

## Review Article

# What Acoustic Studies Tell Us About Vowels in Developing and Disordered Speech

Ray D. Kent<sup>a</sup> and Carrie Rountrey<sup>b</sup>

**Purpose:** Literature was reviewed on the development of vowels in children's speech and on vowel disorders in children and adults, with an emphasis on studies using acoustic methods.

**Method:** Searches were conducted with PubMed/MEDLINE, Google Scholar, CINAHL, HighWire Press, and legacy sources in retrieved articles. The primary search items included, but were not limited to, vowels, vowel development, vowel disorders, vowel formants, vowel therapy, vowel inherent spectral change, speech rhythm, and prosody.

**Results/Discussion:** The main conclusions reached in this review are that vowels are (a) important to speech intelligibility; (b) intrinsically dynamic; (c) refined in both

perceptual and productive aspects beyond the age typically given for their phonetic mastery; (d) produced to compensate for articulatory and auditory perturbations; (e) influenced by language and dialect even in early childhood; (f) affected by a variety of speech, language, and hearing disorders in children and adults; (g) inadequately assessed by standardized articulation tests; and (h) characterized by at least three factors—articulatory configuration, extrinsic and intrinsic regulation of duration, and role in speech rhythm and prosody. Also discussed are stages in typical vowel ontogeny, acoustic characterization of rhotic vowels, a sensory-motor perspective on vowel production, and implications for clinical assessment of vowels.

Vowels make up about half of the acoustic stream of speech, but these sounds are eclipsed by consonants in the literature on speech development and disorders. Davis and MacNeilage (1990) commented that, "Vowels are the poor relations of child phonology. There is perhaps less than one study of vowels for every 20 studies of consonants; and studies ostensibly dealing with a child's complete vocal repertoire usually pay little attention to vowels" (p. 16). Since the time of that comment, the situation has changed, albeit slowly. Speake et al. (2012) wrote that, "Compared to the treatment of consonant segments, the treatment of vowels is infrequently described in the literature on children's speech difficulties" (p. 277). A landmark in this literature is a book devoted to vowels and vowel disorders (Ball & Gibbon, 2013), which addressed a variety of issues pertaining to vowel development and vowel disorders,

but research reports continue to be few in number compared to those on consonants. The main theme of this review is that research in several disciplines points to a number of reasons why vowels are important in understanding how speech develops in childhood and how speech is disrupted in various communication disorders. Acoustic analysis and its advancement through technology have had a major role in these discoveries, and a particular objective of this review article is to show how acoustic studies have enlarged and refined the understanding of vowels. This may lead to improvements in the clinical assessment of vowel disorders and ultimately to improved treatments. Implications for clinical assessment are discussed in a concluding section.

"A vowel is a speech sound that is formed without a significant constriction of the oral and pharyngeal cavities and that serves as a syllable nucleus" (Shriberg et al., 2019, p. 31). This definition hinges on a basic duality of physiology and phonology; that is, vowels have an articulatory basis and a distinct role in phonology. Vowels appear in the neonatal stage of development and are used throughout the normal life span, which makes them one of the earliest to appear and most enduring of human behaviors. They have a central role in phonology given that all syllables—except syllabic consonants—contain a vowel.

<sup>a</sup>Waisman Center, University of Wisconsin–Madison

<sup>b</sup>Department of Communication Sciences and Disorders, University of Cincinnati, OH

Correspondence to Ray D. Kent: Kent@waisman.wisc.edu

Editor-in-Chief: Julie Barkmeier-Kraemer

Editor: Lisa Goffman

Received May 25, 2019

Revision received November 4, 2019

Accepted April 19, 2020

[https://doi.org/10.1044/2020\\_AJSLP-19-00178](https://doi.org/10.1044/2020_AJSLP-19-00178)

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

Much of what we know about vowels in developing and disordered speech is derived from perceptual analysis, especially phonetic transcription and articulation tests (Howard & Heselwood, 2013; Stoel-Gammon & Pollock, 2008). However, questions have been raised about the validity and reliability of perceptual methods in the study of vowels (Cox, 2008; Howard & Heselwood, 2013), the adequacy of commonly used tests of articulation to assess vowels (Eisenberg & Hitchcock, 2010; Pollock, 1991), and the nature of typical vowel development (James et al., 2001). Questions relating to perceptual methods have been raised in language science generally. For example, Sloos et al. (2019) commented that, "...speech sound perception is shaped by what is actually in the speech signal as well as by expectations on the local phonological and global sociolinguistic, and geolinguistic level" (p. 2). These authors go on to distinguish reliability (transcribers' agreement) and validity (the relation between the acoustic signal and the transcription). The clinical literature has emphasized the former, but the latter is equally important. These concerns are good reasons to take a closer look at how vowels develop in children and at the nature of vowel disorders in children and adults. This "look" can be accomplished partly through the lens of acoustic analysis, which is the major substance of this review.

Acoustic studies reviewed here contribute to the contemporary understanding of vowel production and perception in typical and atypical speech. The acoustic properties of vowels relate to recent work on speech perception, auditory neurophysiology, speech articulation, and modeling of speech production. Taken together, these lines of research lead to an integrated sensory-motor conceptualization of vowel sounds and eventually to a framework for clinical intervention targeting this class of sounds. The authors also recognize diversity related to native language and acknowledge these differences in this review, though the perspective of this work is rooted in American English.

### **Research Questions**

The main content of this review article is a review of what acoustic studies reveal about several aspects of vowel development and disorders, leading to a general discussion that bears on implications of vowel assessment. The principle questions addressed in this review are below. Answers to these questions cohere in an improved understanding of both the development of vowels in children and the nature of vowel errors in disordered speech.

Regarding the nature and development of vowels, the following questions are posited: What is the contribution of vowels to speech intelligibility? Are vowels in American English inherently intrinsically dynamic? How can acoustic data help to create a picture of vowel development in children? What are distinctive acoustic properties of rhotic vowels and diphthongs?

Regarding production and perception of vowels, the following questions are posited: What is the interaction between vowel perception and vowel production in early

speech development? Can acoustic measures serve as an index of the precision of vowel production? What do acoustic data add to the knowledge of phonetic mastery of vowels? How is vowel production compensated for in articulatory and auditory perturbations? How and when do language and dialect differences influence vowel production in infant vocalizations and early childhood?

Regarding the functional and clinical application of our knowledge of vowel development and disorders, the following questions are posited: How do speakers adjust their vowel space area (VSA) in accord with listener characteristics and communication settings? In what ways is vowel perception and production vulnerable to speech and language disorders in children and adults? How can sensory-motor mapping serve in a theoretical framework for treating vowel perception and production? How is the clinical assessment of vowels evolving?

This review is intended for two general audiences: clinicians who assess and treat speech sound disorders and researchers who seek to understand the development of vowel sounds and the nature of vowel disorders in various clinical populations. Our goal is not to prescribe clinical methods but rather to describe the implications of research on the eventual refinement of those methods. The basic content of this review article pertains to the lessons learned from acoustic analysis of vowels in developing and disordered speech. A culminating section addresses the clinical assessment of vowels and the need for further study regarding clinical applications of the evidence. In the same vein, an appeal is made for research on the treatment of vowel disorders.

### **Method**

Searches were conducted with PubMed/MEDLINE, Google Scholar, CINAHL, HighWire Press, and legacy sources in retrieved articles. The primary search terms were vowel, vowels, and vocalization (with the following qualifiers or associations: assessment, development, dialect, disorders, duration, formants, intelligibility, phonetic mastery, perception, production, spectral inherent change, spectrum, therapy, treatment). Results of the literature search were organized with respect to the questions listed in the previous section. Retrieved articles were categorized by search terms and by their relationship to the aforementioned questions.

As mentioned earlier, the review was focused on American English and was not intended to be comprehensive of vowels in other languages. However, selected aspects of vowels in different languages are noted in connection with potentially universal principles or tendencies.

### **Review of Methods of Measurement**

This section reviews basic aspects of acoustic theory and analysis that underlie the methods used in the studies under review. Particular attention is given to the estimation of formant frequencies associated with vowel production. Readers who want an introduction or review may find this

section helpful, as it lays out fundamental concepts needed to appreciate acoustic analysis of speech.

### ***Acoustic Analysis Is a Tool to Study Vocal Tract Anatomy, Vowel Articulation, and Vowel Perception***

Vowels are produced with two essential processes: generation of acoustic energy (phonation in the case of ordinary voiced speech) and articulation (vocal tract shaping) to produce distinctive patterns of resonance classically represented by formants. The two processes are largely, but not completely, independent, so that to a first approximation, vowel articulation is unaffected by phonation, and vice versa. This quasi-independence is critical to the dual role of vowels in conveying segmental information (vowel identification) related largely to formant pattern along with prosodic and paralinguistic information signaled by changes in vocal fundamental frequency (F0) and other acoustic modifications. The classic source-filter theory of speech production characterized vowel production in terms of the voicing source and the filter effects of the vocal tract (Fant, 1970).

### ***Methods of Acoustic Analysis***

Different approaches can be taken to analyze the acoustic properties of vowels, but estimation of formant frequencies is the most commonly used and has a long history in speech research (Kent & Vorperian, 2018; Vilain et al., 2015). Conventional notation is to identify individual formants as  $F_n$  where  $n$  is the formant number (e.g., F1 is the first formant, which can be specified in terms of its frequency and bandwidth). A major attraction accruing to formants in studies of speech production is that, at least to a first approximation, individual formants can be associated with articulatory features, as shown in Figure 1. F1 frequency is correlated with vowel height (i.e., tongue position in the superior-inferior axis), such that the higher the vowel, the lower the F1 frequency. F1 frequency also is correlated with vowel duration in many languages, such that vowels with high F1 frequency are longer than vowels with a low F1 frequency. This relationship may be based on physiological factors (especially jaw movement) that have been “phonologized” (i.e., voluntary and extrinsic) in languages such as English but do not necessarily establish a universal pattern (Solé & Ohala, 2010). F2 frequency correlates with the articulatory dimension of tongue advancement or backness (i.e., tongue position in the anterior-posterior axis). Alternatively, the F2-F1 difference correlates with tongue advancement, such that back vowels have a smaller difference than front vowels. Both F1 and F2 (and all formants for that matter) are affected by lip rounding or lip protrusion. Rounding and protrusion have the same acoustic consequence of reducing all formant frequencies. In English, only back vowels are rounded so that rounding and backness tend to co-occur, leading to a low-frequency dominance of acoustic energy. Rounding can be

considered as the third dimension in a three-dimensional phonetic space for vowels.

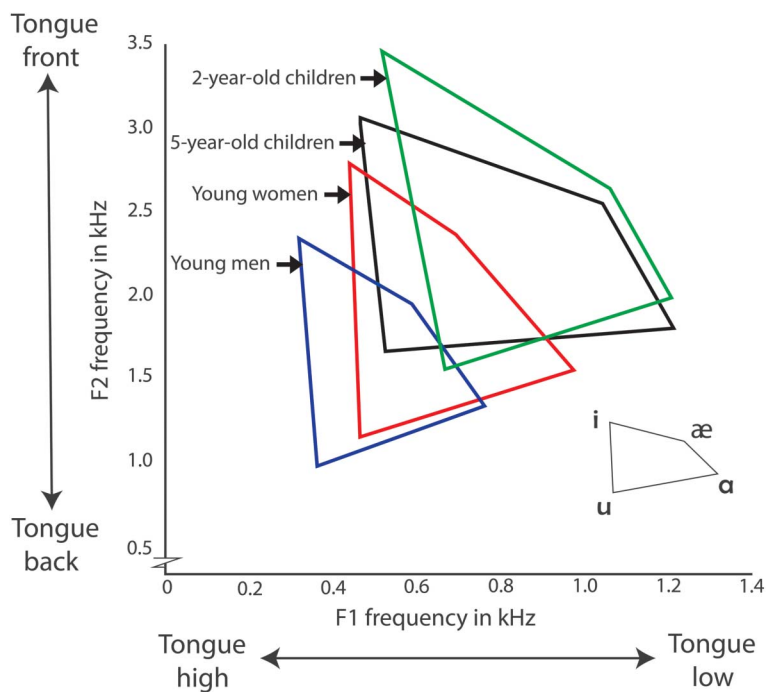
The relationship between vowel articulation and the sensory experience of speech acoustics is further illustrated in Figure 2, which pertains to the three corner vowels /i/, /u/, and /a/, which are remarkably common in the world’s languages (Maddieson, 1984). This illustration shows for each vowel its vocal tract configuration, a simplified vocal tract model consisting of front and back cavities and a stylized representation of the formant pattern (F1, F2, and F3). Figure 2 pertains to vowel production by a 4-year-old child; for example, the formant frequencies are suitable for a child’s vowels and therefore will differ from the majority of formant frequency values reported in the literature, which is dominated by data on adults—especially male adults. Because this review emphasizes vowel development and vowel disorders in children, data for children are included in the discussion of anatomic and acoustic features related to vowel production.

Theoretically, there is an infinite number of formants, but only the first few are needed to identify and discriminate the vowels of a language. The higher formants, such as F3 and F4, are often neglected in discussions of acoustic-anatomic-articulatory relationships, but these formants are important in several respects, including describing rhotic sounds (Hagiwara, 1995); normalizing both rhotic and non-rhotic vowels (Disner, 1980; Hillenbrand & Gayvert, 1993); explaining the speaker’s formant, a local energy maximum in the vicinity of F4 (Bele, 2006; Leino et al., 2011); describing resonances of the hypopharynx (Takemoto et al., 2008); identifying acoustic correlates of maxillary arch dimensions (Hamdan et al., 2018); and describing acoustic consequences of procedures such as tonsillectomy (Švancara et al., 2006) and supracricoid laryngectomy (Buzaneli et al., 2018). Unfortunately, data on the higher formants are not as abundant as those for F1 and F2, so the potential value of a more complete formant description is not established. To be sure, estimation of these higher formant frequencies can be difficult because of their relatively low energy, but improvements in methods of acoustic analysis enhance the likelihood of obtaining data throughout the life span (Kent & Vorperian, 2018).

### ***Vowel Acoustics Related to Speaker Age and Sex***

In general, vowel formant frequencies across all vowels decrease as vocal tract length (VTL) increases. Therefore, the overall developmental pattern from birth to adulthood is one of decreasing formant frequencies, with larger changes in males than females (as can be seen in Figure 1). However, the literature on this topic does not give an entirely coherent view of the acoustic changes. Some studies have reported that vowel formant frequencies change little, if at all, during the first 2–4 years of life (Buhr, 1980; Gilbert et al., 1997; Kent & Murray, 1982; McGowan et al., 2014), although increased range or dispersion of formant frequencies has been observed (Gilbert et al., 1997; Kent & Murray, 1982; Robb et al., 1997). Other studies indicate rapid expansions of the

**Figure 1.** F1–F2 graph of the vowel quadrilateral, showing representative data for adults and children of different ages. F1 frequency varies to a first approximation with tongue height, and F2 frequency varies to a first approximation with tongue advancement. There is almost no overlap of the formant frequencies for 2-year-olds and male adults.



vowel acoustic space during the first 2 years of age (Bond et al., 1982; Ishizuka et al., 2007; Yamashita et al., 2013). Increases in formant frequencies are expected in infancy based on increases in VTL and increased sizes of individual articulators during the first 2 years of life (Vorperian et al., 2005). Because the acoustic data reported to date are based on small numbers of infants and different methods of data collection and analysis, it is difficult to determine with confidence the relationship between acoustic and anatomic changes.

Figure 3 shows the frequencies of the first four formants of vowel /æ/ spoken by a 2-year-old child, a 4-year-old child, a young woman, and a young man. The formants span a considerable range of frequencies, which is important for clinical reasons, such as assessing the effects of hearing loss on the perception of self- and other-produced vowels. For example, frequencies of just over 5 kHz are needed to encompass the first four formants of a 2-year-old child, which implies that a child with a hearing loss affecting frequencies above 4 kHz will have difficulty hearing the higher formants of their vocalizations, as well as the high-frequency energy of fricatives such as /s/.

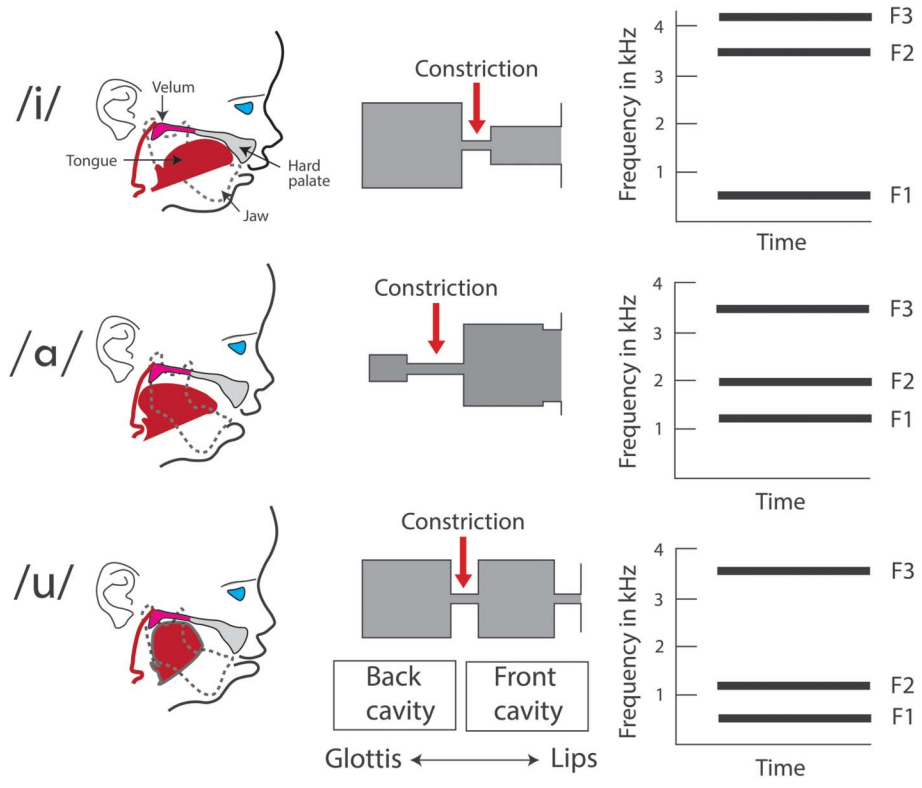
Another unresolved topic in anatomic–acoustic correlations in speech is the emergence of sexual dimorphism, that is, when differences appear between boys and girls. Gender differences in vowel formant frequencies are present by 4 years of age, with boys having lower formant frequencies than girls (Perry et al., 2001; Vorperian & Kent, 2007). However, gender differences in VTL have not been observed until the age of puberty (Fitch & Giedd, 1999;

Markova et al., 2016). Therefore, the formant frequency differences are not explained by increased VTL. Boys and girls may differ in other anatomic and articulatory features, which are a topic of continuing research. Gender differences in speech production may also result from learned behavioral patterns. Evidence for a learning or sociocultural hypothesis of gender differences has been reported in several studies (Cartei et al., 2014, 2012; Cartei & Reby, 2013). The basic idea is that children spontaneously attempt to sound like adults of their own gender. In learning speech, children aspire not only to be understood but also to be identified as to their own gender. Aspects of vowel production, including fundamental and formant frequencies, appear to be important in accounting for gender-related speech patterns (Munson et al., 2015).

VSA, a measure of the area contained within the vowel quadrilateral or vowel triangle, is one of the most frequently used acoustic indices of vowel production in studies of both developing and disordered speech. VSA generally decreases with speaker age, as illustrated in Figure 4 (also see Vorperian & Kent, 2007). This reduction is a consequence of changing VTL, but VSA can be affected by other factors, including communication setting, speech sample, speaking rate and prosody, and speech disorders (as discussed in a later section). The primary point to be made here is that VSA can be construed as the articulatory working space for vowels, usually determined by the quadrilateral (or triangle) formed by the point vowels. Figure 4 shows that VSA values vary widely across studies,

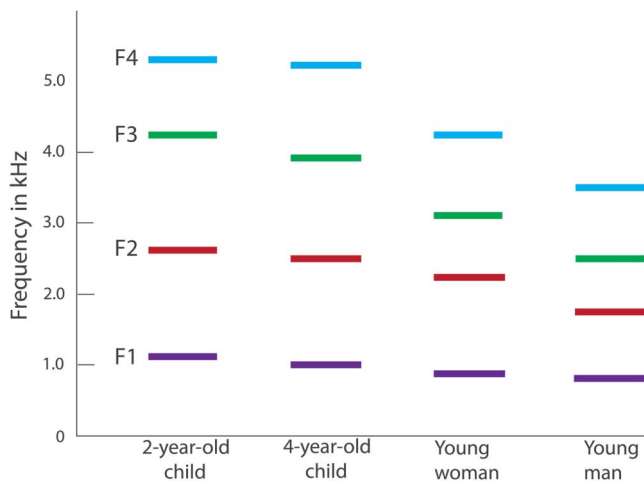


**Figure 2.** Drawings to show for each of the three corner vowels /i/, /u/, and /a/ its vocal tract configuration (left), a simplified vocal tract model consisting of front and back cavities (center), and a stylized representation of the formant pattern (F1, F2, and F3; right). The vowel productions pertain to a 4-year-old child.



and this variation hinders the application of normative data to clinical assessments. Note, for example, the large ranges of VSA values at ages 4 years and adults aged 20 years or more.

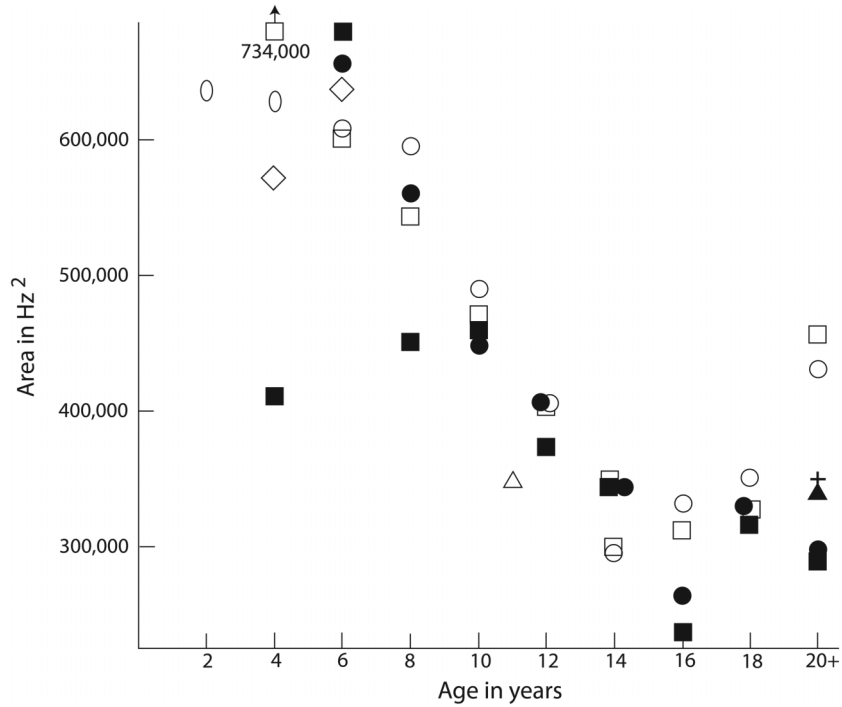
**Figure 3.** Frequencies of the first four formants of vowel /ae/ for four speaker groups.



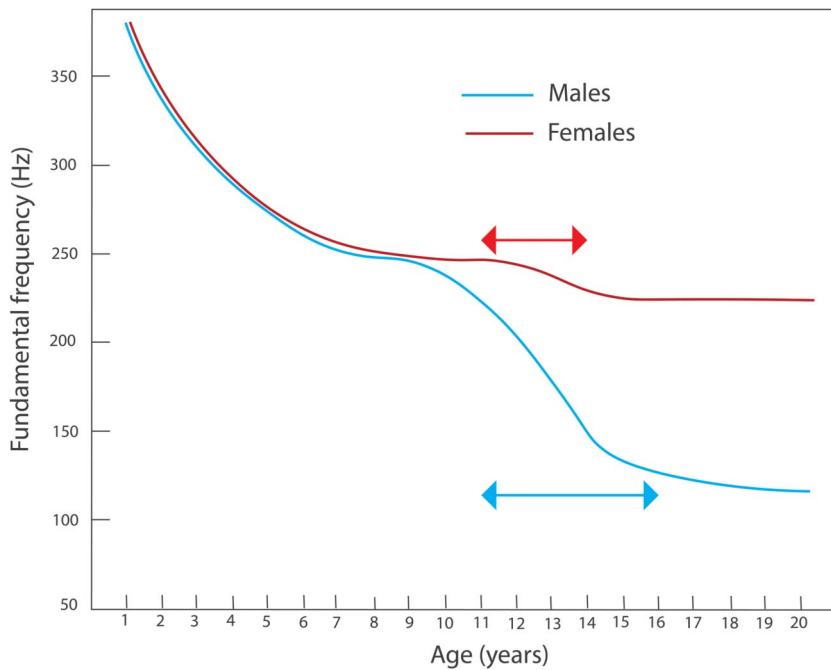
Given that formant frequencies vary with speaker age and gender, studies have used either of two strategies in analyzing and reporting formant data. The first strategy is to report the actual frequency data in hertz, which unavoidably results in aggregates of data corresponding to age–gender characteristics of the speakers. The second is to normalize the formant frequencies in an attempt to render the data from different speakers directly comparable for purposes such as phonetic identification. Both approaches have advantages and disadvantages, depending on the purpose of the study. For a discussion of different approaches to vowel normalization, see Adank et al. (2004). The purposes of this review article are served by using nonnormalized data for formant frequencies, but it is acknowledged that normalization holds considerable value in the ultimate understanding of vowel perception.

With respect to the energy source for vowels, vocal F0 has a generally falling pattern from birth to adulthood (Kent, 1976), but the decrease is not monotonic and appears to depend on the nature of the vocalization. Rothgänger (2003) reported that, during the first year of life, the mean F0 of crying increased (from about 440 to 500 Hz) and the mean F0 of babbling (comfort state vocalizations) decreased (from about 390 to 337 Hz). The data also revealed that the melody (prosody) of babbling bore similarities to the ambient language within the first year of age. Figure 5

**Figure 4.** Vowel space area from several studies of children and adults speaking American English. Data sources are as follows: cross = mean for two males in Bunton and Leddy (2011); filled circles = males in Flipsen and Lee (2012); unfilled circles = females in Flipsen and Lee (2012); diamond = Hustad et al. (2010); ovals = McGowan et al. (2014); filled triangle = G. S. Turner et al. (1995), estimated from graph for males; filled squares = males in Vorperian and Kent (2007); unfilled squares = females in Vorperian and Kent (2007); unfilled triangle = Zajac et al. (2006).



**Figure 5.** Mean vocal fundamental frequency (F0) from ages 1 to 20 years for males and females (based on data in Ludlow et al., 2019). The lines with double arrowheads show the approximate transition periods of F0 for both males and females.



shows average values of F0 over the age interval of 1–20 years for both males and females. Gender differences in F0 occur after about 10 years of age. During puberty, F0 drops sharply for males and more gradually for females (Berger et al., 2019; Maturo et al., 2012). The lines with double arrowheads show the approximate transition periods of F0 for both males and females. The developmental pattern for F0 is important for understanding changes in the pitch of the voice and also to account for age-dependent differences in the accuracy of formant measurements (Kent & Vorperian, 2018).

### *Advantages of Acoustic Analysis*

According to Ciocca and Whitehill (2013), acoustic analysis of speech is relatively inexpensive and accessible, compared to auditory-perceptual analysis and articulatory analysis. Acoustic analysis is also considered more objective than some other methods (e.g., perceptual rating) and serves as a quantifiable “bridge” between articulatory and perceptual analysis, taking into account both kinematics of the source (speaker) and the perception by the recipient (listener). Recent developments in technology are increasing the efficiency and efficacy of acoustic analysis.

These analysis techniques are not only useful in a research context. The visual displays provided by readily available tools, such as waveforms, amplitude spectra, and spectrograms, can be used clinically for biofeedback to clients; analyzing speech and the quantitative analysis that comes from acoustic waveforms and spectrograms is valuable for speech assessment (Neel, 2010). Two no-/low-cost acoustic measurement tools include Praat (Boersma & Weenink, 2018) and WinPitch (P. Martin, 2004). Note that for further discussion of basic speech acoustics, see Gramley (2010), which covers source-filter theory, and for detailed discussion of methods and issues in acoustic studies of speech development and disorders, see Ciocca and Whitehill (2013), Hodge (2013), Kent and Vorperian (2018), and Neel (2010).

## **Results**

We will be answering the posited research questions in the order that they were asked in the following sections.

### *What Is the Contribution of Vowels to Speech Intelligibility?*

The contribution of vowels to speech intelligibility may have been underestimated, perhaps because it has been simply assumed that consonants have a greater effect than vowels on speech intelligibility. One way of determining the relative contributions of vowels and consonants to intelligibility is to use the “noise replacement paradigm,” in which either vowels or consonants are replaced by noise. Studies using this paradigm have shown a 2:1 intelligibility advantage of vowel-only (consonants replaced by noise) over consonant-only (vowels replaced by noise) sentences (F. Chen & Hu, 2019; R. A. Cole et al., 1996; Fogerty & Kewley-Port, 2009;

Kewley-Port et al., 2007). The same effect occurs even in Mandarin, a language with many fewer vowels (F. Chen et al., 2013). Kewley-Port et al. (2007) concluded that, “for spoken sentences, vowels carry more information about sentence intelligibility than consonants for both young normal-hearing and elderly hearing-impaired listeners” (p. 2365). This conclusion should be regarded with some caution, given that it is difficult to isolate consonants and vowels in the acoustic signal of running speech. Inevitably, segments attributed to vowels contain some consonant information if only because of coarticulation, and segments attributed to consonants likewise may contain some vowel information. As noted by Stilp and Kluender (2010), nonlinguistic sensory measures of uncertainty in the speech signal may be better predictors of intelligibility than traditional acoustic measures or linguistic constructs. However, the central point is that, so long as the distinction between vowels and consonants is made, vowels are important to speech intelligibility and should not be regarded as the poor sister of consonants in the goals and means of speech communication.

Moreover, vowels in sentences have the potential to carry information of several kinds, including information on the identity of the vowel itself, identity of flanking consonant(s) and neighboring vowels, syllable pattern based on the amplitude envelope of the utterance, prosodic content (rhythm, stress pattern, and rate), age and gender of the speaker, and affective content related to genuine or feigned emotion. The linkage between vowels and the prosody of an utterance makes them pivotal in the study of multisyllabic utterances, as discussed later. The linkage derives from the capacity of vowels to convey the acoustic cues of prosody but also perhaps from a shared bilateral cortical representation. In a functional magnetic resonance imaging study in which participants were asked to attend to either the vowels or consonants of syllables, generalization maps were bilateral for vowels but unilateral for consonants (Archila-Meléndez et al., 2018). Prosody is also typically presumed to be bilateral in its cortical representation (Wildgruber et al., 2009).

The proportion of time given to vowels in the speech input varies with the rhythmic class of languages. The percentage of the input stream for vowels is about 45% for stress-timed languages such as Dutch and English, about 50% for syllable-timed languages such as French and Italian, and about 55% for mora-timed languages such as Japanese (Ramus et al., 1999). This information may underlie the ability of newborns to discriminate between languages of different rhythmic classes (Ramus et al., 2000) and could provide an early foundation for linking the prosodic and segmental components of a language. It has been proposed that vowels and consonants play different roles in early phonological learning (Hochmann et al., 2011).

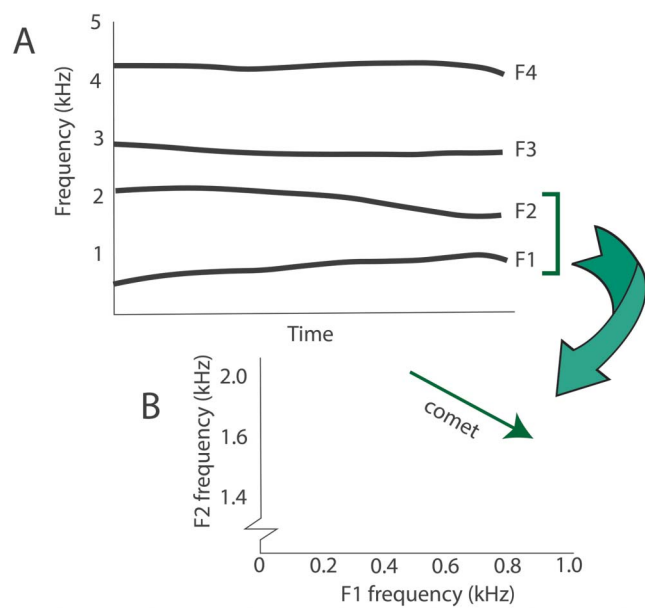
Given the capacity of vowels to signal different types of information, it is not surprising that they have a dynamic structure. The presumed vowel steady state identified as an essentially static formant pattern, as in the production of a sustained vowel, often is not observed in connected speech.

The dynamic nature of vowels derives not only from embedded information on surrounding sounds (coarticulation) and prosody but also from the intrinsic properties of vowels themselves.

### ***Are Vowels in American English Inherently Intrinsically Dynamic?***

Spectral change shows that vowels are intrinsically dynamic. Peeters (2019) wrote that, “In treating vowels like static particles in an articulatory and acoustic space perhaps the most important information to be conveyed by vowels is suppressed” (p. 67). The static particle perspective is reinforced by common practices in articulatory descriptions (such as classifying vowels as fixed patterns of tongue height and advancement) and acoustic analysis (such as representing vowels by single points in the F1–F2 plane). The traditional classification of English vowels into monophthongs and diphthongs has come under question with the recognition that even presumed monophthongs in several dialects of North American English are characterized by substantial spectral change during the vowel segment. In other words, the vowel is not strictly or satisfactorily defined by an invariant acoustic pattern. The concept of “vowel inherent spectral change” (Morrison & Assmann, 2012) appears in natural speech production and has been demonstrated to influence listeners’ perception of vowels. Vowel inherent spectral change is a relatively slow frequency variation compared to the rapid frequency variations associated with consonant–vowel or vowel–consonant transitions. An example is shown in Figure 6 for the formant patterns of vowel

**Figure 6.** Illustration of vowel inherent spectral change. The frequency shifts for F1 and F2 in Part A are shown as a comet in Part B.



/ae/ as produced by an adult male speaker. Note that in the spectrogram (Part A of the illustration), the frequencies of F1 and F2 change during the vocalic nucleus. The implication of such formant shifts is that the acoustic representation of vowels requires more than formant estimation at a single time point (such as the middle of a vowel steady state, if such a segment can even be identified; it frequently cannot). Specification of the spectrotemporal pattern requires that formant trajectories be represented by two time points or, alternatively, by one time point and the slope of the formant trajectory. This idea is shown in part by using a comet (line with an arrowhead) to indicate the changes in F1 and F2 frequencies in a bivariate plot.

The implication is that the vowels of American English should probably be plotted in the F1–F2 plane as comets (line segments) rather than single points. Classic illustrations of vowel formants (e.g., Peterson & Barney, 1952) represent individual vowels as points, but alternative representations are likely to become more common as research into vowel inherent spectral change continues. For examples of averaged vowel formant trajectories determined from two large databases, see Sandoval et al. (2019). Children as young as 3–5 years of age show evidence of vowel inherent spectral change similar to that in adults (Assmann & Katz, 2000; Assmann et al., 2013). This phenomenon applies as well to vowels in second-language learners (Rogers et al., 2012; G. Schwartz et al., 2016) and may be useful in assessing learning progress. Vowel inherent spectral change is a challenge to conventional analyses that focus on vowel steady states or single time points of formant measurement. This phenomenon is also of interest in developing dynamic specifications of articulatory–acoustic features of vowels. For a more complete discussion, see Morrison and Assmann (2012).

Vowel inherent spectral change should be distinguished from context-specific coarticulatory effects, in which production of a target sound is affected by surrounding sounds, especially the flanking consonants but even by nonadjacent vowels (J. Cole et al., 2010). Therefore, the production of a vowel sound incorporates both intrinsic spectral change and coarticulation, a combination that often results in time-varying acoustic properties.

### ***How Can Acoustic Data Help to Create a Picture of Vowel Development in Children?***

The development of vowel perception and production involves a cascade of events beginning with the fetus and proceeding through adolescence. The following discussion addresses both developmental data on vowels and the theoretical interpretations of these data. Selected aspects are summarized in Table 1, which shows developmental events at various ages, beginning before birth and extending to 15 years of age. The discussion in this section amplifies some of these events.

As a theory, acoustic phonetics has to do with the sounds coming from the speech mechanism and how they



**Table 1.** Normative milestones in the development of vowel perception and production.

Age	Perception feature	Production feature
In utero	Exposure to an ambient language affects aspects of speech perception after birth (see examples below).	n/a
1–2 days	Neonates prefer vowels from the native language as opposed to vowels from a foreign language (Moon et al., 2013).	Vowel-like sounds are produced in cry and in comfort vocalizations (humming and cooing). Four basic cries have been identified within the first month. Following the birth cry, these are basic cry, pain cry, and temper cry (Petrovich-Bartell et al., 1982).
2 months	n/a	Production of vowels in early vocalizations in all languages. Especially in the first month, vocalizations may take the form of quasiresonant nuclei (lacking the full resonant quality of vowels). Fully resonant nuclei appear between 2 and 4 months.
3–5 months	Vowel prosody emerges? “Cry” vs. “fuss” perceived by mothers based on peak intensity, F2, and F1:F2 ratio (Petrovich-Bartell et al., 1982).	In infants who imitate adult vowels, the vowels /i u a/ become more tightly clustered from 3 to 5 months (Kuhl & Meltzoff, 1996).
6 months	Discrimination of extrinsic vowel durations in some infants (Eilers et al., 1984). Generalization of vowel exemplars from adult male to adult female or child (i.e., speaker normalization; Kuhl, 1983). Discrimination of spectrally dissimilar vowels (Kuhl, 1979).	Onset of canonical babbling based largely on consonant + vowel (CV) syllables. FSL common in babbling but may disappear only to reappear later (U-shaped developmental curve; Nathani et al., 2003).
10 months	Development of the duration distinction for PVCV (Ko et al., 2009).	Typical onset of jargon babbling, which has characteristics of intonation, rhythm, and pausing carried largely by vowels. Vowel formants in infant babbling take on characteristics of the ambient language (de Boysson-Bardies et al., 1989). 10-month-olds prefer vowels of normal duration over stretched vowels (Kitamura & Notley, 2009).
12 months	Discrimination of tense vs. lax vowels. Theories such as NLM (1993), NLM-e (Kuhl et al., 2008), and PRIMIR (Werker & Curtin, 2005) assert that experience-based perceptual reorganization occurs in the first year, leading to decreased sensitivity to nonnative contrasts and increased sensitivity to native contrasts. Infants are sensitive to mispronunciations of vowels in familiar words by as early as 14 months of age (Mani et al., 2012).	The corner vowels [i u a] appear around this age or shortly after (Davis & MacNeilage, 1990; Selby et al., 2000; Templin, 1957). These vowels correspond to the natural referent vowels (Polka & Bohn, 2011), and they appear to establish the basic vowel triangle of articulation and acoustics. These vowels also meet the dual criteria of dispersion and focalization. Language-specific rhythm begins to emerge at 12 months (Post & Payne, 2018).
24 months	18–24 months—what was perceived in the 7-month time frame regarding native language gives rise to phonetic categorization leading to language learning (Werker & Hensch, 2015).	Production of diphthongs in most children. Vowel duration is adjusted for tense–lax distinction and PVCV (Ko, 2007).
36 months	n/a	PVCV generally present (Krause, 1982; Raphael et al., 1980). Vowel inherent spectral change may be present (Assmann & Katz, 2000).
4 years	n/a	Production of all nonhrotic vowels in most children and production of rhotic vowels by many but not all children. Onset of sexual dimorphism of vowel formants, with boys having lower formant frequencies than girls (Kent & Vorperian, 2018; Vorperian & Kent, 2007).
5 years	Adultlike perception categorization of PVCV (category boundary and category separation; Lehman & Sharf, 1989).	Rhythmic patterning in stress-timed languages such as English is still being refined (Post & Payne, 2018).
8 years	n/a	Postvocalic voicing—category boundary and category separation in production were adultlike by 8 years of age. Supralaryngeal anatomy is now mature (de Boer & Fitch, 2010).
10 years	Adultlike perceptual consistency of PVCV (Lehman & Sharf, 1989).	Variability in production still greater than in adults (Lehman & Sharf, 1989).
11–15 years	n/a	Voice transition in both sexes, larger change in boys (Berger et al., 2019; Maturo et al., 2012).

*Note.* FSL = final syllable lengthening; PVCV = postvocalic consonant voicing; NLM = Native Language Magnet Model; NLM-e = Native Language Magnet Model–Expanded; PRIMIR = Processing Rich Information from Multidimensional Interactive Representations; n/a = not applicable.

are processed by the human ear. This constant interplay between production and perception is at work in such young developing children. The innate need to connect drives development, and reinforcement in communicative exchanges hones these skills.

### General Developmental Features

This section begins with some general features of the developmental pattern of auditory function and vocal tract anatomy pertinent to vowel production. With respect to auditory function, the human ear functions 2.5–3 months before birth, so that the fetus has some auditory exposure well before birth (Pujol et al., 1991; Querleu et al., 1988). A meta-analysis of studies of vowel discrimination in infants (Tsuji & Cristia, 2013) supported the conclusion that native and nonnative discrimination proceed in opposite directions over the first year of life with a distinction evident by about 6 months of age. With respect to vocal tract anatomy, between birth and adulthood, the structures that form the tract undergo changes in size and shape (Kent & Vorperian, 1995). These changes are considerable, as reflected in the assertion that humans begin life with a vocal tract like that of a chimpanzee (P. Lieberman et al., 1972). An adultlike anatomy unfolds gradually so that an essentially mature morphology (but not length) of the vocal tract is present by the age of 6–8 years (de Boer & Fitch, 2010). However, additional growth and reshaping continue until late adolescence, especially in males. More detailed discussion follows with respect to development stages. It should be emphasized that there are substantial individual variations in vowel development (Donegan, 2013). This summary should be considered a general pattern to which exceptions are likely in individual children. Figure 7 gives a graphic summary of typical patterns in vowel acquisition and will be referenced in the following discussion of various developmental stages.

### Prenatal Stage

An account of vowel perception begins before birth. Studies have shown that vowels can be perceived and discriminated in utero (Groome et al., 1997; Lecanuet et al., 1987; Shahidullah & Hepper 1994), but the same has not been shown for consonants (Granier-Deferre et al., 2011). It also has been reported that in utero experience with an ambient language affects vowel perception after birth (Moon et al., 2013) so that babies enter the world with an orientation to the language they will learn from the adult community. Newborns prefer stories heard in the womb over unheard stories (DeCasper & Spence, 1986), probably because of the influence of vowels and the rhythmic pattern of speech. For these and other reasons, neonates can be said to show a vowel advantage in speech processing (Nazzi & Cutler, 2019). The fetus also is exposed to the mother's voice and is biased toward that voice after birth (Fifer, 1987). Spence and DeCasper (1987) presented evidence that prenatal experience with low-frequency characteristics of maternal voices carries over into preferences in postnatal perception of maternal voices. Such low-frequency characteristics likely derive from vowel

sounds, which can be transmitted into the intrauterine environment.

### Neonatal Stage

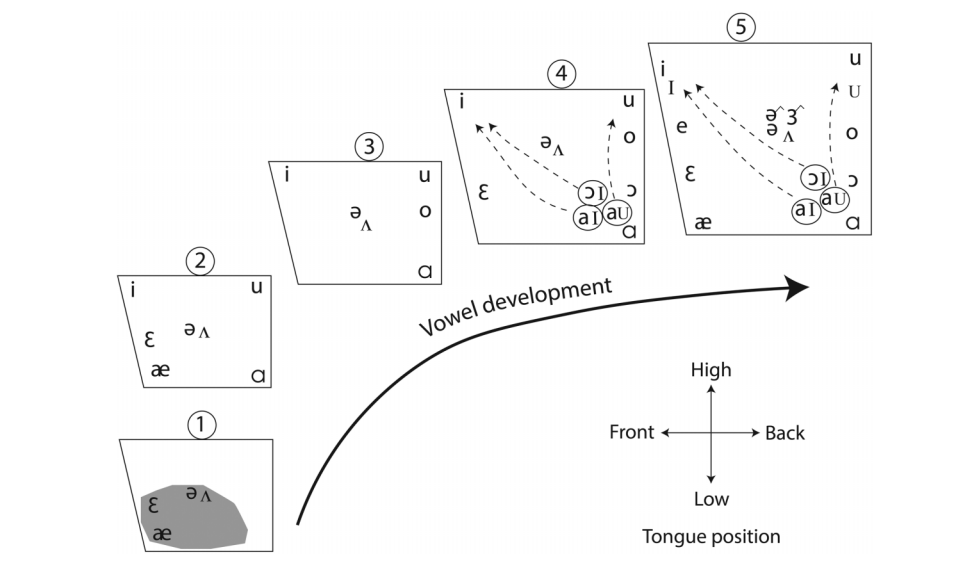
Vocalization of vowels (or vocants, in the terminology introduced by J. A. M. Martin, 1981) is one of the first reliably observed behaviors to emerge in neonates. In an observational study involving four different countries, Ertem et al. (2018) reported that at least 50% of babies vocalized vowels by 1 month of age. The fact that babies produce and hear their own vowel sounds shortly after birth may lay the foundation for auditory–motor correlations that will be expanded and refined with maturation.

### First Year of Life

Vowel development in early infancy is not a process of simple and gradual accretion in which vowels are added one at a time to form a language-specific repertoire. As explained in the following, it is better to consider vowel development in three fundamental aspects: early vocal behaviors favoring low vowels, mostly front or central; establishing the corner vowels; and phonetic mastery of the vowel repertoire. The rationale for this alternative is discussed in the following.

Vowellike sounds produced in early infancy are not isomorphic with vowels produced during later speech development. As Oller et al. (2013) noted, “Protophones occurring before canonical babbling cannot be transcribed sensibly in the International Phonetic Alphabet...because they generally do not contain well-formed and distinguishable consonants and vowels” (p. 6319). Although phonetic symbols often are used as convenient labels for early vowellike sounds, such usage should not imply that the sounds so identified are identical to vowels later used to form words. J. A. M. Martin (1981) used the term “vocant” for such a vowellike sound. Especially in the early stages of infancy, vocants are not necessarily tightly linked to vowel representations in an adult phonetic system but rather may be developmentally specific to particular combinations of perceptual experience, anatomic configuration, and motor capability. From an acoustic point of view, vocants produced in the first year of life are perhaps best regarded as regions of variable density in the vowel articulatory (or acoustic) space. As Donegan (2013) noted, there is an apparent affinity for vowels in the “lower left quadrant” of the vowel space (Stage 1 in Figure 7), a feature evident in the data reported by O. C. Irwin (1948), Davis and MacNeilage (1990), MacNeilage and Davis (1990), Kent and Bauer (1985), and Ménard et al. (2004). The high frequency of occurrence of vocants in this region likely reflects articulatory preferences based on the vocal tract anatomy in which the tongue is relatively wide and flat with a relatively anterior mass, the pharynx is short, dentition is emergent, and the vocal tract shape lacks the 90° craniovertebral angle that is characteristic of adults. These combined features are conducive to tongue carriage within the lower left quadrant, giving rise to vocants that resemble especially the adult vowels /ɪ ε æ ə/. The phonetic symbols are useful to acknowledge some degree of auditory similarity

**Figure 7.** Diagrammatic representation of typical vowel development in children. The numbers correspond generally to age in years, but there is substantial individual variation. Based on patterns described by Donegan (2013), Otomo and Stoel-Gammon (1992), and Stoel-Gammon and Herrington (1990).



between infant and adult vowels but, as argued earlier, should not be taken as evidence of phonemic acquisition. The statistical preponderance of vocants in one quadrant of vowel space is a notable feature that is most confidently interpreted with respect to developmental anatomy and physiology.

A feature that may occur less frequently but is nevertheless of considerable importance is formation of the vowel space determined by the point vowels (the word “vowels” is now used in favor of “vocants”). The corner vowels /i/, /u/, and /a/ appear in the first or second year of life (Buhr, 1980; Davis & MacNeilage, 1990; O. C. Irwin, 1948; Selby et al., 2000; Templin, 1957; Wellman et al., 1931). These vowels, shown in Stage 2 in Figure 7, establish the acoustic and articulatory boundaries of the vowel space, within which other vowels can be produced (Kent, 1992). This aspect of vowel production has a correlate in perception. Polka and Bohn (2003) reported that infants have a perceptual bias for vowels that are close to the periphery of the F1/F2 vowel space and suggested that a bias for these vowels is language universal. J.-L. Schwartz et al. (2005) cast these results in the framework of the dispersion–focalization theory of vowel systems, proposing that focalization (the convergence between two consecutive formants in a vowel spectrum) increases the perceptual salience of the peripheral vowels relative to other vowels not having this property. They commented that “focal vowels, more salient in perception, provide both a stable percept and a reference for comparison and categorization” (p. 425). Focalization can be seen in Figure 1 as the proximity of F2 and F3 for vowel /i/ and F1 and F2 for vowels /a/ and /u/. Polka and Bohn (2011) accepted the J.-L. Schwartz et al. (2005) interpretation of the data in their earlier report and further proposed a natural reference vowel framework to account for

phonetic development in children. The peripheral vowels are anchors within the natural reference vowel framework. The perceptual salience of the focal vowels has an articulatory counterpart insofar as these vowels are produced with extreme positions in the oral cavity (see Figure 2). As infants explore their vocal abilities, they may come to rely on peripheral vowels because of both the perceptual bias of focalization and the anatomic boundaries of vowel production.

Japanese and American English vowels differ in number but share similar spectral and temporal characteristics when they are produced in connected speech (Nishi et al., 2008). Using an acoustic–articulatory inversion model with scalable vocal tract size, Oohashi et al. (2017) noted the following developmental sequence of vowel development in Japanese: At 6–9 months, coordination of the tongue body and lip aperture forms three vowels (front, back, and central); at 10–17 months, the jaw and tongue apex are recruited to differentiate the original three vowels into five; and at 18 months and older, tongue shape is further refined to produce the vowels of Japanese. Research is needed to determine if the same general pattern occurs in other languages.

### Second Year of Life

Most infants make considerable progress in speech acquisition by the age of 24 months. Vowels typically produced at this stage are shown in Stage 3 in Figure 7. Vowel production is likely enhanced by several factors beyond increased familiarity with the ambient language. The vocal tract has been sufficiently remodeled so that the larynx is well separated from the nasopharynx, which contributes to a lengthening of the pharynx and greater motility of the tongue (de Boer & Fitch, 2010; Kent, 1992). Velopharyngeal closure for spontaneous speech production is reliably

accomplished by about 19 months (Bunton & Hoit, 2018), so that vowels are produced with a well-defined oral resonance. In addition to changes in macroanatomy, there appear to be changes in microanatomy. For example, by 2 years of life, the proportion of slow-twitch and fast-twitch fibers in the tongue has reached adult values, which may be evidence that the tongue musculature is being adapted to the requirements of speech production (Sanders et al., 2013). Among these requirements is fatigue resistance during the performance of long stretches of speech. The high proportion of slow-twitch fibers in the posterior part of the tongue may be advantageous for the continuous vowel–vowel movements in conversational speech. This is not to argue for anatomic and biological determinism, but rather to say that musculoskeletal development contributes to the conditions favoring intelligible speech and the identification of phonetic units comparable to those in adult speech.

Vowel development in this period has been studied principally with diary studies (Leopold, 1947; Menn, 1976; Velten, 1943) and cross-sectional studies (Buhr, 1980; Davis & MacNeilage, 1990; O. C. Irwin, 1948; Selby et al., 2000; Templin, 1957; Wellman et al., 1931). The most general conclusion is that vowels at the extremes of the quadrilateral are acquired before those in more central positions, with the exception of an early preference for vowels in the low-front region of the quadrilateral (as discussed previously). This developmental pattern is consistent with establishing the acoustic and articulatory boundaries of the vowel space as a framework for vowel acquisition. Many theories have been advanced to account for phonological development beginning around 2 years of age.

At some point in development, it is appropriate and useful to describe vowel production in terms of phonetic mastery, that is, the age at which a sound produced in a specified context (e.g., a target word in an articulation test). This is judged to be produced correctly by a certain percentage (e.g., 50%, 75%, 90%) of children or productions by a given child. Inferences of mastery tend to be associated with morphology and the lexicon in that assessment tools typically rely on words as the units in which sounds are judged. This aspect of development is discussed in following sections for later ages. As noted earlier, by the end of the first year of life, vocants are being replaced by vowels (i.e., it becomes appropriate to describe development in phonemic terms), so that it becomes appropriate to use terms such as phonetic mastery.

### Third Year of Life

Studies such as those by J. V. Irwin and Wong (1983) and Templin (1957) indicate that nonrhotic vowels are mastered in typical development by the age of 3 years. Although mastery may be delayed in some children, the received wisdom based on these early studies appears to be that vowel development is largely accomplished by this age except for the rhotics. However, more recent studies indicate that mastery of vowels in polysyllabic words and connected speech is not achieved until several years later (James et al., 2001; Wren et al., 2013). Possibly, the refinement

of vowel production depends in part on other aspects of speech development, such as prosody. Acoustic research, summarized in a following section, also indicates that refinement of vowel production continues beyond the age of 3 years. Stage 4 in Figure 7 summarizes typical development at about this age.

### Fourth to Eighth Year of Life

By the age of 4–5 years, the vowel system is typically complete except for rhotic vowels in some children. Stage 5 in Figure 7 represents the essentially mature vowel system. De Boer and Fitch (2010), drawing on research from Fitch and Giedd (1999), D. E. Lieberman and McCarthy (1999), and Vorperian et al. (2005), stated, “Independent studies have shown that a mature supralaryngeal vocal tract anatomy, with a rough match between oral and pharyngeal cavity length is not achieved until age 6–8 years” (p. 43). Although the usual perception-based criteria of phonetic mastery may be satisfied as early as 3 or 4 years, maturation of motor control is a continuing process and likely reflects the emergence of adultlike morphology of the vocal tract.

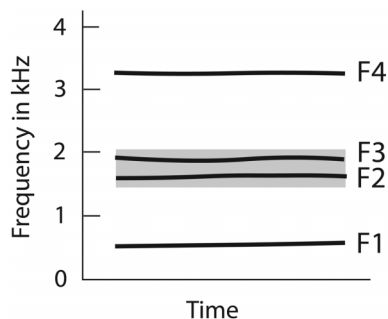
Children of this age have largely consolidated the perceptual, motor, and phonological aspects of speech into adultlike patterns (Ball & Gibbon, 2013). However, refinements continue in many children for both the perceptual and motor skills of speech (as shown in Table 1).

### *What Are the Distinctive Acoustic Properties of Rhotic Vowels and Diphthongs?*

The rhotic vowels (such as the vowels in the word “further”) and the rhotic diphthongs (as in the words “ear,” “oar,” and “our”) have distinctive properties. As discussed, they often are the last vowels to be acquired by children (Stoel-Gammon & Pollock, 2008). The class of rhotics in American English (both consonants and vowels) share a common acoustic feature—a reduced F3 frequency that sets them apart from other sound classes (Alwan et al., 1997). However, it may not be a low F3 frequency per se that is the hallmark of rhotic acoustics. Rather, it may be that a near-merging of F2 and F3 produces a spectral band of energy that is perceived as a “rhotic formant” (i.e., F2 and F3 are not discriminated separately but integrated as a single energy band). Figure 8 shows a stylized spectrogram of the formant pattern for a production of the rhotic vowel /ɜ/. The gray band represents the combined energy of F2 and F3. Heselwood and Plug (2011) conducted two perceptual experiments that showed that F3 contributes to the perception of rhoticity insofar as the proximity of F3 to F2 produces a dominant band of energy in the F2 frequency region. According to the broad-band auditory integration hypothesis (Bladon, 1983), the F2–F3 convergence is fused in perception if the two formants are within 3.5 Bark of each other. A surprising outcome of the perceptual studies reported by Heselwood and Plug was that reducing the amplitude of F3 actually enhanced perceived rhoticity. This result may be important in



**Figure 8.** Stylized spectrogram of the rhotic vowel /ɜ:/ showing the close positioning of F3 and F2. The gray band illustrates the combined energy of these formants and may be considered the rhotic formant (i.e., an integration of the energy in F2 and F3).



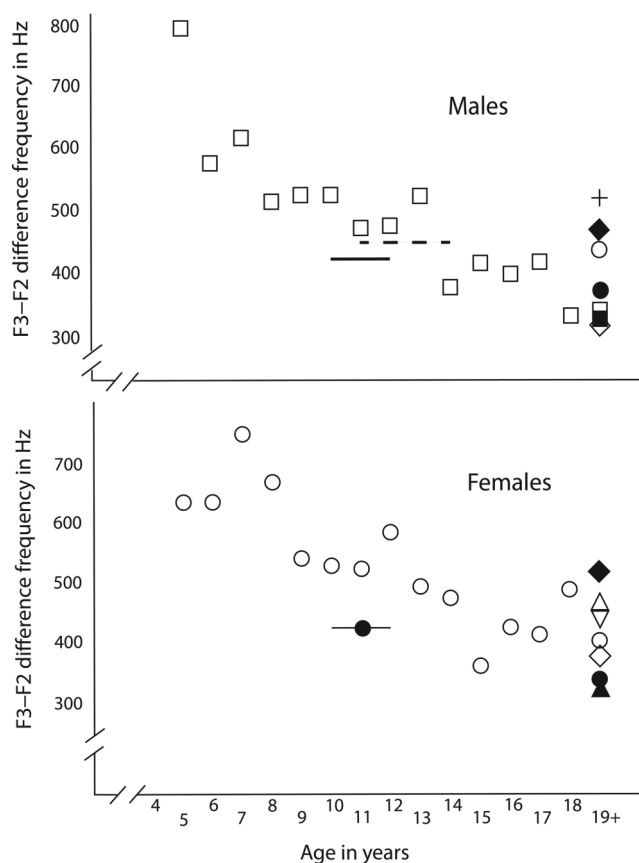
designing systems of visual feedback for the acoustic properties of rhotics.

Rhotics can be characterized acoustically as either a ratio of F3/F2 or the difference in frequency between the two formants (Chung & Pollock, 2019; Flipsen et al., 2001). The frequency difference is useful in estimating the likelihood of broad-band auditory integration. Figure 9 shows for males and females, respectively, the F3–F2 frequency difference for vowel /ɜ:/ across age. The mean difference is about 600–700 Hz for young children and about 400–500 Hz for adults. Across the ages represented, the F3–F2 differences for both males and females fall in rather tight bands, which is evidence of the salience of this acoustic cue of rhoticity.

### What Is the Interaction Between Vowel Perception and Vowel Production in Early Development?

Infants preferentially attend to vowel sounds that have infant-like voice pitch and/or formants over vowel sounds that do not have infant-like properties (Masapollo et al., 2016). The authors interpreted this finding to mean that infants' production of speech sounds influences their perception of infant speech. Research on cortical auditory evoked potentials (McCarthy et al., 2015) revealed that 4- to 5-month-old infants have two-dimensional perceptual maps that reflect F1 and F2 acoustic differences between vowels, but 10- to 11-month-old infants have maps that are less related to acoustic differences but tend to give greater weight to adjacent vowels in the vowel quadrilateral (e.g., /i/–/ɪ/). These results were interpreted to indicate a shift from a primarily acoustic to a more phonologically driven processing. These two studies are examples of perception–production interaction in the development of vowel systems in children. Evidence of this interaction can be seen in a wide range of studies, including speech perception and production in infants (Altwater-Mackensen et al., 2016; Bruderer et al., 2015; Majorano et al., 2014), perception–production transfer from birth language memory (Choi et al., 2017), vowel perception and production in adults (Fox, 1982), and adjustment to an articulatory disruption (Seidl et al., 2018). Taken

**Figure 9.** F3–F2 differences for vowel /ɜ:/ as a function of speaker age in males and females. Data sources for males are as follows: dashed line = 11- to 14-year-olds in Angelocci et al. (1964); cross = Childers and Wu (1991); unfilled circle = Hagiwara (1995); solid line = 10- to 12-year-olds in Hillenbrand et al. (1995); unfilled squares = Lee et al. (1999); unfilled diamond = Peterson and Barney (1952); filled circle = B. Yang (1996); filled diamond = Zahorian and Jagharghi (1993). Data sources for females are as follows: filled diamond = Childers and Wu (1991); inverted triangle = Hagiwara (1995); solid line with filled circle = 10- to 12-year-olds in Hillenbrand et al. (1995); filled circle = adults in Hillenbrand et al. (1995); unfilled circles = Lee et al. (1999); filled triangle = Peterson and Barney (1952); unfilled diamond = B. Yang (1996); unfilled diamond = Zahorian and Jagharghi (1993).



together, these discoveries point to neural processes that link the perceptual experience of speech sounds with the motor processes involved in their production. Computational models incorporating this idea are discussed in a later section, (*How Can Sensory-Motor Mapping Serve in a Theoretical Framework for Treating Vowel Production and Perception?*).

### Can Acoustic Measures Serve as an Index of the Precision of Vowel Production?

Acoustic methods have contributed to the study of variability in both the temporal and spectral aspects of speech production (Kent, 1976; Lee et al., 1999). Work on vowels pertains primarily to formant patterns and segment durations.



Perceptual methods, such as phonetic transcription, are not sensitive to all variations in vowel articulation. Acoustic data on formant patterns and durations evince variability even in tokens that are judged to represent the same phoneme. In motor behavior generally, precision is often assessed by determining the variability in repeated tokens of a behavior, and a typical hypothesis in the study of motor skills is that precision will increase (and variability will decrease) with maturation. Increased precision of vowel production has been reported in several studies (Eguchi & Hirsh, 1969; Gerosa et al., 2007; Lee et al., 1999; J. Yang & Fox, 2013). Adultlike precision is reached at about 12 years of age, about 2 or 3 years later than adultlike precision for temporal aspects of speech.

Low vowels are produced with greater acoustic (and presumably articulatory) variability than high vowels (J. Yang & Fox, 2013). The reasons for this difference may be that production of the high vowels benefits from (a) somatosensory feedback of the tongue against the palate, teeth, or both (Gick et al., 2017) and (b) lateral lingual bracing against the upper structures of the oral cavity (Gick et al., 2017). In addition, the high vowels may have a more critical constriction to achieve the desired acoustic results, which is consistent with greater coarticulatory resistance in these vowels (Recasens & Rodríguez, 2016).

### ***What Do Acoustic Data Add to the Knowledge of Phonetic Mastery of Vowels?***

The developmental primacy of vowels seems desirable from the perspective of motor control given that consonant articulation often is adjusted to the articulatory features of vowels (e.g., the articulatory accommodation of velar stops to the following vowel). From the perspective of phonetic mastery, vowels as a class appear to be produced more accurately than consonants, but perceptual judgments may not be sensitive to all aspects of speech maturation. Formant frequency data show developmental changes in the organization of acoustic vowel categories beyond the age typically given for the phonetic mastery of vowels in English (J. Yang & Fox, 2013); Mandarin (J. Yang & Fox, 2017); Hungarian (Auszmann & Neuberger, 2014); and, in a multilanguage study, Cantonese, American English, Greek, Japanese, and Korean (Chung et al., 2012). Because these languages represent different language families and different vowel inventories, it appears that stabilization of vowels beyond supposed phonetic mastery is universal. The implication is that children continue to refine the characteristics of vowels between the ages of 3 and 7 years and perhaps even later (up to age 13 years according to Auszmann & Neuberger, 2014). Similarly, protracted development of aspects of speech motor control has been reported in acoustic and kinematic studies showing improved precision of speech production until at least 16 years of age (Walsh & Smith, 2002) and perhaps as late as 30 years (Schötz et al., 2013). Fixing an exact age for maturation is problematic because different aspects of speech motor skill may have distinct developmental trajectories. Phonetic mastery, as typically determined by

perceptual judgments, is useful in determining overall conformity with a language's phonological system, but it is not sensitive to the full spectrum of underlying processes of sensory and motor maturation. Refinement of speech motor control is an ongoing process that is adapted to anatomic changes, sociocultural influences, and perhaps other variables yet to be identified.

### ***How Is Vowel Production Compensated for in Articulatory and Auditory Perturbations?***

A given vowel often can be produced with substantially different underlying articulations, so long as an acoustically critical vocal tract shape is formed for the specific vowel. Particularly notable in this respect are different contributions of jaw and tongue positions. Because the jaw provides carriage for the tongue, articulatory movements of the tongue are predicated on the current position of the jaw. In both research and clinical practice, bite blocks are a convenient way to create perturbations or adjustments requiring compensatory articulation (Bahr & Rosenfeld-Johnson, 2010; Crary, 1995; Dworkin, 1978; Netsell, 1985). Adults are capable of making compensations to a bite block on the first glottal cycle of phonation, that is, well before auditory feedback is available to compute and amend motor commands (Gay et al., 1981). Virtually instantaneous compensation occurs even in a condition of bilateral anesthetization of the temporomandibular joint, application of a topical anesthetic to reduce tactile information from the oral mucosa, and white noise masking to reduce auditory information (Kelso & Tuller, 1983). However, Hoole (1987) reports a case study in which auditory masking prevented bite block compensation in a 29-year-old man who had suffered closed-head trauma and whiplash injury to the cervical cord in a sporting accident. The man apparently recovered completely from postaccident dysarthria but had a persisting loss of oral sensation extending from the pharynx to the lips.

Studies that have been aimed at determining when children are capable of bite block compensation have yielded somewhat discrepant results. Gibson and McPhearson (1979/1980) concluded that bite block compensation is incomplete for children aged 6–7 years. However, apparently successful compensation was reported in several later studies, including studies of de Jarnette (1988), Baum and Katz (1988), and Smith and McLean-Muse (1987). De Jarnette found that compensation was accomplished by all participants in three groups (five children with typical speech development, aged 6.4–7 years; five children with moderate articulatory disorders, aged 5.9–8.1 years; and five adults with typical speech, aged 19.3–32.1 years). Baum and Katz observed no differences in F1 or F2 in two groups of children aged 4–5 and 7–8 years in producing both jaw-free and jaw-fixed conditions. In a kinematic study, Smith and McLean-Muse observed essentially no differences between children and adults in compensating to a bite block, concluding that “the ability to produce speech under experimental conditions such as these is apparently acquired by normally developing children by at least

4–5 years of age” (p. 752). As noted earlier, bite blocks are sometimes used in assessing and treating speech disorders, and it is therefore important to know the age at which compensation for bite blocks is accomplished. Apparently, this ability is present by about 4 years of age in typically developing children. Effective acoustic measurement should be considered when making clinical decisions such as what effect a bite block has on an individual speaker’s vowel production.

The effect of auditory perturbations has been investigated by introducing formant changes in vowels produced by talkers. MacDonald et al. (2012) studied responses in a real-time formant perturbation task in three age groups: toddlers, children, and adults. Children and adults reacted by changing their vowels in a direction opposite to the perturbation, that is, correcting for it. In contrast, the toddlers did not change their production in response to altered feedback. Similarly, Ménard et al. (2008) concluded from a study of compensation strategies for a lip-tube perturbation that 4-year-old children did not integrate the auditory feedback in a way that contributed to motor learning, a failure that was attributed to immature internal models. This task may contribute to a deeper understanding of developmental speech sound disorders. Terband et al. (2014) reported that, although most children with this disorder can detect discrepancies in auditory feedback and can adapt their target representations, they fail to compensate for the perturbed auditory feedback.

### ***How and When Do Language and Dialect Differences Influence Vowel Production in Infant Vocalizations and Early Childhood?***

In an early and influential report on cross-language differences in babbling, de Boysson-Bardies et al. (1989) obtained F1 and F2 frequencies of vowels produced by twenty 10-month-old infants from Parisian French, London English, Hong Kong Cantonese, and Algiers Arabic language backgrounds. Significant differences in formants were observed between infants across language backgrounds, and these differences reflected those in adult speech in the corresponding languages. This study showed that, even before the age of 1 year, infants are adjusting their vowel production in ways that accord with the ambient language.

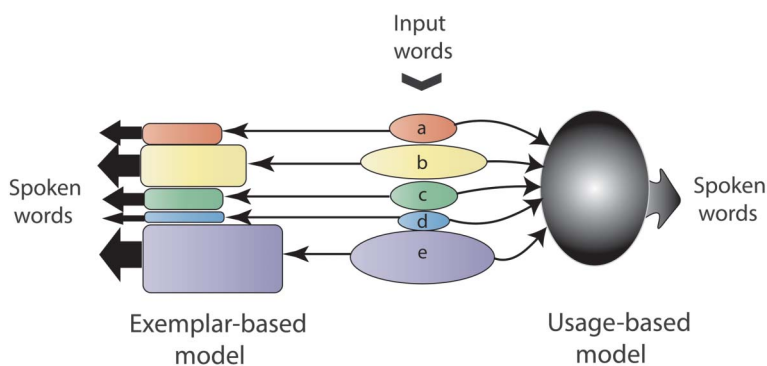
The effect of ambient language on infant vocalizations has been corroborated in a number of studies (Alhaidary & Rvachew, 2018; L. M. Chen & Kent, 2005, 2010; de Boysson-Bardies & Vihman, 1991; Engstrand et al., 2003; Grenon et al., 2007; Rothgänger, 2003; Rvachew et al., 2008, 2006). This line of research shows that infants are aware of their linguistic environment and that they reproduce selected aspects of the ambient language in their own vocalizations. This auditory–motor correspondence is consistent with the hypothesis that, even in the first year of life, infants are shaping their vocalizations in ways that are compatible with the language to be learned. Research on speech perception in infants has given rise to influential theories such as those discussed in the following.

The “Native Language Magnet/Neural Commitment Theory” or “Native Language Magnet Model–Expanded” (Kuhl, 1992; Kuhl et al., 2008) accounts for the developmental processes by which the ability of infants to discriminate speech sounds is progressively adapted to their native language. It is proposed that early auditory experience produces a neural commitment to the phonetic units of the native language, forming prototypical representations of the phonemic inventory. This process enhances the auditory processing of native sounds but interferes with the detection of the sounds in nonnative languages. The theory integrates several processes and abilities, including cognitive and auditory processing skills, connections between speech perception and production, statistical learning, and social factors that affect learning. Therefore, the Native Language Magnet Model–Expanded proposes that speech perception develops through the interplay of several factors in the communicative environment. It can thereby account for the influence of the ambient language and the maturation of sensory and cognitive abilities.

The “Processing Rich Information from Multidimensional Interactive Representations” (Werker & Curtin, 2005) holds that the speech signal is processed by three dynamic filters (initial biases, developmental level of the child, and requirements of the specific language task at hand). The Processing Rich Information from Multidimensional Interactive Representations was developed to address two fundamental issues in infant speech perception. The first, that speech perception is both categorical and gradient, is resolved by the use of multidimensional planes. The second, that perception is influenced by both ontogenetic development and online processing, is handled by assuming that performance is continually changing and flexible as a function of age and task so that processing and representations are interwoven. Exposure to the ambient language is one aspect of the nascent representations.

Acoustic measures of formant patterns show that vowel production is influenced by regional dialects of American English (Clopper et al., 2005; Fox & Jacewicz, 2009) and that children’s vowel systems are regionally distinct by the age of 8–12 years (Jacewicz et al., 2011). Jacewicz et al. (2011) concluded that children acquire not only systemic relations among vowels but also dialect-specific patterns of formant dynamics. Vowel production in children may be influenced by variability of lexical exposure and the frequency of words encountered. Levy and Hanulíková (2019) concluded in an acoustic study of vowel production that children who experience greater input variability produce more variable vowels. This conclusion is particularly important for children in bilingual or bidialectal environments, for whom the diversity of acoustic input for vowels may be reflected in vowel production variability. Such an effect can be explained by either usage-based or exemplar-based models, as graphically conceptualized in Figure 10, which depicts versions of the two models. The input words labeled “a” through “e” represent word productions that vary in frequency of occurrence in a child’s auditory experience. For example, the word production labeled “e” has a high frequency of occurrence.

**Figure 10.** Illustration of exemplar-based and usage-based models to account for the effects of diversity in input on production of vowels. The exemplar-based model assumes that every perceived variant of a word (the boxes labeled “a” through “e”) gives rise to an exemplar in a direct acoustic-to-lexical mapping. Frequency of occurrence is represented by the size of the box for the exemplar. The usage-based model assumes that linguistic units are gradient categories formed in continuous fashion from experienced tokens. For both models, spoken words are produced in variable fashion reflecting the input diversity.



Usage-based models assume that linguistic units are gradient categories formed in continuous fashion from experienced tokens (Bybee & Beckner, 2010). Exemplar-based models assume that every perceived variant of a word gives rise to an exemplar in a direct acoustic-to-lexical mapping. As shown in Figure 10, words that are heard more frequently have a greater number of exemplars (a larger box in the illustration) than infrequent words (Schweitzer et al., 2015) and are therefore more likely to be produced.

### *How Do Speakers Adjust Their VSA in Accord With Listener Characteristics and Communication Settings?*

As noted earlier, VSA is a measure of the area contained within the vowel quadrilateral or vowel triangle. When speakers are asked to speak clearly (or do so spontaneously in an effort to communicate successfully under less than optimal conditions), they adjust their speech in several ways, often including expansion of the vowel space (called vowel hyperarticulation by some authors or over-articulation by some clinicians). Such expansion is a frequently noted characteristic of infant-directed speech, and it has been suggested that parents make this adjustment as a didactic strategy to aid children’s speech development. Talkers alter their speaking patterns for various types of listeners, including pets as well as young children and foreigners (Uther et al., 2007). However, what appears to be unique in infant-directed speech as compared with pet-directed speech is that expansion of vowel space occurs in the former and not in the latter (unless the pet in question is a talking parrot in which case a modest hyperarticulation is performed; Xu et al., 2013).

Vowel hyperarticulation also may be used intentionally or unintentionally by speech-language clinicians, especially when addressing young children. Mothers using

infant-directed speech appear to use a raised larynx to achieve an expanded vowel space (Kalashnikova et al., 2017) even as they deploy an increased vocal pitch, which may be a signal for nonaggressiveness or rapport. Both hypoarticulation expressed as a reduced VSA and hyperarticulation expressed as an increased VSA are of interest in assessing vowel production in developing and disordered speech. The main point to be made is that VSA is modulated by several factors, including characteristics of the listener and the communication environment. In other words, VSA is not an invariant physical value of the vocal tract of a given speaker.

### *In What Ways Is Vowel Perception and Production Vulnerable to Speech and Language Disorders in Children and Adults?*

As reviewed in this section, both the perception and production of vowels are affected by a number of communication disorders in children and adults. We begin with a discussion of VSA (or similar acoustic indices) used to quantify atypical acoustic patterns of vowel production. Then, examples are given of vowel disorders in selected clinical populations.

#### **The Acoustic Vowel Space in Communication Disorders, a Focus on Production**

VSA is perhaps the most frequently reported index of disordered vowel production in both children and adults and is one of the most extensively reported acoustic measures for any aspect of speech articulation. Sandoval et al. (2013) commented that, “Vowel space area (VSA) is an attractive metric for the study of speech production deficits and reductions in intelligibility, in addition to the traditional study of vowel distinctiveness.” However, as discussed in the following, the usual method of calculating VSA has

come under criticism, and it is likely that, in the future, other metrics will gain favor. Alternative metrics are summarized later in this section. In the immediately following discussion, VSA is emphasized because of its frequent mention in the literature under review.

For children's speech, a reduced VSA has been noted for several disorders, including cerebral palsy and other childhood neurological disorders (Higgins & Hodge, 2002; Hustad et al., 2010; Liu et al., 2005; Narasimhan et al., 2016), dyslexia (Bertucci et al., 2003), residual speech sound disorders (Spencer et al., 2017), and Down syndrome (Bunton & Leddy, 2011). These results can be explained largely by auditory, motor, or auditory-motor limitations. However, increased VSA also has been observed in clinical populations, including two groups of children with hearing impairment (deaf with cochlear implants and hearing-impaired with hearing aids; Baudonck et al., 2011). The authors suggested that the enlarged VSA was the consequence of overarticulation (synonymous with hyperarticulation, mentioned earlier), a compensation for reduced auditory feedback by relying on proprioceptive feedback during speech production.

Reduced VSA has also been reported for adults with various disorders and conditions, including, but not limited to, acquired dysarthria (Bang et al., 2013; S. Kim et al., 2014; G. S. Turner et al., 1995; Weismer et al., 2001), glossectomy (Kaipa et al., 2012; Takatsu et al., 2017; Whitehill et al., 2006), oral or oropharyngeal cancer (de Bruijn et al., 2009; van Son et al., 2018), Class III malocclusion (Xue et al., 2011), stuttering (Blomgren et al., 1998; Hirsch et al., 2008), and psychological distress or with self-reported symptoms of depression and posttraumatic stress disorder (Scherer et al., 2016, 2015). It is likely that increased VSA can occur in some clinical populations. Using a measure of dispersion of density, Kelley and Aalto (2019) concluded that head and neck cancer patients may use hyperarticulation strategies to increase the clarity of their speech postsurgery.

It is becoming clear that VSA as typically measured may not be sensitive to all features of clinical interest, and it is advisable to consider other analyses in preference to VSA or to complement it (Karlsson & van Doorn, 2012; Kent & Vorperian, 2018). For example, vowel articulation index and its inverse, formant centralization ratio, effectively reduce noise variability between speakers while maintaining high sensitivity to vowel centralization, which assists in differentiation of disordered speech compared to VSA in persons with reduced articulatory movements seen in hypokinetic dysarthria (Sapir et al., 2011). The point cloud has been posited as a good tool for childlike speech models (Story & Bunton, 2016), as it treats the acoustic space as a three-dimensional cloud (Coen et al., 2015). One advantage of this approach is that little to no information is abstracted or lost as in the case of derived indices such as VSA or vowel articulation index. The point cloud is faithful to the source data and reveals the distribution of values rather than a summary statistic such as the mean. Triangular VSA has been posited to correlate closely to intelligibility in

normal speakers (Bradlow et al., 1996) and to account for dialectal differences in American English (Fox & Jacewicz, 2009). Because different analyses have relative strengths and weaknesses with different speech patterns, clinicians and researchers are called upon to make an educated decision on what tool to use, considering client's age, etiology of disability, and functioning level.

### **The Backbone of Production: Perception**

An earlier section (*What Is the Interaction Between Vowel Perception and Vowel Production in Early Development?*) considered how the perception of vowels is related to vowel production in typical speech development, concluding that the two are closely connected. The perception of vowels can also be affected by communication disorders and may contribute to disorders of vowel production. Vowel perception has been reported to be affected in children with disorders such as hearing impairment (Hack & Erber, 1982), language impairment (Stark & Heinz, 1996), childhood apraxia of speech (CAS; Maassen et al., 2003), and reading disability (Bertucci et al., 2003).

Most of the relevant research has been on the effects of hearing impairment, but the results may generalize to other conditions. It has been shown that formant pattern and segment duration are affected by hearing impairment in infants (Kent et al., 1987; Rvachew et al., 1996) and children (Baudonck et al., 2011). Formant measurements of vowel production in those with hearing impairment reveal that atypical vowel production takes different patterns, with reports of both expanded and constricted vowel spaces. Whether vowel space is expanded or constricted in a given individual may be related to factors such as age of cochlear implant, duration of implant, previous experience with hearing aids, and the type and duration of speech therapy. Altered vowel space contributes to decreased perceptual speech intelligibility in healthy listeners, as reviewed in a previous section. An important concept emerging from research on hearing impairment is that the hearing loss affects the internal representation of a vowel (C. W. Turner & Henn, 1989). Internal representations may be ill-formed or subject to decay in other disorders, whether because of deficiencies in peripheral coding or difficulties with central acoustic-phonetic mapping (Hedrick et al., 2015). Manca and Grimaldi (2016) asserted that studies of auditory neurophysiology in both humans and animals point to two levels of sound coding, a tonotopy dimension for spectral properties and a tonochrony dimension for temporal properties. They further proposed that, in perception of a complex speech sound, the tonotopy and tonochrony data may reveal whether the speech sound parsing and decoding are accomplished solely by bottom-up reflection of acoustic differences or if they also are influenced by top-down processes related to phonological categories. Internal representations that combine tonotopic and tonochronic data may be a unifying concept to explain aspects of vowel perception in different disorders, such as CAS, language impairment, and reading disability. Pollock and Hall (1991) studied five children with CAS (ranging in age from 8;2 to 10;9 [years;



months) and reported that most children had difficulty perceiving tense/lax vowel contrasts (e.g., [i] for /i/ or [ɪ] for /ɪ/). Other common patterns included diphthong reduction (e.g., [a] for /ai/) and backing (e.g., [a] or [ɑ] for /æ/). Language impairment seems to impact perception of vowels, specifically identification of steady-state similar vowels, but not in discrimination of similar vowels (Stark & Heinz, 1996). As for reading disability, children who have lower phonological awareness and reading tend to perceive and produce less well-defined vowel categories than their typical counterparts; however, production and perception were not correlated (Bertucci et al., 2003).

A myriad of different types of impairment seem to impact vowel perception. Further study and clinical attention to both production and perception of vowels in a wide range of disorders, therefore, is warranted.

### ***How Can Sensory-Motor Mapping Serve in a Theoretical Framework for Treating Vowel Production and Perception?***

The evidence reviewed in this review article is generally consistent with contemporary models on sensory-motor mapping as a basis for speech production. The central concept is that children construct experience-driven maps that link sensory and motor properties of speech sounds, an idea that has been incorporated in several computational neural network models of speech production, including Directions Into Velocities of Articulators (Guenther, 1995, 2006; Guenther & Vladusich, 2012), State Feedback Control (Houde & Nagarajan, 2011), Hierarchical State Feedback Control (Hickok, 2012), and the neurocomputational model developed by Kröger et al. (2008). In one way or another, these models use feedback and feedforward signals that enable predictive and adaptive control over articulatory movements in speech. Shiller et al. (2010) took a similar perspective in discussing auditory and articulatory relationships in speech disorders, commenting that, in the model they advocate, "... the auditory-to-articulatory directional and articulator-to-auditory mappings correspond to an internal model that is learned early in life on the basis of accurate sensory feedback. It is posited that the internal model is acquired during babbling as the infant learns to relate articulator movements to their orosensory and acoustic consequences" (p. 182). Similarly, Davis and Redford (2019) propose a model in which perceptuo-motor units of production are established for whole words. The model illustrated in Figure 11 shares this general perspective.

The optimal conditions for speech development in children appear to be the availability of plurimodal sensory information along with ample opportunities for practicing and refining motor patterns of speech production. Linking of different sensory maps (auditory, somatosensory, and visual) in an interactive, complementary manner leads to a robust sensory foundation for flexible and adaptive motor control. Deprivation of or interference with any single sensory modality can, to some degree, be compensated by reliance on intact modalities, either in the short or long term.

However, doing so can compromise efficiency, as in the case of hyperarticulation in individuals with hearing impairment who use amplification or cochlear implants (Baudonck et al., 2011) and perhaps in some individuals receiving intervention for speech disorders.

### ***How Is the Clinical Assessment of Vowel Production Evolving?***

The milestones noted in Table 1 may be useful to guide clinical assessment and intervention, particularly for services that are designed in accord with normative schedules of development. However, timelines of development should be used with the recognition that large individual variations are possible for any particular behavior, and this caution certainly applies to the development of vowels.

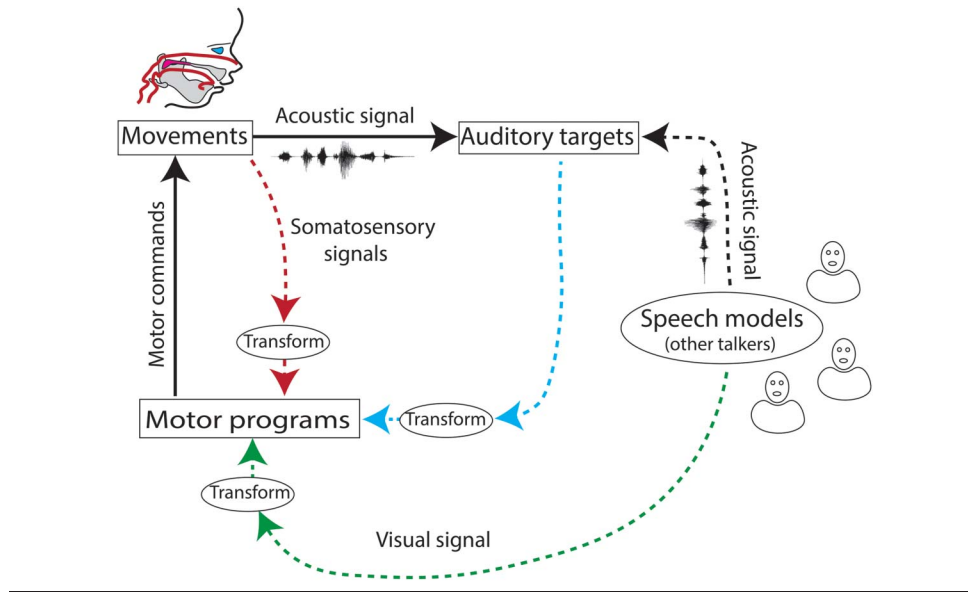
Several caveats should be noted concerning clinical assessment of vowels, as discussed in the following. These caveats carry implications for improvements in clinical practice.

(1) Commonly used tests of articulation often do not provide for comprehensive assessment of vowel production. Pollock (1991) remarked that none of the articulation tests she evaluated provide an adequate sample for analyzing vowel errors. Eisenberg and Hitchcock (2010) noted that only six of 11 standardized tests of speech articulation tested for vowels and that, of these, only two (Fisher-Logemann Test of Articulation, Fisher & Logemann, 1971; Templin-Darley Tests of Articulation, Templin & Darley, 1968) included a phonetically controlled word for all 15 vowels of American English. However, use of a single word does not provide an opportunity to evaluate vowel production in different phonetic environments. It can be concluded that standardized tests are not specifically designed to provide detailed information on the vowel repertoire of children's speech. The two major problems are limitations in phonetic context and unknown reliability in vowel judgments within and across raters.

Therefore, other methods are needed to obtain information on vowel production. One such method is the phonetic transcription of naturalistic speech samples, such as conversational speech or spontaneous utterances. Cox (2008) advocated for phonetic over phonemic transcription, noting that, "Phonetic rather than phonemic transcription of atypical speech is to be preferred because phonemic transcription must be based on familiarity with the phonology of the individual and such familiarity is not possible when describing atypical speech" (p. 5). However, the use of phonetic transcription confronts long-standing questions regarding the reliability of such transcriptions (Amorosa et al., 1985; Cox, 2008; Munson et al., 2012; Oller & Eilers, 1975; Sell & Sweeney, 2020; Shriberg & Lof, 1991; Stoel-Gammon, 2001; Stockman et al., 1981). A combination of acoustic and consensus analysis may be a step forward (Amorosa et al., 1985; Shriberg et al., 2010), but much remains to be done to standardize procedures, establish and maintain requisite analysis skills, and demonstrate that such an approach can



**Figure 11.** Sensory-motor relationships in the development of speech. Motor programs for speech production are associated with auditory (blue line), somatosensory (red line), and visual (green line) information. In addition to the various forms of feedback from self-produced speech, a child also receives acoustic and visual information from other talkers. Typical vowel development therefore proceeds with plurimodal sensory information that is integrated in sensory-motor maps.



be accomplished within the time constraints of clinical services. The question of reliability often resurfaces in the remaining points of discussion.

(2) Compared to consonants, vowels are more likely to be perceived in a continuous rather than categorical mode (Harnad, 1987). Expressed in the strongest terms, this perceptual difference may be taken to mean that consonants, especially stop consonants, are perceived categorically, whereas vowels are perceived continuously. A more conservative statement is that consonants are associated with strong category effects and vowels are associated with weaker category effects (Kronrod et al., 2016). The difference has been explained in various ways. For example, Pisoni (1973) remarked that differences between the discriminability of consonants and vowels may be explained by the assumption that auditory short-term memory for consonants is not maintained as well as that for vowels. If this interpretation is correct, then it might be expected that phonetic transcription for vowels would be more accurate than for consonants. However, in a study of transcription reliability, Shriberg et al. (2010) found a higher level of agreement for vowels than consonants in broad transcription, but the converse for narrow transcription. Howard and Heselwood (2013) surmised that the narrow transcription of vowels may be avoided in part because it is assumed that vowel impairments are uncommon. In addition, as noted by Knight et al. (2018) in their survey of speech-language therapists in the United Kingdom, many clinicians lack confidence in the use of narrow transcription.

(3) Vowel mastery, even in typically developing children, is associated with large individual differences, especially

for diphthongs and rhotics (Pollock, 2002). Therefore, general patterns of development may not always be valid guidelines in the clinical assessment of vowel production. Normative developmental data are particularly lacking for toddlers (DeVeney, 2019), which limits clinical assessment of the birth-to-3 population. Profiles of vowel acquisition, such as that shown in Figure 3, may be helpful in characterizing general patterns but do not necessarily apply to individual children.

(4) Vowels can be strongly influenced by dialect, even in young children, who are influenced by the speech patterns of their speech community. Narrow phonetic transcription, although often challenging, may provide useful information. Heselwood and Howard (2008) make this recommendation: “In principle, a transcription should aim to balance segmental and nonsegmental representations. It should identify, as far as is possible, rhythm group and intonation-group boundaries, speech rate, pauses, and long-domain resonance and voice quality features as well as details about phonation, and articulation” (p. 391). Recommendations for phonetic transcriptions were also made by Pollock and Berni (2001) for vowels and by Teoh and Chin (2009) for the speech of those with hearing impairment. Although a relatively gross measure, such as percentage of vowels correct, can be helpful in distinguishing groups of children with speech disorders (Wren et al., 2013), more detailed descriptions may lead to refinements in planning treatment.

(5) Consonant–vowel interaction is an important consideration in understanding vowel development and vowel disorders. Gierut et al. (1993) concluded from geometric phonological analyses that consonants and vowels are fully

integrated in the earliest stages of development with the place specification of consonants deriving from the vowel. Bates and Watson (1995) also remarked this issue, drawing attention to three types of consonant–vowel interactions that are relevant to clinical assessment and treatment: (a) vowel conditioning of consonant production, (b) consonant conditioning of vowel production, and (c) use of consonantal material to maintain vowel contrasts. They noted that the context conditioning in both consonant and vowel error patterns underscores the need to assess a child’s sound system as a whole.

(6) Traditional phonetic and phonemic transcriptions assume that vowels are isolable static particles, but this assumption runs counter to recent evidence that vowels are intrinsically dynamic, that vowel transcriptions are challenging even to experts, and that early stages of speech development may be based on other units (e.g., whole words; Davis & Redford, 2019). This is not to dismiss the value of conventional vowel descriptions but rather to place a note of caution on the interpretation of data so obtained. Identification of a sound that resembles a vowel in the adult phonetic system is not necessarily an indication that the vowel in question has been acquired as a distinct phoneme in an emerging repertoire.

(7) As noted by Munson et al. (2012), the development of speech sounds is not necessarily categorical but may involve subtle transitions. Before children produce a contrast between two given phonemes, they may go through an intervening stage in which they produce a covert contrast, that is, a subphonemic difference that can be measured acoustically but is not salient enough to warrant identification with a different phonetic symbol. Covert contrasts are difficult if not impossible to discover with typical methods of transcription, which are insensitive to small differences in production. These contrasts have been identified almost exclusively through acoustic and physiological methods (Gibbon & Lee, 2017a; Glaspey & MacLeod, 2010; Macken & Barton, 1980; McAllister et al., 2016; Scobbie et al., 2000; see also the special issue of *Clinical Linguistics & Phonetics*, Gibbon & Lee, 2017b). Further studies are needed to show their frequency and developmental patterns. Gibbon and Lee (2017a) remarked that, “...recent studies have provided convincing new evidence that covert contrasts are likely to be widespread in child speech” (p. 4). These contrasts, which are discoverable through the use of instrumental techniques, may be ubiquitous as developmental phenomena.

It may be possible to overcome some of the problems in transcription by using alternative or complementary methods such as crowdsourcing (Williams et al., 2011), automatic speech recognition (Kabir et al., 2010), or weighted measures of accuracy of speech sounds (Preston et al., 2011). We certainly do not recommend that transcription be abandoned as a clinical tool but rather used with the recognition that it may provide only a superficial rendering of the actual speech behavior. It may not be efficient to perform a detailed acoustic analysis of every utterance, but a selective analysis that targets a selected feature or dimension (e.g., vowel length, F1 frequency) may be practicable.

## Conclusions

It can be concluded from the studies reviewed in this review article that the maturation of vowel production has three major aspects that can be addressed in research and clinical application. The first is the capability to produce articulatory configurations that yield satisfactory acoustic output corresponding to the phonetic characteristics of the ambient language. Developmental status in this aspect is commonly assessed with articulation tests or phonetic transcription, but acoustic and physiological methods can complement or clarify perceptual judgments. As noted previously, perceptual judgments of vowels may lack precision. Formant frequencies are the most commonly used acoustic measurements, but procedures are not yet standardized to be used in general clinical populations (Kent & Vorperian, 2018).

The second aspect is the capability to regulate vowel durations in accord with both intrinsic (e.g., tense vs. lax distinction) and extrinsic (e.g., postvocalic consonant voicing) properties of the parent language. To some extent, this capability can be assessed perceptually, but acoustic measures are needed to ensure accurate determination. Standardization of acoustic assessment should be feasible given that the required measurements are relatively straightforward and can be done with freely available software such as Praat (Boersma & Weenink, 2018) and with agreement on procedures, such as test words used in the assessment.

The third aspect is the capability to deploy vowel sounds to accomplish the requirements of rhythm and prosody. A similar distinction was drawn by James et al. (2001), who described the development of vowels as involving paradigmatic and syntagmatic processes. In a paradigmatic process, children learn to produce a vowel in isolation or in simple monosyllabic words (the usual method of standard articulation tests). A syntagmatic process is the ability to produce sequences of vowels in syllables and words, with regard to suprasegmental features such as stress. Assessments in this domain can be accomplished by perceptual and acoustic methods, but there is relatively little standardization to ensure the comparability of data across ages or settings. For a helpful review of indices of rhythm in developing speech, see Payne et al. (2012). An example of an acoustic index of speech rhythm applicable to developing speech is the vocalic normalized pairwise variability index (Grabe et al., 1999), which examines the variation in duration of successive vocalic intervals. Notable progress in the automatic processing of the acoustic speech signal with applications to speech disorders (Barbosa et al., 2018) holds the promise of easily implemented analysis tools that will complement perceptual assessments.

Advancing technology and programming will make automatic extraction of acoustic measures, such as those mentioned here, more accessible to researchers and clinicians over time. Combined with the proliferation of high-definition audio recording equipment, lab and personal devices with high-powered processors increased accessibility to devices, and even mobile apps for acoustic measurement

and assessment are on the horizon. Opportunities are arising for robust measurement in different environments, with more functional tasks and therefore with increasing ecological validity.

When considering mobile assessment and the challenges and opportunities of the natural environment, the robust vowel signal may emerge as a primary target for assessment of acoustic correlates to intelligibility and other speech functions. In a study showing the variability of intelligibility even when background noise is controlled, researchers found that, in natural background noise, the identity of vowels is the most preserved (Meyer et al., 2013), strengthening our argument for measuring acoustic properties of vowels in functional speech assessments. As assessment moves to more spontaneous speech and natural environments, clinical procedures will likely incorporate advancements in automatic detection of vowels from the speech signal.

The overarching conclusion of this review is that research in two general areas, the development of vowels in children's speech and vowel disorders in children and adults, prompts a fresh look at this class of speech sounds. Acoustic studies have revealed properties of vowels that should be incorporated in accounts of speech development and communication disorders in children and adults. The combination of acoustic and perceptual methods makes it possible to address both reliability and validity in the assessment of vowel production. Vowels in American English are best regarded as having intrinsic dynamic properties and as giving coherence to multisyllabic utterances through their role in expressing rhythm, stress, and paralinguistic aspects of speech. Although phonetic mastery of vowels is accomplished relatively early using conventional perceptual assessments of articulation (e.g., broad phonetic transcription), acoustic studies show that both perceptual and productive properties of vowels develop gradually until at least 6–8 years, perhaps in all languages. The production of vowels is adapted to factors such as communication task, presence of perturbations, and dialect. The importance of vowel disorders is increasingly recognized, along with the need for more sensitive assessments of vowel production, effective programs for treatment of vowel disorders, and a new look at the contribution of vowels to functional speech assessment across the life span and disorders of communication. Movement toward quantifiable acoustic measurement of speech serves to go hand in hand with perceptual clinical measurement for increasingly comprehensive descriptions of function.

## Acknowledgments

This work was supported by National Institute on Deafness and Other Communicative Disorders Grant R01 DC6282 (MRI and CT Studies of the Developing Vocal Tract, Hourii K. Vorperian, Principal Investigator) and National Institute of Child Health and Human Development Grants P30 HD03352 and U54 HD090256 to the Waisman Center. The Department of Communication Sciences and Disorders at the University of Wisconsin–Madison and the Department of Communication Sciences and Disorders at the University of Cincinnati maintained faculty support for one author.

## References

- Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107. <https://doi.org/10.1121/1.1795335>
- Alhaidary, A., & Rvachew, S. (2018). Cross-linguistic differences in the size of the infant vowel space. *Journal of Phonetics*, *71*, 16–34. <https://doi.org/10.1016/j.wocn.2018.07.003>
- Altvater-Mackensen, N., Mani, N., & Grossmann, T. (2016). Audio-visual speech perception in infancy: The influence of vowel identity and infants' productive abilities on sensitivity to (mis)matches between auditory and visual speech cues. *Developmental Psychology*, *52*(2), 191–204. <https://doi.org/10.1037/a0039964>
- Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the Acoustical Society of America*, *101*(2), 1078–1089. <https://doi.org/10.1121/1.417972>
- Amorosa, H., Von Benda, U., Wagner, E., & Keck, A. (1985). Transcribing phonetic detail in the speech of unintelligible children: A comparison of procedures. *International Journal of Language & Communication Disorders*, *20*(3), 281–287. <https://doi.org/10.3109/13682828509012268>
- Angelocci, A. A., Kopp, G. A., & Holbrook, A. (1964). The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys. *Journal of Speech and Hearing Disorders*, *29*(2), 156–170. <https://doi.org/10.1044/jshd.2902.156>
- Archila-Meléndez, M. E., Valente, G., Correia, J. M., Rouhl, R. P., van Kranen-Mastenbroek, V. H., & Jansma, B. M. (2018). Sensorimotor representation of speech perception. Cross-decoding of place of articulation features during selective attention to syllables in 7T fMRI. *eNeuro*, *5*(2). <https://doi.org/10.1523/ENEURO.0252-17.2018>
- Assmann, P. F., & Katz, W. F. (2000). Time-varying spectral change in the vowels of children and adults. *The Journal of the Acoustical Society of America*, *108*(4), 1856–1866. <https://doi.org/10.1121/1.1289363>
- Assmann, P. F., Nearey, T. M., & Bharadwaj, S. V. (2013). Developmental patterns in children's speech: Patterns of spectral change in vowels. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel inherent spectral change* (pp. 199–230). Springer. [https://doi.org/10.1007/978-3-642-14209-3\\_9](https://doi.org/10.1007/978-3-642-14209-3_9)
- Auszmann, A., & Neuberger, T. (2014). Age- and gender-related differences in formant structure during the stabilization process of vowels. In J. Emonds & M. Janebová (Eds.), *Proceedings of the Olomouc Linguistics Colloquium 2014. Olomouc Modern Language Series* (Vol. 5, pp. 663–676). Palacký University.
- Bahr, D., & Rosenfeld-Johnson, S. (2010). Treatment of children with speech oral placement disorders (OPDs): A paradigm emerges. *Communication Disorders Quarterly*, *31*(3), 131–138. <https://doi.org/10.1177/1525740109350217>
- Ball, M. J., & Gibbon, F. E. (Eds.). (2013). *Handbook of vowels and vowel disorders*. Psychology Press. <https://doi.org/10.4324/9780203103890>
- Bang, Y., Min, K., Sohn, Y. H., & Cho, S. R. (2013). Acoustic characteristics of vowel sounds in patients with Parkinson disease. *NeuroRehabilitation*, *32*(3), 649–654. <https://doi.org/10.3233/NRE-130887>
- Barbosa, P. A., Camargo, Z. A., & Madureira, S. (2018). Acoustic-based tools and scripts for the automatic analysis of speech in clinical and non-clinical settings. In H. A. Patil, A. Neustein, & M. Kulshreshtha (Eds.), *Signal and acoustic modeling for speech and communication disorders* (Vol. 5, pp. 69–86). de Gruyter. <https://doi.org/10.1515/9781501502415-004>



- Bates, S., & Watson, J.** (1995). Consonant–vowel interactions in developmental phonological disorder. *International Journal of Language & Communication Disorders, 30*(S1), 274–279. <https://doi.org/10.1111/j.1460-6984.1995.tb01687.x>
- Baudonck, N., Van Lierde, K., Dhooge, I., & Corthals, P.** (2011). A comparison of vowel productions in prelingually deaf children using cochlear implants, severe hearing-impaired children using conventional hearing aids and normal-hearing children. *Folia Phoniatrica et Logopaedica, 63*(3), 154–160. <https://doi.org/10.1159/000318879>
- Baum, S. R., & Katz, W. F.** (1988). Acoustic analysis of compensatory articulation in children. *The Journal of the Acoustical Society of America, 84*(5), 1662–1668. <https://doi.org/10.1121/1.397181>
- Bele, I. V.** (2006). The speaker's formant. *Journal of Voice, 20*(4), 555–578. <https://doi.org/10.1016/j.jvoice.2005.07.001>
- Berger, T., Peschel, T., Vogel, M., Pietzner, D., Poulain, T., Jurkutat, A., Meuret, S., Engel, C., Kiess, W., & Fuchs, M.** (2019). Speaking voice in children and adolescents: Normative data and associations with BMI, Tanner stage, and singing activity. *Journal of Voice, 33*(4), 580.e21–580.e30. <https://doi.org/10.1016/j.jvoice.2018.01.006>
- Bertucci, C., Hook, P., Haynes, C., Macaruso, P., & Bickley, C.** (2003). Vowel perception and production in adolescents with reading disabilities. *Annals of Dyslexia, 53*(1), 174–200. <https://doi.org/10.1007/s11881-003-0009-1>
- Bladon, A.** (1983). Two-formant models of vowel perception: Shortcomings and enhancements. *Speech Communication, 2*(4), 305–313. [https://doi.org/10.1016/0167-6393\(83\)90047-X](https://doi.org/10.1016/0167-6393(83)90047-X)
- Blomgren, M., Robb, M., & Chen, Y.** (1998). A note on vowel centralization in stuttering and nonstuttering individuals. *Journal of Speech, Language, and Hearing Research, 41*(5), 1042–1051. <https://doi.org/10.1044/jshr.4105.1042>
- Boersma, P., & Weenink, D.** (2018). *Praat: Doing phonetics by computer* (Version 6.0.37) [Computer software]. <https://www.praat.org>
- Bond, Z. S., Petrosino, L., & Dean, C. R.** (1982). The emergence of vowels: 17 to 26 months. *Journal of Phonetics, 10*(4), 417–422. [https://doi.org/10.1016/S0095-4470\(19\)31005-8](https://doi.org/10.1016/S0095-4470(19)31005-8)
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B.** (1996). Intelligibility of normal speech. I: Global and fine-grained acoustic–phonetic talker characteristics. *Speech Communication, 20*(3–4), 255–272. [https://doi.org/10.1016/S0167-6393\(96\)00063-5](https://doi.org/10.1016/S0167-6393(96)00063-5)
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F.** (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences, 112*(44), 13531–13536. <https://doi.org/10.1073/pnas.1508631112>
- Buhr, R. D.** (1980). The emergence of vowels in an infant. *Journal of Speech and Hearing Research, 23*(1), 73–94. <https://doi.org/10.1044/jshr.2301.73>
- Bunton, K., & Hoit, J. D.** (2018). Development of velopharyngeal closure for vocalization during the first 2 years of life. *Journal of Speech, Language, and Hearing Research, 61*(3), 549–560. [https://doi.org/10.1044/2017\\_JSLHR-S-17-0208](https://doi.org/10.1044/2017_JSLHR-S-17-0208)
- Bunton, K., & Leddy, M.** (2011). An evaluation of articulatory working space in vowel production of adults with Down syndrome. *Clinical Linguistics & Phonetics, 25*(4), 321–334. <https://doi.org/10.3109/02699206.2010.535647>
- Buzaneli, E. C. P., Zenari, M. S., Kulcsar, M. A. V., Dedivitis, R. A., Cernea, C. R., & Nembr, K.** (2018). Supracricoid laryngectomy: The function of the remaining arytenoid in voice and swallowing. *International Archives of Otorhinolaryngology, 22*(03), 303–312. <https://doi.org/10.1055/s-0038-1625980>
- Bybee, J., & Beckner, C.** (2010). Usage-based theory. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199544004.013.0032>
- Cartei, V., Cowles, H. W., Banerjee, R., & Reby, D.** (2014). Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology, 32*(1), 100–106. <https://doi.org/10.1111/bjdp.12027>
- Cartei, V., Cowles, H. W., & Reby, D.** (2012). Spontaneous voice gender imitation abilities in adult speakers. *PLOS ONE, 7*, Article e31353. <https://doi.org/10.1371/journal.pone.0031353>
- Cartei, V., & Reby, D.** (2013). Effect of formant frequency spacing on perceived gender in pre-pubertal children's voices. *PLOS ONE, 8*(12), Article e81022. <https://doi.org/10.1371/journal.pone.0081022>
- Chen, F., & Hu, Y.** (2019). Segmental contributions to cochlear implant speech perception. *Speech Communication, 106*, 79–84. <https://doi.org/10.1016/j.specom.2018.12.001>
- Chen, F., Wong, L. L., & Wong, E. Y.** (2013). Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility. *The Journal of the Acoustical Society of America, 134*(2), EL178–EL184. <https://doi.org/10.1121/1.4812820>
- Chen, L. M., & Kent, R. D.** (2005). Consonant–vowel co-occurrence patterns in Mandarin-learning infants. *Journal of Child Language, 32*(3), 507–534. <https://doi.org/10.1017/S0305000905006896>
- Chen, L. M., & Kent, R. D.** (2010). Segmental production in Mandarin-learning infants. *Journal of Child Language, 37*(2), 341–371. <https://doi.org/10.1017/S0305000909009581>
- Childers, D. G., & Wu, K.** (1991). Gender recognition from speech: Part II. Fine analysis. *The Journal of the Acoustical Society of America, 90*(4), 1841–1856. <https://doi.org/10.1121/1.401664>
- Choi, J., Cutler, A., & Broersma, M.** (2017). Early development of abstract language knowledge: Evidence from perception–production transfer of birth-language memory. *Royal Society Open Science, 4*(1), 160660. <https://doi.org/10.1098/rsos.160660>
- Chung, H., Kong, E. J., Edwards, J., Weismer, G., Fourakis, M., & Hwang, Y.** (2012). Cross-linguistic studies of children's and adults' vowel spaces. *The Journal of the Acoustical Society of America, 131*(1), 442–454. <https://doi.org/10.1121/1.3651823>
- Chung, H., & Pollock, K. E.** (2019). Acoustic characteristics of rhotic vowel productions of young children. *Folia Phoniatrica et Logopaedica, 1*–12. <https://doi.org/10.1159/000504250>
- Ciocca, V., & Whitehill, T. L.** (2013). The acoustic measurement of vowels. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 113–137). Psychology Press.
- Clopper, C. G., Pisoni, D. B., & De Jong, K.** (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America, 118*(3), 1661–1676. <https://doi.org/10.1121/1.2000774>
- Coen, M. H., Vorperian, H. K., & Kent, R. D.** (2015). High fidelity analysis of vowel acoustic space. *The Journal of the Acoustical Society of America, 137*(4), 2305. <https://doi.org/10.1121/1.4920418>
- Cole, J., Linebaugh, G., Munson, C., & McMurray, B.** (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics, 38*(2), 167–184. <https://doi.org/10.1016/j.wocn.2009.08.004>
- Cole, R. A., Yan, Y., Mak, B., Fanty, M., & Bailey, T.** (1996, May). The contribution of consonants versus vowels to word recognition in fluent speech. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings* (Vol. 7, pp. 853–856). Institute of Electrical and Electronics Engineers.

- Cox, F. (2008). Vowel transcription systems: An Australian perspective. *International Journal of Speech-Language Pathology*, 10(5), 327–333. <https://doi.org/10.1080/17549500701855133>
- Crary, M. A. (1995, May). *Seminars in Speech and Language*, 16(02), 110–125. Thieme. <https://doi.org/10.1055/s-2008-1064114>
- Davis, B. L., & MacNeilage, P. F. (1990). Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing Research*, 33(1), 16–27. <https://doi.org/10.1044/jshr.3301.16>
- Davis, M., & Redford, M. A. (2019). The emergence of discrete perceptual-motor units in a production model that assumes holistic phonological representations. *Frontiers in Psychology*, 10, 2121. <https://doi.org/10.3389/fpsyg.2019.02121>
- de Boer, B., & Fitch, W. T. (2010). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior*, 18(1), 36–48. <https://doi.org/10.1177/1059712309350972>
- de Boysson-Bardies, B., Hallé, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16(1), 1–17. <https://doi.org/10.1017/S0305000900013404>
- de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2), 297–319. <https://doi.org/10.1353/lan.1991.0045>
- de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2), 297–319. <https://doi.org/10.1353/lan.1991.0045>
- de Bruijn, M. J., ten Bosch, L., Kuik, D. J., Quené, H., Langendijk, J. A., Leemans, C. R., & Verdonck-de Leeuw, I. M. (2009). Objective acoustic–phonetic speech analysis in patients treated for oropharyngeal cancer. *Folia Phoniatrica et Logopaedica*, 61, 180–187. <https://doi.org/10.1159/000219953>
- de Jarnette, G. (1988). Formant frequencies ( $F_1$ ,  $F_2$ ) of jaw-free versus jaw-fixed vowels in normal and articulatory disordered children. *Perceptual and Motor Skills*, 67(3), 963–971. <http://doi.org/10.2466/pms.1988.67.3.963>
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, 9(2), 133–150. [https://doi.org/10.1016/0163-6383\(86\)90025-1](https://doi.org/10.1016/0163-6383(86)90025-1)
- DeVeney, S. L. (2019, March). Clinical challenges: Assessing toddler speech sound productions. *Seminars in Speech and Language*, 40(02), 81–93. Thieme. <https://doi.org/10.1055/s-0039-1677759>
- Disner, S. F. (1980). Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America*, 67, 253–261. <https://doi.org/10.1121/1.383734>
- Donegan, P. (2013). Normal vowel development. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 24–60). Psychology Press.
- Dworkin, J. P. (1978). A therapeutic technique for the improvement of lingua-alveolar valving abilities. *Language, Speech, and Hearing Services in Schools*, 9(3), 169–175. <https://doi.org/10.1044/0161-1461.0903.169>
- Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngologica Supplementum*, 257, 1–51.
- Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *The Journal of the Acoustical Society of America*, 75(4), 1213–1218. <https://doi.org/10.1121/1.390773>
- Eisenberg, S. L., & Hitchcock, E. R. (2010). Using standardized tests to inventory consonant and vowel production: A comparison of 11 tests of articulation and phonology. *Language, Speech, and Hearing Services in Schools*, 41, 488–503. [https://doi.org/10.1044/0161-1461\(2009/08-0125\)](https://doi.org/10.1044/0161-1461(2009/08-0125))
- Engstrand, O., Williams, K., & Lacerda, F. (2003). Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12- and 18-month-olds. *Phonetica*, 60(1), 17–44. <https://doi.org/10.1159/000070452>
- Ertem, I. O., Krishnamurthy, V., Mulaudzi, M. C., Sguassero, Y., Balta, H., Gulumser, O., Bilik, B., Srinivasan, R., Johnson, B., Gan, G., Calvoceossi, L., Shabanova, V., & Forsyth, B. (2018). Similarities and differences in child development from birth to age 3 years by sex and across four countries: A cross-sectional, observational study. *The Lancet Global Health*, 6(3), e279–e291. [https://doi.org/10.1016/S2214-109X\(18\)30003-2](https://doi.org/10.1016/S2214-109X(18)30003-2)
- Fant, G. (1970). *Acoustic theory of speech production: With calculations based on X-ray studies of Russian articulations (No. 2)*. de Gruyter. <https://doi.org/10.1515/9783110873429>
- Fifer, W. (1987). Neonatal preference for mother's voice. In N. A. Krasnagor (Ed.), *Perinatal development: A psychobiological perspective* (pp. 111–124). Academic Press.
- Fisher, H. B., & Logemann, J. A. (1971). *Fisher–Logemann Test of Articulation Competence*. Houghton Mifflin.
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522. <https://doi.org/10.1121/1.427148>
- Flipsen, P., Jr., & Lee, S. (2012). Reference data for the American English acoustic vowel space. *Clinical Linguistics & Phonetics*, 26(11–12), 926–933. <https://doi.org/10.3109/02699206.2012.720634>
- Flipsen, P., Jr., Shriberg, L. D., Weismer, G., Karlsson, H. B., & McSweeney, J. L. (2001). Acoustic phenotypes for speech-genetics studies: Reference data for residual /ə/ distortions. *Clinical Linguistics & Phonetics*, 15, 603–630. <https://doi.org/10.1080/02699200110069410>
- Fogerty, D., & Kewley-Port, D. (2009). Perceptual contributions of the consonant–vowel boundary to sentence intelligibility. *The Journal of the Acoustical Society of America*, 126(2), 847–857. <https://doi.org/10.1121/1.3159302>
- Fox, R. A. (1982). Individual variation in the perception of vowels: Implications for a perception–production link. *Phonetica*, 39(1), 1–22. <https://doi.org/10.1159/000261647>
- Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America*, 126(5), 2603–2618. <https://doi.org/10.1121/1.3212921>
- Gay, T., Lindblom, B., & Lubker, J. (1981). Production of bite-block vowels: Acoustic equivalence by selective compensation. *The Journal of the Acoustical Society of America*, 69(3), 802–810. <https://doi.org/10.1121/1.385591>
- Gerosa, M., Giuliani, D., & Brugnara, F. (2007). Acoustic variability and automatic recognition of children's speech. *Speech Communication*, 49(10–11), 847–860. <https://doi.org/10.1016/j.specom.2007.01.002>
- Gibbon, F. E., & Lee, A. (2017a). Electropalatographic (EPG) evidence of covert contrasts in disordered speech. *Clinical Linguistics & Phonetics*, 31(1), 4–20. <https://doi.org/10.1080/02699206.2016.1174739>
- Gibbon, F. E., & Lee, A. (Eds.). (2017b). Covert contrasts [Special issue]. *Clinical Linguistics & Phonetics*, 31(1).
- Gibson, A., & McPhearson, L. (1979/1980). Production of bite-block vowels by children. *Phonetic Experimental Research at the Institute of Linguistics (PERILUS)* (Report II, pp. 26–41). University of Stockholm.
- Gick, B., Allen, B., Roewer-Després, F., & Stavness, I. (2017). Speaking tongues are actively braced. *Journal of Speech,*



- Language, and Hearing Research*, 60(3), 494–506. [https://doi.org/10.1044/2016\\_JSLHR-S-15-0141](https://doi.org/10.1044/2016_JSLHR-S-15-0141)
- Gierut, J. A., Cho, M.-H., & Dinnsen, D. A.** (1993). Geometric accounts of consonant–vowel interactions in developing systems. *Clinical Linguistics & Phonetics*, 7(3), 219–236. <https://doi.org/10.3109/02699209308985559>
- Gilbert, H. R., Robb, M. P., & Chen, Y.** (1997). Formant frequency development: 15 to 36 months. *Journal of Voice*, 11(3), 260–266. [https://doi.org/10.1016/S0892-1997\(97\)80003-3](https://doi.org/10.1016/S0892-1997(97)80003-3)
- Glaspey, A. M., & MacLeod, A. A.** (2010). A multi-dimensional approach to gradient change in phonological acquisition: A case study of disordered speech development. *Clinical Linguistics & Phonetics*, 24(4–5), 283–299. <https://doi.org/10.3109/02699200903581091>
- Grabe, E., Watson, I., & Post, B.** (1999). The acquisition of rhythmic patterns in English and French. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1201–1204). University of California.
- Gramley, V.** (2010). *Acoustic phonetics*. Retrieved May 16, 2019 from [http://www.uni-bielefeld.de/lili/personen/vgramley/teaching/HTHS/acoustic\\_2010.html](http://www.uni-bielefeld.de/lili/personen/vgramley/teaching/HTHS/acoustic_2010.html)
- Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S.** (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, 14(2), 336–352. <https://doi.org/10.1111/j.1467-7687.2010.00978.x>
- Grenon, I., Benner, A., & Esling, J. H.** (2007). Language-specific phonetic production patterns in the first year of life. In J. Trouvain (Ed.), *Proceedings of the 16th International Congress of Phonetic Sciences* (Vol. 3, pp. 1561–1564). Universität des Saarlandes.
- Groome, L. J., Mooney, D. M., Holland, S. B., Bentz, L. S., Atterbury, J. L., & Dykman, R. A.** (1997). The heart rate deceleratory response in low-risk human fetuses: Effect of stimulus intensity on response topography. *Developmental Psychobiology*, 30(2), 103–113. [https://doi.org/10.1002/\(SICI\)1098-2302\(199703\)30:2<103::AID-DEV2>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1098-2302(199703)30:2<103::AID-DEV2>3.0.CO;2-U)
- Guenther, F. H.** (1995). Speech sound acquisition, coarticulation, and rate effects in a neural-network model of speech production. *Psychological Review*, 102(3), 594–621. <https://doi.org/10.1037/0033-295X.102.3.594>
- Guenther, F. H.** (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39(5), 350–365. <https://doi.org/10.1016/j.jcomdis.2006.06.013>
- Guenther, F. H., & Vladusich, T.** (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25(5), 408–422. <https://doi.org/10.1016/j.jneuroling.2009.08.006>
- Hack, Z. C., & Erber, N. P.** (1982). Auditory, visual, and auditory-visual perception of vowels by hearing-impaired children. *Journal of Speech and Hearing Research*, 25(1), 100–107. <https://doi.org/10.1044/jshr.2501.100>
- Hagiwara, R.** (1995). *Acoustic realizations of American /r/ as produced by women and men* [Unpublished doctoral dissertation]. University of California at Los Angeles.
- Hamdan, A. L., Khandakji, M., & Macari, A. T.** (2018). Maxillary arch dimensions associated with acoustic parameters in prepubertal children. *The Angle Orthodontist*, 88(4), 410–415. <https://doi.org/10.2319/111617-792.1>
- Harnad, S.** (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In *Categorical perception: The groundwork of cognition* (pp. 1–52). Cambridge University Press.
- Hedrick, M., Charles, L., & Street, N. D.** (2015). Vowel perception in listeners with normal hearing and in listeners with hearing loss: A preliminary study. *Clinical and Experimental Otorhinolaryngology*, 8(1), 26–33. <https://doi.org/10.3342/ceo.2015.8.1.26>
- Heselwood, B., & Howard, S.** (2008). Clinical phonetic transcription. In M. J. Ball, M. R. Perkins, N. Müller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 381–399). Blackwell. <https://doi.org/10.1002/9781444301007.ch23>
- Heselwood, B., & Plug, L.** (2011). The role of F2 and F3 in the perception of rhoticity: Evidence from listening experiments. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII): August 17–21, 2011* (pp. 867–870). City University of Hong Kong.
- Hickok, G.** (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6), 393–402. <https://doi.org/10.1016/j.jcomdis.2012.06.004>
- Higgins, C. M., & Hodge, M. M.** (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology*, 10, 271–277.
- Hillenbrand, J., & Gayvert, R. T.** (1993). Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech and Hearing Research*, 36, 694–700. <https://doi.org/10.1044/jshr.3604.694>
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K.** (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97, 3099–3111. <https://doi.org/10.1121/1.411872>
- Hirsch, F., Bouarouf, F., Vaxelaire, B., Monfrais-Pfauwadel, M. C., Bechet, M., Sturm, J., & Sock, R.** (2008, December). Formant structures of vowels produced by stutterers in normal and fast speech rates. In R. Sock, S. Fuchs, & Y. Laprie (Eds.), *Proceedings of ISSP 2008—8th International Seminar on Speech Production (Strasbourg, France)*. (p. NC). INRIA.
- Hochmann, J. R., Benavides-Varela, S., Nespore, M., & Mehler, J.** (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science*, 14(6), 1445–1458. <https://doi.org/10.1111/j.1467-7687.2011.01089.x>
- Hodge, M. M.** (2013). Development of the vowel space in children: Anatomic and acoustic aspects. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 1–23). Psychology Press.
- Hoole, P.** (1987, August). Bite-block speech in the absence of oral sensibility. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIth International Conference of Phonetic Sciences* (Vol. 4, pp. 16–19). KTH and Stockholm University.
- Houde, J. F., & Nagarajan, S. S.** (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5, 82. <https://doi.org/10.3389/fnhum.2011.00082>
- Howard, S. J., & Heselwood, B.** (2013). The contribution of phonetics to the study of vowel development and disorders. In M. Ball & F. Gibbon (Eds.), *Vowel disorders* (pp. 79–130). Butterworth–Heinemann.
- Hustad, K. C., Gorton, K., & Lee, J.** (2010). 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*, 53(6), 1496–1513. [https://doi.org/10.1044/1092-4388\(2010\)09-0176](https://doi.org/10.1044/1092-4388(2010)09-0176)
- Irwin, J. V., & Wong, S. P.** (1983). *Phonological development in children: 18–72 months*. Southern Illinois University Press.
- Irwin, O. C.** (1948). Infant speech: Development of vowel sounds. *Journal of Speech and Hearing Disorders*, 13(1), 31–34. <https://doi.org/10.1044/jshd.1301.31>
- Ishizuka, K., Mugitani, R., Kato, H., & Amano, S.** (2007). Longitudinal developmental changes in spectral peaks of vowels produced by Japanese infants. *The Journal of the Acoustical*

- Society of America*, 121(4), 2272–2282. <https://doi.org/10.1121/1.2535806>
- Jacewicz, E., Fox, R. A., & Salmons, J.** (2011). Regional dialect variation in the vowel systems of typically developing children. *Journal of Speech, Language, and Hearing Research*, 54(2), 448–470. [https://doi.org/10.1044/1092-4388\(2010\)10-0161](https://doi.org/10.1044/1092-4388(2010)10-0161)
- James, D., van Doorn, J., & McLeod, S.** (2001). Vowel production in mono-, di- and poly-syllabic words in children 3;0 to 7;11 years. In L. Wilson & S. Hewat (Eds.), *Proceedings of the Speech Pathology Conference* (pp. 127–136). Speech Pathology Australia.
- Kabir, A., Giurgiu, M., & Barker, J.** (2010, June). Robust automatic transcription of English speech corpora. In *2010 8th International Conference on Communications* (pp. 79–82). Institute of Electrical and Electronics Engineers. Curran Associates. <https://doi.org/10.1109/ICCOMM.2010.5509116>
- Kaipa, R., Robb, M. P., O’Beirne, G. A., & Allison, R. S.** (2012). Recovery of speech following total glossectomy: An acoustic and perceptual appraisal. *International Journal of Speech-Language Pathology*, 14(1), 24–34. <https://doi.org/10.3109/17549507.2011.623326>
- Kalashnikova, M., Carignan, C., & Burnham, D. K.** (2017). The origins of babytalk: Smiling, teaching or social convergence? *Royal Society Open Science*, 4(8). <https://doi.org/10.1098/rsos.170306>
- Karlsson, F., & van Doorn, J.** (2012). Vowel formant dispersion as a measure of articulation proficiency. *The Journal of the Acoustical Society of America*, 132(4), 2633–2641. <https://doi.org/10.1121/1.4746025>
- Kelley, M. C., & Aalto, D.** (2019). Measuring the dispersion of density in head and neck cancer patients’ vowel spaces: The vowel dispersion index. *Canadian Acoustics*, 47(3), 114–115. <https://jcaa.caa-aca.ca/index.php/jcaa/article/view/3330>
- Kelso, J. S., & Tuller, B.** (1983). Compensatory articulation under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech and Hearing Research*, 26(2), 217–224. <https://doi.org/10.1044/jshr.2602.217>
- Kent, R. D.** (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 19(3), 421–447. <https://doi.org/10.1044/jshr.1903.421>
- Kent, R. D.** (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development* (pp. 65–90). York Press.
- Kent, R. D., & Bauer, H. R.** (1985). Vocalizations of one-year-olds. *Journal of Child Language*, 12(3), 491–526. <https://doi.org/10.1017/S0305000900006620>
- Kent, R. D., & Murray, A. D.** (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, 72(2), 353–365. <https://doi.org/10.1121/1.388089>
- Kent, R. D., Osberger, M. J., Netsell, R., & Hustedde, C. G.** (1987). Phonetic development in identical twins differing in auditory function. *Journal of Speech and Hearing Disorders*, 52(1), 64–75. <https://doi.org/10.1044/jshd.5201.64>
- Kent, R. D., & Vorperian, H. K.** (1995). Anatomic development of the craniofacial-oral-laryngeal systems: A review. *Journal of Medical Speech-Language Pathology*, 3, 145–190.
- Kent, R. D., & Vorperian, H. K.** (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders*, 74, 74–97. <https://doi.org/10.1016/j.jcomdis.2018.05.004>
- Kewley-Port, D., Burkle, T. Z., & Lee, J. H.** (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 122(4), 2365–2375. <https://doi.org/10.1121/1.2773986>
- Kim, S., Kim, J. H., & Ko, D. H.** (2014). Characteristics of vowel space and speech intelligibility in patients with spastic dysarthria. *Communication Sciences & Disorders*, 19(3), 352–360. <https://doi.org/10.12963/csd.14150>
- Kitamura, C., & Notley, A.** (2009). The shift in infant preferences for vowel duration and pitch contour between 6 and 10 months of age. *Developmental Science*, 12(5), 706–714. <https://doi.org/10.1111/j.1467-7687.2009.00818.x>
- Knight, R.-A., Bandali, C., Woodhead, C., & Vansadia, P.** (2018). Clinicians’ views of the training, use and maintenance of phonetic transcription in speech and language therapy. *International Journal of Language & Communication Disorders*, 53(4), 776–787. <https://doi.org/10.1111/1460-6984.12381>
- Ko, E. S.** (2007). Acquisition of vowel duration in children speaking American English. In *Eighth Annual Conference of the International Speech Communication Association*. International Speech Communication Association. <https://www.isca-speech.org/iscaweb/>
- Ko, E. S., Soderstrom, M., & Morgan, J.** (2009). Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English. *The Journal of the Acoustical Society of America*, 126(5), EL134–EL139. <https://doi.org/10.1121/1.3239465>
- Krause, S. E.** (1982). Developmental use of vowel duration as a cue to postvocalic stop consonant voicing. *Journal of Speech and Hearing Research*, 25(3), 388–393. <https://doi.org/10.1044/jshr.2503.388>
- Kröger, B. J., Schnitker, R., & Lowit, A.** (2008). The organization of a neurocomputational control model for articulatory speech synthesis. In A. Esposito, N. Bourbakis, N. Avouris, & I. Hatzilygeroudis (Eds.), *Verbal and nonverbal features of human-human and human-machine interaction* (pp. 121–135). Springer. [https://doi.org/10.1007/978-3-540-70872-8\\_9](https://doi.org/10.1007/978-3-540-70872-8_9)
- Kronrod, Y., Coppess, E., & Feldman, N. H.** (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin Reviews*, 23, 1681–1712. <https://doi.org/10.3758/s13423-016-1049-y>
- Kuhl, P. K.** (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, 66(6), 1668–1679. <https://doi.org/10.1121/1.383639>
- Kuhl, P. K.** (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior & Development*, 6(2–3), 263–285. [https://doi.org/10.1016/S0163-6383\(83\)80036-8](https://doi.org/10.1016/S0163-6383(83)80036-8)
- Kuhl, P. K.** (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 293–332). American Psychological Association. <https://doi.org/10.1037/10119-012>
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T.** (2008). Phonetic learning as a pathway to language: New data and Native Language Magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 979–1000. <https://doi.org/10.1098/rstb.2007.2154>
- Kuhl, P. K., & Meltzoff, A. N.** (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100(4), 2425–2438. <https://doi.org/10.1121/1.417951>
- Lecanuet, J. P., Granier-Deferre, C., DeCasper, A. J., Maugeais, R., & Andrieu, A. J.** (1987). Perception et discrimination foetales de

- stimuli langagiers; mise en évidence à partir de la réactivité cardiaque; résultats préliminaires. *Comptes rendus de l'Académie des Sciences. Série III, Sciences de la vie*, 305(5), 161–164.
- Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105(3), 1455–1468. <https://doi.org/10.1121/1.426686>
- Lehman, M. E., & Sharf, D. J. (1989). Perception/production relationships in the development of the vowel duration cue to final consonant voicing. *Journal of Speech and Hearing Research*, 32(4), 803–815. <https://doi.org/10.1044/jshr.3204.803>
- Leino, T., Laukkanen, A. M., & Radolf, V. (2011). Formation of the actor's/speaker's formant: A study applying spectrum analysis and computer modeling. *Journal of Voice*, 25(2), 150–158. <https://doi.org/10.1016/j.jvoice.2009.10.002>
- Leopold, W. F. (1947). *Speech development of a bilingual child* (Vol. 2). Northwestern University.
- Levy, H., & Hanulíková, A. (2019). Variation in children's vowel production: Effects of language exposure and lexical frequency. *Journal of the Association for Laboratory Phonology*, 10(1), 9. <https://doi.org/10.5334/labphon.131>
- Lieberman, D. E., & McCarthy, R. C. (1999). The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *Journal of Human Evolution*, 36(5), 487–517. <https://doi.org/10.1006/jhev.1998.0287>
- Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, 74(3), 287–307. <https://doi.org/10.1525/aa.1972.74.3.02a00020>
- Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2005). 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*, 117, 3879–3889. <https://doi.org/10.1121/1.1898623>
- Ludlow, C. L., Kent, R. D., & Gray, L. (2019). *Measuring speech, voice and swallowing in the laboratory and clinic*. Plural.
- Maassen, B., Groenen, P., & Crul, T. (2003). Auditory and phonetic perception of vowels in children with apraxic speech disorders. *Clinical Linguistics & Phonetics*, 17(6), 447–467. <https://doi.org/10.1080/0269920031000070821>
- MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's development of self-regulation in speech production. *Current Biology*, 22(2), 113–117. <https://doi.org/10.1016/j.cub.2011.11.052>
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7(1), 41–74. <https://doi.org/10.1017/S0305000900007029>
- MacNeilage, P. F., & Davis, B. (1990). Acquisition of speech production: Frame, then content. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 453–475). Erlbaum. <https://doi.org/10.4324/9780203772010-15>
- Maddieson, I. (1984). *Patterns of sounds (Cambridge studies in speech science and communication)*. Cambridge University Press.
- Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, 10(2), 179–204. <https://doi.org/10.1080/15475441.2013.829740>
- Manca, A. D., & Grimaldi, M. (2016). Vowels and consonants in the brain: Evidence from magnetoencephalographic studies on the N1m in normal-hearing listeners. *Frontiers in Psychology*, 7, 1413. <https://doi.org/10.3389/fpsyg.2016.01413>
- Mani, N., Mills, D. L., & Plunkett, K. (2012). Vowels in early words: An event-related potential study. *Developmental Science*, 15(1), 2–11. <https://doi.org/10.1111/j.1467-7687.2011.01092.x>
- Markova, D., Richer, L., Pangelinan, M., Schwartz, D. H., Leonard, G., Perron, M., Pike, G. B., Veillete, S., Chakravaty, M. M., Pausova, Z., & Paus, T. (2016). Age- and sex-related variations in vocal-tract morphology and voice acoustics during adolescence. *Hormones and Behavior*, 81, 84–96. <https://doi.org/10.1016/j.yhbeh.2016.03.001>
- Martin, J. A. M. (1981). *Voice, speech and language in the child: Development and disorder*. Springer. <https://doi.org/10.1007/978-3-7091-7042-7>
- Martin, P. (2004). WinPitch LTL II, a multimodal pronunciation software. In *Proceedings of InSTIL/ICALL Symposium 2004*.
- Masapollo, M., Polka, L., & Ménard, L. (2016). When infants talk, infants listen: Pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, 19(2), 318–328. <https://doi.org/10.1111/desc.12298>
- Maturo, S., Hill, C., Bunting, G., Ballif, C., Maurer, R., & Hartnick, C. (2012). Establishment of a normative pediatric acoustic database. *Archives of Otolaryngology—Head & Neck Surgery*, 138(10), 956–961. <https://doi.org/10.1001/2013.jamaoto.104>
- McAllister Byun, T., Buchwald, A., & Mizoguchi, A. (2016). Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical Linguistics & Phonetics*, 30(3–5), 249–276. <https://doi.org/10.3109/02699206.2015.1056884>
- McCarthy, K., Skoruppa, K., & Iverson, P. (2015). Detailing vowel development in infancy using cortical auditory evoked potentials and multidimensional scaling. *The Journal of the Acoustical Society of America*, 138(3), 1810. <https://doi.org/10.1121/1.4933743>
- McGowan, R. W., McGowan, R. S., Denny, M., & Nittrouer, S. (2014). A longitudinal study of very young children's vowel production. *Journal of Speech, Language, and Hearing Research*, 57(1), 1–15. [https://doi.org/10.1044/1092-4388\(2013\)12-0112](https://doi.org/10.1044/1092-4388(2013)12-0112)
- Menn, L. (1976). *Pattern, control, and contrast in beginning speech: A case study in the development of word form and function* [Unpublished doctoral dissertation]. University of Illinois.
- Meyer, J., Dentel, L., & Meunier, F. (2013). Speech recognition in natural background noise. *PLOS ONE*, 8(11), Article e79279. <https://doi.org/10.1371/journal.pone.0079279>
- Ménard, L., Perrier, P., Aubin, J., Savariaux, C., & Thibeault, M. (2008). Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children. *The Journal of the Acoustical Society of America*, 124(2), 1192–1206. <https://doi.org/10.1121/1.2945704>
- Ménard, L., Schwartz, J.-L., & Boë, L.-J. (2004). Role of vocal tract morphology in speech development. *Journal of Speech, Language, and Hearing Research*, 47(5), 1059–1080. [https://doi.org/10.1044/1092-4388\(2004\)079](https://doi.org/10.1044/1092-4388(2004)079)
- Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, 102(2), 156–160. <https://doi.org/10.1111/apa.12098>
- Morrison, G. S. (2012). In P. F. Assmann (Ed.), *Vowel inherent spectral change*. Springer Science & Business Media. <https://doi.org/10.1007/978-3-642-14209-3>
- Munson, B., Crocker, L., Pierrehumbert, J. B., Owen-Anderson, A., & Zucker, K. J. (2015). Gender typicality in children's speech: A comparison of boys with and without gender identity disorder. *The Journal of the Acoustical Society of America*, 137(4), 1995–2003. <https://doi.org/10.1121/1.4916202>
- Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's



- speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, 21(2), 124–139. [https://doi.org/10.1044/1058-0360\(2011/11-0009\)](https://doi.org/10.1044/1058-0360(2011/11-0009))
- Narasimhan, S. V., Nikitha, M., & Francis, N.** (2016). Articulatory working space area in children with cerebral palsy. *International Journal of Health Sciences and Research*, 6, 335–341.
- Nathani, S., Oller, D. K., & Cobo-Lewis, A. B.** (2003). Final syllable lengthening (FSL) in infant vocalizations. *Journal of Child Language*, 30(1), 3–25. <https://doi.org/10.1017/S0305000902005433>
- Nazzi, T., & Cutler, A.** (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, 5, 25–47. <https://doi.org/10.1146/annurev-linguistics-011718-011919>
- Neel, A. T.** (2010). Using acoustic phonetics in clinical practice. *SIG 5 Perspectives on Speech Science and Orofacial Disorders*, 20(1), 14–24. <https://doi.org/10.1044/ssod.20.1.14>
- Netsell, R.** (1985). Construction and use of a bite-block for the evaluation and treatment of speech disorders. *Journal of Speech and Hearing Disorders*, 50(1), 103–106. <https://doi.org/10.1044/jshd.5001.103>
- Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A.** (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The Journal of the Acoustical Society of America*, 124(1), 576–588. <https://doi.org/10.1121/1.2931949>
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R.** (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), 6318–6323. <https://doi.org/10.1073/pnas.1300337110>
- Oller, D. K., & Eilers, R. E.** (1975). Phonetic expectation and transcription validity. *Phonetica*, 31(3–4), 288–304. <https://doi.org/10.1159/000259675>
- Oohashi, H., Watanabe, H., & Taga, G.** (2017). Acquisition of vowel articulation in childhood investigated by acoustic-to-articulatory inversion. *Infant Behavior & Development*, 46, 178–193. <https://doi.org/10.1016/j.infbeh.2017.01.007>
- Otomo, K., & Stoel-Gammon, C.** (1992). The acquisition of unrounded vowels in English. *Journal of Speech and Hearing Research*, 35(3), 604–616. <https://doi.org/10.1044/jshr.3503.604>
- Payne, E., Post, B., Prieto, P., Vanrell, M., & Astruc, L.** (2012). Measuring child rhythm. *Language and Speech*, 55(2), 202–228. <https://doi.org/10.1177/0023830911417687>
- Peeters, W. J. M.** (2019, November). Vowels, diphthongs, and vowel clusters. A quantitative dynamic. In *New Methods in Dialectology: Proceedings of a Workshop held at the Free University of Amsterdam, December 7–10, 1987* (Vol. 33, pp. 67–77). de Gruyter.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H.** (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, 109(6), 2988–2998. <https://doi.org/10.1121/1.1370525>
- Peterson, G. E., & Barney, H. E.** (1952). Control methods used in a study of vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Petrovich-Bartell, N., Cowan, N., & Morse, P. A.** (1982). Mothers' perceptions of infant distress vocalizations. *Journal of Speech and Hearing Research*, 25(3), 371–376. <https://doi.org/10.1044/jshr.2503.371>
- Pisoni, D. B.** (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13, 253–260. <https://doi.org/10.3758/BF03214136>
- Polka, L., & Bohn, O. S.** (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221–231. [https://doi.org/10.1016/S0167-6393\(02\)00105-X](https://doi.org/10.1016/S0167-6393(02)00105-X)
- Polka, L., & Bohn, O. S.** (2011). Natural referent vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39(4), 467–478. <https://doi.org/10.1016/j.wocn.2010.08.007>
- Pollock, K. E.** (1991). The identification of vowel errors using traditional articulation or phonological process test stimuli. *Language, Speech, and Hearing Services in Schools*, 22(2), 39–50. <https://doi.org/10.1044/0161-1461.2202.39>
- Pollock, K. E.** (1994). Assessment and remediation of vowel misarticulations. *Clinics in Communication Disorders*, 4, 23–37.
- Pollock, K. E.** (2002). Identification of vowel errors: Methodological issues and preliminary data from the Memphis Vowel Project. In M. Ball & F. Gibbon (Eds.), *Vowel disorders* (pp. 83–113). Butterworth-Heinemann.
- Pollock, K. E., & Berni, M. C.** (2001). The transcription of vowels. *Topics in Language Disorders*, 21(4), 22–40. <https://doi.org/10.1097/00011363-200121040-00005>
- Pollock, K. E., & Hall, P. K.** (1991). An analysis of the vowel misarticulations of five children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 5(3), 207–224. <https://doi.org/10.3109/02699209108986112>
- Post, B., & Payne, E.** (2018). Speech rhythm in development: What is the child acquiring? In N. Esteve-Gibert & P. Prieto (Eds.), *Prosodic development in first language acquisition* (pp. 125–144). John Benjamins. <https://doi.org/10.1075/tilar.23.07pos>
- Preston, J. L., Ramsdell, H. L., Oller, D. K., Edwards, M. L., & Tobin, S. J.** (2011). Developing a weighted measure of speech sound accuracy. *Journal of Speech, Language, and Hearing Research*, 54(1), 1–18. [https://doi.org/10.1044/1092-4388\(2010/10-0030\)](https://doi.org/10.1044/1092-4388(2010/10-0030))
- Pujol, R., Lavigne-Rebillard, M., & Uziel, A.** (1991). Development of the human cochlea. *Acta Oto-Laryngologica*, 111(Suppl. 482), 7–13. <https://doi.org/10.3109/00016489109128023>
- Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G.** (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology*, 28(3), 191–212. [https://doi.org/10.1016/0028-2243\(88\)90030-5](https://doi.org/10.1016/0028-2243(88)90030-5)
- Ramus, F., Hauser, M. D., Miller, C. T., Morris, D., & Mehler, J.** (2000). Language discrimination by human newborns and cotton-top tamarin monkeys. *Science*, 288(5464), 349–351. <https://doi.org/10.1126/science.288.5464.349>
- Ramus, F., Nespore, M., & Mehler, J.** (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292. [https://doi.org/10.1016/S0010-0277\(99\)00058-X](https://doi.org/10.1016/S0010-0277(99)00058-X)
- Raphael, L. J., Dorman, M. F., & Geffner, D.** (1980). Voicing-conditioned durational differences in vowels and consonants in the speech of three- and four-year old children. *Journal of Phonetics*, 8(3), 335–341. [https://doi.org/10.1016/S0095-4470\(19\)31483-4](https://doi.org/10.1016/S0095-4470(19)31483-4)
- Recasens, D., & Rodríguez, C.** (2016). A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics*, 59, 58–75. <https://doi.org/10.1016/j.wocn.2016.09.002>
- Robb, M. P., Chen, Y., & Gilbert, H. R.** (1997). Developmental aspects of formant frequency and bandwidth in infants and toddlers. *Folia Phoniatrica et Logopaedica*, 49(2), 88–95. <https://doi.org/10.1159/000266442>
- Rogers, C. L., Glasbrenner, M. M., DeMasi, T. M., & Bianchi, M.** (2012). Vowel inherent spectral change and the second-language learner. In G. Morrison & P. Assmann (Eds.), *Vowel inherent*



- spectral change (pp. 231–259). Springer. [https://doi.org/10.1007/978-3-642-14209-3\\_10](https://doi.org/10.1007/978-3-642-14209-3_10)
- Rothgänger, H.** (2003). Analysis of the sounds of the child in the first year of age and a comparison to the language. *Early Human Development*, 75(1–2), 55–69.
- Rvachew, S., Alhaidary, A., Mattock, K., & Polka, L.** (2008). Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French. *Journal of Phonetics*, 36(4), 564–577. <https://doi.org/10.1016/j.wocn.2008.02.001>
- Rvachew, S., Mattock, K., Polka, L., & Ménard, L.** (2006). Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *The Journal of the Acoustical Society of America*, 120(4), 2250–2259. <https://doi.org/10.1121/1.2266460>
- Rvachew, S., Slawinski, E. B., & Williams, M.** (1996). Formant frequencies of vowels produced by infants with and without early onset otitis media. *Canadian Acoustics*, 24(2), 19–28.
- Sanders, I., Mu, L., Amirali, A., Su, H., & Sobotka, S.** (2013). The human tongue slows down to speak: Muscle fibers of the human tongue. *Anatomical Record*, 296(10), 1615–1627. <https://doi.org/10.1002/ar.22755>
- Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., & Spanias, A.** (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*, 134(5), EL477–EL483. <https://doi.org/10.1121/1.4826150>
- Sandoval, S., Utianski, R. L., & Lehnert-LeHouillier, H.** (2019). Average formant trajectories. *Perspectives of the ASHA Special Interest Groups*, 4(4), 719–732. [https://doi.org/10.1044/2019\\_PERS-SIG19-2019-0002](https://doi.org/10.1044/2019_PERS-SIG19-2019-0002)
- Sapir, S., Ramig, L. O., Spielman, J., & Fox, C.** (2011). Acoustic metrics of vowel articulation in Parkinson's disease: Vowel space area (VSA) vs. vowel articulation index (VAI). In C. Manfredi (Ed.), *Seventh International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*. Firenze University Press.
- Scherer, S., Lucas, G. M., Gratch, J., Rizzo, A. S., & Morency, L. P.** (2016). Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. *IEEE Transactions on Affective Computing*, 7(1), 59–73. <https://doi.org/10.1109/TAFFC.2015.2440264>
- Scherer, S., Morency, L. P., Gratch, J., & Pestian, J.** (2015). Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, Australia* (pp. 4789–4793). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ICASSP.2015.7178880>
- Schötz, S., Frid, J., & Löfqvist, A.** (2013). Development of speech motor control: Lip movement variability. *The Journal of the Acoustical Society of America*, 133(6), 4210. <https://doi.org/10.1121/1.4802649>
- Schwartz, G., Aperliński, G., Kazmierski, K., & Weckwerth, J.** (2016). Dynamic targets in the acquisition of L2 English vowels. *Research in Language*, 14(2), 181–202. <https://doi.org/10.1515/rela-2016-0011>
- Schwartz, J.-L., Abry, C., Boë, L.-J., Ménard, L., & Vallée, N.** (2005). Asymmetries in vowel perception, in the context of the dispersion–focalisation theory. *Speech Communication*, 45(4), 425–434. <https://doi.org/10.1016/j.specom.2004.12.001>
- Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G.** (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication*, 66, 65–81. <https://doi.org/10.1016/j.specom.2014.09.006>
- Scobbie, J., Gibbon, F., Hardcastle, W., & Fletcher, P.** (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon* (pp. 194–207). Cambridge University Press.
- Seidl, A., Brosseau-Lapré, F., & Goffman, L.** (2018). The impact of brief restriction to articulation on children's subsequent speech production. *The Journal of the Acoustical Society of America*, 143(2), 858–863. <https://doi.org/10.1121/1.5021710>
- Selby, J. C., Robb, M. P., & Gilbert, H. R.** (2000). Normal vowel articulations between 15 and 36 months of age. *Clinical Linguistics & Phonetics*, 14(4), 255–256. <https://doi.org/10.1080/02699200050023976>
- Sell, D., & Sweeney, T.** (2020). Percent consonant correct as an outcome measure for cleft speech in an intervention study. *Folia Phoniatrica et Logopaedica*, 72(2), 143–151. <https://doi.org/10.1159/000501095>
- Shahidullah, S., & Hepper, P. G.** (1994). Frequency discrimination by the fetus. *Early Human Development*, 36(1), 13–26. [https://doi.org/10.1016/0378-3782\(94\)90029-9](https://doi.org/10.1016/0378-3782(94)90029-9)
- Shiller, D. M., Rvachew, S., & Brosseau-Lapré, R.** (2010). Importance of the auditory perceptual target to the achievement of speech production accuracy [Importance de la cible perceptive auditive dans l'atteinte d'une production adéquate de la parole]. *Canadian Journal of Speech-Language Pathology and Audiology*, 34(3), 181–192.
- Shriberg, L. D., Fourakis, M., Hall, S., Karlsson, H. B., Lohmeier, H. L., McSweeney, J., Potter, N. L., Sxheer-Cohen, A. R., Strand, E. A., Tilkens, C. M., & Wilson, D. L.** (2010). Perceptual and acoustic reliability estimates for the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics*, 24(10), 825–846. <https://doi.org/10.3109/02699206.2010.503007>
- Shriberg, L. D., Kent, R. D., McAllister, T., & Preston, J. L.** (2019). *Clinical phonetics* (5th ed.). Pearson.
- Shriberg, L. D., & Lof, G. L.** (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics*, 5(3), 225–279. <https://doi.org/10.3109/02699209108986113>
- Sloos, M., García, A. A., Andersson, A., & Neijmeijer, M.** (2019). Accent-induced bias in linguistic transcriptions. *Language Sciences*, 76, 101176. <https://doi.org/10.1016/j.langsci.2018.06.002>
- Smith, B. L., & McLean-Muse, A.** (1987). Effects of rate and bite block manipulations on kinematic characteristics of children's speech. *The Journal of the Acoustical Society of America*, 81(3), 747–754. <https://doi.org/10.1121/1.394843>
- Solé, M. J., & Ohala, J. J.** (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kühnert, M. Imperio, & N. Vallée (Eds.), *Papers in laboratory phonology* (Vol. 10, pp. 607–655). de Gruyter Mouton.
- Speake, J., Stackhouse, J., & Pascoe, M.** (2012). Vowel targeted intervention for children with persisting speech difficulties: Impact on intelligibility. *Child Language Teaching and Therapy*, 28(3), 277–295. <https://doi.org/10.1177/0265659012453463>
- Spence, M. J., & DeCasper, A. J.** (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior & Development*, 10(2), 133–142. [https://doi.org/10.1016/0163-6383\(87\)90028-2](https://doi.org/10.1016/0163-6383(87)90028-2)
- Spencer, C. E., Clark, J., Hamilton, S. M., & Boyce, S.** (2017). Vowel space in children with residual speech sound disorders. *The Journal of the Acoustical Society of America*, 142(4), 2642–2642. <https://doi.org/10.1121/1.5014685>
- Stark, R. E., & Heinz, J. M.** (1996). Vowel perception in children with and without language impairment. *Journal of Speech and*

- Hearing Research*, 39(4), 860–869. <https://doi.org/10.1044/jshr.3904.860>
- Stilp, C. E., & Kluender, K. R.** (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences of the United States of America*, 107(27), 12387–12392. <https://doi.org/10.1073/pnas.0913625107>
- Stockman, I. J., Woods, D. R., & Tishman, A.** (1981). Listener agreement on phonetic segments in early infant vocalizations. *Journal of Psycholinguistic Research*, 10(6), 593–617. <https://doi.org/10.1007/BF01067296>
- Stoel-Gammon, C.** (2001). Transcribing the speech of young children. *Topics in Language Disorders*, 21(4), 12–21. <https://doi.org/10.1097/00011363-200121040-00004>
- Stoel-Gammon, C., & Herrington, P.** (1990). Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics & Phonetics*, 4(2), 144–160. <https://doi.org/10.3109/026992090008985478>
- Stoel-Gammon, C., & Pollock, K.** (2008). Vowel development and disorders. In M. J. Ball, M. R. Perkins, N. Müller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 525–548). Blackwell. <https://doi.org/10.1002/9781444301007.ch33>
- Story, B. H., & Bunton, K.** (2016). Formant measurement in children's speech based on spectral filtering. *Speech Communication*, 76, 93–111. <https://doi.org/10.1016/j.specom.2015.11.001>
- Švancara, P., Horáček, J., Vokrál, J., & Cerný, L.** (2006). Computational modelling of effect of tonsillectomy on voice production. *Logopedics, Phoniatrics, Vocology*, 31(3), 117–125. <https://doi.org/10.1080/14015430500342277>
- Takatsu, J., Hanai, N., Suzuki, H., Yoshida, M., Tanaka, Y., Tanaka, S., Hasegawa, Y., & Yamamoto, M.** (2017). Phonologic and acoustic analysis of speech following glossectomy and the effect of rehabilitation on speech outcomes. *Journal of Oral and Maxillofacial Surgery*, 75(7), 1530–1541. <https://doi.org/10.1016/j.joms.2016.12.004>
- Takemoto, H., Kitamura, T., Honda, K., & Masaki, S.** (2008). Deformation of the hypopharyngeal cavities due to F0 changes and its acoustic effects. *Acoustical Science and Technology*, 29(4), 300–303. <https://doi.org/10.1250/ast.29.300>
- Templin, M. C.** (1957). Certain language skills in children. *Institute of Child Welfare Monograph Series* (Vol. 26). University of Minnesota Press. <https://doi.org/10.5749/j.ctttv2st>
- Templin, M. C., & Darley, F. L.** (1968). *The Templin–Darley Tests of Articulation*. University of Iowa Press.
- Teoh, A. P., & Chin, S. B.** (2009). Transcribing the speech of children with cochlear implants: Clinical application of narrow phonetic transcriptions. *American Journal of Speech-Language Pathology*, 18(4), 388–401. [https://doi.org/10.1044/1058-0360\(2009\)08-0076](https://doi.org/10.1044/1058-0360(2009)08-0076)
- Terband, H., Van Brenk, F., & van Doornik-van der Zee, A.** (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders*, 51, 64–77. <https://doi.org/10.1016/j.jcomdis.2014.06.009>
- Tsuji, S., & Cristia, A.** (2013). Fifty years of infant vowel discrimination research: What have we learned? *Journal of the Phonetic Society of Japan*, 17(3), 1–11. <https://doi.org/10.1111/j.2040-0209.2013.00427.x>
- Turner, C. W., & Henn, C. C.** (1989). The relation between vowel recognition and measures of frequency resolution. *Journal of Speech and Hearing Research*, 32(1), 49–58. <https://doi.org/10.1044/jshr.3201.49>
- Turner, G. S., Tjaden, K., & Weismer, G.** (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, 38, 1001–1013. <https://doi.org/10.1044/jshr.3805.1001>
- Uther, M., Knoll, M. A., & Burnham, D.** (2007). Do you speak E-NG-LI-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 49(1), 2–7. <https://doi.org/10.1016/j.specom.2006.10.003>
- van Son, R. J., Middag, C., & Demuynck, K.** (2018). Vowel space as a tool to evaluate articulation problems. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 357–361). International Speech Communication Association (ISCA). <https://doi.org/10.21437/Interspeech.2018-68>
- Velten, H. V.** (1943). The growth of phonemic and lexical patterns in infant language. *Language*, 19(4), 281–292. <https://doi.org/10.2307/409932>
- Vilain, C., Berthommier, F., & Boë, L.-J.** (2015). A brief history of articulatory-acoustic vowel representation. In R. Hoffmann & J. Trouvain (Eds.), *Proceedings of the 1st International Workshop on the History of Speech Communication Research (HSCR 2015)* (pp. 148–159). TUD Press.
- Vorperian, H. K., & Kent, R. D.** (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, 50(6), 1510–1515. [https://doi.org/10.1044/1092-4388\(2007\)104](https://doi.org/10.1044/1092-4388(2007)104)
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S.** (2005). Development of vocal tract length during childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117(1), 338–350. <https://doi.org/10.1121/1.1835958>
- Walsh, B., & Smith, A.** (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research*, 45(6), 1119–1133. [https://doi.org/10.1044/1092-4388\(2002\)090](https://doi.org/10.1044/1092-4388(2002)090)
- Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D., & Kent, J. F.** (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica*, 53, 1–18. <https://doi.org/10.1159/000052649>
- Wellman, B., Case, I., Mengert, I., & Bradbury, D.** (1931). Speech sounds of young children. In G. D. Stoddard (Ed.), *University of Iowa studies in child welfare* (Vol. 5). University of Iowa Press.
- Werker, J. F., & Curtin, S.** (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2), 197–234. <https://doi.org/10.1080/15475441.2005.9684216>
- Werker, J. F., & Hensch, T. K.** (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, 66, 173–196. <https://doi.org/10.1146/annurev-psych-010814-015104>
- Whitehill, T. L., Ciocca, V., Chan, J. C.-T., & Samman, N.** (2006). Acoustic analysis of vowels following glossectomy. *Clinical Linguistics & Phonetics*, 20(2–3), 135–140. <https://doi.org/10.1080/02699200400026694>
- Wild, A., Vorperian, H. K., Kent, R. D., Bolt, D. M., & Austin, D.** (2018). Single-word speech intelligibility in children and adults with Down syndrome. *American Journal of Speech-Language Pathology*, 27(1), 222–236. [https://doi.org/10.1044/2017\\_AJSLP-17-0002](https://doi.org/10.1044/2017_AJSLP-17-0002)
- Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B.** (2009). A cerebral network model of speech prosody comprehension.

- International Journal of Speech-Language Pathology*, 11(4), 277–281. <https://doi.org/10.1080/17549500902943043>
- Williams, J. D., Melamed, I. D., Alonso, T., Hollister, B., & Wilpon, J.** (2011, December). Crowd-sourcing for difficult transcription of speech. In *2011 IEEE Workshop on Automatic Speech Recognition & Understanding* (pp. 535–540). Institute of Electrical and Electronics Engineers (IEEE). Curran Associates. <https://doi.org/10.1109/ASRU.2011.6163988>
- Wren, Y., McLeod, S., White, P., Miller, L. L., & Roulstone, S.** (2013). Speech characteristics of 8-year-old children: Findings from a prospective population study. *Journal of Communication Disorders*, 46(1), 53–69. <https://doi.org/10.1016/j.jcomdis.2012.08.008>
- Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U.** (2013). Vowel hyperarticulation in parrot-, dog- and infant-directed speech. *Anthrozoös*, 26(3), 373–380. <https://doi.org/10.2752/175303713X13697429463592>
- Xue, S. A., Lam, C. W. Y., Whitehill, T. L., & Samman, N.** (2011). Effects of Class III malocclusion on young male adults' vocal tract development: A pilot study. *Journal of Oral and Maxillo-facial Surgery*, 69(3), 845–852. <https://doi.org/10.1016/j.joms.2010.02.038>
- Yamashita, Y., Nakajima, Y., Ueda, K., Shimada, Y., Hirsh, D., Seno, T., & Smith, B. A.** (2013). Acoustic analyses of speech sounds and rhythms in Japanese- and English-learning infants. *Frontiers in Psychology*, 4, 57. <https://doi.org/10.3389/fpsyg.2013.00057>
- Yang, B.** (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *The Journal of the Acoustical Society of America*, 24(2), 245–261. <https://doi.org/10.1006/jpho.1996.0013>
- Yang, J., & Fox, R. A.** (2013). Acoustic development of vowel production in American English children. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 1263–1267). International Speech Communication Association.
- Yang, J., & Fox, R. A.** (2017). Acoustic development of vowel production in native Mandarin-speaking children. *Journal of the International Phonetic Association*, 44(3), 261–282.
- Zahorian, S. A., & Jagharghi, A. J.** (1993). Spectral shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, 94(4), 1966–1982. <https://doi.org/10.1121/1.407520>
- Zajac, D. J., Roberts, J. E., Hennon, E. A., Harris, A. A., Barnes, E. F., & Misenheimer, J.** (2006). Articulation rate and vowel space characteristics of young males with fragile X syndrome: Preliminary acoustic findings. *Journal of Speech, Language, and Hearing Research*, 49(5), 1147–1155. [https://doi.org/10.1044/1092-4388\(2006\)082](https://doi.org/10.1044/1092-4388(2006)082)