

Towards a Self-Rating Tool of the Inability to Produce Soft Voice Based on Nonlinear Events: A Preliminary Study

Peter S. Popolo¹⁾, Ingo R. Titze^{2,3)}, Eric J. Hunter³⁾

¹⁾ National Center for Voice and Speech, Denver, CO, USA

²⁾ Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, IA, USA

³⁾ National Center for Voice and Speech The University of Utah, Salt Lake City, UT, USA. ingo.titze@utah.edu

Summary

The purpose of this preliminary study was to investigate the feasibility of a tool to compare a severity index of nonlinear events and vocal self-rating over a long period of time. One hundred and ninety-seven phonations were analyzed to quantify the severity of instabilities in the voice attributed to nonlinear dynamic phenomena, including voice breaks, subharmonics, and frequency jumps. Instabilities were first counted; then a severity index was calculated for the instabilities in each phonation. The two quantities were compared to the subject's autoperceptual rating. Generally speaking, the measures derived from nonlinear dynamic analysis of the high-pitched, soft phonations followed the subject's own rating of inability to produce soft voice. These preliminary single subject results provide a foundation for future multi-subject studies to formulate acoustic and autoperceptual measures for the fatiguing effects of prolonged speaking in vocally demanding professions. Given the number of observations, the results are useful in showing general relationships. While future work should add additional subjects, a study providing preliminary evidence is useful before attempting to undertake a multi-subject study with complex analysis (i.e., individually selecting the nonlinear events) and with a long observation duration (days, weeks, and months) of subject.

PACS no. 43.58.-e, 43.70.-h

1. Introduction

Inability to produce soft voice (IPSV) was introduced by Bastian *et al.* [1] as a clinical tool to detect vocal fold swelling. The tasks consisted of a patient producing staccato, legato, and trillo-like phrases phonated very softly and at high pitches. IPSV is now being developed as an autoperceptual measure of one component of vocal fatigue [2, 3, 4].

Vocal fatigue has at least two components: (1) muscle fatigue of some or all of the laryngeal and articulatory muscles and (2) material fatigue due to excessive vibration of vocal fold tissues [5]. It is hypothesized that IPSV is useful as a predictor of the second component of fatigue, material fatigue of non-muscular tissue. Lamina propria tissues are targeted because vibration is confined to the lamina propria in soft voice. Also, most benign injuries of vocal fold tissues occur in the lamina propria, where vibrational and collision stresses are on the order of 1–5 kPa [6, 7]. Material fatigue associated with these stresses may bring about fluid imbalances [8] and general viscoelastic changes in the tissue [9]. In turn, these viscoelastic

changes may affect the modes of vibration of the vocal folds and the conditions of self-sustained oscillation [10]. The acoustic manifestations of the vibrational changes can be observed on a narrow-band spectrogram and quantified with nonlinear dynamic (bifurcation) analysis.

Our study is based on the hypothesis that the IPSV autoperceptual rating relates to bifurcation events. The IPSV rating itself is presently undergoing analysis and interpretation [11]. For the current focus, when attempting to determine whether an IPSV rating correlates with bifurcation events (sudden pitch jumps, subharmonics, short aphonic segments), the reliability clearly needs to be tested. It is possible that a single IPSV rating combines too many co-occurring phenomena into one number. Nonlinear dynamic phenomena in vocal fold vibration are perhaps too difficult to quantify on a one-dimensional perceptual scale. The current study may provide at least a partial answer to the reliability issue.

In terms of rating voice quality based on perceptual features, the IPSV rating system has only one or two commonalities with the GRBAS system [12], which uses five parameters: [1] Grade of hoarseness (G), [2] Roughness (R), [3] Breathiness (B), [4] Asthenia (A), and [5] Strained quality (S). The parameters are rated on a four-point scale from 0 to 3, where 0 corresponds to normal, 1

to slight, 2 to moderate, and 3 to severe. The ratings are performed by trained listeners, usually speech language pathologists. The features common to both the IPSV system and the GRBAS systems are roughness, and grade of hoarseness. Subharmonics are perceived as roughness and aperiodicity (such as chaotic vocal fold vibration) is perceived as hoarseness, the grade of which is simply an estimation of magnitude.

Omori *et al.* [13] demonstrated in a study of the acoustic characteristics of rough voice that, along with the traditional measures of jitter and shimmer, roughness is characterized by the presence of subharmonics in the power spectrum of the voice waveform. The subharmonics are produced by atypical mode synchronizations of the vocal folds, which can cause a fundamental periodicity of the waveform to stretch across two or three normal cycles. Subharmonics analysis, using the fast Fourier transform to observe the specific frequency and relative power of the subharmonics, provides an objective evaluation of rough voice. Bergan and Titze [14] conducted a study of perception of roughness in synthesized voice signals, in which amplitude modulation (AM) and frequency modulation (FM) were used to control subharmonic energy. They showed that AM tones with 20% modulation and FM tones with 10% modulation received roughness ratings of 5 or greater (out of 10) from trained subjects listening to the synthesized stimuli.

Breathy voice corresponds to turbulent noise in the glottis. One measure of the noise in the speech signal is the harmonic-to-noise ratio, or HNR [15], which is obtained from analysis of the power spectrum of the waveform. Turbulent airflow noise segments can be visually identified in the narrowband spectrogram as a “fill-in” between the harmonic structure at frequencies well above the fundamental. Chaotic segments may also occur, identified by the lack of periodicity at all frequencies.

The complexity of a self-oscillatory system, nonlinear by nature, can be measured by its internal (embedded) dimensions. A number of recent investigations have gone beyond the traditional perturbation analysis (jitter, shimmer, harmonics to noise ratio, etc.) to quantify the dimensional complexity. Butte *et al.* [16] described and utilized the median correlation dimension D_2 to describe the complexity between different singing styles. Lee *et al.* [17] applied a similar analysis to compare the complexity of acoustic signals of Parkinsonian patients to that of normal groups, and Meredith *et al.* [18] described pediatric dysphonia with the median correlation dimension. In all cases, the metric D_2 showed promise for quantification of the likelihood of occurrence of sudden changes in the vibration patterns of the vocal folds (bifurcations). Yu *et al.* [19] added the Lyapunov exponent to a set of perturbation measures and Nicollas *et al.* [20] experimented with a fractal dimension, both showing promise for differentiating types of voices (dysphonias and age and gender differences among children).

The nonlinear acoustic analysis methods carried out in this study are based on semi-quantitative and visual spec-

trographic methods used by various investigators. The list is as follows: 1) Robb and Saxman [21] studied young children’s non-cry vocalizations; 2) Mende *et al.* [22] studied newborn infant cries; 3) Herzel *et al.* [23] studied vocal disorders, including nodules, polyps, cysts, Reinke’s edema, vocal fold paralysis, and hypo- and hyperfunctional dysphonia; 4) Tokuda *et al.* [24] studied animal vocalizations such as macaque screams and dog-barks; 5) Neubauer *et al.* [25] studied complex vocal improvisations used by contemporary singers; and 6) Titze *et al.* [26] studied male and female amateur singers and non-singers performing vocal exercises involving the crossover of fundamental frequency and first-formant frequency. In all studies, a narrow band spectrogram was the first-level analysis, on which bifurcations were identified by visual inspection. Quantification was then applied semi-automatically. In this context, it should be stated that nonlinear effects are common in acceptable voice production (e.g. [25], pathological production [23], and in animal vocalization [24]).

The purpose of this preliminary study was to determine whether the autoperceptual measure “inability to produce soft voice” (IPSV) correlate with changes in a physical quantification of bifurcations in vocal fold vibration. In other words, can a subject recognize instabilities in his/her own voice and assign a severity number to these changes? The authors chose to conduct a single subject study to investigate the connection between subject ratings, nonlinear effects, and potential vocal fatigue. Such a study is needed before addressing a multi-subject study with complex analysis (i.e., individually selecting the nonlinear events) and long duration times.

2. Methods

2.1. Subject, Phonations and Ratings

The subject was a 38-year-old male who taught a course at a local university (Tuesday and Thursday afternoons, 2.5 hours of lecture). Reported data represents a three-week subset of six weeks of observation. The subject performed daily rating and measurement tasks (described below) in a sound isolation booth (usually one in the morning, one at midday, and one in the evening). Acoustic signals from a head-mounted microphone were recorded directly to digital mass storage. In this manner, seventeen samples of the tasks were collected over a three-week period, several of the samples being on the days the subject was lecturing.

A total of seven soft voice tasks were used in this study, each of which received a subject rating (SR). The first four were adapted from Bastian *et al.* [1] and were previously used in dosimetry studies (e.g., [2, 11, 4]):

1. *comf*: sustaining the vowel /i/ for five seconds as softly as possible on a comfortable pitch,
2. *glid*: gliding on the vowel /i/ from low to high pitch as softly as possible,
3. *stac*: staccato vowel repetitions /i-i-i-i-i/, and
4. *Bday*: a few bars of “Happy Birthday”, extremely soft and high-pitched.

Three additional soft voice tasks were added from the Titze *et al.* study [26], designed to elicit nonlinear source-filter interaction. These were:

1. *grev*: pitch glide on /i/ from high to low, and a reversal,
2. *i æ i*: vowel glide /i-æ-i/ at high pitch,
3. *u a u*: vowel glide /u-a-u/ at high pitch.

The subject rated each task individually on a scale of 1 to 10, with 1 representing the best soft-voice production possible and 10 representing complete inability to produce soft voice. The features he was instructed to consider in his ratings were the presence of hoarse segments (roughness or breathiness), unevenness of repeated phonations, aphonic segments, voice breaks, delayed voice onsets, and reduced range of pitch. While the subject was an author of this paper, he did not perform any of the acoustic analysis (described below). Further, the subject was not part of Titze *et al.* study [26] which presented acoustic nonlinear events of the vocal tract system from the additional three tasks mentioned above. This second set of tasks was chosen because nonlinear source-filter interactions should be more independent of subject control and training.

2.2. Acoustic Measures

As an acoustic correlate to the perceptual rating, the number of instabilities in the recordings of each phonation was first obtained from a narrow-band spectrogram and assigned the number NI. In addition, every instability was classified as one of five types: silence gap, subharmonic, chaotic, modulation with side band frequencies, or F_0 jump. This qualitative classification then received a quantitative weighting coefficient. The weighting coefficients were as follows:

1. α_{11} : energy loss in silence gap vs. harmonic energy on either side (in %),
 α_{12} : duration of phonation gap (normalized by the duration of all voiced segments of the entire phonation),
2. α_{21} : energy in subharmonic vs. harmonic energy in the same segment (in %),
 α_{22} duration of subharmonic segment (normalized as in α_{12})
3. α_{31} : energy of chaos (noise) vs. harmonic energy in the same segment (in %)
 α_{32} : duration of chaotic segment (normalized as in α_{12})
4. α_{41} : energy of sideband frequencies vs. harmonic energy in the same segment (in %)
 α_{42} : duration of sideband segment (normalized as in α_{12})
5. α_{51} : ΔF of frequency jumps (normalized by the average F_0 of the entire phonation).

As seen, each of the first four instability types has two components, a magnitude component and a duration component. Frequency jumps were quantified by only a single component because there was no quasi-steadiness to assign a specific duration.

An overall severity index (SI) was then calculated with the weights as

$$SI = \sum_{i=1}^N (\alpha_{i1} + \alpha_{i2}), \quad (1)$$

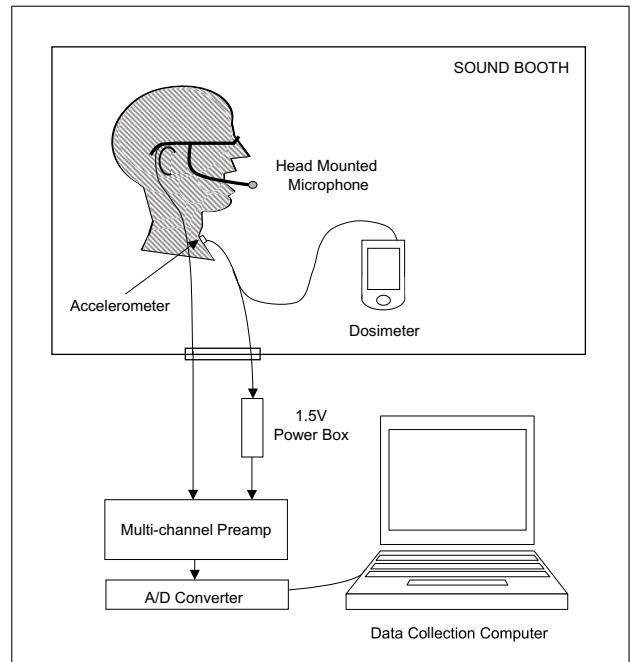


Figure 1. Recording setup.

where the first subscript of the weight is chosen according to the above classification for each instability. In rare cases, there could be two overlapping instabilities (e.g., a subharmonic and a frequency jump), in which case there would be two α values for the same instability, and they would be additive.

It should be noted that it was assumed that the main subharmonic source was the interaction between the vocal tract and the vocal folds. Although the lungs have a resonance and could have some effect on the vocal folds, for the cases detailed here the vocal tract resonances were targeted for interaction.

2.3. Recordings

The recordings were made in a single-wall sound isolation booth (Industrial Acoustics Company, Bronx, NY), 2.2 m wide by 2.3 m high by 2.3 m deep. The subject wore a head-mounted microphone (Countryman Associates omnidirectional B3 Lavalier) mounted on a wire boom attached to a plastic frame, worn like a pair of eyeglasses. The microphone element was about 5 cm from the mouth and slightly to the side, out of the airstream. A schematic diagram of the setup is shown in Figure 1.

2.4. Data Analysis

Analysis of the recordings was conducted after all data collection was complete in an attempt to insulate the subject from the results.

The head-mounted microphone signals obtained were edited using Cool Edit 2000 (Syntrillium Software Corp., Phoenix, AZ) to create separate *wav* files of each phonation. The resulting number of *wav* files was 197.

All phonations were analyzed by computer-assisted inspection of the narrow-band spectrograms to identify instabilities. While it is assumed that these instabilities refer to specific nonlinear events, it is possible that they do not absolutely correlate. The instabilities were identified as:

1. damped oscillation or visually observed as a silent gap in the spectrogram where phonation is expected (note: damped oscillation in phase space corresponds to a point attractor),
2. subharmonic phonation, as evidenced by equally spaced lines between the harmonic lines in a narrow-band spectrum
3. chaos-like, which may appear as random (non-periodic), segments in a narrow-band spectrogram (note: non-periodic is not synonymous with random),
4. modulation, as evidenced by sidebands around the harmonics in the narrow-band spectrograms
5. frequency jumps, as evidenced by sudden discontinuities in the harmonics of the narrow-band spectrogram (note: frequency jump is associated in phase space to the jump from a limit cycle to another coexisting limit cycle).

A set of Matlab scripts was written to perform the acoustic analyses of the phonations. The first script automatically computed and displayed the narrowband spectrograms of the *wav* files of each individual phonation, allowing the user to visually identify and mark the boundaries of each voice instability on the spectrogram. The narrowband spectrograms were generated to have a frequency resolution of about 20 Hz and a time resolution of about 20 ms. For the sampling frequency of 44.1 kHz, at which the *wav* files were recorded, this corresponded to the FFT length and window size of 2048 samples, with an overlap of 1024 samples. The program to generate the spectrograms was a batch-file script, so that all the *wav* files stored in a specified directory were processed automatically. For each file, the time waveform and the narrowband spectrogram were automatically plotted, and the user was prompted to enter the number of instabilities to be marked. The user then marked the start and stop times of each instability, and entered a code number for the description of each (e.g., 1=gap, 2=subharmonic, 3=chaos, 4=sideband, 5=frequency jump). The user also marked the start and stop time of the entire phonation, trimming off any silence portions at the beginning or end.

The program saved the *jpeg* images of all the spectrograms, with the instabilities marked, into a specified folder. Then the program automatically created an Excel spreadsheet tabulating the start times, stop times, and durations of all the instabilities, plus the start time, stop time, duration, and voicing times of each phonation (the voicing time of a phonation was calculated as the total duration minus the sum of all the instability durations). Figure 2 shows a plot of the waveform and the spectrogram of the pitch glide and reversal on the vowel /i/ with five instabilities and the start and stop boundaries marked. Three instabilities are very apparent: voice break or gap, frequency jump, and subharmonics.

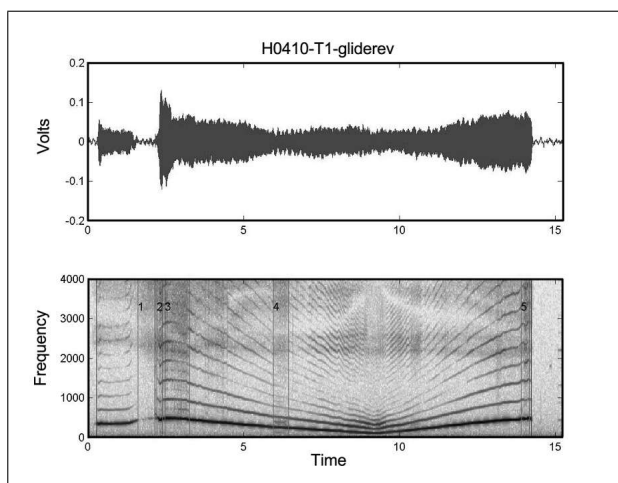


Figure 2. Plot of the waveform and the spectrogram of the pitch glide and reversal on the vowel /i/ with five instabilities and the start and stop boundaries marked. Three instabilities are labeled: voice break or gap (1), frequency jump (2, 5), subharmonics (3, 4).

Three of the five possible types of instabilities were found to occur throughout the set of phonations, these being gaps, subharmonic sections, and frequency jumps. No instances of chaos or sideband instabilities (modulations) were found. The duration component of the severity index for both the gap and the subharmonic segments was calculated as

$$\alpha_{12} = \frac{\text{Duration of segment}}{\text{Duration of all voiced segments in phonation}} \quad (2)$$

To calculate the energy component of the severity index for the gap sections, a batch program was written that automatically selected all the marked gap sections in each phonation and calculated the average RMS energy in the gap as well as the average RMS energy in sections of similar duration both prior to and after the gap. The worst-case energy loss in the gap section was then calculated as

$$\alpha_{i1} = 100 \max (E_1, E_2), \quad (3)$$

where

$$E_1 = \text{anti log} \left[\frac{(\text{Ave pre-gap RMS energy in dB} - \text{Ave gap RMS energy in dB})}{10} \right], \quad (4)$$

and

$$E_2 = \text{anti log} \left[\frac{(\text{Ave post-gap RMS energy in dB} - \text{Ave gap RMS energy in dB})}{10} \right], \quad (5)$$

The batch program generated a new Excel spreadsheet with the gap loss values in the correct row for each gap section of each phonation.

To calculate the energy component of the severity index for the subharmonic sections, a batch program was written that read the spreadsheet generated by the previous program and automatically selected all the user-marked subharmonic sections in each phonation. The average spectrum was plotted for each subharmonic section, calculated

by averaging the power spectra taken at 20 ms intervals within the section, and the user was prompted to mark all the harmonic and subharmonic peaks that could be seen. The program then calculated the value as the subharmonic-to-harmonic ratio for each section [27],

$$\alpha_{i1} = \frac{SS}{SH}, \quad (6)$$

where SH is the sum of harmonic amplitudes, defined as

$$SH = \sum_{n=1}^N A(nf_0), \quad (7)$$

and SS is the sum of subharmonic amplitudes, defined as

$$SS = \sum_{n=1}^N A((n - 1/2)f_0). \quad (8)$$

In the summations, N is the number of harmonics to be considered. Equation (8) is written specifically for a period-2 subharmonic, but was easily generalized to any other subharmonic (e.g., period-3 or period-4) by changing the $1/2$ -fraction to the correct period fraction. The duration component α_{i2} for the subharmonic instabilities was computed in the same way as for the gap instability (equation 2).

The third type of instability that occurred, a frequency jump, was analyzed using the pitch extraction tool in the sound analysis software PRAAT [28]. The segments containing the frequency jumps in each phonation, identified by the first Matlab script, were automatically saved into separate *wav* files in a specified directory. These were opened in PRAAT to measure the maximum and minimum frequency in the jump segment, resulting in the value

$$\alpha_{i1} = \frac{f_{\max} - f_{\min}}{f_{\max}} \cdot 100\%. \quad (9)$$

With all the instability α values quantified as above, equation (1) was now normalized and scaled to be in the range of 1–10. Normalization to all alpha values was necessary because each of the α values by itself is defined in percent (values ranging from 0 to 100). Without normalization, adding several values together could theoretically result in values above 100% if the durations were extremely long and the magnitudes were high. Scaling the result as a numerical integer from 1–10 was only for convenience in comparing the severity index SI in Equation 1 to an autopercptive subject rating SR .

3. Results

In the 197 phonations subjected to the acoustic analysis described above, 473 instabilities were found. The number of instabilities in a single phonation ranged from 0 to 13, and the average number of instabilities over all phonations was 1.99. Figure 3 shows the distribution of instability types over all phonations, and Figure 4 shows the distribution of instability types sorted by test utterance (abbreviations given earlier).

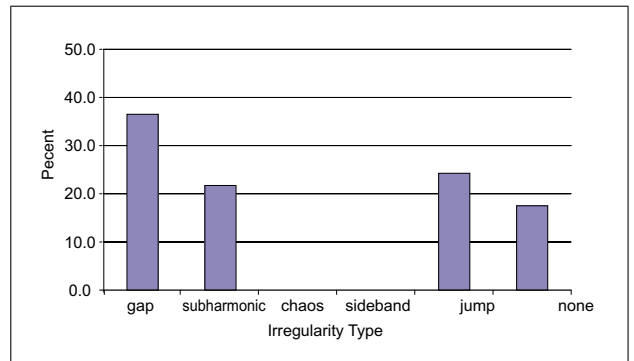


Figure 3. Distribution of instability types over all phonations.

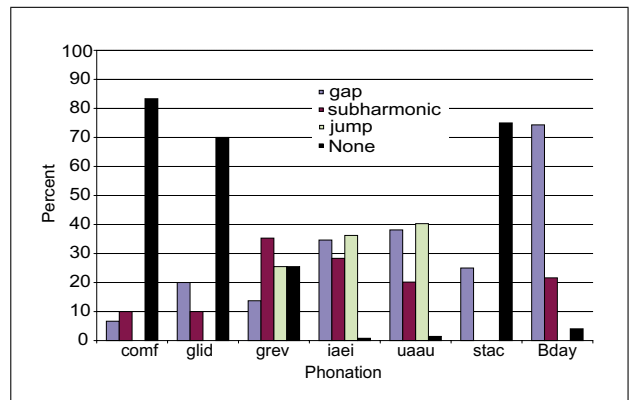


Figure 4. Distribution of instability types sorted by phonation.

The SI values were calculated as in Equation (1) for each of the 197 phonations individually, and were found to lie in the range of 1 to 5 (average value of 1.64 over all phonations). The corresponding subject ratings (SR) were found to lie in the range of 1 to 8 (average 3.11). The severity index SI , the number of instabilities NI , and the subject ratings SR , were then compared using regression techniques.

Two types of statistical analysis were done. First, two-way ANOVAs were performed to examine the trends in the data sets, and to explain the differences and similarities in the variation in the data sets. The NI , SI , and SR values of the 197 individual phonations were sorted into different groups (utterance type, time of day, day of week) to determine if any trends in the data could be found. Second, the average SI , NI , and SR values over a given phonation task were treated as time series. The cross-correlation of these time series was computed.

Figure 5 shows the average SI , NI , and SR values sorted by test utterance. Figure 6 shows the average values of SI , NI , and SR sorted by day of the week. Figure 7 shows the same averages sorted by time of day, divided into morning (7 a.m. to 11 a.m.), midday (11 a.m. to 1 p.m.), afternoon (1 p.m. to 5 p.m.), and evening (5 p.m. to 9 p.m.).

The two-way analysis of variance (ANOVA) on each data set was useful for studying the effects of two different factors in the observed data. A table of the means of the factored data sets was generated, with separate rows for each category of the first factor and separate columns

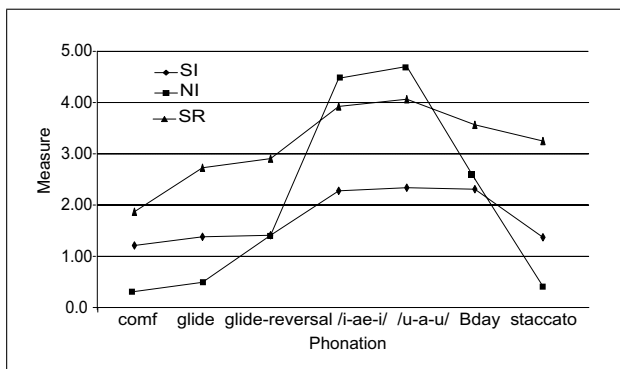


Figure 5. Means of measurement type vs. phonation type..

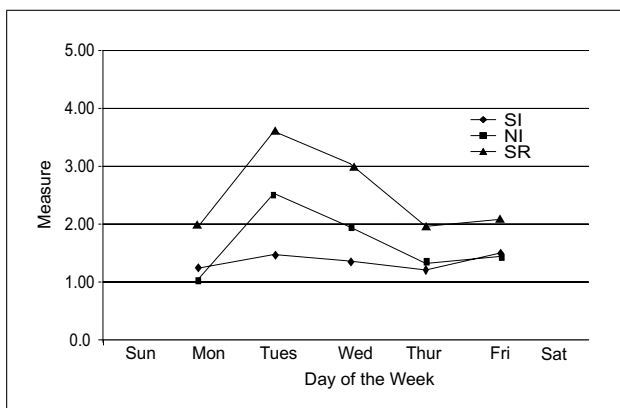


Figure 6. Means of measurement type vs. day if the week.

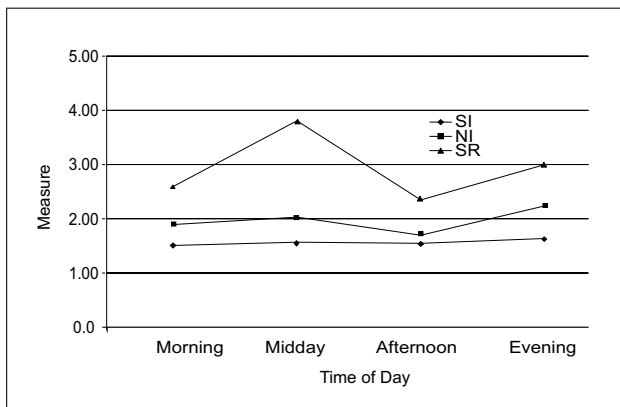


Figure 7. Means of measurement type vs. time of day.

for each category of the second factor. If there was a statistically significant difference between the categories of a factor (as determined by the analysis of variance), then that factor would have a significant main effect on the variations of the observed data, and multiple comparison tests would determine which categories are different. The two-way ANOVA procedure also identified whether there was a statistically significant interaction effect between the two factors on the variations in the observed data.

For the data set shown in Figure 5, the two factors were phonation type and measurement type (SI, NI, and SR). The two-way ANOVA showed that both phonation type and measurement type are significant main effects, with p

Table I. Results of Scheffe's Multiple Comparison Test for days of the week, showing which days are significantly different from which other days. Over-bars signify days that are grouped together.

Day of week	Mon	Thu	Fri	Wed	Tue
MeanRatings	1.687	1.769	1.948	2.374	2.780

Table II. Results of Scheffe's Multiple Comparison Test for measurement types sorted by days of the week.

Measurement type	SI	NI	SR
Mean values	1.635	1.928	2.771

Table III. Results of Scheffe's Multiple Comparison Test for time of day.

Time of day	Afternoon	Morning	Evening	Midday
Mean Rating	1.994	2.090	2.385	2.564

values of less than 10^{-6} , and the interaction effect of these two factors is significant with $p < 10^{-6}$.

For the data set shown in Figure 6, the two factors were day of the week and measurement type, both of which were significant main effects with $p < 10^{-6}$. The interaction of the two factors was not significant at $\alpha = 0.05$, so the next step was to determine which days and measurement types were different from others. Scheffe's Multiple-Comparison test shows that there was no significant difference between the mean values of the data obtained on Monday, Thursday, and Friday, but there is a significant difference between each of these days and Tuesday (the means are significantly higher on Tuesdays, which was one of the teaching days). Wednesday was not significantly different from either Tuesday or the group of Monday, Thursday, and Friday. Scheffe's test of the differences between measurement types showed that SI and NI were both significantly different from the subject ratings SR, but they were not significantly different from each other. Tables I and II show the schematic representation of the results of the multiple comparison test.

The two factors for the data set shown in Figure 7 were time of day and measurement type. In this case, time of day was a significant main effect with $p = 0.03$, measurement type was a significant main effect with $p < 10^{-6}$, and the interaction of the two factors was not significant at $\alpha = 0.05$. Scheffe's test showed that there was a significant difference between midday and afternoon (midday had the highest mean and afternoon had the lowest), but the means for the morning and evening times were not significantly different from each other, or any other times. The differences between measurement types were the same as in the previous two cases. Tables III and IV show the schematic representation of the results of the multiple comparison test of time of day and measurement type.

Table IV. Results of Scheffe's Multiple Comparison Test for measurement types sorted by time of day.

Measurement type	SI	NI	SR
Mean values	1.647	2.052	3.038

Table V. The correlation of the time ratings (SR) and all acoustic measurements (NI and SI).

	SI to NI	SR to NI	SR to SI
Days of Week (Fig. 6)	0.62	0.97	0.53
Time of Days (Fig. 7)	0.77	0.59	0.35

By treating the average subject rating SR and the corresponding physical measures NI and SI as time series, signal correlations were calculated (Table V). For Figure 6, a time-series analysis over days of the week yielded correlations of SI with NI of 0.62, SR with NI of 0.97 and SR with SI of 0.53 (Table V). The time-series analysis over time of day (Figure 7) yielded the correlations of SI with NI of 0.77, SR with NI of 0.59, and SR with SI of 0.39. These results show that, although not a strong correlation, there is some positive correlation between the physical measures and the subjective ratings over time.

4. Discussion

For the data set in which SI, NI, and SR measures were sorted by utterance type, the two-way ANOVA showed there was a significant interaction between the measurement type and the utterance type, suggesting that certain measures are more sensitive to certain utterances (e.g., F_0 jumps to vowel change). Figure 5 showed that the subject had more difficulty producing the vowel glides /i-ae-i/ and /u-a-u/ than any constant vowel phonations. In particular, the number of instabilities NI was high, with perceptual rating SR and severity SI being smaller than NI. The interpretation is that small pitch jumps are discounted perceptually. The α weighting factor also discounts them so that SI and SR are in better agreement with each other than either one is with NI.

For the time series, measurement type versus day of the week and time of day, the Scheffe multiple comparison tests showed that there was no significant difference between the SI and NI measures, although the trend was the same as for utterance type (NI larger than SI). The same test shows that the SI and NI measures were significantly different from the Subject Rating (SR). For the days of the week all three ratings were significantly higher for Tuesdays and Wednesdays than the other days, likely the result of lecturing on Tuesday with some fatigue carryover to Wednesday. However, Thursday did not show an increase in the ratings even though it was a teaching day. An explanation for this could be that moderate vocal activity on Wednesday (with ratings still being high due to time-lag in recovery) had the benefit of a vocal recovery for the remainder of the week. The time-of-day results show that

there is a significant difference between midday and afternoon in all three ratings, with the afternoon ratings being lower than the midday ratings. The pattern is quite similar to the day-of-the-week pattern. A peak around midday suggests a delayed response to morning conversation, with difficulty in phonation increasing until the voice is adequately warmed up. A lunch-time break may then provide a recovery that is ideal for afternoon teaching. A delayed response is then again felt in the evening.

The above analysis would perhaps be most useful to someone with a recurring time schedule, or cyclic voice use (e.g., a telephone worker or teacher). High vocal dose would occur at specific times of the day and specific days of the week. Our subject did not have such cyclic voice use, outside the two teaching responsibilities on Tuesday and Thursday afternoon.

5. Conclusion

This study was a nonlinear dynamic analysis of a series of high-pitched soft phonations designed to serve as a corroboration of the autoperceptive rating *inability to produce soft voice*, a clinical tool used in assessment of vocal function. The nonlinear dynamic analysis was carried out by quantifying instabilities (bifurcations) visible in the narrowband spectrograms of the phonations, using methods previously applied to the characterization of nonlinear phenomena in voice and speech. For each phonation the analysis yielded two numerical ratings: (1) a simple count of the number of instabilities (bifurcations) and (2) a weighted severity index of the same instabilities. These ratings were compared with the subject's auto-perceptual ratings for the same set of phonations. All ratings showed similar trends when sorted by day of the week, time of day, and to a lesser extent, type of utterance (e.g., pitch glides, staccato, and vowel glides). Generally speaking, the measures derived from nonlinear dynamic analysis appear to be usable to predict the subject's own rating of inability to produce soft voice, but strong correlations (90% or more) were not found and cannot be expected with a single numerical rating; nevertheless, all rating methods are in agreement when it comes to describing the short-latency and long-latency temporal effects of prolonged speaking.

It is not clear whether the severity index provides a better accounting of vocal instability than the simple instability count. It does devalue small instabilities (i.e., small F_0 jumps or very short subharmonic segments) the same way that the autoperceptive rating does, but is it possible that small (imperceptible) phenomena capture something below the tip of the iceberg? It is too soon to tell.

There were several shortcomings to the current study. First, this was a single-subject report with a large number of observations. It is possible that the single subject design is not representative of the population as a whole. However, given the number of observations, the results are still useful to show general relationships. While future work should add additional subjects, a study providing

preliminary evidence was useful before attempting to undertake a multi-subject study with complex analysis (i.e., individually selecting the nonlinear events) and long observation duration (days, weeks and months). In addition, this particular subject was not naive to acoustic phenomena. Hence, the source-filter interaction effects may have been within the subject's volitional control, but it is unlikely due to the cognitive load in teaching.

Second, it is possible that overlapping or adjacent instabilities would be perceived as just one instability. Nevertheless, because these events would be rare, the effect on the overall results should be minimal. Also, some instabilities may have a cause outside of source-filter interaction. For example, some modulation instabilities may be introduced by separate motor events [23]. In addition, other modulations may relate to mode-mode interaction [29] as there is always some asymmetry between the two vocal folds [30]. Nevertheless, as there were no chaos or modulation instabilities identified in this preliminary study, it is assumed that these two specific instabilities would be very few and only a minor addition to a self-perception rating system.

Lastly, the subject had no reported pathology or vocal weakness, which limits the ability to translate the results directly to the population with vocal disorders. Nevertheless, it was important to investigate the variations and effects in a normal population to eliminate normal subject. Future work should investigate changes related to voice disorders and fatigue on sizeable subject groups, as well as pre and post intervention assessment.

Acknowledgement

Funding for this work was provided by the National Institute on Deafness and Other Communication Disorders, grant number 1R01 DC04224.

References

[1] R. W. Bastian, A. Keidar, K. Verdolini-Marston: Simple vocal tasks for detecting vocal fold swelling. *J. Voice* **4** (1990) 172–183.

[2] T. Carroll, J. Nix, E. J. Hunter, K. Emerich, I. R. Titze, M. Abaza: Objective measurement of vocal fatigue in classical singers: a vocal dosimetry pilot study. *Otolaryngol. Head and Neck Surg.* **135** (2006) 595–602.

[3] A. Halpern, J. Spielman, E. J. Hunter, I. R. Titze: The inability to produce soft voice (IPSV): a tool to detect vocal change in school teachers. *Logoped. Phon. Voc.* **34** (2009) 117–127.

[4] E. J. Hunter, I. R. Titze: Quantifying vocal fatigue recovery: dynamic vocal recovery trajectories after a vocal loading exercise. *Ann. Otol. Rhinol. Laryngol.* **118** (2009) 449–460.

[5] D. J. McCabe, I. R. Titze: Chant therapy for treating vocal fatigue among public school teachers: a preliminary study. *Am. J. Speech-Lang. Path.* **11** (2002) 356–639.

[6] I. R. Titze: Mechanical stress in phonation. *J. Voice* **8** (1994) 99–105.

[7] J. J. Jiang, I. R. Titze: Measurement of vocal fold intraglottal pressure and impact stress. *J. Voice* **8** (1994) 132–144.

[8] C. Tao, J. J. Jiang, Y. Zhang: A fluid-saturated poroelastic model of the vocal folds with hydrated tissue. *J. Biomech.* **42** (2009) 774–780.

[9] S. Gray, I. R. Titze: Histologic investigation of hyperphoned canine vocal cords. *Ann. Otol. Rhinol. Laryngol.* **97** (1988) 381–388.

[10] I. R. Titze: The myo-elastic aerodynamic theory of phonation. National Center for Voice and Speech, Denver, CO, 2006.

[11] E. J. Hunter: NCVS memo No. 11. General statistics of the NCVS Self-Administered Vocal Rating (SVRa). <http://www.ncvs.org/ncvs/library/tech>, 2008.

[12] M. Hirano: Clinical examination of voice. Springer, New York, 1981.

[13] K. Otori, R. Kojima, D. H. Slavik, S. M. Blaugrund: Acoustic characterizations of rough voice: subharmonics. *J. Voice* **11** (1997) 40–47.

[14] C. C. Bergan, I. R. Titze: Perception of pitch in voice signals with subharmonics. *J. Voice* **15** (2001) 165–175.

[15] E. Yumoto, W. J. Gould, T. Baer: Harmonic-to-noise ratio as an index of the degree of hoarseness. *J. Acoust. Soc. Am.* **71** (1982) 1544–1550.

[16] C. J. Butte, Y. Zhang, H. Song, J. J. Jiang: Perturbation and nonlinear dynamic analysis of different singing styles. *J. Voice* (2011) (in press).

[17] V. S. Lee, X. P. Zhou, I. Rahn, D. A., E. Q. Wang, J. J. Jiang: Perturbation and nonlinear dynamic analysis of acoustic phonatory signal in Parkinsonian patients receiving deep brain stimulation. *J. Comm. Dis.* **41** (2008) 485–500.

[18] M. L. Meredith, S. M. Theis, J. S. McMurray, Y. Zhang, J. J. Jiang: Describing pediatric dysphonia with nonlinear dynamic parameters. *Int. J. Ped. Otorhinolaryngol.* **72** (2008) 1829–1836.

[19] P. Yu, R. Garrel, R. Nicollas, M. Ouaknine, A. Giovanni: Objective voice analysis in dysphonic patients: new data including nonlinear measurements. *Folia Phon. et Logoped.* **59** (2007) 20–30.

[20] R. Nicollas, R. Garrel, M. Ouaknine, A. Giovanni, B. Nazarian, J. M. Triglia: Normal voice in children between 6 and 12 years of age: database and nonlinear analysis. *J. Voice* **22** (2008) 671–675.

[21] M. P. Robb, J. H. Saxman: Acoustic observations in young children's non-cry vocalizations. *J. Acoust. Soc. Am.* **83** (1988) 1876–1882.

[22] W. Mende, H. Herzel, K. Wermke: Bifurcations and chaos in newborn cries. *Phys. Lett. A.* **145** (1990) 418–424.

[23] H. Herzel, D. Berry, I. R. Titze, M. Saleh: Analysis of vocal disorders with methods from nonlinear dynamics. *J. Speech, Lang. Hear. Res.* **37** (1994) 1008–1019.

[24] I. Tokuda, T. Reide, J. Neubauer, M. J. Owren, H. Herzel: Nonlinear analysis of irregular animal vocalizations. *J. Acoust. Soc. Am.* **111** (2002) 2908–2919.

[25] J. Neubauer, M. Edgerton, H. Herzel: Nonlinear phenomena in contemporary vocal music. *J. Voice* **18** (2003) 1–11.

[26] I. R. Titze, T. Riede, P. S. Popolo: Nonlinear source-filter coupling in phonation: Vocal exercises. *J. Acoust. Soc. Am.* **123** (2008) 1902–1914.

[27] X. Sun: A pitch determination algorithm based on subharmonic-to-harmonic ratio. 6th International Conference of Spoken Language Processing, Beijing, China, 2000, 4, 676–679.

[28] P. Boersma, D. Weenick: Doing phonetics by computer. Retrieved from www.praat.org 30 April 2009, 2009.

- [29] R. Laje, D. Sciamarella, J. Zanella, G. B. Mindlin: Bilateral source acoustic interaction in a syrinx model of an oscine bird. Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys. **77** (2008) 011912.
- [30] P. A. Lindestad, S. Hertegard, G. Bjorck: Laryngeal adduction asymmetries in normal speaking subjects. Log. Phon. Vocol. **29** (2004) 128–134.
- [31] P. S. Popolo, J. G. Svec, I. R. Titze: Adaptation of a pocket PC for use as a wearable voice dosimeter. J. Speech Lang. Hear. Res. **48** (2005) 780–791.
- [32] I. R. Titze, J. G. Svec, P. S. Popolo: Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues. J. Speech Lang. Hear. Res. **46** (2003) 919–932.