Using Pitch Height and Pitch Strength to Characterize Type 1, 2, and 3 Voice Signals

*Supraja Anand, †Lisa M. Kopf, ‡Rahul Shrivastav, and *David A. Eddins, *Tampa, Florida, †Cedar Falls, Iowa, and ‡Athens, Georgia

Summary: Objective. Classifying dysphonic voices as type 1, 2, and 3 signals based on their periodicity enables researchers to determine the validity of acoustic measures derived from them. Existing methods of signal typing are commonly performed by listening to the voice sample and visualizing them on narrow-band spectrograms that require training, time, and are subjective in nature. The current study investigated pitch-based metrics (pitch height and pitch strength) as correlates to characterizing voice signal types. The computational estimates were validated with perceptual judgments of pitch height and pitch strength.

Methods. Pitch height and pitch strength were estimated from Auditory-Sawtooth Waveform Inspired Pitch Estimator Prime algorithm for 30 dysphonic voice segments (10 per type). Ten listeners evaluated pitch height through a single-variable matching task and pitch strength through an anchored magnitude estimation task. One way analyses of variance were used to determine the effects of signal type on pitch height and pitch strength estimates. Relationship between computational and perceptual estimates was evaluated using correlation coefficients and their significance.

Results. There was a significant difference between signal types in both computational and perceptual pitch strength estimates. Periodic type 1 signals had greater pitch strength compared to type 2 and 3 signals. Auditory-Sawtooth Waveform Inspired Pitch Estimator Prime produced robust computational estimates of pitch height even in type 3 signals when compared to other acoustic software. Listeners were able to reliably judge pitch height in type 2 and 3 signals despite their lack of a clear fundamental frequency.

Conclusions. Pitch height and pitch strength can be measured in all dysphonic voices irrespective of signal periodicity.

Key Words: Dysphonia−Perception−Acoustics−Signal typing−Pitch height−Pitch strength.

INTRODUCTION

Acoustic analysis of voice is an indispensable part of voice research and clinical assessment of dysphonia. However, the accuracy and validity of most conventional analysis routines is limited by the degree to which the voice being examined has a quasiperiodic waveform. While not universally used, "signal typing" can be a valuable guide for choosing an appropriate set of tools or measures for characterizing the vocal acoustic signal.

Signal typing refers to an evaluative process that allows researchers and clinicians to describe the periodicity of a voice signal, to systematically categorize those signals, and to use the signal type to guide the selection of appropriate acoustic analyses for evaluating dysphonic voices. Signal typing uses both auditory (playback and listen) and visual (waveform and narrowband spectrogram) representations to divide the voice into one of three or four categories. The summary statement of the National Center for Voice and Speech Workshop on Acoustic Voice Analysis (1995), described a scheme for classifying voices into three types (Figure 1).[1] Type 1 signals were defined as nearly-periodic signals with a consistent or nearly consistent fundamental frequency (f_0) throughout a sustained vowel segment. Type 2 signals were defined as those voices containing bifurcations and having subharmonic or modulating frequencies whose magnitude approach the magnitude at the f_0. Type 3 signals were defined as those voices that contain no readily apparent periodic structure. Sprecher et al[2] refined Titze's definition of voice signal types to include a fourth type, which divided type 3 signals into those with finite dimensionality and occasional presence of f_0 (type 3) and those with infinite dimensionality that were stochastic in nature (type 4). These signal typing methods illustrate the complex spectrotemporal features of voices that range along a continuum from normal to disordered and from quasiperiodic to predominantly aperiodic.

Although the importance of signal typing is well-recognized in the scientific community, the methods are complex and include subjective evaluation; therefore, only a limited number of research studies have formally classified voice signals prior to acoustic analyses.[3−9,2,10,11] Among those studies, most have not described the typing procedures in detail (eg, lack of description on rater experience) and have not obtained signal-type classifications from multiple judges. Since the signal typing procedure requires visual judgment and segmentation of dysphonic voices, it is important to evaluate its accuracy through the use of multiple judges. The two studies that did report inter-rater reliability

Accepted for publication August 8, 2019.

Fudning: This research was supported by a grant from NIH (Grant No. R01 DC009029).

From the *Department of Communication Sciences and Disorders, University of South Florida, Tampa, Florida; †Department of Communication Sciences and Disorders, University of Northern Iowa, Cedar Falls, Iowa; and the ‡Office of the Vice President for Instruction, University of Georgia, Athens, Georgia.

Address correspondence and reprint requests to Supraja Anand, 4202 E. Fowler Avenue, PCD 1017, Tampa, FL 33620. E-mail: suprajaanand@usf.edu

Journal of Voice, Vol. 35, No. 2, pp. 181−193
0892-1997

https://doi.org/10.1016/j.jvoice.2019.08.006

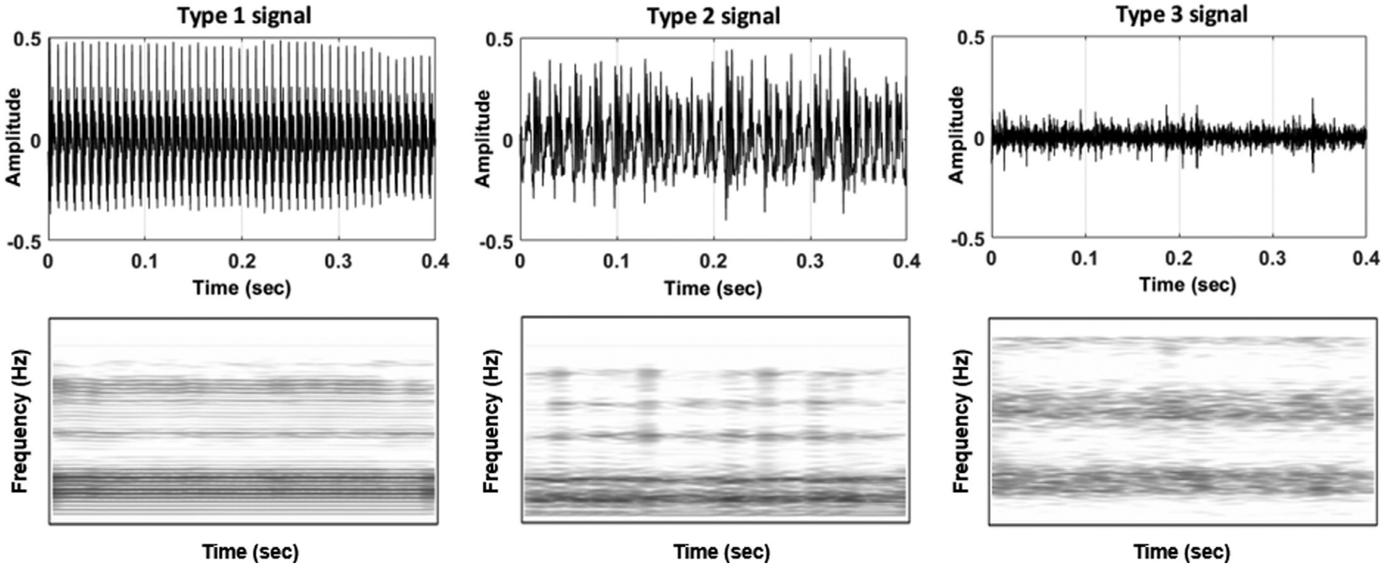

FIGURE 1. Sample waveforms (upper panel) and spectrograms (lower panel) of type 1, 2, and 3 voice signals.

and agreement data indicated only poor to fair correlations among judges.[3,9] For example, Behrman et al[3] reported that judges had difficulty discriminating between 40% of signals considered types 2 and 3. On the other hand, a study by Ma and Yiu[9] reported 77% interjudge exact agreement in categorizing voices into one of the Titze's three voice signal types. It is likely that the subjective nature of the signal typing task results in low reliability/agreement scores and presents challenges during clinical evaluation. Alternatively, quantitative methods through the use of nonlinear dynamic analyses can be used to classify signal types.[12,13] Zhang and Jiang[13] examined correlation dimension analysis as a potential nonlinear dynamic measure and reported that such a measure was able to effectively describe and classify dysphonic voices into type 1, 2, and 3 signals. The correlation dimension (D_2) increased from type 1 to 3 signals and D_2 was statistically different between any two types of signals ($P< 0.001$). Calawerts et al[12] introduced rate of divergence as a potential objective measure that differentiates the four signal types based on Sprecher's classification system. The rate of divergence parameter uses a modified version of Wolf's algorithm for calculating Lyapunov exponents and was extracted from types 1, 2, 3, and type 4 sustained /a/ samples. Similar to D_2, the rate of divergence measure was able to differentiate type 1, 2, 3, and 4 signals with a high level of accuracy. More importantly, unlike D_2, this measure was effective in differentiating type 3 and 4 signals ($P< 0.001$) as it was effective in characterizing type 4 signals which are heavily masked by stochastic signal components. While such computational methods appear to be promising and may provide a suitable replacement for signal typing, their broader adoption by professionals has been constrained by the lack of understanding or intuition regarding their derivation or interpretation along with a lack of easy access to these analytical methods. Consequently, evaluation of the dysphonic voice in the clinic is completed via more conventional acoustic analyses that represent the vocal acoustic signal in temporal, spectral, and/or cepstral domains.

Estimates of fundamental frequency (f_0; defined as the lowest frequency in a harmonic complex) have a long-standing heritage in research and clinical practice as a useful means of indexing vocal fold vibration rate and as a covariate in other complex acoustic analyses. Measures of perturbation (eg, jitter and shimmer) and noise (eg, harmonic-to-noise ratio) in the vocal signal are commonly used to identify dysphonia, to characterize its nature, or to quantify its severity.[14,1] Prior research, however, has demonstrated significant variability in each of these time-based measures[15−20] and, in some cases, contradictory results. For example, Carding et al[4] reported higher perturbation values while Wolfe et al[19] reported lower perturbation values associated with dysphonic voices compared to normal voices. Furthermore, the accuracy of such acoustic measures is dependent on having a nearly perfectly periodic voice signal with a consistently measurable f_0. In reality, however, up to 80% of people with dysphonia have voices that have considerable aperiodicity.[3,9] Due to their dependence on a stable f_0 and resulting periodicity, the validity and reliability of many time-based acoustic analyses of voice are questionable when applied to many dysphonic voices. On the contrary, that very instability, the degree of variability, or the magnitude of change may represent key acoustic features needed to fully characterize a dysphonic voice. Measures based on spectral and cepstral analyses of the voice signal do not require identification of individual cycles as in time-based analyses. Broadly speaking, one class of spectral analyses includes measures that consider local (eg, the amplitude of first harmonic, H1, relative to that of the first formant, A1, or the second harmonic, H2) or global (spectral slope, tilt, low-to-high-frequency (L/H) ratio, high-frequency power ratio) comparisons of spectral magnitude across audio-frequency.[21−24] Another approach is to first use a second Fourier transformation to convert the spectrum to the cepstral domain and then capture variations in the signal characteristics across audio-quefrency (eg, cepstral peak prominence, CPP[25−29]). Significant correlations between spectral/cepstral measures and overall dysphonia severity have been reported in the literature.[30,31] Since these spectral- and cepstral-based measures can overcome the periodicity limitation discussed earlier to a certain extent, their use has gained popularity in the last decade. Some experts have used such measures to develop indices that broadly track changes in voice quality (eg, Acoustic Voice Quality Index, AVQI;[32,33] and Cepstral Spectral Index of Dysphonia, CSID[34,35,29]).

Yet another approach has been to use a bio-inspired computational front-end that can model the nonlinearity of the acoustic signal transduction process in the auditory pathways, prior to any characterization of the vocal acoustic signal. The use of such models has allowed the development of metrics that are highly correlated with perceptual judgments of voice quality.[36−39] This general approach can also apply to the study of f_0 or periodicity, which forms the basis for signal typing. Note that f_0 is a physical property of the signal and is difficult to quantify for signals with bifurcations (type 2) or with aperiodic sound sources (type 3/4). In contrast, pitch characteristics are psychological attributes of a sound, and may be described for all sounds, irrespective of the underlying periodicity (or lack thereof). This relative independence between pitch and periodicity allows the use of pitch to describe a wider set of voice signals than may be possible through the use of f_0. From a perceptual perspective, pitch itself is a three-dimensional construct consisting of “height,” “chroma,” and “strength”.[40−42] Pitch height is that attribute of sound ordered on a scale from low to high.[43,42] In contrast, pitch chroma refers to notes of the musical scale perceived as repeating once per octave.[41] Finally, pitch strength refers to the degree of tonality in a sound or the salience of the pitch sensation on a scale from weak to strong.[42] The concepts of pitch height and pitch strength, along with pitch chroma, have been used for decades to describe a variety of sounds (eg, pure tones, complex tones, amplitude-modulated tones, narrow-band noise, broad-band noise, band-pass noise, and comb filtered noise[44,45,42]). Fastl and Stoll[44] described three separate acoustic bases of the pitch sensation and referred to those as spectral, virtual, and noise pitch. Spectral pitch, described as the pitch sensation the auditory system derives from a pure tone has historically been considered to have the strongest sensation of pitch strength, ranging between 75% and 100%. On the other hand, virtual pitch, described as the pitch sensation (height and strength) that the auditory system derives from a complex signal evokes an intermediate sensation of pitch strength, averaging a pitch strength of 50%. Finally, noise pitch was described as the pitch sensation that the auditory system derives from noise. For narrowband noises, the pitch strength may approach the levels achieved for spectral pitch. However, on average, noises are considered to have the weakest pitch strength, ranging between 0% and 25%.

These concepts of pitch height and pitch strength can be extended from the psychoacoustic literature with synthetic non-speech sounds to normal and dysphonic voices. Pitch height is a parameter that is often assessed in clinical evaluation of dysphonia (ie, CAPE-V[46]). Pitch chroma forms the basis of various notes in most musical styles. The use of pitch strength to describe dysphonic voices is relatively recent[47] and listeners’ ability to reliably provide perceptual judgments of pitch strength (and pitch height) has been documented.[47] Perceived pitch strength varies widely across normal and dysphonic voices, and is strongly correlated with judgments of perceived breathiness (Pearson’s $r = -0.99$, $P < 0.001$) and roughness (Pearson’s $r = -0.90$, $P < 0.005$).[1] While pitch height and pitch strength are perceptual constructs, these can be estimated through the use of appropriate algorithms specifically developed for this purpose[48] (Auditory-Sawtooth Waveform Inspired Pitch Estimator Prime [Aud-SWIPE'[49]]). Note that these approaches differentiate f_0 (a physical property of the signal) from its pitch (its psychoacoustic attribute), even though the two may be highly correlated. In this study, we sought to determine whether such estimates of pitch and pitch strength may be useful in describing dysphonic voices even when these do not have a clearly visible harmonic structure or a sufficiently periodic waveform (ie, type 2 and 3 signals). We also explored whether these might help differentiate dysphonic voices into three distinct signal types as described by Titze.[1] By equating the signal typing descriptions of type 1, 2, and 3 signals to the spectral, virtual, and noise pitches described by Fastl and Stoll,[44] it is hypothesized that type 1 signals will be associated with the strongest pitch sensation or pitch strength and type 3 voice signals will be associated with the lowest pitch sensation or pitch strength. Given that many dysphonic voices are aperiodic in nature, establishing the signal type will support the selection of valid acoustic measures. Likewise, knowledge of signal type can guide the exploration of new analytic methods and help to establish which methods generalize to multiple voice types.

GENERAL METHODS

Stimuli

A total of 36 voices (sustained vowel phonations /a/) from the University of Florida Dysphonic Voice Database were selected for this study based on stratified random sampling procedures. These voices had previously been rated by three expert voice scientists for breathiness, roughness, and strain on a five-point scale (1 being “normal” and 5 being “severe”). For the current study, voices with ratings of either “1” on all voice quality dimensions, or voices with ratings of “2” or higher on at least two voice quality dimensions were included to ensure that the stimuli represented a continuum of signal types. Voices were excluded if the speaker diagnosis was spasmodic dysphonia or vocal tremor to avoid signals with multiple vowel onsets and offsets.

[1] Note that several of the rough voice samples in this study could be characterized as breathy too, while the breathy examples rarely were also considered rough. Thus, it is not clear whether the pitch strength associated with roughness was related to roughness per se, to a breathy component in primarily rough stimuli, or to an overall severity.

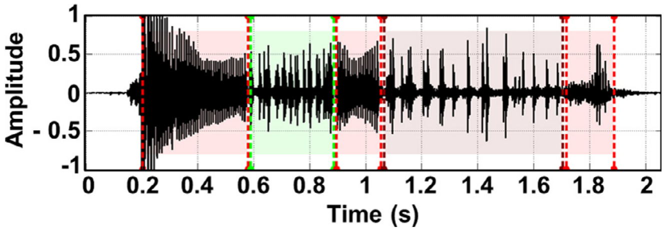

FIGURE 2. Sample graphical output from the MATLAB GUI representing the manual marking capability to identify multiple signal types within the same voice signal.

Signal typing

Three judges (one undergraduate, one masters, one doctoral student) from Michigan State University (all female; ages 20−29) were trained by an expert in signal typing (E.J.H.)* who had greater than 20 years of experience in signal typing. All judges had: (a) at least 6 months of experience with perceptual and acoustic analysis of normal and disordered voice quality and (b) hearing within normal limits as assessed using pure tone audiometry at frequencies ranging from 250 to 8000 Hz.[50] The standard method of signal typing[1] (based on visualization and listening to the signals was expanded to allow parsing of a single signal recording into more than one segment. Thus, rather than constraining the judges to one signal type judgment for the entire recording, they could parse the signal into multiple segments and assign different signal types to different segments. To do so, they were able to visualize the overall waveform and narrowband spectrogram in a two-panel display rendered by the TF32 software.[51] At the same time, they also visualized the same high quality waveform representation and a lesser-quality narrowband spectrogram in a customized MATLAB graphical user interface. Both applications supported listening to the entire signal waveform. The graphical user interface, however, allowed each judge to mark the initial and final time points of segments judged to have different signal types and then to identify the signal type of each of the segments (Figure 2). The separate segments were then saved as individual sound files for subsequent acoustic analyses.

Following the initial signal typing process by all three judges, a consensus session was held to resolve any inconsistencies in judgments. If at least two judges assigned the same signal type for a given segment of a voice signal, that signal type was retained. If each judge assigned a different signal type to a given segment, then they discussed the features that led to their signal typing. If they came to an agreement, the agreed upon signal type was assigned. If an agreement could not be reached, that segment was excluded from further analysis. This process led to a total of 61 segments from the original 36 voices that were agreed upon by the three judges. As shown in Table 1, the number of segments, as well as segment length, varied by signal type. To avoid potential effects of variations in the number of segments on further analysis, the 10 longest segments from each signal type were selected for the study. To avoid the potential influence of variable signal duration, the center 400 ms of each stimulus was extracted from each of the 10 segments using a custom MATLAB script.

EXPERIMENT 1: SIGNAL TYPE AND PITCH STRENGTH

Experiment 1 sought to determine whether computational or perceptual estimates of pitch strength associated with individual voice segments could be used to reliably characterize type 1, 2, and 3 signals in accordance with subjective judgments. Because estimates of pitch strength are proportional to periodicity,[42] the predicted outcome was an inverse relationship between signal type and estimated pitch strength.

Methods

The methods below are presented in an order that is consistent with subsequent presentation of the experimental results.

Computational estimates of pitch strength

The Aud-SWIPE'[49] was used to estimate the pitch height and the pitch strength of each of the 30 stimuli (10 per signal type) chosen from the larger set of 61 segments. This algorithm is outlined in Figure 3. The auditory front-end is comprised of cascaded filters (modeling outer and middle ear transfer functions), a time-aligned gammatone filterbank (simulating cochlear filtering), half-wave rectification (simulating the mechanical to electrical transduction process in the cochlea), and channel-dependent weighted low-pass filters. Following principles detailed by Moore, Glasberg, and Baer,[52] the spectrum is converted to specific loudness on an equivalent rectangular bandwidth scale to transform the representation from a physical to a perceptual one. Subsequent to this auditory processing front-end, the algorithm compares, via cross-correlation, the loudness-based spectrum of a dysphonic voice sample to the loudness-based spectrum of a family of sawtooth kernel functions constructed over a range of f_0 values (pitch candidates). Only the prime harmonics are included in the analysis to

TABLE 1.
Number and Length of Segments for Each Signal Type

Signal Type	Number of Segments	Minimum Length	Maximum Length
Type 1	26	0.07 s	5.47 s
Type 2	16	0.10 s	5.10 s
Type 3	19	0.04 s	5.04 s

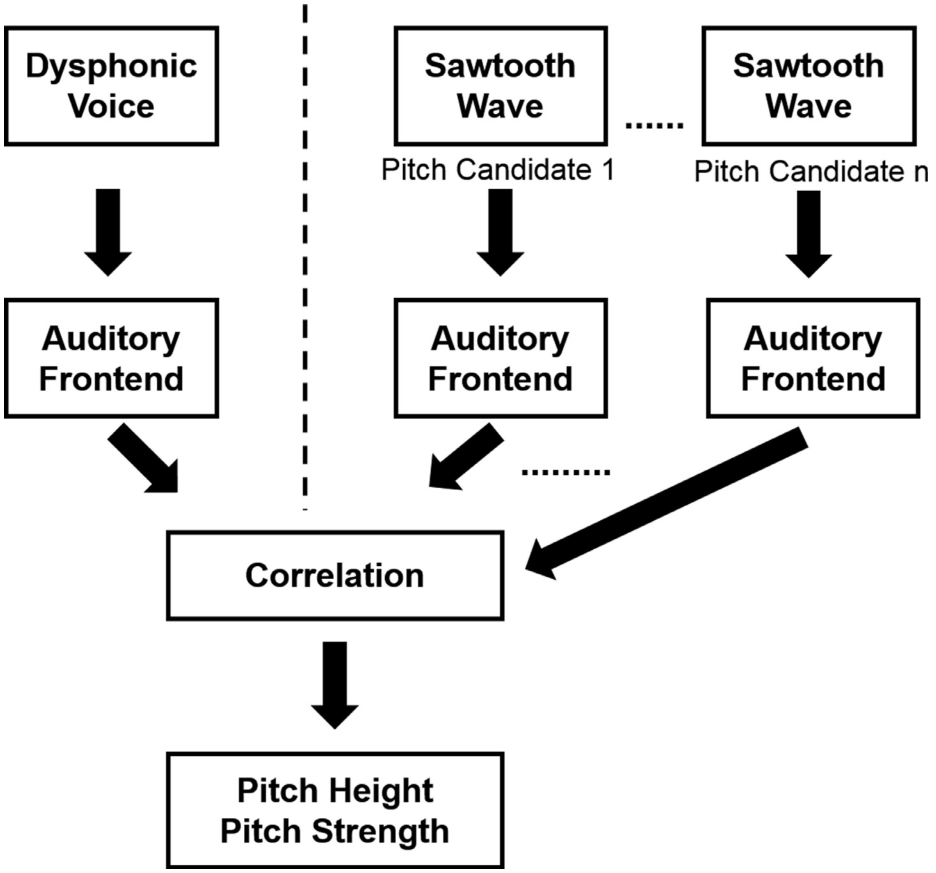

FIGURE 3. A schematic diagram of the Auditory-Sawtooth Inspired Pitch Estimator Prime (Aud-SWIPE') algorithm. As described in the text, the auditory frontend consists of multiple cascaded filters, half-wave rectification, and conversion to a frequency-dependent loudness representation. Also refer to Camacho.[49]

minimize halving and doubling errors associated with estimating f_0. The f_0 of the candidate waveform with the highest degree of similarity (correlation between 0 and 1) between the processed sawtooth waveform and identically processed spectrum of the input signal is taken as the estimated pitch height, and the value of that correlation is taken as the estimate of the pitch strength. Thus, pitch height is in units of Hz and pitch strength is a number between 0 (minimum pitch strength) and 1 (maximum pitch strength).

Perceptual estimates of pitch strength

Listeners: Perceptual estimates of pitch strength for the same 30 stimuli were obtained from 10 young adult listeners (9 female and 1 male; mean age of 25 years). All listeners were native speakers of American English and passed a hearing threshold screening (≤20 dB HL between audiometric frequencies of 250 and 4000 Hz). All procedures were approved by the Institutional Review Board at the University of South Florida and all listeners voluntarily consented to and were compensated for their participation. Prior to this experiment, listeners had no prior experience judging the pitch strength of general sounds or dysphonic voices.

Instrumentation: This experiment was controlled using TDT System III hardware and TDT *Sykofizx* software. Stimuli were delivered at 85 dB SPL in the right ear via ear inserts (ER-2, Etymotic Research) and perceptual testing was performed in a double-walled sound attenuating booth.

Procedure: Perceptual estimates of pitch strength were obtained using an anchored magnitude estimation task.[53,47,54] On a given trial, listeners heard three stimuli separated by 500 ms silent intervals. The first anchor stimulus was a wideband noise with a low pitch strength value (0). The second stimulus was the test stimulus whose pitch strength value was assigned by the listener. The third stimulus was a pure tone (1000 Hz) with a high pitch strength value (1) and served as a second anchor. Listeners judged the pitch strength of the test stimulus on each trial by positioning a continuous slider between the values of 0 and 1 with 101 intervals.

Given that listeners had no prior experience in judging pitch strength, a familiarization task was developed mirroring the main experiment. In this task, listeners judged pitch strength of five iterated rippled noise (IRN) stimuli. IRN stimuli are generated by attenuating and adding a delayed version of a broad-band noise to itself.[44,53–56] Among, the multiple parameters that can be used to manipulate the pitch strength of IRN stimuli, the current study varied the attenuation of each iteration of the noise following Shrivastav et al[47] Judgments of pitch strength were obtained in three separate 50 trial-blocks. For each block, each of the five IRN stimuli, corresponding to five levels of attenuation between −16 and 0 dB, was presented 10 times in random order across listeners.

Thus, each listener completed a total of 150 judgments (5 attenuation levels × 10 repetitions × 3 blocks). The perceived pitch strength for each IRN was based on the average of 30 repetitions (3 blocks × 10 repetitions). Listener responses were reviewed after the familiarization task to verify that perceived pitch strength scores systematically varied across the attenuation levels (from 0 dB or high pitch strength down to −16 dB or low pitch strength). Following the familiarization task, listeners judged the pitch strength of 30 voice stimuli varying in signal type using the same anchored magnitude estimation task. Each stimulus was judged 10 times in random order, resulting in a total 300 stimuli per listener (10 stimuli × 3 signal types × 10 repetitions). Perceived pitch strength was averaged across the 10 repetitions to obtain a single pitch strength value for each of 30 stimuli. The entire testing spanned approximately 1.5 hours. Short breaks throughout the listening session were provided to minimize listener fatigue.

Computational estimates of cepstral peak prominence

This study also evaluated the relationship between the cepstral peak prominence (CPP) and the computational and perceptual estimates of pitch strength. CPP values were estimated for all stimuli twice using the algorithms described by Hillenbrand et al[26] and Boersma and Weenik[57] (as implemented in the PRAAT software).

Results

Computational estimates of pitch strength

Figure 4 depicts the signal type determined by the three judges (abscissa) and the pitch strength estimated from the Aud-SWIPE' model (ordinate). Above each signal type, the mean (solid symbol) and standard error (bar) is shown on the left and the individual data (open circles) are shown on the right. A univariate analysis of variance (ANOVA) revealed a statistically significant difference between pitch strength values for the three signal types ($F_{(2,\ 27)} = 128.05$, $P< 0.001$). *Posthoc* Bonferroni analyses revealed that type 1 signals (mean = 0.45, standard deviation, SD = 0.06) had significantly ($P< 0.001$) higher pitch strength than type 2 (mean = 0.17, SD = 0.07) or type 3 signals (mean = 0.06, SD = 0.03) and that pitch strength was significantly ($P< 0.001$) higher for type 2 than type 3 signals.

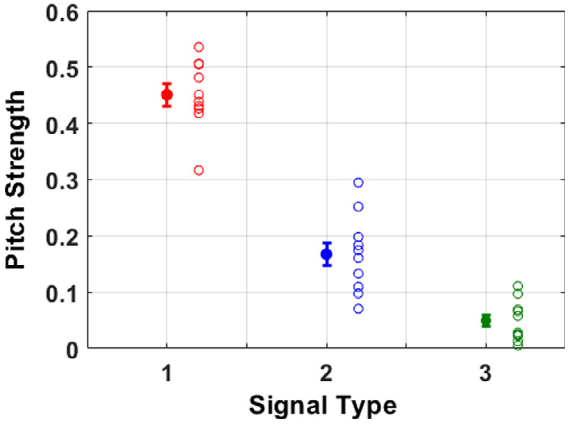

FIGURE 4. Computational estimates of pitch strength by signal type. Error bars indicate mean ± standard error (SE) and the open symbols/markers represent the individual data and range.

Perceptual estimates of pitch strength

Listener reliability: Intra- (comparison across repetitions) and inter- (comparison across listeners) listener reliability were measured using intraclass correlation coefficients[58] (for both the familiarization task and the main experiment, as shown in Table 2. Data from one of the listeners was excluded due to low intralistener reliability, resulting in N = 9. Those 9 listeners were highly reliable in making perceptual judgments of pitch strength in both familiarization task and the main experiment.

Perceived pitch strength: The results of the familiarization task (ie, perceived pitch strength) are shown in Figure 5 with pitch strength (ordinate) shown as a function of IRN attenuation value (abscissa) by red circles (mean) and error bars (standard error). These data indicate that listeners were able to understand the concept of pitch strength and provided judgments that revealed a systematic decrease in perceived pitch strength along the continuum of IRN attenuation values. The blue squares and error bars in Figure 5 show corresponding values from the same task reported by Shrivastav et al,[47] indicating consistency across participant groups. Despite using the same stimuli, measurement methods, and earphones, there were slight differences in the mean absolute pitch strength estimates between the two studies. Examination of the individual data for the current study (unfilled circles) demonstrates considerable overlap with the mean (and presumed distribution) of thresholds reported by Shrivastav et al.[47] The most parsimonious explanation for differences among mean values across studies is simply that they reflect natural individual variation in the respective listener's perceptual judgments.

Perceived pitch strength judgments for the 30 voice stimuli are shown in Figure 6 in the same manner as the computational estimates in Figure 4. Perceived pitch strength decreased with signal type. A univariate ANOVA was used to examine the effects of signal type (independent variable) on perceived pitch strength (dependent variable). The

TABLE 2.
Intra- and Interlistener Reliability for Perceived Pitch Strength Described by Intraclass Correlation, ICC (2, K)

Reliability	Intralistener (Mean ± SD)	Interlistener (Mean)
Familiarization task	0.983 ± 0.011	0.933
Main experiment	0.985 ± 0.003	0.960

For intralistener reliability, K = 30 repetitions for the familiarization task and 10 repetitions for the main experiment. For interlistener reliability, K = 10 listeners.
Abbreviation: SD, standard deviation.

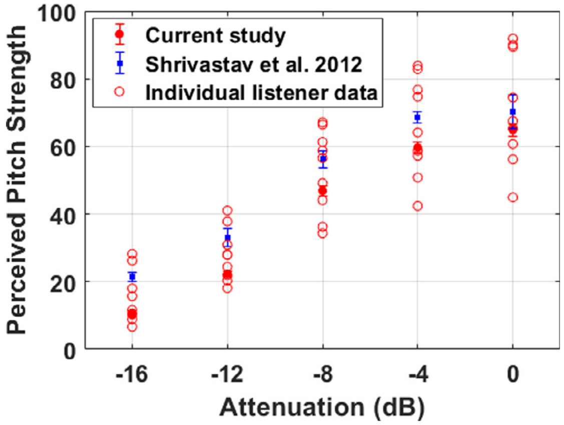

FIGURE 5. Results of the familiarization task using IRN stimuli. Perceived pitch strength is on the *y* axis and the IRN attenuation value is on the x axis. Symbols/markers indicate mean perceived pitch strength using the anchored magnitude estimation task for the current study (circles, bars show standard error, SE) and data from (squares).[47]

Levene's F test revealed unequal variances in perceived pitch strength ($F = 4.993$, $P = 0.007$). Therefore, a *Welch's F* test with an alpha level of 0.05 was used. There was a statistically significant main effect of signal type on pitch strength judgments (Welch's $F_{(2,175.97)} = 339.59$, $P < 0.001$), indicating that perceived pitch strength differed among the three signal types. *Posthoc* comparisons, using Games-Howell posthoc procedure was conducted to determine which pairs of signal types differed significantly. All pair-wise comparisons were significantly different ($P < 0.001$).

Relationship between computational and perceptual estimates of pitch strength

The computational estimates of pitch strength obtained with the Aud-SWIPE′ model were compared to the perceptual estimates of pitch strength for the same type 1, 2, and 3 voice stimuli. There were significant correlations (Pearson's r) between the perceptual and computational estimates for each signal type (type 1: $r = 0.76$; $P = 0.01$; type 2: $r = 0.65$; $P = 0.04$; type 3: $r = 0.89$; $P = 0.001$), illustrating the potential utility of these computational estimates in voice analyses.

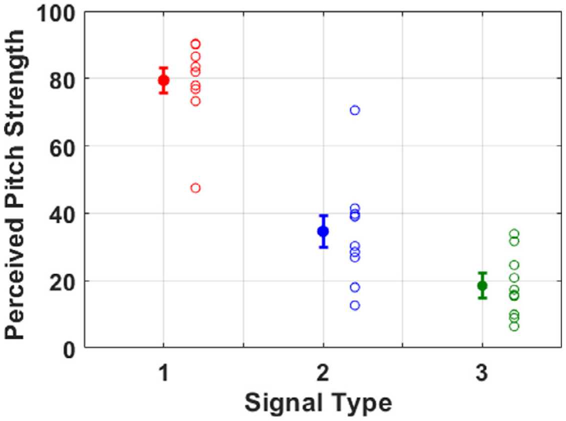

FIGURE 6. Perceptual estimates of pitch strength by signal type. Error bars indicate Mean ± standard error (SE) and the open symbols/markers represent the individual data and range.

Relationship between pitch strength and CPP estimates

Computational estimates of pitch strength and of CPP decreased as signal type increased as expected, reflecting an overall decrease in periodicity from type 1 to type 3. This resulted in strong negative correlations between signal type and computational estimates of pitch strength ($PS = -0.93$, $P < 0.001$) and between signal type and CPP ($PRAAT = -0.81$, $P < 0.001$; $HB = -0.60$, $P < 0.001$). The relationship between computational estimates of pitch strength and CPP, as well as perceptual estimates of pitch strength and CPP, are shown in Table 3. Overall, Pearson's r correlations were moderate to strong across all signal types and the strength of the correlations varied slightly depending on the CPP algorithm. Because pitch strength and CPP should decrease with a decrease in periodicity, one would expect that the correlation among computational measures should not change much with signal type. Surprisingly, however, for type 3 voices, there was a negative correlation between CPP estimates from Hillenbrand et al[26] and both computational and perceptual estimates of pitch strength. In fact, the correlation between CPP estimates from Hillenbrand et al[26] and PRAAT are similarly negative for type 3 voices (not shown in Table 3).

Discussion

The purpose of this experiment was to determine the relationship between voice signal type and perceptual and computational estimates of pitch strength. Computational estimates of pitch strength significantly and monotonically decreased with signal type (ie, as the signal aperiodicity increased), confirming our hypothesis. Similar to the computational estimates, perceived pitch strength decreased systematically with signal type.

Unlike f_0 based algorithms, which increasingly fail with type 2 and 3 voices, pitch strength can be estimated from voices of all signal types. Furthermore, these results reveal that pitch strength can differentiate all signal types, similar to the nonlinear dynamic metrics such as *D2* and rate of divergence.[12,13] The utility of pitch strength estimates in evaluating dysphonic voices was demonstrated previously using different types of synthetic stimuli (varying in level of aspiration noise, open quotient, and the combination of both) and natural dysphonic stimuli.[39] For both stimulus types, the relationship between pitch strength estimates and perceived breathiness revealed correlations ranging from 0.82 to 0.97 (Pearson's r). Stimuli with lower pitch strength values were perceived to be higher in breathiness. Here we show that automated computational pitch strength estimates can be used to characterize severely disordered type 2 and 3 voices where most conventional acoustic analyses fail primarily due to limited periodicity. These results also are consistent with studies that have evaluated pitch strength on a continuum of stimuli including nonspeech tonal complexes and noise stimuli.[45] Per signal

TABLE 3.
Relationship Between CPP (dB) and Computational and Perceptual Pitch Strength Estimates (Pearson's *r*[Significance])

	Computational Pitch Strength Estimates vs CPP		Perceptual Pitch Strength Estimates vs CPP	
Signal Type	Hillenbrand et al[26]	PRAAT	Hillenbrand et al[26]	PRAAT
Type 1	0.525 (P> 0.05)	0.535 (P > 0.05)	0.765 (P = 0.010)	0.600 (P > 0.05)
Type 2	0.713 (P = 0.02)	0.739 (P = 0.02)	0.516 (P > 0.05)	0.729 (P = 0.017)
Type 3	−0.328 (P > 0.05)	0.403 (P > 0.05)	−0.400 (P > 0.05)	0.610 (P > 0.05)

typing and CPP, Stone et al[10] reported a strong, negative correlation between signal typing and CPPS (smoothed version of CPP) value of /a/ (r = −0.85, P< 0.001); participants with type 3 and 4 signals had lower CPPS values. The results of the current study are consistent with those of Stone et al[10] Of the three methods (pitch strength, CPP from Hillenbrand et al,[26] and CPP from PRAAT), the correlation value (correlation with signal type) was highest for pitch strength estimates (r = −0.93, P< 0.001), providing evidence that pitch strength estimates may be better at characterizing signal types than CPP.

EXPERIMENT 2: SIGNAL TYPE AND PITCH HEIGHT

Experiment 2 was designed to determine whether computational and perceptual estimates of pitch height associated with individual voice segments could be used to successfully characterize type 1, 2, and 3 voice signals. Based on the fact that pitch height can be judged for a wide range of periodic and aperiodic stimuli,[47] Experiment 2 tested the hypothesis that type 2 and 3 signals elicit a pitch height that can be reliably estimated through perceptual and computational measures.

Methods

Computational estimates of pitch height

The Aud-SWIPE′ algorithm[49] was used to compute the pitch height values for the 30 stimuli (10 per signal type). A detailed description of the algorithm and its use for estimating pitch height is provided in Experiment 1 methods. In addition to an estimated pitch value from the Aud-SWIPE', f_0 estimates were computed from the TF32 and PRAAT software applications. TF32 estimates f_0 using linear predictive coding, with voicing decisions based on zero crossings and signal amplitude. Additionally, crosscorrelation analysis is used to reduce formant artifacts. PRAAT estimates f_0 from multiple steps. First, a normalized autocorrelation function is used. The autocorrelation of the windowed signal is then divided by the autocorrelation of the window itself. Next, a 'sin x/x' interpolation is used. Finally, multiple pitch candidates are compared (default number is four), removing potential octave errors.

Perceptual estimates of pitch height

Listeners, stimuli, and instrumentation: All were identical to those of Experiment 1 except for software. In this experiment, MATLAB was used to present stimuli.

Procedure: Prior to the single-variable matching task of Experiment 2, listeners completed a short training task that mirrored the main experimental task but with a different set of stimuli. On each trial, listeners were presented with two sounds—a reference sound and a comparison sound. For the training task, a total of four reference sounds were selected; two natural dysphonic /a/ phonations and two clarinet sounds. For the main experiment, the reference sound was one of 30 stimuli that varied in periodicity. The comparison sound was a sawtooth tone with variable f_0 up to 4000 Hz. Pitch height was operationally defined as the fundamental frequency of this sawtooth, in units of cycles per second or Hertz (Hz).[59] For both sets of stimuli, listeners were instructed to increase or decrease the f_0 of the comparison sound (single-variable parameter) such that the perceived pitch of the comparison sound approximated the perceived pitch of the reference sound. The initial f_0 of the comparison sound was randomly chosen over the range of 50−500 Hz. The frequency of the comparison sound was varied according to the listener response in steps of 50, 20, and 2 Hz and the final pitch match value was based on the average of three separate pitch matches for each stimulus. The reference sounds were presented in random order across listeners. The entire testing duration was approximately 1 hour.

Results

Computational estimates of pitch height

The mean pitch height values for each voice stimulus estimated from the Aud-SWIPE' model are shown in column 2 of Table 4 for the type 1 signals along with the f_0 values produced by the TF32 algorithm (third column) and the PRAAT algorithm (fourth column). Columns five, six, and seven show differences among the three estimators. For most of the type 1 signals, estimates of pitch height from the Aud-SWIPE′ algorithm and estimates of f_0 from the two commonly used algorithms were in close agreement (differences ranged from 0 to 6.3 Hz). Correlations among the methods were each r = 0.9997, P< 0.001).

The mean pitch height and f_0 values for type 2 voices are shown in Table 5 in the same manner as for type 1 voices in Table 4. It is clear from the difference values that there is considerable discrepancy among algorithms (differences ranged from 0 to 76 Hz). The correlations among methods for these type 2 voices remained high, however, ranging from r = 0.94 to r = 0.97. The results for type 3 voices (Table 6) reveal even greater absolute differences among

TABLE 4.
Computational Estimates of Mean Pitch Height and f_0 for type 1 Signals Measured From three Algorithms as well as the Difference Between Each Pair of Algorithms

Measure	Mean Results (Hz)			Subtraction of Means (Hz)		
Algorithm	Aud-SWIPE′	TF32	PRAAT	Aud-SWIPE′-TF32	Aud-SWIPE′-PRAAT	TF32-PRAAT
Signal Type/Voice Stimulus	Mean	Mean	Mean	Diff	Diff	Diff
Type 1_Stimulus 1	117.2	117.3	117.3	−0.1	−0.1	0
Type 1_Stimulus 2	133.1	135.5	135.5	−2.4	−2.4	0
Type 1_Stimulus 3	91.1	91.1	91.1	0.0	0.0	0
Type 1_Stimulus 4	92.7	92.8	92.9	−0.1	−0.2	−0.1
Type 1_Stimulus 5	102.9	102.9	102.9	0.0	0.0	0
Type 1_Stimulus 6	100.9	100.9	100.8	0.0	0.1	0.1
Type 1_Stimulus 7	93.7	93.8	93.8	−0.1	−0.1	0
Type 1_Stimulus 8	165.4	168.1	168.1	−2.7	−2.7	0
Type 1_Stimulus 9	174.2	180.5	180.5	−6.3	−6.3	0
Type 1_Stimulus 10	88.3	88.4	88.3	−0.1	0.0	0.1

Abbreviation: Diff, difference.

TABLE 5.
Computational Estimates of Mean Pitch Height and f_0 for type 2 Signals Measured From three Algorithms as well as the Difference Between Each Pair of Algorithms

Measure	Mean Results (Hz)			Subtraction of Means (Hz)		
Algorithm	Aud-SWIPE′	TF32	PRAAT	Aud-SWIPE′-TF32	Aud-SWIPE′-PRAAT	TF32-PRAAT
Signal Type/Voice Stimulus	Mean	Mean	Mean	Diff	Diff	Diff
Type 2_Stimulus 1	129.0	137.4	135.5	−8.4	−6.5	1.9
Type 2_Stimulus 2	236.4	224.6	269.5	11.8	−33.1	−44.9
Type 2_Stimulus 3	82.5	93	106.6	−10.5	−24.1	−13.6
Type 2_Stimulus 4	108.4	108.6	108.4	−0.2	0.0	0.2
Type 2_Stimulus 5	173.2	179	178.8	−5.8	−5.6	0.2
Type 2_Stimulus 6	186.8	196.2	194.6	−9.4	−7.8	1.6
Type 2_Stimulus 7	78.9	91.9	94.7	−13.0	−15.8	−2.8
Type 2_Stimulus 8	105.7	138.3	110	−32.6	−4.3	28.3
Type 2_Stimulus 9	127.8	161.2	204.1	−33.4	−76.3	−42.9
Type 2_Stimulus 10	239.0	255.8	253.6	−16.8	−14.6	2.2

Abbreviation: Diff, difference.

algorithms (ranging from 0 to 330 Hz when values could be estimated). The correlation between values estimated by the Aud-SWIPE' and the TF32 algorithms (for the 9 of 10 stimuli for which values could be obtained by TF32) was $r = 0.54$ and was not computed for the PRAAT algorithm due to the missing data associated with the failure of that algorithm to identify a f_0 candidate.

While conventional software was unable to measure f_0 for many of the aperiodic type 3 signals, the Aud-SWIPE′ algorithm was able to provide pitch height estimates for all stimuli within a range of values that is plausible for the given stimuli. Figure 7 illustrates the performance of each of the algorithms by showing the pitch height trace from Aud-SWIPE′ and f_0 traces from TF32 and PRAAT software for a sample stimulus from each of the three signal types. The quality of the traces for the three signal types is consistent with the differences and correlations from Tables 4–6. It is evident that for type 1 signals, the pitch height and f_0 traces were in close agreement and were consistent over time. For type 2 signals, computational estimates of f_0 using TF32 and PRAAT were less consistent than for type 1 signals while the pitch height estimates from Aud-SWIPE′ were more consistent over time. For type 3 signals, while f_0 could not be estimated using either TF32 or PRAAT for the chosen voice, Aud-SWIPE′ was able to produce a consistent and moderately robust pitch height estimate. Although Aud-SWIPE′ was able to track “pitch height” for some portion of the voice stimulus for all talkers in the current experiment, the accuracy of this measurement will need to be validated against perceptual data.

Perceptual estimates of pitch height

To keep the number of listeners in this experiment consistent with that of Experiment 1 (perceived pitch strength), the same listener was excluded from analyses of this

TABLE 6.
Computational Estimates of Mean Pitch Height and f_0 for type 3 Signals Measured From three Algorithms as well as the Difference Between Each Pair of Algorithms

Measure	Mean Results (Hz)			Subtraction of Means (Hz)		
Algorithm Signal Type/Voice Stimulus	Aud-SWIPE′ Mean	TF32 Mean	PRAAT Mean	Aud-SWIPE′-TF32 Diff	Aud-SWIPE′-PRAAT Diff	TF32-PRAAT Diff
Type 3_Stimulus 1	76.6	155.8	92.7	−79.2	−16.1	63.1
Type 3_Stimulus 2	77.9	0	*	77.9	†	†
Type 3_Stimulus 3	56.2	97.7	161.3	−41.5	−105.1	−63.6
Type 3_Stimulus 4	52.6	44.1	*	8.5	†	†
Type 3_Stimulus 5	45.4	57.3	*	−11.9	†	†
Type 3_Stimulus 6	275.0	152.6	299.3	122.4	−24.3	−146.7
Type 3_Stimulus 7	83.4	171.7	245.5	−88.3	−162.1	−73.8
Type 3_Stimulus 8	97.7	73.6	97.9	24.1	−0.2	−24.3
Type 3_Stimulus 9	157.1	255.6	157.4	−98.5	−0.3	98.2
Type 3_Stimulus 10	59.3	66.8	389.4	−7.5	−330.1	−322.6

* For some voice stimuli, PRAAT was unable to produce a f_0 trace (results = “undefined’).
† Unable to calculate because of “undefined” PRAAT results.
Abbreviation: Diff, difference.

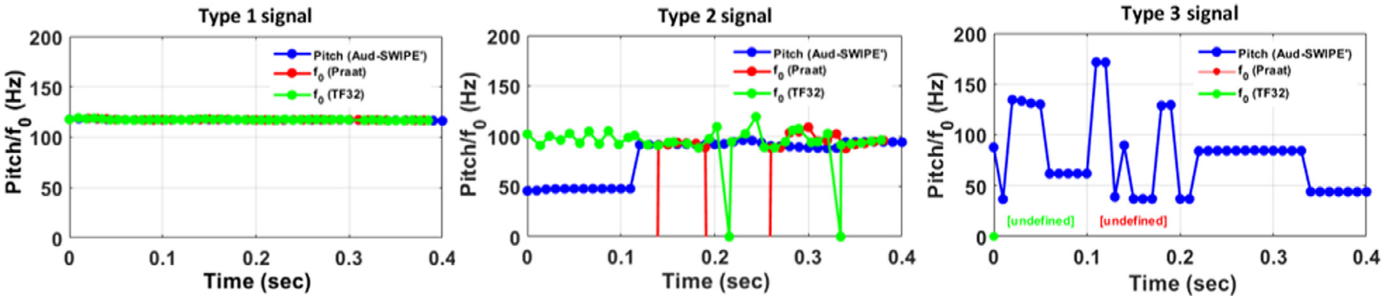

FIGURE 7. Pitch height or fundamental frequency (f_0) traces for a sample voice representative of type 1, 2, and 3 signals. For each signal type, pitch height estimated from Aud-SWIPE’ or f_0 estimated from PRAAT, TF32 software is depicted on y axis. For type 3 signal, f_0 estimates that could not be measured are indicated as (undefined).

experiment. Nine listeners matched the perceived pitch height of type 1, 2, and 3 signals with good accuracy. Figure 8 shows the perceived pitch height/pitch match (ordinate) of each of the 10 stimuli (abscissa) for all signal types (separate panels) in terms of the mean (symbols) and standard error (bars) of the pitch height judgments across nine listeners. The data reveal a wide range of perceived pitch height across the 10 stimuli for each signal type. The variability across listeners, however, differed among signal types (*Levene’s*test of Homogeneity, P= 0.004). Multiple comparisons revealed that variance was greater for type 3 than type 1 signals with no significant difference between type 1 and type 2 or type 2 and type 3 signals.

Relationship between computational and perceptual estimates of pitch height

Correspondence between computational estimates derived from Aud-SWIPE′, TF32, and PRAAT and the perceptual estimates of pitch height from listeners were evaluated using Pearson’s r correlation coefficients. For type 1 voices, the correlation between computational estimates of pitch height, f_0, and the listener judgments were strong (remarkably, the three pairwise correlations each were r > 0.88; P= 0.001). For type 2 voices, there were significant but lower correlations with listener judgments (ranging from r = 0.65 to 0.77; P= 0.001). For type 3 voices, computational estimates of pitch height from Aud-SWIPE′ were significantly correlated with listener judgments (r = 0.75; P= 0.01) but not f_0 estimates (P> 0.05).

Discussion

The purpose of this experiment was to determine the relationship between signal type and perceptual and computational estimates of pitch height. The Aud-SWIPE′ algorithm resulted in pitch height values for all stimuli irrespective of signal type, while conventional f_0 estimators (TF32 and PRAAT) failed to converge consistently on an estimate for type 2 and type 3 signals. The behavioral data revealed that listeners can judge the pitch height of type 1, 2, and 3 voice signal types with high consistency, even when

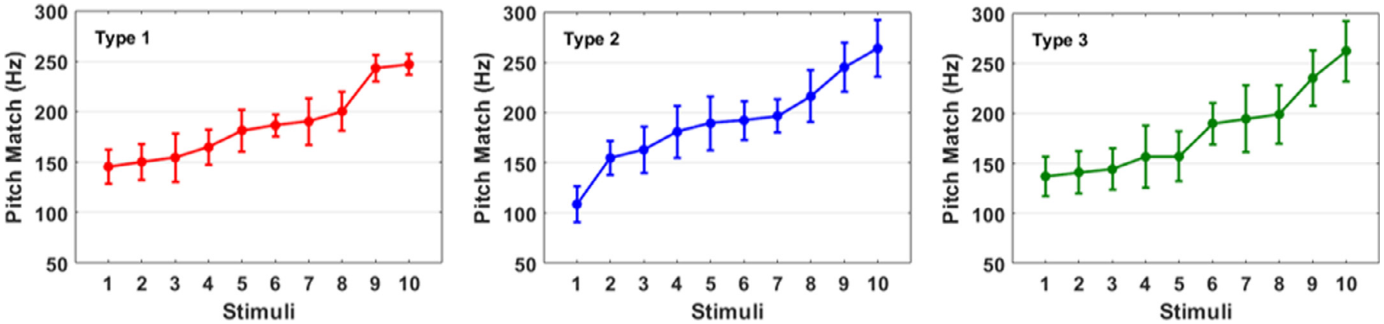

FIGURE 8. Perceptual estimates of pitch height for each of the 10 stimuli per signal type. Error bars indicate Mean±standard error (SE).

these signals may not provide reliable estimates of f_0 due to lack of periodicity (eg, type 2 and 3).

The SWIPE algorithm described by Camacho and Harris[60] estimates pitch height as the as the f_0 of the sawtooth waveform whose spectrum best matches the spectrum of the input signal (dysphonic voice). Thus, it explicitly does not base pitch height on the f_0 of the stimulus under study. Prior research has shown that SWIPE outperformed multiple algorithms in predicting the pitch of both normal and disordered voices.[60] Here, pitch height was estimated using an enhanced version of this model in which signal is preprocessed with a signal processing front-end that mimics processing by the auditory periphery (Aud-SWIPE′). Data from this experiment support the notion that pitch, purely a perceptual phenomenon, does not require the presence of a periodic acoustic signal. Thus, an auditory inspired model of pitch perception can accurately predict human pitch perception even when computational estimates of f_0 occasionally fail.

GENERAL DISCUSSION

Despite the widespread use of acoustic analysis routines to characterize or quantify voices, their use is limited to the subset of voices that have some degree of periodicity (type 1). However, a large number of dysphonic voices, particularly those that are perceived to be most severely dysphonic, are characterized by subharmonics (type 2) or aperiodic sources (type 3), rendering many conventional analysis techniques inaccurate or simply invalid. For such voices, clinicians and researchers are often limited to the use of perceptual judgments such as descriptions of sound quality or visual judgments of their spectrogram. Such measurements create other challenges, including issues around poor accuracy and reliability of measurements, replicability, or speed of measurement. In the last decade, measures or indices of dysphonia severity calculated from the spectral or cepstral domain have gained popularity, partly because these demonstrate greater robustness to the lack of periodicity in acoustic signals. However, these measures also lose sensitivity to changes in vocal signals when voices have aperiodic sound sources. Analytical methods using nonlinear approaches show promise in differentiating amongst the three types of vocal acoustic signals, but these have made limited inroads in routine clinical care or research.

The current study demonstrates a different approach to characterize and quantify vocal acoustic signals that can be applied to all voice types, from the primarily periodic (type 1) signals to the most aperiodic (type 3) signals. These measurements can be made using a typical acoustic recording, and require no additional hardware than is commonly used for conventional acoustic analyses of voices. The primary differentiation between this approach and conventional acoustic analyses is the use of perceptual judgments or psychoacoustic attributes of sound (pitch and pitch strength) as the basis of measurement instead of using a physical attribute (f_0) of the waveform. Since all sounds, irrespective of their periodicity, have a stable percept of pitch height and pitch strength, computational estimates of these attributes can serve as another way to describe and quantify dysphonic voices.

Given that many dysphonic voices are aperiodic in nature, such measurements can directly help quantify these voices or be used indirectly such as in supporting the selection of other acoustic measures that maybe appropriate for a particular voice. For example, voices can be characterized by describing their pitch values, even when these do not have a clear f_0 associated with the waveform. Similarly, pitch strength values can be useful for differentiating voice types, or segments of steady phonation that are likely to represent types 1, 2, or 3, phonation. Once computed, the relative proportion of type 1, 2, or 3 segments may itself serve as a measure of dysphonia. Dysphonic voices characterized by predominately type 1 segments may be closer to "normal" phonation than those which have a greater presence of type 2 and 3 segments.[8] Such segmentation can also help improve the use of other acoustic measurements, such as by limiting the calculation of specific acoustic measurements only to those segments where a particular analysis is likely to be valid and accurate.

Even voices with nonlaryngeal sources, such as in adult and pediatric patients with glottic and supraglottic cancer or laryngotracheal stenosis could benefit from the use of these methods. Note that such conditions often lead to significant reconstructive surgery, resulting in supraglottal voice source (s) and voice source characterized as chaotic. Indeed, several research studies in such populations have demonstrated the predominance of type 2 and 3 aperiodic signals and description of these voices are typically limited to perceptual evaluation of voice quality. For example, over 60% of patients with tracheoesophageal speech have been reported to have aperiodic voices.[5,6] Similar results were reported for adult patients with early glottic cancer (type 2 and 3 signals in 13/14 patients[10]; and stenosis (type 2 and 3 signals in 11/11 patients,[7] as well as pediatric patients postairway reconstruction.[61–63] Only 32% of samples were deemed suitable for conventional acoustic analysis in Brehm et al[61] and 20/21 children were observed to have aperiodic voices in Kelchner et al.[63]

While these still need to formally evaluated, using perceptually-motivated approaches for signal analysis may offer several other advantages. First, these can easily be extended to connected speech,[64] and are likely to be more robust to signal degradation, such as due to environmental noises. Second, these may also be more robust to differences in hardware (eg, differences in microphone type or location), as well as certain software or recording differences (eg, audio signal compression) than many other acoustic measures of voice. Together, the use of pitch height and pitch strength for characterizing voices can open new possibilities for voice analyses including automated analyses of long segments of speech and the analyses of field recordings using lower cost audio recording systems. To improve upon this work, future investigations might distinguish between type 3 and 4 signals,[2,65] use nonlinear dynamic analysis methods to characterize the vocal signal, and attempt to fully automate the signal typing method using pitch strength estimates in conjunction with pitch height and spectral metrics.

CONCLUSIONS

Existing methods for acoustic analysis of voices are limited to signals that are primarily periodic. While signal typing is recommended as an essential first-step in the analyses of dysphonic voices, this is often ignored. When it is conducted, the subjective approach is error-prone without extensive training and is tedious to complete. This study demonstrates the utility and success of pitch-based measures (pitch height and pitch strength) derived from the acoustic signal to characterize a wide range of dysphonic voices, and to differentiate among voice stimuli that vary across signal types. The computational and perceptual estimates of pitch strength were highest for the periodic, type 1 signals and lowest for the aperiodic, type 3 signals. Computational and perceptual estimates of pitch height could be evaluated for all signal types irrespective of periodicity. Pitch-based computational metrics are universally applicable to all dysphonic voices, and can be a valuable tool for clinicians and researchers.

Acknowledgments

The authors wish to thank Dr. Eric J. Hunter (E.J.H)* for training students on signal typing, Dr. Mark D. Skowronski for MATLAB processing, and Ms. Madison Dyjak for collecting the perceptual data.

REFERENCES

1. Titze IR. *Workshop on Acoustic Voice Analysis: Summary Statement*. National Center for Voice and Speech; 1995.
2. Sprecher A, Olszewski A, Jiang JJ, et al. Updating signal typing in voice: addition of type 4 signals. *J Acoust Soc Am*. 2010;127:3710–3716.
3. Behrman A, Agresti CJ, Blumstein E, et al. Microphone and electroglottographic data from dysphonic patients: type 1, 2 and 3 signals. *J Voice*. 1998;12:249–260.
4. Carding P, Steen I, Webb A, et al. The reliability and sensitivity to change of acoustic measures of voice quality. *Clin Otolaryngol Allied Sci*. 2004;29(5):538–544.
5. Clapham RP, van As-Brooks CJ, van Son RJ, et al. The relationship between acoustic signal typing and perceptual evaluation of tracheoesophageal voice quality for sustained vowels. *J Voice*. 2015;29:517.e523–517.e529.
6. D'Alatri L, Bussu F, Scarano E, et al. Objective and subjective assessment of tracheoesophageal prosthesis voice outcome. *J Voice*. 2012;26:607–613.
7. Houlton JJ, De Alarcon A, Johnson K, et al. Voice outcomes following adult cricotracheal resection. *Laryngoscope*. 2011;121:1910–1914.
8. Kopf LM, Jackson-Menaldi C, Rubin AD, et al. Pitch strength as an outcome measure for treatment of dysphonia. *J Voice*. 2017;31:691–696.
9. Ma EPM, Yiu EML. Suitability of acoustic perturbation measures in analysing periodic and nearly periodic voice signals. *Folia Phoniatr Logop*. 2005;57:38–47.
10. Stone D, McCabe P, Palme CE, et al. Voice outcomes after transoral laser microsurgery for early glottic cancer—considering signal type and smoothed cepstral peak prominence. *J Voice*. 2015;29:370–381.
11. Zacharias SR, Myer IV CM, Meinzen-Derr J, et al. Comparison of videostroboscopy and high-speed videoendoscopy in evaluation of supraglottic phonation. *Ann Otol Rhinol Laryngol*. 2016;125:829–837.
12. Calawerts WM, Lin L, Sprott J, et al. Using rate of divergence as an objective measure to differentiate between voice signal types based on the amount of disorder in the signal. *J Voice*. 2017;31:16–23.
13. Zhang Y, Jiang J. Nonlinear dynamic analysis in signal typing of pathological human voices. *Electron Lett*. 2003;39:1021–1023.
14. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. Cengage Learning; 2000.
15. Lin E, Jiang J, Hanson DG. Glottographic signal perturbation in biomechanically different types of dysphonia. *Laryngoscope*. 1998;108:18–25.
16. Ludlow CL, Bassich C, Connor NP, et al. The validity of using phonatory jitter and shimmer to detect laryngeal pathology. *Laryngeal Funct Phonation Respir*. 1987:492–508.
17. Rabinov CR, Kreiman J, Gerratt BR, et al. Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *J Speech Lang Hear Res*. 1995;38:26–32.
18. Rosen CA, Lombard LE, Murry T. Acoustic, aerodynamic, and videostroboscopic features of bilateral vocal fold lesions. *Ann Otol Rhinol Laryngol*. 2000;109:823–828.
19. Wolfe V, Fitch J, Cornell R. Acoustic prediction of severity in commonly occurring voice problems. *J Speech Lang Hear Res*. 1995;38:273–279.
20. Yiu EM. Limitations of perturbation measures in clinical acoustic voice analysis. *Asia Pac J Speech Lang Hear*. 1999;4:155–166.
21. Hanson HM. Glottal characteristics of female speakers: acoustic correlates. *J Acoust Soc Am*. 1997;101:466–481.
22. Kreiman J, Gerratt BR, Precoda K. Listener experience and perception of voice quality. *J Speech Hear Res*. 1990;33:103–115.
23. Pabon JP, Plomp R. Automatic phonetogram recording supplemented with acoustical voice quality parameters. *J Speech Hear Res*. 1988;31:710–722.
24. Shoji K, Regenbogen E, Yu JD, et al. High−frequency power ratio of breathy voice. *Laryngoscope*. 1992;102:267–271.
25. Heman-Ackah YD, Michael DD, Goding Jr. GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice*. 2002;16:20–27.
26. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Lang Hear Res*. 1994;37:769–778.
27. Noll AM. Short-term spectrum and "cepstrum" techniques for vocal pitch detection. *J Acoust Soc Am*. 1964:293–309.

28. Lowell SY, Kelley RT, Awan SN, et al. Spectral-and cepstral-based acoustic features of dysphonic, strained voice quality. *Ann Otol Rhinol Laryngol*. 2012;121:539–548.
29. Watts CR, Awan SN. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *J Speech Lang Hear Res*. 2011;54:1525–1537.
30. Dejonckere PH, Wieneke GH. Cepstra of normal and pathological voices: correlation with acoustic, aerodynamic and perceptual data. In: Ball MJ, Duckworth M, eds. *Advances in Clinical Phonetics*. Amsterdam: John Benjamins; 1996:217–226.
31. Krom G de. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J Speech Lang Hear Res*. 1993;36:254–266.
32. Maryn Y, Corthals P, Van Cauwenberge P, et al. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice*. 2010;24:540–555.
33. Maryn Y, De Bodt M, Roy N. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord*. 2010;43:161–174.
34. Awan SN, Roy N, Zhang D, et al. Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *J Voice*. 2016;30:130–144.
35. Peterson EA, Roy N, Awan SN, et al. Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. *J Voice*. 2013;27:401–410.
36. Shrivastav R. The use of an auditory model in predicting perceptual ratings of breathy voice quality. *J Voice*. 2003;17:502–512.
37. Shrivastav R, Sapienza CM. Objective measures of breathy voice quality obtained using an auditory model. *J Acoust Soc Am*. 2003;114:2217–2224.
38. Shrivastav R, Camacho A, Patel S, et al. A model for the prediction of breathiness in vowels. *J Acoust Soc Am*. 2011;129:1605–1615.
39. Eddins DA, Anand S, Camacho A, et al. Modeling of breathy voice quality using pitch strength estimates. *J Voice*. 2016;30:774.e771–774.e777.
40. Walker KM, Bizley JK, King AJ, et al. Cortical encoding of pitch: recent results and open questions. *Hear Res*. 2011;271:74–87.
41. Warren JD, Uppenkamp S, Patterson RD, et al. Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci*. 2003;100:10038–10042.
42. Zwicker E, Fastl H. Pitch and pitch strength. *In Psychoacoustics: Facts and Models*. New York: Springer-Verlag; 1990:103–132.
43. ANSI. *ANSI Sl.1-1994, American National Standard Acoustical Terminology*. New York: American National Standard Institute; 1994:34.
44. Fastl H, Stoll G. Scaling of pitch strength. *Hear Res*. 1979;1:293–301.
45. Fastl H, Zwicker E. *Psychoacoustics: Facts and Models*. 3rd ed. New York: Springer; 2007.
46. ASHA. *Consensus Auditory-Perceptual Evaluation of Voice(CAPE-V)*. Rockville, MD: American Speech-Language and Hearing Association; 2002.
47. Shrivastav R, Eddins DA, Anand S. Pitch strength of normal and dysphonic voices. *J Acoust Soc Am*. 2012;131:2261–2269.
48. Meddis R, O'Mard L. A unitary model of pitch perception. *J Acoust Soc Am*. 1997;102:1811–1820.
49. Camacho A. On the use of auditory models' elements to enhance a sawtooth waveform inspired pitch estimator on telephone-quality signals. *Paper Presented at the Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*. 2012.
50. ANSI. *ANSI S3.21-2004, Methods for Manual Pure-Tone Threshold Audiometry*. New York: American National Standards Institute; 2004.
51. Milenkovic, P. (2001). TF32 [Computersoftware]. Madison, WI.
52. Moore BC, Glasberg BR, Baer T. A model for the prediction of thresholds, loudness, and partial loudness. *J Audio Eng Soc*. 1997;45:224–240.
53. Shofner WP, Selas G. Pitch strength and Stevens's power law. *Percept Psychophys*. 2002;64:437–450.
54. Yost WA. Pitch strength of iterated rippled noise. *J Acous Soc Am*. 1996;100:3329–3335.
55. Patterson RD, Handel S, Yost WA, et al. The relative strength of the tone and noise components in iterated rippled noise. *J Acoust Soc Am*. 1996;100:3286–3294.
56. Yost WA, Hill R. Models of the pitch and pitch strength of ripple noise. *J Acoust Soc Am*. 1979;66:400–410.
57. Boersma P, Weenink D. *Praat: Doing phonetics by computer, version 4.0. 26*. 2005. *Retrieved September, 24*, 2005.
58. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420.
59. Hartmann WM. *Signals, Sound, and Sensation*. Springer Science and Business Media; 2004.
60. Camacho A, Harris JG. A sawtooth waveform inspired pitch estimator for speech and music. *J Acoust Soc Am*. 2008;124:1638–1652.
61. Brehm SB, Weinrich B, Zieser M, et al. Aerodynamic and acoustic assessment in children following airway reconstruction: an assessment of feasibility. *Int J Pediatr Otorhinolaryngol*. 2009;73:1019–1023.
62. de Alarcón A. Voice outcomes after pediatric airway reconstruction. *Laryngoscope*. 2012;122:S84–S86.
63. Kelchner LN, Weinrich B, Brehm SB, et al. Characterization of supraglottic phonation in children after airway reconstruction. *Ann Otol Rhinol Laryngol*. 2010;119:383–390.
64. Eddins DA, Skowronski MD, Anand S, et al. Acoustic predictors and bio-inspired modeling of the perceived vocal breathiness of sustained phonations and continuous speech. *Paper Presented at the 46th Annual Symposium of the Voice Foundation: Care of the Professional Voice*. Philadelphia, PA; 2017.
65. van As-Brooks CJ, Koopmans-van Beinum FJ, Pols LC, et al. Acoustic signal typing for evaluation of voice quality in tracheoesophageal speech. *J Voice*. 2006;20:355–368.