Toward clinical application of landmark-based speech analysis: Landmark expression in normal adult speech

Keiko Ishikawa[a)]
Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, 901 South 6th Street, Champaign, Illinois 61820, USA
ishikak@illinois.edu

Joel MacAuslan
Speech Technology and Applied Research Corporation, 54 Middlesex Turnpike, Bedford, Massachusetts 01730, USA
JoelM@STARAnalyticalServices.com

Suzanne Boyce
Department of Communication Sciences and Disorders, University of Cincinnati, 3202 Eden Avenue, Cincinnati, Ohio 45267, USA
boycese@ucmail.uc.edu

Abstract: The goal of clinical speech analysis is to describe abnormalities in speech production that affect a speaker's intelligibility. Landmark analysis identifies abrupt changes in a speech signal and classifies them according to their acoustic profiles. These acoustic markers, called *landmarks*, may help describe intelligibility deficits in disordered speech. As a first step toward clinical application of landmark analysis, the present study describes expression of landmarks in normal speech. Results of the study revealed that syllabic, glottal, and burst landmarks consist of 94% of all landmarks, and suggest the effect of gender needs to be considered for the analysis.

[DDO]
Date Received: May 4, 2017 **Date Accepted:** October 18, 2017

1. Introduction

Auditory-perceptual analysis is the main method for clinical assessment of disordered speech; however, results obtained through this type of analysis are susceptible to intra- and inter-rater variabilities. Acoustic analysis of speech can supplement the auditory-perceptual judgment by providing consistent measurements across speakers. Despite this advantage, adoption of acoustic tools into clinical practice has been relatively slow. The hindrance can partly be attributed to the absence of a tool that quickly analyzes and generates measurements that are relevant to abnormality in speech production and perception.

Currently available clinical acoustic-analysis methods require manual analysis of the visual representation of a signal, such as a spectrogram, to describe the signal abnormality. This requires considerable time and labor, making such tools impractical for busy clinicians. Many authors have put substantial research effort in to automate the analysis for clinical assessment of disordered speech (Berisha *et al.*, 2013; Bocklet *et al.*, 2012; He *et al.*, 2013; Lustyk *et al.*, 2014; Maier *et al.*, 2009; Middag *et al.*, 2009; Zhou *et al.*, 2012). Landmark (LM) analysis is a novel approach that characterizes speech with acoustic markers that are developed based on the LM theory of speech perception. This theoretically-driven, knowledge-based approach may serve as the basis of a tool for automatic intelligibility assessment. A few authors have examined the clinical potential of the approach. These studies have shown that the LM-based systems are able to detect acoustic differences between normal and dysarthric speech (DiCicco and Patel, 2008; Chenausky *et al.*, 2011). While their findings are encouraging, these studies have included a small number of speakers [i.e., 1 normal speaker in DiCicco and Patel (2008), and 12 normal speakers in Chenausky *et al.* (2011)], and their outcomes were not reported based on individual acoustic markers. Accordingly, how the LM-based analysis would characterize normal speech has not

[a)]Author to whom correspondence should be addressed.

been well-defined. As the individual markers are designed to correspond to particular articulatory gestures, they would provide information relevant to intelligibility.

Historically, speech production has been described by articulatory features of speech sounds. An example of this is distinctive feature theory by Chomsky and Halle (1968). The theory characterizes a speech sample with a bundle of the articulatory features, each of which describes the sound's place, manner, and laryngeal features. The presence or absence of these features are indicated in a binary system such as [+voice] and [−voice]. The work by Chomsky and Halle (1968) was extended by Kenneth Stevens who sought knowledge-based acoustic correlates of these articulatory features and their relationship with speech sound perception (Stevens, 1989). The LM theory of speech production and perception was born out of this work (Liu, 1996; Stevens, 2002). In LM theory, articulatory movement creates abrupt changes in the speech signal that are called LMs. It is postulated that listeners make judgments of which speech sound was produced based on an acoustic profile of the LMs.

Intelligibility is the primary measurement of communicative effectiveness. Despite an extensive investigation, defining acoustic correlates for intelligibility has been difficult. The difficulty is partly due to a complicated interaction between perceptual and cognitive systems. It has been well documented that the listener's cognitive system can "fill-in" missing acoustic cues, thus the acoustic signal alone cannot account for intelligibility (Cooke, 2006). Yet speech perception does not occur without a response of the auditory system to physical changes in a signal, and there is ample evidence for a contrast in signal serving as the basis of speech perception (see Kluender *et al.*, 2003 for a thorough review on this topic). It has also been well documented that there is a distinct acoustic difference between major classes of speech sounds. For example, while approximants and vowels generate a periodic signal rich in harmonic energy, voiceless fricatives such as /s/ generate an aperiodic signal at high frequency (Stevens, 2000). Increasing acoustic contrast between different classes of articulatory features could result in greater intelligibility. As a LM analysis is designed to detect moments with such contrast in the signal, it is possible that its output would serve as a biomarker for intelligibility.

SpeechMark® is a semi-automated LM-based speech analysis program (Boyce *et al.*, 2012), based on works by Liu (1996) and Howitt (2000). It is a knowledge-based tool, which not only analyzes physical aspects of the signal but also applies acoustic knowledge of articulatory features in the process of analysis. In a typical implementation, the algorithm first computes a spectrogram with a 6 ms Hanning window every 1 ms. The spectrogram is then divided into the six frequency bands, ranging from 0.0–0.4, 0.8–1.5, 1.2–2.0, 2.0–3.5, 3.5–5.0, and 5.0–8.0 kHz (Howitt, 2000; Liu, 1996). The spectrogram then goes through "fine" and "coarse" processing. Both of these processes vary in smoothing values, time frame for detection of band-energy rise, and threshold for peak detection. Subsequently, energy peaks are localized and subjected to the final phase of the program in which a type of LM is determined based on patterns of changes in the frequency bands.

The version of SpeechMark® (Speech Technology and Applied Research Corp., Bedford, MA) used for this study (i.e., SpeechMark® WaveSurfer Plug-in) generates one vowel LM and several classes of abrupt LMs, including glottal, burst, syllabic, unvoiced frication, and voiced frication. These LMs are expressed with positive or negative signs which indicate their onset or offset. The final output of the analysis is a sequence of LMs, such as "[+g] [+s] [−g] [+b] [−b]." An example of displayed output is shown in Fig. 1. The abrupt LMs can be categorized into two types: those that describe laryngeal events and others that describe oral events. Glottal LMs, denoted as [+g] and [−g], describe instances of voicing elicited by vocal fold vibration. It should be noted that onset and offset may not necessarily coincide with the physical onset and offset of the vocal fold vibration. Rather, these LMs are generated at moments where sufficient acoustic evidence for the presence of voicing is indicated. The presence of voicing can be confirmed by two rules: (1) a region having high harmonic to noise ratio (HNR), and/or (2) a region adjacent (within 50 ms) of another region having high HNR *and* having similar or higher power and/or similar spectral tilt to the high HNR region. Burst, syllabicity, unvoiced and voiced frication LMs describe oral events. Acoustic rules of all abrupt LMs are described in Table 1.

It is expected that some variations in the expression of LMs exist among normal speakers even when the analysis is performed on the same speech material. Like other acoustic speech analysis tools, SpeechMark® analyzes speech as uttered. It is well recognized that a number of factors can influence the acoustic profile of speech. Examples include speaker's identity (e.g., age, gender, and dialect), speaker's emotional

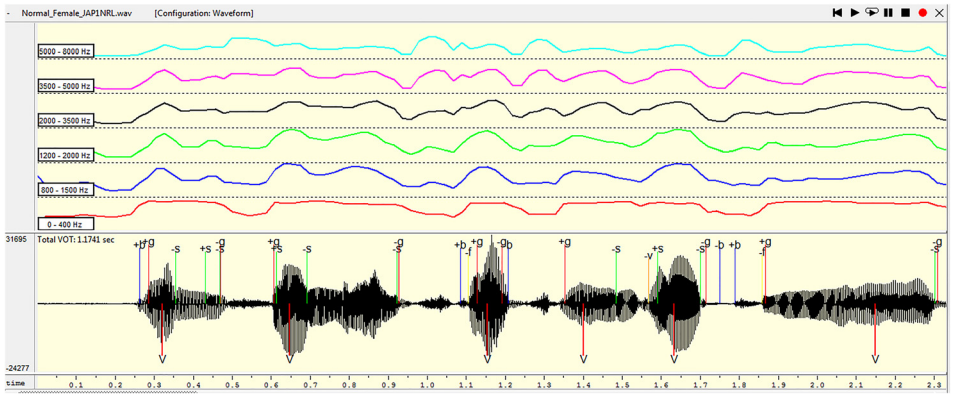

Fig. 1. (Color online) Results of LM analysis with SpeechMark® Plug-in for WaveSurfer, Windows Edition, Version 0.1.36. The top panel shows band energies in the six frequency ranges. The bottom panel shows LMs with a waveform.

state, and speech production style (e.g., conversational vs Lombard speech). Gender-based differences in speech acoustics are especially well-recognized in the literature. The most prominent difference is the fundamental frequency in male and female speech. In order to minimize a possible effect of age and gender on the analysis, SpeechMark® allows adjustment of frequency range based on age and gender of the speaker. This option limits the maximum fundamental frequency for adult male speakers to 220 Hz and female speakers to 350 Hz. Male and female speech also differ in their spectral characteristics, such as formant frequency of vowels (Peterson and Barney, 1952), acoustic parameters relevant to glottal characteristics (Hanson and Chuang, 1999; Klatt and Klatt, 1990), and time-based measurement such as voice onset time (Ryalls *et al.*, 1997; Smith, 1978; Swartz, 1992). Therefore, it is possible that these gender-based acoustic differences would be reflected in an expression of LMs even after adjusting the fundamental frequency for the analysis.

Although acoustic analysis is a powerful tool for describing physical abnormalities in disordered speech, its incorporation into clinical practice has been hindered by lack of an algorithm that quickly yields clinically meaningful data. A LM-based program holds promise as emerging evidence shows that LM-based systems such as SpeechMark® are able to detect acoustic changes in perceptually-different speech. However, the expression pattern of the LMs in normal speech has not been well

Table 1. Acoustic rules for each type of LM. Note that the symbols and mnemonics are not intended to identify underlying articulatory or phonetic events, only to suggest examples: syllabic, voiced frication, etc.

Symbol	Mnemonic	Rule
+g	Glottal onset	Beginning of sustained vocal fold vibration, i.e., of periodicity or of power and spectral slope similar to that of a nearby segment of sustained periodicity
−g	Glottal offset	End of sustained vocal fold vibration
+b	Burst onset	At least 3 of 5 frequency bands show simultaneous power *increases* of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in an unvoiced segment (not between +g and the next −g)
−b	Burst offset	At least 3 of 5 frequency bands show simultaneous power *decreases* of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in an unvoiced segment
+s	Syllabic onset	At least 3 of 5 frequency bands show simultaneous power *increases* of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in a voiced segment (between +g and the next −g)
−s	Syllabic offset	At least 3 of 5 frequency bands show simultaneous power *decreases* of at least 6 dB in both the finely smoothed and the coarsely smoothed contours, in a voiced segment
+f	Frication onset	At least 3 of 5 frequency bands show simultaneous power *increases* at high frequencies *and* decreases at low frequencies (unvoiced segment)
−f	Frication offset	At least 3 of 5 frequency bands show simultaneous power *decreases* at high frequencies and increases at low frequencies (unvoiced segment)
+v	Voiced frication onset	At least 3 of 5 frequency bands show simultaneous power *increases* at high frequencies *and* decreases at low frequencies (voiced segment)
−v	Voiced frication offset	At least 3 of 5 frequency bands show simultaneous power *decreases* at high frequencies and increases at low frequencies (voiced segment)

defined. Clinicians must be able to compare data from their patients to normative data in order to determine the degree of abnormality. Furthermore, the normative data also help to understand the underlying mechanism of the abnormality. Characterizing the expression pattern in normal speech is thus the first step toward developing a LM-based program clinical tool. Therefore, the purpose of this study is to describe LM expression in normal speech using clinical speech material.

2. Methods

2.1 Speech materials

Recordings of the Rainbow passage (Fairbanks, 1960) from Kay Elemetrics' "Disordered Voice Database model 4337" were used for the study (Massachusetts Eye and Ear Infirmary, 1994). The database includes speech samples from 53 adults, who are native speakers of American English with normal voice and speech. It is frequently cited in the disordered voice literature (e.g., Little *et al.*, 2007; Shrivastav and Sapienza, 2003; Zhang and Jiang, 2008). For this study, samples which had sampling rates of less than 11 kHz were excluded. The resultant speaker group consisted of 15 adult females [mean age = 37.8 yrs old, min = 24, max = 52, standard deviation (SD) = 8.1], and 21 adult males (mean age = 38.81 yrs old, min = 26, max = 59, SD = 8.49). The sampling rate of the recordings was standardized at 22 kHz.

2.2 Sample preparation and acoustical analysis

The original speech files contained the first 12 s of the "Rainbow Passage," which is widely used in the evaluation of speech and voice disorders (Fairbanks, 1960; Gilbert and Weismer, 1974; Klostermann *et al.*, 2008). For this study, the files were edited to extract the first sentence: "When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow." The beginning and end of the sentence were visually confirmed on a waveform and spectrogram. These speech samples were analyzed with SpeechMark® Plug-in for WaveSurfer, Windows Edition, Version 0.1.36 for the following parameters: glottis LMs ([+g] and [−g]), burst LMs ([+b] and [−b]), syllabic LMs ([+s] and [−s]), unvoiced frication LMs ([+f] and [−f]), and voiced frication LMs ([+v] and [−v]). In order to minimize the gender effect, the gender option was selected to adjust the range of fundamental frequency for the speaker's gender.

2.3 Statistical analysis

A difference in the average number of LMs between female and male speaker groups was tested with a Welch two-sample *t*-test. A difference in the average number of LMs between onset and offset LMs was tested with a Wilcoxon rank sum test. The level of significance was set as $p < 0.05$. The analyses were performed with R Statistical Software version 3.1.0.

3. Results

The LM analysis of speech samples from all speakers generated a total of 2090 LMs. The average number of LMs was 60.67 [standard error (SE) = 1.11] for female speakers and 56.19 (SE = 1.49) for male speakers (Fig. 2). A two-way analysis of variance was run to examine the effect of gender and age on the total number of LMs. The effect of gender on the total number of LMs was significant, $F(1, 32) = 5.438$, $p = 0.026$. The effect of age on the total number of LMs was not significant, $F(1, 32) = 1.598$, $p = 0.215$. There was no significant interaction between the effects of gender and age on the total number of LMs, $F(1, 32) = 0.348$, $p = 0.559$. A Welch two sample *t*-test indicated that the difference between the male and female groups was significant, $t(34) = 2.41$, $p < 0.02$ (Fig. 2).

The average number of each LM for all speakers combined is shown in Fig. 3. The [+s] was the most frequently occurring LM, followed by [+g], [−g], [−s], [+b], [−b], [−v], [−f], [+v], and [+f]. The data were further analyzed to examine whether there is a difference in the number of onset and offset LMs. A total number of [+g] and [−g] LMs differed only by 1 (407 and 406, respectively), which implies that the numbers of these LMs should be equal to within ±1 for any given recording. A Wilcoxon rank sum test indicated that a greater number of onset LMs than offset LMs were generated in [b] and [s] LMs ($w = 1165$, $p < 0.001$; $w = 931$, $p = 0.001$, respectively). On the other hand, fewer onset LMs than offset LMs were generated in [f] and [v] LMs ($w = 332$, $p < 0.001$; $w = 221$, $p < 0.001$, respectively).

4. Discussion

The purpose of this study was to characterize the expression pattern of LMs in normal speech with clinical speech material. The results of the study showed that [g], [b], and

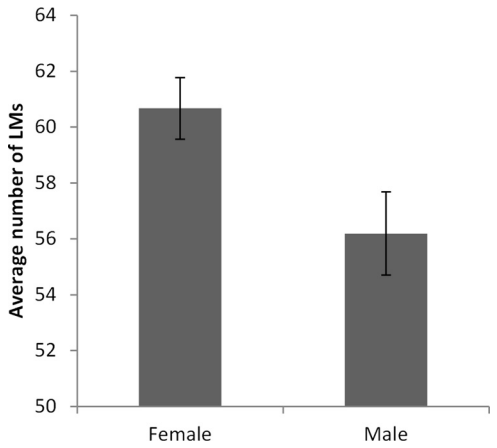

Fig. 2. Average number of LMs in female and male speakers. Error bars indicate standard errors.

[s] LMs comprise 94% of all LMs, occurring considerably more frequently than [f] and [v] LMs. The rarity of [f] and [v] LMs is likely because their acoustic rules are more specific and complex in comparison to the rules for other LMs. The primary condition for detection of [f] and [v] LMs is a simultaneous power change in at least three of the five frequency bands. This rule is also shared by [b] and [s] LMs. Detection of [f] and [v] LMs requires meeting a secondary condition that the simultaneous power change must occur in high frequencies. Additionally, a signal must satisfy a third condition, the presence of a contrary change in low frequencies. The data from this study indicate that it is infrequent to observe a signal change that satisfies all of these three conditions.

It has been reported in the literature that speaker's age and gender are possible influencing factors for intelligibility and acoustic measures (Hazan and Markham, 2004; Jacewicz *et al.*, 2009). The results of this study indicated that the age of our speakers did not affect the total number of LMs. On average, female speech generated a greater number of LMs than male speech, even with the adjustment made for speaker's gender. An interesting hypothesis to consider is that the greater number of LMs in female speech indicates greater intelligibility of their speech. It has been repeatedly demonstrated that female speech is more intelligible than male speech (Bradlow *et al.*, 1996). A study that compared LM expression in conversational and clear speech has shown that clear speech generated a greater number of LMs than conversational speech (Boyce *et al.*, 2013). Furthermore, it has been proposed that intelligibility is influenced by contrast in a speech signal (Kluender *et al.*, 2003). The greater number of LMs in the female speech suggests that there were more abrupt acoustic changes. Accordingly, while intelligibility of our speakers was not measured in this study, it is possible that the number of LMs indicates the superiority intelligibility of female speakers. Further studies are needed to explore the relationship between intelligibility and LM expression.

Acoustic rules for the onset and offset LMs are symmetrically designed; however, the data from this study showed that the acoustic change elicited by articulatory adjustment is not symmetric (except for glottal onset and offset, as expected). During

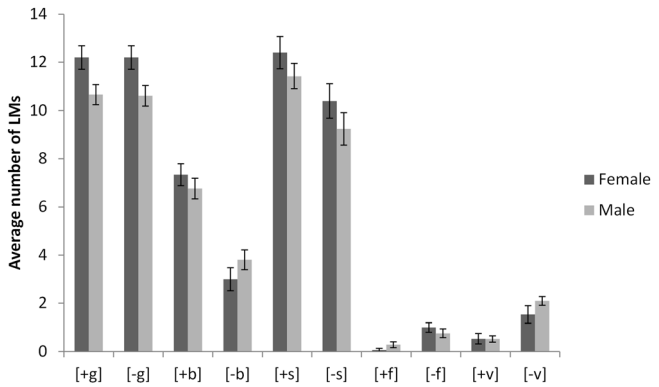

Fig. 3. Average number of LMs for each type of LM in all speakers. Error bars indicate standard errors.

phonation, the onset of vocal fold vibration is always accompanied by the offset of the vibration, creating symmetric rise and fall of acoustic energy. Because glottal LMs are designed to capture these acoustic changes, they occur in pairs. It is not surprising that the number of onset LMs was greater than offset LMs for [b] and [s] LMs. These LMs share acoustic rules for detecting moments of consonant production by vocal tract closure, only differing in the voicing rules. Some consonants have greater energy at their onset than offset. For example, stop consonants are produced upon a release of an obstruction in the vocal tract. The release creates a sudden rise in the acoustic power. Acoustic power then decays after the release, thus the acoustic change is likely more gradual at the end of the consonants. Fricative consonants are produced by turbulent air flow that goes through a constricted point of the vocal tract. Acoustic power created by the airflow is greatest at the onset and decays more gradually as the point of constriction in the vocal tract is widened for the following sound. Consequently, the moments that satisfy the acoustic rule of offset LMs, which is symmetric to the rule of onset LMs, are likely to occur less frequently.

Several limitations should be noted. Because speech samples included in the database were only 10 s long, the analysis was done with only one sentence of the Rainbow passage. The database was designed to illustrate abnormal aspects of speech with voice disorders, and 10 s of speech may be sufficient for this purpose. However, the amount of speech is likely not enough for intelligibility assessment. In a clinical setting, intelligibility is measured with a series of sentences that represent comprehensive phonemic repertoire of a language. One sentence clearly does not include all phonemes of English. For this simple reason alone, it is likely that a greater number of sentences are needed for developing automatized acoustic metrics for describing intelligibility deficits. Furthermore, the distribution of LMs would depend on phonetic content of a speech material, and analysis with the entire passage could have produced a different result. The length of speech sample sufficient for obtaining data that represent normal speech is unknown and needs to be determined for establishing normative data. Another limitation of the study is a lack of evaluation on the effect of possible factors that influence intelligibility. For example, gender, age, and dialect are known to affect intelligibility as well as speech acoustics. The effect of age is not expected in the age range of speakers examined in the study; however, the effect of dialect could not be tested due to the lack of information. Because the sample size of this study is relatively small, the effect of gender on LM expression was examined only for the total number of LMs. It is likely that gender-based differences exist for each particular type of LM. The gender effect on expression of each LM should be examined with a larger sample size in future studies.

To the best of our knowledge, this study is the first to characterize LM expression in normal adult speech with clinical speech material. The results of this study showed that LM expression varies among normal speakers even when standardized speech material is used. The results also illustrated that the gender of a speaker significantly influences LM expression. While the findings of this study should be confirmed by a larger-scale study as noted above, the data from this study may provide a rudimentary "blueprint" for future studies of normal and disordered speech.

Acknowledgments

The authors gratefully acknowledge the support of the United States National Institutes of Health, including NIH Grant Nos. R43/44 DC010104, R42 AG033523, R41/42 DC005534, R21 HL086689, and R41/42 HD034686. The authors also wish to thank Katherine Elkind who assisted in the proofreading of the manuscript.

References and links

Berisha, V., Utianski, R., and Liss, J. (**2013**). "Towards a clinical tool for automatic intelligibility assessment," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, pp. 2825–2828.

Bocklet, T., Riedhammer, K., Nöth, E., Eysholdt, U., and Haderlein, T. (**2012**). "Automatic intelligibility assessment of speakers after laryngeal cancer by means of acoustic modeling," J. Voice **26**(3), 390–397.

Boyce, S., Fell, H. J., and McAuslan, J. (**2012**). "SpeechMark: Landmark detection tool for speech analysis," Paper presented at the INTERSPEECH.

Boyce, S., Krause, J., Hamilton, S., Smilijanic, R., Bradlow, A. R., Rivera-Campos, A., and MacAuslan, J. (**2013**). "Using landmark detection to measure effective clear speech," Proc. Mtgs. Acoust. **19**, 060129.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**(3), 255–272.

Chenausky, K., MacAuslan, J., and Goldhor, R. (**2011**). "Acoustic analysis of PD speech," Parkinson's Disease **2011**, 1–13.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper & Row, New York).

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**(3), 1562–1573.
DiCicco, T., and Patel, R. (**2008**). "Automatic landmark analysis of dysarthric speech," J. Med. Speech-Lang. Path. **16**(4), 213–219.
Fairbanks, G. (**1960**). "The rainbow passage," in *Voice and Articulation Drillbook*, 2nd ed. (Harper Bros., New York).
Gilbert, H. R., and Weismer, G. G. (**1974**). "The effects of smoking on the speaking fundamental frequency of adult women," J. Psycholing. Res. **3**(3), 225–231.
Hanson, H. M., and Chuang, E. S. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**(2), 1064–1077.
Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**(5), 3108–3118.
He, L., Zhang, J., Liu, Q., Yin, H., and Lech, M. (**2013**). "Automatic evaluation of hypernasality and speech intelligibility for children with cleft palate," in *2013 IEEE 8th Conference on Industrial Electronics and Applications* (*ICIEA*), Melbourne, Victoria, pp. 220–223.
Howitt, A. W. (**2000**). "Automatic syllable detection for vowel landmarks," Doctoral dissertation, Massachusetts Institute of Technology.
Jacewicz, E., Fox, R. A., O'Neill, C., and Salmons, J. (**2009**). "Articulation rate across dialect, age, and gender," Lang. Var. Change **21**(2), 233–256.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**(2), 820–857.
Klostermann, F., Ehlen, F., Vesper, J., Nubel, K., Gross, M., Marzinzik, F., Curio, G., and Sappok, T. (**2008**). "Effects of subthalamic deep brain stimulation on dysarthrophonia in Parkinson's disease," J. Neurol., Neurosurg. Psych. **79**(5), 522–529.
Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Commun. **41**(1), 59–69.
Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., and Moroz, I. M. (**2007**). "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," BioMed. Eng. OnLine **6**(1), 23.
Liu, S. A. (**1996**). "Landmark detection for distinctive feature-based speech recognition," J. Acoust. Soc. Am. **100**(5), 3417–3430.
Lustyk, T., Bergl, P., and Cmejla, R. (**2014**). "Evaluation of disfluent speech by means of automatic acoustic measurements," J. Acoust. Soc. Am. **135**(3), 1457–1468.
Maier, A., Hönig, F., Bocklet, T., Nöth, E., Stelzle, F., Nkenke, E., and Schuster, M. (**2009**). "Automatic detection of articulation disorders in children with cleft lip and palate," J. Acoust. Soc. Am. **126**(5), 2589–2602.
Massachusetts Eye, and Ear Infirmary. (**1994**). Disordered Voice Database Model 4373 (Kay Elemetrics), Voice and Speech Lab, Boston, MA.
Middag, C., Martens, J.-P., Van Nuffelen, G., and De Bodt, M. (**2009**). "Automated intelligibility assessment of pathological speech using phonological features," EURASIP J. Adv. Signal Process. **2009**, 3629030.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**(2), 175–184.
Ryalls, J., Zipprer, A., and Baldauff, P. (**1997**). "A preliminary investigation of the effects of gender and race on voice onset time," J. Speech, Lang., Hear. Res. **40**(3), 642–645.
Shrivastav, R., and Sapienza, C. M. (**2003**). "Objective measures of breathy voice quality obtained using an auditory model," J. Acoust. Soc. Am. **114**(4), 2217–2224.
Smith, B. L. (**1978**). "Effects of place of articulation and vowel environment on 'voiced' stop consonant production," Glossa **12**(2), 163–175.
Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonet. **17**, 3–45.
Stevens, K. N. (**2000**). *Acoustic Phonetics*, Vol. 30 (MIT Press, Cambridge, MA).
Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**(4), 1872–1891.
Swartz, B. L. (**1992**). "Gender difference in voice onset time," Percept. Motor Skills **75**(3), 983–992.
Zhang, Y., and Jiang, J. J. (**2008**). "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," J. Voice **22**(1), 1–9.
Zhou, X., Garcia-Romero, D., Mesgarani, N., Stone, M., Espy-Wilson, C., and Shamma, S. (**2012**). "Automatic intelligibility assessment of pathologic speech in head and neck cancer based on auditory-inspired spectro-temporal modulations," in *Thirteenth Annual Conference of the International Speech Communication Association*. Retrieved from http://www.isr.umd.edu/Labs/SCL/publications/conference/xinhui_Interspeech_2012.pdf (Last viewed October 27, 2017).