Predicting Intelligibility of Dysphonic Speech with Automatic Measurement of Vowel Related Parameters

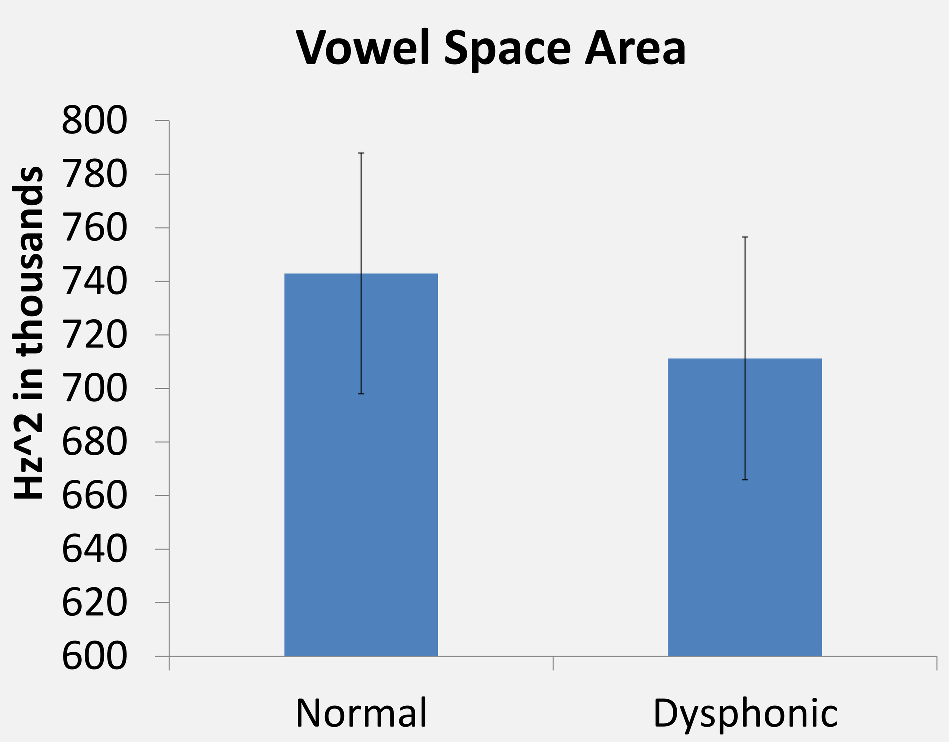
Keiko Ishikawa[1], Meredith Meyer[1], Joel MacAuslan[2] and Suzanne Boyce[1]

University of Cincinnati, College of Allied Health, Dpt. of Communication Sciences and Disorders[1]; Speech Technology & Applied Research[2]

Introduction

- Reduced intelligibility is a common complaint among people with dysphonia.[1]
- Vowels carry information that greatly contributes to intelligibility.[2-3]
- A formant is a cluster of frequencies amplified by the vocal tract. The first two formants are critical for perception of vowels.[4]
- A greater amount of noise and a lack of harmonic power are common characteristics of dysphonic speech signals.[5-6] These acoustic abnormalities can negatively affect perceptual resolution of formants.
- Formant bandwidth can affect intelligibility.[7-8] Listeners may not be able to identify vowels correctly with formants that have abnormally large bandwidth.
- Vowel Space Area (VSA) is a standard acoustic measure used to estimate distinctiveness among vowels in speech production. It is positively correlated with intelligibility in normal and dysarthric speech.[9]
- VSAs are calculated using the measured first two formant frequencies of a set of vowels.
- VSA measurement is traditionally done by hand and this process is highly laborious and time consuming. Automation of this process would dramatically increase its efficiency.
- SpeechMark® is an automatic speech analysis program based on landmark theory of speech production. Instances at which the maximum harmonic power is reached are identified as vowels. Center frequencies of the first two formants are calculated and used to generate VSA.[10]

Results



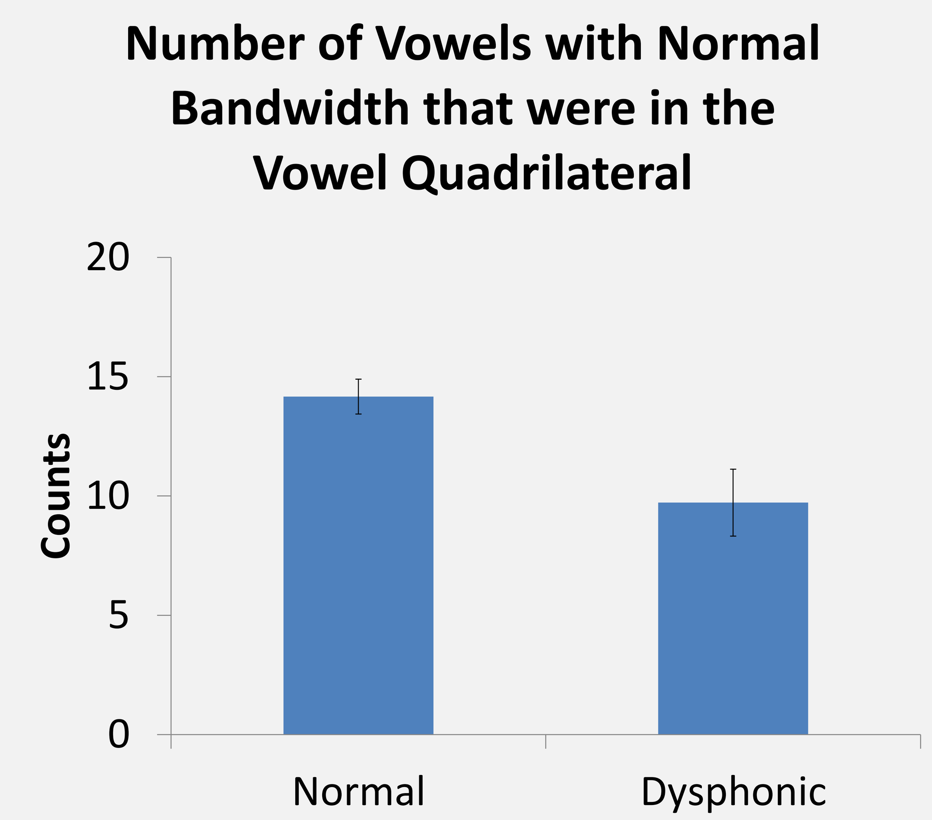


Figure 2:
Top: Bar plot comparing average vowel space area of normal and dysphonic groups

Bottom: Bar plot comparing average number of vowels with normal formant bandwidth that fell in the standard vowel quadrilateral

For both figures, error bars indicate the standard errors.

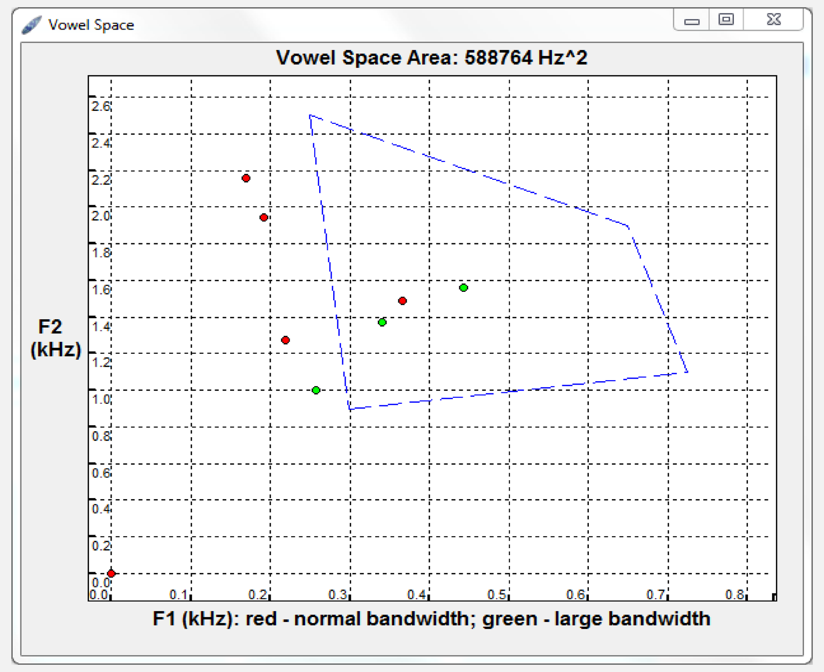


Figure 3: An example of vowel space area display for a dysphonic adult male speaker

Measure	Noise level	Rho	S	P-value	Significance
Vowel space area	SNR-3	0.352	628	0.152	
	SNR+0	0.235	740.882	0.347	
	SNR+3	0.354	625.823	0.149	
Number of Vowels with Normal Bandwidth that were in the Vowel Quadrilateral	SNR-3	0.513	471.457	0.029	
	SNR+0	0.643	345.747	0.004	*
	SNR+3	0.486	497.797	0.041	

Table 1: Results of Spearman rank correlation tests examining the relationship between intelligibility ratings and acoustic parameters. Significance level was set at 0.01 with Bonferroni correction.

Discussions & Conclusions

- There was no statistically significant difference in VSA between normal and dysphonic speech. VSA also did not predict listeners' impression of intelligibility. This result supports a well-recognized clinical observation: the speech production mechanism of dysphonic speakers is normal, while that of dysarthric speakers is disordered.
- There was a significant difference between normal and dysphonic groups in the number of vowels with normal bandwidth that fell in the standard vowel quadrilateral. This measure was a significant predictor for intelligibility at noise level of SNR+0. This result supports the clinical observation that the dysphonic speech source is noisier, leading to formants with greater bandwidth, which may contribute to the intelligibility deficit.

Limitations:

- The study utilized a relatively small number of speakers and listeners. The intelligibility measure is subjective. The speech materials include asymmetric numbers of vowels.

Conclusions:

- A bandwidth measure-- the bandwidth of the first and second formants within a normal VSA ---based on automatic, landmark-based detection of vowel-related parameters is a potentially useful biomarker for the intelligibility deficit in dysphonic speech. Further studies with larger numbers of speakers and listeners, more comprehensive speech materials and objective measures of intelligibility are needed.

Methods

- **Speakers:** 18 speakers with dysphonic voice (6 adult females, 6 adult males, and 6 children between 6-12yo); 3 speakers with normal voice, one from each age/gender group. All speakers are native speakers of American English with no history of hearing loss, speech sound disorders, and neurological speech, language and voice disorders.
- **Speech Material:** Six sentences from Consensus Auditory Perception Evaluation of Voice (CAPE-V) – 1) The blue spot is on the key again; 2) How hard did he hit him; 3) We were away a year ago; 4) We eat eggs every Easter; 5) My mama made lemon muffins; 6) Peter will keep at the peak.
- **Recording of Speech Samples:** The speech samples were recorded in a sound proof booth using a unidirectional microphone (Neumann, TLM 103). The recordings were digitized at 44.1kHz with a solid state recorder (TASCAM SS-R200).
- **Listeners:** 45 native speakers of American English with normal hearing between 20 – 48 years of age (mean = 22.8, SD = 4.85).
- **Stimuli preparation:** Cafeteria noise was added to the speech samples at three noise levels (SNR-3, SNR+0 and SNR+3). Every dysphonic stimulus was paired with an anchor stimulus provided by the age/gender-matched normal speakers.
- **Stimuli presentation:** The listening experiment was conducted in a single-wall sound-treated booth. The stimuli were presented through a headphone (Senhheiser HD 380 pro) at an average output level of 65 dB SPL. All stimuli were presented twice in random order.
- **Intelligibility measurement:** 7-point Likert scale (0 = unable to understand any words; 7 = as intelligible as a model speaker)

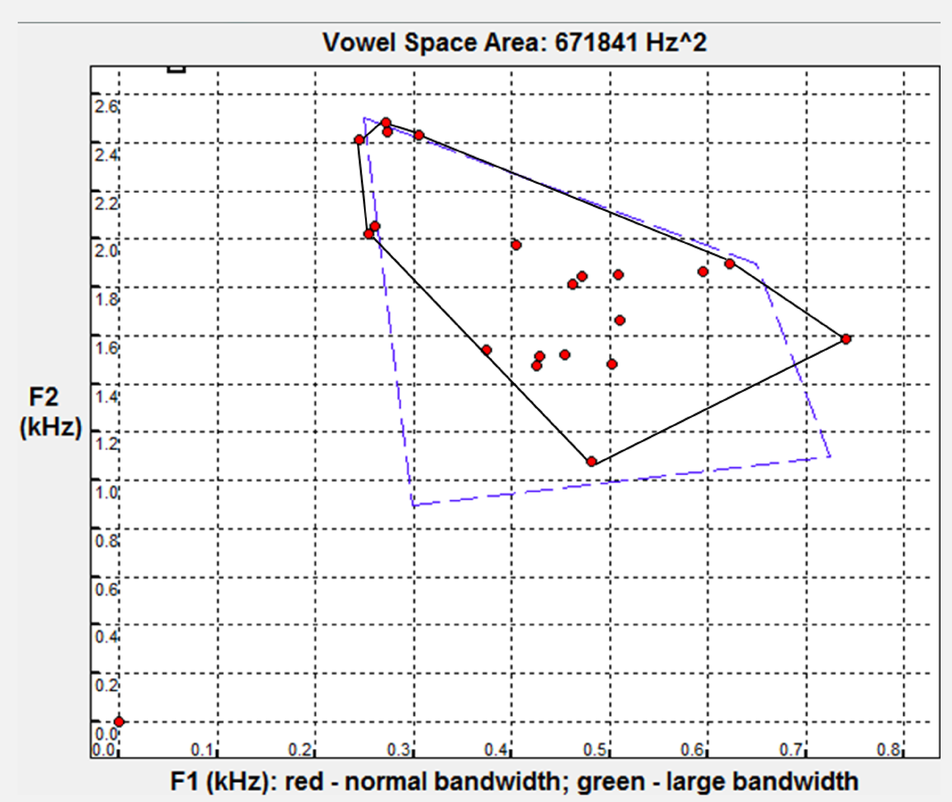


Figure 1: An example of vowel space area display for a normal adult male speaker. The quadrilateral with blue dashed lines shows the standard vowel quadrilateral based on values reported by previous literature.
Black line was drawn to show VSA of this sample. Vowels with normal and large bandwidth are shown in red and green, respectively. The bandwidth is considered large when it exceeds 1/6 of the central frequency (Fc/6).

References

1. Ishikawa, K., Boyce, S., Kelchner, L., Golla Powell, M., Schieve, H., de Alarcon, A., Khosla, S. (in press) The Effect of Background Noise on Intelligibility of Dysphonic Speech, *Journal of Speech Language Hearing Research.*
2. Cole R, Yan Y, Mak B, Fanty M, Bailey T (1996) The contribution of consonants versus vowels to word recognition in fluent speech. *Proceedings of International Conference on Acoustics, Speech and Signal Processing,* 853–856.
3. Kewley-Port D, Burkle TZ, Lee JH (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,* 122:2365–2375.
4. Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America,* 24(2), 175-184.
5. Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *The journal of the Acoustical Society of America,* 71(6), 1544-1550.
6. Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research,* 37(4), 769-778.
7. De Cheveigné, A. (1999). Formant bandwidth affects the identification of competing vowels. *ICPhS, 2093, 2096*.
8. Dubno, J. R., & Dorman, M. F. (1957). Effects of spectral flattening on vowel identification. The Journal of the Acoustical Society of America, 82(5), 1503-1511.
9. Weismer, G., Yi-Jeng, J., Laures, J. S., Kent, R. D., & Kent, intelligibility characteristics of sentence production in neurogenic speech disorders. Folia Phoniatrica et Logopaedica, 53, 1-18.
10. Boyce, S., Fell, H. J., & MacAuslan, J. (2012). SpeechMark: Landmark Detection Tool for Speech Analysis. In *INTERSPEECH* (pp. 1894-1897).

Acknowledgment

This work was funded by United States National Institutes of Health (NIH)grants R43 and R44 DC010104, R42 AG033523, and R42 HD34686 to S.T.A.R. Corp.

For more information and SpeechMark downloads (currently freeware in beta version): www.speechmrk.com