Acoustic and aerodynamic correlates in paralytic dysphonia

Dana M. Hartl, M.D.¹, Stéphane Hans, M.D. ¹, Jaqueline Vaissière, PhD. ², Daniel F. Brasnu, M.D. ¹

¹ Voice, Biomaterials and Head and Neck Oncology Research Laboratory, CNRS-UPRESA 7018, University Paris V, Laënnec Hospital, 42 rue de Sèvres, 75007 Paris, France.

² CNRS UPRESA 7018, Institut de Linguistique et de Phonétique Générales et Appliquées, Université de la Sorbonne Nouvelle-Paris III, 19 rue des Bernardins, 75005 Paris, France.

Corresponding author : Daniel F. Brasnu, M.D. Chief of Department. Department of Otorhinolaryngology, Head and Neck Surgery. Laënnec Hospital. 42, rue de Sèvres. 75007 Paris. France. Tel : 33-1-44-39-66-58  Fax : 33-1-44-39-66-46  E-mail: DBrasnu@aol.com
Abstract

The aim of this study was to analyze the correlations between objective acoustic and aerodynamic parameters available in clinical practice and perceptual characteristics in a group of 8 subjects with unilateral vocal fold paralysis (UVFP). Twelve subjects with normal larynx of the same age distribution and smoking history were studied as controls. Values from 5 normal non-smoking subjects were obtained for comparison with the control group. Only male subjects were included in this prospective study. Acoustic parameters were measured for the vowel /a/ at comfortable frequency and intensity: jitter, shimmer, harmonics-to-noise ratio (H/N), cepstral peak prominence (CPP), the difference between the levels of the first two harmonics (H1-H2), and the relative energy above 6 kHz (HFR). Aerodynamic parameters included the mean flow rate during a maximally sustained /a/ (MFRs) and during a 3-second /a/ (MFRc), at comfortable frequency and intensity levels. Intraoral pressure was measured during the phoneme /pi:/ to approximate phonatory subglottic pressure (Ps). Six judges rated a mid-/a/ sample using a 5-parameter scale with 4 levels of severity.

Intra-judge reliability averaged 84%. Nonparametric statistical analysis revealed significant differences (p<.05) between the UVFP group and the control group, for all of the subjective and objective parameters studied except H1-H2 and H/N. Normal smokers and non-smokers only differed significantly in intraoral pressure. Correlation between objective parameters and perceptual characteristics differed in the two subject groups. Correlations were as expected, based on previous literature reports, but only for the UVFP group. Breathiness and roughness coexisted and were intercorrelated in UVFP. H1-H2 and airflow measurements seemed to be the most specifically related to breathiness. The other acoustic parameters were affected by both turbulent noise and dysperiodicity.

These objective acoustic and aerodynamic measurements seem to lack specificity in terms of voice quality in a homogeneous cohort of patients with UVFP. Despite the lack of specificity and quantification, these measurements allow provide reproducible and objective qualification of voice in patients with UVFP.

Key words: laryngeal paralysis, voice quality, acoustics, aerodynamics, smoking
Introduction

Dysphonia in unilateral vocal fold paralysis (UVFP) is principally characterized by breathiness, or the perception of turbulent laryngeal airflow. This air leak leads to weak voice and increased phonatory effort. The vibrational asymmetry caused by UVFP can lead to the appearance of subharmonics and the perception of diplophonia or roughness.

Acoustic correlates of breathiness have been studied in dysphonia and in normal subjects. Increased high-frequency spectral energy (Shoji et al., 1992), low harmonics-to-noise ratio, and a decrease in cepstral peak prominence (Hillenbrand et al., 1994) have been shown to correlate with breathiness. The incomplete closing phase in the glottic cycle has been shown to increase the intensity of the first harmonic (or the fundamental frequency) as compared to the second harmonic, and to contribute to the perception of breathiness (Klatt and Klatt, 1990). The perturbation measurements of jitter and shimmer have been correlated with perceived roughness (Imaizumi, 1986) and shown to increase in UVFP (Kim et al., 1982). Increased phonatory air flow have been shown in patients with UVFP (Tanaka et al., 1991).

The aim of this study was to evaluate the correlation between objective acoustic and aerodynamic parameters available in clinical practice and the subjective voice quality in patients with UVFP. As our patients were heavy smokers, and smoking may be related to dysphonia, we compared them to a group of smoking controls.

Materials and methods

The dysphonic group consisted of eight patients presenting with UVFP. Etiologies were as follows: thoracic surgery for malignant neoplasm (n=5), malignant thoracic neoplasm (n=1), cervical malignancy (n=1), and blunt cervical trauma (n=1). Average age was 61 years (range: 38-78). Average cigarette consumption was 35 pack years (range: 0-70). UVFP was confirmed using fiberoptic videolaryngoscopy (Crumley, 1994). The laryngeal mucosa was normal, and there was no evidence of Reinke's edema. The average time lapse between onset of UVFP and voice recording was 6 weeks (range: 0.5 to 19 weeks), but the delay was one week or less in 4 of the 8 patients.

Twelve heavy smokers, presenting with pulmonary carcinoma, without UVFP, comprised the control group "smokers". Average age was 62 years (range: 49-75). Average cigarette consumption was 53 pack years (range: 16-165). Five non-smoking healthy subjects (hospital personnel) were concurrently studied for comparison with the control group. Their average age was 35 years (range: 28-49). For reasons related to pitch and harmonic detection, only male patients were included in this study. Fiberoptic laryngoscopy confirmed the normalcy of the larynx in the smokers and non-smokers. No edema, keratosis or supraglottic hyperfunction were noted.

Acoustic recordings were performed in a quiet room on a SONY DTC-60ES Digital Audio Tape (DAT) deck with a sampling rate of 48 kHz. Mouth to microphone distance was 15 cm. Subjects were asked to phonate the vowel /a/ for approximately 3 seconds, at comfortable pitch and intensity. Three tokens of the vowel were recorded.

Perceptual testing

Subjective voice quality was evaluated by 6 judges: 3 speech therapists and 3 otolaryngologists specialized in voice disorders. For each subject, a 1000 ms mid-vowel /a/, at
least 50 ms from voicing onset and 50 ms from offset, was resampled at 16 kHz and numerized at 16 bits. Tokens were presented individually to each judge using a personal computer (Compaq Armada 5300) and headset. The characteristics of the grade of dysphonia (G), roughness (R), breathiness (B), asthenia (A) and strainedness (S) were each rated on a scale of 0 (absent) to 3 (strongly present) (Imaizumi, 1986). Fifteen tokens were duplicated to determine intra-judge reliability.

**Objective acoustic parameters**

For the acoustic measurements, the three vowel samples were introduced by direct connection into a Sonograph model 4300B (Kay Elemetrics). After resampling at 50 kHz, jitter (in percent), shimmer (in decibels) and harmonic-to-noise ratio (H/N) were analyzed for a 1.75 second mid-/a/ segment using the Multidimensional Voice Program (MDVP version 1.34, Kay Elemetrics, 1993). The results from the sample with the highest H/N were retained. Cepstral peak prominence (CPP) was calculated from one mid-vowel point of the token with the highest H/N. The cepstrum was calculated automatically using the Computer Speech Lab (CSL version 5.05, Kay Elemetrics), without preemphasis. After resampling at 20 kHz, data were obtained every .05 ms, from .5 to 20 ms. Calculation of CPP was performed as described by Hillenbrand et al. (1994). CPP was defined as the height of the principal cepstral peak (in decibels), as compared to the regression line through the remaining points.

The difference in intensity between the fundamental and the second harmonic (H1-H2) was calculated by averaging the values obtained from 9 different points. Three points were taken at 200 ms intervals from a 1000 ms mid-vowel portion of each of the three tokens of the vowel /a/. Levels were recorded manually from a 1024-point Fourier transform spectrum as calculated by the Computer Speech Lab (CSL version 5.05, Kay Elemetrics), using a Hamming window, without preemphasis.

An average spectrum was calculated from a 1000 ms mid-vowel portion of the token with the highest H/N. From this average spectrum, the average intensity in the 6 to 10 kHz range was divided by the total average intensity from 0 to 10 kHz, to obtain a "high-frequency ratio" (HFR) (Shoji et al., 1992).

**Aerodynamic parameters**

Laryngeal aerodynamics were evaluated using the Aerophone II (Kay Elemetrics). Data were available for all 8 patients with UVFP and for 9 subjects from the smoking control group. Mean flow rate during the phonation of /a/ at comfortable pitch and intensity, for approximately 3 seconds (MFRc), and the mean flow rate during a maximally sustained voicing of the vowel /a/ (MFRs) were recorded. Each procedure was repeated 3 times and the average was retained for final analysis. Intraoral pressure was measured during the phonation of the phoneme /pi:/ as described by Smitheran and Hixon (1981). The maximum pressure during /p/ was averaged over 5 tokens of the phoneme. This approximates the subglottic pressure in comfortable phonation, and was referred to as "Ps." Results were compared with normative values obtained from 48 normal French-speaking males, using the same apparatus and the same testing protocol (Hans, in press).

Average values for each parameter were calculated within each group (UVFP, smokers and non-smokers). Non-parametric statistical analysis (Mann-Whitney U-test) was performed between each pair of subject groups. The correlation coefficient (Spearman’s non-parametric correlation) was then calculated between each perceptual trait and each objective parameter.
Strong correlations (rho>.50) and trends were compared with results from other literature reports.

**Results**

Table 1 shows the values obtained for the parameters analyzed. Perceptual, acoustic and aerodynamic parameters were significantly different between patients with UVFP and the smoking group (Mann Whitney U-test, p<.05) with the exception of the parameters H1-H2 and H/N. The age and tobacco consumption of the groups UVFP and smokers were not significantly different (Mann Whitney U-test, p>.05). There was no statistical difference between the smokers and the group of 5 non-smokers, concerning the acoustic and perceptual data (Mann-Whitney U-test, p>.05).

**Perceptual ratings**

Intra-listener reliability averaged 84% (Spearman's correlation coefficient). All of the voice characteristics, GRBAS, were significantly more severe in the UVFP group as compared to the smoking control group (Mann-Whitney U-test, p<.01).

**Objective acoustic parameters**

Jitter and shimmer were significantly higher in the UVFP group. The difference in H/N was not statistically significant. No significant difference was found between the UVFP group and the smokers for the parameter H1-H2. CPP was significantly lower, and HFR significantly higher in the UVFP group, as compared to the smoking control group.

**Aerodynamic parameters**

Airflow during sustained phonation (MFRs) and during comfortable phonation (MFRc) in patients with UVFP surpassed the values for the smokers by more than two standard deviations. This difference was statistically significant (Mann-Whitney U-test, p<.01). The aerodynamic data for the smokers was compared to normative data, measured in the same conditions, on the same machine (Hans, in press). Flow values were not statistically different between the groups (non-parametric sign test, p<.01). However, Ps was significantly smaller in smokers (non parametric sign test, p>.05).

**Correlations**

Non-parametric Spearman’s correlation coefficients are shown in tables 2 and 3. Correlations varied between the two groups, both in direction and amplitude. Correlations having a coefficient of .50 or greater were generally retained as pertinent, and general trends were noted.

Among the perceptual voice quality parameters, G was strongly related to both R and B for the UVFP group. G was strongly related to R, B and A, for the smoking group. B and A were strongly related in the smoking group, and less strongly related in the UVFP group. R and B were strongly related in both groups.

Jitter correlated with G and R for the UVFP group. Shimmer correlated with G, R and B for the UVFP group. H/N correlated most strongly with R, for the UVFP group. CPP was strongly and inversely related to A in the smoking group. The correlation was similar but less strong for the UVFP group. For H1-H2, in the UVFP group, a negative correlation was noted.
with G, R and S, and a weak but positive correlation with B and A. In the smoking control group, H1-H2 was positively and moderately correlated to all of the perceptual traits. A strong correlation for HFR with R and G was found only for the UVFP group.

For the UVFP group, B and A were strongly related to airflow rates. A was inversely related to Ps. For the smokers, the flow parameters were inversely related to R.

No one objective parameter was strongly correlated in both groups to any one perceptual voice quality characteristic.

**Discussion**

Except for H1-H2 and H/N, the objective acoustic and aerodynamic parameters significantly distinguished the UVFP group from the age- and tobacco-matched group with normal larynx. There was no difference between the smoking group and the group of 5 non-smokers concerning the acoustic and perceptual parameters. We can thus safely say that the acoustic differences between the smoking and the UVFP group was due to the UVFP, and not to smoking history.

**Perceptual ratings**

The correlation tendencies were different in the two groups. Globally, the correlations in the UVFP group corresponded with expected results previously reported in the literature, whereas the correlations in the smoking group were less logical and more difficult to explain. This difference implies that, perhaps, voice quality is not perceived in a linear fashion, or that “dysphonic breathiness” is not perceived in the same fashion as “euphonic breathiness,” for example. If this is the case, it would be logical to analyze dysphonic voices separately from normal voices, in further studies. Perhaps even distinguishing different voice pathologies for separate perceptual, acoustic and aerodynamic studies would be warranted.

No one objective acoustic measurement was specifically correlated to any one vocal quality characteristic. G and B had high correlation coefficients, implying the breathiness was a major contributor to dysphonia in these groups. But the correlation between G and R was just as high. It is not clear to what point B and R interact perceptually (deKrom, 1995). Breathiness and roughness are not mutually exclusive voice quality characteristics. There may be interaction on the perceptual level, as reported by Kreiman et al. (1994). Breathiness can contribute to increased roughness scores, and vice-versa. This implies that the human ear's capability of distinguishing turbulence noise and dysperiodicity is limited. On the objective acoustic level, many parameters may be measuring both noise and dysperiodicity at the same time. For example, CPP is lowered by both phenomena. Likewise, measurements based on the detection of the fundamental frequency (such as jitter and shimmer) have been shown to encompass both noise and dysperiodicity (Murphy, 1999). Finally, from a purely statistical point of view, the multicollinearity of B and R in our study may contribute to the lack of specificity of the objective measurements. More powerful statistic analyses, such as factor analysis or principal component analysis, were not permitted, given the limited cohort.

**Objective acoustic parameters**

Jitter, shimmer and H/N correlated significantly with each other. This is not surprising, in that presently used algorithms cannot distinguish additive turbulent "white" noise in the signal from the "noise" caused by dysperiodicity (Murphy, 1999; Ezkenazi et al., 1990). In addition,
as the severity of dysphonia increases, the correlation between perturbation as measured by current algorithms and perceptual ratings decreases (Rabinov et al., 1995). In patients with UVFP, jitter, shimmer and H/N were frankly pathologic, whereas the values for the smokers and the non-smokers were within the range considered normal (MDVP Operations Manual; Woodson and Cannito, 1998). The H/N was not significantly different between the UVFP and smoking groups. This could be due to errors in harmonic detection, to the small number of patients and/or to the large standard deviation in the UVFP group.

Normative values for CPP, H1-H2 and HFR have not been established. However, the high values for H1-H2 and HFR, and the low value for CPP correlate with trends reported in the literature concerning breathy dysphonia (Hillenbrand and Houde, 1996; Shoji et al. 1992; Hammarberg et al., 1986; Stevens and Hanson, 1995).

CPP has been correlated with breathiness in normal (Hillenbrand, 1994) and dysphonic voices (Hillenbrand, 1996). In theory, CPP measures harmonic regularity. The more regular the spacing of the harmonics (i.e., the more periodic the signal) the higher the cepstral peak. CPP is adversely affected by spectral noise. The peak is proportional to the relative amplitude of the harmonics. When noise "drowns out" the harmonics, the CPP is lowered. Likewise, when subharmonics lower the harmonic intensities, CPP is lowered. Thus CPP is affected by both turbulent noise and by dysperiodicity, explaining our finding a correlation with roughness, as well as with breathiness, in both subject groups.

An increase in H1-H2 has been correlated with breathy phonation in normal voices (Klatt nad Klatt, 1990; Ladefoged et al., 1988; Stevens and Hanson, 1995). Furthermore, an increase in H1 as compared to H2 was shown to correlate with a hypofunctional or asthenic breathy voice, whereas a decrease in H1-H2 corresponded to a hyperfunctional, tense or strained and breathy voice quality (Hammarberg et al., 1986). H1-H2 is in theory related to the closing phase of the glottal cycle (Ni Chasaide and Gobl, 1997). A slow or incomplete closing phase has been shown to increase the ratio H1-H2 in synthetic voice, and to increase the perceived breathiness (Klatt and Klatt, 1990).

In our study, statistical analysis failed to reveal a significant difference between the UVFP group and the smoking control group. The value H1-H2 tended to increase in the UVFP group, however. There are several possible explanations for these observations. A simply statistical explanation is that the small number of subjects and the large standard deviation weakened the statistical tests. Several acoustic phenomena may have also contributed to our observations. An increased F1 bandwidth from tracheal coupling could contribute to an increased H2 level. H1-H2 would in this case no longer reflect exclusively the vocal source characteristics, but also the vocal tract resonances. Stevens (1995) employed a “corrected H1-H2” parameter, correcting H2 for the “boosting” from the first formant. Such a calculation is possible in an all-pole model. But in severely breathy voice, the tracheal coupling introduces zeroes and an increased first formant bandwidth that render further “corrections” too subjective. Subharmonics or low-frequency noise may also have contributed to an “artificial” lowering of the first two harmonics, altering their relative amplitudes. Finally, a certain degree of involuntary, compensatory vocal tract constriction may occur secondary to the glottic incompetence of UVFP (Van Doersten et al., 1992). The resonant effects of these adjustments are unknown. Involuntary vocal tract constriction secondary to the glottic incompetence of UVFP may have lowered the first formant frequency just enough to indirectly increase the relative intensity of H2, and "artificially" lower H1/H2.
We expected to find a correlation between HFR and breathiness or asthenia, at least in the UVFP group. The higher correlation observed with roughness may again be a manifestation of the "multidimensionality" of voice quality perception (Kreiman et al., 1994). The high-frequency noise was perceived as roughness, rather than breathiness. It could also mean that the HFR does not exclusively measure turbulent noise.

**Aerodynamic measurements**

The lower than "normal" Ps in smokers is difficult to explain, in that the measurements were obtained under identical conditions. It is possible, however, that the average sound pressure level (SPL) during the utterance of /pi:/ was lower for the smoking group. Subglottic pressure, and thus intraoral pressure, is directly related to SPL (Holmberg et al., 1988). This parameter (SPL) was visually controlled on the computer screen during measurement for all patients. The visual approximation of SPL was subjective, and a certain degree of error was possible.

A limited number of reports in the literature have evaluated the aerodynamics of phonation in UVFP. The normal distribution of Ps in the UVFP group had already been reported (Kitajima and Tanaka, 1993). It has been shown that airflow and phonation threshold pressure increase in UVFP (Kitajima and Tanaka, 1993; Makiyama et al., 1998). However, the effects of turbulence on flow and pressure transducers have not been taken into account. Could high airflow, as in the case of UVFP, provoke artifacts in flow or pressure measurements? For the approximation of subglottal pressure by intraoral pressure, one must wonder if the aerodynamics and laryngeal mechanisms of devoicing are the same in UVFP as in normal subjects. High airflow during vowel voicing may interfere with intervocalic intraoral pressure measurements. Supraglottic vocal tract adjustments, such as increased supraglottic laryngeal or pharyngeal muscle tone may also affect aerodynamic measurements differently in different UVFP patients. The limits of automatic aerodynamic measurements in UVFP still need to be explored (Kitajima and Tanaka, 1993).

The high positive correlation between breathy (B) and asthenic (A) voice qualities and phonatory airflow was as expected, in the UVFP group. The negative correlation between flow and strainedness (S) was also as expected. For the smoking group, a high negative correlation was found between airflow and roughness (R), and to a lesser extent strainedness, possibly corresponding to a hyperfunctional vocal mechanism in these subjects. No supraglottic hyperfunction was evident upon fiberoptic laryngoscopy, however. In all cases, the same trends were observed for the flow rate in comfortable phonation (MFRc) and in sustained phonation (MFRs). The different phonatory tasks were not individually distinctive for any particular voice quality.

**Comments**

Our study seems to raise more questions than it answers. The neurological and psychological mechanisms of the perception of voice quality remain poorly understood (deKrom, 1995). The GRBAS rating system does not necessarily reflect anything measurable from the brute spectrums of the voice signal (Rabinov et al., 1995). One then should not be surprised to find that the objective and subjective correlations are weak. It may be beneficial, on a practical level, to define and measure objective acoustic and aerodynamic parameters, independently of perceptual voice rating. Both approaches are complementary. Perceptual rating is considered the "gold standard" in terms of voice quality. But it is subjective, difficult to reproduce and difficult to quantify. Acoustic and aerodynamic measurements are not as sensitive or specific as the ear, but are easily reproducible and quantifiable.
Turbulent noise, subharmonics and dysperiodicity all contribute to the decrease in harmonic prominence in paralytic dysphonia. This leads to difficulties in measuring, even by hand, certain parameters such as H1-H2, jitter, shimmer and H/N. These phenomena in the signal lead to errors in current algorithms for detecting fundamental frequency, jitter and shimmer. Reliability of current algorithms for jitter decreases as the degree of dysphonia increases, for example (Rabinov et al., 1995; Woodson and Cannito, 1998). It seems thus that current “automatic” calculations of certain parameters are not adapted to pathologic voice.

Our findings are in agreement with previous reports correlating low CPP (Hillenbrand and Houde, 1996) and high H1-H2 (Ladefoged et al., 1988; Stevens and Hanson, 1995; Hammarberg et al., 1986) to breathiness, and jitter to roughness (Imaizumi, 1986). This was only true, however, within the group of patients with UVFP. In addition, the correlation coefficients were lower than expected. Few reports have been limited to UVFP, and often include different organic and functional dysphonia. Some previous studies only included normal subjects mimicking a "breathy" voice (Hillenbrand et al., 1994). Not all UVFP have the same dysphonia, due to the etiology of the UVFP, its duration, residual innervation or synkinesis (Crumley, 1994; Woodson, 1996). Our study was comprised mostly of patients with severe denervation from sugery or neoplasms. The acoustic and aerodynamic recordings were performed soon after the onset of UVFP, precluding, in theory, the onset of synkinesis, which can take place after several weeks or months. It may be that the laryngeal denervation in our group of patients was more severe than in other literature reports. In any case, our study is one of the few to analyse a homogeneous group of patients with the same organic dysphonia. Because voice quality is multidimensional, different types of breathy dysphonia, for example, may be perceived differently according to the perceptual characteristics of the voice (Kreiman et al., 1994). It may be more judicious in future studies to define different groups of dysphonia according to the etiology or mechanism, rather than to group all in one.

Conclusions

Breathiness, the perception of turbulent laryngeal airflow, is considered to be a principal characteristic of paralytic dysphonia. Acoustic parameters related to breathiness (H1-H2, CPP, HFR and H/N) were more strongly related to perceived breathiness in the group of patients with UVFP than in the smoking control group. Laryngeal airflow (MFRc and MFRs) correlated with perceived breathiness in the UVFP group, and inversely with perceived roughness in the smoking group. These measurements lacked specificity in distinguishing breathiness from roughness. The measurements, however, allowed to distinguish the UVFP group from the controls. Measurements were not significantly different between smokers and non-smokers with normal larynx. Correlations between objective and subjective parameters were different in the two subject groups studied. Voice quality parameters (GRBAS) may be perceived differently in dysphonic voice as compared to normal voice.

The objective measurements currently at our disposal lack specificity in distinguishing between the phenomena of turbulent laryngeal air flow and of laryngeal waveform irregularities. As measures of general tendency, in intra-individual comparisons, these measures retain their advantage of objectivity and reproductibility, as compared to perceptual voice judgements. Subjective and objective, acoustic and aerodynamic, measurements continue to be complementary, as no one objective parameter perfectly replaces any one voice quality characteristic.

For the spectral and aerodynamic measurements to be useful in current practice, it will be necessary to define standards for voice recording (frequency and intensity), signal processing
(filtering and sampling rates) and parameter analyzing. Until then, normative values, will be lacking, and a quantitative use of these parameters unjustified. A final problem is that of analyzing pathologic voice, as opposed to normal voice, with the technical problems of noise and dysperiodicity.

Acknowledgement

The authors thank Dr. Paul Landais for the generous contribution of his time and statistical expertise.
References


Hans, S., (in press);


