

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/263966063>

# Spectral Moments of the Long-term Average Spectrum: Sensitive Indices of Voice Change After Therapy?

Article in *Journal of Voice* · June 2005

DOI: 10.1016/j.jvoice.2004.02.005

CITATIONS

36

READS

326

4 authors, including:



**Kristine Tanner**

Brigham Young University - Provo Main Campus

36 PUBLICATIONS 798 CITATIONS

[SEE PROFILE](#)



**Nelson Roy**

University of Utah

136 PUBLICATIONS 6,093 CITATIONS

[SEE PROFILE](#)



**Eugene Buder**

The University of Memphis

77 PUBLICATIONS 1,207 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Personality in Children with Vocal Fold Nodules [View project](#)



Personality, Psychological Factors, and Behavioral Tendencies in Children with Vocal Nodules: A Systematic Review [View project](#)

# Spectral Moments of the Long-term Average Spectrum: Sensitive Indices of Voice Change After Therapy?

\*Kristine Tanner, †Nelson Roy, ‡Andrea Ash, and §Eugene H. Buder

*Salt Lake City, Utah, Lawrence, Kansas, and Memphis, Tennessee*

---

**Summary:** Voice clinicians require an objective, reliable, and relatively automatic method to assess voice change after medical, surgical, or behavioral intervention. This measure must be sensitive to a variety of voice qualities and severities, and preferably should reflect voice in continuous speech. The long-term average spectrum (LTAS) is a fast Fourier transform-generated power spectrum whose properties can be compared with a Gaussian bell curve using spectral moments analysis. Four spectral moments describe features of the LTAS: Spectral mean (Moment 1) and standard deviation (Moment 2) represent the spectrum's central tendency and dispersion, respectively. Skewness (based on Moment 3) and kurtosis (based on Moment 4) represent the spectrum's tilt and peakedness, respectively. To examine whether the first four spectral moments of the LTAS were sensitive to perceived voice improvement after voice therapy, this investigation compared pretreatment and post-treatment voice samples of 93 patients with functional dysphonia using spectral moments analysis. Inspection of the results revealed that spectral mean and standard deviation lowered significantly with perceived voice improvement after successful behavioral management ( $p < 0.001$ ). However, changes in skewness and kurtosis were not significant. Furthermore, lowering of the spectral mean uniquely accounted for approximately 14% of the variance in the pretreatment to posttreatment changes observed in perceptual ratings of voice severity ( $p < 0.001$ ), indicating that spectral mean (ie, Moment 1) of the LTAS may be one acoustic marker sensitive to improvement in dysphonia severity.

**Key Words:** Spectral moments—Long-term average spectrum (LTAS)—Spectral mean—Dysphonia—Voice analysis.

---

---

Accepted for publication February 17, 2004.

From the \*The Department of Communication Disorders, The University of Utah, Salt Lake City, Utah; †The Department of Communication Disorders & Division of Otolaryngology—Head & Neck Surgery, The University of Utah, Salt Lake City, Utah; ‡The Department of Child Language, The University of Kansas, Lawrence, Kansas; §School of Audiology and Speech-Language Pathology, The University of Memphis, Memphis, Tennessee.

Address correspondence and reprint requests to Kristine Tanner, c/o Nelson Roy, Department of Communication Disorders, The University of Utah, 390 South 1530 East, Room 1219, Salt Lake City, UT 84112-0252. E-mail: [kristine.tanner@hsc.utah.edu](mailto:kristine.tanner@hsc.utah.edu)

*Journal of Voice*, Vol. 19, No. 2, pp. 211–222

0892-1997/\$30.00

© 2005 The Voice Foundation

doi:10.1016/j.jvoice.2004.02.005

## INTRODUCTION

In clinical circles, voice practitioners have traditionally used auditory-perceptual judgments to evaluate or document the effectiveness of voice therapy. However, results from an expanding literature suggest that—even among experienced listeners—perceptual judgments of voice quality and severity suffer from poor interjudge and intrajudge reliability.<sup>1–10</sup> In an attempt to avoid problems related to auditory-perceptual measures, acoustic analysis techniques have been offered to “objectively” assess dysphonia severity and track treatment outcomes. Recently, however, the reliability and validity of several popular acoustic measures have also been questioned. Measures such as jitter and shimmer require periodic or quasi-periodic voice signals to permit reliable analysis. These measures apply peak-picking or zero-crossing algorithms in an attempt to statistically quantify cycle-to-cycle variations in amplitude or frequency. Because many of the dysphonic voices that clinicians encounter (ie, more severely disordered voices) are highly *aperiodic*, the appropriateness, reliability, and validity of these measures when applied to such voices remain uncertain.<sup>11–20</sup>

To complicate matters, most clinically available signal processing packages require sustained vowel productions for analysis. However, certain voice disorders—such as spasmodic dysphonia (SD)—are characterized by sustained vowel productions that are often only mildly affected as compared with connected speech. Consequently, the acoustic analysis of sustained vowels can, in some cases, grossly underestimate the severity of the voice problem as well as its response to treatment. It is apparent that voice clinicians need a robust, fairly automatic index of voice severity that is representative of a diverse set of voice qualities in connected speech, and is sensitive to improvements after treatment.<sup>21,22</sup> It is unlikely, however, that a single acoustic parameter will suffice; rather, if such an index of severity is developed, it will almost certainly be composed of a relatively small subset of acoustic parameters combined in a weighted algorithm or equation. Each acoustic variable in the algorithm would be uniquely sensitive to specific characteristics of the voice signal, and when applied to a person’s voice, an

overall severity index or rating could be established.<sup>23</sup> Ideally, the acoustic index would relate well with listener ratings of voice disorder severity. The question remains, however: Which measures should be included in such an algorithm?

Research over the last several decades has shown that analysis of the long-term average spectrum (LTAS) holds promise as an index of voice quality and may be one possible acoustic marker that is sensitive to voice improvement.<sup>11,21,24–30</sup> The LTAS is a fast Fourier transform (FFT)–generated power spectrum of frequencies represented in the acoustic voice signal (either the mean of all spectra during a relatively long sample or a long-term running spectrum obtained from a long sample), and it has the advantage of not requiring a periodic or quasi-periodic voice signal to permit reliable analysis. Early work by Hartman and von Cramon found that the presence of spectral energy above 5 kHz was a strong predictor of perceived breathiness.<sup>25</sup> Kitzing examined spectral characteristics of the LTAS as compared with perceptual ratings of voice quality in patients with functional dysphonia.<sup>26</sup> Several predictors, including the relative amount of low-frequency spectral energy, successfully discriminated these pathological voice qualities. Results from Löfqvist and Mandersson showed that spectral features may be linked to breathiness and perceived hyper/hypofunction.<sup>31</sup> Spectral slope, in combination with other acoustic parameters, has also been linked to perceived breathiness and roughness.<sup>21</sup>

Hillenbrand et al<sup>32</sup> investigated spectral tilt (ie, skewness) of the LTAS, as well as signal periodicity and first harmonic amplitude, as possible predictors of perceptual ratings of breathiness. Using direct magnitude estimation (DME), listeners rated overall breathiness of sustained vowels produced by 15 vocally healthy speakers. Speakers were instructed to simulate varying levels of breathiness during sustained vowel productions. Results indicated that spectral tilt, signal periodicity, and the first harmonic amplitude were each correlated with auditory-perceptual ratings of breathiness. Expanding on this work, Hillenbrand and Houde<sup>33</sup> examined spectral tilt of the LTAS, signal periodicity, and the first harmonic amplitude as possible predictors of perceptual ratings of breathiness. Sustained vowel and connected speech samples of 20 dysphonic and 5 normal

participants were analyzed using each of the three acoustic parameters. Connected speech samples consisted of identical excerpts from a short reading passage. Because of the large ratio of vowel-to-consonant energy in the connected speech sample, the authors reasoned that it was unnecessary to remove consonant energy before the acoustic analysis. The results showed that spectral tilt of the LTAS accounted for 40% of the variance in perceptual ratings of breathiness in sustained vowels, and 70% for connected speech. Thus, a decrease in spectral energy in the higher frequency ranges predicted decreases in perceptual ratings of voice severity.

Although the LTAS holds promise as one possible marker of voice quality and severity, its acoustic quantification has been somewhat problematic. One way to quantify the LTAS output is to compare it with a Gaussian bell curve (ie, a normal random probability distribution), and then characterize the LTAS distribution using spectral moments analysis. The first four spectral moments are described as follows: Spectral mean (M1) and standard deviation (M2) describe the spectrum's central tendency and variation, respectively. M1 is the mean, or average value, of the LTAS distribution. M2 describes the variance of the distribution and may be represented in terms of standard deviation. Skewness (M3) describes the spectral tilt of the LTAS (ie, negative or positive). Positively skewed distributions appear to have a "tail" of scores or values extended to the right of the bell curve. Negatively skewed distributions have a tail extending from the left of the bell curve. Kurtosis represents the "peakedness" of the LTAS, with steep distributions producing positive kurtosis values, and flat distributions producing negative values (Figure 1).<sup>34</sup>

Recently, spectral moments have been used in analyses of both normal and disordered speakers, primarily to characterize consonant production.<sup>35–40</sup> Collectively, the results from these studies indicate that spectral moments are sensitive to specific acoustic properties in both normal and disordered speech. But the application of spectral moments analysis of the LTAS for the purpose of evaluating changes in voice disorder severity has been limited. Voice clinicians need a measure to "objectively" assess dysphonia severity and track treatment outcomes. This measure must be sensitive to heterogeneous voice

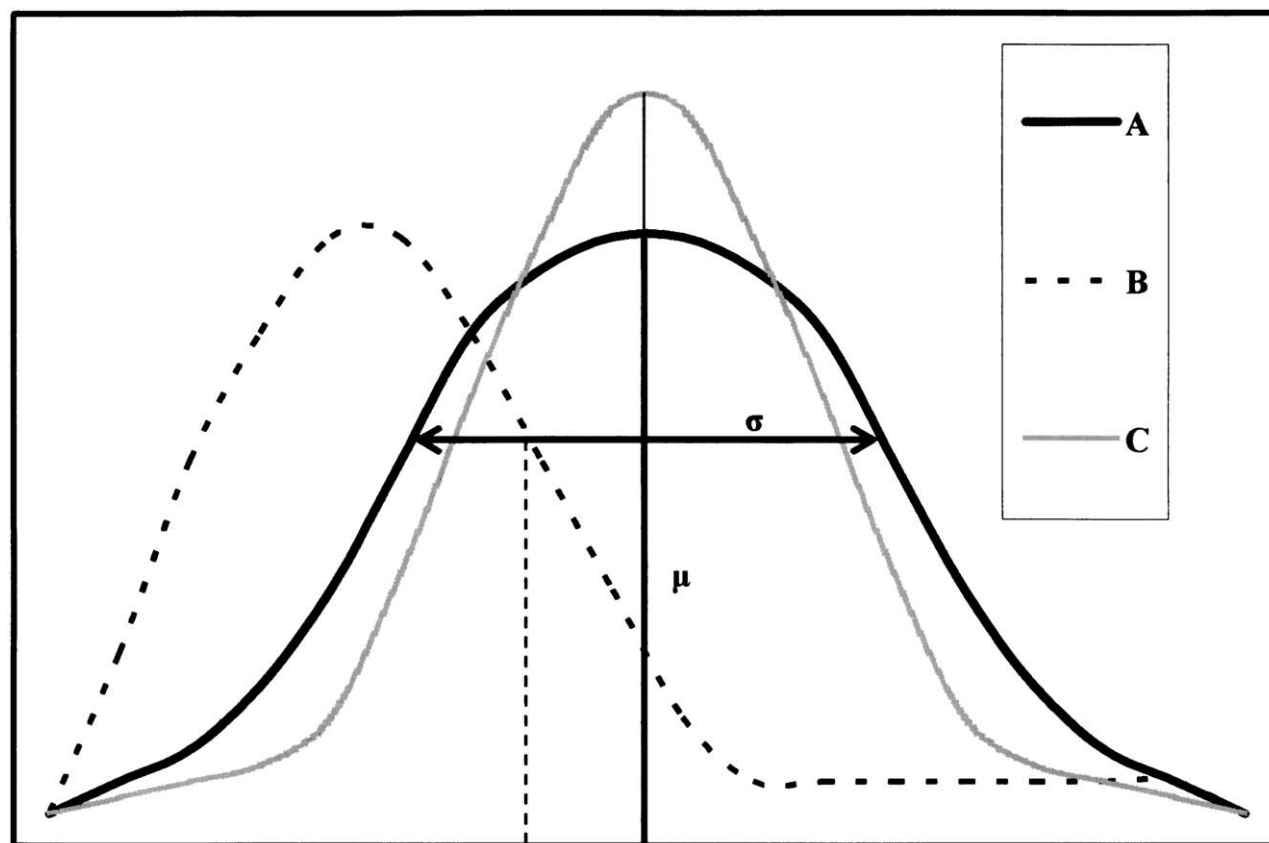
qualities and severities and be reliable, accurate, and relatively automatic. As well, it should reflect voice quality in continuous speech. To this end, the present investigation examined spectral moments of the LTAS as possible objective markers of voice change in continuous speech after voice therapy.

As a related—but secondary research question—we explored the effects of different frequency ranges and analyzing bandwidths on the LTAS results. Spectral moments of the LTAS may be generated using several frequency range selections and analyzing bandwidth parameters. The frequency range of the LTAS analysis may be adjusted to include only those frequencies represented in the acoustic speech signal.<sup>28,33,35</sup> Likewise, several analyzing bandwidths may be selected to control the number of spectral points within each frequency window of the LTAS. As window length increases and the resulting bandwidth of analysis decreases (as is the case in the Fourier analysis of acoustic signals), the number of spectral points also increases.<sup>11</sup> In theory, the default method for spectral analysis would be to use a wide analyzing bandwidth and the entire frequency range available. At present, however, it is unclear what effect frequency range selections and analyzing bandwidth parameters have on the ability of spectral moments of the LTAS to track voice change. Therefore, frequency range and analyzing bandwidth were manipulated to determine their effects, if any, on changes in spectral moments after voice therapy.

## METHOD

### Participants

Pretreatment and posttreatment voice samples from 93 patients with functional dysphonia (female, range 18–79 years, mean 47.5 years, SD 13.3 years) were selected for analysis from an archival database. All patients were found to have functional dysphonia (FD) after a comprehensive evaluation, including videolaryngoscopy by an otolaryngologist and a speech-language pathologist. Functional dysphonia is a voice disturbance in the absence of visible structural or neurological laryngeal pathology, and it is characterized by myriad voice qualities and severities. Voice samples consisted of patients reading



**FIGURE 1.** Mean ( $\mu$ ), standard deviation ( $\sigma$ ), skewness, and kurtosis of a probability distribution. “A” illustrates a normal distribution, with corresponding mean (solid vertical line) and standard deviation (horizontal arrowed lines). “B” illustrates a positively skewed distribution and corresponding shift in mean (dotted vertical line). “C” illustrates a peaked distribution with a corresponding positive kurtosis (note that distributions A and C have the same mean).

the second and third sentences of *The Rainbow Passage* (“The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch with its path high above and its two ends apparently beyond the horizon”).<sup>41</sup> Pretreatment and posttreatment recordings were collected for each of the 93 participants, for a total of 186 samples. The pretreatment and posttreatment recordings were made before and after a single session of voice therapy involving manual circumlaryngeal techniques.<sup>42</sup> Each participant was recorded in the same quiet room under identical recording conditions for the pretreatment and posttreatment samples. All samples were collected using research quality recording equipment. Although recording devices varied among patients, each individual patient’s pretreatment and posttreatment recordings were made

using identical equipment within the same session. The pretreatment and posttreatment voice samples were digitized at a sampling rate of 25 kHz using the *Computerized Speech Lab*, Model 4300 (Kay Elemetrics Corp., Lincoln Park, New Jersey).

### Acoustic Analysis

Spectral moments were generated from the LTAS of each voice sample using *Multi-Speech*, Model 3700, version 2.3 (Kay Elemetrics Corp.). To examine the effects of varying frequency ranges and analyzing bandwidths on the LTAS results, spectral moments of the LTAS were generated using different frequency ranges and analyzing bandwidths. In order to accomplish this, it was necessary to modify the *Multi-Speech* software to allow for frequency range selection (modifications provided by Speech



Technology Research Limited, Victoria, B.C., Canada and Kay Elemetrics, Corp.). (More recent versions of *Multi-Speech* include frequency range selection as a standard feature of the LTAS procedure.) Six long-term average (LTA) power spectra were generated for each of the 186 pretreatment and posttreatment voice samples. Each of the six power spectra was generated using one of two frequency ranges (0–8 kHz or 0–12.5 kHz) and one of three analyzing bandwidths (128-point/5.12 ms, 512-point/20.48 ms, or 1024-point/40.96 ms). Voice samples were not filtered, and they were spectrally analyzed using a Hamming window weighting.

### Auditory-Perceptual Evaluation

Because the purpose of this investigation was to assess how well each spectral moment changed with perceived voice change after treatment, it was necessary to obtain a perceptual change score related to voice disorder severity for each patient before and after treatment. Each pretreatment and posttreatment pair of digitized, coded samples was dubbed onto a master tape using a Tascam digital audio-recorder, Model DA-P1 (TEAC Corp., Tokyo, Japan). Five speech-language pathology students who had completed graduate coursework in voice disorder assessment and management rated each pretreatment and posttreatment voice sample. Voice samples were presented at a comfortable loudness level over loud speakers in a quiet room. Five sample pairs were presented at the beginning of the session to orient the listeners to the range of severity and to the listening task. Samples were presented to the listeners as a group; however, the listeners were not permitted to discuss their ratings during the listening session. Ten percent of the samples were repeated at the end of the session for later analysis of intrajudge reliability.

Pretreatment and posttreatment voice samples were presented to the listeners as a set, with the order randomized within each pair. Listeners were asked to indicate the overall severity of each sample by placing a mark on a 10-cm visual analog scale (VAS), ranging from “normal voice” on the extreme left of the scale to “profoundly abnormal voice” on the extreme right (see the Appendix). Listeners were asked to label the first sample in the pair as “Sample A” and the second sample in the pair as “Sample

**TABLE 1.** Estimates of Interjudge and Intrajudge Reliability for Perceptual Ratings of Voice Severity

Statistic	Pretreatment	Posttreatment
ICC (95% C.I.)	0.92 (0.86–0.95)	0.89 (0.84–0.93)
Pearson $r$ ( $r^2$ )	0.86 (0.74)	0.87 (0.76)

ICC, Intraclass correlation coefficient.

B.” The Appendix illustrates the listener rating protocols and instructions. These instructions were read aloud as the listeners followed along on their protocols.

Severity ratings, ranging from 0 to 10, were measured from the marks on the VAS, with values approaching 0.0 cm indicating “normal voice,” and values approaching 10.0 cm indicating “profoundly abnormal voice.” Pretreatment ratings ranged from 0.8 to 10.0, with a mean severity rating of 7.3. Posttreatment ratings ranged from 0.1 to 9.7, with a mean severity of 1.9. A paired-samples  $t$  test indicated a significant improvement in perceptual ratings after treatment ( $t(92) = 22.51$ ,  $p < .001$ ) (Table 1). Because the purpose of this study was to track voice change after treatment and not treatment efficacy per se, only pretreatment and posttreatment sample pairs that demonstrated voice improvement were included in the study. That is, only sample pairs that showed a lowering of voice severity as indicated on the VAS were included in the acoustic analysis. Of the original corpus of 188 pretreatment and posttreatment samples (ie, 94 pairs), 186 (or 93 pairs) showed voice improvement after treatment as measured perceptually. Change scores were generated from the differences between each of the 93 pretreatment and posttreatment severity ratings. These change scores were then compared with changes in each of the four spectral moments after treatment.

### Listener Reliability

Intraclass correlation coefficients (ICC) were calculated to determine the overall level of reliability among listeners (ie, interjudge reliability) for both pretreatment and posttreatment samples. For pretreatment samples, listeners had an ICC of 0.92, with a 95% confidence interval (CI) of 0.86–0.95 (Table 1). For posttreatment samples, listeners had

**TABLE 2.** Spectral Mean Values (in Hertz) for Each of the Six LTA Spectra, with Corresponding *t* and *p* Values

Frequency range	Analyzing bandwidth	Pretreatment mean (std. error)	Posttreatment mean (std. error)	<i>t</i> ( <i>df</i> )	<i>p</i> <
0–8 kHz	128 points	1187 (93)	848 (48)	3.81 (92)	0.001*
	512 points	1107 (87)	758 (28)	4.15 (92)	0.001*
	1024 points	1085 (86)	743 (28)	4.09 (92)	0.001*
0–12.5 kHz	128 points	1406 (118)	918 (56)	4.40 (92)	0.001*
	512 points	1324 (109)	812 (36)	4.80 (92)	0.001*
	1024 points	1282 (110)	794 (35)	4.59 (92)	0.001*

\*Statistically significant,  $\alpha < 0.05$ .

an ICC of 0.89, with a 95% CI of 0.84–0.93. Results from this analysis indicated a relatively high level of reliability among listeners. A Pearson product moment correlation coefficient of 0.86 ( $r^2 = 0.74$ ) was obtained from the 10% repeated severity ratings as a measure of reliability within each listener (ie, intrajudge reliability). The average absolute difference between original and repeated severity ratings was 0.8 cm (range 0–3.7 cm), indicating acceptable levels of intrajudge reliability.

### Acoustic Measurement Reliability

Once all analysis parameters are set for generation of the LTAS and each of the spectral moments, the acoustic analysis is completely automatic. Even so, 10% of the LTA spectra were regenerated to verify acoustic measurement reliability. An average absolute difference of 0 (range = 0; Pearson  $r = 1.0$ ) indicated perfect reliability for the acoustic analysis procedure.

## RESULTS

To examine changes in each spectral moment after treatment, paired-samples *t* tests were performed. Six LTA power spectra were generated for each pretreatment and posttreatment voice sample. Each of the six power spectra was generated using one of two frequency ranges and one of three analyzing bandwidths. Individual statistical analyses were conducted for each of the six power spectra. The results are presented in Tables 2–5.

The results were qualitatively similar across the two frequency ranges and the three analyzing bandwidths. For each of the six power spectra, M1 (spectral mean) and M2 (standard deviation) decreased

significantly after treatment. Changes in M3 (skewness) and M4 (kurtosis) were not significant. Figure 2 shows the average (composite) pretreatment and posttreatment LTAS curves for the 0 to 8-kHz, 128-point/5.12-ms (wide bandwidth) power spectra. Comparison of pretreatment and posttreatment LTAS curves revealed an apparent decrease in high-frequency energy in the power spectrum after treatment. As illustrated in Figure 2, a greater proportion of acoustic energy was represented in the lower frequencies after voice therapy.

To evaluate the relationship between each spectral moment and its contribution to the explained variance in change scores on the severity rating scale (ie, the difference between pretreatment and posttreatment ratings), several multiple regression analyses were undertaken. Results from these analyses were also qualitatively similar for each of the six LTA power spectra. Changes in M1 uniquely accounted for a significant amount of the variance in the change scores on the severity dimension (Adj.  $R^2 = 0.131$ – $0.153$ ,  $F = 14.86$ – $17.57$ ,  $p < 0.001$ ) (Table 6). For the 0 to 12.5-kHz frequency range, the addition of M3 to the regression equation increased the explained variance in perceptual change scores by 2% to 3% using the wide or narrow analyzing bandwidths (ie, 128 or 1024 points). However, M3 did not consistently add unique information to the regression equation across the six power spectra. It is also important to recall that M3 did not change significantly from pretreatment to posttreatment as indicated by the paired-samples *t* tests. Thus, it is likely that M3 only approached statistical significance in the multiple regression analyses as the analyzing frequency range was increased. That is, the

**TABLE 3.** *Standard Deviation Values for Each of the Six LTA Spectra, with Corresponding  $t$  and  $p$  Values*

Frequency range	Analyzing bandwidth	Pretreatment mean (std. error)	Posttreatment mean (std. error)	$t(df)$	$p <$
0–8 kHz	128 points	1231 (63)	962 (49)	4.18 (92)	0.001*
	512 points	1267 (64)	933 (40)	5.72 (92)	0.001*
	1024 points	1227 (64)	908 (39)	5.62 (92)	0.001*
0–12.5 kHz	128 points	1598 (86)	1149 (62)	5.46 (92)	0.001*
	512 points	1617 (82)	1096 (52)	6.94 (92)	0.001*
	1024 points	1569 (83)	1067 (50)	6.88 (92)	0.001*

\*Statistically significant,  $\alpha < 0.05$ .

**TABLE 4.** *Skewness Values for Each of the Six LTA Spectra, with Corresponding  $t$  and  $p$  Values*

Frequency range	Analyzing bandwidth	Pretreatment mean (std. error)	Posttreatment mean (std. error)	$t(df)$	$p =$
0–8 kHz	128 points	4.16 (.49)	4.22 (.26)	−0.11 (92)	0.912
	512 points	3.88 (.39)	4.33 (.20)	−1.27 (92)	0.207
	1024 points	4.04 (.36)	4.40 (.20)	−1.19 (92)	0.234
0–12.5 kHz	128 points	4.62 (.52)	4.91 (.32)	−0.52 (92)	0.604
	512 points	4.27 (.40)	4.93 (.25)	−1.78 (92)	0.078
	1024 points	4.39 (.38)	4.97 (.23)	−1.77 (92)	0.078

**TABLE 5.** *Kurtosis Values for Each of the Six LTA Spectra, with Corresponding  $t$  and  $p$  Values*

Frequency range	Analyzing bandwidth	Pretreatment mean (std. error)	Posttreatment mean (std. error)	$t(df)$	$p =$
0–8 kHz	128 points	48.64 (18)	33.66 (6)	0.80 (92)	0.422
	512 points	35.06 (10)	29.73 (3)	0.58 (92)	0.560
	1024 points	34.11 (7)	29.68 (3)	0.67 (92)	0.500
0–12.5 kHz	128 points	55.97 (20)	47.93 (8)	0.38 (92)	0.703
	512 points	40.99 (10)	40.56 (5)	0.04 (92)	0.965
	1024 points	40.93 (9)	39.84 (4)	0.12 (92)	0.901

farther the power spectrum was extended to the right, the more the LTAS curve became positively skewed. M2 and M4 were not significant predictors of voice severity after voice therapy.

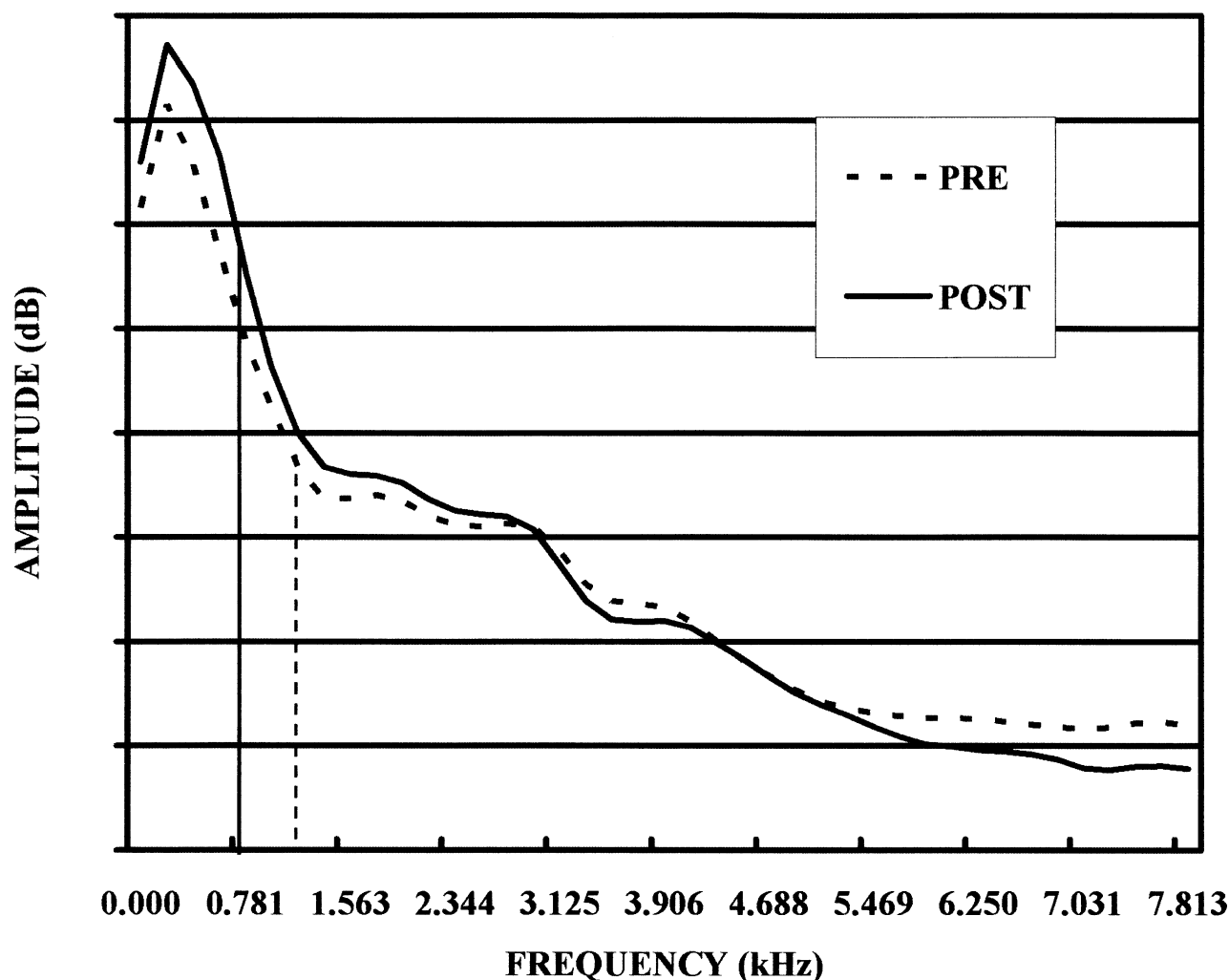
To summarize, M1 consistently accounted for a significant amount of the variance in perceptual ratings after voice therapy. Once M1 entered the regression equation, no other moments contributed uniquely and substantively to the residual variance. Pearson correlations revealed relationships among the pattern of change for M1–M4 from pretreatment to posttreatment samples (Table 7). Based on the

regression analyses, it is clear that each spectral moment provided similar information regarding voice change after treatment.

## DISCUSSION

The present investigation examined spectral moments of the LTAS as possible objective markers of voice change after voice therapy. To examine whether the first four spectral moments of LTAS were sensitive to perceived voice improvement after





**FIGURE 2.** The average LTA power spectra, pretreatment and posttreatment, for the 93 participants using the 128-point analyzing bandwidth and 0 to 8-kHz frequency range. Average pretreatment values for M1–M4 are as follows (in Hertz): M1 = 1232, M2 = 1418, M3 = 4.23, M4 = 42.62. Average posttreatment values for M1–M4 are as follows: M1 = 812, M2 = 1019, M3 = 4.63, M4 = 36.90. The vertical dashed line indicates the average spectral mean for each of the 93 pretreatment voice samples. The vertical solid line represents the average spectral mean for each of the 93 posttreatment samples.

voice therapy, this investigation compared pretreatment and posttreatment voice samples of patients with functional dysphonia using spectral moments analysis. Inspection of the results revealed that spectral mean and standard deviation lowered significantly with perceived voice improvement after successful behavioral management ( $p < 0.001$ ), whereas changes in skewness and kurtosis were not significant. Furthermore, lowering of the spectral mean uniquely accounted for approximately 14% of

the variance in the pretreatment to posttreatment changes observed in perceptual change scores related to dysphonia severity ( $p < 0.001$ ).

At first glance, the results from the present investigation may seem somewhat incongruous with previous reports using spectral shape to characterize voice quality. Hillenbrand and Houde<sup>33</sup> recently reported that spectral tilt accounted for 70% of the variance in perceptual ratings using connected speech voice samples. However, it is important to

**TABLE 6.** Regression Variables for Each of the Six LTA Spectra, with Corresponding Adjusted  $R^2$ ,  $F$ , and  $p$  Values

Fz. range (Hertz)	Bandwidth (points)	Variable (moments)	Adj. $R^2$	$df$	$F$	$p <$
0–8 kHz	128	M1	0.144	1, 91	16.47	0.001*
	512	M1	0.131	1, 91	14.86	0.001*
	1024	M1	0.143	1, 91	16.40	0.001*
0–12.5 kHz	128	M1 (+M3)	0.139 (0.167)	1, 91 (2, 90)	15.89 (10.20)	0.001*
	512	M1	0.153	1, 91	17.57	0.001*
	1024	M1 (+M3)	0.132 (0.164)	1, 91 (2, 90)	15.03 (10.00)	0.001*

\*Statistically significant,  $\alpha < 0.05$ .

**TABLE 7.** Average Pearson Correlations for Changes in Each Spectral Moment from Pretreatment to Posttreatment, with Corresponding Ranges and  $p$  Values

Spectral moment pair	Average Pearson $R$	Pearson $r$ range	$p <$
M1, M2	0.75	0.69 to 0.79	0.001*
M1, M3	−0.53	−0.46 to −0.61	0.001*
M1, M4	−0.24	−0.20 to −0.30	0.002–0.029*
M2, M3	−0.66	−0.62 to −0.72	0.001*
M2, M4	−0.37	−0.34 to −0.45	0.001*
M3, M4	0.89	0.87 to 0.91	0.001*

\*Statistically significant,  $\alpha < 0.05$ .

note that the dominant perceptual feature in Hillenbrand and Houde's study was breathiness. That is, disordered voices were selected based on different levels of perceived breathiness in the voice signal. As the authors acknowledged, findings may be different for voices that possess other dominant perceptual features (eg, rough, hoarse, pressed, etc.). It is important to note that although all of the participants in the present study were from the diagnostic category of functional dysphonia, a wide range of voice severities was represented (see Table 1). Given the perceptually diverse nature of voice severities included in the present study, these findings are an encouraging first step toward identifying one objective marker of perceived voice change.

It is interesting that M1 (spectral mean) was the only consistent predictor of perceived voice change in the present investigation, accounting for approximately 14% of the variance in perceptual change scores related to dysphonia severity. The reduction in spectral mean may reflect a number of phonatory changes that occurred with successful treatment. Two likely phonatory changes are (1) a reduction

of high-frequency energy associated with excessive vocal fold impact and/or turbulent noise, and (2) an increase in the magnitude of the fundamental frequency. The latter effect can also be measured by the difference in magnitudes of the first two harmonics, although this metric may not be so easily automated as the current LTAS procedure.<sup>43,44</sup> Buder et al<sup>45</sup> observed a significant change in the average LTAS distribution from women with spasmodic dysphonia after their first BOTOX injection, and isolated an increase in the relative magnitude of the first harmonic as one source of this change.

Although M3 (skewness) increased the explained variance by 2% to 3% using the 0 to 12.5-kHz frequency range and wide (128-point) and narrow (1024-point) analyzing bandwidths, M3 was not a consistently significant predictor of perceived voice change. It is possible that the statistical effect of M3 was an artifact of increasing the analyzing frequency range. When the upper limit of the analyzing frequency range was extended from 8 to 12.5 kHz, the right tail of the LTAS curve was extended. By continuing to extend the LTAS curve to the right,

assuming that acoustic energy was still present, M3 eventually became statistically significant. Results from Pearson correlations indicated statistically significant relationships among each of the four spectral moments. Therefore, it was not surprising that higher moments failed to explain any significant variance in perceptual ratings beyond that already contributed by M1. Collectively, the first four spectral moments seem to provide similar information regarding changes in the LTAS curve after voice treatment.

It was somewhat surprising that adjusting the analyzing frequency range did not have larger substantive effects on how well spectral moments tracked voice change. Theoretically, lowering the upper end of the analyzing frequency range from 12.5 kHz to 8 kHz should isolate the voice signal from ambient high-frequency noise that might be contained in the audiorecordings.<sup>28,38</sup> However, the ability of spectral moments to track perceived voice change did not significantly improve by lowering the frequency range. Statistically significant increases in skewness using the 0 to 12.5-kHz frequency range did not contribute substantively to explained variance in perceptual change scores. Although it is possible that the audiorecordings in the present study were relatively free of extraneous environmental noise, it is also possible that spectral moments of the LTAS are less sensitive to recording conditions than previously hypothesized.

This investigation was undertaken to examine spectral moments of the LTAS as possible objective markers of voice change after treatment. Analysis of the results indicated that large, statistically significant changes in spectral mean occurred from pretreatment to posttreatment voice samples. However, these changes only accounted for 14% of the variance in perceptual change scores. It is clear that although valuable information is gained by spectral moments analysis, particularly changes in spectral mean, other acoustic markers will be required to provide a comprehensive acoustic assessment of voice quality in continuous speech. Given the wide range of voice severities under the diagnostic category of functional dysphonia in the present investigation, these results are an encouraging first step toward the development of an objective and reliable acoustic index of voice severity.

## REFERENCES

1. Bassich CJ, Ludlow CL. The use of perceptual methods by new clinicians for assessing voice quality. *J Speech Hear Disord.* 1986;51:125–133.
2. Kreiman J, Gerratt BR. The perceptual structure of pathologic voice quality. *J Acoust Soc Am.* 1996;100:1787–1795.
3. Kreiman J, Gerratt BR, Precoda K. Listener experience and perception of voice quality. *J Speech Hear Res.* 1990;33:103–115.
4. Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception. *J Speech Hear Res.* 1992;35:512–520.
5. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res.* 1993;36:21–40.
6. Kempster GB, Kistler DJ, Hillenbrand J. Multidimensional scaling analysis of dysphonia in two speaker groups. *J Speech Hear Res.* 1991;34:534–543.
7. Kent RD. Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders. *Am J Speech Pathol.* 1996;5:7–23.
8. Kreiman J, Gerratt BR. Sources of listener disagreement in voice quality assessment. *J Acoust Soc Am.* 2000;108:1867–1876.
9. Gerratt BR, Kreiman J. Measuring vocal quality with speech synthesis. *J Acoust Soc Am.* 2001;110:2560–2566.
10. Yumoto E, Sasaki Y, Okamura H. Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *J Speech Hear Res.* 1984;27:2–6.
11. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice.* 2nd ed. San Diego, CA: Singular; 2000:225–295.
12. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res.* 1996;39:126–134.
13. Gelfer MP. Fundamental frequency, intensity, and vowel selection: effects on measures of phonatory stability. *J Speech Hear Res.* 1995;38:1189–1199.
14. Viera MN, McInnes FR, Jack MA. Comparative assessment of electroglottographic and acoustic measures of jitter in pathological voices. *J Speech Lang Hear Res.* 1997;40:170–183.
15. Rabinov CR, Kreiman J, Gerratt BR, Bielamowicz S. Comparing reliability of perceptual ratings of roughness and acoustic measure of jitter. *J Speech Hear Res.* 1995;38:26–32.
16. Karnell M. Laryngeal perturbation analysis: minimum length of analysis window. *J Speech Hear Res.* 1991;34:544–548.
17. Hillenbrand J, Houde R. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res.* 1996;39:311–321.
18. Titze IR. *Workshop on Acoustic Voice Analysis: Summary Statement.* Denver, CO: National Center for Voice and Speech; 1995.

19. Titze IR, Liang H. Comparison of  $F_0$  extraction methods for high-precision voice perturbation measurements. *J Speech Hear Res.* 1993;36:1120–1133.
20. Titze IR, Winholtz WS. Effect of microphone type and placement on voice perturbation measurements. *J Speech Hear Res.* 1993;36:1177–1190.
21. de Krom G. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J Speech Hear Res.* 1995;38:794–811.
22. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Hear Res.* 2001;44:327–339.
23. Awan SN, Roy N. Acoustic prediction of voice type and severity in adult females. *Clin Linguist Phon.* 2004 (in press).
24. Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol.* 1980;90:441–451.
25. Hartmann E, von Cramon D. Acoustic measurement of voice quality in central dysphonia. *J Commun Disord.* 1984;17:425–440.
26. Kitzing P, Akerlund L. Long-time average spectrograms of dysphonic voices before and after therapy. *Folia Phoniatri.* 1993;45:53–61.
27. Linville SE. Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatri Logop.* 1998;50:35–48.
28. Mendoza E, Muñoz J, Valencia Naranjo N. The long-term average spectrum as a measure of voice stability. *Folia Phoniatri Logop.* 1996;48:57–64.
29. Mendoza E, Valencia N, Muñoz J, Trujillo H. Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *J Voice.* 1996;10:59–66.
30. Pittam J. Discrimination of five voice qualities and prediction to perceptual ratings. *Phonetica.* 1987;44:38–49.
31. Löfqvist A, Mandersson B. Long-time average spectrum of speech and voice analysis. *Folia Phoniatri.* 1987;39:221–229.
32. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res.* 1994;37:769–778.
33. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res.* 1996;39:311–321.
34. Kent RD, Read WC. *The Acoustic Analysis of Speech.* 2nd ed. San Diego, CA: Singular; 2002:124–128.
35. Buder EH, Kent RD, Kent JF, Milenkovic P, Workinger M. FORMAFFA: an automated formant, moment, fundamental frequency, amplitude analysis of normal and disordered speech. *Clin Ling Phon.* 1996;10:31–54.
36. Forrest K, Weismer G, Elbert M, Dinnsen DA. Spectral analysis of target-appropriate /t/ and /k/ produced by phonologically disordered and normally articulating children. *Clin Ling Phon.* 1994;8:267–281.
37. Forrest K, Weismer G, Hodge M, Dinnsen DA, Elbert M. Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clin Ling Phon.* 1990;4:327–340.
38. Forrest K, Weismer G, Milenkovic P, Dougall RN. Statistical analysis of word-initial voiceless obstruents: preliminary data. *J Acoust Soc Am.* 1988;84:115–123.
39. Nittrouer S. Children learn separate aspects of speech production at different rates: evidence from spectral moments. *J Acoust Soc Am.* 1995;97:520–530.
40. Tjaden K, Turner GS. Spectral properties of fricatives in amyotrophic lateral sclerosis. *J Speech Lang Hear Res.* 1997;40:1358–1372.
41. Fairbanks G. *Voice and Articulation Drillbook.* New York: Harper and Brothers; 1960.
42. Roy N, Bless DM, Heisey D, Ford CN. Manual circumlaryngeal therapy for functional dysphonia: an evaluation of short- and long-term treatment outcomes. *J Voice.* 1997;11:321–331.
43. Hanson HM. Glottal characteristics of female speakers: acoustic correlates. *J Acoust Soc Am.* 1997;101:466–481.
44. Hanson HM, Chuang ES. Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *J Acoust Soc Am.* 1999;106:1064–1077.
45. Buder EH, Cannito MP, Taylor ML, Woodson GE, Murry T. Assessment of BOTOX effects on adductor spasmodic dysphonia using LTAS guided analysis. *J Acoust Soc Am.* 1999;5:1246.

## APPENDIX

### SCORESHEET FOR AUDITORY-PERCEPTUAL EVALUATION

#### *Identifying Information*

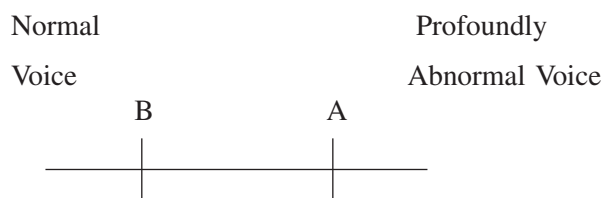
Listener Number \_\_\_\_\_

Age \_\_\_\_\_

Date \_\_\_\_\_

#### *Instructions*

This recording contains a series of voice samples that have been played randomly in pairs. Before each pair of samples, you will hear a voice telling you the number that corresponds with the pair. The first voice sample in each pair will be Sample A; the second will be Sample B. After hearing the first voice sample, use a vertical slash to indicate the point on the horizontal line that represents the overall vocal quality of the sample. The horizontal line represents a continuum, with the extreme left representing normal voices and the extreme right representing profoundly abnormal voices. Clearly write an "A" above the slash to indicate that it represents the first voice sample in the pair. Then use a vertical slash on the same horizontal line to represent the overall vocal quality of the second sample that you hear. Clearly write a "B" above this vertical slash. Please see the example below.



KEY A = First Sample Presented in the Pair

B = Second Sample Presented in the Pair

For example, before the first pair, you will hear "Sample 1" spoken on the tape. Then you will hear the first sample in the pair, Sample A. Immediately after the first sample, mark your rating on the visual analog scale and label it "A." Then you will hear the second sample in the pair. Mark your rating of the second sample and label it "B." Then you will hear a voice saying "Sample 2." Repetitions of the samples will be permitted on request. Several practice samples will be presented to orient you, the listener, to the range of severity and the listening task. The task will take approximately 2 hours to complete. There will be a short break after each half-hour of the listening task.

#### *Auditory-Perceptual Rating*

Normal	Profoundly
Voice	Abnormal Voice

PAIR # 1 \_\_\_\_\_