PERCEPTUAL AND ACOUSTIC ANALYSIS OF THE VOICE IN ACUTE LARYNGITIS

Anders Löfqvist and Lucyna Schalén

Department of Logopedics and Phoniatrics, Lund University, Lund, Sweden

ABSTRACT
Acoustic and perceptual analyses of the voice were made in 20 cases of laryngitis during the acute stage and after recovery. The acoustic analysis indicated considerable variability in the pattern of change between the acute and the control conditions. However, the perceptual analysis showed consistent, and significant, differences in the ratings of the two conditions. Correlations between acoustic measures and perceptual ratings were generally low.

1. INTRODUCTION
One goal of applied voice analysis is the development of acoustic measurements that are useful for the clinical management of voice disorders. Such measures may supplement current perceptual evaluations used in clinical analysis of dysphonia. Their main advantage is that results obtained at different places can be compared. This is not necessarily the case for the results of perceptual analysis, where the background and training of the listeners influence the results [1, 2]. The present study compares perceptual evaluations and acoustic measurements of dysphonia in acute laryngitis.

2. METHODS
2.1 Material
Voice samples from 20 adults (11 females and 9 males) with dysphonia due to acute infectious laryngitis were analyzed.

2.2 Procedure
Voice recordings were made under standardized conditions in a sound-proof room during the acute stage, and a control recording at least two months later. A short story served as the speech material. The duration of the recorded speech was approximately 40 s.

The perceptual evaluation was made by a group of four experienced clinicians using a 5-point rating scale, where 0 represented normal, and 4 maximal deviation. The evaluation comprised 12 different voice qualities. Of these, only those were used in the present study that met two criteria: a significant test-retest correlation, and a significant interjudge reliability (Kendall W). The qualities used here were: diplophonia, breathiness, roughness, aphonia, and voice breaks. In addition, vocal fry was also included, although it failed to show a significant test-retest correlation.

Two different acoustical analyses were made. First, long-time-average spectra were calculated using the procedure described in [7]. This analysis was made of the whole recording, excluding pauses and voiceless segments of the speech signal. Based on this analysis, a rough measure of the tilt of the source spectrum was obtained by the ratio of energy in the frequency bands 0-1 and 1-5 kHz. In addition, the relative energy level in the frequency range 5-8 kHz was calculated; this measure is related to the presence or absence of noise in the voice [9]. Second, the relationship between non-harmonic to harmonic energy (N/S) was estimated using the procedure described in [8]. Due to the computational complexity of this procedure, this analysis only covered a single stressed vowel in the recording; its duration was in the range 100-150 ms.

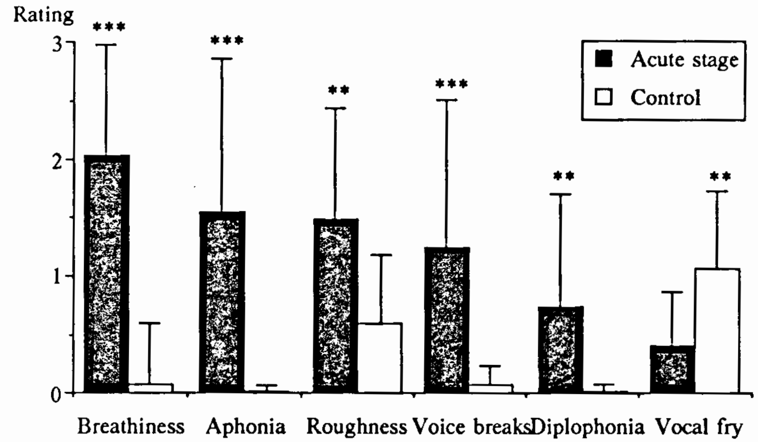

Figure 1. Results of perceptual analysis. *** $p<0.001$; ** $p<0.005$.

For both analyses, a 12-bit A/D conversion was used. The sampling rate was 20 kHz for the long-time spectral analysis, and 10 kHz for the N/S analysis.

3. RESULTS
The results of the perceptual analysis are shown in Figure 1. For all six voice qualities, there were significant differences between voice in the acute and the control conditions. All but one of the qualities showed a decrease from the acute to the control stage; the exception was vocal fry.

While the perceptual analysis indicated that there were significant group differences between the acute and the control conditions, the results of the acoustic analysis showed non-significant group differences between the two conditions. The reason is that different voices showed different acoustic patterns of change between the acute and the control condition. This is illustrated in Figures 2 and 3. Here, the voices have been divided into two groups based on the pattern of change revealed by the long-time spectral analysis. Thus, the top part of Figure 2 shows 9 voices where the predominant change is a decrease in the relative energy between 5-8 kHz. The difference between conditions is significant, $t(16) = 4.123$, $p<0.05$.

The lower part of Figure 2 shows the remaining 11 voices, where the major change is a decrease in the ratio of energy 0-1/1-5 kHz; also this change is significant, $t(20) = 4.539$, $p<0.01$. Similar results were found for the relationship between harmonic and non-harmonic components in the voice. The top and lower panels of Figure 3 plots the results of N/S for the acute and control conditions for two groups of voices. These groups correspond to the ones shown in the top and lower parts of Figure 2, respectively. As shown in the top panel of Figure 3, 8 voices in this group showed a decrease in the N/S from the acute to the control condition. The difference between conditions is significant, $t(16) = 2.168$, $p<0.05$. For the remaining 11 voices, the lower panel of Figure 3 shows an increase of N/S from the acute to the control conditions for 8 of them; the difference is not significant, however.

Pearson product-moment correlations were calculated between the acoustical measures and the perceptual ratings. Significant correlations were found between the rating of breathiness and the relative energy level between 5-8 kHz ($r = .43$, $p<0.01$), vocal fry and the relative energy level between 5-8 kHz ($r = -0.38$, $p<0.05$), and roughness and N/S ($r = 0.5$, $p<0.01$). The correlations between the

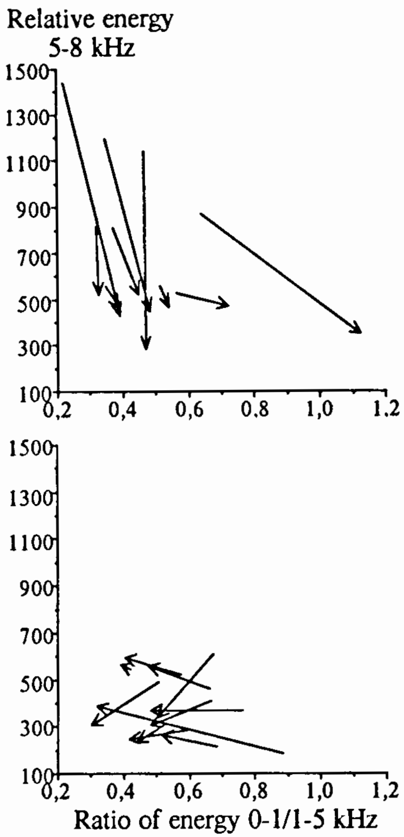

Figure 2. Results of long-time-spectral analysis. The arrows indicate the direction and magnitude of change between the acute and control conditons.

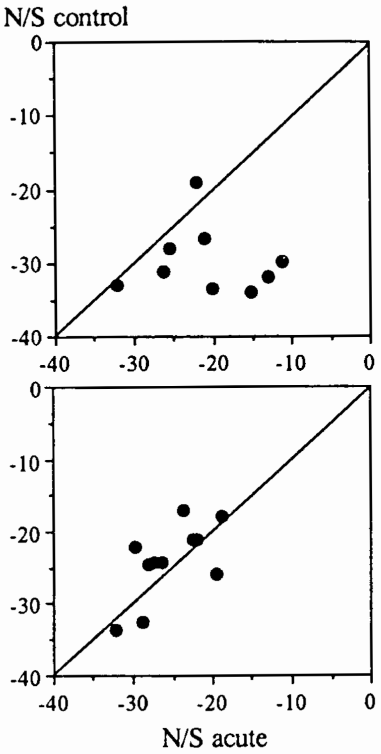

Figure 3. Results of N/S analysis.

ratings of vocal fry and all acoustic measures were negative, although only one was statistically significant.

4. DISCUSSION

The results of the present study indicate that the perceptual ratings of the voices differed between the acute and control conditions. The acoustic analysis did not reveal any overall consistent findings. Rather, two patterns of change were identified.

With the exception of vocal fry, all other perceptual qualities showed a decrease from the acute to the control condition. The changes are most likely related to inflammatory changes in the laryngeal mucosa: edema and a decreased mucosal wave. Vocal fry is presumably a common characteristic of normal voices.

The acoustic analyses suggest two patterns of change between conditions. In one of them, shown in the top panels of Figures 2 and 3, the relative energy between 5-8 kHz decreases as well as the amount of non-harmonic energy. These voices thus contain less noise in the control condition, presumably due to a

better glottal closure during phonation. The other group, shown in the bottom panels of Figures 2 and 3, shows a reduction in the measure of spectral tilt.

Signifciant correlations were found between some acoustic and perceptual results. Breathiness was positively correlated with relative energy between 5-8 kHz. This is reasonable, given that this acoustic measure is an indicator of noise. Vocal fry was negatively correlated with relative energy between 5-8 kHz. Again, this is reasonable given that voices characterized by vocal fry can be assumed to contain less noise. Roughness was positively correlated with N/S. Roughness is most likely related to both the amount of noise and to perturbations in time and amplitude. Interestingly, the N/S measure is sensitve to both these aspects of the voice source. That is, a high degree of perturbations will increase the value of N/S.

We should note, furthermore, that the acoustic measures we have applied are related to the frequency domain. The perceptual qualities of voice breaks, aphonia, and diplophonia are most likely related to temporal properties of the voice source. Hence, we should not expect them to be highly correlated with the present set of acoustic measures. In addition, the psychophysics of voice evaluation is far from understood, given the complexity of the signal.

Some studies have shown quite significant correlations between acoustic measurements and perceputal ratings [3, 4, 5, 6]. However, the highest correlations are usually found between acoustic measurements and perceptual "supercategories", based on factor analysis or composite measures. When simple perceptual qualities are used, as in the present study, correlations tend to be reduced.

5. ACKNOWLEDGMENTS

This work was supported by funds from the Faculty of Medicine, Lund University.

6. REFERENCES

[1] ANDERS, L., HOLLIEN, H., HURME, P., SONNINEN, A. & WENDLER, J. (1988), "Perception of hoarseness by several classes of listeners", *Folia Phoniatrica, 40,* 91-100.
[2] GELFER, M. P. (1988), "Perceptual attributes of voice: Development and use of rating scales", *Journal of Voice, 2,* 320-326.
[3] HAMMARBERG, B. & ASKENFELT, A. (1986), "Speech waveform perturbation analysis: A perceptual - acoustical comparison of seven measures" *Journal of Speech and Hearing Research, 29,* 50-64.
[4] HAMMARBERG, B., FRITZELL, B., GAUFFIN, J. & SUNDBERG, J. (1986), "Acoustic and perceptual analysis of vocal dysfunction", *Journal of Phonetics, 14,* 533-547.
[5] HAMMARBERG, B., FRITZELL, B., GAUFFIN, J., SUNDBERG, J. & WEDIN, L. (1980), "Perceptual and acoustic correlates of abnormal voice qualities", *Acta Otolaryngologica, 90,* 441-451.
[6] HAMMARBERG, B., FRITZELL, B. & SCHIRATZKI, H. (1984), "Teflon injection in 16 patients with paralytic dysphonia: Perceptual and acoustic evaluations", *Journal of Speech and Hearing Disorders, 49,* 72-82.
[7] LÖFQVIST, A. & MANDERSSON, B. (1987), "Long-time-average spectrum of speech and voice analysis", *Folia Phoniatrica, 39,* 221-229.
[8] MUTA, H., BAER, T., WAGATSUMA, K., MURAKOA, T. & FUKUDA, H. (1988), "A pitch-synchronous analysis of hoarseness in running speech, *Journal of the Acoustical Society of America, 75,* 224-230.
[9] YANAGIHARA, N. (1967), "Significance of harmonic changes and noise components in hoarseness", *Journal of Speech and Hearing Research, 10,* 531-541.