LTAS criteria pertinent to the measurement of voice quality

Kitzing, Peter

Published in:
Journal of Phonetics

1986

Journal of Phonetics (1986) **14**, 477–482

LTAS criteria pertinent to the measurement of voice quality

Peter Kitzing

Department of Phoniatrics, University ENT-Clinic, General Hospital, Malmö, Sweden

Earlier attempts to establish LTAS criteria to quantify pathological voice qualities in dysphonic patients have mainly centered on organic voice disorders or qualities signalling organic lesions, like breathiness or roughness. The present investigation emphasizes qualities of disturbed voice due to aberrant function in organically healthy vocal organs. The same passage of continuous speech was produced by 10 experienced voice therapists in four different recognizable voice qualities. The speech samples were analysed by means of a commercially available narrow band analyser with a facility to produce long time average spectrograms. Using a personal computer connected to the analyser, a number of criteria were tested to differentiate among the voice samples. From the results it can be concluded that criteria expressing the slope of the LTA-spectrum or the energy ratio between the fundamental and first formant regions differentiate with statistical significance between samples of systematically different voice quality. However, there is also a substantial influence of voice intensity on the LTAS configuration, and the noted differences were in the magnitude of only 3–6%. Furthermore, inter-individual differences between LTA-spectra seem to be substantial.

1. Introduction

Most of the research on LTAS criteria pertinent to pathological voice quality has been based on patients with dysphonia caused by organic disease of the larynx, like paresis or cancer (Frøjkaer-Jensen & Prytz, 1974, 1976; Gerull, Giesen, Hippel, Mrowinski & Schweers, 1977; Dejonckere, 1983; Hammarberg, Fritzell & Schiratzki, 1984; Hiraoka, Kitazoe, Ueta, Tanaka & Tanabe, 1984; Hurme & Sonninen, 1985) or concentrated on voice qualities signalling organic pathology, like roughness and breathiness (Wendler *et al.*, this issue). Reasons might be that as a rule, aberrant voice qualities caused by laryngeal pathology are expected to differ more from normal voices and obviously to show less variability than just "functional" dysphonias. Therefore, signs of organic dysphonia may be more prominent in the acoustical analysis than criteria of "functional" voice disorders. Indeed, Hammarberg *et al.* (this issue) in their most thorough study of the acoustical correlates to perceptually determined dimensions of voice quality succeeded in finding acoustic variables corresponding to the perceptual factors of instability and coarseness, qualities characteristic of organic dysphonia, whereas the "functional" dimensions of sonority and strain were not satisfactorily explained by their acoustic data.

0095–4470/86/030477 + 05 $03.00/0

It is a fact that, in clinical practice, decisions concerning the diagnostics and therapy of patients with "functional" voice disorders essentially are founded on subjective evaluations of vocal function. In such evaluations of the voice quality in speech samples, the intersubjectivity or consensus among experienced voice therapists has been shown to be surprisingly high (Hammarberg, Fritzell, Gauffin, Sundberg & Wedin, 1980, evaluation of quality; Kitzing, 1979, evaluation of pitch). The common method to accomplish evaluations of voice function is to imitate more or less audibly the voice sample in question and thus to experience it by one's own vocal apparatus. The method is known as "functional listening" in the German literature and as "creative listening" in the American. Usually, voice therapists can also point out to a patient what he does wrong with his voice by mimicking his dysfunctional voice quality. Thus a considerable capacity to produce different, recognizable kinds of vocal dysfunction can be expected from experienced voice therapists.

Therefore, in the present investigation, aimed to test different LTAS measurements applicable to the clinical evaluation of disturbed voice function it was found adequate to use speech samples produced by a group of experienced voice therapists. Traditionally, comparisons between estimations of voice quality and acoustical measurements are based on voice material from pathological cases evaluated by a panel of listeners. However, as is pointed out by Hurme & Sonninen (1985; this issue) in their description of the observational levels of voice quality within the framework of the speech chain, a similar element of subjectivity enters not only at the level of perception (of others' voices) but also in the production (of one's own voice). Furthermore, in an earlier investigation by this author on vocal strain, the correlation between voice experts' intention to produce certain degrees of straining and the voice quality evaluation by an expert group of listeners with good inter-rater reliability showed to be high [$r = 0.95$ for a female voice and 0.93 for a male voice (Kitzing, 1983)].

2. Procedure

Ten experienced voice therapists (eight females, two males) aged 32–54 (average 41) years, with healthy vocal function read a continuous text (*Nordanvinden och solen . . .*; 99 words; about 40 s) under standardized conditions in a sound treated studio. The subjects were instructed to read the passage four times with different voice qualities: (1) normal quality, with optimal sonority, (2) leaky quality, (3) strained quality, and (4) normal quality, but soft intensity.

Calibrated tape recordings were made with high quality equipment: AKG CE 10 microphone (20 cm distance to the mouth controlled by a head-attachment) and an A 700 Revox tape recorder (tape speed $7\frac{1}{2}$ in/s). On the second track of the tape the signal from an electroglottograph was recorded.

The recorded signals were analysed in the following ways:

(1) Mean and distribution of the fundamental frequency by means of a Glottal Frequency Analyser, GFA (Model 05, Teltec Company, Lund/Sweden).

(2) Total intensity level by means of an $L_{eq.}$-analysis integrating over the entire reading passage (Brüel & Kjaer, Sound Level Meter, 2218).

(3) Long Time Average Spectrographs by means of a Brüel & Kjaer Signal Analyser 2033, analysing 400 frequency lines in the chosen range, using linear averaging over 128 triggered spectra by flat weighting.

The analyses were accomplished in two series with different baseband frequency ranges, viz. 0–5 kHz and 0–2 kHz, respectively. In order to avoid noise from consonant articulation, unvoiced parts of the signal were eliminated by a gate, controlled either by the EGG-signal or by a low-pass filter device.

3. Determination of LTAS-criteria

The signal analyser was attached to a personal computer (ABC 80, Luxor/Sweden). Using a special program (Computer Aided Recording and Analysis of Human Voice, CARAHV.—Patrik Lorentzon, Dept. of Medical Engineering, General Hospital, Malmö), the computer could store the LTA-spectra and the criteria (1)–(3) mentioned below. From these data the criteria (4)–(11) below were computed and printed out:

(1) F_0 and L_0, fundamental frequency: manual determination of frequency and level of peak in F_0 region by means of a cursor in the signal analyser, storage by the computer.

(2) F_1 and L_1, first formant: manual determination, storage by computer.

(3) F_{min}, minimum level between F_0 and F_1 regions: manual determination, storage by computer.

(4) F_{max} and L_{max}, peak level (mode) of entire spectrum.

(5) F_0/F_1, quotient between F_0 and F_1.

(6) $< 1/> 1$ kHz, quotient between area under the spectrum in the frequency range from 0 to 1 kHz and the area above 1 kHz.

(7) 0.3–0.8/1.5–3.0 (2.0) kHz, quotient between area within frequency range from 300 to 800 Hz and 1.5 to 3.0 kHz (2.0 kHz in second series of analyses).

(8) $F_0 \pm 50/F_{min} + 400$ Hz, quotient between a range of 100 Hz around the F_0 line and a 400 Hz range starting from F_{min}. Established only for the second (0–2 kHz) series of analyses.

(9) Median from zero, frequency line dividing the area under the entire spectrum into two equal parts, the line being defined as percentage of entire bandwidth (400 lines) starting from 0 Hz.

(10) Median from F_{min}, frequency line dividing the area from F_{min} upwards into two equal parts, defined as percentage of total bandwidth.

(11) F_{min} of normalized range, same as (10), but defining the line as percentage of the range from F_{min} upwards.

4. Results and discussion

The average results are summarized in Table I. A more detailed discussion will be found in the original extended version which may be requested from the author.

4.1. *Fundamental frequency*

In normal sonorous voice quality the mean fundamental frequency as measured by GFA was 186 Hz as an average for the female voices and 114 Hz for the two male voices. This is in accordance with earlier findings in normal Swedish subjects (Kitzing, 1979).

4.2. *Intensity level*

The average total intensity of the normal sonorous voices was 73 dB ($L_{eq.}$ value as measured by the Brüel & Kjaer 2218 integrating sound level meter). During leaky voice

TABLE I. Average results ($n = 10$)

	Normal sonorous voice (female)	% Difference from normal sonorous		
		Leaky	Strained	Soft
F_0 measured by GFA, mean (Hz)	186/114	+5	+13	−13
F_0 range, ±1 SD (semitones)	6.0	−1.1	−1.6	−0.9
$L_{eq.}$ of speech sample (dB)	73	−1	+2**	−5***
L_0 (dB)	101.6	+2.5*	+0.6	−1.1
L_1 (dB)	99.2	−0.1	+4.9***	−3.4*
L_{max} (dB)	99.8	+1.5	+1.7	−1.1
Correlation (r) $L_{eq.}$ vs. L_{max}	0.61	0.87*	0.57	0.62
Correlation (r) $L_{eq.}$ vs. L_1	0.81**	0.95***	0.70*	0.66*
F_1 (Hz)	488	+24	+9	−53
F_{min} (Hz)	312/193	−17	−12	−30
0–1/1–5 kHz ratio	0.292	+0.007*	−0.007	+0.002
0–1/1–2 kHz ratio	1.09	+0.027*	−0.027*	+0.052**
0.3–0.8/1.5–3.0 kHz ratio	0.415	+0.002	−0.011*	+0.003
0.3–0.8/1.5–2.0 kHz ratio	1.21	+0.013	−0.029*	+0.045*
0–5 kHz median (% of range)	46.9	−0.4	+0.3	+0.2
0–2 kHz median (% of range)	46.7	−0.4	+1.1**	−1.0
F_{min}–5 kHz median (% of range)	47.2	−0.1	+0.2	+0.5
F_{min}–2 kHz median (% of range)	46.9	−0.1	0.0	−0.4*
F_{min}–5 kHz median (% of normalized range)	41.8	−1.3	+0.7	+0.4
F_{min}–2 kHz median (% of normalized range)	41.1	−1.4	+2.4*	−3.6**
L_0/L_1 ratio	1.0	+0.04*	−0.03*	+0.06**
0–F_{min}/F_{min}–2 kHz ratio	0.173/0.890	+0.004	−0.015	−0.015
$F_0 \pm 50/F_{min} + 400$ Hz ratio	1046.9	+48.9***	−24.6*	+44.8**

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

there was no difference of intensity compared to normal sonorous voice quality, whereas there was a small but statistically significant average increase of 2 dB ($p < 0.01$) during strained voice quality and a clear decrease of 5 dB ($p < 0.001$) during soft voicing. The $L_{eq.}$ values were positively correlated to the L_1 values of the LTAS analyses. This agrees with reports by Gauffin & Sundberg (1980) who for sung vowels found that the maximum amplitude of the differentiated glottogram (equivalent to the L_1 in the spectrum) predicted the SPL of the resulting vowel with an accuracy of ±3 dB.

4.3. *Measures of the LTAS slope*

From a theoretical standpoint, the psycho-acoustic voice quality of "sonority" as opposed to dull or husky voice should be reflected in the LTAS as an increase of the harmonics, i.e. a decreased steepness of the overall slope of the spectrum. Indeed, Hammarberg *et al.* (1980) showed that the LTAS-correlates of "breathy" and "hypo-functional" voice quality besides noise components in the region above 5 kHz consist of a high amplitude of the fundamental and a marked drop of energy of the harmonics above the F_1 region, which means a steeper slope of the entire spectrum.

One method to quantify this change in the LTAS slope has been suggested by Frøkjaer-Jensen & Prytz (1976), who related the overall intensity above 1.000 Hz to the intensity below 1.000 Hz, calling this parameter alpha. In the present investigation the ratio between the area below and above 1 kHz has been calculated, i.e. an inverted alpha. When based on 0–5 kHz spectra, the computations resulted in a slightly significant difference only between the groups of sonorous and leaky voices. The inadequate discrimination from the other voice qualities was felt to be due to non-systematic variations of energy in the 3–4 kHz range of the spectra. This frequency range is reported to be influenced by factors of resonance in the vocal tract ("singing formant" according to Blomberg & Elenius, 1970), rather than by the source signal depending on laryngeal function. Therefore, the voice samples were analysed in a second series of LTAS with a base band frequency range of only 0–2 kHz. Now by the criterion of a quotient between the area below and above 1 kHz, the leaky, strained and soft voice samples differed significantly from the normal sonorous. However, the difference was only of a magnitude of 2% for the leaky and strained voices. Only for the soft voices there was a 5% difference. As mentioned earlier, the overall level of the soft voices ($L_{eq.}$) was 5 dB lower than that of the sonorous voices.

A different way to define the slope of the LTAS would be to compute a measure of central tendency of the frequency distribution like the median or the mean. The location of these measures in the frequency range of the LTAS could be stated as the percentage of the range, starting from its lowest frequency. A percentage of less than 50 would thus mean a distance shorter than half the frequency range and the smaller the percentage, the steeper the slope of the spectrum.

Indeed, the median computation applied to the 0–2 kHz spectra indicated a steeper slope in the less sonorous leaky and soft voices, but only the latter statistically (slightly) significant, and a clearly significant decrease of slope steepness for the strained voices. Again however, the magnitude of the differences was only about 2%.

Other methods to define the spectral slope inclination would be to compute the linear regression or to determine the decrease of energy at higher frequencies in terms of dB/octave.

4.4. *Measures of the fundamental relative the first formant*

Investigations of the spectral correlates to physiological function of phonation in singers have been published by Gauffin & Sundberg (1979, 1980). They have demonstrated that the amplitude of the fundamental in the spectrum correlates with the vibratory amplitude of the glottis and that the amplitude of the differentiated glottogram, which is an expression of the velocity of vibratory glottal closure, correlates with the SPL and thereby with the amplitude of the harmonics close to the first formant in normal phonation.

From these findings, hypothetically, it could be inferred that one may expect to find a relatively high L_0 and a low L_1 in leaky voice quality, and conversely a relatively low L_0 and a high L_1 in strained phonation, whereas the optimum of normal phonation should be characterized by a balance between L_0 and L_1. To test this hypothesis, the L_0/L_1 ratio was computed. As can be seen in Table I, the results showed the expected tendencies.

5. Conclusions

In the present investigation, four different qualities of voicing during continuous speech have been compared by means of long time average spectrograms from a commercially available narrow band analyser (Brüel & Kjaer 2033). From the results it can be concluded, that—even in the absence of gross organic anomalies of the larynx like laryngeal paresis and cancer, which have been in focus in many earlier LTAS studies —it is possible to demonstrate statistically significant differences with respect to the psychoacoustic dimensions of sonority and straining. The most potent criteria for such a differentiation have been shown to be (1) the below vs. above 1 kHz ratio, (2) a measure of spectral slope inclination in the first formant range and (3) the ratio between the peak level of the fundamental and the first formant region.

However, even if significant, the differences are only in the order of 3–6%, in spite of the fact that the differences of voice quality in the samples were very conspicuous. Furthermore, as noted by earlier authors, the quality of a voice and thereby the configuration of an LTAS depends highly on the overall intensity. For clinical applications of LTAS analysis in phoniatrics and voice therapy it is therefore recommended to keep the overall voice intensity under control. Useful results can primarily be expected from intra-individual comparisons, as the inter-individual variation between LTA-spectra is substantial.

References

Blomberg, M. & Elenius, K. (1970) Statistical analysis of speech signals, *Speech Transmission Laboratory —Quarterly Progress and Status Report* (STL-QPSR), No. 4, 1–8.

Dejonckere, P. H. (1983) Recognition of hoarseness by means of LTAS, *International Journal of Rehabilitation Research*, **6**(3), 343–345.

Frøkjaer-Jensen, B. & Prytz, S. (1974) Evaluation of speech disorders by means of long-time-average-spectra, *Annual Report of the Institute of Phonetics, University, Copenhagen* (ARIPUC), **8**, 227–237.

Frøkjaer-Jensen, B. & Prytz, S. (1976) Registration of voice quality, *Brüel & Kjaer, Technical Review*, **3**, 3–17.

Gauffin, J. & Sundberg, J. (1980) Data on the glottal voice source behavior in vowel production, *Speech Transmission Laboratory—Quarterly Progress and Status Report* (STL-QPST), Nos 2–3, 61–70.

Gerull, G., Giesen, M., Hippel, K., Mrowinski, D. & Schweers, H. (1977) Untersuchung verschiedener Gruppen von Stimmstörungen durch statistische Spektralanalyse, *HNO*, **25**, 14–22.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J. & Wedin, L. (1980) Perceptual and acoustic correlates of abnormal voice qualities, *Acta Otolaryngologica*, **90**, 441–451.

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.

Hiraoka, N., Kitazoe, Y., Ueta, H., Tanaka, S. & Tanabe, M. (1984) Harmonic-intensity analysis of normal and hoarse voices, *Journal of the Acoustic Society of America*, **76**(6), 1648–1651.

Hurme, P. (1980) Auto-monitored speech level and average speech spectrum. *Papers in Speech Research* (University of Jyväskylä), **2**, 121–127.

Hurme, P. & Sonninen, A. (1985) Normal and disordered voice quality: listening tests and long-term spectrum analyses. In *Publications of the Department of Communication*, vol. 1. Sweden: University of Jyväskylä.

Kitzing, P. (1979) Glottografisk frekvensindikering. Thesis, University Lund, Malmö.

Kitzing, P. (1983) Simultaneous photo- and electroglottographic measurements of voice strain. In *Vocal fold physiology* (I. Titze & R. Scherer), pp. 221–229. Denver: Denver Center of Performing Arts.

Sonninen, A. & Hurme, P. (1982) Clinical and acoustic observations of normal and hoarse voices. *Papers in Speech Research* (University of Jyväskylä), **4**, 1–16.

Sundberg, J. & Gauffin, J. (1979) Waveforms and spectrum of the glottal voice source, pp. 301–320 in *Frontiers of Speech Communication Research*, Festschrift for Gunnar Fant (B. Lindblom & S. Öhman, editors), pp. 301–320. London: Academic Press.