Why a phenomenology of vowel sounds is needed

Dieter Maurer

Institute for the Performing Arts and Film, Zurich University of the Arts

dieter.maurer@zhdk.ch

Abstract

In literature, there is an extensive and often controversial debate on the primary acoustic and perceptual cues of vowel quality, resulting in two main viewpoints that these cues are contained in either the formants, or, alternatively, in the spectral shape. However, in our understanding, one aspect is highly underestimated: the fact that any spectral representation of vowel-quality is directly or indirectly related to pitch and to the fine structure of the spectrum. Hence, a given formant pattern as well as a given spectral envelope is in many – if not all – cases ambiguous in terms of representing sounds of different vowel qualities, if speakers equal in size and gender or even single speakers produce the sounds on very different pitches. If they in addition vary vocal effort and phonation type, this ambiguity will be further enhanced. Neither of the two above-mentioned viewpoints can account for this issue, however. – The present paper (i) summarises the ongoing debate, (ii) describes the empirical evidence for spectral representation of vowel quality being related to pitch, spectral shape and spectral fine structure of a sound, (iii) concludes that existing approaches to determine the acoustic cues for vowel quality do not account for all recognisable vowel sounds, both conceptually and methodologically, and (iv) argues for the need of a phenomenology of the acoustics of vowels in terms of building up large-scale, language-specific sound descriptions, addressing all variations of production parameters and their possible extension relevant for perceived vowel quality.

Background

Phonetic summaries generally state that vowel sounds exhibit spectral peaks (termed formants) as the primary acoustic and perceptual cue for the perceived vowel quality, and that these peaks are the consequence of vowel-specific resonance characteristics of the vocal tract. However, different conceptual understandings of formants exist side by side, and there is an extensive and often controversial debate in the literature addressing topics that are considered either as aspects of methodology, or as additional cues, or as aspects that are difficult to understand in the framework of a formant concept. (For excellent overviews, see e.g. Harrington, 2012, Ciocca and Whitehill, 2013, Kiefte et al., 2013; for overviews and exemplary discussions of single aspects, see the corresponding references given below).

Formant concept: As Titze et al. (2015) state: “Unfortunately, the common definition between a formant and a resonance is yet to be established.” Above all, formants are understood in terms of either resonances of the vocal tract, or peaks of the spectral envelope, or filters resulting from an acoustic analysis and related to a corresponding algorithm (e.g. LPC; see Wolfe, n.d.).

Formant estimation: Up to now, no objective method of formant estimation exists, regardless of the algorithm applied: formant patterns are generally estimated by means of an interactive measurement procedure involving general phonetic knowledge and analytical skill of the examiner, context information (size and gender of the speaker), visual crosschecks of calculated values on the basis of the sound spectrum and spectrogram, sometimes related to changes of parameter settings and recalculation of the patterns, and manual interpolations of calculated formant tracks (see e.g., Hillenbrand et al., 1995). Even though LPC analysis has replaced spectrographic measurement, there remains an inherent circularity in the method of formant pattern estimation (Ladefoged, 1967, Hillenbrand et al., 1995). In addition to this general methodological problem, incongruency between expected and the actual numbers of spectral peaks occurs, understood as “formant merging” or as “spurious formants” (Ladefoged, 2003, pp. 114–115, 119–120). Further, and most important, formant estimation loses methodological substantiation with increasing fundamental frequency (f_o). Some scholars consider the critical f_o frequency level as corresponding to approximately half of the first formant frequency (F_1) of a sound (see Ladefoged, 1967, pp. 80–81, Sundberg, 1987, pp. 124–125; Swerdlin et al., 2010), others assume an f_o level in the F_1 region of the closed vowels, i.e. an f_o level of c. 350 Hz (Monson & Engebretson, 1983, Fereirra, 2007) as representing that limit.

Formants and additional cues: The debate on additional cues that potentially affect the acoustics of vowel sounds and the perception of vowel quality, concerns different types of phonation (see Swerdlin et al., 2010), speaker characteristics (above all size and gender differences; see Johnson, 2008) and f_o (see Cheveigné & Kawahara, 1999, Barreda & Nearey, 2012), duration (Hillenbrand et al., 2000), vowel-inherent spectral change, context and transitions (see Morrison & Assmann, 2013), formant amplitude (see Kiefte et al., 2010), spectral contrast and spectral tilt (see Liu & Eddins, 2008), and auditory spectral averaging process (see Chistovich & Lublinskaya, 1979).

Aspects difficult to understand in the framework of formants: Besides the lack of an objective method to estimate formant patterns, the debate on aspects that are difficult to understand in the framework of a formant concept concerns, above all, the lack of evidence that the data reduction process, implied by this concept, corresponds to the auditory processing of speech sounds, as well as observed nonlinearities in the relation between shifts of formant frequencies and shifts in the perceived vowel quality (Bladon, 1982), and the lack of evidence for a peak picking mechanism of perception as indicated by recognisable vowel sounds with suppressed single formants (Ito et al., 2001) or flat spectra (Carpenter & Morton, 1962, Gooding, 1986, Maurer, 2016, pp. 147–157; Maurer & Suter, 2017a; see also Ito, 2001).

Formants versus spectral shape: Referring to Hillenbrand and Houde (2003) and Swanepoel et al. (2012), we conclude that the entire debate on the multitude of aspects mentioned and their often controversial appraisal still have left us with only two main viewpoints, that the major acoustic and perceptual cues are contained in either the formants – more precisely the formant frequency patterns (Carlson et al, 1975, Kasturi et al., 2002) – or, alternatively, in the spectral shape (Bladon, 1982, Zahorian & Jagharghi, 1993; see Ito et al., 2001, Molis, 2005, for a relativisation of a complete opposition), all other aspects of minor or additional effect. Thereby, spectral shape is commonly understood as the envelope of the spectrum derived from some kind of smoothing operation (Hillenbrand, 2003).

Methodological limitations of spectral envelope estimation: With rising f_o, as is true for formant estimation, spectral smoothing becomes also problematic because of spectral undersampling and interrelated distortions. The problem is severe for $f_o \geq 300$ Hz (Cheveigné & Kawahara, 1999, Hillenbrand, 2003).

Core problem: vowel quality-specific spectral representation is related to pitch and to spectral fine structure

However, in our understanding of the matter, two aspects are highly underestimated: Firstly, the fact that vowel quality-specific spectral representation is directly or indirectly pitch-related, and that, as a consequence, a given formant pattern as well as a given spectral envelope is in many – if not all – cases ambiguous in terms of acoustically and perceptually representing sounds of different perceived vowel qualities, if the sounds are produced by speakers *equal in size and gender* on very different pitches. *Thus, pitch-related spectral representation of vowel quality as such cannot primarily be tied to speaker differences in size and gender or to paralinguistic variation.* Secondly, the fact that this pitch-related spectral representation of vowel quality is not systematic or uniform, but also depends on the fine structure of the sound spectrum. – Details are given in the following paragraphs.

Formants and f_o: Most scholars conclude for a marginal or very limited effect of f_o on vowel quality of sounds of speakers equal in size and gender (see Cheveigné & Kawahara, 1999, Barreda & Nearey, 2012). However, most of the studies related to this conclusion reported values for f_o variation below 300 Hz. Yet, the few studies which included higher f_o levels, concluded for a substantial effect of f_o on vowel recognition (Fujisaki & Kawashima, 1968; Traunmüller 1981, 1988, Hirahara & Kato, 1992, Maurer & Landis, 1995, 1996, Ménard et al, 2002, Maurer, 2016, pp. 158–169). This finding was either interpreted as calling for some kind of intrinsic normalisation of f_o and formants, possibly also related to paralinguistic variation of vocal effort, or as an indication of pitch-related spectral representation of vowel quality, a perspective adopted here.

"Oversinging" F_1 as generally given in formant statistics: Pätzold and Simpson (1997) reported statistical F_1 for six of the eight long Standard German vowels /i-y-e-ø-o-u/ < 400 Hz for men, and < 450 Hz for women. Summarising studies on vowel recognition in Western classical singing, Sundberg (2013, pp. 86–88) concluded that recognition can be maintained for all vowels up to C5 (523 Hz). Studies on vowel sounds produced in other artistic styles or involving untrained speakers, however, showed even higher f_0 limits for general vowel recognition up to f_0 in the range of 660–1046 Hz (dependent on the conditions of vowel production and of the listening tests; see Smith & Scott, 1980, Maurer & Landis, 1996, Maurer et al., 2014, Friedrichs et al., 2015, Maurer, 2016, pp. 158–166). Further, the corner vowels were found to be recognisable up to 1046 Hz (Friedrichs et al., 2017). All these studies show that, at least for a substantial part of vowels of a language, they can be produced and recognised on f_0 above statistical F_1 obtained for relaxed speech or for citation-form words.

"Oversinging" the f_0 frequency limit for formant and spectral envelope estimation: The finding that vowel sounds can generally be recognised at f_0 of c. 600 Hz and even above indicates a discrepancy between perception and methods of acoustic analysis: vowels can be recognised at pitches for which no formant frequency and no spectral envelope estimation is methodologically substantiated; further, the assumption of a direct relation between "spectral undersampling" and degradation of vowel quality perception (Ryalls & Lieberman, 1982) is also contradicted.

Significance of extensive f_0 variation in vowel production and perception: There is a strong tendency in the phonetic literature to describe the acoustic characteristics of vowel sounds on f_0 levels related to relaxed speech and to citation-form words, and to consider extensive f_0 variation as a phenomenon of either size and age-differences of the speakers, or specific (strong) emotions, or to singing. Moreover, most investigations on singing concern Western classical singing style. However, we assume that the significance of f_0 variation should be reflected on differently: (i) f_0 ranges of speakers different in size and gender substantially overlap. (ii) There is no pitch-related difference of spoken and sung vowels (for corresponding examples, see the pitch contours of speech of actresses and actors shown in Maurer, 2016, pp. 170–182) and, in art, the transition between speaking and singing can be fluid (traditional Chinese opera style may serve here as an excellent example). (iii) Western classical opera style cannot be regarded as providing a general reference for vowel production and recognition, because the style-specific need for vocal power and instrumental sound timbre is often superordinated to vowel differentiation. (iv) Roughly spoken, according to Hollien (1972) and his terminology, vocal expressions can be experienced up to f_0 = c. 500 Hz for men and c. 800 Hz for women in modal register, and up to f_0 = c. 800 Hz for men and even substantially above 1 kHz for women in loft or falsetto register. Thus, voice range profiles of untrained speakers and trained speakers and singers cover often two to three octaves. Further, the intensity range is usually greatest at intermediate f_0 levels (Titze, 1992). (v) Noteworthy, expressions with register changes and/or with strong emotional variations, strong vocal efforts and shouting, as well as specific speaking styles (ethnolects, infant directed speech, speech in a large audience, artistic speaking and singing styles etc.) include extensive f_0 variation.

Spectral representation of vowel quality is f_0 or pitch-related: This all comes down to the conclusion that, concerning the acoustics and perception of (radiated) voiced sounds, spectral representation of vowel quality is directly f_0 or pitch-related (for the difference, see below). For unvoiced sounds (whisper phonation), this relation is indirect in the sense, that their estimated formant patterns and spectral envelopes correspond to patterns and envelopes of only a part of voiced sounds of the same vowel quality, within a limited f_0 range of the latter.

Ambiguity of formant patterns and spectral envelopes: If formant patterns and spectral envelopes for sounds with different f_0 differ, then what is to be expected are sounds with quasi-identical formant patterns or even quasi-identical spectral envelopes which, however, represent different vowel qualities, the main acoustic difference being their level of f_0. This kind of ambiguity was already indicated in early studies of vowel synthesis (Potter & Steinberg, 1950, Miller, 1953), as is true (even not discussed explicitly) for the later study of Hirahara and Kato (1992). Recently, Maurer et al. (2017) confirmed the ambiguity in vowel synthesis. Further, and most importantly, the ambiguity was also demonstrated for natural vocalisations, including sounds produced by speakers equal in size and gender or even by single speakers (Maurer & Landis, 2000; Maurer, 2016, pp. 64–65 and 187–216).

Noteworthy, Maurer et al. (2017) also demonstrated that open-tube resonance patterns, in their turn, are perceptually not "neutral", i.e. not exclusively related to the "neutral" Schwa sound, but that they are also ambiguous for vowel recognition if f_o is varied.

***f_o* versus pitch:** Because the two phenomena discussed here can also be observed in cases of a "missing fundamental" (Maurer & Suter, 2017b), strictly speaking, we consider the phenomena as related to pitch perception. In most cases, however, both f_o and pitch are concerned.

Non-systematic relation between f_o/pitch and vowel quality-related sound spectrum – the role of the spectral fine structure: As discussed earlier (Maurer & Landis, 1995, 2000, Maurer, 2016, p. 59 and pp. 158–169; see also Bladon, 1984), the relation between f_o/pitch, spectral peaks and envelope of the sound (if methodologically substantiated), and vowel quality is not systematic. It varies according to f_o/pitch range and course of the spectral envelope in general, and according to frequencies, levels and harmonic resolution of the spectral peaks in particular, the peaks represented, e.g., in calculated values of formant frequencies, bandwidths and levels. However, roughly speaking, ambiguous spectral peaks and envelopes occur if f_o/pitch is varied substantially above c. 200 Hz, and they primarily concern sounds of closed and mid-open vowel categories.

A lack of conception and methodology

The concept of formants as being the major acoustic and perceptual cue for vowel quality does not account for the phenomena related to pitch-dependent spectral representation of vowel quality: (i) According to this concept, f_o is considered as an aspect of phonation, i.e. an aspect of the source, and formants are considered as an aspect of articulation and vowel differentiation, i.e. an aspect of the filter. These two parameters are assumed as quasi-independent, and the finding of nonlinear dynamics in the source-filter relationship (see Maxfield et al., 2017) does not principally contradict this assumption. (ii) Concerning the method of formant estimation, as said, no methodological substantiation exists to estimate formant patterns for the entire f_o/pitch range of recognisable vowel sounds.

The formant concept represents a powerful model for the description and prediction of vowel quality-related acoustic characteristics of a part of vowel sounds, namely, sounds produced within limited ranges of certain production parameters (above all within certain f_o/pitch limits, but also with regard to other parameters such as phonation type and vocal effort), but it cannot account for vowel sounds independently of these parameters and the limits set by methods of formant estimation (Maurer, 2016).

The same holds true for a corresponding concept of spectral shape as being the major acoustic and perceptual cue for vowel quality: the acoustic representation of vowel-quality as spectral envelope also comes with a pitch-constraint and its determination also loses methodological substantiation with rising f_o.

A phenomenology is needed

Against this background, we argue that there is no robust approach to determine and predict acoustic sound characteristics directly related to perceived vowel quality for all recognisable vowel sounds, neither conceptually nor methodologically. We further conclude that a phenomenological approach to the acoustics of vowel sounds is needed, with three major aims: (i) reaffirming a pitch-related spectral representation of vowel quality and providing corresponding evidence for different fields of the scientific community, (ii) demonstrating the extension of spectral variation of recognisable vowel sounds, and (iii) building up empirical references for future competing approaches to assess the major acoustic cues of vowel quality, generally valid for all recognisable vowel sounds.

General structure and form: A phenomenology of the acoustics of vowels is considered here in terms of large-scale sound descriptions related to a given language, which include and interrelate sounds of all size and gender-related speaker groups, and address all variations of production parameters and their possible relevance for vowel quality.

Vowels: Principally, all vowels of a language are subject of investigation. However, at first, long vowels may be brought into focus because of their duration and their often quasi-constant (steady-state) sound nucleus.

Production parameters to vary: The primary variable parameters required in building up a within-speaker subsample of sounds are phonation types, f_0 including register change, vocal effort, and phoneme context (isolated sounds, CVCV or CVC, minimal pairs, read speech).

Artistic speech and singing styles are of high interest because of the vocal abilities of the artists and the expressed variation of production parameters, including style-specific aspects. Therefore, sounds of non-professionals and professionals, and untrained and trained speakers and singers have to be included in a phenomenology, and production style as well as the differentiation of speaking and singing (and corresponding subtypes of vocal expression) have to be added as parameters to vary.

Acknowledgement

This work was supported by the Swiss National Science Foundation SNSF Grant No. 100016_159350.

References

Barreda, S., & Nearey, T. M. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *The Journal of the Acoustical Society of America*, *131*(1), 466–477.

Bladon, A. (1982). Arguments against formants in the auditory representation of speech. In R. Carlson, & B. Granström (Eds.), *The representation of speech in the peripheral auditory system*, (pp. 95–102). Amsterdam: Elsevier Biomedical Press.

Bladon, R.A.W., Henton, C.G., Pickering, J.B., 1984. Towards an auditory theory of speaker normalization. Lang. Commun. 4, 59–69.

Carlson, R., Fant, G., & Granström, B. (1975). Two-formant models, pitch and vowel perception. In G. Fant, & M. A. A. Tathum (Eds.), *Auditory analysis and perception of speech* (pp. 55–82). London: Academic Press.

Carpenter, A., & Morton, J. (1962). The perception of vowel colour in formantless complex sounds. *Language and Speech*, *5*(4), 205–214.

de Cheveigné, A., & Kawahara, H. (1999). Missing-data model of vowel identification. *Journal of the Acoustical Society of America*, *105*(6), 3497–3508.

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). New York, London: Psychology Press, Taylor & Francis Group.

Chistovich, L. A., & Lublinskaya, V. V. (1979). The ''center of gravity'' effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research, 1,* 185–195.

Ferreira, A. J. (2007). Static features in real-time recognition of isolated vowels at high pitch. *The Journal of the Acoustical Society of America*, *122*(4), 2389–2404.

Friedrichs, D., Maurer, D., & Dellwo, V. (2015). The phonological function of vowels is maintained at fundamental frequencies up to 880 Hz. *The Journal of the Acoustical Society of America*, *138*(1), EL36–EL42.

Friedrichs, D., Maurer, D., Rosen, S., & Dellwo, V. (2017): Vowel recognition at fundamental frequencies up to 1kHz reveals point vowels as acoustic landmarks. *Journal of the Acoustical Society of America, 142*(2), 1025–1033.

Fujisaki, H., & Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE transactions on audio and electroacoustics*, *16*(1), 73–77.

Gooding, F. (1986). Formantless vowels and theories of vowel perception. *The Journal of the Acoustical Society of America*, *80*(S1), S126–S126.

Harrington, J. (2012). Acoustic phonetics. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 81–129). (2nd ed.). Malden MA: Wiley-Blackwell

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022.

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*(5), 3099–3111.

Hillenbrand, J. M., & Houde, R. A. (2003). A narrow band pattern-matching model of vowel perception. *The Journal of the Acoustical Society of America*, *113*(2), 1044–1055.

Hirahara, T., & Kato, H. (1992). The effect of F0 on vowel identification. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka, Y. (Eds.), *Speech perception, production and linguistic structure* (pp. 89–112). Tokyo: Ohmsha.

Hollien, H. (1974). On vocal registers. Communication Sciences Laboratory Quarterly Report, 10(1), 1–33.

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

Johnson, K. (2008): Speaker normalization in speech perception. In D. B. Pisoni, R. E. Remez, *The handbook of speech perception* (pp. 363–389). Malden MA, Wiley-Blackwell.

Kasturi, K., Loizou, P. C., Dorman, M., & Spahr, T. (2002). The intelligibility of speech with "holes"in the spectrum. *The Journal of the Acoustical Society of America*, *112*(3), 1102–1111.

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

Kiefte, M., Nearey, T. M., & Assmann, P. F. (2013). Vowel perception in normal speakers. *Handbook of vowels and vowel disorders* (pp. 160–185). New York NY: Psychology Press.

Ladefoged, P. (2003): Phonetic Data Analysis–An Introduction to Fieldwork and Instrumental Techniques. Malden MA, Wiley-Lackwell.

Ladefoged, P. (1967): Three Areas of Experimental Phonetics. London, Oxford University Press.

Liu, C., & Eddins, D. A. (2008). Effects of spectral modulation filtering on vowel identification. *The Journal of the Acoustical Society of America*, *124*(3), 1704–1715.

Maurer, D. (2016). *Acoustics of the Vowel – Preliminaries*. Bern: Peter Lang.

Maurer, D., Dellwo, V., Suter, H., & Kathiresan, T. (2017): Formant pattern and spectral shape ambiguity of vowel sounds revisited in synthesis: changing perceptual vowel quality by only changing the fundamental frequency. *Journal of the Acoustical Society of America*, *141*(5), 3469–3470.

Maurer, D., d'Heureuse, C., & Landis, T. (2000). Formant pattern ambiguity of vowel sounds. *International Journal of Neuroscience, 100*(1–4), 39–76.

Maurer, D., & Landis, T. (1995). FO-dependence, number alteration, and non-systematic behaviour of the formants in German vowels. *International Journal of Neuroscience*, *83*(1–2), 25–44.

Maurer, D., & Landis, T. (1996). Intelligibility and spectral differences in high-pitched vowels. *Folia phoniatrica et logopaedica*, *48*(1), 1–10.

Maurer, D., Mok, P., Friedrichs, D., & Dellwo, V. (2014). Intelligibility of high-pitched vowel sounds in the singing and speaking of a female Cantonese Opera singer. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2132–2133.

Maurer, D., & Suter, H. (2017a, abstract). "Flat" vowel spectra revisited in vowel synthesis. *The Journal of the Acoustical Society of America*, *141*(5), 3469–3469. For details, see http://www.phones-and-phonemes.org/asa/Poster-ASA-2017b-170611.pdf (retrieved February 10th, 2018).

Maurer, D., & Suter, H. (2017b, abstract). Vowel synthesis related to equal-amplitude harmonic series in frequency ranges > 1 kHz combined with single harmonics < 1 kHz, and including variation of fundamental frequency. *The Journal of the Acoustical Society of America*, *141*(5), 3469–3469. For details, see http://www.phones-and-phonemes.org/asa/Poster-ASA-2017c-170611.pdf (retrieved February 10th, 2018).

Maurer, D., Dellwo, V., Suter, H., Kathiresan, T. (2017, abstract): Formant pattern and spectral shape ambiguity of vowel sounds revisited in synthesis: changing perceptual vowel quality by only changing the fundamental frequency. Journal of the Acoustical Society of America, 141(5):3469–3470. For details, see http://www.phones-and-phonemes.org/asa/Poster-ASA-2017a-170611.pdf (retrieved February 10th, 2018).

Maxfield, L., Palaparthi, A., & Titze, I. (2017). New evidence that nonlinear source-filter coupling affects harmonic intensity and Fo stability during instances of harmonics crossing formants. *Journal of Voice*, *31*(2), 149–156.

Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S., Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *Journal of the Acoustical Society of America 111*(4), 1892–1905.

Miller, R. L. (1953). Auditory tests with synthetic vowels. *The Journal of the Acoustical society of America*, *25*(1), 114–121.

Molis, M. R. (2005). Evaluating models of vowel perception. *The Journal of the Acoustical Society of America*, *111*(2), 2433–2434.

Monsen, R. B., & Engebretson, A. M. (1983). The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction. *Journal of Speech and Hearing Research, 26*(3), 89–97.

Morrison, G. S., & Assmann, P. F. (Eds., 2013). *Vowel inherent spectral change*. Berlin: Springer.

Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel*, *32*, 215–247.

Potter, R. K., & Steinberg, J. C. (1950). Towards the specification of speech. *The Journal of the Acoustical society of America*, *22*(6), 807–820.

Ryalls, J. H., & Lieberman, P. (1982). Fundamental frequency and vowel perception. *The Journal of the Acoustical Society of America*, *72*(5), 1631–1634.

Smith, L. A., & Scott, B. L. (1980). Increasing the intelligibility of sung vowels. *The Journal of the Acoustical Society of America*, *67*(5), 1795–1797.

Sundberg, J. (1987). *The science of the singing voice*. DeKalb IL: Northern Illinois University Press.
Sundberg, J. (2013). Perception of singing. In D. Deutsch (Ed.), *The Psychology of music*. (3rd ed.). London: Academic Press.
Swanepoel, R., Oosthuizen, D. J., & Hanekom, J. J. (2012). The relative importance of spectral cues for vowel recognition in severe noise. *The Journal of the Acoustical Society of America, 132*(4), 2652–2662.
Swerdlin, Y., Smith, J., & Wolfe, J. (2010). The effect of whisper and creak vocal mechanisms on vocal tract resonances. *The Journal of the Acoustical Society of America*, *127*(4), 2590–2598.
Titze, I.R. (1992). Acoustic interpretation of the voice range profile (phonetogram). *Journal of Speech and Hearing Research, 35*(1), 21–34,
Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., ... & Kreiman, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America*, *137*(5), 3005–3007.
Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *The Journal of the Acoustical Society of America*, *69*(5), 1465–1475.
Traunmüller, H. (1988). Paralinguistic variation and invariance in the characteristic frequencies of vowels. *Phonetica*, *45*(1), 1–29.
Wolfe J. (n.d.). Formant: what *is* a formant? http://newt.phys.unsw.edu.au/jw/formant.html Accessed August 28, 2017.
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.