



Published in final edited form as:

J Commun Disord. 2018 ; 74: 74–97. doi:10.1016/j.jcomdis.2018.05.004.

Static Measurements of Vowel Formant Frequencies and Bandwidths: A Review

Raymond D. Kent and Hourii K. Vorperian

Waisman Center, University of Wisconsin-Madison

Abstract

Purpose—Data on vowel formants have been derived primarily from static measures representing an assumed steady state. This review summarizes data on formant frequencies and bandwidths for American English and also addresses (a) sources of variability (focusing on speech sample and time sampling point), and (b) methods of data reduction such as vowel area and dispersion.

Method—Searches were conducted with CINAHL, Google Scholar, MEDLINE/PubMed, SCOPUS, and other online sources including legacy articles and references. The primary search items were vowels, vowel space area, vowel dispersion, formants, formant frequency, and formant bandwidth.

Results—Data on formant frequencies and bandwidths are available for both sexes over the lifespan, but considerable variability in results across studies affects even features of the basic vowel quadrilateral. Origins of variability likely include differences in speech sample and time sampling point. The data reveal the emergence of sex differences by 4 years of age, maturational reductions in formant bandwidth, and decreased formant frequencies with advancing age in some persons. It appears that a combination of methods of data reduction provide for optimal data interpretation.

Conclusion—The lifespan database on vowel formants shows considerable variability within specific age-sex groups, pointing to the need for standardized procedures.

1.0. Introduction

Vowel formant frequencies are among the most frequently reported acoustic measures of speech and are used in a variety of applications including automatic speech recognition, studies of speech production and speech perception in various populations of speakers, and clinical assessments in a range of speech, voice, and language disorders. This review summarizes major sources of data on formant frequencies and bandwidths for American English and also addresses (a) sources of variability in these data (focusing on speech sample and time sampling point used for formant measurement), and (b) methods of data reduction such as vowel area and dispersion. Given the relatively long history of work on vowels, it may be tempting to regard the topic of vowel acoustics as basically settled and

closed in contemporary science and practice. To the contrary, Maurer (2016) wrote, “On the one hand, the existing documentation of vowel sounds hitherto published is no more than fragmentary and on the other, the methods for describing their acoustic characteristics have substantial shortcomings and limitations” (p. 84).

1.1. Formants as vowel descriptors

Speech is a fluid phenomenon, characterized by rapid changes in articulation and its acoustic product. The dynamics of speech pose great challenges to its analysis, and one solution has been to make measurements at selected time points thought to represent targets, goals, or steady states. This approach has been taken with vowel formant measurements, which have a long history in the study of speech production, especially because formant descriptions are suited to articulatory interpretations of acoustic data and are therefore fundamental to discovery of features in articulatory-acoustic conversion. In apparently independent work, Joos (1948) and Delattre (1948) were the first to show that vowels depicted as points in the articulatory diagram of the International Phonetic Alphabet were associated with their point locations in the acoustic F1-F2 space, thereby setting the stage for the articulatory-acoustic interpretation of vowels and the recovery of vocal tract shape from the formant pattern. F1 and F2 values have been used to construct an acoustic working space and to discover how this space relates to an articulatory working space based on kinematic data or an auditory decision space for vowel identification. A primary goal is to develop principles by which the acoustic working space can be used to infer the articulatory working space.

The specification of speech by its formant patterns can be challenging, even with the development of powerful methods of digital signal processing. The uncertainties of formant analysis are widely acknowledged (Bickley, 1989; Hillenbrand, Getty, Clark, & Wheeler, 1995; Ladefoged, 1967; Maurer, 2016). Almost inevitably, the search for formants is aided by knowledge of acoustic phonetics, even if this approach suffers from a fundamental circularity (Ladefoged, 1967). Formant estimation can be a process in which knowledge about likely formant locations leads to an efficient inspection of the acoustic data and can serve as a check on automatic analyses that may generate spurious formants or miss a formant altogether.

Reaching back to some of the earliest studies of vowel acoustics (Peterson & Barney, 1952), it has been the practice to represent vowels acoustically as a single point in the plane defined by the first- and second-formant frequencies. This tradition of a single-point representation can be called *static* in the sense that it ignores changes in formant frequency pattern occurring at other time points in the vowel. A static acoustic description can be given in the two-dimensional formant plane ($F1 \times F2$), three-dimensional formant space ($F1 \times F2 \times F3$), or potentially the four-dimensional formant hypercube ($F1 \times F2 \times F3 \times F4$), but the first of these is the most commonly used in in phonetics, speech science, automatic speech recognition, speech pathology, forensics, and other fields. Reliance on a static description is seen in many current metrics of vowel acoustics, such as vowel space area (considered later).

Data obtained with the static approach have the advantage of simplicity and economy, but this approach is challenged by the fact that vowel features vary with speaker, phonetic context, and speaking rate (Carré, 2009). It also is difficult to reconcile with concepts such

as vowel inherent spectral change, which is the concept that slowly varying changes in formant frequencies are associated with vowel production, even without the influence of consonantal context (Nearey & Assmann, 1986). There is substantial evidence that the dynamics of formant-frequency patterns are highly relevant to vowel perception and to the acoustic characterization of vowels in context (Hillenbrand, 2013; Lindblom, 1963; Nearey, 1989; Strange, 1989). Efforts to capture such changes over time can be categorized as *nonstatic* or *dynamic* and include 2-point representations (e.g., vowel onset and offset; Morrison & Nearey, 2007), specification of formant dynamics (Fox & Jacewicz, 2009), and indices of spectral distance and spectral angle (Jin & Liu, 2013). Although static formant data are of primary concern in this paper, both static and dynamic analyses of formants confront threats to validity and reliability of measurement associated with the factors under discussion.

1.2. Purpose

The purpose of this review paper is to (a) describe variations in the methodology used for static or single-point formant frequency or bandwidth estimation, especially the selection of speech samples and the selection of a time measurement point, (b) summarize major datasets of formant frequencies and bandwidths for American English, and (c) compile derived measures of vowel acoustics based on static formant values (e.g., vowel space area). The discussion is directed particularly toward applications in the analysis of developing and/or disordered speech, for which acoustic methods play an increasingly large role, particularly in the inference of articulatory behaviors from acoustic data. First, we consider terminology and symbolic notation used in the acoustic analysis of speech.

1.3. Terminology and symbolic notation

According to Vilain, Berthommier, and Boë (2015) the term formant was introduced by Hermann (1894) to refer to the resonance frequencies of the vocal tract. Since then, various definitions of formant have appeared in the literature on speech acoustics (Maurer, 2016; Titze et al., 2015), with two dominant definitions being: (1) a spectral peak in the radiated sound spectrum, and (2) a resonance of the vocal tract. The first definition implies direct measurement, whereas the second implies an inference or estimate based on physical measurements. The second definition is favored here, and that is why the analysis of formant frequencies is termed “estimation.” The task of estimation is complicated by several factors, two of which are quite commonly encountered. First, peaks in the output spectrum do not necessarily correspond to vocal tract resonances. An example is the occurrence of “interformant fill,” or spectral energy that falls between two adjacent formants. Another example is a very strong first harmonic that hinders detection of a closely spaced first formant. Second, vocal tract resonances are not always realized as peaks in the output spectrum, especially because (a) voices with a high f_0 (and therefore wide spacing of the harmonics) may obscure a formant location, and (b) closely spaced formants can merge to form a single peak in the spectrum. For these and other reasons, the identification of formants is not simply a matter of looking for peaks in a particular spectral analysis.

Unfortunately, symbolic notations for acoustic entities in speech analysis are not consistent. As a step toward standardization, Titze et al. (2015) recommended the following symbolic

notations. Formant frequency is symbolized as f_{F_i} , where f is the center frequency of formant F_i and i is the formant number. Hence, the first formant frequency is f_{F_1} . Formant bandwidth is symbolized as B_{F_i} where B is the bandwidth of formant F_i and i is the formant number. Hence, the bandwidth of the first formant is B_{F_1} . Although we believe these recommendations are well-founded we will follow here the notations used in the majority of the original publications on which this review is based. Accordingly, the following symbolic notations are used:

Fundamental frequency is symbolized as f_0 . Lower case is used to distinguish vocal fundamental frequency from formant frequencies.

Formant frequency is symbolized as F_i , where F is the center frequency of the formant and i is the formant number. For example, F_1 is the first formant frequency.

Bandwidth is symbolized as B_i where B is formant bandwidth and i is the formant number. For example, B_1 is the first formant bandwidth.

Harmonics of the voice source are symbolized as H_i , where i is the harmonic number beginning with H_1 as the fundamental frequency.

The *formant frequency for a particular vowel* is expressed as $F_i/x/$ where F_i is the formant frequency for formant i and x is a phonetic symbol. For example, the first-formant frequency of vowel /i/ is expressed as $F_1/i/$.

In describing target words in speech samples, we use the following conventions: V represents a vowel category and C represents a consonant category. For example, $/bVC/$ represents the phoneme /b/ followed by a vowel and a consonant. Any other symbol within virgules is a phoneme of the International Phonetic Alphabet.

2.0. Method

Searches were conducted with CINAHL, Google Scholar, MEDLINE/PubMed, SCOPUS and other online sources including legacy articles and references. The primary search items were vowels, vowel space area, vowel dispersion, formants, formant frequency, and formant bandwidth. Literature search was focused on American English but data for other languages were considered for topics such as formant bandwidth. A main objective was to describe (a) the procedures used in determination of formant frequencies and formant bandwidths, and (b) the normative data available for various age-sex combinations of speakers and approaches used to compile them. Findings on the first objective are summarized and discussed in sections 3 to 7, and the second objective in sections 8 to 10. Results and discussion are included in each of these major sections.

3.0. General Issues of Speech Recording and Analysis

Obtaining a high fidelity recording of the speech signal and appropriate application of spectral analysis are essential to the acoustic analysis of speech. Information on microphone and recorder specifications needed for speech recordings is available in several papers (Hunter, Spielman, Starr, and Popolo, 2007; Plichta, n.d.; Sustainable Heritage Network, 2006; Švec & Granqvist, 2010; Vogel, Rosen, Morgan, & Reilly, 2015). For spectral analysis

of formant pattern, the commonly used methods are the Fast Fourier Transform (FFT) and Linear Predictive Coding (LPC). Detailed discussions of these and other types of spectral analysis are provided by Fulop (2011) and Harrington and Cassidy (1999) and issues arising in vowel formant estimation are considered later in this paper. Algorithms for both FFT and LPC are included in several speech analysis systems that are available for free or low cost (Burriss et al., 2014; Llisterri, 2015), although few neutral comparisons have been reported on the efficiency and accuracy of these systems for various types of speech analysis. LPC appears to be the most commonly used analysis for formant measurement but it is vulnerable to errors in estimating both formant frequencies (Burriss et al., 2014; Vallabha & Tuller, 2002) and formant bandwidths (Burriss et al., 2014; Mehta & Wolfe, 2015). No single method is accurate for all speakers and speech samples; nor is any given set of analysis parameters suitable for all speakers and speech samples (Burriss et al, 2014; Derdemezis et al., 2016). Cepstral analysis and time-frequency reassignment may offer improvements over conventional FFT and LPC analyses for at least some speakers (Shadle, Nam, & Whalen, 2016; Story & Bunton, 2016). Discussion of pitfalls in acoustic analysis are noted in various sections of the paper, particularly Sections 6 and 7.

4.0. Selection of Speech Samples

Data on formant frequencies and other acoustic variables have been reported for vowels produced in a variety of contexts, including isolated sustained phonation; syllables or words with a /hVd/ or /bVC/ phonetic structure (with or without a carrier phrase); word lists of different types; and connected speech using sentences, reading passages, or conversational samples. No single type of sample is likely to be ideal for all purposes, and different samples may yield different results for some acoustic measures of vowel production such as vowel space area (Chesworth, Coté, Shaw, Williams, & Hodge, 2003; Sandoval, Berisha, Utianski, Liss, & Spanias, 2013). Isolated, sustained vowels sometimes have been used to determine the vowel target, or the vowel uninfluenced by phonetic context and speaking rate (Joos, 1948). Use of words is complicated by the effect of phonetic context on vowel properties (Hillenbrand, Clark, & Nearey, 2001; Stevens & House, 1965), and even the phonological context of a vowel (Munson & Pearl-Solomon, 2004). Therefore, in selecting single-word or monosyllabic samples for test words, the glottal /h/ has been recommended as a neutral context because it minimizes supraglottal articulations that might influence vowel characteristics. Studies have confirmed that it is indeed neutral for vowel articulation compared to other contexts (Chesworth, Cote, Shaw, Williams, & Hodge, 2003; Robb & Chen, 2009). The study of vowels in the /hVd/ context is a major source of data for the vowels of American English, beginning with the seminal study by Peterson and Barney (1955) and continuing to the more recent studies of Hillenbrand et al. (1995) and Perry, Ohde, and Ashmead (2001). In their study of children's speech, S. Lee, Potamianos, and Narayanan (1999) used the /bVC/ target words *bead*, *bit*, *bet*, *bat*, *pot*, *ball*, *but*, *put*, *boot*, *bird*. These words were produced in the carrier sentence "I say uh — again" except for children of ages 5 and 6 years, who produced the words in isolation. Both /hVd/ and /bVC/ words have been used to achieve a neutral context but the influence of the final consonant, typically an alveolar, cannot be ignored. A different approach is to use several words to represent each vowel of interest. For example, Eichhorn et al. (in press) studied the corner

vowels of the traditional quadrilateral by using 5 words for each vowel; the words for vowel /u/ were *boo*, *boot*, *zoo*, *hoot*, *shoe*. The words were selected to be familiar to young children and to have high phonological neighborhood density to obtain maximal vowel acoustic space since lexically difficult words are produced with more expanded vowel acoustic space (Wright, 2004; Munson & Pearl Solomon, 2004). Therefore, these words are suited to lifespan studies of vowel acoustics. Vowels occurring in connected speech are perhaps more interesting for the purposes of understanding speech communication but pose a number of complicating factors such as coarticulation, adjustment to speaking rate and stress pattern, and influence of lexical and syntactic variables. Moreover, in the case of samples that are highly unintelligible, it can be difficult to identify target sounds.

Studies have shown that formant-derived measures such as vowel space area are influenced by several factors, including speaking rate and word stress (Fletcher, McAuliffe, Lansford, & Liss, 2015; Fourakis, 1991; Tsao, Weismer, & Iqbal, 2006; Turner, Tjaden, & Weismer, 1995), speaking style (e.g. clear versus conversational) (Lam, Tjaden, & Wilding, 2012), phonetic context (Chesworth et al., 2003), and, as noted above, phonological neighborhood density of test words (Munson & Solomon, 2004). Several of these factors influence the coarticulation between a vowel segment and its surrounding phonetic elements. Careful control and consideration of the factors just listed, including the characteristics of the speech sample and the instructions to the speaker (e.g., speaking rate, vocal effort) are important to ensure that valid and reliable acoustic data can be obtained, and to enable valid data comparisons, whether across studies of different speakers or across performances of an individual in repeated clinical assessments.

There does not appear to be a speech sample that is standardized in content or method of administration for purposes such as clinical assessment. If the goal is to determine a speaker's maximal performance (i.e., greatest dispersion of the sampled vowels to obtain the maximal vowel space), then consideration should be given to high phonological neighborhood density, slow speaking rate, strong stress on target words, a neutral phonetic context, and instructions designed to elicit clear speech. A change in any of these factors can result in reduction or centralization of the vowel space and, therefore, an underestimation of the maximal size of the vowel space. There has not been a comprehensive assessment of the stability of different speech samples over testing intervals or across different times. Vogel, Fletcher, Snyder, Fredrickson, and Maruff (2011) concluded that formant-frequency measurements from sustained vowels were stable, but the stability of formant data from other kinds of stimuli has not been established. Although no single speech sample will satisfy all purposes, there is good reason to standardize speech samples that could be applicable for many clinical and research purposes involving children or individuals with speech disorders. Steps taken to ensure greater uniformity would enhance the feasibility and validity of comparing data across settings, tasks, and individuals.

5.0. Selection of Measurement Time Point or Interval for Formant Estimation

What time point is most representative of a vowel? Ideally, the time point (or a narrow time interval) is selected to avoid the influence of flanking consonants and to approximate the steady-state portion of the vowel, that is, an interval in which the formant pattern is static. However, there is no guarantee that such a stationary pattern will occur and it frequently does not. Among the measurement points that have been used in formant analyses are the following: (1) midpoint or middle section of the vowel duration (Chung, Kong, Edwards, Weismer, & Fourakis, 2012; Heald & Nusbaum, 2015; McAuliffe, Fletcher, Kerr, O'Beirne, & Anderson, 2016; Perry et al., 2001); (2) one-third and two-third points of the vowel duration (Clopper, Pisoni, & De Jong, 2005); (3) global median of LPC formant tracks (S. Lee et al., 1999); (4) temporal locations corresponding to the 20-35-50-65-80%-point over a vowel's duration (Jacewicz, Fox, & Salmons, 2007); and (5) articulatory-referenced locations based on stability of formant pattern and/or acoustic properties of individual vowels where formants reach the presumed value characteristic of the intended vowel, such as the vowel inflection point (Derdemezis et al., 2016; Eichhorn et al., in press; Fletcher, McAuliffe, Lansford, & Liss, 2016). In the above listing, the items 1 through 4 rely on either time-specific or whole-vowel measures applied uniformly across vowels. In contrast, item 5 is vowel specific and narrowly focused in time in relation to the formant pattern.

Hillenbrand et al. (1999) compared several methods and concluded that the best results were obtained when the steady state was defined as the center of the sequence of seven analysis frames (56 ms) with the minimum slope in log F2- log F1 space. Duckworth, McDougall, de Jong, and Schockey (2011) suggested that formant measurements be made at a relatively stable portion of the vowel (when the formants are static or approximately so) and near the maximum intensity of the vowel (which usually is early in the vowel segment). A complication is that different formants, even the main vowel formants F1 and F2, do not necessarily exhibit stability at the same sampling point. Furthermore, in some cases, the time of maximum intensity does not coincide with the region of greatest stability in formant pattern. Criteria for selection of a measurement point differ across studies and these differences can contribute to differences in formant frequency estimation.

Presumed monophthongs such as /æ/ and /u/ may have formants that change substantially throughout the syllable nucleus. For vowel /æ/, it is not unusual to observe a diphthongal pattern in which F1 and/or F2 change during the vowel, apparently reflecting an articulatory change toward backing and lowering of the tongue, as illustrated in Figure 1 for production of this vowel by a young woman. Very different estimates of formant frequencies are obtained from the three sampling times labeled in the spectrogram: (a) near vowel onset, (b) vowel midpoint, and (c) near vowel end. The spectrogram also shows another complication that is frequently encountered, the appearance of vocal fry at the end of the word, which is a vocal characteristic commonly observed in young men and women (Abdelli-Beruh, Wolk, & Slavin, 2014; Wolk, Abdelli-Beruh, & Slavin, 2012). The change to a fry phonation can complicate spectral analyses especially when it disrupts the continuity of the formant pattern. For vowel /u/, the F2 frequency often declines in frequency from the initial portion

of the vowel to its termination, as shown in Figure 1 for a production of this vowel by a young man. Formant frequencies measured at the three labeled time points would yield quite different values. As shown in the spectrogram, a predetermined measurement point such as vowel midpoint often lands in an interval of F2 change. The spectrogram for vowel /u/ illustrates another common obstacle in that the point at which the F2 minimum frequency (and the minimum value of the F2-F1 difference) is reached is also a point when the amplitude of the higher formants is greatly reduced, which hinders estimation of the overall formant pattern.

One approach to dealing with complication such as those just discussed is to select the time point of analysis based on properties of the formant pattern. Some possible criteria are as follows. Vowel /i/ is associated with a low F1 frequency and a high F2 frequency (or maximal F1-F2 separation). Vowel /u/ is associated with a low F1 frequency and a low F2 frequency (or minimal F1-F2 separation). Vowel /æ/ is characterized by a fairly uniform spacing of formants, including separation of F1 and F2. Vowel /ɑ/ is marked by a high F1 frequency and a low F2 frequency (or small F1-F2 separation). The dynamics of the formant pattern vary with phonetic context. In the case of a CVC syllable, the usual pattern is that F1 frequency rises from the initial consonant to the value for the vowel and then decreases for the final consonant. The pattern for F2 depends on the place of consonant articulation (the F2 locus). In the case of CV syllable, the vowel /u/ may have a F2 frequency that continues to decrease into the end of the vowel. As already noted, a possible complication is that different formants do not always reach a stability or inflection at the same time point, so that priority may have to be given to the main vowel formants (F1 and F2). Ideally, all formant frequencies would be measured at the same time point but exceptions may be made, for example, when a higher formant is too weak to be evident in the acoustic analysis at the selected time point but has sufficient energy at a nearby point.

Ultimately, the static representations that have dominated acoustic analysis of vowels may be replaced by more dynamic approaches based on formant trajectories or combinations of measurement points that sample the vowel formant pattern.

6.0. Sources of Error in Formant Frequency Estimation

Minimizing or controlling sources of error helps to ensure that formant estimates are valid and reliable. Listed below are some of the more common sources of error and ways to handle them. See Maurer (2016) for a more detailed discussion.

6.1. Speaker's fundamental frequency

For voiced speech, in which the source excitation of the vocal tract is the quasi-periodic glottal pulse train, the transfer function of the vocal tract is effectively sampled at multiples of f_0 , or the harmonic source spectrum. Harmonics are selectively reinforced by formants, so that estimation of formants is a process of determining how harmonic amplitudes are shaped by formants. This process is made more difficult as f_0 increases and the spacing between harmonics becomes wider. It is for this reason that formant estimation for the speech of women and children is especially uncertain. For both spectrogram measurements and LPC analysis, the accuracy of formant-frequency estimation is about $f_0/4$ (Lindblom, 1962; Kent,

1976; Monsen & Engebretson, 1983; Vallabha & Tuller, 2002). For example, an accuracy of about 60 Hz is expected for a child with a f_0 of 250 Hz. However, accuracy can be affected by other factors, so that $f_0/4$ often is a best-case estimate of accuracy. In a comparison of formant-frequency estimation by LPC or spectrography, Monsen and Engebretson (1983) concluded that the frequencies of F1 and F2 could be measured to within approximately ± 60 Hz by either method. A comparable accuracy was achieved for F3 by LPC but spectrographic measurement had nearly twice as large an error. These results were obtained with synthesized isolated vowels modeled on an adult male talker. Hillenbrand et al. (1995) reported an absolute difference for repeated measurements of formant frequency of between 12 and 60 Hz. It is expected that accuracy would decrease for very young children with a high f_0 . A f_0 as high as 500 Hz may occur with high vocal effort in both children and women (Wang & Quatieri, 2010). Solutions to overcome the problem of formant estimation in the presence of a high f_0 include sweep-tone methods (White, 1999), use of an artificial larynx (Huggins, 1980), analysis-by-synthesis based on LPC (Traunmüller & Eriksson, 1997), weighted linear prediction (Alku, Pohjalainen, Vainio, Laukkanen, & Story, 2013), modified cepstral analysis (Story & Bunton, 2016), and analyses based on temporal changes in f_0 (Wang & Quatieri, 2010). Another problem related to f_0 is discussed in the following section.

6.2. F1- f_0 congruence

A particular challenge occurs when the first formant equals or approximates the vocal fundamental frequency ($F1 = f_0$). In this situation, care should be taken first of all to ensure that the analysis procedure has not confused f_0 for F1. The likelihood of F1 matching f_0 is greatest for the high vowels, which have a low F1 frequency. LPC analysis occasionally locks onto the first harmonic rather than the actual first formant, and spectrograms can be similarly deceiving in this respect. This is not to say that F1 can never coincide with f_0 . In fact, there may be an acoustic advantage to this coincidence in that F1 amplitude can be enhanced when its frequency falls on the first harmonic (H1). In singing, this is known as whoop timbre (Bozeman, 2013). Young children may exploit this advantage, but this convergence can sometimes be fallacious in acoustic analysis. Detection of error often can be accomplished by reference to multiple analyses including LPC and FFT spectra and wide-band and narrow-band spectrograms, with the object of determining if evidence of F1 is observed away from the H1 peak. Inspection of the FFT spectrum can be helpful for this purpose.

6.3. Checking for bias

The measurement bias resulting from laryngeal harmonics is not easily avoided and can be observed in the classic vowel formant data of Peterson and Barney (1952). Turner, Walters, Monaghan, & Patterson (2009) pointed out that the formant-frequency values in the Peterson and Barney data tend to assume integer multiples of the f_0 (i.e., as spikes in the histogram occurring at harmonic intervals). One approach to detecting this bias is that done by Turner et al. (2010), to plot a histogram of all formant-frequency values and look for unusual features in the data. If the histograms show spikes at intervals corresponding to harmonics or to step sizes in cursor control, then a bias is confirmed. Depending on the source of the bias, it may be possible to make adjustments for a more refined analysis. In a comparison of

several methods of formant analysis, Shadle, Nam, and Whalen (2016) concluded that f_0 bias affected all automatic methods. Manual measurements from pruned reassigned spectrograms were superior to the automatic methods not only in controlling f_0 bias but also in dealing with weak formants and glottal fry.

6.4. Frequency range of analysis

Especially for infants and young children, speech energy may extend well beyond the range typically considered in the analysis of the speech of adults of either sex (Bauer & Kent, 1987). Although this problem is most severe for fricatives (which can have energy extending to 16 kHz or higher in infants), it can also affect vowel formant patterns, particularly for the higher formants (F3 and higher). Care should be taken to ensure that the frequency range of analysis is appropriate to the intended goals of measurement. Reference to published normative data on formant frequencies is essential in setting the frequency range to accommodate the desired number of formants. Guidance on this issue is given in section 7.0. Another problem, considered in section 9.2, is that children appear to have wider formant bandwidths, and this can also contribute to inaccuracy in formant estimation. Fortunately, the greatly increased processor speeds and memory of modern computers facilitates the use of extended frequency ranges in acoustic analysis.

6.5. Nasalization

Nasalization is the consequence of coupling the oral portion of the vocal tract with the nasal cavities. Nasalization of a vowel can hinder formant analysis, especially in developing or disordered speech when nasalization may occur unexpectedly. Unfortunately, detection and quantification of nasalization are not straightforward, especially if oral vowels produced by the same speaker are not available for comparison. From the point of view of acoustic theory, nasalized vowels possess extra pole-zero pairs near the F1 region and across most of the spectrum, giving them a high density of formant patterns and a concentration of energy in the low frequencies (Fujimura, 1962; Hawkins & Stevens, 1985; Qian et al., 2017). However, the acoustic correlates of nasalization in a given speech sample can vary with anatomical differences across individuals, the area of velar coupling, vowel identity, and phonetic context (Pruthi, Espy-Wilson, & Story, 2007).

Acoustic correlates of nasalization can be identified in both the time and frequency domains. In the time domain, nasalization generally appears as a relative reduction of the overall amplitude of a vowel largely because of increased damping of a sound that is transmitted through the nasal cavity. In the frequency domain, several indexes have been proposed as correlates of nasalization, including the (a) amplitude of F1 minus the amplitude of H1 (Huffman, 1990), (b) amplitude of F1 minus the amplitude of the first nasal formant (P0), which is frequently below F1 (Chen, 1997), (c) amplitude of F1 minus the amplitude of the second nasal formant (P1) between F1 and F2 (Chen, 1997), and (d) bandwidth of the first formant (B1) (Styler, 2015).

Given the difficulty of identifying correlates of nasalization that apply across different vowels, studies often select the vowel /i/ for analysis (Haque, Ali, & Haque, 2016; Kataoka, Michi, Okabe, Miura, & Yoshida, 1996; A. S. Y. Lee, Ciocca, & Whitehill, 2003). The

reasons for this selection are: (a) for a given velopharyngeal area, the effect of acoustical coupling is greater for high than low vowels; (b) high oral vowels adjacent to nasal consonants have a longer duration of nasalization than low vowels in the same context; and (c) the nasal pole near 1 kHz in /i/ usually is prominent because it is distant from F1 and F2. But detection of nasalization in an unrestricted sample of vowels is problematic.

7.0. Adjustment of Analysis Parameters

Difficulties in formant estimation can occur for any speaker, child or adult, male or female, typical or atypical speech. Derdemezis et al. (2016) made several recommendations to improve the accuracy of formant estimation using FFT and LPC analyses. Some of the most challenging situations arise when formants are close in frequency (e.g. proximity of F1 and F2 in vowel /a/ and F2 and F3 in vowel /i/) or when F1 falls very close to f_0 , as often happens with vowel /i/. LPC sometimes will identify a single formant when, in fact, there are two formants. For example, Mauer (2016) commented, "...the sound spectra of back vowels and of /a- α / can exhibit only one single vowel-specific spectral energy maximum, although formant analysis using an analytical model (e.g. LPC analysis)—under involvement of "phonetic knowledge" and sometimes with interactive manual adjustment of parameter settings—indicates two vowel specific formants, often close in frequency" (p. 45). When two formants are not easily separated, either by visual inspection of a FFT spectrogram or by LPC, adjustments of analysis parameters may be helpful. For the former, it may help to decrease the number of FFT points, thereby decreasing the effective analysis bandwidth. Comparison of the spectral slice (the isolated FFT and LPC spectrum at given time points) with the LPC formant tracks overlaid on a spectrogram may help to resolve the formants. Reference to a narrow-band spectrogram also sometimes can be beneficial. For LPC, it may help to increase the number of filter coefficients, usually in steps of 2, to resolve more spectral detail, at least until the additional detail begins to obscure the formant pattern. In some cases, it is useful to examine the overall formant pattern in a word or syllable, as formants can converge or diverge even during a presumed monosyllable. Above all, it should be kept in mind that analysis methods and parameters can be adjusted to enhance the analysis. But there is no fail-safe solution, and the decision ultimately lies with the human analyst who must take into account acoustic-phonetic knowledge and the overall spectral pattern. If the estimate cannot be made with reasonable confidence, it is better not to record a measurement. Having reference values for speakers of the same age and sex can help to guide the estimation process. Published data are not available for F1, F2, F3, and F4 for speakers of different ages, and F4 in particular is rarely reported for children. Figure 2 gives an example of the frequency ranges reported in several studies and may serve as a general guideline.

Table 1 summarizes developmental static data on vowel acoustics that are relevant to the estimation of formant frequencies and to comparisons with normative values. The data for this heuristic approach include: ranges for the first three formant frequencies, vowel space area calculated for the vowel quadrilateral (qVSA), and mean vocal fundamental frequency (data sources are listed in the table caption). The values given in the table should not be taken as absolutes but rather as general guidelines to be considered in the adjustment of analysis parameters (such as analyzing bandwidth in a spectrogram or number of

coefficients in LPC analysis) or in detecting errors in formant analysis (such as merged or missing formants).

8.0. Lifespan Data on Vowel Formant Frequencies

This section appraises the database on vowel formants for both sexes across the lifespan, with an eye toward establishing normative or reference data for clinical applications. Formant frequencies have been estimated through various methods, primarily (1) visual inspection of a FFT spectrogram or spectrum to identify formants, or (2) automatic formant detection by means of LPC or cepstral analysis. The automatic methods generally use a root solving or peak picking procedure (Vallabha & Tuller, 2002; Welling & Ney, 1998) to identify formants after the initial analysis is completed. No one of these methods is error-free, and it is often prudent to use two or three of them in complementary fashion to ensure the most accurate results.

The great majority of data on vowel formant frequencies are for the first two formants, F1 and F2, but the higher formants F3 and F4 also are of interest, for example, in describing rhotic sounds (Hagiwara, 1995), statistical approaches to categorization and normalization of both rhotic and non-rhotic vowels, perhaps because F3, like f0, can serve as a normalizing factor (Disner, 1980; Hillenbrand & Gayvert, 1993), accounting for the speaker's formant (Bele, 2006; Leino, Laukkanen, & Radolf, 2011), specifying resonances of the hypopharynx (Takemoto, Kitamura, Honda, & Masaki, 2008), and describing acoustic consequences of procedures such as tonsillectomy (Švancara, Horáček, Vokrál, & Cerný, 2006). To meet these needs, it is desirable to have lifespan data on at least the first four formants in both males and females. The sexual dimorphism in speech is proportionately (i.e., male:female ratio) one of the largest observed in physical measurements of humans (Rendall, Kollias, Ney, & Lloyd, 1995), which makes speaker sex a critical factor in compiling and comparing data on speech production. A complication is that the methods of analysis are not equally suited to both sexes or to adults and children. As Klatt and Klatt (1990, p. 820) remarked, "informal observations hint at the possibility that vowel spectra obtained from women's voices do not conform as well to an all-pole [i.e. all formant] model, due perhaps to tracheal coupling and source/tract interactions." By this reasoning, sexual dimorphism affects not only the data but also the optimal ways by which the data are obtained.

8.1. Normative vowel formant-frequency databases for young to middle-age adults

For adults from young to middle-age, the most frequently cited formant-frequency databases for American English appear to be those of Hillenbrand et al. (1995), S. Lee et al. (1999), and Peterson and Barney (1952). Basic descriptions of these and 4 smaller datasets (some with a larger age range of adult speakers) are given in Table 2. For the purposes of this paper, vowel quadrilaterals are used to compare the data from various studies. The quadrilateral represents most, but not necessarily all, of the acoustic working space for vowel production and can be used to derive summary indices such as vowel space area (as considered later in the paper). This is not to assume that vowel quadrilaterals and measures such as vowel space area are preferred over other approaches but rather to examine a conventional and highly used approach in vowel formant studies. Another advantage to the

vowel quadrilateral is that most studies include data for the four corner vowels. F1-F2 vowel quadrilaterals derived from the mean data in 7 studies are shown for men and women in Figure 3a and Figure 3b, respectively. There is notable variance in the mean formant frequencies of the corner vowels. For example, there are differences of approximately 500 Hz for men's F2/u/, 400 Hz for women's F1/æ/, and 700 Hz for women's F2/u/. For the moment, the emphasis is on means, as these have been used in compiling the major datasets on vowel formants. Among the likely factors that account for the variation in results are speaker size, speaker dialect, speech sample, and analysis method. These illustrations show that the F1-F2 values of the corner vowels from different studies do not so much converge on specific points as they fall in regions of the plane.

Because it is difficult to visualize individual quadrilaterals in the composite data in Figures 3a and 3b, selected isolated quadrilaterals are shown for men in Figure 4a and women in Figure 4b. The four panels in these figures represent data from: (a) Hillenbrand et al., (1995), (b) Zahorian and Jagharghi (1993), (c) Peterson and Barney (1952), and (d) Childers and Wu (1991). Intersecting diagonals have been drawn from the corner points. For both men and women, the quadrilaterals are convex (i.e., the diagonals are internal and intersect) in all studies except Hillenbrand et al. (1995), where the quadrilateral collapses to a nearly triangular shape, especially for women (panel a in Figure 4b). The positions of the quadrilaterals in the F1-F2 plane also vary. Notable differences among the vowel quadrilaterals shown in Figures 4a and 4b are that (a) the high vowels in Hillenbrand et al. (1995) have higher F1 values than in other studies, (b) a relatively restricted F2 range for the high vowels is seen in the data of both men and women in Zahorian & Jagharghi (1993), and (c) overall, the quadrilateral of Peterson and Barney (1952) is the largest or nearly the largest in both F1 and F2 ranges, virtually encompassing the quadrilaterals from other studies for men and somewhat less so in women.

Dialect variations are likely one source of the variability in formant patterns. For example, the variance in results for vowel /u/ may be related to the phenomenon of /u/ fronting (also called "GOOSE fronting" in dialect studies because *goose* is a keyword for this vowel). Fronting of vowel /u/ has been noted in nearly all varieties of North American English (Labov, Ash, & Boberg 2006; Thomas, 2001). Furthermore, it appears that the fronting can take two different forms, monophthongal and diphthongal (Koops, 2010). Either form presents a challenge to the assumption that vowel /u/ can be regarded as a high-back monophthong. Possibly, the presence or absence of /u/ fronting accounts for some of the variance in the F1-F2 data in Figs. 3a and 3b. Dialect variations can affect other vowels as well, as detailed by Labov et al. (2006). A series of systematic studies is shedding light on the acoustic properties of vowels in various dialects of American English (Fox & Jacewicz, 2008, 2009, 2010, 2015; Jacewicz & Fox, 2002; Jacewicz, Fox, & Salmons, 2011), and these data are an important background in interpreting vowel formant data.

8.2. Normative vowel formant-frequency databases for older adults

Summaries of studies reporting formant data for older adults are compiled in Table 3 for measures of f0, F1 and F2 frequencies, and the overall formant pattern in vowel space. The most consistent result from both cross-sectional and longitudinal studies is that f0 and F1

frequencies decrease with age in women and that f_0 is either unchanged or increases in men (Eichhorn et al., in press). Because published studies involve different ages of participants and different procedures of analysis, it is difficult to fix a particular age or age range at which changes occur in either sex. Several explanations have been advanced to account for age-related changes in vowel formant frequencies, including lengthening of the vocal tract (Endres, Bambach, & Flosser, 1971; Linville & Rens, 2001), altered dimensions of the back cavity (Scukanec, Petrosino, & Squibb, 1991), diachronic or intergenerational phonetic change (Fox & Jacewicz, 2010), reduction of articulatory movement (Watson & Munson, 2007), and adjustments of lingual articulation (Linville & Fisher, 1985). Further studies are needed to determine the relative contributions of these factors, which may have different effects in different individuals. It is possible that acoustic changes occur for subgroups within the aging population (e.g., individuals with compromised health). Normative data are needed clinically to interpret acoustic changes that may occur in age-related conditions such as dentofacial alterations or neurodegenerative diseases. The data also are needed for the development of methods for the automatic detection of speaker age (Schötz, 2007).

The studies summarized in Table 3 report on only a small number of individuals compared to the millions of people in the United States who are 65 years or older. Therefore, only cautious or tentative conclusions can be drawn from the data now available. An ambitious program of research is needed to determine age-related effects in speech. As Schötz (2007) concluded, “speaker age is a very complex characteristic of speech...[that] leaves traces in all acoustic-phonetic dimensions and it is influenced by numerous other factors, such as physiological condition” (p. 15).

8.3. Normative vowel formant-frequency databases for children

The major sources of formant frequency data for typically developing children are listed in Table 4, which shows for several published studies the ages of participants for whom data were obtained. Presumably, all participants in these studies were learning American English as their first language so that the composite data can be used to construct a picture of the development of vowel acoustics in that particular language (neglecting dialects) (as reported by Vorperian & Kent, 2007). Normative developmental data are useful for many functional, clinical, and theoretical purposes such as: the development of automatic speech recognition systems for unrestricted speaker populations, the specification of age-typical values for the interpretation of clinical data, the understanding of speech development in children including the determination of age of emergence of sexual dimorphism, and the ontogenic patterns of vowel mastery (which can also be instructive for clinical purposes). In addition, these data are helpful in guiding the estimation of formant frequencies in children’s speech and to guard against the uncritical acceptance of data from automatic analyses such as LPC. Vorperian and Kent (2007) provide graphical summaries of the development of the first three formants in F1-F2 and F1-F3 plots. However, the data from different sources may not be completely compatible owing to differences in speech samples, analysis methods, and speaker selection. Vorperian and Kent used the composite data F1-F2 plots to characterize major developmental features, such as decreases in both formant frequencies and vowel space area. They also pointed to possible evidence of growth spurts in the vowel quadrilaterals, that is, abrupt shifts in F1-F2 patterns at certain ages. Shifts can be seen

especially in the longitudinal data for males reported by Kohn and Farrington (in press). The ability to detect such developmental changes depends on the accuracy of the measurements at representativeness of the data at each age. In Table 3, studies A through H report data on infants and toddlers, and the remaining studies pertain to children aged preschool through adolescence. Studies providing data over several different years that are particularly valuable in determining the developmental pattern of vowel acoustics are: J – Eguchi and Hirsh (1969); K – Perry, Ohde, and Ashmead (2001); M – Assmann, Nedarey, and Bharadwaj (2008); and N – S. Lee et al. (1999). The prospect is that new data will be published, given the availability of databases such as the Arizona Child Acoustic Database Repository, a longitudinal collection of audio samples from children between the ages of 2-7 years. (Bunton & Story, 2016) and HomeBank (VanDam et al., 2016). Perhaps the most valuable database would be in the form of raw data that could be used to construct point clouds or to derive other measures such as those considered in a later section.

An important question relating to children's speech is: when does sexual dimorphism appear in formant datasets? In their review of formant-frequency data, Vorperian and Kent (2007) noted that although f_0 differences do not emerge until age 10 or 11, sexual dimorphism in formant frequencies emerges by about 4 years. By the age of 8 years, boys have lower formant frequencies than girls across all vowels, but sex differences are not uniform across formants. Yang and Mu (1989) reported that sex differences in F3 frequency appeared at the age of 3 years and were marked by the age of 6 years. Whiteside (2001) observed a prepubescent sex difference in F3 of the rhotic vowel, leading her to conclude that there are prepubertal sex differences in the volume of the oral cavity. Accordingly, reference data for formant frequencies should separate the data for boys and girls by at least 4 years of age for F1 and F2 and as young as 3 years for higher formants.

The origin of the prepubertal sex differences is not entirely clear. In a widely cited study, Fitch and Giedd (1999) concluded that the vocal tract anatomy is not sexually dimorphic until puberty. Possibly, the reported anatomic measurements (mostly measures of the length of the tract and its divisions) are not sufficient to account for the acoustic properties of the vocal tract. Information on regional vocal tract volumes would provide a more complete picture of the anatomic correlates of acoustic properties of speech. Vorperian et al. (2011) concluded from their imaging studies that there were not only postpubertal dimorphisms for most vocal tract structures, but also prepubertal sex differences for some structures at particular ages that have not been documented to date due to apparent growth rate differences between males and females. Thus, it is feasible that the anatomic differences may contribute to pre-pubertal acoustic differences, but in addition, it is possible that the acoustic differences between boys and girls are the result of learning gender-specific speech patterns. Support for a learning or sociocultural hypothesis comes from research on voice gender (Cartei, Cowels, Banerjee, & Reby, 2014; Cartei, Cowles, & Reby, 2012; Cartei & Reby, 2013). A complete explanation of developmental differences in formants for boys and girls likely will have to take into account both biological (sex) aspects and sociocultural (gender) aspects.

9.0. Lifespan Data on Vowel Formant Bandwidths

Formants can be described by their frequencies, bandwidths, and amplitudes. The classic linear source-filter theory holds that formant amplitudes are determined by formant frequencies, formant bandwidths, and the effective source energy. Therefore, formant amplitudes are largely predictable given these other types of information. The bandwidths of formants are determined physically by the combined effects of radiation, compliance of the vocal tract walls, viscosity, heat conduction, and glottal opening (Bickley, 1989; Fant, 1962; Fleischer, Pinkert, Mattheus, Mainka, & Mürbe, 2014). These effects can vary with speaker characteristics such as age, sex, and state of health, and perhaps specific control of relaxation and strain of the mucosa and muscle tissues of the vocal tract (Fleischer et al., 2014). Formant bandwidths have only a small effect on vowel identity in perceptual studies but they are of interest in determining relationships between the biomechanical properties of the vocal tract and the acoustic signal of speech, especially in the study of speech development and speech disorders related to craniofacial dysmorphologies or neurologic disturbances.

9.1. Methods of bandwidth estimation

Commonly used speech analysis software can be used to make automatic measurements of formant bandwidth from LPC, but these data are not necessarily valid nor reliable, as shown empirically (Burriss et al., 2014) and mathematically (Mehta & Wolfe, 2015). The main alternatives to LPC-derived bandwidth values are either the logarithmic decrement method in the time domain (Bickley, 1989) or the half-power point in the frequency domain (Burriss et al., 2014). For the latter method, the half-power points (or 3db down points) are not always easily determined when formants are in close proximity in frequency, because the formant curves may be asymmetric. One solution is to measure the half-point on the side of the formant that is better defined and assume symmetry (Burriss et al., 2014). Both of these methods are time consuming when done manually, which is perhaps one reason why formant bandwidth data are rarely reported.

9.2. Normative data on formant bandwidth

Values of formant bandwidth from several studies are shown in Table 5. The studies differed in the speakers' sex, age, and language, the method of measurement; and the number of vowels that were examined. Because the vowels varied across studies and in some cases were not specified, vowel identity is not included in the table but it should be noted that bandwidths can differ across vowels produced by the same speaker. Differences in methods may account for the considerable spread of values across studies. For English-speaking adults, reported bandwidths generally range from 50 to 140 Hz for B1, 62 to 149 Hz for B2, and 67 to 223 Hz for B3. Reasons for the variation in results across studies are not immediately clear but probably are associated largely with procedural differences in bandwidth measurement.

Children appear to have formant bandwidths larger than those in adults, but the limited data make it difficult to construct a clear picture of developmental trends (Bickley, 1989; Krishna & Rajashekar, 2013; Robb, Chen, & Gilbert, 1997; Whiteside & Hodgson, 1999). Robb et

al. (1997) reported mean bandwidths ranging from 170 to 350 Hz (B1) and 345 to 671 Hz (B2) for children aged 4 to 25 months. According to Oller et al. (2010), typical bandwidths of vowel-like sounds produced with normal phonation by young children are less than 400 Hz for F1 and less than 600 Hz for F2. Bandwidths exceeding these values were used to identify infant vocalizations with a growl quality. For 6- to 10-year-old children, Whiteside and Hodgson (1999) reported B1 and B2 means of about 190 Hz and a B3 mean of 263 Hz. A tentative conclusion from these limited data is that the main formant bandwidths decline from about 400 Hz in infants, to about 200 Hz in 6-to-10 year-olds, and to about 50 to 200 Hz in adults, with women having larger bandwidths than men. The suggestion by Fleischer et al. (2014) that speakers exercise specific control of relaxation and strain of the mucosa and muscle tissues of the vocal tract has interesting developmental implications, including the possibility that children learn to make such adjustments to enhance the intelligibility and quality of speech. A related possibility of clinical interest is that children with hypotonic musculature, associated with conditions such as cerebral palsy or Down syndrome, may have atypically large formant bandwidths.

Based on the available data, formant bandwidths are larger in children even when they are considered relative to formant center frequency, for example, using the Q formula that characterizes filter tuning:

$$Q = F_i/B_i$$

Where F is the formant frequency for formant i , and B is the bandwidth for that formant.

Values of Q calculated for data reported for adult male speakers by Bogert (1953), Fant (1962), and House and Stevens (1958) generally range from about 8 to 30 for the bandwidths of the first 3 formants. By comparison, values of Q calculated from data reported by Robb et al. (1997) for infants range from about 2 to 7 for the first 2 formants. The larger formant bandwidths in infants carry implications for both the perception and acoustic analysis of speech. Generally, speech produced with larger formant bandwidths is likely to be less intelligible and more difficult to analyze acoustically. A high f_0 frequency combined with large formant bandwidths could contribute to errors in estimating formant frequencies and bandwidths in children's speech.

To summarize, formant bandwidth data from LPC should be considered as tentative and potentially highly inaccurate. Estimates of formant bandwidth for a given age-sex group of speakers differ by as much as a factor of 4. Until analysis algorithms are improved, it is prudent to rely on measurements derived from half-power points. However, promising newer methods of bandwidth calculation include using properties of group delay functions (Medabalimi, Seshadri, & Bayya, 2014) or using LPC root extraction combined with a root classification algorithm (Qian et al., 2017). Based on the limited lifespan data that have been published, it appears that formant bandwidths decrease from infancy to adulthood. Additional collection of bandwidth data from speakers of both sexes and different ages would be a valuable normative database that could help quantify developmental changes in

speech acoustics and have clinical applications for the assessment of biomechanical properties of the vocal tract.

10.0. Derived Metrics and Data Displays

Several different data reduction strategies have been developed for the purpose of revealing general patterns and tendencies in formant-frequency data. Some of these measures are indices of central tendency and some are indices of the size and geometry of the acoustic space for vowel production. This section lists and defines the most commonly used measures, which are predicated on the assumption that vowels can be characterized by a static pattern of formant frequencies (e.g., a point in the F1-F2 plane). However, with some adjustments, many of these measures could be used with nonstatic approaches such as 2-point measures of vowel onset and offset.

10.1. Centroid

An example of a measure of central tendency is calculation of the centroid based on either the vowel triangle or quadrilateral. The centroid, S , of a triangle is the grand mean of the formant frequencies for the corner vowels:

$$S(F_i) = (F_i/i/ + F_i/a/ + F_i/u/)/3$$

A variation of centroid calculation used in vowel normalization is:

$$S(F_i) = (F_i/i/ + F_i/a/ + F_i/u'/)/3$$

Where $/u'/$ is a hypothetical extreme vowel point derived from the coordinates of $/i/$ such that both F1 and F2 of $/u'/$ are equal to the F1 of $/i/$.

Because this calculation can result in a skewing of values in the lower region of the vowel space (Fabricius, Watt, & Johnson, 2009; Thomas & Kendall, 2007), Fabricius et al. (2009) proposed a modified formula:

$$S(F_i) = (F_i/i/ + F_i/a/ + F_i/u'/)/3, i = 1$$

$$(F_i/i/ + F_i/u'/)/2, i = 2$$

The centroid has been used for several purposes, including vowel normalization and characterization of speaker differences. An important caveat is that the location of the centroid can be strongly affected by the number of vowels used in its calculation and by systematic directional changes in vowel articulation (Karlsson & van Doorn, 2008). A lack of stability in the centroid can have substantial effects on measures that are based on its calculation.

10.2. Long-term formant distribution (LTF)

This measure, introduced by Nolan and Grigoras (2005) for forensic applications, determines the average formant-frequency values of a given speaker. The procedure is to calculate for each formant of interest the average of all formant measurements of all vowels produced across a recorded sample. This average is termed the LTF value for the formant. Accordingly, a given speaker has a LTF value and standard deviation for each formant (LTF1, LTF2...etc.). Because the measurements are made frame-by-frame, vowels of long duration carry more weight than vowels of short duration. A particular value of LTF is that it is potentially independent of f_0 , dialect, and speaking rate. Aside from its forensic application, LTF may have a more general value to characterize speakers without needing to identify the phonetic identity of vowels in a sample. For example, it could be used in developmental studies to chart formant changes with maturation.

10.3. Area and distance measures

This section pertains to measures that have been used as summary indices of formant data. Many of these are area or distance measures calculated for vowels plotted in the F1-F2 plane. They have been applied to the study of several topics in speech and language, including (a) language or dialect differences in vowel systems, (b) speaking registers and prosodic patterns, (c) articulatory impairment in various speech disorders, (d) speech development in children, and (e) speech intelligibility in various speaker groups, both normal and disordered. Generally, these efforts seek to determine how the distribution of formant values relates to the particular question under study. A number of area or distance measures in the F1-F2 plane have been proposed as numerical indices. The great majority of these assume a static acoustic representation of vowels. The formulas shown in Table 6 follow the formant notation of $F_n/x/$ where n is the formant number and x is the phonetic symbol for the vowel (e.g., $F_1/i/$ is the first formant frequency of vowel $/i/$). A fundamental question in choosing among the alternatives in Table 5 is whether it is preferable to maintain the full set of data (e.g., a point cloud that represents all data points) or to reduce the data to a representative value such as a mean. One disadvantage to a mean is that it is highly responsive to extreme values in the data and does not reveal details such as bimodality or multimodality in the data distribution.

10.3.1. Vowel space area—Probably the most frequently reported acoustic measure of the vowel F1-F2 plot is the Vowel Space Area (VSA), calculated as the area of the polygon (generally either a triangle, [tVSA or VSA-3] or a quadrilateral [qVSA or VSA-4]) formed by the point vowels and therefore supposedly reflecting the articulatory extrema of vowel production. Formulas for the calculation of VSA are included in Table 6. VSA has also been called Maximal Vowel Space Area (MVSA), vowel working space, articulatory working space, and area of the vowel loop. Normative data on VSA include its development in both male and female children (Flipsen & Lee, 2012; Kohn & Farrington, in press; Pettinato, Tuomainen, Granlund, & Hazan, 2016; Vorperian & Kent, 2007) as well as information for male and female adults (Kwon, 2010; Neel, 2008; Vorperian & Kent, 2007). A general decrease in VSA is noted across age in children but reversals may occur as the result of factors such as articulatory overshoot (Kohn & Farrington, in press; Pettinato et al., 2016). Substantial variation in qVSA is observed across studies. For example, the VSA calculated

from the adult male data of Peterson and Barney (1952) is 411.5 kHz^2 compared to only about 330 kHz^2 in the data that Lee et al. (1999) reported for 18-year-old males. Given this large difference, caution should be observed in using any particular source of normative data for qVSA.

The clinical relevance of VSA has been noted across a spectrum of disorders affecting communication, including: (1) children with neurogenic speech disorders (Higgins & Hodge, 2002; Hustad, Gorton, & Lee, 2010; Liu, Tsao, & Kuhl, 2005; Narasimhan, Nikitha, & Francis, 2016); (2) adults with acquired dysarthria (Bang, Min, Sohn, & Cho, 2013; S. Kim, Kim, & Ko, 2014; Turner, Tjaden, & Weismer, 1995; Weismer, Jeng, Laures, Kent, & Kent, 2001); (3) adults with Down syndrome (Bunton & Leddy, 2011), (4) individuals with hearing loss (Palethorpe & Watson, 2003); (5) individuals with hearing loss (Palethorpe & Watson, 2003); (6) adults who have undergone glossectomy (Kaipa, Robb, O'Beirne, & Allison, 2012; Whitehill, Ciocca, Chan, & Samman, 2006); (7) adults who have undergone glossectomy (Kaipa, Robb, O'Beirne, & Allison, 2012; Takatsu et al. 2017; Whitehill, Ciocca, Chan, & Samman, 2006); (8) individuals undergoing treatment for oral or oropharyngeal cancer (de Bruijn et al., 2009); (9) individuals with Class III malocclusion (Xue, Lam, Whitehill, & Samman, 2011); (8) people who stutter (Blomgren, Robb, & Chen, 1998; Hirsch et al., 2008); (9) individuals with hypernasality associated with cleft palate (Haque, Ali, & Haque, 2016); and (10) individuals in psychological distress or with self-reported symptoms of depression and post-traumatic stress disorder (Scherer, Lucas, Gratch, Rizzo, & Morency, in press; Scherer, Morency, Gratch, & Pestian, 2015).

VSA or similar measures also have been used to evaluate the effects of treatment on voice and speech production (Eliasova et al., 2013; Lin, Hornibrook, & Ormond, 2012; Mahler, & Ramig, 2012; Roy, Nissen, Dromey, & Sapir, 2009; Sapir, Spielman, Ramig, Story, & Fox, 2007; Takatsu et al., 2017; Wenke, Cornwell, & Theodoros, 2010), to study disease progression (Skodda, Grönheit, & Schlegel, 2012), and to serve as an early marker of disease (Rusz, et al., 2013). These reports across a spectrum of speech disorders indicate that VSA has value as one component of an acoustic profile of disordered speech.

Several studies have shown that VSA is correlated with speech intelligibility, with larger values of VSA being associated with higher levels of intelligibility in normal native speech (de Boer, 2009; Neel, 2008; Smiljani & Bradlow, 2009), normal non-native speech (Chen, Evanini, & Sun, 2010), and disordered speech (de Bruijn et al., 2009; Higgins & Hodge, 2002; Turner, Tjaden, & Weismer, 1995). VSA is only one factor that accounts for the intelligibility of speech, and its importance in any given speaker or group of speakers depends on other factors, such as the availability of different types of acoustic cues.

Questions have been raised regarding the sensitivity and validity of VSA measures (Karlsson & van Doorn, 2012; Neel, 2008; Sandoval et al., 2013; Sapir, Polczynska, & Tobin, 2009) and the vowel quadrilateral representation itself (Fox & Jacewicz, 2015). From an analytic perspective, a major shortcoming of VSA is that it may not be sensitive to variations in the location of any given vowel or pair of vowels in F1-F2 space. A fuller interpretation of differences in VSA requires an examination of alterations in the location of the point vowels. Contraction of VSA can result from an overall centralization of vowels (i.e., all point vowels

are displaced centrally in the F1-F2 plane) or from specific changes in formant patterns (e.g., reduction of the F2 frequency range, or centralization of only one vowel.) In addition, area alone discards much of the information contained in the original data. VSA often is calculated from F1-F2 means either for individual speakers or a group of speakers but using the mean may not always be sufficient. If the objective is to use VSA as a measure of the maximum vowel space of which an individual speaker is capable (i.e., the acoustic envelope of possible vowel production), the mean may underestimate the speaker's actual maximum performance if there is considerable variability in the formant-frequency values for one or more vowels. This problem affects most measures of maximum performance such as maximum phonation time, maximum expiratory pressure, and maximum rate of syllable repetition (Kent, Kent, & Rosenbek, 1987). One solution is use the "best of three" (the maximum value observed in three trials) Sørensen, Durck, Bork, & Rasmussen, 2016).

Intra-speaker variations in F1-F2 values are not represented in Figs. 3a and 3b, as the data points used in constructing the figures are the means from the respective studies. Point clouds that preserve individual data would illuminate the actual variability within and between speakers. One alternative is to calculate the area of the convex hull containing the formant-frequency values for a set of vowels. This approach considers all data points and is well-suited to samples of conversational speech or other large speech samples, for which the vowel space is a higher-dimensional polygon than the traditional vowel quadrilateral (Sandoval et al., 2013). An algorithm for determining the area of a convex hull is available in software systems including MATLAB[®] and R (R Core Team, 2016). Another approach is calculation of the Articulatory-Acoustic Vowel Space (AAVS) which is based on continuously sampled formant trajectories in connected speech (Whitfield & Goberman, 2016). Progress also has been made in developing measures that express the similarity between vowel quadrilaterals. Amir and Amir (2007) describe two measures for this purpose, vowel space similarity and vowel space skewedness. These measures were used to evaluate the degree of vowel reduction that occurred in continuous speech.

VSA appears to be the most frequently reported measure of the planar distribution of vowels, but the interpretation of VSA values may be enhanced by other measures, such as those discussed in the next section. A given value of VSA is not fully interpretable without reference to the actual vowel pattern (quadrilateral or triangle). In addition, VSA usually is reported for mean values only, without standard deviations or other measures of variability. VSA is not alone in this disadvantage, as measures such as VAS, FCR, and VAI all reduce formant data into a single metric, with a resulting loss of statistical power and a reduction in the quality of intra-speaker models of articulatory proficiency (Karlsson & van Doorn, 2008). When the VSA is accompanied by other measures, it may be possible to understand sources of variability and to interpret more fully the effects of factors such as development, speaker sex, disorder, or dialect. In a study of dysarthria in amyotrophic lateral sclerosis, Fougeron and Audibert (2011) concluded that, "Results support the need for different acoustic metrics in order to capture the large interspeaker and inter-sex variation observed, and to reflect the various types of alteration possible" (p. 687). These authors found that differences between speakers with dysarthria and control speakers were best accounted for by two measures of vowel space area, two measures of centralization, two measures related to reduction of the front-back tongue dimension, and a global measure of overlap between

vowel pairs. Efforts should be made to determine which combination of measures (or graphic displays) is most revealing of differences in formant data across speaker groups or across individuals in repeated observations.

F1-F2 plots have been the predominant graphical presentation of vowel formant data, often on the assumption that the first two formants are sufficient for recognition of non-rhotic vowels. However, for reasons noted in Section 8.0 and in particular to help understand developmental changes and the emergence of sexual dimorphism in formant data, a more complete specification of vowels can be obtained by consideration of data for a least the first four formants. Graphical summaries would take the form of 3- or 4-dimensional displays (see Vorperian et al., 2015, for an example of a 3-dimensional display for the first three formants).

10.3.2. Other measures of vowel configurations—A number of alternatives or complements to VSA are listed in Table 6. The suitability of a given measure depends on the purpose of analysis. Representation of all points in the F1-F2 plane is required to establish the ranges for automatic speech recognition or to determine the maximally achievable vowel space for an individual talker or group of talkers. Means and standard deviations may be satisfactory for characterizing the typical performance of a clinical or dialect group, but as noted earlier in this section, means do not necessarily reflect maximal performance. Several of the measures in Table 6 focus on dispersion or distance features and are useful in characterizing the configuration and dimensions of vowel production. It is a task for future research to determine the most effective combinations of these measures for various purposes. Although some of these likely provide redundant information, some combinations of them or indices yet to be determined may be particularly useful in depicting formant-frequency patterns. It may be premature to recommend particular combinations but the need for such an approach is evident. The selection of measures may depend on the particular application; for example, the F2/i-/F2/u/ ratio can help in the understanding of reduced tongue movement in dysarthria or Down syndrome. The tabled indices rely almost entirely on F1 and F2 frequencies that are most critical for vowel identification. However, as discussed in the previous section, higher formants (at least F3 and F4) may help to understand developmental changes and the emergence of sexual dimorphism in formant data and therefore metrics or visualizations of higher formants is needed.

The focus of this review is on static vowel representations based mostly on simple statistics such as means but the unfolding story about vowel acoustics necessarily includes dynamic properties, such as those recognized in vowel inherent spectral change (Morrison & Assmann, 2013) and the consideration of individual variability in the form of analyses such as point clouds. Measures such as VSA may retain some usefulness but they probably will not be sufficient in themselves to address important issues in vowel production. The status quo regarding static measures of vowel formants is not entirely satisfactory but it is a foundation for a more informative analysis of vowel production.

11.0. General Recommendations pertaining to Current and Future Clinical Applications

Aside from the recommendations to establish more standardized methods and reference data for clinical use, the following recommendations pertain especially to the clinical application of vowel acoustics, specifically the estimation of formant frequencies in developing, aging, and disordered speech.

If it is desired to compare results with normative data, then it is recommended to consult the articles mentioned in this review, taking into account factors such as speaker sex and age, speech sample, and dialect. Although data on vowel formant frequencies have been reported in multiple studies cited in this review, it is not always straightforward to find a suitable normative database for all speakers. Data on vowel formant frequencies for any age-sex group vary somewhat across published studies, and the sources of this variation may include differences in speech samples, methods of analysis, and dialect. The development and use of standardized speech samples is one step that would help to establish uniformity. One approach is to construct speech samples appropriate for different age groups, for example, one for very young children with immature language capability, one for children with language capability typical for early school grades, and one for adults. Another is to construct speech samples that can be used by speakers of various ages and levels of linguistic competence (Eichhorn et al., in press). The challenges of lifespan research for either typical or atypical speech are substantial but not surmountable. Hazan (2017, p. 41), commented, "...it is still the case that a majority of researchers within the field of speech sciences who have an interest in the effect of age on speech communication specialize in either development studies or studies into ageing, with few having the practical experience of running studies with different age ranges, which each have specific demands and challenges." Clinical application would be enhanced by the availability of lifespan normative data for both sexes.

Procedures of formant-frequency estimation should be outlined at the outset, including selection of the speech sample, time points of measurement, methods of acoustic analysis, and steps to be taken in case the initial analysis is considered inadequate or erroneous. It is particularly important to be aware that a vowel steady state may not be evident in the formant pattern, in which case criteria are needed to select a single-time point for measurement or to use an alternative such as two time points or trajectories. It is important to anticipate the need for adjustment of analysis parameters, especially the number of points in FFT, the number of coefficients in LPC, and the dynamic range in spectrograms (Derdemezis et al, 2016). The major typical speaker characteristics that affect the adjustment of analysis parameters are sex and age, but within a particular age-sex group there can be large differences in f_0 and other acoustic properties.

Particularly for young children, female speakers, and speakers with disordered speech, due note should be taken of vowel samples with a high f_0 , apparent nasalization, phonatory irregularities, or other features that may compromise the validity and reliability of acoustic analysis. Inspection of the wide-band spectrogram can be helpful in identifying potentially troublesome features. Although it is not possible to anticipate all complications in acoustic

analysis, guidelines can be developed for some of the more commonly occurring challenges. Suggestions are included in this paper and in other sources for particular speaker groups: typical and atypical speech development in children (Derdemezis et al., 2016; Kent, 1976; Lee et al. 1999); autism spectrum disorder (Lyakso, Frolova, & Grigorev, 2016); cleft palate and other velopharyngeal disorders (Philips & Kent, 1983); dysarthria (Kent & Kim, 2003; Kent, Weismer, Kent, Vorperian, & Duffy, 1999); and speech disorders in general (Ludlow, Kent, & Gray, 2019).

It is not always possible to obtain data for all formants in a given sample. Among the most frequent problems are merging of formants (e.g., merging of F1 and F2 for back vowels), weak energy in the higher formants, and formant-harmonic interaction (most likely to occur when f_0 is high). Criteria should be developed to decide when measurement of one or more formant frequencies should be abandoned after complementary methods to estimate formants have been exhausted. For example, when formant LPC tracks lack continuity or are otherwise questionable, then alternative methods, such as visual examination of the wide-band or narrow-band spectrogram, may be tried.

Comparison of observed formant-frequency values with normative data (Table 1 and Figure 2) can be helpful in detecting errors of estimation. Discrepancies do not always mean that an error was made but they do signal the need to check on the accuracy of the data. Errors can arise for a number of reasons, and LPC analysis is by no means immune to errors such as mislabeled formants (e.g., F3 identified as F2).

Quantifying vowel space using only mean data for formant frequencies can lead to reduced discrimination of differences between speaker groups (Fougeron & Audibert, 2011; Karlsson & van Doorn, 2012). This limitation is particularly important for the computation of indices such as Vowel Space Area, Vowel Articulation Index, and Formant Centralization Ratio. Therefore, consideration should be given to the use of two or more metrics to make the most effective use of the data. Further research is needed to identify metrics that are both efficient and complementary in their application to disordered speech. In the meantime, consideration can be given to measures with different computational approaches, such as Vowel Space Area complemented by Vowel Formant Dispersion or Mean Vowel Cluster Size.

Standardization of procedures is a feasible goal but it requires careful consideration of alternatives and a consensus among interested parties. Titze (1994) outlines general steps toward standardization of acoustic measures of voice, and his comments are an excellent starting point for an effort to standardize procedures of acoustic analysis of speech for clinical and other purposes. This effort is best undertaken by a working group affiliated with a professional society. As suggested in this review, reasonable objectives are standardization of speech samples, time points for acoustic analysis (static and/or dynamic), and data graphing and reduction.

12.0. Conclusions

Substantial progress has been made in the collection of formant frequencies and bandwidths for vowels produced by speakers of both sexes over most of the lifespan. The main spectral

analysis methods used are FFT and LPC, but other methods such as cepstral analysis and time-frequency reassignment may be effective alternatives or complementary tools if they were more available in commonly used acoustic analysis software packages. Studies have focused on F1 and F2 frequencies, generally neglecting the higher formants F3 and F4 and all formant bandwidths, which would enrich the lifespan perspective of vowel acoustics. Among the summary indexes of vowel articulation, the F1-F2 VSA appears to be the most commonly used despite questions about its validity and the availability of a number of other measures of vowel distance or dispersion that may overcome some of the limitations of VSA. One attraction of VSA in its application to speech disorders is its association with the articulatory working space, which aids in articulatory-acoustic interpretation of vowels. However, given the heavy reliance of VSA on the corner vowel means without accounting for variability in formant measurements, other approaches or indices should be explored, including ones that also account for the higher formants. Clinical application would be enhanced by standardized methods (especially speech sample and measurement procedure), consideration of alternative metrics and data displays, and more extensive normative data for all four formants (including bandwidths as well as formant frequency).

Findings from the various studies considered in this review offer a lifespan perspective showing that (a) both formant frequencies and bandwidths decrease during typical speech development (summarized in tables 4 and 6), (b) additional decreases in f0 and F1 frequencies with aging are more likely in women than men, and (c) there is considerable variability in the results for particular age-sex groups, including adults. Major sources of variability across the lifespan probably can be attributed largely to methodological differences and dialect variations. Included in the former are differences in speech sample, selection of time point for measurement, and type of spectral analysis. Standardization of methods should improve the accuracy of formant estimation (i.e., decrease in measurement error) to help establish valid normative database across the lifespan for both sexes, as well as to enhance the comparability of data across studies of groups and individuals. These steps would facilitate the clinical application of formant measures. By minimizing the confounding of analysis error with true developmental variability, it should be possible to construct a more valid gauge of developmental variability related to factors such as anatomic growth or improved motor control.

Acknowledgments

This work was supported by NIH Research Grant R01 DC6282 (MRI and CT Studies of the Developing Vocal Tract, from the National Institute on Deafness and other Communicative Disorders (NIDCD) and by a core grant P30 HD03352 and U54 HD090256 to the Waisman Center, University of Wisconsin-Madison, from the National Institute of Child Health and Human Development (NICHD).

References

- Abdelli-Beruh NB, Wolk L, Slavin D. Prevalence of vocal fry in young adult male American English speakers. *Journal of Voice*. 2014; 28:185–190. [PubMed: 24315658]
- Alku P, Pohjalainen J, Vainio M, Laukkanen AM, Story BH. Formant frequency estimation of high-pitched vowels using weighted linear prediction. *The Journal of the Acoustical Society of America*. 2013; 134:1295–1313. [PubMed: 23927127]

- Amir, N., Amir, O. Novel measures for vowel reduction. In: Trouvain, J., editor. Proceedings of 16th International Congress of Phonetic Sciences. Saarbrücken: University des Saarlandes; 2007. p. 849-852.
- Andrianopoulos MV, Darrow K, Chen J. Multimodal standardization of voice among four multicultural populations formant structures. *Journal of Voice*. 2001; 15:61–77. [PubMed: 12269636]
- Angelocci AA, Kopp GA, Holbrook A. The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys. *Journal of Speech and Hearing Disorders*. 1964; 29:156–170. [PubMed: 14147499]
- Assmann PF, Nearey TM, Bharadwaj S. Analysis and classification of a vowel database. *Canadian Acoustics*. 2008; 36:148–149.
- Auger F, Flandrin P. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions in Signal Processing*. 1995; 43:1068–1089.
- Bang Y, Min K, Sohn YH, Cho SR. Acoustic characteristics of vowel sounds in patients with Parkinson disease. *NeuroRehabilitation*. 2013; 32:649–654. [PubMed: 23648619]
- Bauer HR, Kent RD. Acoustic analyses of infant fricative and trill vocalizations. *The Journal of the Acoustical Society of America*. 1987; 81:505–511. [PubMed: 3558968]
- Benavides AM, Murillo JLB, Pozo RF, Cuadros FE, Toledano DT, Alcázar-Ramírez JD, Gómez LAH. Formant frequencies and bandwidths in relation to clinical variables in an obstructive sleep apnea population. *Journal of Voice*. 2016; 30:21–29. [PubMed: 25795368]
- Bele IV. The speaker's formant. *Journal of Voice*. 2006; 20:555–578. [PubMed: 16325374]
- Bennett S. Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*. 1981; 69:231–238. [PubMed: 7217521]
- Berisha V, Sandoval S, Utianski R, Liss J. Characterizing the distribution of the quadrilateral vowel space area. *The Journal of the Acoustical Society of America*. 2014; 135:421–427. [PubMed: 24437782]
- Bickley, CA. Acoustic evidence for the development of speech. Cambridge, MA: Massachusetts Institute of Technology; 1989. Research Laboratory of Electronics (RLE) Technical Report No. 548
- Blomgren M, Robb M, Chen Y. A note on vowel centralization in stuttering and nonstuttering individuals. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1042–1051.
- Bogert BP. On the bandwidth of vowel formants. *Journal of the Acoustical Society of America*. 1953; 25:791–792.
- Bogert, BP., Healy, MJR., Tukey, JW. The quefrency analysis of time series for echoses: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking. In: Rosenblatt, M., editor. Time series analysis. 1963. p. 209-243.
- Bozeman, KW. Practical vocal acoustics. Whitehill, NY: Pendragon Press; 2013.
- Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech. I. Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*. 1996; 20:255–272. [PubMed: 21461127]
- Bunton K, Leddy M. An evaluation of articulatory working space in vowel production of adults with Down syndrome. *Clinical Linguistics & Phonetics*. 2011; 25:321–334. [PubMed: 21091205]
- Bunton K, Story B. Arizona child acoustic database repository. *Folia Phoniatrica et Logopaedica*. 2016; 68:107–111. Available at: <http://arizona.openrepository.com/arizona/handle/10150/316065>. [PubMed: 27784009]
- Burris C, Vorperian HK, Fourakis M, Kent RD, Bolt DM. Quantitative and descriptive comparison of four acoustic analysis systems: vowel measurements. *Journal of Speech, Language, and Hearing Research*. 2014; 57:26–45.
- Carré, R. Signal dynamics in the production and perception of vowels. In: Chitoran, I. Coupé, C. Marsico, E., Pellegrino, F., editors. Approaches to phonological complexity. Berlin: Mouton de Gruyter; 2009. p. 59-81.
- Cartei V, Cowels W, Banerjee R, Reby D. Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology*. 2014; 32:100–106. [PubMed: 24372318]

- Cartei V, Cowles HW, Reby D. Spontaneous voice gender imitation abilities in adult speakers. *PLoS one*. 2012; 7:e31353. [PubMed: 22363628]
- Cartei V, Reby D. Effect of formant frequency spacing on perceived gender in pre-pubertal children's voices. *PLoS ONE*. 2013; 8(12):e81022.doi: 10.1371/journal.pone.0081022 [PubMed: 24312517]
- Chen MY. Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*. 1997; 102:2360–2370. [PubMed: 9348695]
- Chen L, Evanini K, Sun X. Assessment of non-native speech using vowel space characteristics. *Spoken language technology workshop (SLT)*. 2010; :139–144. 2010 IEEE. DOI: 10.1109/SLT.2010.5700836
- Chesworth J, Coté K, Shaw C, Williams S, Hodge M. Effect of phonetic context on women's vowel area. *Canadian Acoustics/Acoustique canadienne*. 2003; 31:20–21.
- Childers DG, Wu K. Gender recognition from speech: Part II. Fine analysis. *The Journal of the Acoustical Society of America*. 1991; 90:1841–1856. [PubMed: 1755877]
- Chung H, Kong EJ, Edwards J, Weismer G, Fourakis M. Cross-linguistic studies of children's and adults' vowel spaces. *The Journal of the Acoustical Society of America*. 2012; 131:442–454. [PubMed: 22280606]
- Clopper CG, Pisoni DB, De Jong K. Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*. 2005; 118:1661–1676. [PubMed: 16240825]
- Cooley JW, Tukey JW. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*. 1965; 19:297–301.
- de Boer, B. Why women speak better than men (and its significance for evolution). In: Botha, R.Rudolf, Knight, C., editors. *The prehistory of language*. Oxford: Oxford University Press; 2009. p. 255-265.
- de Bruijn MJ, ten Bosch L, Kuik DJ, Quené H, Langendijk JA, Leemans CR, Verdonck-de Leeuw IM. Objective acoustic-phonetic speech analysis in patients treated for oropharyngeal cancer. *Folia Phoniatrica et Logopaedica*. 2009; 61:180–187. [PubMed: 19571552]
- Debruyne F, Decoster W. Acoustic differences between sustained vowels perceived as young or old. *Logopedics Phoniatrics Vocology*. 1999; 24:1–5.
- Delattre P. Un triangle acoustique des voyelles rales du Français. *The French Review*. 1948; 21:477–484.
- Derdemezis E, Vorperian HK, Kent RD, Fourakis M, Reinicke EL, Bolt DM. Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups. *American Journal of Speech-Language Pathology*. 2016; 25:335–354. [PubMed: 26501214]
- Disner SF. Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America*. 1980; 67:253–261. [PubMed: 7354193]
- Duckworth M, McDougall K, de Jong G, Schockey L. Improving the consistency of formant measurement. *International Journal of Speech, Language and the Law*. 2011; 18:35–51.
- Dunn HK. Methods of measuring vowel formant bandwidths. *The Journal of the Acoustical Society of America*. 1961; 33:1737–1746.
- Eguchi S, Hirsh IJ. Development of speech sounds in children. *Acta otolaryngologica Supplementum*. 1968; 257:1–51.
- Eichhorn, J., Kent, R., Austin, D., Vorperian, HK. Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice*. 2017. Available online: <https://doi.org/10.1016/j.jvoice.2017.08.003>
- Eliasova I, Mekyska J, Kostolova M, Marecek R, Smekal Z, Rektorova I. Acoustic evaluation of short-term effects of repetitive transcranial magnetic stimulation on motor aspects of speech in Parkinson's disease. *Journal of Neural Transmission*. 2013; 120:597–605. [PubMed: 23254927]
- Endres W, Bambach W, Flosser G. Voice spectrograms as a function of age, voice disguise, and voice imitation. *The Journal of the Acoustical Society of America*. 1971; 49:1842–1847. [PubMed: 5125731]
- Fabricius AH, Watt D, Johnson DE. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change*. 2009; 21:413–435.

- Fant G. Formant bandwidth data. Quarterly Progress and Status Report, KTH (Royal Institute of Technology, Stockholm, Sweden). 1962; 3:001–002.
- Flandrin P, Auger F, Chassande-Mottin E. Time-Frequency reassignment from principles to algorithms. Applications in Time-frequency Signal Processing. 2002; 5:179–203.
- Fleischer M, Pinkert S, Mattheus W, Mainka A, Mürbe D. Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall. Biomechanics and Modeling in Mechanobiology. 2014; :1–15. published online 23 November 2014. DOI: 10.1007/s10237-014-0632-2
- Fletcher AR, McAuliffe MJ, Lansford KL, Liss JM. The relationship between speech segment duration and vowel centralization in a group of older speakers. The Journal of the Acoustical Society of America. 2015; 138:2132–2139. [PubMed: 26520296]
- Fletcher AR, McAuliffe MJ, Lansford KL, Liss JM. Assessing vowel centralization in dysarthria: A comparison of methods. Journal of Speech, Language, and Hearing Research. 2016; 60:341–354.
- Flipsen P, Lee S. Reference data for the American English acoustic vowel space. Clinical Linguistics and Phonetics. 2012; 26:926–933. [PubMed: 23057793]
- Fourakis M. Tempo, stress, and vowel reduction in American English. The Journal of the Acoustical Society of America. 1991; 90:1816–1827. [PubMed: 1960277]
- Fougeron, C., Audibert, N. Testing various metrics for the description of vowel distortion in dysarthria. In: Lee, W.-S., Zee, E., editors. Proceedings of the International Congress of Phonetic Sciences (ICPHs'11). Hong Kong: 2011. p. 687-690.
- Fox, RA., Jacewicz, E. Analysis of total vowel space areas in three regional dialects of American English; Acoustics'08. 2008. p. 495-500. www.acoustics08-paris.org
- Fox RA, Jacewicz E. Cross-dialectal variation in formant dynamics of American English vowels. The Journal of the Acoustical Society of America. 2009; 126:2603–2618. [PubMed: 19894839]
- Fox, RA., Jacewicz, E. Dialect and generational differences in vowel space areas. Proceedings of the Third ISCA Tutorial and Research Workshop on Experimental Linguistics, ExLing 2010; 25-27 August, 2010; Athens, Greece. International Speech Communication Association. 2010.
- Fox RA, Jacewicz E. Reconceptualizing the vowel space in analyzing regional variation. The Journal of the Acoustical Society of America. 2015; 138:1922–1922.
- Fujimura O. Analysis of nasal consonants. The Journal of the Acoustical Society of America. 1962; 34:1865–1875.
- Fulop SA. Accuracy of formant measurement for synthesized vowels using the reassigned spectrogram and comparison with linear prediction. The Journal of the Acoustical Society of America. 2010; 127:2114–2117. [PubMed: 20369992]
- Fulop, S. Speech spectrum analysis. Berlin: Springer-Verlag; 2011.
- Fulop SA, Kari E, Ladefoged P. An acoustic study of the tongue root contrast in Degema vowels. Phonetica. 1998; 55:80–98. [PubMed: 9693345]
- Gargouri, D., Kammoun, MA., Hamida, AB. In: Demiralp, M.Aan, A., Mastorakis, N., editors. A comparative study of formant frequencies estimation techniques; Proceedings of the 5th World Scientific and Engineering Academy and Society (WSEAS) International Conference on signal Processing; Istanbul, Turkey. May 27-29; 2006. p. 15-19.
- Gilbert HR, Robb MP, Chen Y. Formant frequency development: 15 to 36 months. Journal of Voice. 1997; 11:260–266. [PubMed: 9297669]
- Halberstam B, Raphael LJ. Vowel normalization: the role of fundamental frequency and upper formants. Journal of Phonetics. 2004; 32:423–434.
- Hagiwara, R. Acoustic realizations of American /r/ as produced by women and men Unpublished PhD thesis. University of California at Los Angeles; 1995.
- Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice quality. Acta Otolaryngologica. 1980; 90:441–451.
- Hanna N, Smith J, Wolfe J. Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation. The Journal of the Acoustical Society of America. 2016; 139:2924–2936. [PubMed: 27250184]

- Hanson HM. Glottal characteristics of female speakers: acoustic correlates. *Journal of the Acoustical Society of America*. 1997; 101:466–481. [PubMed: 9000737]
- Hanson HM, Chuang ES. Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America*. 1999; 106:1064–1077. [PubMed: 10462811]
- Haque S, Ali MH, Haque AF. Variability of acoustic features of hypernasality and its assessment. *International Journal of Applied Engineering Research*. 2016; 11:9262–9268.
- Harrington, J., Cassidy, S. *Techniques in speech acoustics*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1999.
- Harrington, J., Palethorpe, S., Watson, CI. Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech 2007). International Speech Communication Association; 2007. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers; p. 2753-2756. available at <http://www.isca-speech.org/iscaweb/index.php/archive/online-archive>
- Hawkins S, Stevens KN. Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels. *The Journal of the Acoustical Society of America*. 1985; 77:1560–1575. [PubMed: 3989111]
- Hazan V. Speech communication across the lifespan. *Acoustics Today*. 2017; 13:36–43.
- Hazan V, Markham D. Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America*. 2004; 116:3108–3118. [PubMed: 15603156]
- Heald SLM, Nusbaum HC. Variability in vowel production within and between days. *PLOS*. 2015; September 2, 2015. doi: 10.1371/journal.pone.0136791
- Hermann L. Beiträge zur Lehre von der Klangwahrnehmung. *Pflügers Archive*. 1894; 56:467–499.
- Higgins CM, Hodge MM. Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology*. 2002; 10:271–277.
- Hillenbrand, JM. Static and dynamic approaches to vowel perception. In: Morrison, GS., Assmann, PF., editors. *Vowel inherent spectral change*. Berlin, Heidelberg: Springer; 2013. p. 9-30.
- Hillenbrand J, Clark M, Nearey T. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*. 2001; 109:748–763. [PubMed: 11248979]
- Hillenbrand J, Gayvert RT. Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech, Language, and Hearing Research*. 1993; 36:694–700.
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*. 1995; 97:3099–3111. [PubMed: 7759650]
- Hirsch, F., Bouarourou, F., Vaxelaire, B., Monfrais-Pfauwadel, MC., Bechet, M., Sturm, J., Sock, R. Formant structures of vowels produced by stutterers in normal and fast speech rates. 8th International Seminar on Speech Production; Dec 2008; France. 2008. p. NC
- House AS, Stevens KN. Estimation of formant bandwidths from measurements of transient response of the vocal tract. *Journal of Speech and Hearing Research*. 1958; 1:309–315.
- Huffman MK. The role of F1 amplitude in producing nasal percepts. *The Journal of the Acoustical Society of America*. 1990; 88:S54–S54.
- Huggins AWF. Better spectrograms from children's speech. *Journal of Speech, Language, and Hearing Research*. 1980; 23:19–27.
- Hunter EJ, Spielman J, Starr AC, Popolo PS. Acoustic Voice Recording, "I am seeking recommendations for voice recording hardware...". *Perspectives on Voice and Voice Disorders*. 2007; 17:7–14. 2007. <http://div3perspectives.asha.org/content/17/3/7.full.pdf>. Downloaded 17 January 2016. Also see: http://www.ncvs.org/e-learning/NCVSONlineTechnicalMemo15-vs1_1.pdf. DOI: 10.1044/vvd17.3.7
- Hustad KC, Gorton K, Lee J. Classification of speech and language profiles in 4-year-old children with cerebral palsy: a prospective preliminary study. *Journal of Speech, Language, and Hearing Research*. 2010; 53:1496–1513.
- Jacewicz E, Fox RA. Dialectal and age-related acoustic variation in vowels in spontaneous speech. *The Journal of the Acoustical Society of America*. 2002; 132:2002.

- Jacewicz E, Fox RA. Assessing vowel space area metrics in the context of dialect variation. *The Journal of the Acoustical Society of America*. 2017; 141:3980–3980.
- Jacewicz E, Fox RA, Salmons J. Regional dialect variation in the vowel systems of typically developing children. *Journal of Speech, Language, and Hearing Research*. 2011; 54:448–470.
- Jin SH, Liu C. The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *The Journal of the Acoustical Society of America*. 2013; 133:EL363–EL369. [PubMed: 23656095]
- Joos, M. Supplement to *Language*. Vol. 24. Baltimore: The Linguistic Society of America; 1948. Acoustic phonetics.
- Jung, Y. Acoustic articulatory evidence for quantal vowel categories: the features [low] and [back] PhD dissertation. Massachusetts Institute of Technology; 2009.
- Kaipa R, Robb MP, O’Beirne GA, Allison RS. Recovery of speech following total glossectomy: An acoustic and perceptual appraisal. *International Journal of Speech-language Pathology*. 2012; 14:24–34. [PubMed: 22257069]
- Karlsson F, van Doorn J. Vowel formant dispersion as a measure of articulation proficiency. *The Journal of the Acoustical Society of America*. 2012; 132:2633–2641. [PubMed: 23039456]
- Karlsson HB, Shriberg LD, Flipsen PF Jr, McSweeney JL. Acoustic phenotypes for speech-genetics studies: Toward an acoustic marker for residual/s/distortions. *Clinical Linguistics & Phonetics*. 2002; 16:403–424. [PubMed: 12469448]
- Kataoka R, Michi KI, Okabe K, Miura T, Yoshida H. Spectral properties and quantitative evaluation of hypernasality in vowels. *The Cleft Palate-craniofacial Journal*. 1996; 33:43–50. [PubMed: 8849858]
- Kent RD. Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech, Language, and Hearing Research*. 1976; 19:421–447.
- Kent RD, Kent JF, Rosenbek JC. Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders*. 1987; 52:367–387. [PubMed: 3312817]
- Kent RD, Kim Y. Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics*. 2003; 17:427–445. [PubMed: 14564830]
- Kent RD, Murray AD. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*. 1982; 72:353–365. [PubMed: 7119278]
- Kent RD, Weismer G, Kent JF, Vorperian HK, Duffy JR. Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders*. 1999; 32:141–186. [PubMed: 10382143]
- Kim H, Hasegawa-Johnson M, Perlman A. Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*. 2011; 63:187–194. [PubMed: 20938200]
- Kim S, Kim JH, Ko DH. Characteristics of vowel space and speech intelligibility in patients with spastic dysarthria. *Communication Sciences & Disorders*. 2014; 19:352–360.
- Kochanski G, Grabe E, Coleman J, Rosner B. Loudness predicts prominence: fundamental frequency lends little. *The Journal of the Acoustical Society of America*. 2005; 118:1038–1054. [PubMed: 16158659]
- Kohn, M., Farrington, C. Longitudinal sociophonetic analysis. What to expect when working with child and adolescent data. In: Wagner, SE., Buchstaller, I., editors. *Panel studies of variation and change*. New York: Routledge; 2018. p. 122-152.
- Koops C. /u/-fronting is not monolithic: Two types of fronted /u/ in Houston Anglos. *University of Pennsylvania Working Papers in Linguistics*. 2010; 16:113–122.
- Krause JC, Braid LD. Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*. 2004; 115:362–378. [PubMed: 14759028]
- Krishna Y, Rajashekhar B. Association between bandwidths of vowel formants and age, gender and consonant context in Telugu. *Language in India*. 2013; 13:574–584. www.languageinindia.com.
- Kuhl PK, Meltzoff AN. Infant vocalizations in response to speech: vocal imitation and developmental change. *The Journal of the Acoustical Society of America*. 1996; 100:2425–2438. [PubMed: 8865648]

- Kwon HB. Gender difference in speech intelligibility using speech intelligibility tests and acoustic analyses. *Journal of Advanced Prosthodontics*. 2010; 2:71–76. [PubMed: 21165272]
- Labov, W., Ash, S., Boberg, C. *The atlas of North American English: phonetics, phonology, and sound change: a multimedia reference tool*. Berlin: Mouton de Gruyter; 2006.
- Ladefoged, P. *Three areas of experimental phonetics*. London: Cambridge University Press; 1967.
- Lam J, Tjaden K, Wilding G. Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*. 2012; 55:1807–1821.
- Lee ASY, Ciocca V, Whitehill TL. Acoustic correlates of hypernasality. *Clinical Linguistics & Phonetics*. 2003; 17:259–264. [PubMed: 12945600]
- Lee S, Potamianos A, Narayanan S. Acoustics of children’s speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*. 1999; 105:1455–1468. [PubMed: 10089598]
- Leino T, Laukkanen AM, Radolf V. Formation of the actor’s/speaker’s formant: a study applying spectrum analysis and computer modeling. *Journal of Voice*. 2011; 25:150–158. [PubMed: 20456915]
- Liljencrants J, Lindblom B. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*. 1972; 48:839–862.
- Lin E, Hornibrook J, Ormond T. Evaluating iPhone recordings for acoustic voice assessment. *Folia Phoniatrica et Logopaedica*. 2012; 64:122–130. [PubMed: 22584162]
- Lindblom, B. Accuracy and limitations of sonagraph measurements. In: Sovijärvi, A., Aalto, P., editors. *Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki*. The Hague: Mouton; 1962. p. 188-202.
- Lindblom B. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*. 1963; 35:1773–1781.
- Linville SE, Fisher HB. Acoustic characteristics of women’s voices with advancing age. *Journal of Gerontology*. 1985; 40:324–330. [PubMed: 3989246]
- Linville SE, Rens J. Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice*. 2001; 15:323–330. [PubMed: 11575629]
- Liu HM, Tsao FM, Kuhl PK. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*. 2005; 117:3879–3889. [PubMed: 16018490]
- Llisterri, J. *Speech analysis and transcription tools*. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona; 2015. Available at: http://liceu.uab.cat/~joaquim/phonetics/fon_anal_acus/herram_anal_acus.html. Downloaded 17 January 2016
- Ludlow, CL., Kent, RD., Gray, LC. *Measuring voice, speech and swallowing in the clinic and laboratory*. San Diego: Plural Publishing; 2019.
- Lulich SM. Subglottal resonances and distinctive features. *Journal of Phonetics*. 2010; 38:20–32.
- Lulich, SM., Arsikere, H., Morton, JR., Leung, GK., Alwan, A., Sommers, M. *Proceedings of the 12th Annual Conference of the International Speech Communication Association 2011 (INTERSPEECH 2011)*. Baixas, France: International Speech Communications Association; 2011. Analysis and automatic estimation of children’s subglottal resonances; p. 2817-2820.
- Lulich SM, Morton JR, Arsikere H, Sommers MS, Leung GK, Alwan A. Subglottal resonances of adult male and female native speakers of American English. *The Journal of the Acoustical Society of America*. 2012; 132:2592–2602. [PubMed: 23039452]
- Lyakso, E., Frolova, O., Grigorev, A. A comparison of acoustic features of speech of typically developing children and children with autism spectrum disorders. In: Potapova, AR., Németh, G., editors. *Speech and computer SPECOM 2016 Lecture notes in computer science*. Vol. 981. Berlin: Springer, Cham; 2016. p. 43-50.
- Mahler LA, Ramig LO. Intensive treatment of dysarthria secondary to stroke. *Clinical Linguistics & Phonetics*. 2012; 26:681–694. [PubMed: 22774928]
- Markel, JD., Gray, AH, Jr. *Linear prediction of speech*. New York: Springer-Verlag; 1976.
- Maurer, D. *Acoustics of the vowel: Preliminaries*. Bern: Peter Lang; 2016.

- McAuliffe MJ, Fletcher AR, Kerr SE, O'Beirne GA, Anderson T. Effect of dysarthria type, speaking condition, and listener age on speech intelligibility. *American Journal of Speech-Language Pathology*. 2016; 26:113–123.
- McCloy D, Wright R, Souza P. Modeling intrinsic intelligibility variation: vowel-space size and structure. *Proceedings of Meetings on Acoustics*. 2014; 18:060007.doi: 10.1121/1.4870070
- McGowan RW, McGowan RS, Denny M, Nittrouer S. A longitudinal study of very young children's vowel production. *Journal of Speech, Language, and Hearing Research*. 2014; 57:1–15.
- Medabalimi AJX, Seshadri G, Bayya Y. Extraction of formant bandwidths using properties of group delay functions. *Speech Communication*. 2014; 63:70–83.
- Mehta DD, Wolfe PJ. Statistical properties of linear prediction analysis underlying the challenge of formant bandwidth estimation. *The Journal of the Acoustical Society of America*. 2015; 137:944–950. [PubMed: 25698026]
- Milenkovic, P. Time-frequency analysis software program for 32-bit Windows (Alpha Version 1.2) [Computer software]. 2010. Available from <http://userpages.chorus.net/cspeech/>
- Monahan PJ, Idsardi WJ. Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and Cognitive Processes*. 2010; 25(6):808–839. [PubMed: 20606713]
- Monsen RB, Engebretson AM. The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction. *Journal of Speech, Language, and Hearing Research*. 1983; 26:89–97.
- Monsen RB, Engebretson AM. The Accuracy of Formant Frequency Measurements: A Comparison of Spectrographic Analysis and Linear Prediction. *Journal of Speech, Language, and Hearing Research*. 1983; 26:89–97.
- Morrison, GS., Assmann, PF., editors. *Vowel inherent spectral change*. Berlin: Springer Verlag; 2013.
- Morrison GS, Nearey TM. Testing theories of vowel inherent spectral change. *The Journal of the Acoustical Society of America*. 2007; 122:EL15–EL22. DOI: 10.1121/1.2739111 [PubMed: 17614372]
- Moura CP, Cunha LM, Vilarinho H, Cunha MJ, Freitas D, Palha M, Poeschel SM, Pais-Clemente M. Voice parameters in children with Down syndrome. *Journal of Voice*. 2008; 22:34–42. [PubMed: 17014985]
- Munson B, Solomon NP. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*. 2004; 47:1048–1058.
- Narasimhan SV, Nikitha M, Francis N. Articulatory working space area in children with cerebral palsy. *International Journal of Health Sciences and Research (IJHSR)*. 2016; 6:335–341.
- Nearey TM. Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*. 1989; 85:2088–2113. [PubMed: 2659638]
- Neel AT. Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*. 2008; 51:574–585.
- Nolan F, Grigoras C. A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law*. 2005; 12:143–173.
- Noll AM. Cepstrum pitch determination. *The Journal of the Acoustical Society of America*. 1967; 41:293–309. [PubMed: 6040805]
- Oller DK, Niyogi P, Gray S, Richards JA, Gilkerson J, Xu D, Yapanel U, Warren SF. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*. 2010; 107:13354–13359.
- Oppenheim AV, Shafer RW. From frequency to quefrequency: a history of the cepstrum. *IEEE Signal Processing Magazine*. Sep.2004 :95–106.
- Palethorpe S, Watson CI. Acoustic analysis of monophthong and diphthong production in acquired severe to profound hearing loss. *The Journal of the Acoustical Society of America*. 2003; 114:1055–1068. [PubMed: 12942984]
- Penny, L., Russell, A., Pemberton, C. Some speech and acoustic measures of the aging voice. In: Brockhaus-Grand, W., Togneri, R., editors. *Proceedings of the 5th Australian international conference on speech science & technology*. Perth, Canberra: ASSTA; 1994. p. 76-80. Available online at: <http://www.assta.org/?q=sst-conferences>

- Perry TL, Ohde RN, Ashmead DH. The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*. 2001; 109:2988–2998. [PubMed: 11425141]
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Pettinato M, Tuomainen O, Granlund S, Hazan V. Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics*. 2016; 54:1–14.
- Philips, BJ., Kent, RD. Acoustic phonetic descriptions of speech production in persons with cleft palate and other velopharyngeal disorders. In: Lass, NJ., editor. *Speech and language: Advances in basic research and practice (Vol 10)*. New York: Academic Press; 1984.
- Plichta, B. Best practices in the acquisition, processing, and analysis of acoustic speech signals. (no date given) Available from: <http://flint.matrix.msu.edu/extras/Audio-technology.pdf>. Downloaded 17 January 2016
- Potamianos A, Narayanan S. A review of the acoustic and linguistic properties of children's speech. *IEEE 9th Workshop on Multimedia Signal Processing (MMSP), 2007*. 2007:22–25.
- Pruthi T, Espy-Wilson CY, Story BH. Simulation and analysis of nasalized vowels based on magnetic resonance imaging data. *The Journal of the Acoustical Society of America*. 2007; 121:3858–3873. [PubMed: 17552733]
- Qian J, Fu F, Liu X, He L, Yin H, Zhang H. The analysis and detection of hypernasality based on a formant extraction algorithm. *Journal of Physics: Conference Series*. 2017; 887:012082.doi: 10.1088/1742-6596/887/1/012082
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2016. URL <https://www.R-project.org/>
- Rastatter MP, Jacques RD. Formant frequency structure of the aging male and female vocal tract. *Folia Phoniatica*. 1990; 42:312–319. [PubMed: 2081583]
- Rastatter MP, McGuire RA, Kalinowski J, Stuart A. Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatica et Logopaedica*. 1997; 49:1–8. [PubMed: 9097490]
- Rendall D, Kollias S, Ney C, Lloyd P. Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size. *The Journal of the Acoustical Society of America*. 1995; 117:944–955.
- Robb MP, Cacace AT. Estimation of formant frequencies in infant cry. *International Journal of Pediatric Otorhinolaryngology*. 1995; 32:57–67. [PubMed: 7607821]
- Robb MP, Chen Y. Is /h/ phonetically neutral? *Clinical Linguistics and Phonetics*. 2009; 23:842–855. [PubMed: 19891524]
- Robb MP, Chen Y, Gilbert HR. Developmental aspects of formant frequency and bandwidth in infants and toddlers. *Folia Phoniatica et Logopaedica*. 1997; 49:88–95. [PubMed: 9197091]
- Robb MP, Saxman JH. Developmental trends in vocal fundamental frequency of young children. *Journal of Speech and Hearing Research*. 1985; 28:421–427. [PubMed: 2981049]
- Robb MP, Yates J, Morgan EJ. Vocal tract resonance characteristics of adults with obstructive sleep apnea. *Acta Oto-Laryngologica*. 1997; 117:760–763. [PubMed: 9349877]
- Roy N, Nissen SL, Dromey C, Sapir S. Articulatory changes in muscle tension dysphonia: evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of Communication Disorders*. 2009; 42:124–135. [PubMed: 19054525]
- Rusz J, Cmejla R, Tykalova T, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E. Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. *The Journal of the Acoustical Society of America*. 2013; 134:2171–2181. [PubMed: 23967947]
- Sandoval S, Berisha V, Utianski RL, Liss JM, Spanias A. Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*. 2013; 134:EL477–EL483. [PubMed: 24181994]

- Sapir S, Raming LO, Spielman JL, Fox C. Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research*. 2010; 53:114–125.
- Sapir S, Polczynska M, Tobin Y. Why does the vowel space area as an acoustic metric fail to differentiate dysarthric from normal vowel articulation and what can be done about it? *Poznan Studies in Contemporary Linguistics*. 2009; 45:301–311.
- Sapir S, Spielman JL, Ramig LO, Story BH, Fox C. Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research*. 2007; 50:899–912.
- Scherer S, Lucas G, Gratch J, Rizzo A, Morency LP. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. *IEEE Transactions on Affective Computing*. (in press).
- Scherer S, Morency LP, Gratch J, Pestian J. Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015:4789–4793.
- Schötz, S. Acoustic analysis of adult speaker age. In: Müller, C., editor. *Speaker classification I. Fundamentals, features, and methods*. Berlin: Springer; 2007. p. 88-107.
- Scukanec G, Petrosino L, Squibb K. Formant frequency characteristics of children, young adult, and aged female speakers. *Perceptual and Motor Skills*. 1991; 73:203–208. [PubMed: 1945691]
- Sebastian S, Babu S, Oommen NE, Ballraj A. Acoustic measurements of geriatric voice. *Journal of Laryngology and Voice*. 2012; 2:81–84.
- Shadle CH, Nam H, Whalen DH. Comparing measurement errors for formants in synthetic and natural vowels. *The Journal of the Acoustical Society of America*. 2016; 139:713–727. [PubMed: 26936555]
- Skodda S, Grönheit W, Schlegel U. Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS ONE*. 2012; :e32132.doi: 10.1371/journal.pone.0032132 [PubMed: 22389682]
- Sluijter AMC, van Heuven V. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*. 1966; 100:2471–2485.
- Smiljani R, Bradlow AR. Speaking and hearing clearly: talker and listener factors in speaking style changes. *Language and Linguistics Compass*. 2009; 3:236–264. [PubMed: 20046964]
- Sørensen MK, Durck TT, Bork KH, Rasmussen N. Normative values and interrelationship of MDVP voice analysis parameters before and after endotracheal intubation. *Journal of Voice*. 2016; 30:626–630. [PubMed: 26234917]
- Stevens K, House A. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*. 1965; 6:111–121.
- Story BH, Bunton K. Formant measurement in children's speech based on spectral filtering. *Speech Communication*. 2016; 76:93–111.
- Story BH, Bunton K. Vowel space density as an indicator of speech performance. *The Journal of the Acoustical Society of America*. 2017; 141:EL458.doi: 10.1121/1.4983342 [PubMed: 28599542]
- Strange W. Evolving theories of vowel perception. *The Journal of the Acoustical Society of America*. 1989; 85:2081–2087. [PubMed: 2659637]
- Styler, W. On the acoustical and perceptual features of vowel nasality, Unpublished Ph D dissertation. University of Colorado; Boulder, CO: 2015.
- Sundberg J, Nordenberg M. Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average-spectra of speech. *The Journal of the Acoustical Society of America*. 2006; 120:453–457. [PubMed: 16875241]
- Sustainable Heritage Network. Digital audio best practices Version 2.1. 2006. Available online at <http://sustainableheritagenetwork.org/content/digital-audio-best-practices-version-21>
- Švancara P, Horáček J, Vokrál J, Cerný L. Computational modelling of effect of tonsillectomy on voice production. *Logopedics Phoniatrics Vocology*. 2006; 31:117–125.
- Švec JG, Granqvist S. Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*. 2010; 19:356–368. [PubMed: 20601621]

- Takatsu J, Hanai N, Suzuki H, Yoshida M, Tanaka Y, Tanaka S, Hasegawa Y, Yamamoto M. Phonologic and acoustic analysis of speech following glossectomy and the effects of rehabilitation on speech outcomes. *Journal of Oral and Maxillofacial Surgery*. 2017; 75:1530–1541. [PubMed: 28039737]
- Takemoto H, Kitamura T, Honda K, Masaki S. Deformation of the hypopharyngeal cavities due to F_0 changes and its acoustic effects. *Acoustical Science and Technology*. 2008; 29:300–303.
- Thomas, Erik R., Kendall, Tyler. NORM: The vowel normalization and plotting suite [software]. 2007. Available at: <http://ncslaap.lib.ncsu.edu/tools/norm/index.php>
- Titze IR. Toward standards in acoustic analysis of voice. *Journal of Voice*. 1994; 8:1–7. [PubMed: 8167782]
- Titze IR, Baken RJ, Bozeman KW, Granqvist S, Henrich N, Herbst CT, Howard DM, Hunter EJ, Kaelin D, Kent RD, Kreiman J, Kob M, Löftqvist A, McCoy S, Miller DG, Noë H, Scherer RC, Smith JR, Story BH, Švec JG, Ternström S, Wolfe J. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America*. 2015; 137:3005–3007. [PubMed: 25994732]
- Tiwari M. Speech acoustics: How much science? *Journal of Natural Science, Biology, and Medicine*. 2012; 3:24.
- Torre P, Barlow JA. Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*. 2009; 42:324–333. [PubMed: 19394957]
- Traunmüller, H., Eriksson, A. A method of measuring formant frequencies at high fundamental frequencies. In: Kokkinakis, G, Fakotakis, N., Dermatas, E., editors. *Proceedings of EuroSpeech'97, Fifth European Conference on Speech Communication and Technology*. Vol. 1. Rhodes, Greece: ISCA; 1997. p. 477-480.
- Tsao YC, Weismer G, Iqbal K. The effect of intertalker speech rate variation on acoustic vowel space. *The Journal of the Acoustical Society of America*. 2006; 119:1074–1082. [PubMed: 16521769]
- Turner GS, Tjaden K, Weismer G. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*. 1995; 38:1001–1013.
- Turner RE, Walters TC, Monaghan JJM, Patterson RD. A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *The Journal of the Acoustical Society of America*. 2009; 125:2374–2386. [PubMed: 19354411]
- Vallabha GK, Tuller B. Systematic errors in the formant analysis of steady-state vowels. *Speech Communication*. 2002; 38:141–160.
- VanDam M, Warlaumont AS, Bergelson E, Cristia A, Soderstrom M, De Palma P, MacWhinney B. HomeBank: an online repository of daylong child-centered audio recordings. *Seminars in Speech and Language*. 2016; 37:128–142. Epub 2016 April 25. DOI: 10.1055/s-0036-1580745 [PubMed: 27111272]
- Vilain C, Berthommier F, Boë LJ. A brief history of articulatory-acoustic vowel representation. 1st International Workshop on the History of Speech Communication Research (HSCR 2015), Sep 2015, Dresede, France. *Proceedings of the 1st International Workshop on the History of Speech Communication Research 2015*. 2015
- Vogel AP, Fletcher J, Snyder PJ, Fredrickson A, Maruff P. Reliability, stability, and sensitivity to change and impairment in acoustic measures of timing and frequency. *Journal of Voice*. 2011; 25:137–149. [PubMed: 20171828]
- Vogel AP, Rosen KM, Morgan AT, Reilly S. Comparability of modern recording devices for speech analysis: smartphone, landline, laptop, and hard disc recorder. *Folia Phoniatrica et Logopaedica*. 2015; 66:244–250.
- Vorperian, HK. Anatomic development of the vocal tract structures as visualized by MRI. University of Wisconsin-Madison; Madison, WI: 2000. Unpublished doctoral dissertation
- Vorperian HK, Kent RD. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*. 2007; 50:1510–1545.
- Vorperian HK, Kurtzweil SL, Fourakis M, Kent RD, Tillman KK, Austin D. Effect of body position on vocal tract acoustics: Acoustic pharyngometry and vowel formants. *The Journal of the Acoustical Society of America*. 2015; 138:833–845. [PubMed: 26328699]

- Vorperian HK, Wang S, Schimek EM, Durtschi RB, Kent RD, Gentry LR, Chung MK. Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research*. 2011; 54:995–1010.
- Wang TT, Quatieri TF. High-pitch formant estimation by exploiting temporal change of pitch. *IEEE Transactions on Audio, Speech, and Language Processing*. 2010; 18:171–187.
- Watson, PJ., Munson, B. A comparison of vowel acoustics between older and younger adults. In: Barry, WJ., Trouvain, J., editors. *Proceedings of the XVIth International Congress of Phonetic Sciences: ICPhS XVI*; 6-10 August; Saarbrücken, Germany. Saarbrücken: Univ. des Saarlandes; 2007.
- Weismer G, Jeng JY, Laures JS, Kent RD, Kent JF. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniarica et Logopaedica*. 2001; 53:1–18.
- Welling L, Ney H. Formant estimation for speech recognition. *IEEE Transactions on Speech and Audio Processing*. 1998; 6:36–48.
- Wenke RJ, Cornwell P, Theodoros DG. Changes to articulation following LSVT® and traditional dysarthria therapy in non-progressive dysarthria. *International Journal of Speech-Language Pathology*. 2010; 12:203–220. [PubMed: 20433339]
- White P. Formant frequency analysis of children's spoken and sung vowels using sweeping fundamental frequency production. *Journal of Voice*. 1999; 13:570–582. [PubMed: 10622522]
- Whiteside SP. Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America*. 2001; 110:464–478. [PubMed: 11508971]
- Whiteside SP, Hodgson C. Acoustic characteristics in 6- 10-year-old children's voices: some preliminary findings. *Logopedics Phoniatrics Vocology*. 1999; 24:6–13.
- Whitehill TL, Ciocca V, Chan JCT, Samman N. Acoustic analysis of vowels following glossectomy. *Clinical Linguistics & Phonetics*. 2006; 20:135–140. [PubMed: 16428229]
- Whitfield JA, Goberman AM. Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *International Journal of Speech-Language Pathology*. 2016:1–11.
- Wolk L, Abdelli-Beruh NB, Slavin D. Habitual use of vocal fry in young adult female speakers. *Journal of Voice*. 2012; 26:e111–e116. [PubMed: 21917418]
- Wright, R. Factors of lexical competition in vowel articulation. In: Local, J., Ogden, R., Temple, R., editors. *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press; 2004. p. 26-50.
- Yang B. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*. 1996; 24:245–261.
- Yang S, Mu L. An investigation of the third formant of /a/ in prepubertal children. *Journal of Voice*. 1989; 3:321–323.
- Yang J, L FL, Nehorai A. Spectral contrast enhancement: algorithms and comparisons. *Speech Communication*. 2003; 39:33–46.
- Yildirim, S., Narayanan, S., Byrd, D., Khurana, S. Acoustic analysis of preschool children's speech. In: Solé, MJ., Recasens, D., Romero, J., editors. *Proceedings of the 15th International Congress of Phonetic Sciences*; 15th ICPhS; Barcelona. 3-9 August 2003; Barcelona: Universitat Autònoma de Barcelona; 2003. p. 949-952.
- Xue A, Jiang J, Lin E, Glassenberg R, Mueller PB. Age-related changes in human vocal tract configurations and the effects on speakers' vowel formant frequencies: a pilot study. *Logopedics Phoniatrics Vocology*. 1998; 24:132–137.
- Xue SA, Hao GJ. Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. *Journal of Speech, Language, and Hearing Research*. 2003; 46:689–701.
- Xue SA, Lam CW, Whitehill TL, Samman N. Effects of Class III malocclusion on young male adults' vocal tract development: A pilot study. *Journal of Oral and Maxillofacial Surgery*. 2011; 69:845–852. [PubMed: 20674124]
- Zahorian SA, Jagharghi AJ. Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*. 1993; 94:1966–1982. [PubMed: 8227741]

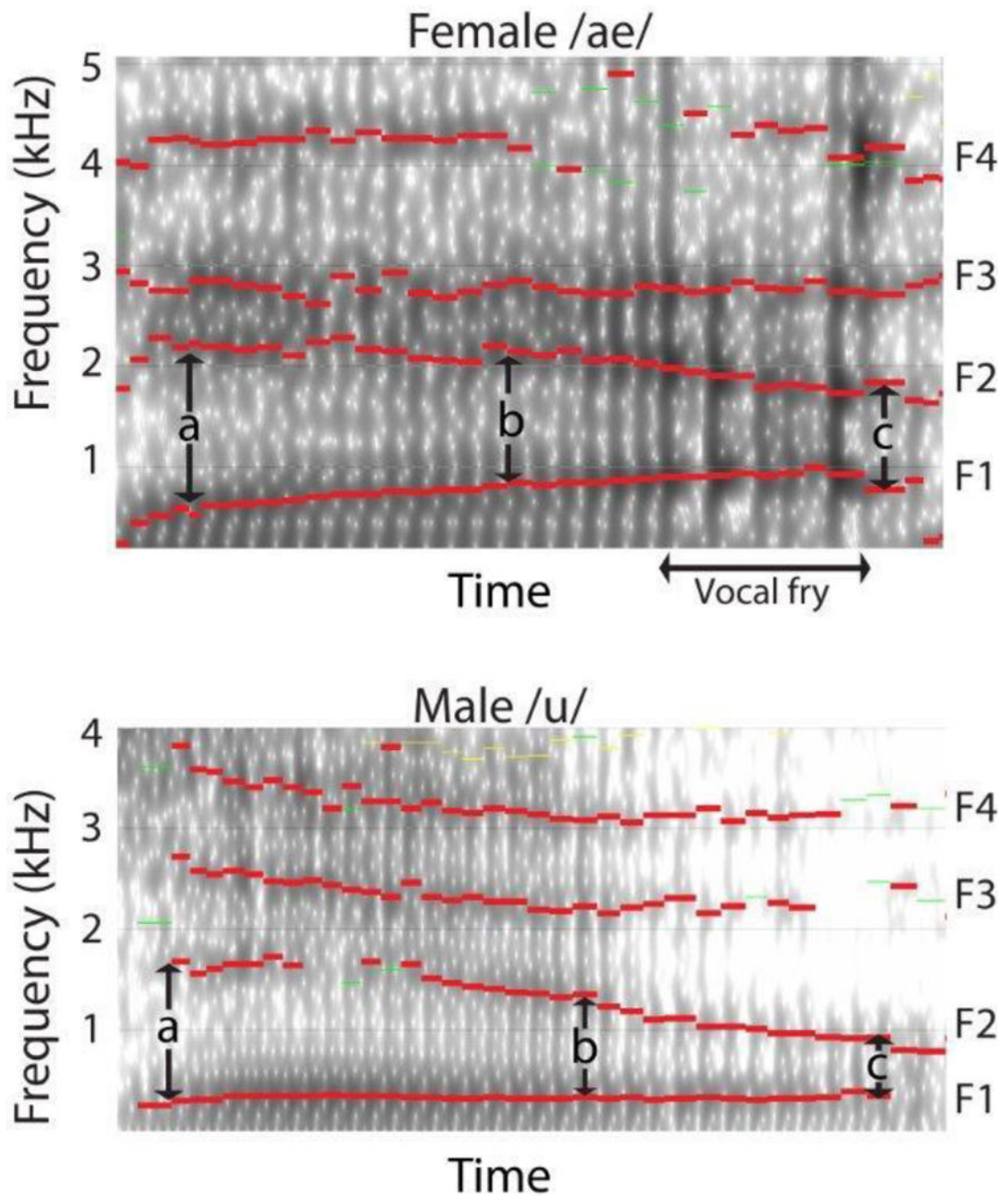


Figure 1.

Upper part: spectrogram with LPC formant tracks for vowel /ae/ produced by a young woman, illustrating a diphthongal formant pattern and vocal fry at the end of the vocalic segment. *Lower part:* Spectrogram with LPC formant tracks for vowel /u/, produced by a young man, illustrating continuous decline in F2 frequency over the duration of the vowel. In both spectrograms, formants are labeled F1 to F4 and frequency in kHz is shown at the left margin. The vertical lines indicate three potential time points of formant measurement: (a) near vowel onset, (b) midpoint, and (c) near vowel end. The double-headed arrows

indicate F2-F1 separation at different time points. Spectrograms are from analyses done with the software TF32 (Milenkovic, 2010), with unsmoothed LPC formant tracks.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

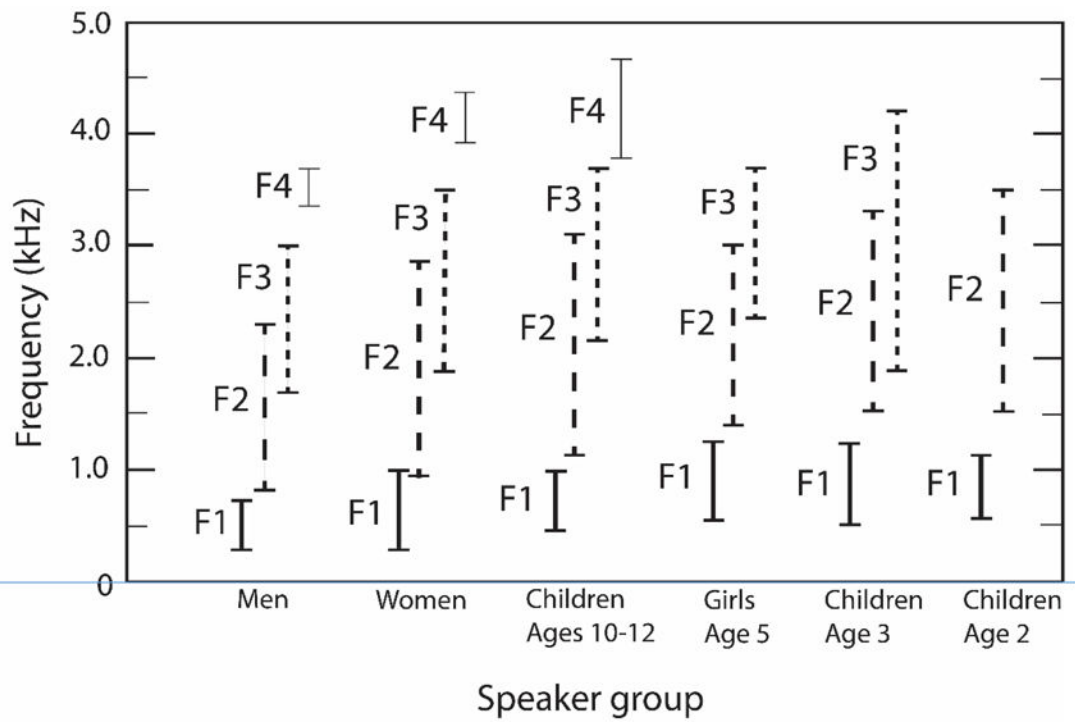
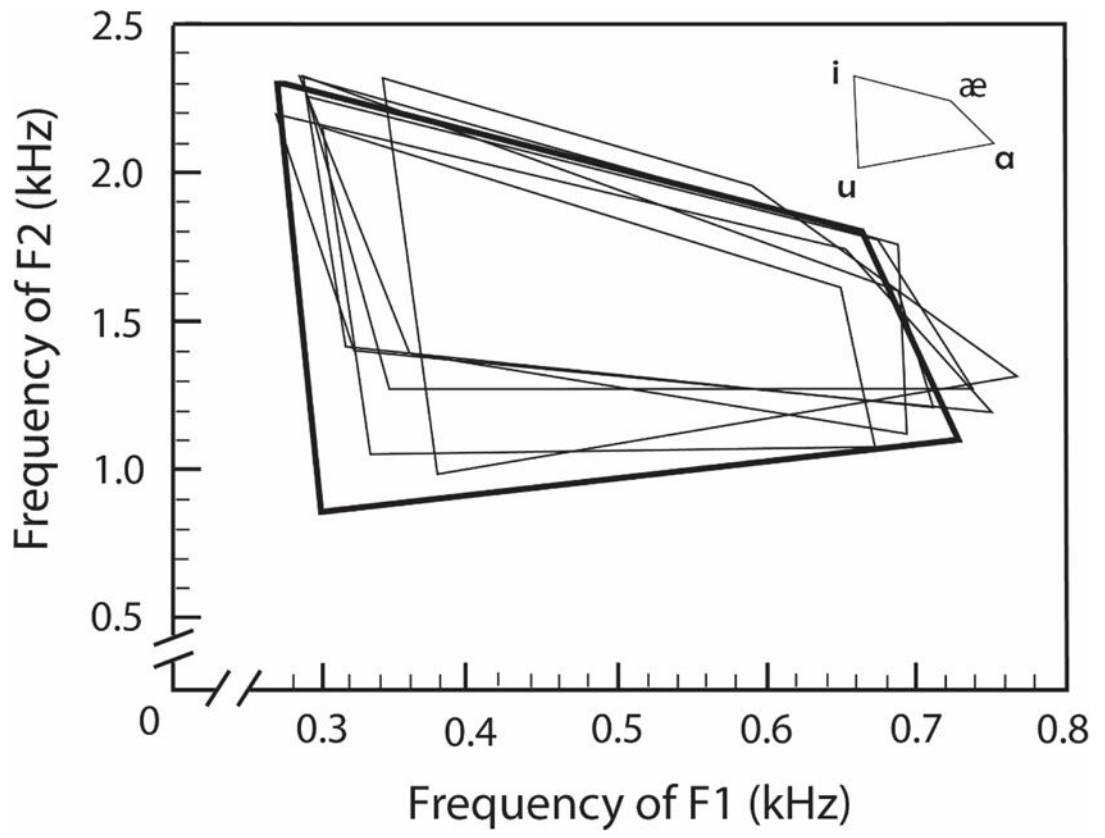


Figure 2. Frequency ranges for F1, F2, F3, and F4 (as available) for men, women, and children. Data sources are as follows: men, women, and children ages 10-12 (Hillenbrand et al., 1995); girls age 5 (S. Lee et al., 1997); Children age 3 and Children age 2 (McGowan et al., 2014).



Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

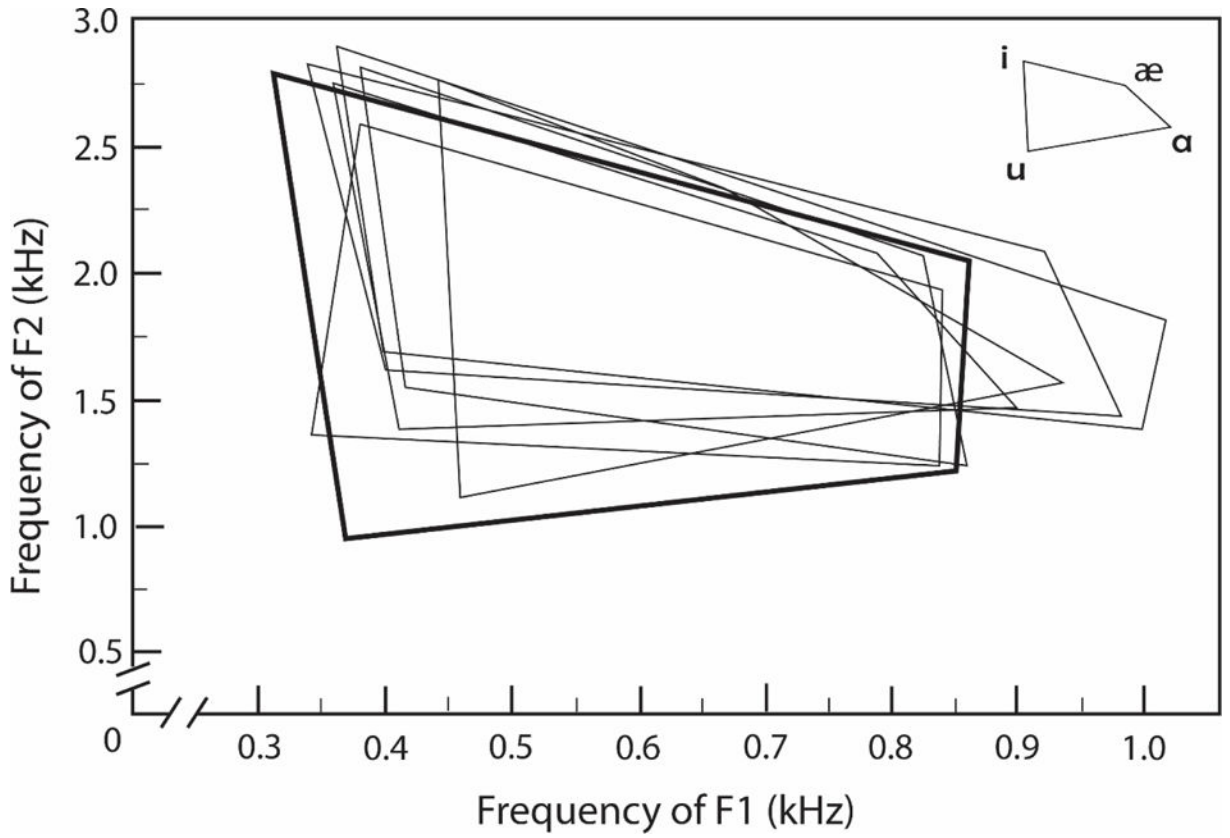
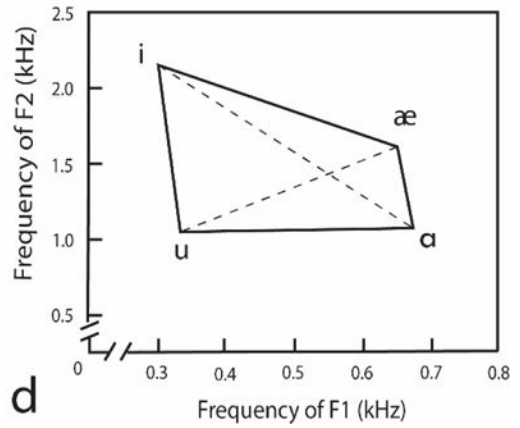
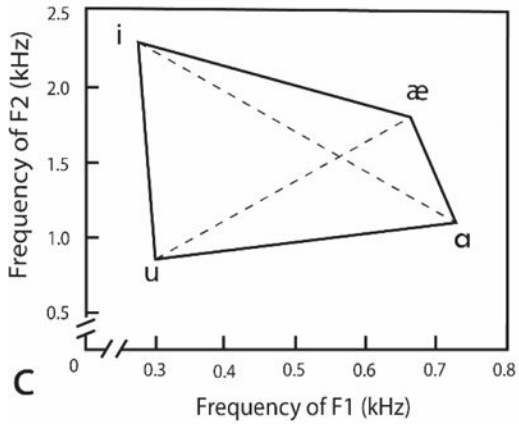
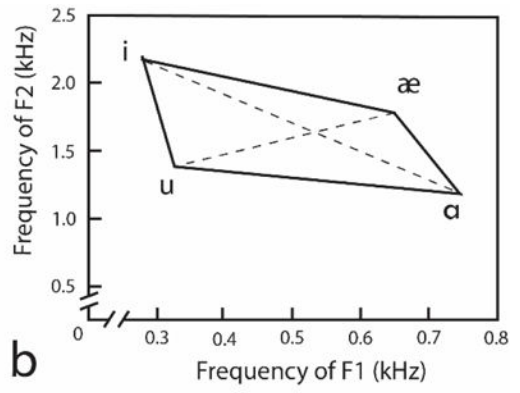
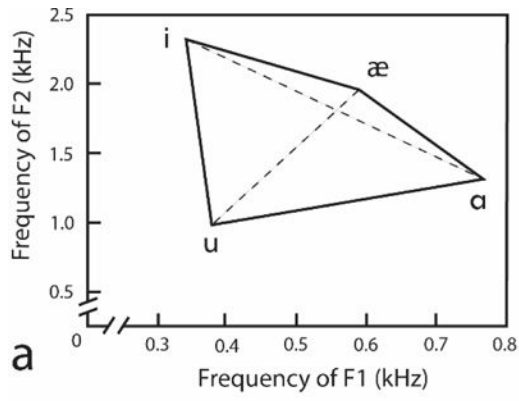


Figure 3.

- a. F1-F2 plots for men showing values for the point vowels reported in the 7 studies listed in Table 1. The thick line is the vowel quadrilateral from Peterson and Barney (1952).
- b. F1-F2 plots for women showing values for the point vowels reported in the 7 studies listed in Table 1. The thick line is the vowel quadrilateral from Peterson and Barney (1952).



Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

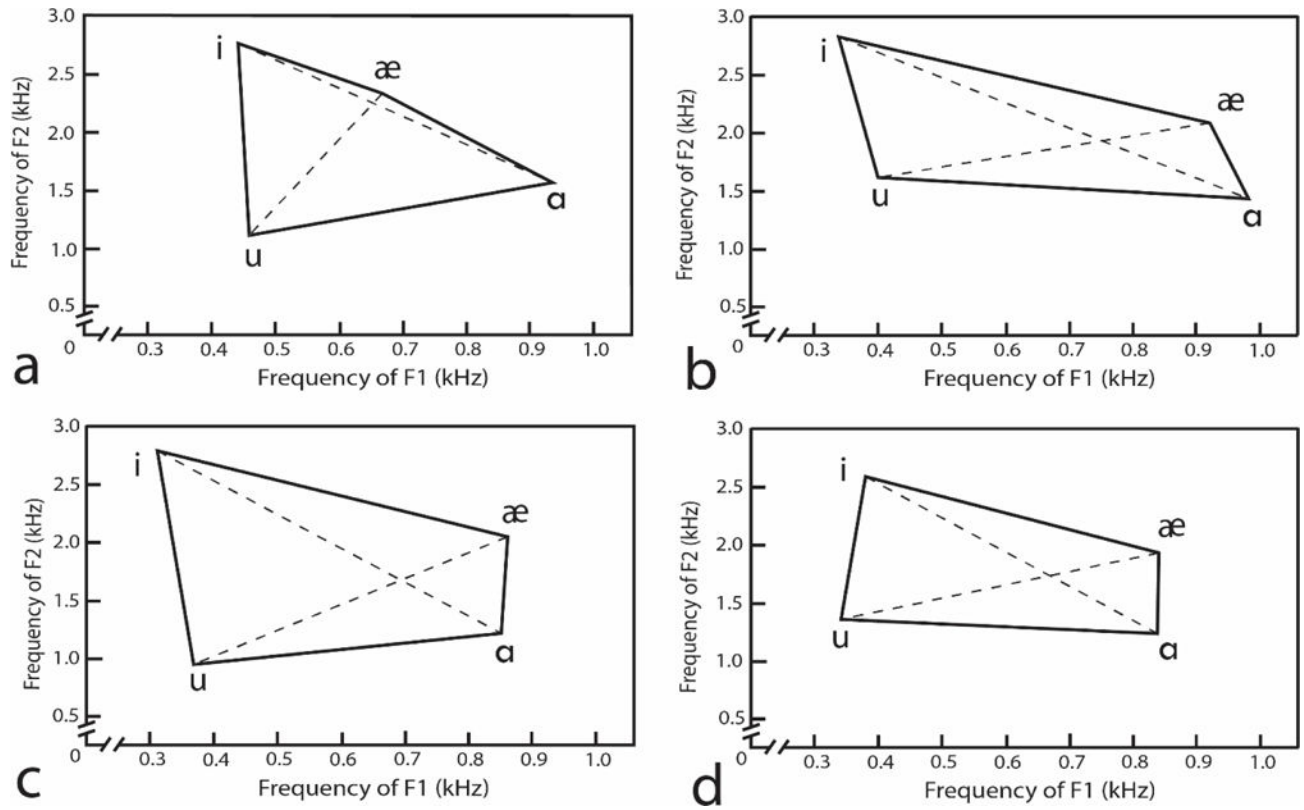


Figure 4.
 a. Vowel quadrilaterals for men derived from the data in 4 studies: (a) Hillenbrand et al., (1995), (b) Zahorian and Jagharghi (1993) (c) Peterson and Barney (1952), and (d) Childers and Wu (1991). Intersecting diagonals have been drawn from the corner points.
 b. Vowel quadrilaterals for women derived from the data in 4 studies: (a) Hillenbrand et al., (1995), (b) Zahorian and Jagharghi (1993) (c) Peterson and Barney (1952), and (d) Childers and Wu (1991). Intersecting diagonals have been drawn from the corner points.

Table 1

Shown for selected age-sex groups in childhood are values of acoustic measures for vowel production: range of F1 frequency (Hz), range of F2 frequency (Hz), range of F3 frequency (Hz); vowel space area (kHz²), and mean f0. The values are approximate and are intended as a heuristic for formant estimation. The data sources are: F1, F2 and F3 frequency ranges (S.Lee et al., 1999; Vorperian & Kent, 2007); vowel space area

Age-sex Group	F1 range (Hz)	F2 range (Hz)	F3 range (Hz)	Vowel space area (kHz ²)	Mean f0 (Hz)
2-year-old children, both sexes	400-1000	1500-3500	No data		300
4-year-old children, both sexes	480-1200	1700-3600	3400-4300	573 ^a	280
6-year-old children, both sexes	350-1200	1500-3200	3300-3700	639 ^a 633 ^b	260
12-year-old boys	325-900	1300-2700	2700-3300	374 ^a 406 ^b	230
12-year-old girls	375-950	1600-3000	3000-3500	349 ^a 408 ^b	230
14-year-old boys	325-850	1300-2700	2700-3300	346 ^a 344 ^b	180
14-year-old girls	400-900	1600-2700	2800-3300	298 ^a 295 ^b	230
16-year-old boys	300-750	1250-2300	2300-3000	240 ^a 264 ^b	130
16-year-old girls	400-950	1800-2800	2800-3300	313 ^a 334 ^b	230
18-year-old boys	270-750	900-2300	2300-3000	318 ^a 330 ^b	130
18-year-old girls	350-900	1300-2800	2700-3400	335 ^a 352 ^b	230

^aVorperian & Kent, 2007

^bFlipsen & Lee, 2013), mean f0 (Kent, 1994; Hunter, 2009; Lee et al., 1999; Robb & Saxman, 1985; Vorperian & Kent, 2007)

Table 2

Description of 7 data sets for vowel formant-frequencies of adult speakers of American English.

Source	Speakers	Speech material
Childers & Wu (1991)	27 men, 25 women (adults 20 to 80 years)	Sustained vowels
Hagiwara (1997)	6 men, 9 women (adults 18 to 26 years)	CVC syllables with 3 consonantal environments
Hillenbrand et al. (1995)	45 men, 48 women, 46 children (ages 10 to 12 years)	/hVd/ syllables
S. Lee et al. (1999)	29 men, 27 women, 436 children (ages 5 to 17 years)	CVC syllables and 5 sentences
Peterson & Barney (1952)	33 men, 28 women, 15 children	/hVd/ syllables
Yang (1996)	10 men, 10 women (adults 18 to 27 years)	/hVd/ words
Zahorian & Jagharghi, (1993)	10 men, 10 women, 10 children (ages 7 to 11 years)	CVC syllables with various consonants

Table 3

Summary of studies of the effects of aging on fundamental frequency and first two formant frequencies of vowels.

Source Article	Change in f0	Change in F1	Change in F2	Centralization or reduction?
Debruyne & Decoster (1999)	Decreased in women	Decreased in men and women for one vowel	Decreased in men and women for one vowel	–
Eichhorn et al. (in press)	Decreased in women	Vowel-specific changes for men and women	Vowel-specific changes for men and women	No
Endres, Bambach, & Flosser (1971)	Decreased in men and women	Decreased in men and women	Decreased in men and women	–
Fisher & Linville (1985)	–	Decreased in women for one vowel studied	–	–
Fletcher et al. (2015)	–	No change for men or women in 65- to 90-year-old age range	No change for men or women in 65- to 90-year-old age range	No
Harrington, Palethorpe, & Watson (2007)	Decreased in both men and women	Decreased in both men and women	–	–
Linville & Ross	–	Decreased in both men and women	Decreased in women; tended to decrease in men	–
Rastatter & Jaques (1990)	–	Varied with vowel for both men and women	Varied with vowel for both men and women	Yes
Scukanec & Petrosino (1991)	–	Decreased for 4 vowels studied in women	Decreased for back vowels in women	Yes
Sebastian, Babu, Oommen, & Ballraj (2012)	No change for men or women in 60- to 80-year-old age range	No change for men or women in 60- to 80-year-old age range	No change for men or women in 60- to 80-year-old age range	–
Torre & Barlowx (2009)	Increased in men, decreased in women	–	–	No
Watson & Munson (2007)	–	Decreased in men and women	Decreased in men and women	Yes, for one dimension
Xue, Jiang, Lin, Glassenberg, & Mueller (1998)	–	Decreased in men and women	Varied with vowel	–
Xue & Hao (2003)	–	Decreased in men and women	Decreased depending on vowel in men and women	–

Table 4

Published studies of vowel formant frequencies for children of different ages (in years). The values in the cells are the numbers of participants studied at each age or age group. All studies are of children presumably learning American English as their first language. A – Robb & Cacace (1995); B – Kuhl & Meltzoff (1996); C – Kent & Murray (1982); D – Robb, Chen, & Gilbert (1997); E – Bickley (1989); F – Gilbert, Robb, & Chen (1997); G – McGowan, Denny, & Nittrouer (2014); H – Chung et al. (2012); I – Yildirim, Narayanan, Byrd, & Khurana (2003); J – Eguchi & Hirsh (1969); K – Perry, Ohde, & Ashmead (2001); L – Kohn & Farrington (in press); M – Assmann, Nedarey, & Bharadwaj (2008); N – S. Lee, Potamianos, & Narayanan (1999); O – Bennett (1981); P – Jacewicz, Fox, & Salmons (2011); Q – Hillenbrand et al. (1999); R – Angelocci, Kopp, & Holbrook (1964). 83

Age/ Study	<1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	20																		
B	72																		
C	21																		
D	20	-	-																
E	40	-	-																
F		4	-	-															
G		6	6	6	6														
H			10			10													
I				12	-	-	-												
J		5	5	5	5	5	5	5	5	5	5	10	10	10	10	10	10	10	10
K					20				20				20			20			
L					10						20			20		20			
M						10	10	10	10	10	10	10	10	10	10	10	10	10	10
N						32	27	35	36	48	39	43	43	39	21	22	22	19	20
O								42	-										
P									94	-									
Q											46								
R												18							

Table 5

Formant bandwidth reported in several studies. Shown are the language and speaker description, number of vowels studied, and formant bandwidths (B1, B2, and B3). Range and standard deviations (s.d.) are included for studies reporting these values.

Source	Language and Speakers	Number of Vowels	B1	B2	B3
Benavides et al. (2015)	Spanish, adults	5	84 (range of 5 to 588)	117 (range of 12 to 1485)	183 (range of 22 to 1650)
Bogert (1953)	English, male adults	10	130	100	185
Childers & Wu (1993)	English, adults	10	140	149	223
Dunn (1961)	English, adults	10	50	64	115
Fant (1962)	Swedish, adults	10	48	50	98
Hana, Smith, & Wolfe (2016)	Australian English, male adults	1	85 s.d. = 30	90 s.d. = 25	85 s.d. = 20
Hana et al. (2016)	Australian English, female adults	1	75 s.d. = 30	90 s.d. = 25	90 s.d. = 20
House & Stevens (1958)	English, adults	9	52	62	67
Krishna & Rajashekhar (2013)	Teluga, adults	10	57	136	230
Penny, Russell, & Pemberton (1994)	Australian English, female adults recorded in 1945 and 1992	11	1945 recording: 227 s.d. = 45 1992 recording: 234 s.d. = 52	1945 recording: 407 s.d. = 80 1992 recording: 238 s.d. = 98	NA
Robb, Chen, & Gilbert (1997)	English, children (ages 4, 8, 15, 18, 25 months)	Not specified	4 mo. – 350 8 mo. – 260 15 mo. – 247 18 mo. – 191 25 mo. – 170	4 mo. – 671 8 mo. – 490 15 mo. – 416 18 mo. – 345 25 mo. – 360	NA
Robb, Yates, & Morgan (1997)	English, adults	3	79	99	NA
Whiteside & Hodgson (1999)	English, 6- to 10-year-old children	1	193	190	263

Table 6

Measures of area, dispersion, or distance within the F1-F2 plane.

Measure or plot	Formula or method	References
Point cloud	Plot of all F1-F2 points for a sample of vowels	Hillenbrand et al. (1995); Peterson & Barney (1952); Story & Bunton (2017)
Quadrilateral Vowel Space Area (qVSA or VSA-4)	$qVSA = 0.5 * (F1/i/(F2/a/-F2/u) + F1/a/(F2/u/-F2/i) + F1/u/(F2/i/-F2/a/) $	Higgins & Hodge (2002)
Triangular Vowel Space Area (tVSA or VSA-3)	$tVSA = 0.5 * [(F2/u/+F2/i/) * (F1/u/-F1/i/) - (F2/a/ +F2/u/) * (F1/u/ -F1/a/) - (F2/a/+F2/i/) * (F1/a/-F1/i/)] $ $tVSA = [s(s-a) (s-b) (s-c)]^{1/2}$ Where $s = 1/2 (a + b + c)$ and a, b, c are the lengths of the sides of the triangle. (Heron's formula based on the perimeter s and the sides of the triangle)	Bradlow, Torretta, & Pisoni (1996) Fox & Jacewicz (2008)
Vowel Articulation Index (VAI)	$VAI = (F2/i/+F1/a/) / (F1/i/+F1/u/+F2/u/+F2/a/)$.	Roy et al., 2009)
Formant centralization ratio (FCR)	$FCR = (F2/u/ + F2/a/ + F1/i/ + F1/u/) / (F2/i/ + F1/a/)$ Note: FCR is the reciprocal of VAI	Sapir, Ramig, Spielman, & Fox (2010)
Vowel Formant Dispersion (VFD)	VFD = the Euclidean length of vectors for vowels originating from a defined center point of the vowel space fixed to a basic three-vowel frame.	Karlsson & van Doorn (2012)
Formant Spacing: compact-diffuse (C-D) and grave-acute (G-A). C-D are an index of tongue elevation and G-A is an index of tongue advancement.	$C-D = F2 - F1$ $G-A = (F1 + F2)/2$.	Blomgren, Robb, & Chen (1998)
Area of convex hull encompassing all vowel tokens	Area contained within convex hull	de Boer (2009), McCloy, Wright, & Souza (2014), Story & Bunton (2017)
Mean vowel cluster size	Mean of the areas of the 95% confidence ellipses for each vowel category	McCloy et al (2014)
Mean Euclidean distance from center	The mean of the distances between vowel tokens and a point in the center of a speaker's vowel space	Bradlow et al. (1996)
Repulsive force (or total energy)	Sum of inverse squared distances between all pairs of vowel tokens not belonging to the same vowel phoneme.	Liljencrants & Lindblom (1972), McCloy et al (2014)
F1 or F2 frequency range	Maximum range of formant frequency across vowel tokens or vowel category means	Bradlow et al. (1996), Hazan & Markham (2004)
F2/i//ssF2/u/ ratio	The ratio between the F2 frequencies for /i/ and /u/	Moura et al. (2008), Sapir et al. (2007)