Utility of Formant-Based Measures for Describing Vowel Intelligibility in Dysphonic Speech

Keiko Ishikawa, M.M., Ph.D., CCC-SLP, Jossemia Webster, B.S.
Department of Speech and Hearing Science, College of Applied Health Sciences, University of Illinois at Urbana-Champaign

Voice and Speech Rehabilitation Research Lab

INTRODUCTION

Dysphonia negatively affects the intelligibility of a speaker, especially in noisy communication environments (Evitts, et al., 2016; Ishikawa, et al., 2017). While the intelligibility deficit is clinically well-recognized, there is no method for evaluating this aspect of dysphonia. Consequently, clinicians have not been able to determine the baseline performance of their patients nor keep track of their progress in improving intelligibility.

Correct perception of vowels is important for overall, sentence-level intelligibility (Kewley-Port, Burkle, & Lee, 2007). Speakers produce different vowels by changing the shape of their vocal tract. This articulatory difference creates energy peaks in the acoustic spectrum, called formants. Formants are characterized by their central frequency and bandwidth. It has been well-demonstrated that central frequency of the first and second formants (F1 and F2, respectively) are acoustic cues for listeners to differentiate one vowel from another (Peterson & Barney, 1952). Previous studies with synthesized speech have shown that widening of formant bandwidth reduces a listener's ability to identify the vowels presented with competing sounds (De Cheveigné, 1999).

Dysphonia is often caused by laryngeal pathology that leads to incomplete glottic closure. Contribution of the glottic closure to formant bandwidth has been theoretically implied (Steven, 1998). An acoustic-phonetic model by Stevens (1998) predicts that incomplete glottic closure increases the bandwidth of F1, and the effect can be especially substantial for low vowels. The wider bandwidth is accompanied by reduction in acoustic energy of F1, which makes the vowel less distinct in competing sounds.

The incomplete glottic closure by the laryngeal pathology may affect the formant bandwidth. If so, these formant-based measures may have the potential to be the acoustic markers for the unintelligibility in dysphonic speech.

PURPOSE

To examine the effect of dysphonia on formant-based measures, and correlation between the measurements and vowel intelligibility determined by listeners

METHODS

Subjects: 10 adult females with healthy voices, and 10 adult females with mild to severe dysphonia. All speakers were native speakers of American English and between 20 to 60 years of age.

Speech Samples:

- 11 vowels of American English in /h/-vowel-/d/ format
- Speech was recorded with a unidirectional microphone (Neumann, TLM 103) in a single-wall sound-proof room. The recordings were digitized at 44.1 kHz with a solid-state recorder (TASCAM SS-R200)
- The intensity of words were standardized, and the words were embedded into the carrier phase "I'm going to say ___."
- Cafeteria noise was added to these samples at signal-to-noise (SNR) -6 dB.

Listening Experiment:

- A total of 15 adults with normal hearing evaluated the identifiability of these vowels with and without background noise. Hearing status of these listeners was confirmed by pure-tone audiometry with tones presented at 500, 1,000, 2,000, 3,000 and 4,000 Hz at 25 dBA. All listeners were native speakers of American English.
- Listeners were asked to choose the word they heard by clicking the word on a computer screen.

Acoustic Analyses:

- Frequency and bandwidth for the vowels were measured using a semi-automatic speech analysis program, PRAAT.
- Acoustic measurements were done on vowel segments. The examiners visually determined vowels on a waveform and spectrogram, and manually selected a "steady-portion" of the vowel.

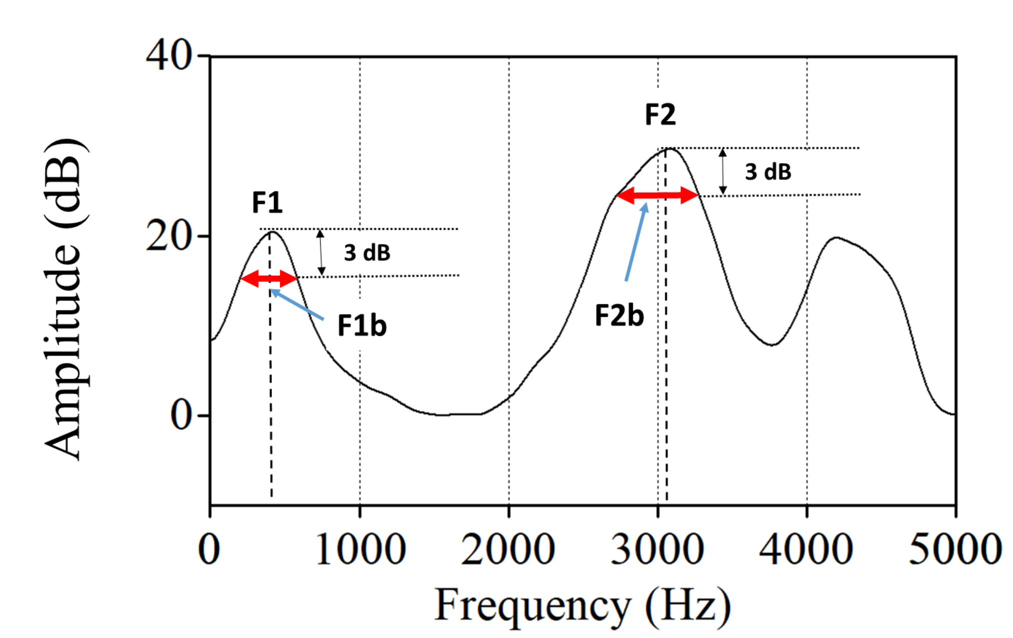

Figure 1. A schematic drawing showing measurement of formant frequency and bandwidth

RESULTS

- F1 frequency and bandwidth were significant predictors for dysphonia (p << 0.01). F2 frequency was approaching statistical significance (p = 0.054)

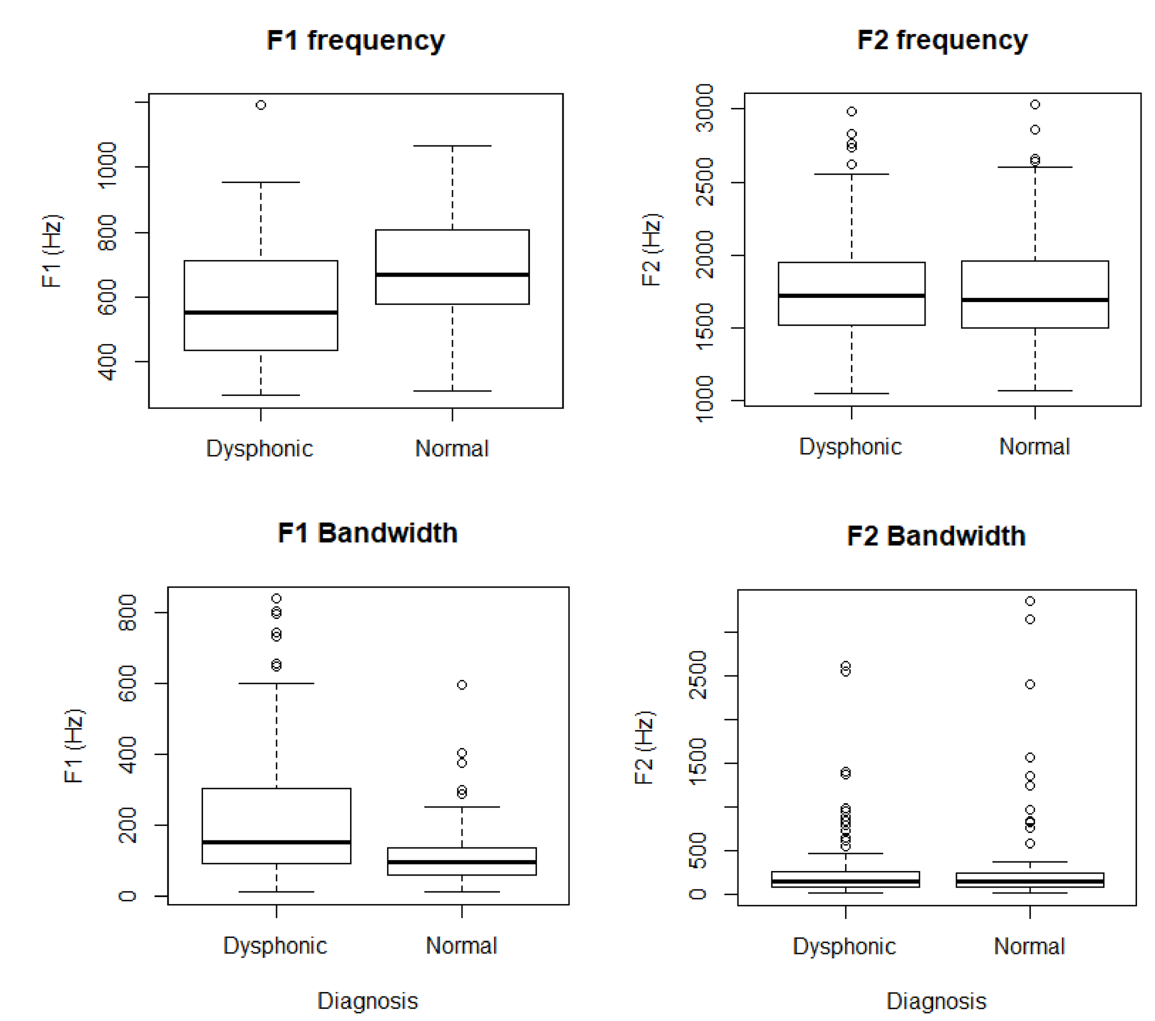

Figure 2. Box plots showing the difference in formant-based measurements between dysphonic and normal groups

- F1 frequency was moderately, positively correlated with vowel intelligibility (r_s = 0.66, p = 0.001)
- F2 frequency was not correlated with vowel intelligibility (p > 0.05)
- F1 bandwidth was negatively correlated with vowel intelligibility (r_s = -0.55 , p = 0.01)
- F2 bandwidth was not correlated with vowel intelligibility (p > 0.05)

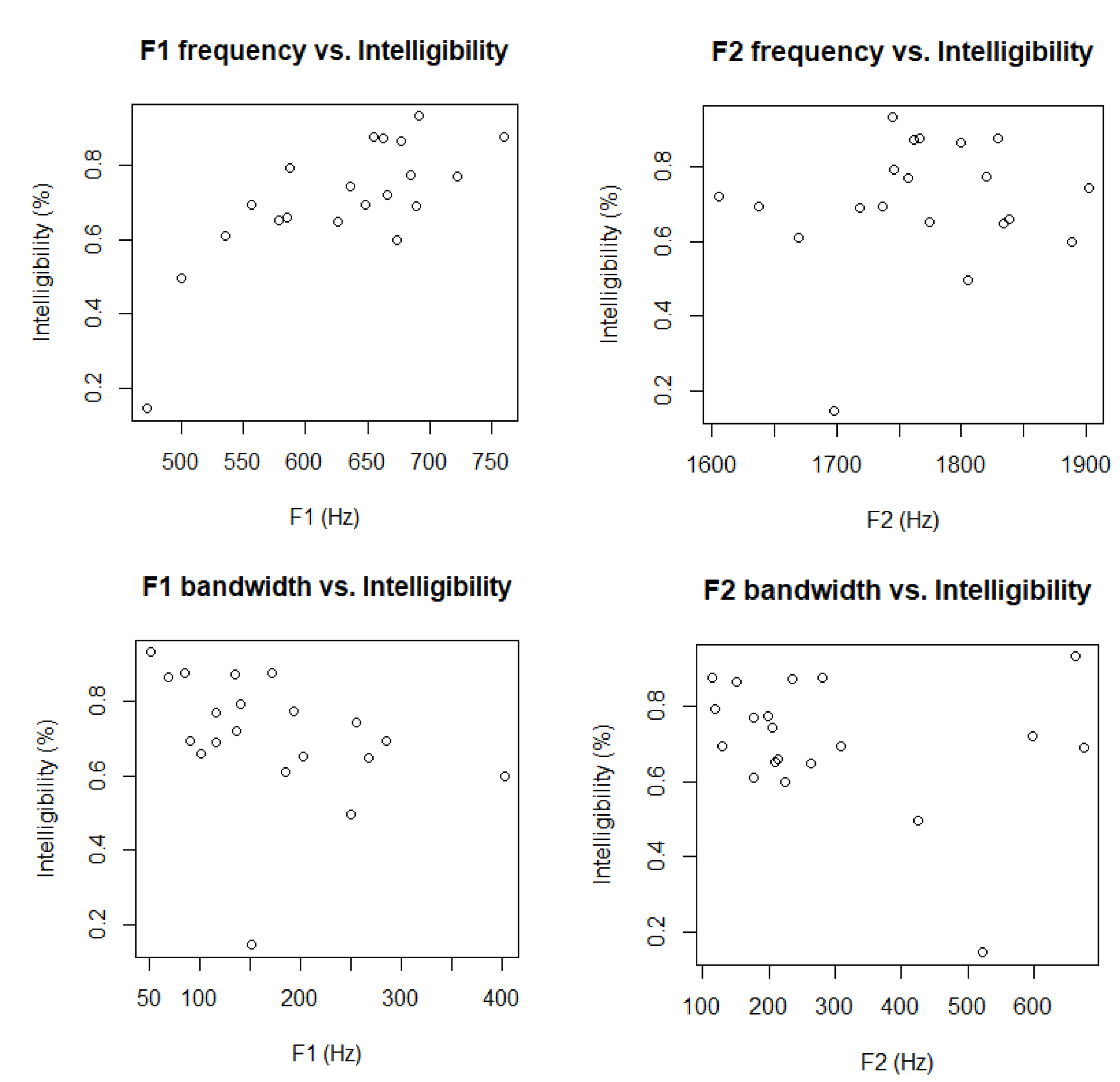

Figure 3. Scatter plots showing a correlation between formant-based measurements and vowel intelligibility

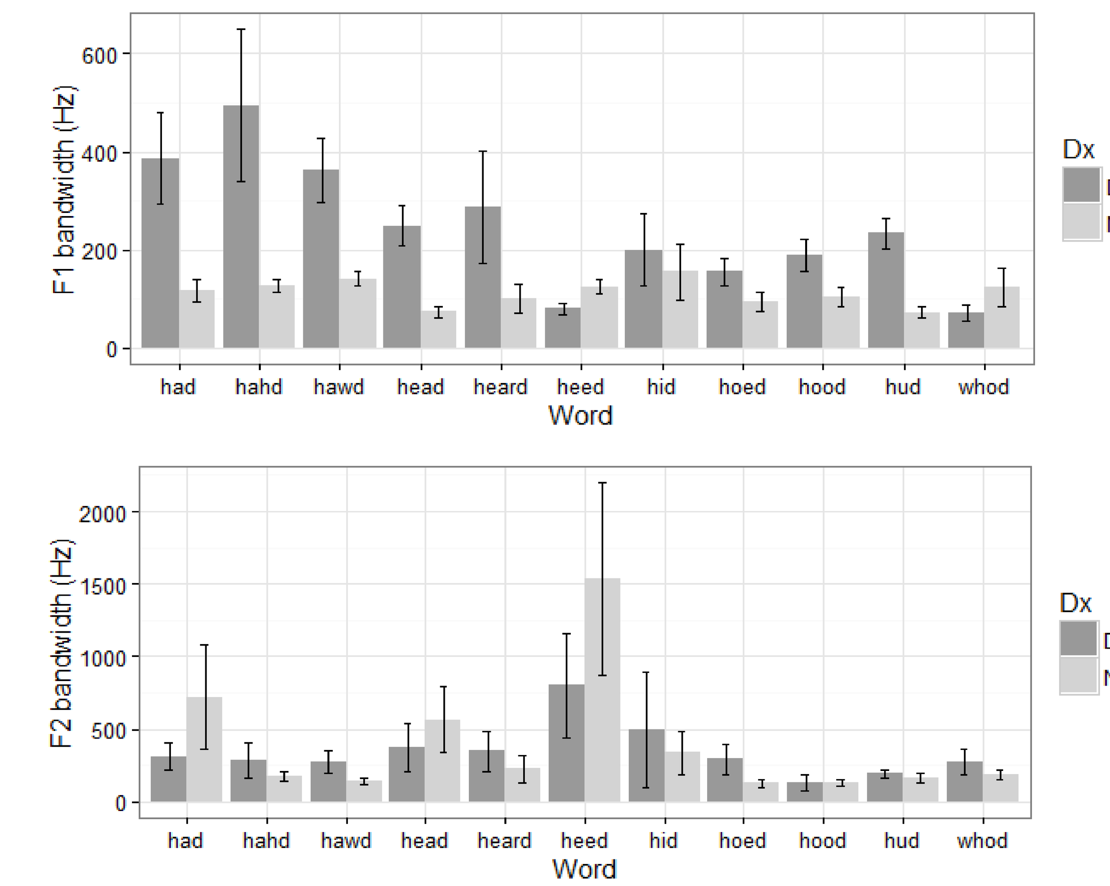

Figure 4. Bar plots showing the difference in formant-bandwidth measurements between dysphonic and normal groups for all vowels. Top: F1, Bottom: F2

DISCUSSION & CONCLUSIONS

F1 frequency and F1 bandwidth measurements may serve as acoustic markers for vowel intelligibility in dysphonic speech. The F1 frequency was significantly lower and F1 bandwidth was significantly wider in dysphonic group. As predicted by Steven's model, both of these acoustic changes may be attributed to the incomplete glottal closure.

The effect of dysphonia on F1 bandwidth was significantly greater for the lower vowels. Most of the acoustic energy at F1 is stored in the pharyngeal region of the vocal tract for these vowels. The incomplete glottic closure occurs immediately below this region, resulting in energy loss from the pharyngeal region. Ultimately, the energy loss leads to widening of the F1 bandwidth, reducing intelligibility of these vowels.

Praat uses Linear Predictive Coding for formant estimation, which may have produced errors in calculating the frequency and bandwidth (Kent & Vorperian, 2018). Manual analysis would be time-consuming but may yield different results.

References

Bless, D. (1987). Videostroboscopic evaluation of the larynx. *ENTJ*, *66*, 289.
Chiba, T., & Kajiyama, M. (1958). *The vowel: Its nature and structure* (Vol. 652). Tokyo: Phonetic society of Japan.
De Cheveigné, A. (1999). Formant bandwidth affects the identification of competing vowels. *ICPhS, 2093, 2096*.
Evitts, P. M., Starmer, H., Teets, K., Montgomery, C., Calhoun, L., Schulze, A., ... & Adams, L. (2016). The impact of dysphonic voices on healthy listeners: listener reaction times, speech intelligibility, and listener comprehension. *American journal of speech-language pathology*, *25*(4), 561-575.
Ishikawa, K., Boyce, S., Kelchner, L., Powell, M. G., Schieve, H., de Alarcon, A., & Khosla, S. (2017). The effect of background noise on intelligibility of dysphonic speech. *Journal of Speech, Language, and Hearing Research*, *60*(7), 1919-1929.
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.Kent, R. D., & Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press.
Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of communication disorders*, *74*, 74-97.

ILLINOIS