressure to build up somewhat in the cavity behind it. Assuming a constant subglottal pressure, the oral pressure build-up causes the transglottal pressure drop to decrease, thereby weakening the glottal sound source. The glottal waveform becomes decreased in peak amplitude and skewed to the right, causing a decrease in A_{RMS} and an increase in OQ. As the harmonic amplitude is decreased, aspiration noise remains relatively constant, leading to a decrease in HNR. F0 of the glottal source may also be decreased due to acoustic loading from the narrow vocal tract constriction, and to the lengthening of the glottal period caused by the increased length of the open phase. The observation of all of these acoustic effects in the production data from Chapter 2 lends support to the hypothesis that the narrow constriction formed during a glide has an aerodynamic effect on the glottal source excitation, and that multiple acoustic correlates may be used to detect this effect.

In Chapter 3, a perceptual study using synthetic speech was undertaken to investigate the relative perceptual salience of the source-filter interaction and filter-only acoustic cues to the distinction between glides and vowels. It was found that listeners give more perceptual weight to voicing amplitude (AV) cues (representative of glottal source effects) than to F1 cues (representative of vocal tract filter shape only) when deciding whether they heard glides or simply high vowels. In fact, the AV decrease was sufficient to cue the presence of a glide in the absence of any F1 cue, whereas the F1 decrease was not sufficient on its own. This indicates that the amplitude cue to the glottal source effects is more central to the perceptual definition of the glide category and its distinction from high vowels. The perceptual experiment found that a phonetic category boundary for the perception of glides occurred at $\Delta AV = 8$ dB, which is also consistent with the majority of the natural glide productions analyzed in Chapter 2. It is suggested that future articulatory studies pinpoint the precise constriction areas and airflows used during the production of glides, in order to directly relate this acoustic category boundary to the articulatory movements that condition it.

4.2 Variation: overlap and enhancement

All of the acoustic cues studied in this thesis are correlates of a single articulatory gesture for glides. That gesture is the formation of a narrow constriction in the oral part of the vocal tract using either the tongue body or the lips. The articulatory target is most constricted for glides that are adjacent to high vowels, since the constriction degree (with its aerodynamic effects on the glottal source) is the factor that separates the glides and high vowels into two different sound segment classes. However, as for many other types of sound segments, the articulatory gesture for glides may exhibit some variation in other environments due to contextual effects and overlap. Particularly apparent in the acoustic data in this thesis have been the coarticulatory effects of neighboring vowel heights on the minimum F1 target reached during glides.

In Section 2.3 it was shown that the minimum F1 ($F1_{min}$) reached at the glide landmark is significantly higher when the glide is flanked by low vowels than when the glide is flanked by high vowels. This could be explained by the coarticulatory "pull" of the low vowel's high F1 on the $F1_{min}$ of the glide; i.e., the lowered tongue body gesture for the surrounding vowels has an overlapping lowering effect on the constriction gesture for the glide. It could also be explained by a conservation of articulatory effort; since the perceptual distance between the glide and the low vowel is much larger than that between it and a high vowel, the effort need not be made to form as narrow a glide constriction in low vowel contexts than in high vowel contexts. In either case, the result is that $F1_{min}$ is significantly higher in low vowel contexts, even higher than typical F1 values for high vowels.

The fact that $F1_{min}$ is so raised in glides in low vowel contexts indicates that the vocal tract constriction in glides in these contexts is not made as narrow as in high vowel contexts. Since $F1_{min}$ in low vowel contexts is often higher than typical F1 for high vowels, it is unlikely that the glide constriction is narrow enough to cause significant aerodynamic effects on the glottal source such as those described above for glides in high vowel contexts. However, the acoustic amplitude correlates of such glottal source effects are preserved in low vowel contexts, even though the physical source of those amplitude characteristics is different. For glides in low vowel contexts, it is the movement of F1 that contributes primarily to the decrease in A_{RMS}, rather than the glottal source effects that are the primary contribution in high vowel contexts. Notwithstanding the coarticulation or conservation of articulatory effort that serves to decrease the constriction narrowing and raise $F1_{min}$ for glides in low vowel contexts, sufficient narrowing is produced to ensure that ΔA_{RMS} is not significantly different from its value in high vowel contexts. The potential invariance of ΔA_{RMS} to articulatory overlap was posited previously by Stevens (1998, p. 520):

> *"It should be noted, however, that there is not a requirement for this minimum constriction size to be achieved each time a glide is produced. This kind of precision is not needed, since the principal requirement for a glide (which always occurs immediately preceding a vowel, at least in English) appears to be that there is a sufficiently reduced low-frequency amplitude relative to the vowel."*

This thesis has provided acoustic data from a comprehensive database of canonical glide productions to support Stevens's claim, as well as a preliminary understanding of the articulatory basis of the amplitude reduction defining glides in different vocalic contexts. In high vowel contexts, glides seem to be produced with their most canonical constriction degree, to maintain articulatory and perceptual separation from nearby high vowels, and the defining amplitude reduction is caused by aerodynamic effects on the glottal source. In glides in low vowel contexts, there seems to be undershoot in the degree of vocal tract constriction, but the F1 movement is maintained at a sufficient size to approximate the defining amplitude reduction derived from the production in high vowel contexts.

Another source of acoustic variation is the effect of the prosodic context in which the glide occurs. The database of glide recordings on which acoustic analyses were performed in this thesis, presented in Chapter 2, allowed for intonational prosodic effects to be controlled through the systematic variation of the location of pitch accent with respect to the target glide. Prosodic context was shown to have some effect on the acoustic and production characteristics of glides, especially on the measurable decrease in F0 often observed around the glide landmark. Since intonational prosody surfaces in part through its determination of the F0 contour of an utterance, acoustic measurements of this parameter during a sound segment could not escape being affected by the surrounding prosodic context. Significant effects of prosody were also observed on acoustic measures independent of F0, however, including ΔA_{RMS} and $F1_{min}$. That these measures were most extreme in glides located before pitch-accented vowels indicates a certain amount of articulatory strengthening (i.e., more extreme constriction) in glides that begin pitch-accented syllables.

The acoustic analyses in this thesis have also identified potential avenues for enhancement of the glide/vowel contrast, should a speaker wish to increase the perceptual salience of glides in certain situations through articulatory mechanisms other than the defining constriction gesture. Such potential avenues present themselves when an acoustic correlate of the glide constriction gesture could also receive an independent contribution through a separate but concurrent articulatory gesture. For example, the lowered F0 often present in glides from loading on the glottal source could possibly be enhanced by a deliberate F0 decrease through slackening of the vocal folds. Also, the decreased HNR of glides could potentially be enhanced by the deliberate addition of frication noise through increased airflow or subglottal pressure. It is suggested that future studies designed to increase articulatory effort, perhaps incorporating communicative intent or emphatic speech, investigate the potential use of such enhancing gestures for the glide segments.

4.3 Glides and distinctive features

This study's work on the acoustic characteristics of glides suggests that a boundary between glides and high vowels may be defined in terms of the degree of constriction of the vocal tract and the resulting aerodynamic effects on the glottal excitation source. Glides are distinguished from high vowels by a narrower constriction, of such a degree that some pressure is built up in the oral/pharyngeal cavity behind it. This oral pressure causes a weakening and skewing of the glottal pulse waveform, resulting in increased open quotient and decreased overall amplitude of the voicing source. The fact that the acoustic characteristics presented in this thesis all refer to the target articulation of glides, with no need of durational references or specific rates of change, suggests that glides deserve their own classification within any distinctive feature framework. That their potential articulator-free feature specification can be related specifically to the aerodynamic production characteristics of their unique articulatory configuration is consistent with recent ideas offered by Stevens & Hanson (in press) regarding the articulatory basis for distinctive features.

It is of particular importance to the articulatory/acoustic basis of the potential glide feature that the category boundaries for glides are defined with reference to the same type of articulatory gesture as that for other types of consonants, that of a single constriction created at a certain place along the vocal tract. The cross-sectional area of that constriction for a glide is smaller than that of a vowel, but larger than that of a fricative produced at the same place of articulation, which is in turn larger than that of a stop consonant produced at the same place of articulation. In fact, from an aerodynamic standpoint, glides have more in common with fricatives than with vowels, given that the Reynolds number for glides is probably, along with fricatives, on the opposite side of a critical threshold from vowels. The Reynolds number of the airflow in a uniform tube is given by (Stevens, 1998, p. 28):

$$Re = \frac{Uh\rho}{\mu A_c}$$

where:

U = the volume velocity of the air particles

h = a characteristic dimension roughly equal to the diameter of the circular tube

$\rho = 0.0011\ \text{g/cm}^3$ = the air density in the vocal tract

$\mu = 1.94 \times 10^{-4}$ dyne-s/cm^2 = the air viscosity in the vocal tract

A_c = the cross-sectional area of the tube

If A_c is taken to be between 0.2 cm^2 and 0.4 cm^2, as Stevens (1998) has postulated for glides, and U is assumed to be 200 cm^3/s, which is given as the low end of the range of volume velocities for non-vowels, then the Reynolds number for glides is expected to be between 2023 and 4046. (If U is assumed to be higher, the Reynolds number would also be higher.) Since this theoretical Reynolds number for glides is greater than the critical Reynolds number of 2000, glides are expected to be produced with turbulent, rather than laminar, airflow. Unlike vowels, whose Reynolds number is less than 2000, glides are expected to share the turbulent property of fricatives, allowing a pressure drop to form across the oral tract constriction that affects the shape of the glottal source waveform.

Although glides are expected to have the capacity for turbulent airflow, however, it seems from the acoustic data in Chapter 2 that they are not typically produced with a measurable increase in frication noise. The power of the frication noise source that does develop at the glide constriction may not be large enough to emerge over the aspiration noise already present in the signal from the glottal constriction, even though the oral pressure may be large enough to influence glottal behavior. This contrastive noise characteristic serves to separate glides from fricative consonants at the same place of articulation. Along the continuous scale of constriction degrees, glides are apparently separated from high vowels on one side by the critical Reynolds number, allowing pressure to build up behind the glide constriction when it would not for a high vowel constriction. Glides are separated from fricatives on the other side of the constriction degree scale by the lack of salient frication noise with abrupt onset. The important point, however, is that glides are defined with reference to the same constriction degree scale on which the manner features of obstruent consonants are defined. Since distinctive features may be used to classify obstruent consonants as separate from vowels, the use of a distinctive feature to also classify glides as separate from vowels seems called for. If, however, the case is made for the elimination of all major class features, in favor of a sonority index, for instance (Selkirk, 1984a), then the glides should receive their own sonority value distinct from that of vowels, since they differ in terms of constriction degree.

Within the distinctive feature framework, if it is agreed that glides should be described by a distinctive feature that differentiates them from vowels, it must still be decided what such a feature should be named, and how it should be organized within the larger feature system. The [vocalic] feature has recently been in disuse, having been replaced by classification tools related to syllabicity. The classification of glides as [+consonantal], along with other consonants, has therefore been considered as a means of differentiating them from vowels without requiring the creation or resurrection of additional features (Padgett, 2008). Levi (2008) writes that the use of [consonantal] has been the most common method of differentiating glides from vowels; however, this use goes against the definition of the feature. From an articulatory standpoint, [+consonantal] has been defined by the production of a radical obstruction (i.e., occlusion or near-occlusion) in the midsagittal region of the vocal tract (Chomsky & Halle, 1968), and the vocal tract constriction in glides is not of this radical degree. Sound segments classified as [+consonantal] have also been identified acoustically with landmarks of abrupt discontinuities in spectral energy (Stevens, 2002), which glides do not normally exhibit. In the words of Fant (1986), “the hunt for maximum economy often leads to solutions that impair the phonetic reality of features.” We should not be hesitant to include an additional feature that is well supported by the physical data, just because it would increase the size of the theoretical feature inventory.

A new feature [glide] has been proposed by Stevens & Hanson (in press) in order to simultaneously differentiate glides from both vowels and all other types of consonants. The feature is part of a hierarchical system they propose, based on articulatory/acoustic relations that define the distinctive features. In such a hierarchy, the distinctive features used to specify any particular sound segment are sparse, such that particular features are only specified if they are applicable to the articulatory configuration at hand. The specification of a feature at a low node in the hierarchy implies the specification of the features at its parent nodes, while features at cousin nodes need not be specified. For example, Figure 43(a) shows the hierarchy proposed by Stevens & Hanson for all articulator-free features stemming from the [+sonorant] node, along with the types of sound segments they specify. For any sound specified by a node in the [+sonorant] tree, all of the articulator-free features stemming from the [-sonorant] node, such as [continuant] and [strident], are unspecified because their articulatory definitions do not apply to [+sonorant] sounds. In the tree in Figure 43(a), the [glide] feature is low in the articulator-free hierarchy; [+glide] specifies glides, [-glide] specifies vowels, and both classes are specified [-consonantal]. Nasals and liquids are unspecified for [glide], since they are [+consonantal], and [glide] does not apply to its cousin nodes.

Stevens & Hanson's (in press) articulator-free feature hierarchy in Figure 43(a) adequately differentiates glides from vowels and other consonants through the proposed [glide] feature. However, this feature is problematic in that it does not allow for the grouping of glides and consonants into a larger non-vowel class when necessary for the description of phonological processes. For example, glides have been shown to pattern with other consonants in blocking nasal harmony in Sundanese, while vowels and laryngeals do not (Padgett, 2008). If Stevens & Hanson's feature hierarchy is used, the set of

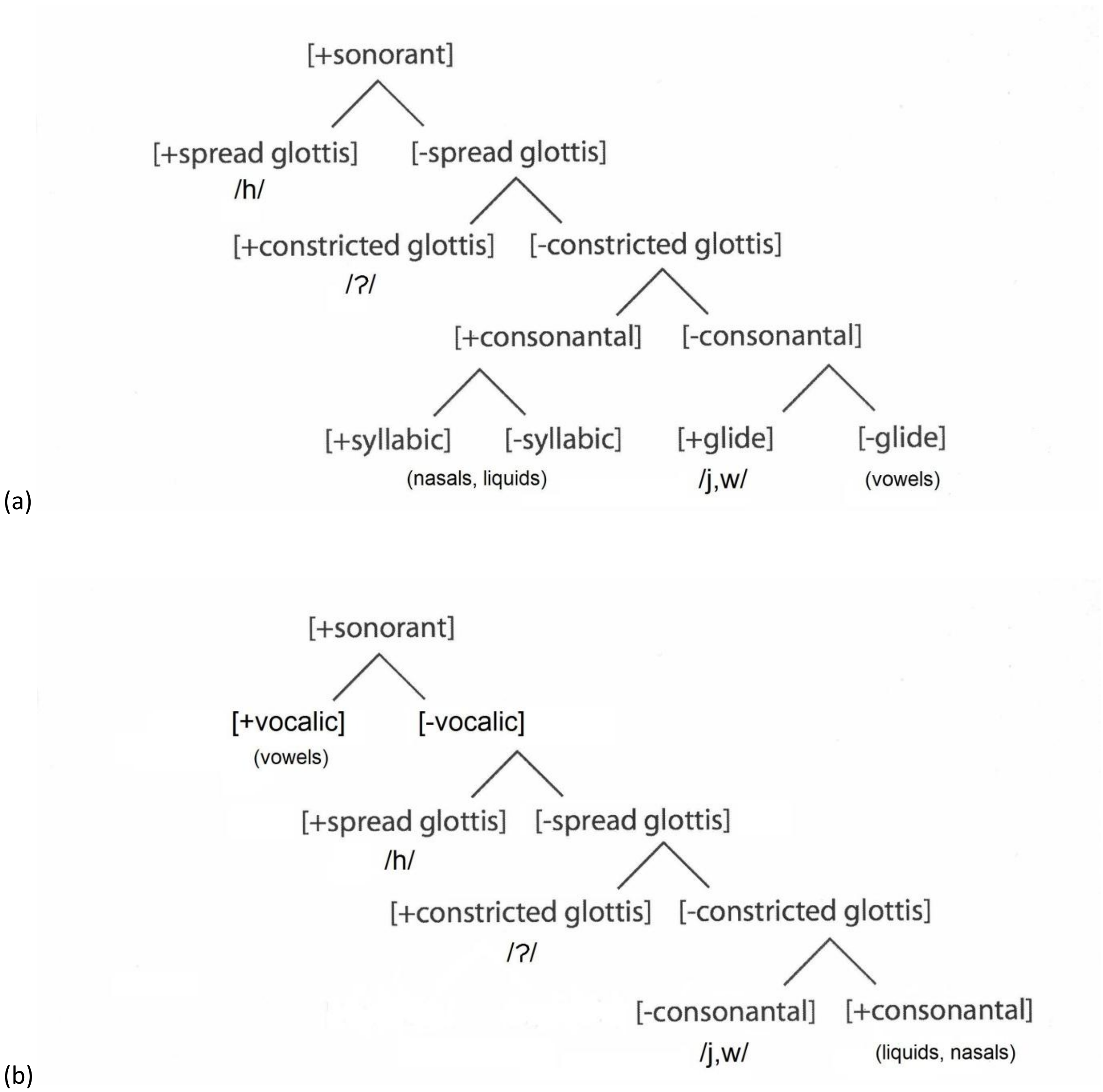


Figure 43: Possible hierarchies of articulator-free distinctive features, based on principles of aero-mechanical interactions, for the [+ sonorant] sounds. (a) Feature hierarchy proposed by Stevens & Hanson (in press), including the feature [± glide]. (b) Proposed modification of the feature hierarchy in (a), with [± glide] replaced by [± vocalic].

sounds that block nasal harmony cannot be specified by a single combination of features; rather, the rule must refer to the union of two disconnected feature sets [+consonantal] and [+glide]. Likewise, the set of sounds that allow nasal harmony (vowels, /h/, and /ʔ/) must be specified through disconnected sets that are far apart on the feature tree of Figure 43(a).

In addition to the requirement that the distinctive feature inventory and hierarchy be applicable to the phonological processes that make reference to it, it is valuable for the feature inventory to reflect the articulatory/acoustic relations that are supposed to be the defining basis for the features themselves, as Stevens & Hanson (in press) point out. Acoustic evidence such as that compiled in this thesis, along with the production methods it implies, should therefore be considered in any feature choice. The acoustic analyses undertaken in this thesis indicate that glides share an articulatory/acoustic relation with other consonants in the narrow constriction they create in the vocal tract. This constriction causes oral pressure build-up with weakening effects on the glottal source, differentiating the glides and consonants as a group from vowels, for which airflow through the vocal tract is unimpeded. This calls for the addition (or re-addition) of a feature such as [vocalic], which can be used to group glides and other consonants in opposition to vowels. The [consonantal] feature should also be retained according to the acoustic evidence from this study, since glides are separated from other consonants by their lack of occlusion, abruptness, or significant frication noise. The use of both [-vocalic] and [+consonantal] to specify glides is also phonologically warranted, since glides have been found to pattern with consonants rather than vowels for some processes, but with vowels rather than consonants for others (Nevins & Chitoran, 2008).

Stevens & Hanson's (in press) feature hierarchy could be modified to replace [glide] with [vocalic], as well as reordered to better reflect the articulatory and phonological relationships between the sound segments represented. A possible reorganization of their articulator-free feature tree for the

[+sonorant] sounds is given in Figure 43(b). Here, the [vocalic] feature is placed high in the hierarchy and the [consonantal] feature is placed low in the hierarchy, with the laryngeal features [spread glottis] and [constricted glottis] in between. The set of sound segments that block Sundanese nasal harmony are now grouped together as children of the same parent node, and can easily be classified together as [-constricted glottis]. To address phonological processes in which the laryngeal consonants pattern together with glides and other sonorant consonants (for example, in Karuk gemination (Levi, 2008)), these sounds can easily be classified together as [-vocalic] (n.b., [-nasal] would also be required in the specification for Karuk gemination). Note also that this proposed modification to Stevens & Hanson's hierarchy allows the sound segment classes to "fall out" of the feature tree in the following order, which corresponds to their relative degree of constriction of the vocal tract:

vowels > laryngeals > glides > liquids, nasals

This corresponds also to the order of an "openness" scale proposed to condition processes of consonant lenition, and can also be equated to a sonority scale (Kingston, 2006).

The placement of the laryngeal consonants /h, ʔ/ together with glides under the [-vocalic] node in the proposed feature tree is supported by the acoustic and perceptual data presented in this thesis, in combination with other researchers' data on the acoustics of /h, ʔ/. The evidence implies that glides and laryngeal consonants may be defined by similar aerodynamic effects on the glottal source from constrictions in the oral tract or at the glottis. Perceptual experiments have shown that the same voicing source amplitude cues can signal the presence of glides or /ʔ/ as distinguished from vowels. Note, however, that /h, ʔ/ remain unspecified for the [consonantal] feature in the hierarchy in Figure 43(b), as they were in Figure 43(a), specified only by the features [spread] and [constricted] specifying their laryngeal configurations (Stevens, 1977). This may make sense from a production standpoint, since realizations of the laryngeal consonants exhibit a great deal of variation, and /ʔ/ in particular can be produced with occlusion or without (Priestly, 1976). Phonologically, the laryngeals have proven extremely difficult to classify, since studies have not agreed about whether they pattern for the most part as sonorants or obstruents (Parker, 2002). Their status with respect to features such as [consonantal] therefore remains somewhat unclear; their classification as [-vocalic] along with the glides and other constricted sonorant consonants, however, seems well supported by this study and others.

It should be remembered that this thesis is intended to present an acoustic phonetic study of glides and the characteristics that differentiate them from other sound segments in the speech signal. It is not in any way meant to approximate a phonological treatise, and the author is certainly not the most qualified to predict the phonological ramifications of recommending a particular classification scheme for glides. However, the debate as to the phonological status of glides has been so unresolved that a few comments concerning the potential phonological implications of this acoustic study seemed called for. The author is also a proponent of the essential value of phonetics to phonology, as expounded quite well by Fant (1986). It is hoped that phonologists will consider the new acoustic evidence on the production and perception of glides presented here and in future, and allow it an integral part in informing more complete future systems for the description of these important and widely used sounds.

5 Suggestions for future work

Since the full range of target acoustic characteristics of glides have been relatively unstudied before now, there is a wealth of future work that could be done to expand and refine the results and conclusions of the current study. First and foremost, the articulatory hypotheses presented based on the acoustic data compiled in this study would be well served by confirmation from further production-oriented studies. The acoustic evidence in this thesis has supported a definition of glides based on the narrowness of their vocal tract constrictions producing pressure build-up in the oral cavity. This articulatory target configuration could be physically confirmed through studies using electropalatography or possibly MRI to measure the oral constriction dimensions, electroglottograph or laryngoscope recordings to directly characterize the glottal source waveform, or airflow measurements using specialized masks. In addition, the acoustic measurements made in this thesis could be repeated with more subjects or improved; in particular, it would be valuable to obtain first formant bandwidth (B1) measurements directly from glide tokens. Perhaps this could best be undertaken through an analysis-by-synthesis procedure using an external excitation source, similar to that followed by Fujimura & Lindqvist (1971) for vowels. The perceptual experiments presented in this thesis could also be expanded to test more of the potential acoustic cues to glidehood identified in the production study, in isolation and in combination with each other.

In addition to expansion and improvement of the study of glides in intervocalic contexts covered in this thesis, there are many other aspects of glides in everyday language that could not be investigated within the time constraints of this study. It is hoped that future work will explore many of these additional contexts and applications, some of which are introduced in this chapter.

Glides in varied segmental contexts

This thesis has focused exclusively on glides in intervocalic contexts, specifically with identical vowels flanking either side of the glide under study. Since glides are known always to occur in pre-vocalic position, this was a reasonable context in which to begin an acoustic analysis. In addition, an effort was made to pinpoint the acoustic targets of the glide landmark, and these are easiest to discern when similar transitions into and out of the glide can be observed. However, our understanding of the acoustic characteristics of glides cannot be complete without studying glides in less limited segmental contexts as well.

The glide landmarks and transitions are likely to differ from the behavior observed in this study when they occur after consonants rather than vowels. Glide landmarks may be difficult to locate when overlapped by articulatory and spectral transitions out of preceding consonants. Glides following obstruent consonants may be particularly changed, as the frication noise from the release of the preceding consonant may be elongated through the narrow constriction of the glide. Such elongated noise may go so far as to mutate into affrication, as when “that you” is pronounced “thatchoo”, for instance.

Aspiration noise may also be incorporated into glide productions, to the point that many speakers produce what may be termed voiceless glides. In some dialects of English, this voiceless quality has been claimed to differentiate words such as “hue” from “you”, or “whether” from “weather”. Voiceless /j, w/ have also been reported to occur as contrastive segments in a number of other languages (Ladefoged & Maddieson, 1996). All of the above variations can occur in canonical glide productions in speech of normal effort, and their acoustic characteristics should be measured in order to complete the acoustic picture for canonical glides.

Conversational and emphatic speech

Once the production and acoustics of canonical glides are well understood, future studies can proceed to investigate glides in running speech and conversational corpora. Like all other sound segments, it is expected that glides would exhibit greater degrees of gestural overlap, articulatory undershoot, and lack of acoustic clarity in connected speech than in the controlled productions elicited in the current study. Research on such hypo-articulated glides in running speech would further complete our understanding of the production and acoustics of these sound segments in everyday human language, as well as better prepare us for speech technology applications.

In addition to glides produced with less than canonical articulatory effort, it would also be interesting for future research to consider how glide production changes with stronger articulatory effort. This could be studied through the elicitation of emphatic speech, speech conveyed over noise, or speech intended to overcome a communicative hurdle (such as addressing a child, or a hearing-impaired person). This thesis has identified potential avenues for strengthening or enhancement of glides, such as deliberate F0 lowering or the addition of frication noise, that could possibly be used by speakers in such situations. These or others yet to be identified could be investigated in future work.

Classification of other glide-like sounds

An open phonological question that could be addressed in future research is how to classify certain sound segments which seem to share similarities with the class of glides. These other sounds may be argued to belong with the glides due to similarities in production, acoustics, or phonological patterning. The laryngeal consonants /h, ʔ/ have already been identified as sharing similar acoustic cues with glides, stemming from the glottal source effects that differentiate them from adjacent vowels. It was argued in Chapter 4 that the glides and laryngeals should share the [-vocalic] distinctive feature, while the laryngeals' status with respect to [consonantal] was left speculative. Future production studies, as suggested above for glides, could also be applied to determine the specific aerodynamic and acoustic characteristics of laryngeal consonants, to shed more light on the similarities and dissimilarities between the two types of sounds.

The liquids /r, l/ are generally assumed to occupy their own feature class distinct from glides, at least in many languages. However, the approximant /r/ in American English is produced very differently than in other languages in which it is flapped or trilled. Ladefoged & Maddieson (1996) point out that American English /r/ has something in common with glides, in that it bears the same semivowel relationship to the vowel in "bird" as /j/ does to the vowel in "bee". Selkirk (1982) notes that /r/ is different from other consonants in its pattern of preventing aspiration in following voiceless stops, suggesting that /r/ be considered a glide. Fant (1962) describes the lateral /l/ as possessing the "vowellike" feature (as do glides), since its spectrum exhibits a clear formant pattern. Fant also classifies variants of /r, l, j, w/ together as glides since their rate of spectrum change is relatively slow, but faster than that of two adjacent vowels.

Phonological evidence also points to the grouping of liquids and glides by some similarities of patterning. In English, /r, l, j, w/ are the only segments that can constitute the third member of a syllable-initial three-segment consonant cluster (e.g., "screw", "splint", "skew", "square"), and each of them is required to occur immediately before the vowel, without any other intervening consonant (O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957). Productional or acoustic similarities between glides and liquids are also implied by the fact that the articulatorily difficult liquids often surface as glides in child speech (Haelsig & Madison, 1986; Inkelas & Rose, 2008). Perhaps the application of the glide acoustic parameters identified in this thesis to the study of liquids could inform the discussion of how best to classify the relationship between glides and liquids, at least in English.

The acoustic analyses in this thesis provide evidence for an articulatory-acoustic mapping defining the category boundary between glides and vowels based on aeromechanical interactions. By giving a physical basis to the proposed [vocalic] feature differentiating glides from vowels, these data offer not only a newly detailed characterization of glides, but also a more complete understanding of vowels and the articulatory limits of their feature class. With a new concept of the [vocalic] feature marking the vowel-glide boundary, the acoustic analyses from this study could be further used to investigate other constriction-related boundaries between vowels themselves, which include the relations for the features [high], [low], and [tense]. These are classified as articulator-bound features, which, according to Stevens & Hanson (in press), tend to be defined by acoustic resonator coupling relations. It is proposed that the feature [low] is defined by the relation between the first formant frequency and the first subglottal resonance, but the features [high], [glide] (proposed [vocalic] in this thesis), and [tense] are listed as "features for which defining attributes have not been clearly worked out" by Stevens & Hanson. As the current study has made progress in discovering the defining relations for [vocalic], it is hoped that future work will find renewed success in describing features such as [high] and [tense].

Cross-linguistic studies

The analyses and experiments presented in this thesis are limited to data from glides in the American English language, and it would be of great value to our understanding of language universals to expand this work to cross-linguistic studies. It is expected that many of the acoustic characteristics of American English glides studied in this thesis would be similar for glides in other languages. However, the articulatory targets of glides may be slightly different in other languages, especially if they have more or fewer categorical distinctions to be made along the same articulatory gesture continuum. Padgett (2008) claims that cross-linguistic glides are of two types, "semivocalic" and "consonantal", which are differentiated by their constriction degree. If "consonantal" glides, for instance, are produced with narrower constrictions than typical American English glides, then it is possible that they would exhibit greater pressure drops and more observable frication noise than was found in the current study. This would explain Padgett's description of palatalizing mutation conditioned by the frication characteristics of Slavic consonantal glides, for example, while maintaining the separation between American English glides and fricatives by a frication-based boundary.

Although /j, w/ are the only glides in English, there are also many more sounds that could be classified as glides cross-linguistically. In addition to /j, w, ɥ, ɰ/ (the glide cognates of the rounded and unrounded, front and back high vowels), Ladefoged & Maddieson (1996) list glide cognates of the mid vowels /e, o/, as well as bilabial and labiodental approximants. Future acoustic studies of these other glide sounds could help to determine how representative the American English glides are of the cross-linguistic class as a whole.

Speech recognition and synthesis

The acoustic and perceptual data presented in this thesis have very direct applications for speech technologies such as automatic recognition and synthesis. For knowledge-based approaches to speech recognition, the available information about the expected acoustic patterns for glide segments is now significantly expanded from its previous scope, and will hopefully continue to grow. Close to home, for example, the Speech Communication Group at MIT has been developing a model for human speech recognition called Lexical Access from Features (LAFF) (Stevens, 2002; Slifka *et al.*, 2004; Park & Chen, 2008; Park, 2008). The model searches for the acoustic landmarks of the articulator-free features of sound segments, and then measures acoustic cues to the articulator-bound features in the vicinity of the landmarks, theoretically resulting in the complete specification of all the discrete phonological segments in any received utterance signal. The most recent implementation of the automatic glide landmark detection module of the LAFF model reported a detection rate of 88% and an insertion rate of 9.4%, using only the measurement of cues related to RMS amplitude and first formant frequency (Sun, 1996). Based on the results of the acoustic analyses conducted in this thesis, this module could now be improved, combining additional cues such as open quotient (OQ), harmonics-to-noise ratio (HNR), and fundamental frequency (F0) in a probabilistic framework along with attention to local prosodic and segmental contexts to improve glide detection.

The acoustic characteristics of glides identified and quantified in this thesis could also be applied to the improvement of the naturalness of glides in speech synthesis. In particular, the perceptual experiments conducted in this thesis have shown the importance of the amplitude reduction characteristic to the perception of glides, suggesting that glide synthesis employing only formant frequency movements may sound unnatural, or even fail to elicit the percept of a glide. The results of this study also have applications for other types of speech technology; for example, the importance of local amplitude characteristics to the perception of glides raises an important issue for electrolarynx speech, which may need to incorporate the capability for more detailed control of the source amplitude in order to better approximate natural sounding speech.

All of the above potential applications and avenues for future work of interest on the subject of glides, as well as the implications for models of speech production and phonological systems already discussed in detail, make the case for the importance of the acoustic and articulatory study of glides. It is hoped that such work will progress so that the glides will continue to become better represented in the speech science literature.

Bibliography

Aoyama, K., & Reid, L. A. (2006). Cross-linguistic tendencies and durational contrasts in geminate consonants: an examination of Guinaang Bontok geminates. *Journal of the International Phonetic Association , 36* (2), 145-157.

Bickley, C., & Stevens, K. (1986). Effects of a vocal-tract constriction on the glottal source: experimental and modelling studies. *Journal of Phonetics , 14*, 373-382.

Bothorel, A., Simon, P., Wioland, F., & Zerling, J.-P. (1986). *Cinéradiographie des Voyelles et Consonnes du Français.* Strasbourg, France: L'Institut de Phonetique de Strasbourg.

Catford, J. (1988). *A Practical Introduction to Phonetics.* Oxford: Clarendon Press.

Chitoran, I. (2002). A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association , 32* (2), 203-222.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English.* Cambridge, Massachusetts: MIT Press.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics , 24*, 423-444.

Fant, G. (1997). Acoustical Analysis of Speech. In M. Crocker (Ed.), *Encyclopedia of Acoustics* (Vol. 4, pp. 1589-1597). John Wiley.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. *LOGOS , 5* (1), 3-17.

Fant, G. (1986). Features - fiction and facts. In J. Perkell, & D. Klatt (Eds.), *Invariance and Variability of Speech Processes* (pp. 481-491). Lawrence Erlbaum Ass. Publ.

Fant, G. (1983). *Preliminaries to analysis of the human voice source.* Quarterly Progress and Status Report, Royal Institute of Technology (KTH, Speech Transmission Laboratory, Stockholm.

Fant, G. (1972). *Vocal tract wall effects, losses, and resonance bandwidths.* Quarterly Progress and Status Report, Royal Institute of Technology (KTH), Speech Transmission Laboratory, Stockholm.

Fant, G., Nord, L., & Branderud, P. (1977). *A note on the vocal tract wall impedance.* Quarterly Progress and Status Report, Royal Institute of Technology (KTH), Speech Transmission Laboratory, Stockholm.

Fujimura, O., & Lindqvist, J. (1971). Sweep-Tone Measurements of Vocal-Tract Characteristics. *Journal of the Acoustical Society of America , 49* (2B), 541-558.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI: constraints on phonetic interpretation.* Cambridge: Cambridge University Press.

Haelsig, P. C., & Madison, C. L. (1986). A Study of Phonological Processes Exhibited by 3-, 4-, and 5-Year-Old Children. *Language, Speech, and Hearing Services in Schools , 17*, 107-114.

Halle, M. (1992). Features. In W. Bright (Ed.), *Oxford International Encyclopedia of Linguistics.* New York: Oxford University Press.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America , 125* (1), 425-441.

Hanson, H. M. (1995). *Glottal Characteristics of Female Speakers.* PhD Thesis, Harvard University, Division of Applied Sciences, Cambridge, Massachusetts.

Hillenbrand, J. M., & Houde, R. A. (1996). Role of F0 and Amplitude in the Perception of Intervocalic Glottal Stops. *Journal of Speech and Hearing Research , 39*, 1182-1190.

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *Journal of the Acoustical Society of America , 84* (2), 511-529.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language , 55* (1), 37-58.

Inkelas, S., & Rose, Y. (2008). Positional neutralization: A case study from child language. *Language , 83* (4), 707-736.

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*, *1*, pp. I669-I672.

Isshiki, N. (1964). Regulatory mechanism of voice intensity variation. *Journal of Speech and Hearing Research , 7*, 17-29.

Jackson, P. J., & Shadle, C. H. (2000). Frication noise modulated by voicing, as revealed by pitch-scaled decomposition. *Journal of the Acoustical Society of America , 108* (4), 1421-1434.

Jackson, P. J., & Shadle, C. H. (2001). Pitch-Scaled Estimation of Simultaneous Voiced and Turbulence-Noise Components in Speech. *IEEE Transactions on Speech and Audio Processing , 9* (7), 713-726.

Jakobson, R., & Halle, M. (1956). *Fundamentals of Language.* The Hague: Mouton & Co.

Jakobson, R., Fant, C. G., & Halle, M. (1952). *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates.* Technical Report No. 13, Massachusetts Institute of Technology, Acoustics Laboratory, Cambridge, Massachusetts.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative Phonology.* San Diego: Academic Press.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language , 82* (1), 33-63.

Kingston, J. (2006). Lenition. In L. Colantoni, & J. Steele (Ed.), *Proceedings of the Third Conference on Laboratory Approaches to Spanish Phonology.* Cascadilla Press.

Klatt, D. H. (1986). Representation of the first formant in speech recognition and in models of the auditory periphery. In P. Mermelstein (Ed.), *Montreal Satellite Symposium on Speech Recognition, 12th International Congress on Acoustics. Proceedings.* Toronto.

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

Klatt, D., & Klatt, L. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America , 87* (2), 820-857.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages.* Oxford: Blackwell.

Lehiste, I., & Peterson, G. E. (1961). Transitions, Glides, and Diphthongs. *Journal of the Acoustical Society of America , 33* (3), 268-277.

Levi, S. V. (2008). Phonemic vs. derived glides. *Lingua , 118*, 1956-1978.

Mack, M., & Blumstein, S. E. (1983). Further evidence of acoustic invariance in speech production: The stop-glide contrast. *Journal of the Acoustical Society of America , 73* (5), 1739-1750.

Maddieson, I. (2008). Glides and gemination. *Lingua , 118*, 1926-1936.

Maddieson, I., & Emmorey, K. (1985). Relationship between Semivowels and Vowels: Cross-Linguistic Investigations of Acoustic Difference and Coarticulation. *Phonetica , 42*, 163-174.

Mehta, D. (2006). *Aspiration noise during phonation: synthesis, analysis, and pitch-scale modification.* SM Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, Massachusetts.

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* (5), 1751-1755.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics , 25* (6), 457-465.

Nevins, A., & Chitoran, I. (2008). Phonological representations and the variable patterning of glides. *Lingua , 118*, 1979-1997.

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

Ohala, J. J. (1990). Respiratory activity in speech. In W. Hardcastle, & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 23-53). The Netherlands: Kluwer Academic Publishers.

Okobi, A. O. (2006). *Acoustic Correlates of Word Stress in American English.* PhD Thesis, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts.

Padgett, J. (2008). Glides, vowels, and features. *Lingua , 118*, 1937-1955.

Park, C. (2008). *Consonant landmark detection for speech recognition.* PhD Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, Massachusetts.

Park, C., & Chen, N. (2008). Consonant landmarks: Automatic detection and interpretation. *Journal of the Acoustical Society of America , 124* (4), 2527.

Parker, S. G. (2002). *Quantifying the Sonority Hierarchy.* PhD Thesis, University of Massachusetts, Amherst.

Pierrehumbert, J. B., & Frisch, S. (1997). Synthesizing Allophonic Glottalization. In J. P. Van Santen, R. W. Sproat, J. P. Olive, & J. Hirschberg (Eds.), *Progress in Speech Synthesis* (pp. 9-26). New York: Springer.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty, & R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90-119). Cambridge, England: Cambridge University Press.

Priestly, T. M. (1976). A Note on the Glottal Stop. *Phonetica , 33*, 268-274.

Repp, B. H. (1982). Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception. *Psychological Bulletin , 92* (1), 81-110.

Selkirk, E. O. (1984a). On the Major Class Features and Syllable Theory. In M. Aronoff, & R. T. Oehrle (Eds.), *Language Sound Structure* (pp. 107-136). Cambridge, Massachusetts: MIT Press.

Selkirk, E. O. (1984b). *Phonology and syntax: the relation between sound and structure.* Cambridge, Massachusetts: MIT Press.

Selkirk, E. O. (1982). The Syllable. In H. van der Hulst, & N. Smith (Eds.), *The structure of phonological representations (Part II)* (pp. 337-383). Cinnaminson: Foris Publications.

Slifka, J., Stevens, K. S., Manuel, S., & Shattuck-Hufnagel, S. (2004). A landmark-based model of speech perception: History and recent developments. *From Sound to Sense: Fifty+ Years of Discoveries in Speech Communication. Proceedings*, (pp. C85-C90). Cambridge, Massachusetts.

Smits, R. (1994). Accuracy of quasistationary analysis of highly dynamic speech signals. *Journal of the Acoustical Society of America , 96* (6), 3401-3415.

Stathopoulos, E. T., & Sapienza, C. (1993). Respiratory and Laryngeal Function of Women and Men During Vocal Intensity Variation. *Journal of Speech and Hearing Research , 36*, 64-75.

Stevens, K. N. (1998). *Acoustic Phonetics.* Cambridge, Massachusetts: MIT Press.

Stevens, K. N. (1971). Airflow and Turbulence Noise for Fricative and Stop Consonants: Static Considerations. *Journal of the Acoustical Society of America , 50* (4B), 1180-1192.

Stevens, K. N. (1977). Physics of Laryngeal Behavior and Larynx Modes. *Phonetica , 34*, 264-279.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America , 111* (4), 1872-1891.

Stevens, K. N., & Hanson, H. M. (in press). Articulatory-acoustic relations as the basis of distinctive contrasts. In W. Hardcastle, & J. Laver (Eds.), *Handbook of Phonetic Sciences* (2nd ed.). Malden, Massachusetts: Wiley-Blackwell.

Sun, W. (1996). *Analysis and Interpretation of Glide Characteristics in Pursuit of an Algorithm for Recognition.* SM Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, Massachusetts.

Sundberg, J., Elliot, N., Gramming, P., & Nord, L. (1993). Short-Term Variation of Subglottal Pressure for Expressive Purposes in Singing and Stage Speech: A Preliminary Investigation. *Journal of Voice , 7* (3), 227-234.

Sundberg, J., Fahlstedt, E., & Morell, A. (2005). Effects on the glottal voice source of vocal loudness variation in untrained female and male voices. *Journal of the Acoustical Society of America , 117* (2), 879-885.

Sundberg, J., Titze, I., & Scherer, R. (1993). Phonatory Control in Male Singing: A Study of the Effects of Subglottal Pressure, Fundamental Frequency, and Mode of Phonation on the Voice Source. *Journal of Voice , 7* (1), 15-29.

Tanaka, S., & Gould, W. J. (1983). Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects. *Journal of the Acoustical Society of America , 73* (4), 1316-1321.

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: Theory. *Journal of the Acoustical Society of America , 123* (5), 2733-2749.

Titze, I., Riede, T., & Popolo, P. (2008). Nonlinear source-filter coupling in phonation: Vocal exercises. *Journal of the Acoustical Society of America , 123* (4), 1902-1915.

Whalen, D., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics , 23*, 349-366.

Whalen, D., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America , 93* (4), 2152-2159.

Zañartu, M., Mongeau, L., & Wodicka, G. R. (2007). Influence of acoustic loading on an effective single mass model of the vocal folds. *Journal of the Acoustical Society of America , 121* (2), 1119-1129.