Average Formant Trajectories

Steven Sandoval[a,*], Rene L. Utianski[b]

[a] *School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA*
[b] *Department of Neurology, Mayo Clinic, Rochester, MN 55902 USA*

Abstract

The use and study of formant frequencies for the description of vowels is commonplace in acoustical phonetics, with uses ranging from quality description, to identification/classification, and perception. However, numerous studies have shown that vowels are more effectively separated when the acoustic parameters are based on spectral information extracted at multiple time points, rather than at a single time instance. This suggests that spectral dynamics play an integral part in phonetic specification. In this paper, we provide an analysis of the average trajectories of the first two formant frequencies using two popular speech databases. Unlike previous studies of formant trajectories, we analyze speech samples that exhibits a wide range of speakers, dialects, and coarticulation contexts. We illustrate how the formant trajectories vary with gender and, to a lesser extent, with age. Additionally, we provide average formant trajectories for phoneme groups that are not typically considered. Furthermore, we point out that phonemes which have close $F1$ and $F2$ values at the temporal midpoint, often exhibit formant trajectories progressing in different directions, promoting the importance or formant trajectory progression. Finally, we briefly consider three-dimensional average formant trajectories.

Keywords: Formant trajectory, Formant dynamics, Fine phonetics, Dynamics of speech

Highlights

- Speech material from different ages, genders, dialects, and contexts was employed.
- In general, average formant trajectories displayed consistent trends across speakers.
- Average formant trajectories were considered for phonemes other than vowels.
- Dynamic formant measurements offer possible explanations of perceptual consequences.
- Three-dimensional average formant trajectories are visualized and briefly discussed.

* Corresponding author
Email addresses: spsandov@asu.edu (Steven Sandoval), Utianski.Rene@mayo.edu (Rene L. Utianski)
URL: http://StevenSandoval.info (Steven Sandoval)

1. Introduction

The use of formant frequencies has played a central role in the development and testing of theories of vowel recognition since popularized by the seminal study of vowels by Peterson and Barney (1952). Over the last 60 years, there have been many different kinds of studies that have established the role of the first two formant frequencies, $(F1/F2)$, as the main determiners of vowel quality (Peterson and Barney, 1952; Fant, 1973; O'Shaughnessy, 1987; Watson and Harrington, 1999; Quatieri, 2002). These various studies range from research of vowel recognition (Nearey, 1978; Nearey et al., 1979; Syrdal, 1985; Syrdal and Gopal, 1986; Lippmann, 1989; Miller, 1989; Nearey, 1992; Hillenbrand and Gayvert, 1993b; McDougall and Nolan, 2007), and speech perception (Delattre et al., 1952; Klein et al., 1970) to articulatory-to-acoustic modeling (Stevens et al., 1953; Fant, 1960), and acoustic phonetic cues (Peterson and Barney, 1952; Ladefoged, 1972). All of the aforementioned studies have shown high correlation between the first two formant frequencies and phonetic height and backness. Since relative values of the first and second formants roughly relate to the size and shape of the cavities created by jaw opening $(F1)$ and tongue position $(F2)$, the formant frequencies are an acoustic proxy for the kinematic displacements of the articulators (Lee and Shaiman, 2012). The preceding insights have led to a convenient phonetic/acoustic/perceptual portrayal of vowels, called a vowel diagram, which is formed by arranging the vowel tokens in the $F2/F1$ space (Essner, 1947; Joos, 1948; Watson and Harrington, 1999). An example of a vowel diagram and corresponding words in /hVd/ context is shown in Fig. 1.

As useful as $F1/F2$ measurements and the illustrative vowel diagram have proven to be, there is also a large body of evidence indicating that dynamic properties such as duration (Bennett, 1968; Ainsworth, 1972; Jenkins et al., 1983; Nearey, 1989) and spectral change (Jenkins et al., 1983; Strange et al., 1983; Nearey and Assmann, 1986; Nearey, 1989; Benedetto, 1989; Strange, 1989a; Whalen, 1989; Hillenbrand and Gayvert, 1993a; Hillenbrand et al., 1995) play an important role in vowel perception. For example, some vowels may have long or short vowel onglides or offglides, resulting in a considerable displacement of the formant frequencies across duration from the values at the temporal midpoint (Lehiste and Peterson, 1961; Huang, 1986; Strange, 1989b; Bernard, 1981; Cox, 1996, 1998; Harrington and Cassidy, 1994; Harrington et al., 1997; Watson and Harrington, 1999). Although the effectiveness of the first two formant frequencies in vowel identification is indisputable, it has also been recognized that information derived from beyond the temporal midpoint provides many kinds of cues to vowel quality (Watson and Harrington, 1999). For example, acoustic classification studies (Harrington and Cassidy, 1994; Hillenbrand et al., 1995; Huang, 1992; Zahorian and Jagharghi, 1993; Neel, 2004; Hillenbrand, 2013) have shown that 1) vowels are more effectively separated when the acoustic parameters are based on spectral information extracted at multiple time points, rather than at a single time instance; 2) spectral change patterns aid in the statistical separation of vowels in both fixed and variable phonetic environments (Hillenbrand, 2013); and 3) static vowel targets are not necessary for vowel identification, nor are they sufficient to explain the very high levels of vowel intelligibility reported in studies such as Peterson and Barney (1952) and Hillenbrand et al. (1995). Additionally, it was demonstrated that formant trajectory is beneficial for the within-class separation of the tense/lax monophthong

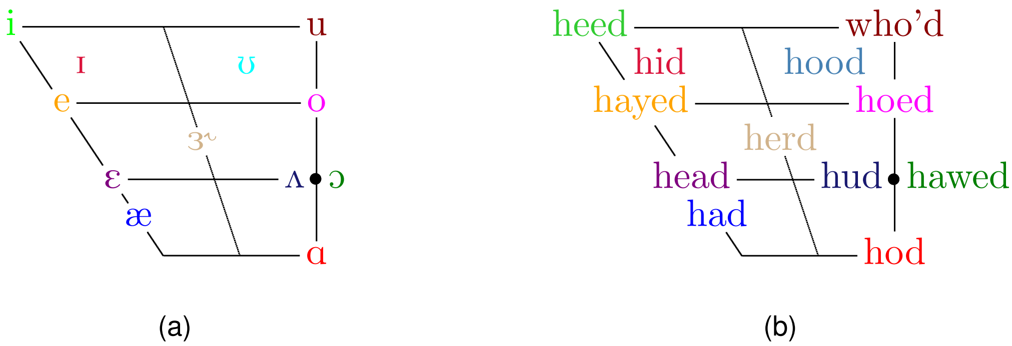


Fig. 1. An IPA vowel trapezium showing (a) American English vowels; and (b) the corresponding /hVd/ context words; used by Hillenbrand et al. (1995).

pairs (Watson and Harrington, 1999). The need to study the spectral changes associated with the vowels that are typically regarded as monophthongs, rather than using information from a single time point, has long been recognized (Peterson and Barney, 1952; William, 1953). Nearey and Assmann (1986) coined a term, vowel inherent spectral change, that specifically includes the formant changes associated with monophthongs (Morrison and Assmann, 2012; Nearey, 2013). In fact, all but a few nominally monophthongs show a significant amount of spectral movement through the courses of the vowel, even when those vowels are spoken in isolation (Hillenbrand, 2013). However, the discussion of formant changes is far more prevalent in studies of diphthongs (Morrison, 2009) than monophthongs, where vowel duration is typically used as an additional feature to classify vowels, rather than considering the formant trajectories (Watson and Harrington, 1999).

The long standing practice of static vowel representation in phonetic/acoustic/perceptual space, rather than trajectories through that space, remains in use despite several authors pointing out that this oversimplification has fundamental limitations which are not always acknowledged in interpretation (Hillenbrand, 2013). Although it has been suggested in the literature that spectral change, such as the trajectory of vowel formants, may be useful in the identification and classification of vowels, very little work has been done to quantify the progression of formant trajectories. Many works which seek to quantify formant trajectories utilize only a coarsely sampled two point trajectory (Klatt, 1980; Nearey and Assmann, 1986; Assmann and Katz, 2000), and while other studies have considered more detailed trajectories, these studies are limited to only a few speakers (Broad and Clermont, 2002; Neel, 2004; Kewley-Port and Neel, 2006; Broad and Clermont, 2010), a single dialect region (Fox and Jacewicz, 2009; Nearey, 2013), a specific range of ages (Morrison and Assmann, 2012), or a single word context (e.g. isolated vowels or single consonant-vowel or consonant-vowel-consonant context) (Broad and Fertig, 1970; Broad and Clermont, 1987; Nearey, 2013). To the best knowledge of the authors, no studies have attempted to quantify formant trajectories using a wide range of speakers, dialects, and coarticulation contexts, while also assessing the formants throughout full duration of phoneme production.

The purpose of this paper is to provide an initial analysis of the trajectories of formants using two popular speech databases to offer average formant trajectories that are representative of standard American English. The paper is organized into two studies. The first utilizes the Hillenbrand database, allowing for the comparison of this method to a widely cited assessment of vowel characteristics. The second study examines formant trajectories on the comprehensive TIMIT database, which offers several dialects and coarticulation contexts, and allows the examination of not only vowels but also other phoneme types. Briefly, we illustrate that phoneme tokens which lie close to each other in the $F2/F1$ space, preventing easy discrimination based on the $F2/F1$ at the temporal midpoint, often exhibit formant trajectories progressing in different directions, allowing easy visual discrimination when a formant trajectory in utilized. Use of the third formant, $F3$, in average formant trajectories is also succinctly examined.

2. Experiment 1

The first study examines the average formant trajectories present in the database provided by Hillenbrand et al. (1995). Average formant trajectories for each vowel token were computed for four classes of speakers based on gender and age. Results are provided in the form of figures showing the average formant trajectories.

2.1. Method

2.1.1. Speech Material

The Hillenbrand et al. (1995) database consists of recordings of /hVd/ utterances spoken by a 45 men, 48 women, and 46 children (27 boys, 19 girls) sampled at 16 kHz. Measurements of the formant frequencies are provided with the Hillenbrand database that were calculated using Linear Predictive Coding (LPC) analysis using a 16 ms window hamming window and an 8 ms frame advance. The formant frequencies were estimated using a three-point parabolic interpolator, yielding a finer resolution than the 61.5-Hz frequency quantization. The results were verified and hand edited to correct and tracking errors that occurred. The formant frequencies are provided for 10-80% vowel duration at 10% increments. However, limitations of this database include: 1) the relatively small database size (139 subjects); 2) limited dialect variation (87% were raised in Michigan's lower peninsula); 3) words spoken only in /hVd/ context; and 4) utilization of only one instance of each word per speaker.

2.1.2. Trajectory Averaging

For the Hillenbrand data, values of the formant frequencies are pre-computed and provided with the database, therefore, only trajectory averaging must be performed to obtain the average format trajectories. Using MATLAB (2014), utterances corresponding to a common vowel token are collected and the mean formant values across the utterances, at each temporal point relative to the vowel duration, are computed. This results in a mean trajectory in the $F2/F1$ space for each of the tokens in the database.

2.2. Results and Discussion

2.2.1. Vowel Formant Trajectories

The mean trajectory for each token in the database can be plotted in the $F2/F1$ space resulting in a plot similar to the standard IPA vowel trapezium. However, unlike standard vowel diagrams in which each token is represented as a point in the $F2/F1$ space, here each token is represented by a curve in the $F2/F1$ space. Fig. 2 shows the average formant trajectories for each of the tokens in the Hillenbrand database (i.e., 12 American English vowels) for each of the speaker groups.

The Hillenbrand database can be used to highlight the difference in average formant trajectories based on age group, in addition to gender. The female and male children have very similar vowel trajectories; however, there is notably more variation and higher formant values among the female children when compared to the male children. Previously, Pettinato et al. (2016) found that the two-dimensional vowel space area, derived from the first and second formant frequency coordinates of vowels, was significantly larger for children compared to adults. In contrast, we found the female adult trajectories exhibit only slight compression and slightly lower formant values than the male children; however, the male adult trajectories exhibit a very noticeable compacting and lowering of the trajectory values compared to all groups. As expected, the trajectory arrangement of the vowels is, in general, consistent across age and gender, exhibiting only shifts in value and changes in scale. Importantly, the average trajectories are nearly identical in direction of progression across the four groups.

Hillenbrand et al. (1995) has pointed out that the frequencies of $F1$ and $F2$, taken at a single time point, are not good predictors of vowel identification results. His example, the /æ/ - /ɛ/ pair, are identified quite well by listeners despite very poor separation in static $F1/F2$ space. We note that when the vowel trajectory is considered, we find that these tokens are nearly perpendicular to each other. Similarly, /ʊ/ and /ɝ/ appear very close to one another at the temporal midpoints; however, they also exhibit trajectories that progress at $\sim 45°$ from one another. This offers an explanation for listeners' ability to accurately identify these tokens that is eluded by utilizing only midpoint measurements.

When considering the results from this experiment, is important to note several limitations. First, the Hillenbrand database, albeit widely used, is relatively small and the speakers are quite homogeneous, in that they are all from the same dialectical region of the United States. Further, the vowels utilized in the study are all spoken in the /hVd/ context, providing a single articulatory and coarticulatory context. While this database provides an important foundational ground for the study of acoustical phonetics, it provides limited ecological validity for extrapolating findings. The results of this experiment provide substantial proof of concept of this method and a point of comparison for the use of a much larger, representative database, that it utilized in the second experiment, below.

3. Experiment 2

The second study examines the average formant trajectories present in the TIMIT database (Fisher et al., 1986) for adult female and adult male speakers. The phonemes considered include vowels, similar to above, along with diphthongs, semivowels, glides, stops, fricatives, and affricates. Results are provided in the form of figures showing the average formant trajectories, as well as tables with descriptive statistics.

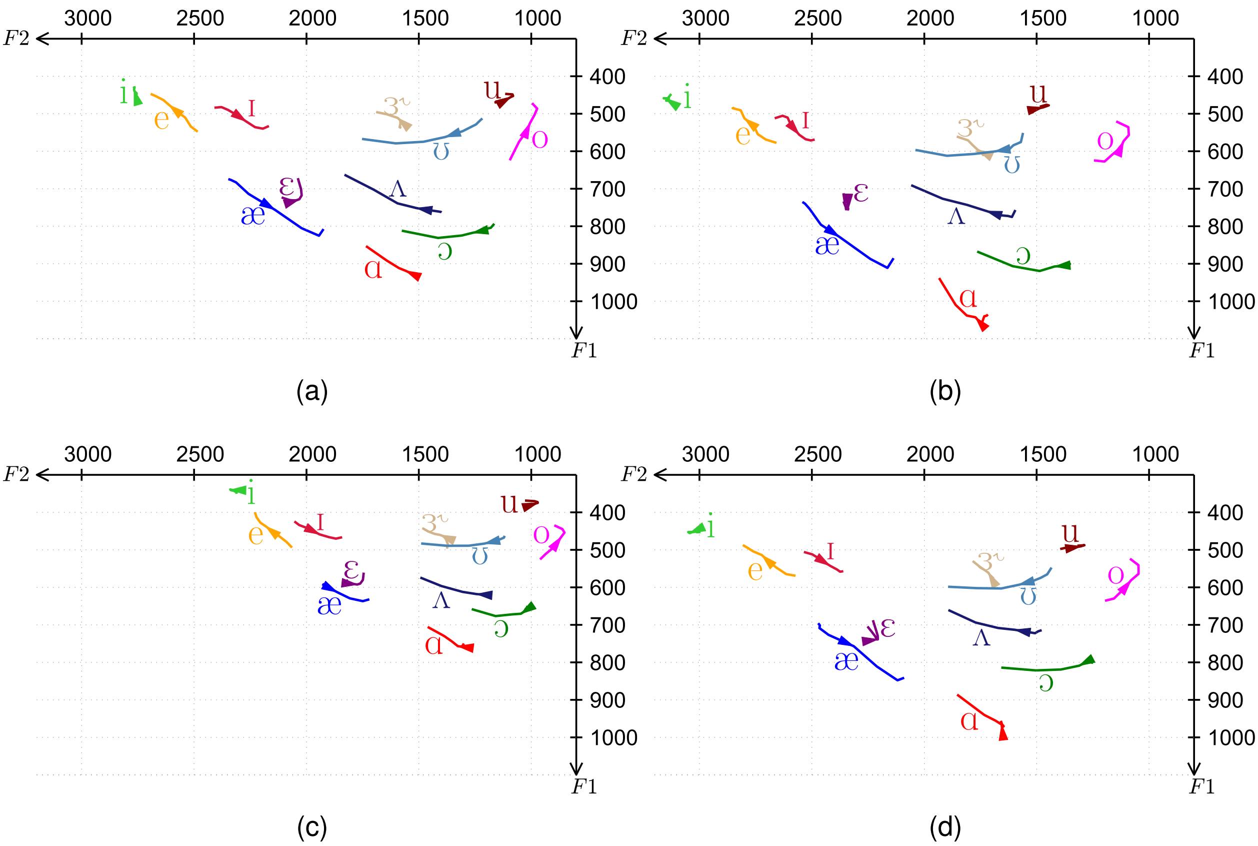


Fig. 2. The mean formant trajectories for (a) female adults; (b)female children; (c) male adults; (d) male children; taken from the Hillenbrand database. The same axis limits are used in in each of the plots to facilitate comparison and have been chosen so that the plots have the same orientation as the standard IPA vowel trapezium. Direction is indicated by an arrow (→—) which is placed at the mean $F2/F1$ value at 50% vowel duration. Note that this may not be centrally located along the length of the trajectory, thus this can be used to infer if there is more variation early in the trajectory or later in the trajectory.

3.1. Method

In order to determine the average formant trajectories for each phoneme token, three steps are necessary. First, the formant frequencies must be extracted from the acoustic signal. Second, the value of the formant frequencies must be determined at the relative temporal increments across the duration of each utterance. Finally, the average formant frequency must be computed across utterances at each of the temporal points. This is performed for a series of sounds, described in detail below. Moreover, although formants are usually only discussed in relation to vowels, if a formant merely defined as a concentration of acoustic energy around a particular frequency, then they can be similarly discussed for other phoneme types. As such, we provide the average formant trajectories for phonemes beyond vowels.

3.1.1. Speech Material

In an attempt to succeed the limitations of the Hillenbrand database, speech samples were drawn from the TIMIT (Fisher et al., 1986) database commissioned by DARPA. The

TIMIT database consists of 6300 sentences, with 10 sentences spoken by 630 speakers from 1 of 8 major dialect regions (Colby et al., 1982) of the United States. Although the database consists of only adults, it contains a wide variety of speakers. The TIMIT database includes hand verified and time-aligned orthographic and phonetic word transcriptions, as well as 16-bit, 16kHz speech waveform files for each utterance. Database design was a joint effort among the Massachusetts Institute of Technology (MIT), Stanford Research Institute (SRI) International, and Texas Instruments (TI), Inc. The speech material consists of phonetically-diverse sentences intended to expose dialectal variants of the speech. In the TIMIT database, speech material consists of sentences, in contrast to the isolated word /hVd/ productions in the Hillenbrand database. In the analysis and figures below, we have maintained the grouping of the phoneme classes (vowel, semivowel or glide, stop, fricative or affricate, nasal) specified in the TIMIT documentation. However, we have chosen to separate the diphthongs and vowel variants (rhotic, centralized, fronted, and voiceless) from the rest of the vowels to allow for more discernible figures and also to facilitate a closer comparison to the Hillenbrand database.

3.1.2. Formant Extraction

Formant extraction closely follows the procedure used in a recently presented algorithm for automatic assessment of vowel space area (Sandoval et al., 2013). A Praat (Boersma, 2001) script is used to automatically extract formant frequencies on a frame-by-frame basis. The Praat script assesses voicing on a frame-by-frame basis by estimating periodicity using an autocorrelation-based method. In this study, we only consider the first three formants; however using the recommended Praat values, 5 formants were extracted per frame below a ceiling value (5000 male, 5500 female) in Hz. Other settings were as follows: 5 ms frame advance; 50 ms analysis window; pre-emphasis starting from 50 Hz. Internally, Pratt computes estimates of the formants by resampling to twice the ceiling of the formant search range, then applying a pre-emphasis filter, windowing the speech in the time domain using a Gaussian window, and estimating the LPC coefficients using the algorithm by Burg (Childers and Kesler, 1978; Press et al., 1992).

3.1.3. Trajectory Derivation

Due to the variation in phoneme duration both across individual utterances and across speakers, we utilize time points corresponding to each utterance's relative phoneme duration to temporally capture the formant trajectory (e.g., formant values at 20 percent of phoneme duration). Using MATLAB (2014) and the meta-data provided with the TIMIT database, the start and end times of each vowel utterance were determined and used to calculate the times corresponding to 0-100% vowel duration at increments of 10%. The time corresponding to relative phoneme durations are likely to fall between the frames in which the formant frequencies are sampled (every 5 ms). As a result, we interpolate the values of the formant frequencies between analysis frames using a cubic spline in order to get more precise temporal values. Processing all input speech results in an $N \times 20$ matrix, $\mathbf{F}$, that stores all $F1$ and $F2$ pairs for a particular phoneme token at each of the 10 temporal points, where N is the number of phoneme observations.

3.1.4. Trajectory Averaging

Utterances corresponding to a particular phoneme token are collected and the mean formant values across the utterances, at each temporal point relative to the phoneme duration, were computed. This results in a mean trajectory in the $F2/F1$ space for each of the tokens in the database.

3.2. Results and Discussion

3.2.1. Vowel Formant Trajectories

Table 1 summarizes the number of occurrences for each vowel in the TIMIT database. Fig. 3(a) and (b) show the average formant trajectories for each of the vowels in the TIMIT database. Tables 2 and Table 3 show a summary of the average vowel formant values in the TIMIT database at 20%, 50%, and 80% duration for adult female and adult male speakers, respectively.

Although the arrangement of the vowel trajectories is similar in both the Hillenbrand and the TIMIT databases, there are some key differences. Particularly, the trajectories in the TIMIT database exhibit more of a curved trajectory and are more tightly arranged with smaller average $F1$ and $F2$ values. It is not apparent whether these differences result from speaker dialect or coarticulation effects, or the difference in methods for formant computation. Unlike the vowels in the Hillenbrand database, which had some formant values very close to one another, here each of the vowels appear to have a distinct region of occurrence. As expected, the male trajectories exhibit a very noticeable compacting and lowering of the trajectory values, compared to both women and young children in either database.

3.2.2. Diphthong and Vowel Variant Formant Trajectories

Table 4 summarizes the number of occurrences for diphthongs and vowel variants in the TIMIT database. Fig. 3(c) and (d) show the average formant trajectories for each of the diphthongs and vowel variants in the TIMIT database overlaid on the vowel trajectories from the same database [originally shown in Fig. 3(a) and (b)] for adult female and adult male speakers, respectively. Tables 5 and 6 show a summary of the average diphthong and vowel variant formant values in the TIMIT database at 20%, 50%, and 80% duration for adult female and adult male speakers, respectively. We note that due to a lack of a standard IPA symbol for fronting, /u̳/ has been used to denote a fronted allophone of /u/.

In general, the female and male trajectories are in agreement; however, the male trajectories exhibit a very noticeable compacting and lowering of formant values. Additionally, there were some noticeable differences in the shape and direction of formant trajectories for male and females. For example: /ɪ̈/ resembled a upward angled cup for females (◡) and a downward angled cup for males (◠); /ə̥/ is mostly one directional for males, while distinctly two directional for females; /ɝ/ and /ɚ/ start and end closer to the center of the vowel space for males than females.

Again, we observe that similar to the trajectory of the Hillenbrand study vowels, tokens that are close in $F2/F1$ space travel in different directions. For example, /aɪ/ and /aʊ/ have very close formant values especially at the temporal midpoint. However, they traverse in oppo-

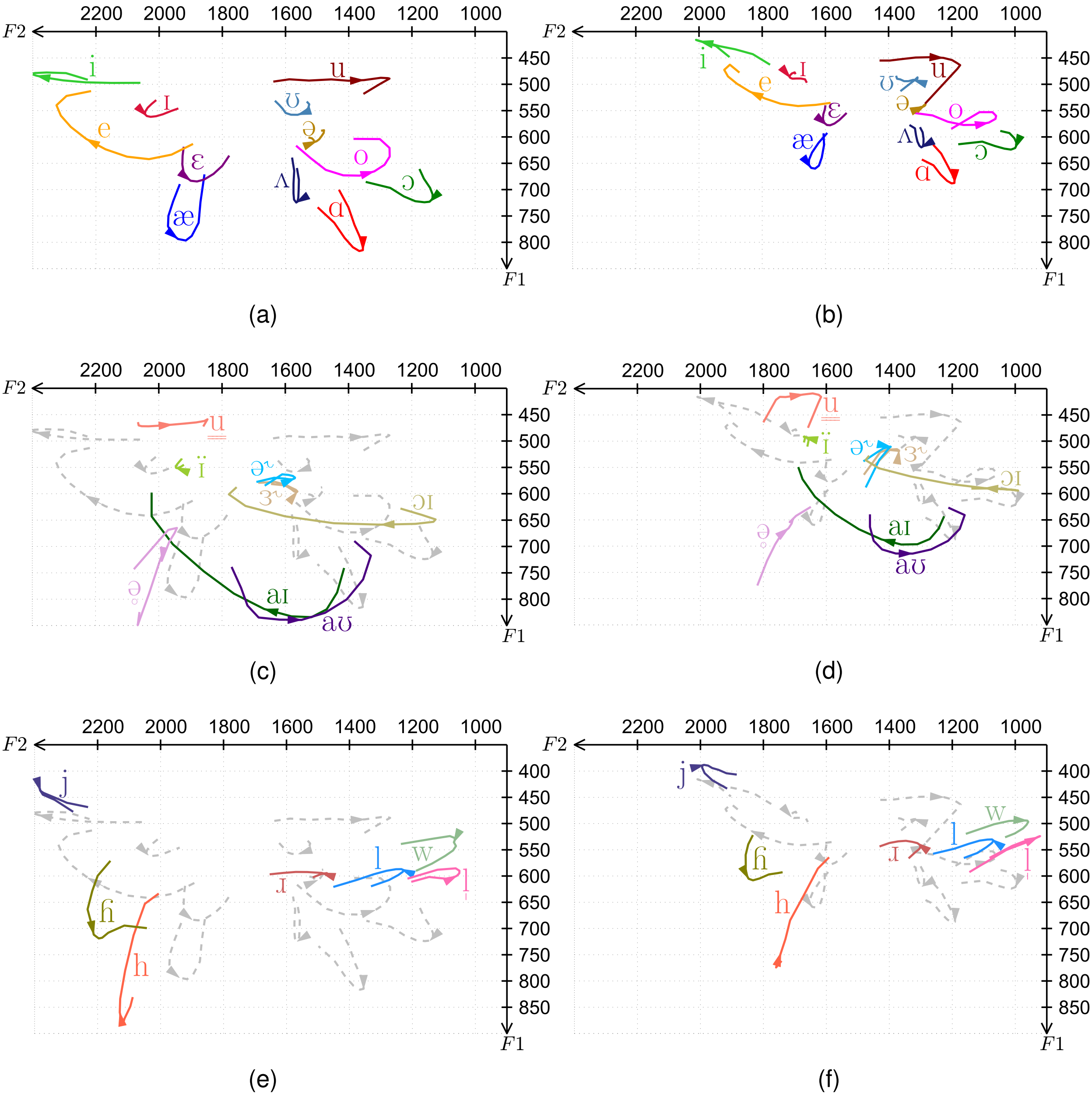
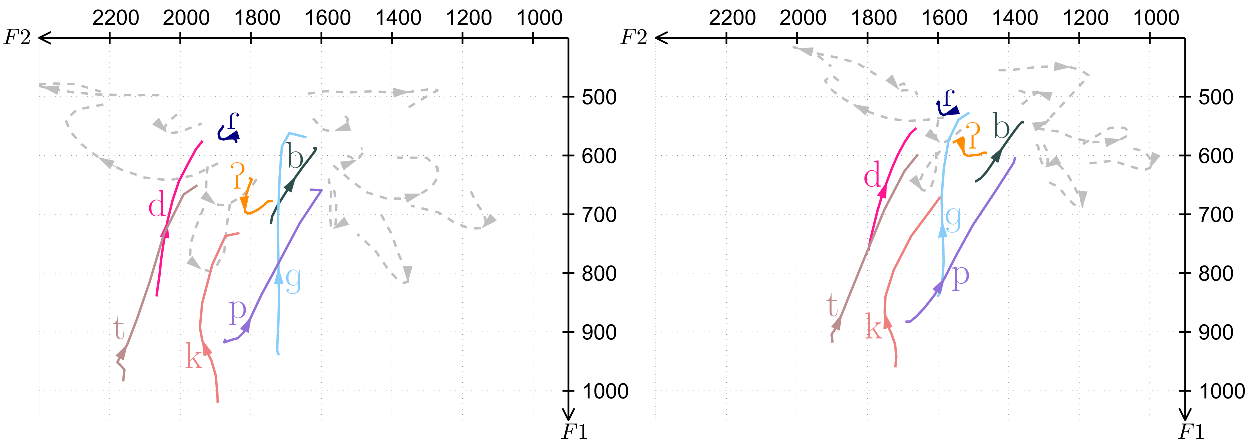


Fig. 3. The average formant trajectories in the TIMIT database for adult (a) female vowels; (b) male vowels; (c) female diphthongs and vowel variants; (d) male diphthongs and vowel variants; (e) female semivowels and glides; (f) male semivowels and glides. Direction is indicated by an arrow (→) which is placed at the mean $F2/F1$ value at 50% vowel duration. Note that this may not be centrally located along the length of the trajectory, thus this can be used to infer if there is more variation early in the trajectory or later in the trajectory. Plots (c), (e) and (d), (f) have been overlaid for comparison on the vowels from (a) and (b), respectively, and are displayed using a grey dashed line (- - -).



(a)

(b)

Fig. 4. The average stop formant trajectories for adult (a) female; (b) male; speakers in the TIMIT database. The stops have been overlaid on the vowels from Figure 3 and are displayed using a grey dashed line (- - -). Direction is indicated by an arrow (→—) which is placed at the mean $F2/F1$ value at 50% vowel duration. Note that this may not be centrally located along the length of the trajectory, thus this can be used to infer if there is more variation early in the trajectory or later in the trajectory.

site directions and the formant values of /aɪ/ have overall greater deviation from the temporal midpoint.

3.2.3. Semivowel and Glide Formant Trajectories

Table 7 summarizes the number of occurrences for semivowels and glides in the TIMIT database. Fig. 3(e) and (f) shows the average formant trajectories for each of diphthongs and vowel variants in the TIMIT database overlaid on the vowel trajectories [originally shown in Fig. 3(a) and (b)] for adult female and adult male speakers, respectively. Tables 8 and 9 show a summary of the average semivowel and glide formant values in the TIMIT database at 20%, 50%, and 80% duration for adult female and adult male speakers, respectively.

Again the female and male trajectories are very similar with the male trajectories exhibiting a very noticeable compacting and lowering of the trajectory values. Similar to the Hillenbrand vowels, tokens that are close in $F2/F1$ space travel in different directions. For example, /h/ and /ɦ/ are relatively close in the $F2/F1$ space, but traverse in opposite directions. The same can be said about /l̩/ and /w/.

3.2.4. Stop Formant Trajectories

Table 13 summarizes the number of occurrences of stops in the TIMIT database. Fig. 4 shows the average formant trajectories for each of stops in the TIMIT database overlaid on the vowel trajectories from the same database [originally shown in Fig. 3(a) and (b)]. Tables 14 and 15 show a summary of the average stop formant values in the TIMIT database at 20%, 50%, and 80% duration for adult female and adult male speakers, respectively.

Again the female and male trajectories are very similar with the male trajectories exhibiting a very noticeable compacting and lowering of the trajectory values. The average formant

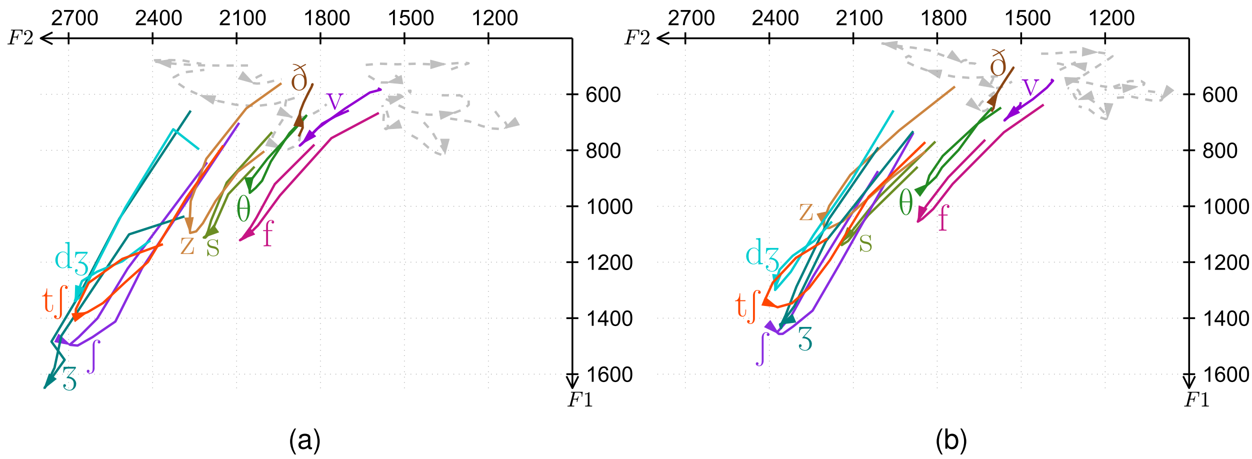


Fig. 5. The average fricative and affricate formant trajectories for adult (a) female; (b) male; speakers in the TIMIT database. The diphthong and vowel variants have been overlaid on the vowels from Figure 3(a) and (b) respectively and are displayed using a grey dashed line (---). Direction is indicated by an arrow (→) which is placed at the mean $F2/F1$ value at 50% vowel duration. Note that this may not be centrally located along the length of the trajectory, thus this can be used to infer if there is more variation early in the trajectory or later in the trajectory.

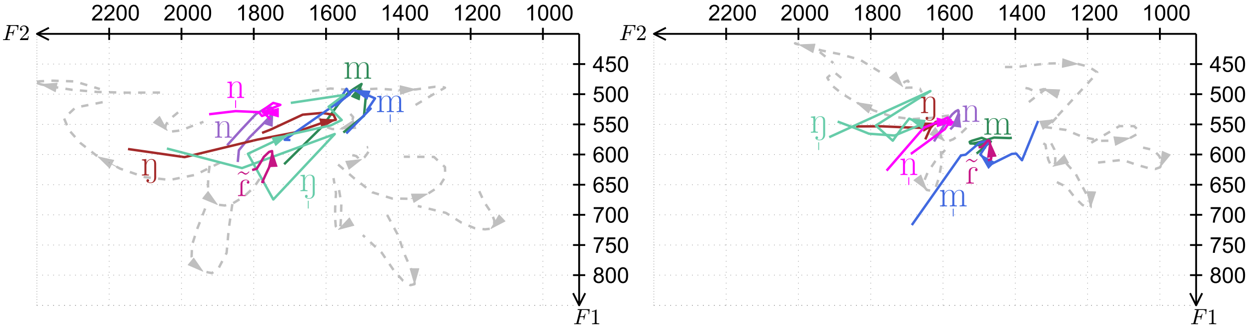
trajectories of stop phonemes seem to appear in two categories: 1) /d/, /g/, /p/, /t/, and /k/ all begin with rather large $F21$ and $F2$ values which decrease significantly during the duration of the phoneme; and 2) /ɾ/, /ʔ/, and /b/ are located in the frequency range of the vowel trajectories and exhibit a relatively small amount of movement during the duration of the phoneme compared to other stop consonants.

3.2.5. Fricative and Affricate Formant Trajectories

Table 10 summarizes the number of occurrences for fricatives and affricates in the TIMIT database. Fig. 5 shows the average formant trajectories for each of fricatives and affricates in the TIMIT database overlaid on the vowel trajectories from the same database [originally shown in Fig. 3(a) and (b)]. Tables 11 and 12 show a summary of the average fricative and affricate formant values in the TIMIT database at 20%, 50%, and 80% duration for adult female and adult male speakers, respectively.

Most fricatives and affricates exhibit a positive swing in both $F1$ and $F2$, which is expected because these phonemes are traditionally characterized by relatively high frequency noise. Unlike the previous phoneme types, which exhibit a very noticeable compacting and lowering of the formant values for the male trajectories, this trend is less robust for fricatives and affricates. We conjecture that this is because the predominate determiner of fricative and affricate acoustics is the manner of articulation and place of constriction; this is in contrast to other phonemes (namely vowels), for which variation of the overall vocal tract results in these differing characteristics. In other words, affricates and fricatives are possibly less influenced by the differences in male and female anatomies.

The fricative and affricate average formant trajectories seem to appear in three categories: 1) /v/ and /ð/ have all $F1$ values less than 800 Hz and all $F2$ values less then 1900 Hz; 2) /z/,



(a) (b)

Fig. 6. The average nasal formant trajectories for adult (a) female; (b) male; speakers in the TIMIT database. The diphthong and vowel variants have been overlaid on the vowels from Figure 3(a) and (b) respectively and are displayed using a grey dashed line (- - -). Direction is indicated by an arrow (→) which is placed at the mean $F2/F1$ value at 50% vowel duration. Note that this may not be centrally located along the length of the trajectory, thus this can be used to infer if there is more variation early in the trajectory or later in the trajectory.

/s/, /θ/, and /f/ have all $F1$ values less than 1200 Hz and all $F2$ values less then 2300 Hz; 3) /ʃ/, /ʒ/, /dʒ/, and /tʃ/ have all $F1$ values less than 1700 Hz and all $F2$ values less then 2800 Hz. Most of the fricative and affricates begin with relatively low $F1$ and $F2$ values which rapidly increase to a maximum near the temporal mid point, before then rapidly falling and returning to lower $F1$ and $F2$ values. This is in stark contract to the stop formant trajectories where most of the phonemes begin with large $F1$ and $F2$ values that decrease during the duration of the phoneme. Also, unlike previous phoneme types considered, there is considerable overlap in the formant trajectories of phonemes within the class of fricatives and affricates. Interestingly, most of the overlapping trajectories have similar progressions and do not diverge in different directions.

3.2.6. Nasal Formant Trajectories

Table 16 summarizes the number of occurrences of nasals in the TIMIT database. Fig. 6 shows the average formant trajectories for each of the nasals in the TIMIT database overlaid on the vowel trajectories from the same database [originally shown in Fig. 3(a) and (b)]. Tables 17 and 18 show a summary of the average nasal formant values in the TIMIT database at 20%, 50%, and 80% duration for adult female and adult male speakers, respectively.

Unlike the rest of the phoneme types considered in this report, the configuration of nasal trajectories is substantially different when comparing female and male speakers. Only /r̃/ seems to appear with some consistency in the two speaker groups. This may be the case because /r̃/ is not a formal nasal but rather a nasalized flap. We conjecture that this is secondary to the retention of the stop-like qualities of the flap, as this is consistent with the patterns seen when examining the non-nazalized version of this stop consonant. Furthermore, we conjecture that the substantial variation of the rest of the nasal trajectories results from the fact that the predominate determiner of nasal quality, the nasal cavity, cannot be reconfigured like the rest of the vocal tract, and, as a result, could exacerbate the speaker dependence of these sounds.

3.2.7. Three-dimensional Trajectories utilizing $F3$

In this study, $F3$ values were computed but not reported, primarily due to the limitations of displaying 3-Dimensional (3-D) data using 2-Dimensional (2-D) media; nevertheless, $F3$ values have been found to be useful for distinguishing certain phoneme types, e.g., rhotic vowels and velar consonants. With this in mind, we have provided animated visualizations of the formant trajectories in 3-D space using the first three formants ($F2/F1/F3$). The 3-D average formant trajectories for females and males are given in Vid. 1 and Vid. 2, respectively. These illustrations utilize all of the phonemes previously considered using the TIMIT database (Experiment 2). They are plotted in a similar fashion to the previous figures, but in 3-D space with labels omitted.

As the illustration shows, most of the phonemes lie very close to a 2-D hyperplane of the 3-D space. This simple representation, while somewhat lacking in mathematical scrutiny, shows a general lack of independence between the first three formants, and suggests that inclusion of $F3$ is superfluous for many, but not all, phonemes. Markedly, /ɝ/, /ɚ/, and /ɹ/ appear with drastically lower $F3$ values than other phonemes with similar $F2/F1$ values. There is also a noticeable lowering in $F3$ at the start of /g/ and /k/, and the the end of /ɑ/. Likewise, /l/, /l̩/, and /j/ appear with observable larger $F3$ values than other phonemes with similar $F2/F1$ values. A relative increase in $F3$ values is also true for /z/ and /s/; interestingly, the extent of this difference is far exacerbated in the trajectories of the male compared to the female speakers.

4. Conclusion

Since the introduction of formant analysis, Peterson and Barney (1952) found that increased crowding of vowels in static $F2/F1$ space was not accompanied by an increase in perceptual confusions among vowels. Hillenbrand et al. (1995) speculated that this could be explained by spectral change, which further supports the idea that formant trajectories are important for phoneme perception. This is further elucidated as we consider the formant trajectories of phonemes other than vowels. Even though phonemes can appear considerably crowded in $F2/F1$ space, most have distinct trajectories across this space during the duration of the production, presumably lending to accurate perception.

Although the effectiveness of spectral information provided by the first two formant frequencies in vowel identification is indisputable, it has also been recognized that temporal information provides additional cues. Static measurements alone do not explain why vowels are perceived correctly despite having similar temporal midpoints; however, it is the change across time that provides insight into this perceptual process. We have illustrated that is holds true for other phoneme types as well. Furthermore, the use of duration as an additional feature to differentiate between phoneme with closely spaced $F2/F1$ values is common. Although the increase in classification performance when including duration cannot be denied, this does not imply that duration is the best or even the most relevant way to discriminate between these sound types. For example, vowel duration is sensitive to speaking rate, where as a formant trajectory computed relative to vowel duration, as performed in this study, is not. This suggests that formant trajectories may be a measurement robust to dialectical variations or pathological changes in rate, while still capturing variations in phoneme productions. Furthermore, because $F1/F2$ values roughly relate to jaw/tongue excursion, the use of formant trajectories as a proxy for kinematic movement may be useful as a means to track improvement of therapy or progression of disease for pathological speakers. This could further be validated in experiments to determine how trajectories with reduced/increased variation relate to perceptual errors and the communication disorder associated with such changes in formant trajectories.

The current study utilized two different databases, with varying contexts: isolated vowels from the Hillenbrand database and vowels taken from productions of sentences in the TIMIT database. It is possible the differences seen between the TIMIT and Hillenbrand trajectories are due to contextual differences or coarticulation effects, or the difference in methods for formant computation. The significance of these differences is uncertain and should be explored in further research to ascertain true differences of formant trajectories among children and adults of different genders. Other topics include more closely examining the influence of regional dialect on the vowel trajectories, or comparing individual trajectories to an average population trajectory.

It is acknowledged that some of the methodology used as part of this study could be criticized. In particular, automated formant extraction was not hand verified or individually optimized due to the vast amount to speech material and speakers utilized. Also, raw formant values in Hertz were directly averaged rather than averaging the value subsequent to formant normalization. Nevertheless, we believe that the reported values are well representative of formant trajectories in American English, and can serve as a basis for progressing the investigation of formant trajectories in acoustical phonetics.

Bibliography

Ainsworth, W. A., 1972. Duration as a cue in the recognition of synthetic vowels. J. Acoust. Soc. Am. 51 (2B), 648–651.

Assmann, P. F., Katz, W. F., 2000. Time-varying spectral change in the vowels of children and adults. J. Acoust. Soc. Am. 108 (4), 1856–1866.

Benedetto, M.-G. D., 1989. Vowel representation: Some observations on temporal and spectral properties of the first formant frequency. J. Acoust. Soc. Am. 86 (1), 55–66.

Bennett, D. C., 1968. Spectral form and duration as cues in the recognition of english and german vowels. Lang. Speech 11 (2), 65–85.

Bernard, J. R., 1981. Australian pronunciation. The Macquarie Dictionary 18, 27.

Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot International 5 (9/10), 341–345.

Broad, D. J., Clermont, F., 1987. A methodology for modeling vowel formant contours in cvc context. J. Acoust. Soc. Am. 81 (1), 155–165.

Broad, D. J., Clermont, F., 2002. Linear scaling of vowel-formant ensembles (vfes) in consonantal contexts. Speech Commun. 37 (3), 175–195.

Broad, D. J., Clermont, F., 2010. Target–locus scaling methods for modeling families of formant transitions. J. Phon. 38 (3), 337–359.

Broad, D. J., Fertig, R. H., 1970. Formant-frequency trajectories in selected cvc-syllable nuclei. J. Acoust. Soc. Am. 47 (6B), 1572–1582.

Childers, D. G., Kesler, S. B., 1978. Modern spectrum analysis. Vol. 331. IEEE Press New York.

Colby, C. J., Wallace, R., Jolly, C. (Eds.), 1982. Language Files. Ohio State University Press.

Cox, F., 1996. An acoustic study of vowel variation in australian english. Ph.D. thesis, Macquarie University.

Cox, F., 1998. The bernard data revisited. Aust. J. Linguis. 18 (1), 29–55.

Delattre, P., Liberman, A. M., Cooper, F. S., Gerstman, L. J., 1952. An experimental study of the acoustic determinants of vowel color; observations on one-and two-formant vowels synthesized from spectrographic patterns. Word.

Essner, C., 1947. Recherche sur la structure des voyelles orales. Archives Néerlandaises de Phonétique Expérimentale 20, 40–77.

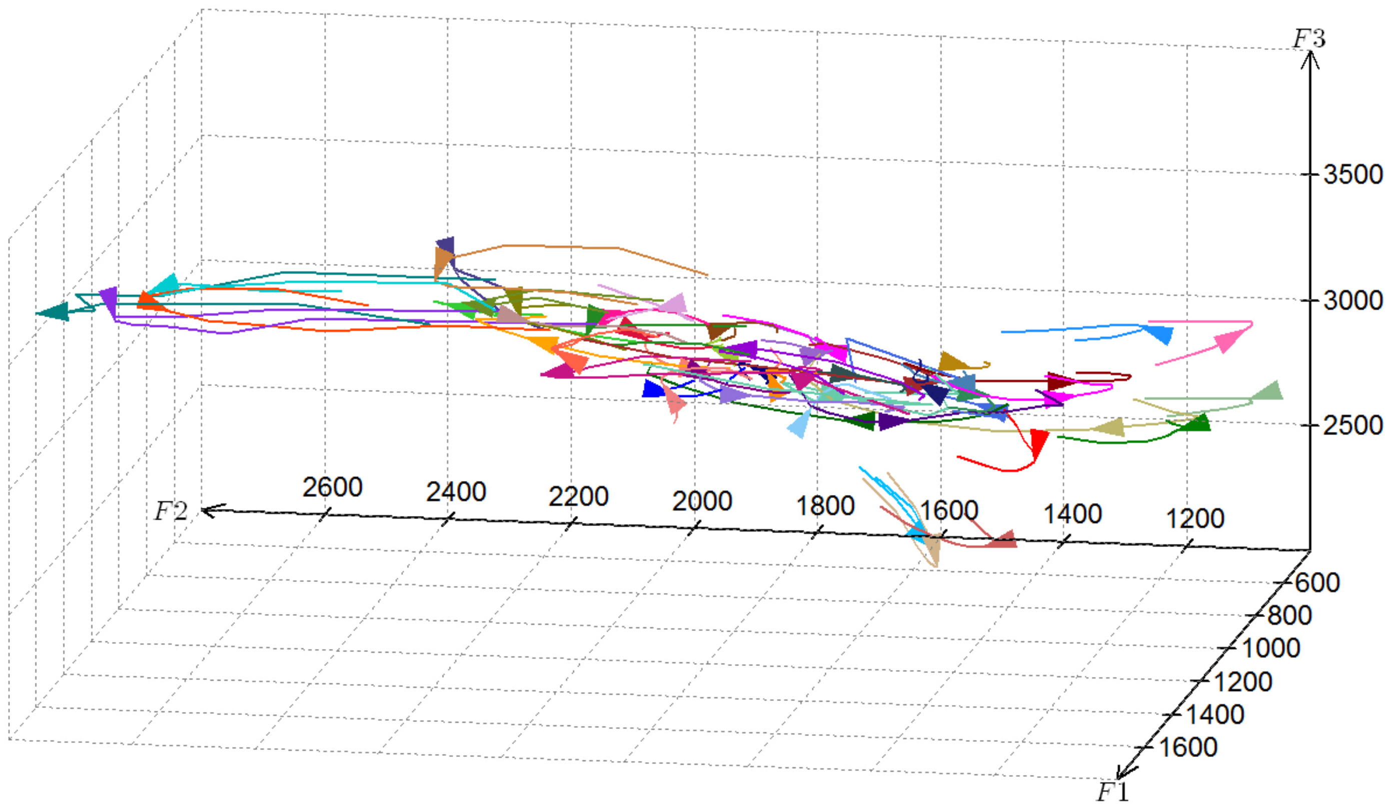
Fant, G., 1960. Acoustic theory of speech production. Gravenhage: Mouton and Co.

Fant, G., 1973. Speech sounds and features. The MIT Press.

Fisher, W. M., Doddington, G. R., Goudie-Marshall, K. M., 1986. The DARPA Speech Recognition Research Database: Specifications and Status. In: Proceedings of DARPA Workshop on Speech Recognition. pp. 93–99.

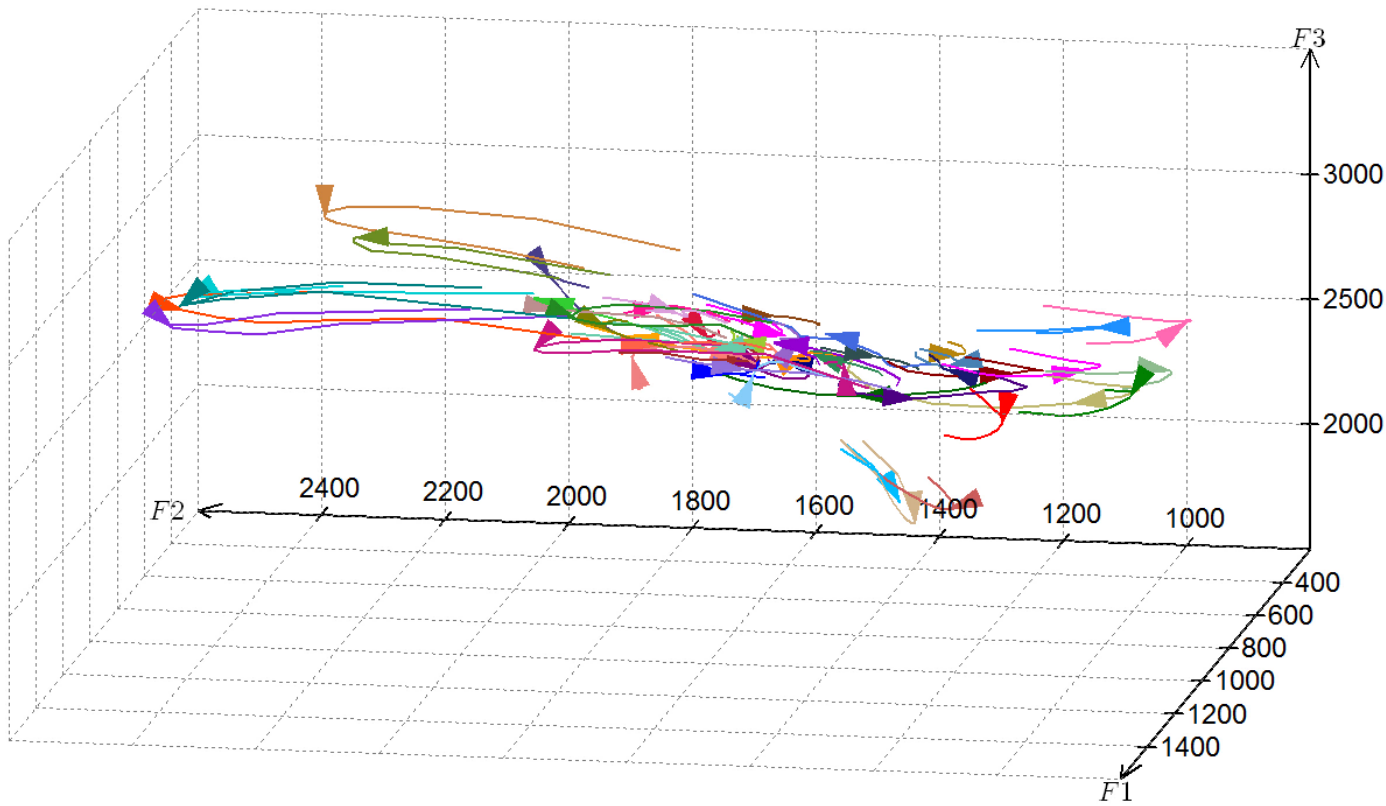
Fox, R. A., Jacewicz, E., 2009. Cross-dialectal variation in formant dynamics of american english vowels. J. Acoust. Soc. Am. 126 (5), 2603–2618.

Harrington, J., Cassidy, S., 1994. Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in australian english. Lang. Speech 37 (4), 357–373.
Harrington, J., Cox, F., Evans, Z., 1997. An acoustic analysis of cultivated, general and broad australian english speech. Aust. J. Linguis. 17, 155–84.
Hillenbrand, J., Gayvert, R. T., 1993a. Identification of steady-state vowels synthesized from the peterson and barney measurements. J. Acoust. Soc. Am. 94 (2), 668–674.
Hillenbrand, J., Gayvert, R. T., 1993b. Vowel classification based on fundamental frequency and formant frequencies. J. Speech Lang. Hear. Res. 36 (4), 694–700.
Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K., 1995. Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97 (5), 3099–3111.
Hillenbrand, J. M., 2013. Static and dynamic approaches to vowel perception. In: Vowel inherent spectral change. Springer, pp. 9–30.
Huang, C. B., 1986. The effect of formant trajectory and spectral shape on the tense/lax distinction in american vowels. In: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. Vol. 11. IEEE, pp. 893–896.
Huang, C. B., 1992. Modelling human vowel identification using aspects of formant trajectory and context. Speech perception, production and linguistic structure, 43–61.
Jenkins, J. J., Strange, W., Edman, T. R., 1983. Identification of vowels in “vowelless” syllables. Percept. Psychophys. 34 (5), 441–450.
Joos, M., 1948. Acoustic phonetics. Lang. 24 (2), 5–136.
Kewley-Port, D., Neel, A., 2006. Perception of dynamic properties of speech: Peripheral and central processes. Listening to Speech: An Auditory Perspective, 49.
Klatt, D. H., 1980. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67 (3), 971–995.
Klein, W., Plomp, R., Pols, L. C., 1970. Vowel spectra, vowel spaces, and vowel identification. J. Acoust. Soc. Am. 48 (4), 999–1009.
Ladefoged, P., 1972. Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech. Vol. 15. Oxford University Press.
Lee, J., Shaiman, S., 2012. Relationship between articulatory acoustic vowel space and articulatory kinematic vowel space. J. Acoust. Soc. Am. 132 (3), 2003.
Lehiste, I., Peterson, G., 1961. Some basic considerations in the analysis of intonation. J. Acoust. Soc. Am. 33 (4), 419–425.
Lippmann, R. P., 1989. Review of neural networks for speech recognition. Neural Comput. 1 (1), 1–38.
MATLAB, 2014. version 8.3.0.532 (R2014a). The MathWorks, Inc., Natick, Massachusetts.
McDougall, K., Nolan, F., 2007. Discrimination of speakers using the formant dynamics of/u/in british english. In: Proc. Int. Congr. Phonetic Sci. pp. 1825–8.
Miller, J. D., 1989. Auditory-perceptual interpretation of the vowel. J. Acoust. Soc. Am. 85 (5), 2114–2134.
Morrison, G. S., 2009. Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongsa). J. Acoust. Soc. Am. 125 (4), 2387–2397.
Morrison, G. S., Assmann, P. F., 2012. Vowel inherent spectral change. Springer Science & Business Media.
Nearey, T. M., 1978. Phonetic feature systems for vowels. Vol. 77. Indiana University Linguistics Club.
Nearey, T. M., 1989. Static, dynamic, and relational properties in vowel perception. J. Acoust. Soc. Am. 85 (5), 2088–2113.
Nearey, T. M., 1992. Applications of generalized linear modeling to vowel data. In: Int. Conf. Spoken Lang. Process.
Nearey, T. M., 2013. Vowel inherent spectral change in the vowels of north american english. In: Vowel inherent spectral change. Springer, pp. 49–85.
Nearey, T. M., Assmann, P. F., 1986. Modeling the role of inherent spectral change in vowel identification. J. Acoust. Soc. Am. 80 (5), 1297–1308.
Nearey, T. M., Hogan, J., Rozsypal, A., 1979. Speech signals, cues and features. Perspect. Exp. linguist., 73–96.
Neel, A. T., 2004. Formant detail needed for vowel identification. Acoust. Res. Lett. Online 5 (4), 125–131.
O’Shaughnessy, D., 1987. Speech communication: human and machine. Addison-Wesley Pub. Co.
Peterson, G. E., Barney, H. L., Mar. 1952. Control Methods Used in a Study of the Vowels. J. Acoust. Soc. Am. 24 (2), 175–184.
Pettinato, M., Tuomainen, O., Granlund, S., Hazan, V., 2016. Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. J. Phonetics 54, 1–14.
Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 1992. Numerical recipes in C (2nd ed.): the art of scientific computing. Cambridge University Press, New York, NY, USA.
Quatieri, T. F., 2002. Discrete-Time Speech Signal Processing. Prentice Hall.
Sandoval, S., Berisha, V., Utianski, R., Liss, J., Spanias, A., 2013. Automatic assessment of vowel space area. J. Acoust. Soc. Am. 134 (5), EL477–EL483.
Stevens, K. N., Kasowski, S., Fant, G., 1953. An electrical analog of the vocal tract. J. Acoust. Soc. Am. 25 (4), 734–742.
Strange, W., 1989a. Dynamic specification of coarticulated vowels spoken in sentence context. J. Acoust. Soc. Am. 85 (5), 2135–2153.
Strange, W., 1989b. Evolving theories of vowel perception. J. Acoust. Soc. Am. 85 (5), 2081–2087.
Strange, W., Jenkins, J. J., Johnson, T. L., 1983. Dynamic specification of coarticulated vowels. J. Acoust. Soc. Am. 74 (3), 695–705.
Syrdal, A. K., 1985. Aspects of a model of the auditory representation of american english vowels. Speech Commun. 4 (1), 121–135.
Syrdal, A. K., Gopal, H. S., 1986. A perceptual model of vowel recognition based on the auditory representation of american english vowels. J. Acoust. Soc. Am. 79 (4), 1086–1100.
Watson, C. I., Harrington, J., 1999. Acoustic evidence for dynamic formant trajectories in australian english vowels. J. Acoust. Soc. Am. 106 (1), 458–468.
Whalen, D. H., 1989. Vowel and consonant judgments are not independent when cued by the same information. Percept. Psychophys. 46 (3), 284–292.
William, T. R., 1953. Vowel recognition as a function of duration, frequency modulation and phonetic context. J.Speech Hear. Disord. 18 (3), 289–301.
Zahorian, S. A., Jagharghi, J. A., 1993. Spectral-shape features versus formants as acoustic correlates for vowels. J. Acoust. Soc. Am. 94 (4), 1966–1982.



Vid. 1. The average 3-D vowel formant trajectories for adult female speakers in the TIMIT database. Direction is indicated by an arrow (→—) which is placed at the mean $F2/F1/F3$ value at 50% vowel duration. Note that this may not be centrally located along the length of the trajectory, thus this can be used to infer if there is more variation early in the trajectory or later in the trajectory. Observe that many of the phonemes lie approximately on a 2-D hyperplane of the 3-D space.

(Animated in online version)



Vid. 2. The average 3-D vowel formant trajectories for adult male speakers in the TIMIT database. Direction is indicated by an arrow (→) which is placed at the mean $F2/F1/F3$ value at 50% vowel duration. Note that this may not be centrally located along the length of the trajectory, thus this can be used to infer if there is more variation early in the trajectory or later in the trajectory. Observe that many of the phonemes lie approximately on a 2-D hyperplane of the 3-D space.

(Animated in online version)

Table 1: Number of vowel occurrences in TIMIT database.

Token	/æ/	/ɑ/	/ɔ/	/ɛ/	/e/	/ʊ/	/ɪ/	/i/	/o/	/ə/	/ʌ/	/u/
Female	1651	1335	1152	1593	935	256	2258	3057	860	1369	1031	199
Male	3753	2859	2942	3700	2152	500	4498	6604	2051	3584	2152	524
Total	5404	4194	4094	5293	3087	756	6756	9661	2911	4953	3183	723

Table 2: The mean, μ, and standard deviation, σ, for the average vowel formant trajectories for the female speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/æ/	750	87	794	86	763	93	1971	274	1940	250	1875	255
/ɑ/	767	93	815	83	792	78	1382	225	1355	155	1413	181
/ɔ/	697	107	723	105	713	90	1143	201	1144	162	1225	192
/ɛ/	653	80	683	75	670	78	1927	287	1900	246	1828	270
/e/	642	76	604	77	540	78	2032	280	2231	258	2325	273
/ʊ/	545	64	558	71	546	68	1528	312	1562	289	1609	297
/ɪ/	544	79	560	79	557	81	2036	310	2039	274	1988	289
/i/	496	74	484	70	478	78	2234	310	2388	264	2356	309
/o/	662	69	665	81	621	97	1473	274	1319	231	1271	252
/ə/	606	91	608	93	592	94	1512	256	1501	247	1479	262
/ʌ/	701	86	724	86	689	89	1559	237	1566	196	1575	218
/u/	493	53	492	71	489	77	1526	325	1352	281	1271	261

Table 3: The mean, μ, and standard deviation, σ, for the average $F1/F2$ formant trajectories for the male speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/æ/	629	59	659	58	639	70	1642	179	1645	157	1611	171
/ɑ/	653	70	687	62	672	59	1215	183	1192	125	1235	135
/ɔ/	602	75	621	75	617	71	1002	202	999	160	1064	164
/ɛ/	557	59	577	57	568	61	1607	206	1595	177	1556	202
/e/	541	57	514	56	473	60	1687	212	1840	173	1920	188
/ʊ/	487	78	492	61	491	75	1309	282	1322	243	1363	256
/ɪ/	480	71	488	63	489	72	1706	234	1711	206	1678	228
/i/	434	82	418	75	420	92	1885	230	1999	197	1984	221
/o/	571	59	573	70	553	92	1213	207	1083	174	1071	213
/ə/	543	81	542	79	538	89	1297	203	1288	191	1284	227
/ʌ/	603	66	620	64	597	68	1289	196	1296	148	1313	168
/u/	452	104	450	93	464	114	1359	263	1227	226	1174	246

Table 4: Number of diphthong and vowel variant occurrences in TIMIT database.

Token	/aɪ/	/aʊ/	/ɜ˞/	/ə˞/	/u̳/	/ï/	/ə̥/	/ɔɪ/
Female	998	298	952	1339	750	3603	88	292
Male	2243	647	1894	3451	1738	7979	405	655
Total	3241	945	2846	4790	2488	11582	493	947

Table 5: The mean, μ, and standard deviation, σ, for the average diphthong and vowel variant formant trajectories for the female speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/aɪ/	819	83	819	96	696	106	1469	167	1667	191	1955	263
/aʊ/	812	95	839	87	763	102	1720	264	1548	214	1354	205
/ɜ˞/	591	73	596	72	583	83	1581	250	1562	206	1599	225
/ə˞/	571	80	571	73	564	84	1605	260	1570	232	1590	262
/u̳/	472	57	469	58	462	61	2054	267	1957	265	1856	282
/ï/	548	83	551	84	539	86	1929	288	1945	276	1935	293
/ə̥/	789	423	727	474	665	415	2030	421	1991	462	1942	448
/ɔɪ/	649	78	659	64	623	68	1125	191	1299	216	1726	280

Table 6: The mean, μ, and standard deviation, σ, for the average diphthong and vowel variant formant trajectories for the male speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/aɪ/	687	63	687	71	606	84	1271	142	1425	141	1636	190
/aʊ/	690	69	714	60	667	69	1450	192	1326	157	1176	149
/ɜ˞/	518	72	518	67	516	89	1386	202	1369	160	1406	173
/ə˞/	514	94	511	86	523	124	1419	217	1395	194	1414	217
/u̳/	416	111	411	99	416	138	1748	211	1672	212	1617	239
/ï/	494	83	492	79	491	101	1658	226	1670	212	1668	232
/ə̥/	696	350	666	349	654	347	1762	357	1719	354	1716	376
/ɔɪ/	593	78	589	63	552	60	1002	201	1098	166	1433	223

Table 7: Number of semivowel and glide occurrences in TIMIT database.

Token	/l/	/ɹ/	/w/	/j/	/h/	/ɦ/	/l̩/
Female	2481	2773	1334	709	368	490	401
Male	5671	6288	3043	1640	945	1033	893
Total	8152	9061	4377	2349	1313	1523	1294

Table 8: The mean, μ, and standard deviation, σ, for the average semivowel and glide formant trajectories for the female speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/l/	598	108	587	105	597	103	1245	255	1235	256	1324	312
/ɹ/	594	118	594	111	592	99	1473	244	1488	242	1571	246
/w/	524	98	536	90	562	88	1081	280	1068	291	1096	242
/j/	445	166	438	98	453	78	2374	278	2381	267	2326	298
/h/	872	185	860	179	713	224	2114	396	2130	384	2086	456
/ɦ/	628	200	712	199	708	173	2217	435	2212	403	2166	416
/l̩/	604	74	599	73	585	79	1113	186	1059	157	1063	179

Table 9: The mean, μ, and standard deviation, σ, for the average semivowel and glide formant trajectories for the male speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/l/	545	94	530	86	539	88	1091	269	1074	240	1148	272
/ɹ/	549	130	537	110	534	105	1306	201	1311	193	1372	198
/w/	495	107	494	89	509	78	1020	365	962	336	975	278
/j/	404	175	388	118	398	96	1987	200	1988	196	1943	202
/h/	770	178	746	183	637	190	1755	319	1744	328	1672	384
/ɦ/	560	143	605	148	597	121	1856	347	1847	328	1783	334
/l̩/	535	76	525	73	532	85	975	245	927	241	950	274

Table 10: Number of fricative and affricate occurrences in TIMIT database.

Token	/s/	/ʃ/	/z/	/ʒ/	/f/	/θ/	/v/	/ð/	/dʒ/	/tʃ/
Female	3062	915	1560	57	943	324	849	1182	495	332
Male	7051	2118	3483	168	2184	694	1855	2691	1085	748
Total	10113	3033	5043	225	3127	1018	2704	3873	1580	1080

Table 11: The mean, μ, and standard deviation, σ, for the average fricative and affricate variant formant trajectories for the female speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/s/	1050	229	1112	166	1080	221	2191	273	2217	234	2184	272
/ʃ/	1401	421	1496	382	1412	401	2597	338	2694	289	2533	360
/z/	831	426	1096	311	982	342	2208	416	2266	354	2175	359
/ʒ/	1320	755	1652	491	1463	599	2664	410	2789	359	2726	358
/f/	1041	237	1122	195	962	221	2029	243	2088	190	1945	259
/θ/	830	274	952	199	840	231	1990	266	2053	216	1977	249
/v/	594	201	778	395	706	299	1623	349	1867	454	1794	382
/ð/	700	230	660	215	574	94	1865	258	1872	245	1832	223
/dʒ/	1234	394	1339	506	1005	623	2598	333	2678	308	2497	339
/tʃ/	1274	310	1378	406	1199	481	2629	245	2632	299	2417	368

Table 12: The mean, μ, and standard deviation, σ, for the average fricative and affricate formant trajectories for the male speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/s/	1078	275	1129	217	1124	259	2096	363	2141	352	2119	376
/ʃ/	1391	353	1456	317	1373	342	2321	274	2366	249	2245	322
/z/	888	441	1078	351	992	377	2111	520	2202	458	2070	445
/ʒ/	1287	651	1422	482	1357	525	2303	382	2362	298	2311	330
/f/	1000	203	1055	169	920	212	1840	227	1869	188	1745	255
/θ/	820	253	918	183	798	218	1780	275	1829	215	1723	230
/v/	576	201	690	286	658	258	1408	326	1557	400	1525	367
/ð/	637	231	589	195	515	98	1597	270	1582	238	1534	192
/dʒ/	1177	426	1299	508	1101	614	2317	309	2380	289	2241	325
/tʃ/	1273	294	1361	346	1192	440	2387	226	2371	245	2183	326

Table 13: Number of stop occurrences in TIMIT database.

Token	/b/	/d/	/g/	/p/	/t/	/k/	/ɾ/	/ʔ/
Female	943	1510	909	1124	1822	2015	1019	1862
Male	2074	3253	1856	2417	4070	4468	2629	2969
Total	3017	4763	2765	3541	5892	6483	3648	4831

Table 14: The mean, μ, and standard deviation, σ, for the average stop formant trajectories for the female speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/b/	687	236	636	193	593	155	1726	351	1673	356	1618	393
/d/	803	229	714	246	617	212	2059	266	2037	249	1981	247
/g/	915	217	792	218	585	189	1724	376	1721	390	1712	429
/p/	914	228	874	227	715	200	1855	271	1799	302	1661	394
/t/	955	201	921	281	733	283	2172	255	2150	271	2046	279
/k/	973	165	914	192	786	253	1900	395	1938	415	1909	436
/ɾ/	572	109	573	152	558	105	1844	319	1877	278	1888	296
/ʔ/	651	167	696	183	685	154	1805	425	1816	459	1764	504

Table 15: The mean, μ, and standard deviation, σ, for the average stop formant trajectories for the male speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/b/	628	219	586	187	548	161	1463	355	1417	356	1371	373
/d/	720	234	646	246	577	233	1783	236	1746	233	1695	249
/g/	808	234	713	252	574	251	1587	348	1589	351	1570	357
/p/	873	215	809	212	660	189	1660	275	1582	302	1437	364
/t/	905	215	870	284	692	279	1902	237	1873	267	1752	272
/k/	921	166	868	210	737	268	1722	362	1752	358	1676	369
/ɾ/	521	130	531	161	513	133	1554	254	1584	242	1598	246
/ʔ/	573	150	597	150	599	137	1528	364	1526	391	1485	426

Table 16: Number of nasal occurrences in TIMIT database.

Token	/m/	/n/	/ŋ/	/m̩/	/n̩/	/ŋ̩/	/ɾ̃/
Female	1701	3099	535	45	246	16	281
Male	3725	6466	1207	126	728	27	1050
Total	5426	9565	1742	171	974	43	1331

Table 17: The mean, μ, and standard deviation, σ, for the average nasal formant trajectories for the female speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/m/	503	168	483	188	517	193	1490	318	1502	335	1567	364
/n/	572	126	527	138	533	164	1812	347	1746	383	1794	381
/ŋ/	575	129	542	153	534	186	1790	585	1573	540	1665	582
/m̩/	526	188	493	178	500	256	1486	306	1523	324	1534	419
/n̩/	530	129	531	137	514	163	1786	347	1774	379	1746	426
/ŋ̩/	566	195	567	211	521	178	1576	460	1706	543	1593	534
/ɾ̃/	634	82	594	96	615	83	1767	249	1749	233	1783	244

Table 18: The mean, μ, and standard deviation, σ, for the average nasal formant trajectories for the male speakers in TIMIT at 20%, 50%, and 80% vowel duration.

	$F1$						$F2$					
	20%		50%		80%		20%		50%		80%	
Token	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
/m/	572	235	582	273	572	234	1459	394	1525	426	1493	406
/n/	540	147	528	187	534	184	1556	292	1557	343	1575	344
/ŋ/	555	132	548	162	546	172	1699	450	1615	454	1620	450
/m̩/	599	210	600	279	600	269	1398	348	1497	428	1537	433
/n̩/	543	182	546	204	543	216	1603	314	1573	355	1578	377
/ŋ̩/	565	155	540	157	552	175	1805	397	1694	413	1786	477
/ɾ̃/	602	108	578	130	585	104	1466	228	1471	218	1488	202