Comparison of voice relative fundamental frequency estimates derived from an accelerometer signal and low-pass filtered and unprocessed microphone signals

Yu-An S. Lien[a)]
Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Cara E. Stepp
Departments of Speech, Language, and Hearing Sciences and Biomedical Engineering, Boston University, Boston, Massachusetts 02215

(Received 24 September 2013; revised 18 February 2014; accepted 25 March 2014)

The relative fundamental frequency (RFF) surrounding the production of a voiceless consonant has previously been estimated using unprocessed and low-pass filtered microphone signals, but it can also be estimated using a neck-placed accelerometer signal that is less affected by vocal tract formants. Determining the effects of signal type on RFF will allow for comparisons across studies and aid in establishing a standard protocol with minimal within-speaker variability. Here RFF was estimated in 12 speakers with healthy voices using unprocessed microphone, low-pass filtered microphone, and unprocessed accelerometer signals. Unprocessed microphone and accelerometer signals were recorded simultaneously using a microphone and neck-placed accelerometer. The unprocessed microphone signal was filtered at 350 Hz to construct the low-pass filtered microphone signal. Analyses of variance showed that signal type and the interaction of vocal cycle × signal type had significant effects on both RFF means and standard deviations, but with small effect sizes. The overall RFF trend was preserved regardless of signal type and the intra-speaker variability of RFF was similar among the signal types. Thus, RFF can be estimated using either a microphone or an accelerometer signal in individuals with healthy voices. Future work extending these findings to individuals with disordered voices is warranted. © *2014 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4870488]

PACS number(s): 43.70.Jt, 43.70.Gr [CYE] Pages: 2977–2985

I. INTRODUCTION

Relative fundamental frequency (RFF) is a measure estimated using the instantaneous fundamental frequencies (*F*0s) in a sonorant—voiceless consonant—sonorant instance (RFF instance). Specifically, RFF is defined as the ten instantaneous *F*0s preceding and subsequent to a voiceless consonant, normalized to nearby steady state *F*0 in semitones (e.g., upper panel of Fig. 1).

RFF has been of interest to a variety of different research studies (Watson, 1998; Robb and Smith, 2002; Goberman and Blomgren, 2008; Stepp *et al.*, 2010; Stepp and Eadie, 2011; Stepp *et al.*, 2011; Stepp *et al.*, 2012; Eadie and Stepp, 2013; Stepp, 2013). Several of these investigated the characteristics of RFF in populations with voice disorders, including vocal hyperfunction (VH) and Parkinson's disease (PD), and found that individuals with voice disorders have RFF that differs from those of individuals with healthy voices (Goberman and Blomgren, 2008; Stepp *et al.*, 2010; Stepp *et al.*, 2012; Stepp, 2013).

The characteristic pattern of RFF in young speakers with healthy voices is a stable [near 0 semitone (ST)] or slightly decreasing (as a function of cycle) RFF prior to the consonant and decreasing RFF after the consonant (Watson, 1998; Robb and Smith, 2002). For the last cycle preceding the voiceless consonant (cycle 10 of the offset vowel) and the first cycle following the consonant (cycle 1 of the onset vowel), mean RFF values are −0.84 to 0.44 ST and 2.3 to 2.8 ST, respectively (Watson, 1998; Robb and Smith, 2002). However, for speakers with VH and PD, both offset and onset RFF tend to be lower compared to age-matched controls (Goberman and Blomgren, 2008; Stepp *et al.*, 2010; Stepp, 2013). In individuals with VH prior to voice therapy (Stepp *et al.*, 2010) and individuals with PD while off medication (Goberman and Blomgren, 2008; Stepp, 2013), mean RFF values for offset cycle 10 and onset cycle 1 are −1.0 to −2.2 ST and 1.8 to 2.7 ST, respectively.

The difference in RFF between individuals with healthy voices and those with disordered voices (VH or PD) has been hypothesized to be caused by differences in baseline laryngeal tension (Goberman and Blomgren, 2008; Stepp *et al.*, 2011; Stepp, 2013). A previous study has shown that the activity of the cricothyroid muscle increases preceding or during the voiceless consonant and decreases immediately after (Lofqvist *et al.*, 1989). An increase in cricothyroid muscle activity is associated with an increase in laryngeal tension, which consequently leads to increased *F*0 (Arnold, 1961; Roubeau *et al.*, 1997). Individuals with VH and PD are thought to have excessive laryngeal tension, i.e., higher baseline tension (Berardelli *et al.*, 1983; Hillman *et al.*, 1989; Roy *et al.*, 1996; Gallena *et al.*, 2001). Thus, their ability to

[a)]Author to whom correspondence should be addressed. Electronic mail: slien@bu.edu

use changes in tension during devoicing and revoicing may be limited due to a ceiling effect, and this decreased ability to modulate laryngeal tension could explain the lowered RFF seen in these individuals (Goberman and Blomgren, 2008; Stepp *et al.*, 2011; Stepp, 2013).

The difference in RFF between individuals with healthy voices and individuals with VH suggests that RFF may be adapted for a clinical assessment of VH, a condition defined as the "abuse and/or misuse of the vocal mechanism due to excessive and/or 'imbalanced' muscular forces" (Hillman *et al.*, 1989, p. 373). Current clinical assessment of VH primarily depends on clinicians' subjective interpretations based on auditory and visual perception, patient history, palpation of neck musculature, and patient report of self-perceived fatigue or discomfort (Morrison *et al.*, 1986; Roy *et al.*, 1996; Behrman, 2005). RFF is a promising measure for *objective* assessment of VH. In fact, in individuals with spasmodic dysphonia, onset cycle 1 RFF values have been shown to significantly correlate with listeners' perception of vocal effort (Eadie and Stepp, 2013), a primary subjective diagnostic indicator. However, for these individuals, the current protocol for RFF estimation requires at least six RFF instances to attain a stable estimate (Eadie and Stepp, 2013), which is not currently feasible for inclusion into clinical protocols due to the time-consuming nature of manual RFF estimation. The number of RFF instances necessary may be reduced with the use of optimized signals resulting in more reliable RFF estimation. This would reduce the time required for RFF estimation, allowing RFF to be implemented in clinical protocols for objective assessment of VH.

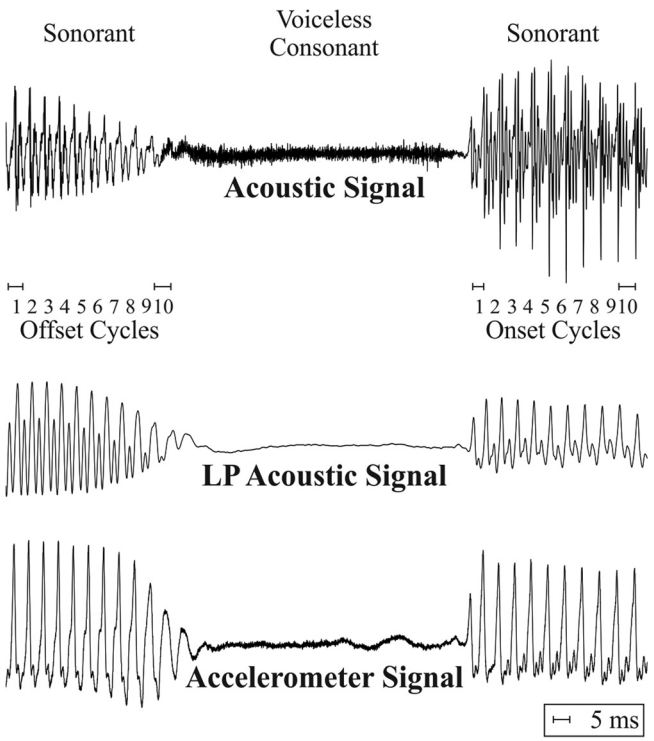

FIG. 1. Upper: A waveform of the RFF instance /ifa/ recorded using a microphone. This RFF instance is extracted from the sentence "Nelly found new fabric while Ray fell down." The bar scales directly below the waveform denote the first and tenth cycles for both offset and onset vowels. Center: A LP filtered microphone waveform of the instance /ifa/, constructed by LP filtering the unprocessed microphone signal at 350 Hz. Lower: The accelerometer waveform of the instance /ifa/, recorded using a neck-placed accelerometer. The time calibration bar below the waveform denotes the time scale of a 5 ms interval.

In previous studies, RFF has been estimated from the sound pressure waveforms recorded using a microphone in an experimental, low-noise environment. In these studies, RFF was estimated either directly from the microphone signal (Robb and Smith, 2002; Goberman and Blomgren, 2008; Stepp *et al.*, 2010; Stepp and Eadie, 2011; Stepp *et al.*, 2011; Stepp *et al.*, 2012; Eadie and Stepp, 2013; Stepp, 2013) or from a low-pass (LP) filtered waveform of the microphone signal (Watson, 1998). The effect of LP filtering on the reliability or mean values of RFF estimates is unknown. The reliability of RFF largely depends on the reliability of estimates of instantaneous $F0$, which is determined primarily by the glottal source (Fant, 1970), although evidence suggests that glottal source is somewhat dependent on the vocal tract filter (Titze *et al.*, 2008; Titze, 2008). LP filtering can "reduce the amplitude of the vocal tract resonances to facilitate measurement of vocal $F0$" (Watson, 1998, p. 3644); thus RFF estimates using the LP filtered sounds pressure waveform may be more reliable compared to estimates using the unprocessed waveform.

However, rather than post-processing the microphone signal to reduce the vocal tract resonances, an accelerometer signal recorded from the neck surface could also potentially be used for RFF estimation. Several studies have examined the advantages and disadvantages of using the accelerometer for voice assessments and monitoring (Coleman, 1988; Cheyne, 2002; Cheyne *et al.*, 2003; Popolo *et al.*, 2005; Hillman *et al.*, 2006; Zanartu *et al.*, 2009; Mehta *et al.*, 2012). It has been shown that a neck-placed accelerometer can provide accurate measurements of $F0$, sound pressure level, and phonation duration in both individuals with healthy voices and individuals with disordered voices (Svec *et al.*, 2005; Hillman *et al.*, 2006). The accelerometer signal is dependent on vibrations passing through the neck surface, so some high frequency components present in the microphone signal may be lost (Coleman, 1988), but this simplifies the procedure for $F0$ extraction due to the reduced harmonic content (Popolo *et al.*, 2005). Additionally, a previous study by Cheyne *et al.* (2003) that examined a signal from a neck-placed accelerometer found that the accelerometer signal contains minimal vocal tract formants, even those low frequency formants likely to remain in the microphone signal after LP filtering. Another potential benefit of using neck surface acceleration is that, unlike sound pressure, vocal cycles preceding or following voiceless consonant will not be masked by the burst of high energy in frication/aspiration that occurs due to coarticulation. Coarticulation is defined as "changes in the articulation of a speech segment depending on preceding and upcoming segments" (Cohen and Massaro, 1993, p. 94), which results in "an eventual obscuration of the boundaries between units at the articulatory or acoustic levels" (Kent and Minifie, 1977, p. 116). Last, the accelerometer signal is less affected by environmental noise, but is more sensitive to movement artifacts (Popolo *et al.*, 2005; Mehta *et al.*, 2012).

In this study, we carried out a systematic investigation to determine the effect of signal type (unprocessed microphone signal, LP filtered microphone signal, and unprocessed accelerometer signal) on RFF means and standard deviations (SDs). We also analyzed the intra-rater and inter-rater reliabilities for each signal type. We hypothesized that there would be minimal differences in RFF means across the signals because the calculation of RFF is based on the *F*0 which is similar when estimated using microphone and accelerometer signals (Hillman *et al.*, 2006). In addition, we hypothesized that RFF SDs would be lowest when estimated using unprocessed accelerometer signal and highest in unprocessed microphone signal, because masking of vocal cycles in the unprocessed microphone signal may interfere with the technicians' abilities to reliably estimate *F*0. Examining the effects of signal types on mean RFF values will determine whether it is feasible to estimate RFF using a LP filtered microphone signal or an accelerometer signal and support comparisons across studies that utilize different signal types to estimate RFF. Determining the effects of signal type on RFF SDs will aid in establishing a standard protocol with minimal within-speaker variability, allowing for more reliable and clinically feasible RFF-based voice assessments.

II. METHODOLOGY

A. Participants

Participants were 12 young adults (6 females) aged 18–28 yrs. All participants reported no prior history of speech, language, or hearing disorders and were native speakers of American English. All participants completed written consent in compliance with the Boston University Institutional Review Board.

B. Experimental design

Each participant was fitted with a head-mounted microphone (PC131, Sennheiser, Wedemark, Germany) and a miniature accelerometer (BU Series Knowles Acoustics, Itasca, IL). The microphone was positioned at a 45° angle from the mouth and the accelerometer was placed on the surface of the neck just above the sternal notch using medical grade adhesive (3M Double Stick Discs, 3M, St. Paul, MN). A previous study has shown that neck surface acceleration measured at this location provides estimates of *F*0, sound pressure level, and phonation duration that are similar to those using a microphone (Hillman *et al.*, 2006). Both microphone and accelerometer signals were recorded at 44.1 kHz and 16-bit resolution using a digital audio recorder (Olympus, model LS-10, Center Valley, PA).

Each participant was instructed to read the same set of stimuli (see Table I) in their typical pitch and loudness. The stimuli consisted of six sentences specifically designed for RFF analysis. Each sentence was purposefully loaded with three RFF instances (e.g., "We feel you do fail in new fallen dew"). RFF estimation requires that there are at least ten vocal cycles before and after the voiceless consonant and that the instantaneous *F*0s at the vocal cycles furthest away from the voiceless consonant (the reference cycles) are at steady-state. A reference cycle was considered to be at steady state if the RFF for the adjacent cycle (offset cycle 2 and onset cycle 9) has a magnitude less than 0.8 ST. To ensure that most RFF instances were usable, the stimuli were developed such that the voiceless consonants were flanked on both sides by stressed voiced phonemes whose durations tend to be longer than unstressed phonemes (Parmenter and Trevino, 1935).

The experiment took place in a sound-treated room. If a sentence was misarticulated or obviously glottalized, the experimenter instructed the participant to repeat the sentence.

C. Estimation of RFF

RFF was estimated from three types of signals: The unprocessed microphone signal, a LP filtered microphone signal, and the unprocessed accelerometer signal. The unprocessed microphone signal and the unprocessed accelerometer signal were the signals recorded during the experiment without any post-processing. The LP filtered microphone signal was constructed by filtering the unprocessed microphone signal using a LP fifth order Butterworth filter with a cutoff frequency of 350 Hz in MATLAB (The MathWorks Inc., Natick, MA, 2012). Similar to a previous study (Watson, 1998), this cutoff frequency was selected because it removes most of the vocal tract resonances while still maintaining energy at the *F*0, which typically averages to be about 117 to 137 Hz for males and 200 to 217 Hz for females (Fitch and Holbrook, 1970).

Three individuals, including the first author, were trained in RFF analysis by the final author. All three technicians independently estimated the RFF using Praat (Boersma and Weenink, 2012) and Microsoft Excel (Microsoft, Redmond, WA). To estimate RFF, each signal waveform was first examined in Praat using a default pitch range of 60–300 Hz for male recordings and 90–500 Hz for female recordings, but this range was altered on an individual basis. Default settings were used for all other Praat parameters. Next, each technician selected the waveform of an instance, using Praat to determine the 11 pulse timings before and after the voiceless consonant. These pulse timings were exported to Excel to calculate the periods of each vocal cycle as the difference between adjacent pulse times. Instantaneous *F*0s, the inverse of the periods, were computed for ten vocal cycles preceding and following the voiceless consonant. To calculate RFF in semitones, each *F*0 was normalized to the reference

TABLE I. RFF stimuli.

Phoneme	Sentence
/s/	**We sa**ng a jol**ly so**ng all d**ay Su**nday morning.
/f/	Nel**ly fo**und n**ew fa**bric while R**ay fe**ll down.
/k/	**You kno**ck away m**y ca**ke and Ne**lly ca**me along.
/t/	**I te**ll you, m**y tea** is w**ay too** warm.
/ʃ/	I w**ish I** would w**ash o**n m**y sho**re one day.
/p/	I'm h**appy** w**e pa**y our n**ew pa**l Nelly.

*The RFF instances used for estimation are bolded and underlined.

fundamental frequencies ($F0_{ref}$) using Eq. (1). Similar to previous studies, the instantaneous $F0$s for offset cycle 1 and onset cycle 10 were selected to be the $F0_{ref}$ for offset cycles and onset cycles, respectively. These reference cycles were selected because they were the ones closest to the mid-portions of the vowel and furthest from the consonant. Consequently, they are more likely to be at steady-state and to facilitate the capture of the changes in instantaneous $F0$s during devoicing and revoicing:

$$\mathrm{ST} = 39.86 \times \log_{10}(F0/F0_{\mathrm{ref}}). \quad (1)$$

Offset or onset RFF in an instance was rejected by a technician if the phoneme was misarticulated, or if the voiced section was glottalized or did not contain at least ten voicing cycles. In addition, to ensure that the reference cycle of the vowel was near steady-state, a technician also rejected the offset or onset RFF if the magnitude of the RFF for the cycle adjacent to the reference cycle (i.e., offset cycle 2 or onset cycle 9) was greater than 0.8 ST. The RFF values estimated from the three RFF instances in each sentence were used to calculate the sentence-level RFF means and SDs.

D. Reliability procedures and analysis

To determine the intra-rater reliability, each technician re-estimated >15% of the total samples in each measurement type (unprocessed microphone signal, LP filtered microphone signal, and unprocessed accelerometer signal) roughly 1 month after the initial analysis. The Pearson-product moment correlation coefficients were calculated and are displayed in Table II. For all technicians, the intra-rater reliabilities were lowest for the LP filtered microphone signal, although all were greater than or equal to 0.87. Two out of three technicians had slightly higher intra-rater reliability for the unprocessed accelerometer signal than for the unprocessed microphone signal.

Inter-rater reliability of RFF estimates was analyzed using the intraclass correlation (Shrout and Fleiss, 1979), type (2, k). The reliabilities for unprocessed microphone signals, LP filtered microphone signals, and unprocessed accelerometer signals were all high, reaching 0.94, 0.95, and 0.96, respectively.

Statistical analyses were performed using Minitab Statistical Software (Minitab Inc., State College, PA). A three-factor repeated-measures analysis of variances (ANOVA) was used to determine the effect of signal type (unprocessed microphone signal, LP filtered microphone signal, and unprocessed accelerometer signal) on both the sentence-level RFF means and SDs. Factors were signal type, vocal cycle (offset 1–10 and onset 1–10, rater (technicians 1–3), and all interactions: Vocal cycle × signal type, vocal cycle × rater, signal type × rater, and vocal cycle × rater × signal type. Effect sizes were quantified using the square partial curvilinear correlation (η_p^2) and interpreted as small, medium, or large (Witte and Witte, 2010). A predetermined level of statistical significance ($p < 0.05$) was used for all analyses. All *post hoc* analyses were completed using Tukey's Honestly Significant Difference tests.

TABLE II. Intra-rater reliability using Pearson-product moment correlation coefficients.

	Unprocessed microphone	LP microphone	Unprocessed accelerometer
Technician 1	0.90	0.87	0.91
Technician 2	0.99	0.90	0.96
Technician 3	0.94	0.93	0.96

*LP = Low-pass filtered.

III. RESULTS

A three-factor repeated-measures ANOVA (see Table III) indicated statistically significant effects ($p < 0.001$) of vocal cycle (offset 1–10 and onset 1–10), signal type (unprocessed microphone signal, LP filtered microphone signal, and unprocessed accelerometer signal), rater (technician 1–3), and the interactions of vocal cycle × signal type and vocal cycle × rater on the sentence-level RFF means. The effect sizes of signal type and the interaction of vocal cycle × signal type were small ($\eta_p^2 \leq 0.02$) in comparison to the effect size of vocal cycle ($\eta_p^2 = 0.51$). Similarly, the effect sizes of rater and the interaction of vocal cycle × rater were also small ($\eta_p^2 \leq 0.02$). *Post hoc* testing revealed that the sentence-level RFF means determined from the LP filtered microphone signal and unprocessed accelerometer signal were significantly ($p_{adj} < 0.05$) lower than those estimated using the unprocessed microphone signal, and the RFF means for the LP filtered microphone signal were not significantly different from those for the unprocessed accelerometer signal. To explore these differences in terms of the statistically significant interaction found between signal type and cycle, the sentence-level RFF means for each signal type were plotted as a function of cycle (Fig. 2). Significant differences ($p_{adj} < 0.05$) were observed among the sentence-level RFF means in offset cycles 7–10 and onset cycle 2. For offset cycles 7–9, the RFF means for the unprocessed microphone signal were significantly higher than those for the LP filtered microphone signal and unprocessed accelerometer signal, but no significant difference was observed between the RFF means for the LP filtered microphone signal and the unprocessed accelerometer signal. For offset cycle 10, the RFF means for the unprocessed

TABLE III. Results of 3-factor repeated-measures analysis of variance on sentence-level RFF means.

Effect	DF	η_p^2	F	p
Vocal cycle	19	0.51	693.6	<0.001
Signal type (Microphone, LP Microphone, Accelerometer)	2	<0.01	10.5	<0.001
Rater (Technician 1–3)	2	<0.01	9.1	<0.001
Vocal cycle × signal type	38	0.02	7.1	<0.001
Vocal cycle × rater	38	0.02	5.1	<0.001
Signal type × rater	4	<0.01	1.6	0.184
Vocal cycle × rater × signal type	76	<0.01	0.7	0.959

*LP = Low-pass filtered.

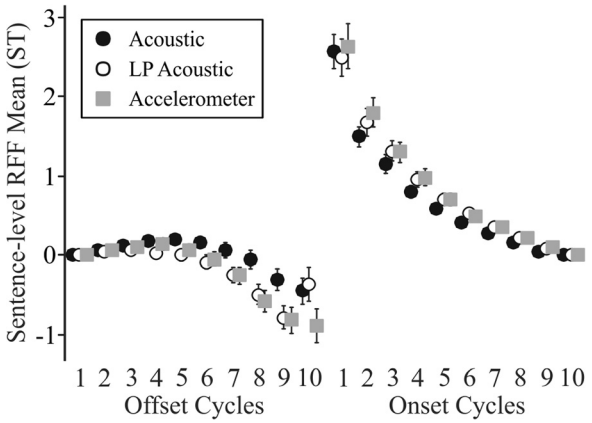

FIG. 2. Sentence-level RFF means as a function of signal type (unprocessed microphone signal—Microphone, LP filtered microphone signal—LP Microphone, and unprocessed accelerometer signal—Accelerometer) and vocal cycle (offset 1–10 and onset 1–10) in STs. Error bars indicate 95% confidence intervals for the means.

accelerometer signal were significantly lower than those for the unprocessed microphone signal and the LP filtered microphone signal, but no significant difference was observed between RFF means of the unprocessed microphone signal and the LP filtered microphone signal. For onset cycle 2, the RFF means for the unprocessed accelerometer signal were significantly higher than those for the unprocessed microphone signal, but no significant difference was observed between RFF means of the unprocessed microphone signal and the LP filtered microphone signal or between the RFF means of the LP filtered microphone signal and the unprocessed accelerometer signal.

A three-factor repeated-measures ANOVA (see Table IV) indicated statistically significant effects ($p < 0.001$) of vocal cycle, signal type, rater, and the interaction of vocal cycle × signal type on the sentence-level RFF SDs. The effect sizes of signal type and the interaction of vocal cycle × signal type were small ($\eta_p^2 \leq 0.01$) in comparison to the effect size of vocal cycle ($\eta_p^2 \leq 0.42$). Similarly, the effect size of the rater ($\eta_p^2 \leq 0.01$) was small. *Post hoc* testing revealed that the sentence-level RFF SDs for the unprocessed microphone signal were significantly ($p_{adj} < 0.05$) lower than those for the LP filtered microphone signal and the unprocessed accelerometer signal, but no statistically significant difference in sentence-level RFF SDs was found

TABLE IV. Results of 3-factor repeated-measures analysis of variance on RFF standard deviations.

Effect	DF	η_p^2	F	p
Vocal cycle	19	0.42	440.4	<0.001
Signal type (Microphone, LP Microphone, Accelerometer)	2	<0.01	23.5	<0.001
Rater (Technician 1–3)	2	<0.01	17.6	<0.001
Vocal cycle × signal type	38	0.01	4.2	<0.001
Vocal cycle × rater	38	<0.01	1.2	0.229
Signal type × rater	4	<0.01	0.9	0.450
Vocal cycle × rater × signal type	76	<0.01	0.6	0.997

*LP = Low-pass filtered.

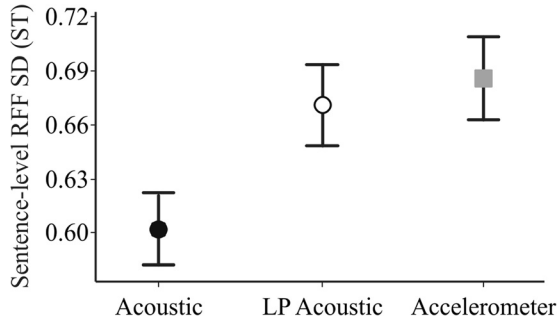

FIG. 3. Sentence-level RFF SDs as a function of signal type (Microphone, LP Microphone, and Accelerometer) in ST. Error bars indicate 95% confidence intervals for the means.

between the LP filtered microphone signal and the unprocessed accelerometer signal. Figure 3 shows a plot of the mean values of the sentence-level RFF SDs for each signal type. To explore these differences in terms of the statistically significant interaction found between signal type and cycle, the sentence-level RFF SDs for each signal type were plotted as a function of cycle (Fig. 4). Significant differences ($p_{adj} < 0.05$) were observed among the sentence-level RFF SDs in offset cycle 10 and onset cycles 1–3. For offset cycle 10, RFF SDs for the unprocessed microphone signal were significantly ($p_{adj} < 0.05$) lower than those for the LP filtered microphone signal and the unprocessed accelerometer signal, but no statistically significant difference was observed between standard deviations of the LP filtered microphone signal and the unprocessed accelerometer signal. Although no statistically significant difference was found for other offset cycles, there was a general trend in which the offset RFF SDs tended to be lowest in the unprocessed accelerometer signal. For onset cycles 1–3, RFF SDs for the unprocessed microphone signal were significantly lower than those for the unprocessed accelerometer signal, but no statistically significant differences were observed between SDs of the unprocessed microphone signal and the LP filtered microphone signal or between standard deviations of the LP filtered microphone signal and the accelerometer signal. There was also a general trend in the onset cycles in which the RFF SDs tended to be lowest for the unprocessed microphone signal and highest for the unprocessed accelerometer signal.

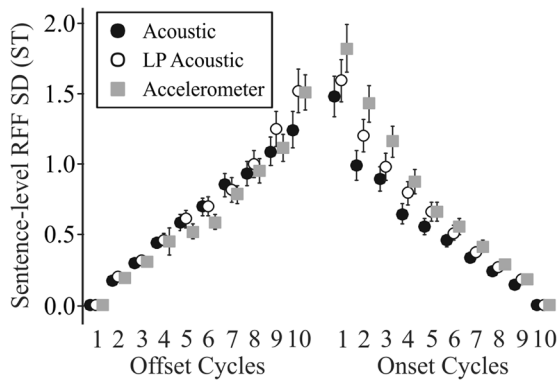

FIG. 4. Sentence-level RFF SDs as a function of signal type (Microphone, LP Microphone, and Accelerometer) and vocal cycle (offset 1–10 and onset 1–10) in ST. Error bars indicate 95% confidence intervals for the means.

IV. DISCUSSION

The goal of this study was to understand how signal type affects RFF. Statistically significant effects of signal type and the interaction of vocal cycle × signal type were found for sentence-level RFF means; however, the effect sizes were quite small ($\eta_p^2 \leq 0.02$), with most of the variance in the data explained by the effect of vocal cycle. The overall RFF trend was preserved regardless of the signal used for RFF estimation, suggesting that RFF can be accurately estimated from either an unprocessed microphone signal, a LP filtered microphone signal, or an unprocessed accelerometer signal.

We also found statistically significant effects of signal type and the interaction of vocal cycle × signal type on sentence-level RFF SDs. Again, compared to the effect size of vocal cycle, the effect sizes were all small ($\eta_p^2 \leq 0.01$). Significant differences in RFF SDs among signal types tended to occur in the cycles closest to the voiceless consonant (offset cycle 10 and onset cycles 1–3). Offset RFF SDs tended to be lowest when estimated using the unprocessed accelerometer signal and onset RFF SDs tended to be lowest when estimated using the unprocessed microphone signal. This suggests that offset RFF estimated from an unprocessed accelerometer signal will have the lowest intra-speaker variability and onset RFF estimated from an unprocessed microphone signal will have the lowest intra-speaker variability. However, given the small effect sizes ($\eta_p^2 \leq 0.01$) of the interaction of vocal cycle × signal type, this slight difference in intra-speaker variability is probably not important for most applications.

A. RFF estimates using neck surface accelerometry

This study shows that RFF mean values estimated from the accelerometer signal are significantly lower than those estimated using the unprocessed microphone signal. *Post hoc* testing indicated that this difference in RFF means occurs in offset cycles 7–10 and onset cycle 2. In agreement with our initial hypothesis, the difference in RFF means is small as indicated by the small effect sizes ($\eta_p^2 \leq 0.02$) of signal type and the interaction of vocal cycle × signal type. These effect sizes are much smaller than the effect size of vocal cycles ($\eta_p^2 \leq 0.51$). Consequently, the overall trend of RFF as a function of cycle estimated using accelerometer signals is essentially similar to those estimated using microphone signals.

This study also demonstrates that RFF SDs estimated from an accelerometer signal are significantly higher than those estimated using an unprocessed microphone signal. There was also a significant effect of the interaction of vocal cycle × signal type on RFF SD values. *Post hoc* analysis indicated that in the offset cycles, RFF derived from accelerometer signals tends to have a *lower* SD than those derived from microphone signals, while onset RFF derived from accelerometer signals tends to have a higher SD than those derived from microphone signals. This is in contrast with our initial hypothesis that both offset and onset RFF SDs should be higher in RFF derived from a microphone signal relative to an accelerometer signal due to masking by frication/aspiration. Even though masking may increase the difficulty for technicians to reliably estimate RFF resulting in higher RFF SDs, missing the cycles close to the voiceless consonant will result in lower RFF SDs since RFF standard deviations tend to be lower for cycles further away from the voiceless consonant. Due to these opposing effects, the resulting effect of signal type on RFF SD should be small, which is consistent with our results.

Overall, differences in both mean and SD RFF values between accelerometer signals and microphone signals are small. The pattern of RFF as a function of cycle when estimated using an accelerometer signal is similar to that estimated using a microphone signal. These findings imply that RFF can be both accurately and reliably estimated from an accelerometer signal. In fact, in comparison to the microphone signal, the accelerometer signal may be preferred for RFF estimation due to the slightly higher technician intra- and inter-rater reliabilities.

B. Differences between accelerometer and microphone signals

Although small, we found significant differences in RFF between microphone and accelerometer signals, which is likely due to the inherent differences between the two signals. Microphone signals capture the vocal fold vibrations, the vocal tract formants, radiation characteristics of the mouth, and any environmental noise, while accelerometer signals are comprised of vocal fold vibrations, the subglottal formants, and neck surface transmission properties (Svec *et al.*, 2005). Aside from the vocal fold vibrations, all of these other properties may result in the differences seen in RFF estimated using a microphone signal versus an accelerometer signal. Moreover, since the accelerometer signal will not respond to sound sources above the glottis, its response to unvoiced segments will be minimal compared to the microphone signal (Cheyne *et al.*, 2003). The effect of this difference on RFF is not direct, since RFF is calculated based on the *F*0 from the *voiced* segments. However, because of coarticulation, vocal cycles immediately preceding and following a fricative may be masked by simultaneous frication in the microphone signal, which is not present in the accelerometer signal. In addition, the vocal cycles following stops may also be masked in the microphone signal due to frication/aspiration. Thus, the resulting RFF estimation may be affected because the vocal cycles that are present in the accelerometer signal are masked in the microphone signal.

To explore this hypothesis, we compared the RFF cycle times (the start and ending times of vocal cycles) of the nearest cycles preceding and following the voiceless consonant between the unprocessed microphone signal and the accelerometer signal (i.e., the ending time of offset cycle 10 and the starting time of onset cycle 1; see Fig. 1). The offset and onset RFF for the unprocessed microphone signal were taken on average 13.6 ms (SD = 19.3 ms) and 5.9 ms (SD = 10.2 ms) further away from the consonant compared to the unprocessed accelerometer signal, respectively. The mean *F*0 of the speakers was 155 Hz (average period

6.5 ms). Thus, on average, offset RFF was taken approximately two cycles further away from the voiceless consonant and onset RFF was taken less than one cycle further away from the consonant for the unprocessed microphone signal relative to the accelerometer signal. This supports our hypothesis that some offset cycles are masked in the unprocessed microphone signal and this may result in the higher offset RFF in the microphone signal relative to the accelerometer signal. Little masking occurs in the onset cycles of the microphone signal, so onset RFF values derived using the unprocessed microphone signals and the unprocessed accelerometer signals are more similar.

We explored this hypothesis further by examining the LP filtered microphone signal, which should have reduced masking effects compared with the unprocessed microphone signal. When we compared the average RFF cycles times of the LP filtered microphone signal to the accelerometer signal (see Fig. 1), offset cycle 10 and onset cycle 1 RFF for the LP filtered microphone signal were taken on average only 2.0 ms (SD = 15.5 ms) and 1.9 ms (SD = 7.3 ms) further away from the consonant than the unprocessed accelerometer signal, respectively. Thus, both offset and onset RFF are taken less than one cycle further away from the voiceless consonant for the LP filtered microphone signal relative to the accelerometer signal. Since filtering removes most of the masking, RFF values estimated using the LP filtered microphone signal were more similar to those estimated using the unprocessed accelerometer signal; however, some masking effects still remain, resulting in the differences seen in offset cycle 10 RFF between the LP filtered microphone signal and the unprocessed accelerometer signal.

C. Advantages/disadvantages of accelerometer-based RFF measurements

There are several advantages to using an accelerometer for RFF-based measurements. Individuals with voice disorders often find voice therapy exercises difficult to perform outside of the clinic and some have reported that feedback is helpful in these situations (van Leer and Connor, 2010). Real-time feedback related to vocal function such as RFF-based assessments could be used to assist these individuals. Use of an accelerometer to collect the RFF instances would allow for monitoring in noisy environments. In addition, the privacy of monitored individuals would be maintained since the messages cannot be reconstructed with an accelerometer signal (Cheyne *et al.*, 2003).

However, there are also disadvantages associated with the use of accelerometers for RFF-based measurements. The accelerometer is attached to the surface of the jugular notch with an adhesive. This adhesive may detach due to perspiration or movement. Moreover, adhesives may be more difficult to attach in older individuals as the elastic properties of skin change with aging (Escoffier *et al.*, 1989). The fidelity of the accelerometer signal depends not only on the adhesive, but also on the neck surface transmission properties (Svec *et al.*, 2005). Thus, the signal will be degraded in individuals with thick tissue (skin or adipose) layers.

D. Implications and limitations

The RFF means and SDs for the microphone signals were similar to those observed in the previous study by Robb and Smith (2002) who also studied young adults with healthy voices. In both studies mean offset and onset RFF values were found to decrease as a function of cycle. For this study, the mean RFF values for offset cycle 10 and onset cycle 1 were −0.45 and 2.6 ST, respectively. Robb and Smith (2002) found mean RFF values for offset cycle 10 and onset cycle 1 of −0.84 and 2.8 ST, respectively. However, offset RFF means estimated from the LP filtered microphone signal in the present study were slightly lower than those reported by Watson (1998). The mean RFF values for offset cycle 10 and onset cycle 1 were −0.37 and 2.5 ST, respectively. Conversely, Watson (1998) found mean offset cycle 10 and onset cycle 1 RFF values of 0.44 and 2.27 ST, respectively. The studies differed slightly in terms of the age and number of participants; however, the most prominent differences between the two studies were the stimuli and signal post-processing employed. In the study by Watson (1998), the stimuli consisted of RFF instances with the phoneme /s/, whereas the stimuli in this study contained RFF instances with several voiceless consonants (/f/, /s/, /ʃ/, /p/, /t/, /k/). Furthermore, in contrast to this study, the sounds pressure waveform in Watson's study was down-sampled at 5 kHz to expedite the filtering process. Finally, the filter employed was not specified in the study by Watson (1998), thus there may be slight differences between the exact filtering characteristics.

The results of the current study imply that RFF in young adults with healthy voices can be accurately and reliably estimated using either an unprocessed microphone signal, a LP filtered microphone signal, or an unprocessed accelerometer signal, since the effect sizes of signal on both the mean and SD RFF values were small ($\eta_p^2 \leq 0.02$). The magnitude of the mean RFF difference between the accelerometer signal and LP filtered microphone signal (the two signals with the greatest difference) at offset cycle 10 and onset cycle 1 were 0.53 and 0.10 ST, respectively. For offset cycle 10, this magnitude difference between signals is approximately one-half as much as the magnitude difference that has been found between individuals with healthy voices and individuals with voice disorders (Goberman and Blomgren, 2008; Stepp *et al.*, 2010). For example, previous research has found that this magnitude difference between individuals with healthy voices and individuals with VH before voice therapy was 1.0 ST and the magnitude difference between individuals with healthy voices and individuals with PD before medication was 1.1 ST (Goberman and Blomgren, 2008; Stepp *et al.*, 2010). For onset cycle 1, this magnitude difference between signal is less than one-tenth as much as the magnitude difference between individuals with healthy voices and individuals with voice disorders (Goberman and Blomgren, 2008; Stepp *et al.*, 2010). For instance, previous research has found that this magnitude difference between individuals with healthy voices and individuals with VH before voice therapy was 1.4 ST and the magnitude difference between individuals with healthy voices and individuals with PD before medication was 3.8 ST (Goberman and Blomgren, 2008; Stepp

et al., 2010). Only offset cycle 10 RFF and onset cycle 1 RFF are compared here, because they are the cycles that resemble the greatest differences between the signal types and the cycles that resemble the greatest differences between disordered voices and controls. In addition, the results suggest that when comparing across studies that estimate RFF utilizing different signals, one needs to account for the slight differences in offset cycles 7–10 and onset cycle 2.

In this study, we compared the accelerometer and microphone signals, both of which have been shown to provide accurate estimates of *F*0. Another signal that provides accurate measures of *F*0 is the electroglottagraph (EGG) signal (Kitzing, 1980; Colton and Conture, 1990; Baken, 1992). An EGG signal measures the change in electrical impedance of a current that passes through two electrodes placed slightly apart on the middle of the thyroid lamina (Kitzing, 1980; Colton and Conture, 1990; Baken, 1992). Since EGG signals provide accurate estimates of *F*0, we expect the signal can also be used for RFF estimation. The advantages of using the EGG signal is that the *F*0 can be extracted using simple zero-crossing methods (Baken, 1992). In addition, similar to the accelerometer signal, the EGG signal is not influenced by vocal tract formats and the vocal cycles adjacent to voiceless consonants will not be masked by coarticulation (Baken, 1992). The major disadvantage of using this signal is that the signal-to-noise is even lower than the accelerometer signal when it is used in individuals with thick adipose layers (Kitzing, 1980). Additionally, the signal is affected by the exact placement of the sensor, the degree of electrode-to-skin contact, movements of the various neck structures (e.g., larynx, extrinsic neck muscle contraction), and mucus bridges (Colton and Conture, 1990; Baken, 1992; Golla *et al*., 2009).

Although we have shown that different signal types may be used for RFF estimation, the optimal signal is still unknown. If RFF is to be adapted for the assessment of VH, then the "gold standard" signal for RFF estimation should be the signal that provides reliable RFF estimates that most accurately distinguish the voices of healthy individuals from voices with VH. The results of this study shows that in healthy speakers, either signal can provide a reliable and accurate measure of RFF, and that the mean RFF values for individuals with healthy voices will be slightly lower when estimated from the unprocessed accelerometer signal relative to the microphone signal. However, a limitation of the study is that no individuals with voice disorders were used, thus the results do not yet directly apply to these individuals. Future studies should be performed to determine the effect of signal type on RFF in individuals with *disordered* voices. We hypothesize that RFF can be accurately estimated from an accelerometer signal in disordered voices, since estimation of RFF requires accurate estimation of *F*0, which has previously been accomplished in individuals with disordered voices (Hillman *et al*., 2006). The RFF estimated from a microphone signal for individuals with excessive tension or strain in their voice is lower compared to individuals with healthy voices (Goberman and Blomgren, 2008; Stepp *et al*., 2010; Stepp, 2013). Since the physiological mechanisms underlying this difference are the same regardless of the signal type, we hypothesize that a similar pattern will be observed in RFF estimated using an accelerometer signal as the pattern observed in RFF estimated using a microphone. Future studies utilizing both microphone and neck surface acceleration in speakers with healthy and disordered voices are necessary to confirm this hypothesis.

V. CONCLUSIONS

Signal type showed a significant effect on RFF means and SDs, but the effect sizes of signal type and the interaction of vocal cycle × signal type were small in comparison to the effect size of vocal cycle. The overall RFF trend was preserved regardless of the signal used to estimate RFF and the differences in intra-speaker variance between the RFF estimated using microphone signal and accelerometer signal were fairly small. Thus, for individuals with healthy voices, RFF can be accurately and reliably estimated from either a microphone signal or an accelerometer signal. Future studies are necessary to determine the effect of signal type on RFF in speakers with disordered voices.

ACKNOWLEDGMENTS

This work was supported in part by Grant No. DC012651 from the National Institute of Deafness and Other Communication Disorders and a New Century Scholar grant from the American Speech-Language-Hearing Foundation. Thanks to Lauren Kalfin and Emma Billard for help with RFF analysis, Caitlin Gattuccio for assistance with data recording, and Carolyn Michener and Joe Mendoza for support with data analysis.

Arnold, G. E. (**1961**). "Physiology and pathology of the cricothyroid muscle," Laryngoscope **71**, 687–753.
Baken, R. J. (**1992**). "Electroglottography," J. Voice **6**, 98–110.
Behrman, A. (**2005**). "Common practices of voice therapists in the evaluation of patients," J. Voice **19**, 454–469.
Berardelli, A., Sabra, A. F., and Hallett, M. (**1983**). "Physiological mechanisms of rigidity in Parkinson's disease," J. Neurol., Neurosurg., Psych. **46**, 45–53.
Boersma, W., and Weenink, D. (**2012**). "Praat: Doing phonetics by computer [Computer program]," http://www.praat.org/ (Last viewed January 31, 2012).
Cheyne, H. A. (**2002**). "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," J. Acoust. Soc. Am. **112**, 2445–2446.
Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., and Hillman, R. E. (**2003**). "Development and testing of a portable vocal accumulator," J. Speech, Lang., Hear. Res. **46**, 1457–1467.
Cohen, M. M., and Massaro, D. W. (**1993**). "Modeling coarticulation in synthetic visual speech," Models Tech. Comput. Animation **92**, 139–156.
Coleman, R. F. (**1988**). "Comparison of microphone and neck-mounted accelerometer monitoring of the performing voice," J. Voice **2**, 200–205.
Colton, R. H., and Conture, E. G. (**1990**). "Problems and pitfalls of electroglottography," J. Voice **4**, 10–24.
Eadie, T. L., and Stepp, C. E. (**2013**). "Acoustic correlate of vocal effort in spasmodic dysphonia," Ann. Otol. Rhinol. Laryngol. **122**, 169–176.
Escoffier, C., de Rigal, J., Rochefort, A., Vasselet, R., Leveque, J. L., and Agache, P. G. (**1989**). "Age-related mechanical properties of human skin: An in vivo study," J. Invest. Dermatol. **93**, 353–357.
Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton and Co., The Netherlands), pp. 15–26.
Fitch, J. L., and Holbrook, A. (**1970**). "Modal vocal fundamental frequency of young adults," Arch. Otolaryngol. **92**, 379–382.

Gallena, S., Smith, P. J., Zeffiro, T., and Ludlow, C. L. (**2001**). "Effects of levodopa on laryngeal muscle activity for voice onset and offset in Parkinson disease," J. Speech, Lang., Hear. Res. **44**, 1284–1299.
Goberman, A. M., and Blomgren, M. (**2008**). "Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease," J. Voice **22**, 178–191.
Golla, M. E., Deliyski, D. D., Orlikoff, R. F., and Moukalled, H. J. (**2009**). "Objective comparison of the electroglottogram to synchronous high-speed images of vocal-fold contact during vibration," in *Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications MAVEBA*, pp. 141–144.
Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., and Cheyne, H. A. (**2006**). "Ambulatory monitoring of disordered voices," Ann. Otol. Rhinol. Laryngol **115**, 795–801.
Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., and Vaughan, C. (**1989**). "Objective assessment of vocal hyperfunction: an experimental framework and initial results," J. Speech Hear. Res. **32**, 373–392.
Kent, R. D., and Minifie, F. D. (**1977**). "Coarticulation in recent speech production models," J. Phonetics **5**, 115–133.
Kitzing, P. (**1980**). "A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency," J. Speech Hear. Res. **23**, 258–273.
Lofqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.
Mehta, D. D., Zanartu, M., Feng, S. W., Cheyne, H. A., and Hillman, R. E. (**2012**). "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," IEEE Trans. Biomed. Eng. **59**, 3090–3096.
Morrison, M. D., Nichol, H., and Rammage, L. A. (**1986**). "Diagnostic criteria in functional dysphonia," Laryngoscope **96**, 1–8.
Parmenter, C., and Trevino, S. (**1935**). "The length of the sounds of a Middle Westerner," Am. Speech **10**, 129–133.
Popolo, P. S., Scaronvec, J. G., and Titze, I. R. (**2005**). "Adaptation of a Pocket PC for use as a wearable voice dosimeter," J. Speech, Lang., Hear. Res. **48**, 780–791.
Robb, M. P., and Smith, A. B. (**2002**). "Fundamental frequency onset and offset behavior: a comparative study of children and adults," J. Speech, Lang., Hear. Res. **45**, 446–456.
Roubeau, B., Chevrie-Muller, C., and Lacau Saint Guily, J. (**1997**). "Electromyographic activity of strap and cricothyroid muscles in pitch change," Acta Oto-Laryngol. **117**, 459–464.
Roy, N., Ford, C. N., and Bless, D. M. (**1996**). "Muscle tension dysphonia and spasmodic dysphonia: the role of manual laryngeal tension reduction in diagnosis and management," Ann. Otol. Rhinol. Laryngol. **105**, 851–856.
Shrout, P. E., and Fleiss, J. L. (**1979**). "Intraclass correlations: Uses in assessing rater reliability," Psychol. Bull. **86**, 420–428.
Stepp, C. E. (**2013**). "Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson's disease," J. Acoust. Soc. Am. **133**, 1637–1643.
Stepp, C. E., and Eadie, T. L. (**2011**). "Relative fundamental frequency as an acoustic correlate of vocal effort in spasmodic dysphonia," J. Acoust. Soc. Am. **129**, 2526–2526.
Stepp, C. E., Hillman, R. E., and Heaton, J. T. (**2010**). "The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset," J. Speech, Lang., Hear. Res. **53**, 1220–1226.
Stepp, C. E., Merchant, G. R., Heaton, J. T., and Hillman, R. E. (**2011**). "Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction," J. Speech, Lang., Hear. Res. **54**, 1260–1266.
Stepp, C. E., Sawin, D. E., and Eadie, T. L. (**2012**). "The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset," J. Speech, Lang., Hear. Res. **55**, 1887–1896.
Svec, J. G., Titze, I. R., and Popolo, P. S. (**2005**). "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," J. Acoust. Soc. Am. **117**, 1386–1394.
Titze, I., Riede, T., and Popolo, P. (**2008**). "Nonlinear source-filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**, 1902–1915.
Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.
van Leer, E., and Connor, N. P. (**2010**). "Patient perceptions of voice therapy adherence," J. Voice **24**, 458–469.
Watson, B. C. (**1998**). "Fundamental frequency during phonetically governed devoicing in normal young and aged speakers," J. Acoust. Soc. Am. **103**, 3642–3647.
Witte, R. S., and Witte, J. S. (**2010**). *Statistics* (Wiley, Hoboken, NJ), pp. 383–384.
Zanartu, M., Ho, J. C., Kraman, S. S., Pasterkamp, H., Huber, J. E., and Wodicka, G. R. (**2009**). "Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface," IEEE Trans. Biomed. Eng. **56**, 443–451.