Clinical Linguistics & Phonetics, April–May 2006; 20(2/3): 181–186

The clinical utility of the soft phonation index

NANCYE C. ROUSSEL, & MARY LOBDELL

University of Louisiana, Lafayette, LA, USA

(Received 15 July 2004; accepted 14 October 2004)

Abstract

This paper investigates the applicability of the soft phonation index (SPI) as an indicator of vocal fold adduction. The SPI is one of 33 acoustic parameters calculated and displayed from a single vocalization by the multi-dimensional voice program (MDVP). A review of pertinent literature revealed that while increased values of SPI are presumed to be associated with incomplete adduction of the vocal folds, results should be interpreted cautiously as the index has not been sufficiently studied. The purpose of this pilot study was to determine how consistently SPI tracks systematic changes in vocal fold adduction. Samples of normal, breathy and pressed phonation produced by female speakers were analysed. As predicted, a significant increase in SPI for breathy productions was demonstrated. However, increases in SPI were also noted for pressed phonation. Possible explanations and implications for clinical application are discussed.

Keywords: *Vocal fold adduction, breathiness, phonation*

Introduction

The diagnosis and subsequent treatment of underlying vocal pathologies and/or vocal dysfunction by otolaryngologists and speech-language pathologists is dependent on multiple sources of information about the integrity of the laryngeal mechanism. One aspect of laryngeal function that is of great interest to both groups of professionals is the extent of vocal fold closure or adduction during phonation. This information is often used for diagnostic purposes but is also helpful when teaching or evaluating the effectiveness of a particular treatment strategy. For example, one popular treatment strategy to counteract the effects of laryngeal hyperfunction is to encourage the use of a breathy voice (Boone & McFarlane, 1988). The rationale being that consistent use of a breathy voice will encourage incomplete adduction of the vocal folds and, in individuals with hyperfunctional voice patterns, will eventually result in a normalization of laryngeal closure for voicing. Conversely, for patients with hypofunctional voice patterns, techniques such as pushing or pulling against an object during phonation have been used to encourage more forceful closure of the glottis during voicing. Being able to determine if a particular strategy is having the desired effect would allow the clinician to change treatment focus if necessary. This information could also be used to track change over time or even to serve as

Correspondence: Nancye C. Roussel, University of Louisiana, PO Box 43170, Burke-Hawthorne Hall, Rm 215, Lafayette, LA70504, USA. E-mail: ncroussel@louisiana.edu

ISSN 0269-9206 print/ISSN 1464-5076 online © 2006 Taylor & Francis
DOI: 10.1080/02699200400026942

biofeedback for the patient when learning the technique. While direct visualization using videostroboscopy is the preferred method for judging adduction patterns, not all speech-language practices or ENT offices have access to the required instrumentation. In addition, the procedure requires specialized training and because of the invasive nature of the procedure, should be performed in settings with proper medical backup. Noninvasive methods of estimating vocal fold closure, which include electroglottography, airflow measures and acoustic analysis of the vocal signal would be more appropriate for many clinical settings. Of these, acoustic-based analysis has several advantages.

Modern acoustic analysis systems are relatively inexpensive, available for use on Windows and Mac-based computers and are quite user-friendly. They provide immediate quantitative information that can be used in both diagnosis and treatment of a variety of voice disorders. The multidimensional voice program (MDVP) is a clinical application for the Computerized Speech Lab (CSL) by Kay Elemetrics. This program automatically extracts and displays 33 separate voice parameters that can then be compared to threshold values and displayed graphically. One of the parameters is the soft phonation index (SPI). According to publications from the developers of the software, SPI is an evaluation of the weakness of high-frequency harmonic components that may be an indication of loosely adducted vocal folds during phonation (Deliyski, 1993). It reflects the average ratio of lower frequency (70–1600 Hz) harmonic energy to the higher frequency (1600–4500 Hz) harmonic energy for a sequence of 81.92 ms (4096 pts) windows of the voiced areas in the analysed signal with a half-window overlap. Though the spectral slope of a voiced signal may be influenced by several factors including the rigidity of the vocal tract walls, it is known that the vocal source also affects the resultant spectral shape. Weaker or less complete vocal fold closure, produces a steeper spectral slope (Koreman & Putzer, 1997). Therefore, increased values of SPI are thought to indicate incomplete or loose vocal fold adduction.

Previous research involving the SPI has focused on this index as one of many potential indicators of variations in vocal quality. In an investigation of aging female voices, Bruckl and Sendlmeier (2003) found that SPI was not correlated with chronological age but showed a small, but significant, correlation with perceived age of speakers based on sustained /a/ samples. This would indicate that listeners related a proportional increase of the modulated energy in the lower frequency range to the perception of increasing age. The relationship was not seen for connected speech leading these authors to exclude SPI from the list of primary acoustic measures of aging in the female voice. Munoz, Mendoza, Fresneda, Carballo, and Lopez (2003) studied a group of acoustic and perceptual parameters as potential indicators of normal and pathological voices. They performed cluster analyses using the perceptual ratings to group 69 male and female subjects into three groups, and then used discriminant analysis to isolate acoustic parameters with discriminative power. Results indicated that SPI was not a significant factor for female voices but was one of several significant factors for the male clusters. Mean values for SPI ranged from 11.52 for the normal group to 15.81 for the hoarse group and 13.89 for the rough group. SPI values were not reported in the study for the female speakers. In a study investigating the reliability of the MDVP program for the analysis of voices of dysarthric speakers, Kent, Vorperian and Duffy (1999) concluded that SPI was not a parameter that showed a high degree of abnormality in dysarthric speakers. In the only study to date to look at SPI as a possible indicator of vocal fold adduction, Koreman and Putzer (1997) attempted to determine which of group of voice parameters (based on both the acoustic signal and EGG signal) could be used to distinguish normal speakers from breathy speakers and speakers with organic pathology (unilateral vocal cord paralysis and cordectomy). They found that the parameters derived from the EGG were, in general, better than those from the microphone signal with the exception of SPI. The effect of SPI was not strong, however, and could only distinguish patients showing unilateral vocal fold paralysis with compensation from the normal and breathy groups for the vowel /u/. Of note was the fact that the normal group had an average SPI of 64, which is much higher than the threshold value of 14.12 used by the MDVP program, as well as the results from the study by Munoz et al. (2003). In addition, the patients with uncompensated unilateral vocal fold paralysis and the cordectomized patients, groups that presumably had less vocal fold adduction, actually showed intermediate SPI values, i.e., higher than the normal and breathy groups, but lower than those with unilateral vocal fold paralysis with compensation. Therefore, it is does not appear, based on these data, that SPI is a reflection of strength of vocal fold adduction, at least not exclusively.

The focus of the present research was to determine if SPI could be used reliably to track systematic changes in vocal fold adduction. Our hypothesis was that SPI would increase when normal subjects produced breathy voice quality under the assumption that vocal fold adduction would be incomplete in this condition. Conversely, we should see a decrease in SPI for productions of "pressed" voice in which vocal fold adduction would be increased.

Method

Twenty-eight female subjects without a history of speech, hearing or neurologic disorders served as subjects for the study. Their ages ranged from 21–44 with a mean age of 25.65. A prescreening questionnaire was utilized to control for factors that might have an adverse effect on voice quality. Subjects who had symptoms for upper respiratory involvement or were premenstrual at the time of testing were rescheduled. A voice screening performed by the first author indicated that all subjects had perceptually normal voices on the day of the study. Three second voice samples of sustained /a/ were recorded directly into the CSL 4400 using a 50 kHz sampling rate and 16 bit quantization. Intensity for each production averaged 85 dB with a two-inch mouth to microphone distance. Subjects were asked to produce /a/ at a comfortable pitch, and this pitch was matched to a note on a digital keyboard. This reference note was used to control pitch across samples. Recordings were begun after initiation of voicing and ended before the subject terminated voicing. Each subject provided three repetitions under three conditions for a total of nine voice samples per subject. Normal voice was defined as phonation at a comfortable pitch with efforts to maintain pitch across conditions as described above. In the breathy condition, subjects were asked to initiate voicing with /h/ and maintain breathy quality throughout the phonation. Models and practice were given prior to recording. Subjects were instructed to use a relaxed throat and make their voice "fuzzy". For the pressed voice condition, voicing was initiated with a hard glottal attack and subjects either pressed down or lifted up on the edge of the table. Subjects were instructed to maintain tightness throughout the laryngeal area during voicing.

Two of the subjects could not maintain pitch across the nine repetitions and seven of the subjects could not produce the breathy voice condition. Data from these subjects were not included in the analysis reducing the number of data sets to nineteen. A one sample t-test of mean differences across repetitions within voicing conditions revealed no significant differences, therefore mean SPI for each subject in the three voicing conditions was computed and used for further analysis (Table I). To test our hypothesis (increased SPI

Table I. Mean SPI ratios for individual subjects across three voicing conditions. Group means for each condition are also shown.

Subject	Age	Breathy	Normal	Pressed
1	22	23.14	16.78	23.20
2	28	11.53	10.87	6.42
3	24	19.51	7.65	27.85
4	35	35.87	28.43	33.15
5	23	8.93	11.58	10.83
6	44	9.26	8.15	13.14
7	30	10.13	6.95	7.09
8	32	15.70	16.28	13.78
9	24	25.84	10.58	17.81
10	29	19.17	12.94	11.25
11	23	13.38	16.50	13.34
12	23	16.64	7.14	11.78
13	23	20.91	11.83	14.66
14	39	26.50	34.35	16.40
15	21	9.01	5.84	6.65
16	22	11.49	11.78	11.37
17	22	19.75	10.94	15.58
18	22	16.27	8.13	9.71
19	21	21.05	11.26	12.89
Means	26.68	17.58	13.05	14.57
SD		7.14	7.27	6.95

with decreased vocal fold adduction and vice versa), we chose to analyse the data using paired comparisons (Maxwell & Delaney, 1990). Paired comparisons allow us to test the significance of differences among the three voicing conditions and specify the direction of any differences. Tests of significance (f-ratio) were computed for mean differences in SPI for the following comparisons: breathy vs. normal, normal vs. pressed and breathy vs. pressed.

Results

Results of one-tailed f ratios computed using the difference scores across each of the three comparisons are given in Table II. The mean difference of 5.134 for the breathy vs. normal comparison was significant ($f=19.106$, $p<.001$). The difference was in the predicted direction indicating that SPI in the breathy condition was significantly higher than SPI in the normal condition. Average SPI for the normal condition was 13.05, while average SPI for the breathy condition was 17.58. The mean difference of 2.945 for the breathy vs. pressed comparison was significant ($f=7.734$, $p<.05$). Again the difference was in the predicted direction indicating that SPI was significantly higher in the breathy condition when compared to the pressed condition. Average SPI for the pressed condition was 14.57.

Table II. Comparison of SPI scores across the three voicing conditions based on difference scores.

	f	df	Sign. Level	Mean Diff
Breathy vs Normal	19.106	1,18	<.001	5.1343
Normal vs. Pressed	−2.789	1,18	.056	−2.1891
Breathy vs. Pressed	7.734	1,18	.006	2.9452

The mean difference of −2.1891 for the normal vs. pressed condition was not significant (f=−2.789, p=.056). This indicates no difference between pressed and normal phonation with respect to SPI levels. In addition, the mean difference was in the opposite direction than predicted indicating that the trend was for SPI values to be higher in the pressed condition than in the normal condition. Visual analysis of the individual data revealed that only two subjects (nos. 2 and 10) showed the predicated pattern of highest SPI for breathy phonation, intermediate values for normal phonation and lowest values for pressed phonation. Additionally, only five subjects (nos. 5, 8, 11, 14, and 16) showed reductions in SPI for the pressed condition. Conversely, all but five subjects (nos. 5, 8, 11, 14 and 16) showed higher SPI values for the breathy condition than for the normal condition. Only four subjects (nos. 1, 3, 5, and 6) failed to show higher SPI for the breathy condition when compared with the pressed condition (see Table I).

Discussion

Based on the preliminary findings of this pilot study, the relationship between SPI and strength of vocal fold adduction is not clear. The fact that SPI showed predicted changes when comparing breathy phonation to both normal and pressed phonation seems to indicate that SPI does reflect, in part, decreased vocal fold closure. Examination of the actual SPI values for the six subjects in the study who did not show increased SPI for the breathy condition reveals very little change in SPI across the conditions. This may indicate that these subjects were making very minor alterations in the physiology of voice production, which would correspond to barely perceptible changes in the acoustic properties of the signal. Though subjects that had obvious difficulty producing the breathy voice condition were eliminated from the study, we did not include perceptual ratings for the breathy and pressed conditions. Our clinical impressions were that many of the subjects produced barely perceptible differences in voice quality across the three conditions. Perhaps, SPI reflects only robust changes in vocal physiology and is not able to track the subtle changes as our subjects stimulated the different vocal adduction conditions. Our results are in agreement with those of Koreman and Putzer (1997) who found that SPI could not distinguish speakers from the normal group from those with nonpathological breathy voice. Of interest from a clinical point of view was the difficulty that many of the subjects had both producing a breathy voice and judging whether or not their own voice was breathy. The subjects were all graduate students currently enrolled in a voice disorders course. This confirms the importance of providing perceptual training in our graduate courses, in addition to training in the use of various facilitative techniques.

The trend in the current data toward increased SPI values for the pressed condition cannot be explained given our current understanding of the nature of this vocal parameter. One explanation is that the subjects were not getting tighter and/or firmer vocal fold closure during the pressed voice simulation despite the increased external tension induced by the pushing/pulling gesture. Instead, this gesture may have actually induced more supraglottic constriction. In that case, our results are in agreement with those of Koreman and Putzer (1997), who report highest levels of SPI for their subjects with unilateral vocal fold paralysis with supraglottal compensation. Average SPI for these subjects was 209, while SPI for subjects with unilateral vocal fold paralysis without compensation was 102. Though Koreman and Putzer did not describe the type of compensation used by their patients, we would expect spectral slope (which is reflected in the SPI measure) to be affected by the rigidity of the vocal tract walls, which was likely increased in our subjects through the application of the pushing/pulling gesture in the current study. Another explanation for increased SPI for the pressed condition is that SPI is sensitive to other changes in vocal fold movement that may have been induced during the pressed voice simulation.

The mean SPI for the normal condition in our group of female speakers was 13.05, which is below the suggested threshold of 14.12 given in the documentation from the manufacturer. There are no published means for female voices exclusively. Koreman and Putzer (1997) reported a mean SPI of 69 for their normal speakers (two male and two female). The only other published SPI levels for non-disordered speakers were given by Munoz et al. (2003), and though they collected data for females only the male values were reported. The sample of 47 normal male speakers had a mean SPI of 10.42. Without more published values for normal subjects, it is hard to determine if the threshold of 14.12 is clinically relevant.

With regard to the clinical utility of the SPI parameter, the data does not support a clear role in either identifying disordered voices or in measuring change pre-post therapy. The significant relationship between SPI and the breathy condition as simulated in this experiment indicates a potential for clinical use as a feedback during voice training when increased breathy quality is the desired response. Additionally, measures could be used to indicate post-therapy or post-surgical change. Clinicians must be aware that SPI is sensitive to changes in both fundamental frequency and intensity, and both must be controlled for comparisons across time to be meaningful. However, based on our experiences, requiring subjects to keep intensity stable when attempting breathy phonation adds to the difficulty of the task and may have affected the results reported here. Studies that focus on correlations between SPI values and other indicators of vocal fold adduction, such as closed quotient and skew from EGG measures, may shed light on the exact relationship between SPI and vocal fold adduction.

References

Boone, D., & McFarlane, S. (1988). *The voice and voice therapy*, fourth edition. Englewood Cliffs: Prentice-Hall.

Bruckl, M., & Sendlmeir, W. (2003). Aging female voice: an acoustic and perceptive analysis. *Proceedings of VOQUAL'03*, Geneva, August 27–29.

Deliyski, D. (1993). Acoustic model and evaluation of pathological voice production. *Proceedings of 3rd conference on speech communication and technology* (pp. 1969–1972).

Kent, R., Vorperian, H., & Duffy, J. (1999). Reliability of the multi-dimensional voice program for the analysis of voice samples of subjects with dysarthria. *American Journal of Speech and Language Pathology*, *8*, 129–136.

Koreman, J., & Putzer, M. (1997). Finding correlates of vocal fold adduction deficiencies. *Phonus 3* (pp. 155–178). Institute of Phonetics, University of the Saarland.

Maxwell, S., & Delaney, H. (1990). *Designing experiments and analyzing data* (pp. 516–542). Belmont, CA: Wadsworth.

Munoz, J., Mendoza, E., Fresneda, M., Carballa, G., & Lopez, P. (2003). Acoustic and perceptual indicators of normal and pathological voice. *Folia Phoniatrica*, *55*, 102–114.