Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 1999)
Firenze, Italy, September 1-3, 1999

ANALYSIS OF NEUROLOGICAL DYSPHONIAS: METHODOLOGICAL CONSIDERATIONS

Ingo Hertrich & Hermann Ackermann
Department of Neurology, University of Tübingen, Hoppe-Seyler-Str. 3, D-72076 Tübingen, Germany
E-mail: ingo.hertrich@uni-tuebingen.de

Abstract

This paper provides a short introduction to the various methods available for the analysis of the voice of dysphonic patients suffering from neurological disorders. Both auditory-perceptual ratings as well as conventional acoustic perturbation measures are of limited relevance to the specific characterization of patient groups. However, the addition of further techniques such as electroglottography (EGG) as well as new methods of signal processing such as fractal dimension analysis allows to delineate characteristic profiles of performance to some degree.

Introduction

Neuromotor diseases affecting the various subsystems of the brain such as the basal ganglia or the cerebellum often give rise to voice and speech disorders, and dysphonia may even emerge as an early symptom [1]. Specific evaluation of these voice abnormalities requires elaborated analysis techniques. The present paper summarizes some of the methods that have been used to analyze dysphonic voices with respect to both theoretical motivation and technical performance.

Auditory perceptual evaluation

In terms of auditory-perceptual characteristics, the voices of neurological patients have been labeled as 'high-pitched', 'low-pitched', 'breathy', 'harsh', 'strained', irregular', 'quivering', or even 'aphonic' [2]. Many of these perceptual dimensions, however, are highly intercorrelated representing, therefore, neither statistically independent variables nor providing a salient basis for subgroup classification. For example, a correlation analysis among the six perceptual dimensions 'harsh', 'strained', 'creaky', 'quiver', 'pitch fluctuations', 'loudness fluctuations') was performed using ratings from a study including four groups of neurological patients (Parkinson's disease, Huntington's disease, Friedreich's disease, cerebellar degeneration; see [3]). Sixteen of the 21 computed correlation coeffecients were significant at the level of $p < 0.01$, and diagnostic group separation on the basis of these perceptual data was poor, confirming previous findings [4]. Irrespective of their unspecificity, however, perceptual ratings seem to be superior to objective acoustic parameters with respect to the quantification of overall voice impairment. As concerns the distinction between patients and healthy controls, statistical group separation on the basis of perceptual data was better than on the basis of the parameters of a 'multidimensional' voice program (MDVP program, Kay Elemetrics, Lincoln Park, NJ) [3].

Conventional perturbation measures

Most commonly, three types of acoustic parameters have been applied to assess the various aspects of signal abnormalities in dysphonic voices: (1) jitter, i.e. the temporal irregularity of vocal fold cycle durations, (2) shimmer, i.e. period-to-period irregularity of signal amplitude, and (3) the noise-to-harmonics ratio (NHR), i.e. the proportion of spectral non-harmonic to harmonic energy [5] [6] [7] [8]. However, since the exact methods underlying these measures may vary considerably across studies [9] [10] [11] it is hardly possible to perform statistical comparisons across data sets that are reported in the literature. In particular, perturbation values depend on their definition as relative or absolute values, their smoothing level and correction for global trends of pitch movements (see [12]), and the method and frequency resolution of pitch extraction [13] [14] [15] [16]. The latter aspect raises difficulties for the assessment of severely impaired voices: Automatic F0 computation may result in more or less high values of all perturbation measures, reflecting problems of the pitch algorithm rather than specific properties of the acoustic signal.

A further problem is that these measures are not mathematically independent of each other [17]. For example, if pitch extraction is based on detection of peaks or zero crossings, any noise component arising from aspiration or friction does not only influence the NHR but also increases jitter and shimmer. Similarly, in the presence of high durational period-to-period variability (jitter) the spectrum does not show a sharp harmonic structure, giving rise to higher NHR values as well. Therefore, an improved set of perturbation measures has recently been developed [18], the so-called 'hoarseness diagram' including two parameters that are supposed to be orthogonal to each other. The first one is obtained by jitter analysis measuring the temporal irregularity of glottal source pulses based on an elaborated inverse filtering technique. The more innovative second measure provides the glottal-to-noise excitation ratio (GNE). Contrary to conventional noise-to-harmonics measures, it estimates only that amount of noise that cannot be accounted for by temporal or amplitude-related aperiodicity of the glottal source.

The results obtained with any of the above mentioned perturbation measures are valid only under the assumption that the analyzed voices are quasi-stationary signals representative of a person's voice. Neurological dysphonias, however, are often characterized by discontinuities in terms of abrupt changes in vocal fold activity such as jumps of vocal register. It is obvious that perturbation parameters obtained across such discontinuities do not reflect valid measurements. For example, period-to-period perturbation during vocal fry is considerably higher as compared to modal voice [19].

Also from a theoretical point of view these perturbation measures do not seem to be well-motivated. They measure three principally different aspects of aperiodicity but consider these deviations to be random influences [20]. In fact, however, vocal fold vibrations are quite complex events, and deviation from strict periodicity can be the result of regular and predictable subsystem interactions rather than just 'perturbation'.

Considering all these limitations, it is not surprising that voice perturbation measures have been of limited value with respect to both differential diagnosis of dysphonias [21] [22] and quantitative assessment of normal voice quality [23].

Expanded sets of acoustic voice quality measures

Some approaches have been made to amplify the number of voice parameters in order to account for some aspects of phonation that are not adequately represented by the above three perturbation measures. For example, it has been found that the spectral energy distribution above 6 kHz is particularly sensitive to the perception of breatiness [24] . The commercially available MDVP program claims to provide a multidimensional voice profile by computing more than 20 parameters [25]. They relate to a) various aspects of shimmer and jitter applied to different smoothing levels of the underlying pitch contour, b) aspects of long-term variability such as tremor and the occurrance of subharmonic segments, and c) the spectral energy distribution of harmonic and inharmonic signal components in various frequency regions. With respect to the latter aspects, the 'soft phonation index' (ratio of harmonic energy between 70 and 1600 Hz to harmonic energy between 1600 and 4500 Hz) might explicitly be mentioned here. This parameter turned out to be sensitive to the kind of breathiness occurring in male Parkinsonian patients [3] [26].

Unfortunately, the user's possibilities to adjust the MDVP program are quite limited. When analyzing dysphonic voices, for example, the pitch algorithm often fails to extract a plausible fundamental frequency contour, but no tool has been provided for correction of F0 errors. In particular, the program fails to reliably analyze severely dysphonic voices that are characterized by discontinuities in time. The best way to investigate such cases seems to be a descriptive analysis of the pathological events. The occurrance of subharmonic segments and voice breaks in patients with basal ganglia diseases, for example, has been documented by visual inspection of the speech oscillogram [27] [28] and by semi-quantitative evaluation of spectrograms of the electroglottographic signal (EGG) [29]. The latter method has also been applied to show instances of voice tremor in cerebellar patients [30].

Fractal dimension analysis

Biomechanical models [31] [32] [33] as well as analysis of high-speed films [34] [35] indicate that seemingly irregular phonatory activity may reflect complex vibratory behavior of the laryngeal system rather than random perturbations of an underlying periodic signal. Therefore, methods of chaos analysis have been introduced in order to obtain quantitative parameters of voice signal complexity. While some studies applied these methods to fundamental frequency contours [36], again being dependent on reliable pitch extraction, others used the acoustic signal itself as input [37] [38] [39] [40] [41].

The so-called correlation dimension [42] [43] represents a measure which is derived by embedding the signal in a multidimensional space and estimating the power to which the number of signal points within a small hypercube increases with the edge length of this hypercube. Roughly, this kind of analysis indicates the number of independent oscillating structures required to model the signal. It has been shown, e.g., that laryngeal asymmetry, in terms of desynchronization of the left and right vocal fold, leads to an increase in dimensional complexity [44]. However, since fractal dimension analysis just provides one further parameter of the acoustic signal, it cannot specifically describe all aspects of voice quality in the various kinds of dysphonias. It has been shown, for example, that dysphonic voices do not consistently show higher dimension values than the voices of healthy controls [37] [38].

One general limitation of acoustic voice analysis is that this signal reflects both the laryngeal source as well as vocal tract filtering and that the eigenfrequencies of the vocal tract, being controlled independently of the glottal source signal, may interact with the latter [45]. Therefore, the results of acoustic fractal dimension analysis may depend on vowel type (eg., /a/ versus /i/), whereas signals such as the electroglottogram (EGG), that are 'closer' to the glottal source signal, fail this effect [38]. Therefore, acoustic analysis analysis should be supplemented by EGG recordings.

Gender-specific performance in neurological dysphonias

In general, considerable differences between male and female voices have been described [46] [47]. Furthermore, it has to be taken into accout that neurological disorders may give rise to gender-specific phenomena of voice impairment as documented in studies on dysphonia in Parkinson's disease (PD), Huntington's disease (HD), and cerebellar atrophy (CA) [3] [26] [29] [38]. Male PD voices, for example, sound breathy and have high jitter values as well as a lack of higher harmonic energy. Regarding fractal dimension analysis, their acoustic signal is relatively low-dimensional as compared to the one of age-matched controls. PD females, in contrast, tend toward harsh, strained voice quality rather than breathiness, show long-term instabilities and highly increased fractal dimension values, whereas within segments of continuous phonation their jitter is relatively low. As concerns breathiness and jitter, HD patients show the reverse gender-specific pattern as the one obtained in PD patients: Breathy voice quality concomitant with high jitter values was observed in the female subgroup. Dysphonia caused by cerebellar dysfunction seems to be more severe in females as compared to males.

Conclusion

Voice analysis of neurological dysphonias requires a multidimensional approach including auditory-perceptual ratings as well as various acoustic and physiological measurements. Within some limits, characteristic profiles of the various types of neurogenic dysphonias can be worked out by combining several methods. However, separation of these syndromes on the basis of automatic voice analysis systems is not yet possible. Since, furthermore, neurological dysphonias show gender-specific features to a considerable extent, male and female samples should not be pooled for analysis.

References

[1] Hartmann, D.E. (1984). Neurogenic dysphonia. *Annals of Otology, Rhinology, and Laryngology*, **93**, 57-64.

[2] Darley, F.L., Aronson, A.E., & Brown, J.R. (1975). *Motor Speech Disorders*. Philadelphia: Saunders, 0.

[3] Hertrich, I., Spieker, S., & Ackermann, H. (1998). Gender-specific phonatory dysfunctions in disorders of the basal ganglia and the cerebellum: Acoustic and perceptual characteristics. In W. Ziegler, & K. Deger (Eds.), *Clinical Phonetics and Linguistics* (pp 448-457). London: Whurr.

[4] Zyski, B.J., & Weisiger, B.E. (1987). Identification of dysarthria types based on perceptual analysis. Journal of *Communication Disorders*, **20**, 367-378.

[5] Horii, Y. (1980). Vocal shimmer in sustained phonation. *Journal of Speech and Hearing Research*, **23**, 202-209.

[6] Klingholz, F. (1991). Jitter. *Sprache - Stimme - Gehör*, **15**, 79-85.

[7] Koike, Y. (1973). Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica*, **7**, 17-23.

[8] Yumoto, E., Gould, W.J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, **81**, 1544-1550.

[9] Karnell, M.P., Scherer, R.S., & Fischer, L.B. (1991). Comparison of acoustic voice perturbation measures among three independent voice laboratories. *Journal of Speech and Hearing Research*, **34**, 781-790.

[10] Pinto, N.B., & Titze, I.R. (1990). Unification of perturbation measures in speech signals. *Journal of the Acoustical Society of America*, **87**, 1278-1289.

[11] Rabinow, C.R., Kreiman, J., Gerratt, B.R., & Bielamowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of Jitter. *Journal of Speech and Hearing Research*, **38**, 26-32.

[12] Ackermann, H., & Ziegler, W. (1994). Acoustic analysis of vocal instability in cerebellar dysfunctions. *Annals of Otology, Rhinology, and Laryngology*, **103**, 98-104.

[13] Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. Journal of Speech and Hearing Research, 22, 5-19.

[14] Read, C., Buder, E.H., & Kent, R.D. (1992). Speech analysis systems: An evaluation. *Journal of Speech and Hearing Research*, **35**, 314-332.

[15] Titze, I.R., Horii, Y., & Scherer, R. (1987). Some technical considerations in voice perturbation measurements. *Journal of Speech and Hearing Research*, **30**, 252-60.

[16] Titze, I.R., & Liang, H. (1993). Comparison of Fo extraction methods for high-precision voice perturbation measurements. *Journal of Speech and Hearing Research*, **36**, 1120-1133.

[17] Cox, N.B., Ito, M.R., & Morrison, M.D. (1989). Technical considerations in computation of spectral harmonics-to-noise ratios for sustained vowels. *Journal of Speech and Hearing Research*, **32**, 203-218.

[18] Michaelis, D., Fröhlich, M., & Strube, H.W. (1998). Selection and combination of acoustic features for the description of pathologic voices. *Journal of the Acoustical Society of America*, **103**, 1628-1639.

[19] Horii, Y. (1985). Jitter and shimmer in sustained vocal fry phonation. *Folia Phoniatrica*, **37**, 81-86.

[20] Laver, J., Hiller, S., & Mackenzie Beck, J. (1992). Acoustic waveform perturbations and voice disorders. *Journal of Voice*, **6**, 115-126.

[21] Feijoo, S., & Hernández, C. (1990). Short-term stability measures for the evaluation of vocal quality. *Journal of Speech and Hearing Research*, **33**, 324-334.

[22] Zwirner, P., Murry, T., & Woodson, G.E. (1991). Phonatory function of neurologically impaired patients. *Journal of Communication Disorders*, **24**, 287-300.

[23] Eskenazi, L., Childers, D.G., & Hicks, D.M. (1990). Acoustic correlates of vocal quality. *Journal of Speech and Hearing Research*, **33**, 298-306.

[24] Shoji, K., Regenbogen, E., Yu, J.D., & Blaugrund, S.M. (1992). High-frequency power ratio of breathy voice. *Laryngoscope*, **102**, 267-271.

[25] Deliyski, D. (1993). Acoustic model and evaluation of pathological voice production. *Proceedings of the 3rd Conference on Speech Communication and Technology, Eurospeech'93*, Berlin, Germany.

[26] Hertrich, I., Ackermann, H., Braun, S., & Spieker, S. (1996). Geschlechtsdimorphismus pathologischer Stimmmerkmale bei zentralnervösen Dysphonien: Eine vergleichend auditiv-akustische Studie. *Sprache - Stimme - Gehör*, **20**, 169-174.

[27] Ramig, L.A. (1986). Acoustic analyses of phonation in patients with Huntington's disease. *Annals of Otology, Rhinology and Laryngology*, **95**, 288-293.

[28] Ramig, L.A., Titze, I.R., Scherer, R.C., & Ringel, S.P. (1988). Acoustic analysis of voices of patients with neurologic disease: Rationale and preliminary data. *Annals of*

Otology, Rhinology and Laryngology, 97, 164-172.

[29] Hertrich, I., & Ackermann, H. (1995). Gender-specific vocal dysfunctions in Parkinson's disease: Electroglottographic and acoustic analyses. *Annals of Otology, Rhinology & Laryngology*, **104**, 197-202.

[30] Hertrich I., & Ackermann, H. (1996). Electroglottographic assessment of vocal instabilities in patients with Parkinson's disease and cerebellar atrophy. In T. Powell (Ed.), *Pathologies of Speech and Language: Contributions of Clinical Phonetics and Linguistics* (pp 199-202). ICPLA, New Orleans.

[31] Herzel, H., Berry, D., Titze, I., & Steinecke, I. (1995). Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling. *Chaos*, **5**, 30-34.

[32] Herzel, H., & Knudsen, C. (1995). Bifurcations in a vocal fold model. *Nonlinear Dynamics*, **7**, 53-64.

[33] Lucero, J.C. (1993). Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation regions. *Journal of the Acoustical Society of America*, **94**, 3104-3111.

[34] Isshiki, N., Tanabe, M., Ishizaka, K., & Broad, D. (1977). Clinical significance of asymmetrical vocal cord tension. *Annals of Otology*, **86**, 58-66.

[35] Kiritani, S. (1995). Recent advances in high-speed digital image recording of vocal cord vibration. In K. Elenius & P. Branderud (Eds.), *Proceedings ICPHS 95*, Vol. 4 (pp 62-67). Stockholm: Stockholm University,

[36] Baken, R.J. (1990). Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice. *Journal of Voice*, **4**, 185-197.

[37] Behrman, A. (1999). Global and local dimensions of vocal dynamics. *Journal of the Acoustical Society of America*, **105**, 432-443.

[38] Hertrich, I., Lutzenberger, W., Spieker, S., & Ackermann, H. (1997). Fractal dimension of sustained vowel productions in neurological dysphonias: An acoustic and electroglottographic analysis. *Journal of the Acoustical Society of America*, **102**, 652-654.

[39] Herzel, H., Berry, D., Titze, I.R., & Saleh, M. (1994). Analysis of vocal disorders with methods from nonlinear dynamics. *Journal of Speech and Hearing Research*, **37**, 1008-1019.

[40] Mende, W., Herzel, H., & Wermke, K. (1990). Bifurcations and chaos in newborn infant cries. *Physics Letters A*, **145**, 418-424.

[41] Narayanan, S.S., & Alwan, A.A. (1995). A nonlinear dynamical systems analysis of fricative consonants. *Journal of the Acoustical Society of America*, **97**, 2511-2524.

[42] Farmer, J.D., Ott, E., & York, J.A. (1983). Dimension of chaotic attractors. *Physica D*, **7**, 153-180.

[43] Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D*, **9**, 189-208.

[44] Steinecke, I., & Herzel, H. (1995). Bifurcations on an asymmetric vocal-fold model. *Journal of the Acoustical Society of America*, **97**, 1874-1884.

[45] Kröger, B.J. (1991). Zur Auswirkung der Glottis-Sprechtrakt-Kopplung auf die Stimmreinheit. *Sprache - Stimme - Gehör*, **15**, 139-142.

[46] Nittrouer, S., McGowan, R.S., Milenkovic, P.H., & Beehler, D. (1990). Acoustic measurements of men's and women's voices: A study of context effects and covariation. *Journal of Speech and Hearing Research*, **33**, 761-775.

[47] Sussman, J.E., & Sapienza, C. (1994). Articulatory, developmental, and gender effects on measures of fundamental frequency and jitter. *Journal of Voice*, **2**, 145-156.