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

Keiko Ishikawa¹, Meredith Meyer², Joel MacAuslan² and Suzanne Boyce¹

University of Cincinnati, College of Allied Health, Dpt. of Communication Sciences and Disorders¹; Speech Technology & Applied Research²

---

**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.⁷
- Formant bandwidth can affect intelligibility.⁷⁻⁸ Listeners may not be able to identify vowels correctly with formants that have abnormally large bandwidths.
- 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 dysphonic 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. At instances where the maximum harmonic power is reached are identified as vowels. Center frequencies of the first two formants are calculated and used to generate VSA.¹⁰⁻¹¹

**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]. Each 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 (Sennheiser 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)