Acoustic speech analysis of patients with decompensated heart failure: A pilot study

Olivia M. Murton,[a] **Robert E. Hillman,[b]** **and Daryush D. Mehta[b]**
Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital,
Boston, Massachusetts 02114, USA
omurton@g.harvard.edu, hillman.robert@mgh.harvard.edu,
mehta.daryush@mgh.harvard.edu

Marc Semigran,[c] **Maureen Daher, Thomas Cunningham,**
Karla Verkouw, Sara Tabtabai, Johannes Steiner, G. William Dec,
and Dennis Ausiello
Institute for Heart, Vascular and Stroke Care, Massachusetts General Hospital, Boston,
Massachusetts 02114, USA
msemigran@gmail.com, mdaher@mgh.harvard.edu, tfcunningham@mgh.harvard.edu,
kverkouw@gmail.com, stabtabai@gmail.com, jksteiner@mgh.harvard.edu,
gdec@partners.org, ausiello.dennis@mgh.harvard.edu

Abstract: This pilot study used acoustic speech analysis to monitor patients with heart failure (HF), which is characterized by increased intracardiac filling pressures and peripheral edema. HF-related edema in the vocal folds and lungs is hypothesized to affect phonation and speech respiration. Acoustic measures of vocal perturbation and speech breathing characteristics were computed from sustained vowels and speech passages recorded daily from ten patients with HF undergoing inpatient diuretic treatment. After treatment, patients displayed a higher proportion of automatically identified creaky voice, increased fundamental frequency, and decreased cepstral peak prominence variation, suggesting that speech biomarkers can be early indicators of HF.
© 2017 Acoustical Society of America
[DDO]

1. Introduction

Heart failure (HF) is a large and growing concern, consuming significant clinician time and major healthcare expenditures. Preventing the escalation of existing disease is critical to avoiding frequent hospitalizations, lowering healthcare costs, and reducing patient mortality. The 30-day readmission rate after a HF-related hospitalization is 21.9%,[1] increasing to over 50% within six months.[2] There are over 5.3×10^6 HF patients in the U.S., with 550 000 new cases diagnosed annually and almost 300 000 deaths each year.[3]

Patients with HF can be stable for long periods of time. However, progression of the underlying disease or other changes in health can cause those compensatory strategies to fail. This loss of stability is known as *decompensated* HF and is an acute crisis requiring immediate intervention.[4] The course of a patient's disease may include multiple episodes of decompensation separated by stable periods of varying duration.[2] Thus, a major goal of HF management is to maintain stability by predicting and preventing episodes of decompensation.

A hallmark of HF is the presence of *edema*, which is characterized by swelling from fluid retention in body tissues. Edema is typically found in the legs and feet, but can also occur in the lungs and throughout the body. Leading HF specialists have emphasized the importance of managing edema in patients and the need for technologies that identify edema early.[4] As decompensation approaches, HF-related edema causes enough of a weight increase—often termed *volume overload*—to be detectable with regular weighing, which patients can do easily and non-invasively at home.[5] Weight is used for first-line monitoring of volume status and as an early warning for decompensation. However, edema-related weight increases occur relatively late in the disease progression and may not be detected in time to prevent an episode of decompensation.[5] There are also surgically implantable devices which may predict impending

[a]Also at: Speech and Hearing Bioscience and Technology, Division of Medical Sciences, Harvard Medical School, Boston, MA 02115, USA.
[b]Also at: Department of Surgery, Harvard Medical School, Boston, MA 02115, USA.
[c]Also at: Department of Medicine, Harvard Medical School, Boston, MA 02115, USA.

decompensation by monitoring pressures within the heart and lungs.[6] However, these monitors are expensive and invasive. There is currently an unmet need for a reliable signal that can be identified earlier than weight increase, but can be measured just as easily at home by patients or caregivers.

Voice has the potential to provide an easily obtained, non-invasive way to monitor physiological changes throughout the body, as long as those changes also affect the larynx.[7] In particular, the vocal folds consist of thin tissue layers that might be particularly sensitive to HF-related edema. For example, Verdolini *et al.* found that a dose of the diuretic Lasix (furosemide), which is widely used to treat decompensated HF, induced a 23% increase in the phonation threshold pressure in healthy adults.[8] Since the amount of edema required to measurably change the voice is small (whereas the amount needed to increase body weight is large), voice monitoring may allow us to detect and track HF-related edema at an earlier stage than weight does.

In this pilot study, we analyzed the voices of HF patients as they underwent treatment for decompensated HF and returned to a stable clinical state. Decompensation is treated with diuretics, which cause patients to rapidly lose fluid and decrease their edema until they reach their *target weight* (typical body weight without extra fluid). This decrease in volume overload during treatment mirrors the increasing edema that occurs during the asymptomatic pre-decompensation phase. Similarly, voice changes during treatment may also mirror corresponding changes before decompensation. By looking for changes in voice that correlate with in-hospital improvements, we can use this new knowledge to direct future studies that may predict decompensation in stable HF patients at home. Our approach is based on hypotheses about how voice and speech production may be impacted by HF pathophysiology. This approach is fundamentally different from studies with similar goals that have used traditional machine learning methods to analyze voice acoustic signals, because such investigations have often lacked hypotheses and theories related to the underlying physical systems.[9]

The goal was to determine whether measurable changes in voice occur during treatment for acute decompensated HF. Specifically, we hypothesized that the fluid retention, dyspnea, and fatigue associated with HF will cause increased edema and surface dehydration of the vocal folds' phonatory mucosa and reduced respiratory support for speech production, resulting in increased phonatory irregularity, decreased fundamental frequency, decreased durations of speech breath groups, and increased pausing.

2. Methods

Ten patients (8 male, 2 female) with acute decompensated HF were enrolled in the study with a mean $\pm$ standard deviation (SD) age of 70 ± 13 years. Inclusion criteria called for patients to be at least 4.5 kg above their target weight on admission and expected to need diuretics for over 48 h. The average length of stay was 7 ± 3 days during which weight was measured each day and the average total weight loss was 8.5 ± 5 kg. Blood levels of N-terminal pro-brain b-type natriuretic peptide (NT-proBNP) were tested at the beginning and end of each patient's hospitalization since high levels of NT-proBNP have been associated with HF.[10] The average change (last minus first measurement) in NT-proBNP level was -1379 ± 3100 pg/mL. Patients also used visual analog scales to evaluate their dyspnea and global symptoms from 0 (worst) to 100 (best). Average changes in dyspnea and global symptom rating were 8.4 ± 21 and 22 ± 30, respectively.

Each day, patients recorded a standard speech protocol consisting of four different utterance types: sustained vowels, CAPE-V sentences,[11] the Rainbow Passage,[12] and spontaneous speech. Recording sustained vowels allows us to evaluate a patient's ability to produce stable, regular phonation. The other tasks, which elicit connected speech, are designed to evaluate voice characteristics under more natural speaking conditions and provide information about prosody and speech breathing that cannot be recovered from short non-linguistic vowel productions. The CAPE-V sentences and the Rainbow Passage are widely used in voice research to elicit various speech contexts, including soft/hard glottal attacks, vowel coarticulation, and nasality. Both an acoustic microphone (H1 Handy Recorder, Zoom Corporation, Tokyo, Japan) and neck-mounted accelerometer (BU-27135; Knowles Corp., Itasca, IL) were used to record the speech tasks.[13] This paper presents analysis of the microphone signal, which was sampled at 44.1 kHz and 16-bit quantization. All acoustic recordings were high-pass filtered at 70 Hz to remove low-frequency noise artifacts. We then computed the change in voice and speech measures from the first recording session to the last.

Each sustained vowel recording was divided into overlapping 1-s segments starting 100 ms apart. PRAAT[14] was used to extract the 1-s segment with the lowest

percent jitter (degree of cycle-to-cycle variation in glottal pulse spacing) that was used for all further vowel analyses. The segment with the lowest jitter was selected to compare each patient's best voice sample and to minimize artifacts from environmental noise or irregular voicing at onsets and offsets.

The fundamental frequency (F0) was extracted using Praat's cross-correlation pitch-tracking method (40-ms Hanning window every 3.3 ms). For each vowel utterance, the F0 track was used to generate a PRAAT voice report including mean F0, median F0, F0 SD, and jitter percent. Following Awan and Roy, a low-high spectral energy ratio was calculated from the spectrum as the band energy difference, in dB, between 0 and 4 kHz and 4 and 10 kHz to reflect perceptual ratings of dysphonia severity.[15] The PRAAT pitch track was also used to calculate mean F0, median F0, and F0 SD for each Rainbow Passage recording. To minimize potential distortion from pitch tracking artifacts, F0 measures were only calculated over frames with F0 between the 5th and 95th percentile for each recording.

Cepstral peak prominence (CPP) and CPP SD were calculated for both vowel and Rainbow Passage utterances following Awan *et al.*[16] CPP was computed as the difference, in dB, between the magnitude of the highest peak and the noise floor in the power cepstrum (window length 40.96 ms, computed every 10.24 ms). The location of the CPP was limited to quefrencies between 3.3 and 16.7 ms (equivalent to F0 between 60 and 300 Hz). For Rainbow Passage utterances, CPP was only calculated on voiced frames (those with a non-zero F0 in the PRAAT pitch track). These measures were used to test our hypothesis that, pre-treatment, vocal fold edema would cause increased acoustic perturbation and decreased F0.

The probability of creak was computed as a measure of glottal cycle irregularity related to vocal fry for sentence and Rainbow Passage recordings. This method uses short-term power contours, intra-frame periodicity, and inter-pulse similarity,[17] as well as measures indicating secondary or widely spaced glottal pulses as inputs to an artificial neural network that generates frame-by-frame creak probabilities.[18] Creak percent for an utterance was given by (# creaky frames)/(# voiced frames) × 100, where # voiced frames was the sum of the number of frames considered voiced by PRAAT and the number of frames with a very high creak probability (above 0.8). Adding high-probability creak frames accounted for frames with highly irregular phonation—often on the margins of voiced regions—where PRAAT's pitch tracking did not identify an underlying F0. Creaky frames were defined as voiced frames with creak probability above 0.3. The creak threshold of 0.3 was based on an iterative process involving perceptual evaluation of the acoustic signal and balanced the requirement of including frames with clear creaky voice quality and minimizing artifacts from regions with low creak probability (pseudo-periodic signal).

Rainbow Passage readings were transcribed and aligned to sound files using the Penn Phonetics Lab Forced Aligner.[19] Intra-passage pause durations longer than 500 ms were assumed to exhibit an inhalation; thus, a speech breath group was defined between successive inhalations, excluding the pause duration. Respiratory-related measures within each Rainbow Passage reading included total breath group duration (sum of the durations of each breath group); breath group duration SD (SD of all the breath group durations); and mean, max, and SD of the number of phonemes per breath group (based on the phoneme counts given by the forced aligner). Unlike the measures described earlier, which evaluated phonatory irregularity, these measures were used to test the effects of edema on speech breathing support.

3. Results

Since the goal was to determine voice measures that correlated with cardiac status, Table 1 lists acoustic speech measures that trended in the same direction for a majority of patients. Creak percent increased after treatment in the sentence and Rainbow Passage conditions for eight and nine patients, respectively. Total breath group duration within the Rainbow Passage decreased, and measures related to the number of phonemes per breath group increased, for seven of ten patients. Measures of acoustic irregularity (sustained-vowel F0 SD, jitter percent, and CPP SD) decreased for most patients. Median F0 increased for a majority of patients in the Rainbow Passage condition (but not the sustained-vowel task). Most patients required less speaking time to read the Rainbow Passage after treatment and produced more phonemes per phrase. These results indicate that patients spoke faster or breathed less frequently after HF treatment (potential effects of increased familiarity with the reading passage require further investigation). As

Table 1. Summary of measures for which seven or more patients changed in the same direction during hospitalization. Includes first-day mean values across all ten patients, means and ranges of changes from first to last day, and *p*-values from paired, two-tailed t-tests comparing first and last day values.

Measure name	Utterance type	Majority change	Day 1 mean	Mean change	Range of change	*p*
Creak (%)	passage	9 +	3.0	6.9	[−0.69, 33]	0.065
F0 SD (Hz)	vowel	8 −	3.6	−0.95	[−6.5, 1.2]	0.20
Jitter (%)	vowel	8 −	0.88	−0.25	[−1.5, 0.25]	0.18
CPP SD (dB)	vowel	8 −	2.5	−0.41	[−1.1, 0.15]	0.022[a]
Creak (%)	sentence	8 +	3.6	5.5	[−0.75, 19]	0.031[a]
F0 SD (Hz)	passage	8 +	16	2.4	[−24, 16]	0.50
Low-high spectral ratio	vowel	7 +	25	1.8	[−8.7, 12]	0.39
Median F0 (Hz)	passage	7 +	141	3.7	[−24, 20]	0.38
Total breath group duration (s)	passage	7 −	35	−1.9	[−9.8, 1.5]	0.11
Mean phonemes per phrase	passage	7 +	20	1.1	[−7.9, 12]	0.49
Max phonemes per phrase	passage	7 +	41	1.7	[−33, 16]	0.73
SD phonemes per phrase	passage	7 +	9.9	0.89	[−7.7, 6.4]	0.48

[a]$p < 0.05$.

hypothesized, these preliminary results indicate that most patients showed increased irregularity and decreased F0 at admission compared to discharge.

Figure 1 displays each patient's creak percent on their first and last day of treatment. Creak percent was higher on the last day than the first for nine patients in the Rainbow Passage condition (mean change 6.9 pp, range −0.69 to 33 pp, $p = 0.065$) and eight patients on the CAPE-V sentence condition (mean change 5.5 pp, range −0.75 to 19 pp, $p = 0.031$).

Figure 2, which is based on data from Rainbow Passage recordings, shows median F0 increasing in seven patients and F0 SD increasing in eight. The mean change in median F0 was 3.7 Hz (0.54 semitones), whereas the mean change in F0 SD was 2.4 Hz (0.23 semitones).

Figure 3 is based on data from sustained vowel recordings. The left panel shows that CPP increased in six of ten patients. However, the CPP increased for all four patients who began with a comparatively low CPP (<20 dB). The mean change among these patients was 3.9 dB, compared to −0.7 dB among those who started with CPP >20 dB (indicating good voice quality to begin with). The right panel shows decreases in sustained-vowel F0 SD for eight of ten patients after treatment. The mean change in F0 SD was −1 Hz, with one patient having an F0 SD 6.5 Hz lower on their last day.

4. Discussion

This pilot analysis studied patients who were being treated for acute decompensated HF and monitored their voices as their volume status improved. The ultimate goal, though, is the reverse: to monitor stable HF patients and identify impending decompensation before it becomes a crisis. Our main purpose in this pilot project was to demonstrate that voice features have the potential to reflect HF status and may have use in at-home or continuous ambulatory monitoring contexts.

Voice measures examined can be divided into two categories based on their response to treatment. The first and most promising category for future analysis includes measures for which most patients exhibited large changes in the same direction. Creak percent increased for most patients in both the Rainbow Passage and
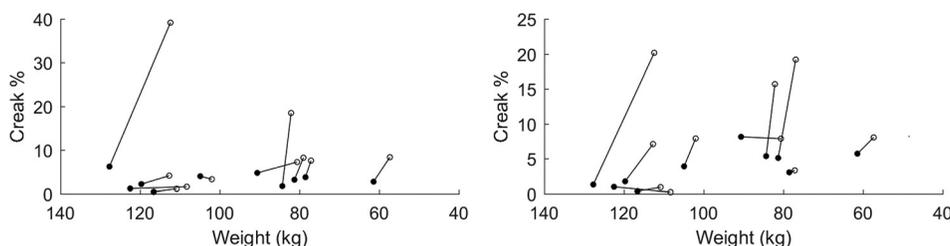

Fig. 1. Creak percent versus weight for each patient's Rainbow Passage (left) and CAPE-V sentences (right), with first-day (filled) and last-day (open) values. Note the horizontal axis shows weight decreasing toward the right, indicating direction of HF status improvement.
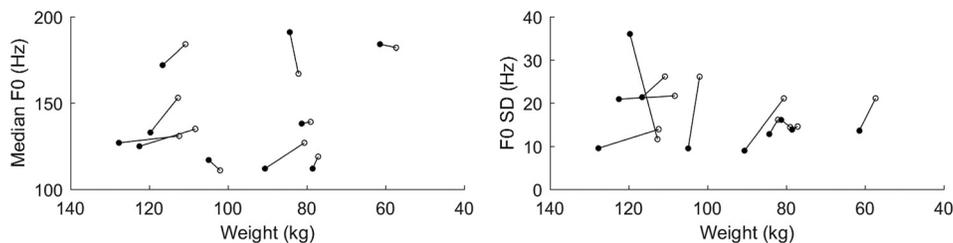
Fig. 2. Median (left) and SD (right) F0 across voiced frames for each patient's Rainbow Passage.

sentence conditions, with large effect sizes (mean change 6.9 and 5.5 pp, respectively, relative to starting values of 3.0% and 3.6%). Also, median F0 in the Rainbow Passage task increased after treatment for seven of the ten patients. In the sustained-vowel condition, CPP SD decreased for eight of ten patients with another fairly large effect size (mean change −0.4 dB relative to a mean starting value of 2.5 dB).

The second category of measures includes those for which most patients saw large changes due to treatment, but not all in the same direction. For example, median F0, maximum and mean phonemes per breath group (in the Rainbow Passage) and low-high spectral ratio (in the vowel condition) had small average changes relative to baseline because large increases in some patients were cancelled out by large decreases in others. These measures may be useful for HF monitoring in future if we can understand what causes a given patient's voice to change in a given direction.

Creaky voice quality can be associated with a pathological voice condition.[20] However, unlike most other voice quality measures, creak is also partially governed by the linguistic system, and some amount of creak can be normal in any voice.[21] It is possible that HF-related edema in the vocal folds affects the phonatory mechanism by reducing the capability of the vocal folds to become slack or flaccid (necessary for creaky voice production), which decreases the likelihood of creaky voice. Further examination of the distributions of creak in reading passages would allow us to see where creaky regions are relative to linguistically expected ones.

The direction of CPP change during treatment was not uniform; six of ten patients had higher CPP at discharge than at admission. However, four patients began their hospitalization with CPPs below 20 dB, and all improved by discharge. Our eventual goal is to predict decompensation in stable patients, rather than observing the return to stability after decompensation. Our CPP results indicate that, if most stable patients have high CPP, some patients will have lower CPPs during decompensation and some will not. Reduced CPP might be a sign of increasing fluid volume in some HF patients who exhibit a large change.

A potential confound for acoustic voice analysis is that the voice production mechanism can compensate for physiological vocal fold changes, such as increased edema. Therefore, changes in voice output may not be directly related to the amount of edema in the vocal folds since compensatory mechanisms allow talkers to produce voices of similar quality even under different physiological conditions.[22] In this population, direct visualization of edema in the larynx via laryngoscopy is not feasible due to their comprised medical condition and increased fragility.

Additionally, seven of the ten patients in this study were 65 years of age or older. Older patients may have aging-related deterioration of vocal function that is unrelated to HF. For example, a common cause of degraded voice quality in older individuals is presbyphonia, which is described as vocal fold bowing (incomplete closure) secondary to loss of tissue and/or diminished laryngeal neuromotor control.[23]
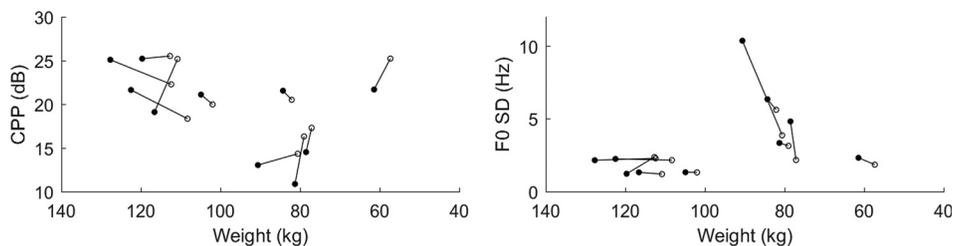

Fig. 3. Mean CPP (left) and F0 SD (right) vs weight across all sustained-vowel utterances in each patient's first- and last-day recordings.

HF-related edema might actually improve voice quality by bulking up the vocal folds and facilitating better contact and closure (reduced bowing) during phonation.

5. Conclusion

Reliable methods of monitoring fluid overload in stable HF patients are critically important for preventing decompensation. Voice monitoring has the potential to provide a non-invasive, easily obtained way for patients to track their own status. Several measures of voice quality, including creak percent, F0, and CPP SD, may correlate well with improvements in HF symptoms during decompensation treatment.

Acknowledgments

References and links

[1]Centers for Medicare and Medicaid Services, "Readmissions and deaths—National" [data file], https://data.medicare.gov/d/qqw3-t4ie (Last viewed 4/2/2017).

[2]A. S. Desai and L. W. Stevenson, "Rehospitalization for heart failure," Circulation **126**(4), 501–506 (2012).

[3]C. M. O'Connor, W. G. Stough, D. S. Gallup, V. Hasselblad, and M. Gheorghiade, "Demographics, clinical characteristics, and outcomes of patients hospitalized for decompensated heart failure: Observations from the IMPACT-HF registry," J. Cardiac Failure **11**(3), 200–205 (2005).

[4]M. Gheorghiade, F. Zannad, G. Sopko, L. Klein, I. L. Piña, M. A. Konstam, B. M. Massie, E. Roland, S. Targum, S. P. Collins, G. Filippatos, and L. Tavazzi, "Acute heart failure syndromes," Circulation **112**(25), 3958–3968 (2005).

[5]S. M. Joseph, A. M. Cedars, G. A. Ewald, E. M. Geltman, and D. L. Mann, "Acute decompensated heart failure," Texas Heart Inst. J. **36**(6), 510–520 (2009), available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2801958/.

[6]W. T. Abraham, P. B. Adamson, R. C. Bourge, M. F. Aaron, M. R. Costanzo, L. W. Stevenson, W. Strickland, S. Neelagaru, N. Raval, S. Krueger, S. Weiner, D. Shavelle, B. Jeffries, and J. S. Yadav, "Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: A randomised controlled trial," Lancet **377**(9766), 658–666 (2011).

[7]J. H. Van Stan, D. D. Mehta, and R. E. Hillman, "Recent innovations in voice assessment expected to impact the clinical management of voice disorders," Perspectives of the ASHA Special Interest Groups, SIG 3 (2017), Vol. 2, 4–13.

[8]K. Verdolini, Y. Min, I. R. Titze, J. Lemke, K. Brown, M. Van Mersbergen, J. Jiang, and K. Fisher, "Biological mechanisms underlying voice changes due to dehydration," J. Speech Lang. Hear. Res. **45**(2), 268–281 (2002).

[9]E. Maor, J. D. Sara, L. O. Lerman, and A. Lerman, "The sound of atherosclerosis: Voice signal characteristics are independently associated with coronary artery disease," Circulation **134**, A15840 (2016), available at http://circ.ahajournals.org/content/134/Suppl_1/A15840.

[10]Q. Dao, P. Krishnaswamy, R. Kazanegra, A. Harrison, R. Amirnovin, L. Lenert, C. Clopton, J. Alberto, P. Hlavin, and A. S. Maisel, "Utility of B-type natriuretic peptide in the diagnosis of congestive heart failure in an urgent-care setting," J. Am. College Cardiol. **37**(2), 379–385 (2001).

[11]G. B. Kempster, B. R. Gerratt, K. Verdolini Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol," Am. J. Speech Lang. Pathol. **18**(2), 124–132 (2009).

[12]G. Fairbanks, "The rainbow passage," in *Voice and Articulation Drillbook* (1960), Vol. 2.

[13]D. D. Mehta, J. H. Van Stan, and R. E. Hillman, "Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer," IEEE/ACM Trans. Audio Speech Lang. Process. **24**(4), 659–668 (2016).

[14]P. Boersma and D. Weenink, "Praat: Doing phonetics by computer" [computer program], version 5.4.17, http://www.praat.org/ (Last viewed 8/28/2015).

[15]S. N. Awan and N. Roy, "Toward the development of an objective index of dysphonia severity: A four-factor acoustic model," Clin. Ling. Phon. **20**(1), 35–49 (2006).

[16]S. N. Awan, N. Roy, M. E. Jetté, G. S. Meltzner, and R. E. Hillman, "Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V," Clin. Ling. Phon. **24**(9), 742–758 (2010).

[17]C. T. Ishi, K. I. Sakakibara, H. Ishiguro, and N. Hagita, "A method for automatic detection of vocal fry," IEEE Trans. Audio Speech Lang. Process. **16**(1), 47–56 (2008).

[18]J. Kane, T. Drugman, and C. Gobl, "Improved automatic detection of creak," Comput. Speech Lang. **27**(4), 1028–1047 (2013).

[19]J. Yuan and M. Liberman, "Speaker identification on the SCOTUS corpus," J. Acoust. Soc. of Am. **123**(5), 3878 (2008).

[20]E. B. Holmberg, R. E. Hillman, B. Hammarberg, M. Södersten, and P. Doyle, "Efficacy of a behaviorally based voice therapy protocol for vocal nodules," J. Voice **15**(3), 395–412 (2001).

[21]M. Gordon and P. Ladefoged, "Phonation types: A cross-linguistic overview," J. Phon. **29**(4), 383–406 (2001).

[22]R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, "Objective assessment of vocal hyperfunction: An experimental framework and initial results," J. Speech Lang. Hear. Res. **32**(2), 373–392 (1989).

[23]P. B. Mueller, "The aging voice," in *Seminars in Speech and Language* (Thieme Medical, Stuttgart, Germany, 1997), Vol. 18, No. 02, pp. 159–169.