TOWARDS A CLASSIFICATION OF PHONATORY FEATURES OF DISORDERED VOICES

J. Schoentgen
National Fund for Scientific Research, Belgium
Department "Signals and Waves", Faculty of Applied Sciences, Université Libre de Bruxelles, Brussels
jschoent@ulb.ac.be

The purpose of the presentation is to give an overview of phonatory features of disordered voices and propose a classificatory framework. Generally speaking, phonatory features are numerical cues that summarize properties of speech signals or other voice-related signals that are obtained non-invasively, and which are clinically relevant. The object of a classificatory framework is to ease the planning of future experiments and the exploitation of the existing literature, which is diverse with regard to pathologies studied, speaker tasks, speaker performance, speech material, vocal symptoms, sensors and sought for relations with other levels of description.

I. INTRODUCTION

The purpose of the presentation is to give an overview of phonatory features of disordered voices and propose a classificatory framework. Generally speaking, phonatory features are numerical cues or measurements that are clinically relevant and that summarize properties of speech signals or other signals that are obtained non-invasively, and which report on a speaker's voice. Typically, the acquisition of phonatory features involves the (non-invasive) recording of signals that are relevant to laryngeal function, the signal processing that discards irrelevant signal properties and the summary of the clinically-relevant properties by means of a handful of numbers.

The purpose of the extraction of phonatory features in a clinical framework is the documentation of the voice of patients, their longitudinal follow-up during treatment (e.g. before and after surgery) as well as comparisons with normophonic speakers.

The goal of the presentation is not to discuss a classification of the mathematical forms of clinical cues per se. This would be an ineffectual exercise because most of the extant features are heuristically defined and their mathematical or statistical properties have been explored superficially only. The goal is rather to classify phonatory cues in relation to the use they have been put to.

Indeed, the scientific as well as clinical literature devoted to phonatory features of voice disorders is abundant. However, its diversity is impressive with regard to the pathologies or handicaps that have been studied, the vocal symptoms that have been described, the vocal tasks speakers have been asked to carry out and the linguistic, paralinguistic and extralinguistic performances that have been examined, the sensors that have been used and the speech material that has been recorded, as well as the correlations that have been examined.

As a consequence, it is difficult to distil general rules or compare results obtained in different frameworks. The purpose of a classification is therefore to ease the planning of future experiments and the exploitation of the existing literature.

II. CLASSIFICATORY FRAMEWORK

The following is a proposal of a grid that may be used to classify different feature-based approaches to the assessment of voice disorders.

Etiology

The types of pathologies or handicaps the effects on voice of which have been investigated are wide and varied. Typically, one distinguishes voice problems that are the consequences of organic alterations of the vocal folds from those that are dysfunctional, i.e. voice disorders that are not the consequence of observable structural changes of the vocal folds [1]. Voice problems caused by motor disorders are a third category. An example of the latter is the voice of Parkinson speakers. A separate category are substitution voices, the purpose of which is to enable speech communication by speakers who have lost the capacity of producing voice by means of the vibration of the true vocal folds.

Whether voice disorders that have different causes must be described by different sets of phonatory features is not clear at present. Generally speaking, speech and voice problems owing to motor disorders are kept separate from voice disorders that are the consequence of laryngeal pathologies [6]. A possible exception is vocal fold paralysis.

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)

Speech material

One major distinction between approaches to voice assessment rests on the speech material. Indeed, the speech material that is analyzed may be connected speech or sustained speech sounds. Sustained speech sounds may again be subdivided according to whether onsets and offsets are included in the analysis frame or not. Connected speech is often presented as ideal; analyses of stationary speech fragments are the rule, however [5]. The reason is that many signal processing schemes are based on assumptions of local stationarity and local periodicity. These assumptions may not be valid in the case of hoarse speakers emitting connected speech [2].

Speaker tasks

Tasks refer to what is requested from a speaker during vocal assessment. Tasks that subjects are the most frequently asked to carry out are speaking, which includes sustaining speech sounds, singing (when appropriate), vocal loading, as well as profiling.

Vocal loading consists in recording the phonatory features of a speaker, followed by reading out loud a text for some time (e.g. 45 minutes) and recording the same features again. The purpose is to track vocal alterations that are the consequence of burdening the larynx [7].

Finally, profiling is the discovery of the limits of phonatory performance, e.g. loudest possible voice, softest possible voice, highest possible pitch, lowest possible pitch, maximum phonation time, etc [9].

Speaker performance

Speaker performance refers to the actual capacity that is examined. Phonatory performance may be subdivided into registers, phonation types, voicing, prosody and vocal quality.

Known speech registers are creak, modal voice and falsetto. Examples of phonation types are breathy voice, soft voice, modal voice, loud voice, pressed voice and so forth [3].

Voicing is the capacity of the speaker to voice and un-voice speech sounds. Prosody refers to the capacity to control intonation, accentuation, and rhythm, as well as speech rate. Voice quality, finally, designates vocal timbre, e.g. hoarseness, roughness, vocal tremor or quaver, and so forth.

Instrumentation

Instrumentation refers to the equipment that is used to obtain signals non-invasively that report on the phonatory performance of speakers. The microphone signal is the most often used; it evolves proportionally to acoustic pressure and therefore proportionally to the speech signal that is recorded by the ear of a listener. Other signals that can be obtained non-invasively are the electroglottogram and photoglottogram. The former is reported to evolve proportionally to the vocal fold contact surface and the latter to the glottal area. One other sensor that is used frequently is the flow mask that enables recording airflow rate, as well as, occasionally, intra-oral pressure.

Signals

One major distinction is the one between features that describe the phonatory source signal and those that describe the speech signal. The phonatory signal is the acoustic signal that is generated at the glottis via the vibration of the vocal fold and pulsatile airflow. The speech signal is emitted at the mouth consequent to the propagation of the acoustic wave through the vocal tract. Observing the glottal source signal directly is difficult. Often, it is replaced by auxiliary signals such as the photoglottographic or electroglottographic signals that report on glottal properties directly.

Transform domains

At present, a systematic classification of clinical signal processing schemes is not possible because most schemes involve a heuristic processing stage that may differ from task to task and from study to study.

A possible processing-related categorization is based on the type of signal transform that is involved. Examples are Fourier, Hilbert or Wavelet transforms. When no signal transformation is carried out, the corresponding phonatory features are temporal; otherwise they acquire properties that are typical of the corresponding transform domains [8].

Vocal symptoms

One core distinction between phonatory features is the one that pertains to vocal symptoms. Vocal symptoms are the speech properties that are believed to be clinically relevant, that report on the state of the glottis and that the signal processing is aimed at. Typically one distinguishes between signal dysperiodicity, signal morphology, and supra-segmental as well as coordinative features.

Coordinative features refer to the onset and offset of voicing in relation to supra-glottal events. Examples of relevant supra-glottal events are obstruent check and release or resonant onset and offset. The most often studied coordinative cue is vocal onset time, which is the signed time interval between the release of an obstruent and the onset of voicing, on which hinges the distinction between voiced and unvoiced obstruents. This language-typical interval, which may be short, requires a fine control of glottal adduction and abduction with regard to supra-glottal articulation. Voice onset time is therefore frequently studied in relation to motor speech disorders or substitution voices, which are suspected to impede fine control of voicing [10].

Supra-segmental features pertain to intonation, accentuation, rhythm, speech rate as well as average phonatory frequency, the variability of the phonatory frequency, and average loudness, i.e. sound pressure level. In a clinical framework, speech rate, the average and spread of phonatory frequency, as well as sound pressure level are the most popular.

Morphological features refer in practice to the shape attributes of the glottal source signal. Examples are the open quotient, closing quotient, speed quotient, the amplitude of the volume velocity as well as the amplitude of the negative peak of the differentiated volume velocity [3]. An example of a spectral morphological feature is the spectral balance, which quantifies harmonic richness. Morphological features have been mainly used to characterize phonation types.

Morphological, supra-segmental as well as coordinative features are not confined to clinical applications. These features have been widely studied by phoneticians, linguists, psychologists and engineers because they report on speech and voice production as well as perception in general.

Features that are typically clinical are those that describe irregularities of the movement of the vocal folds. Generally speaking, one distinguishes between non-modal vibratory regimes that cause diplophonia, biphonation and random cycles, and external perturbations (i.e. modulation noise), turbulence noise and breathiness (i.e. additive noise), unsolicited vibrations of the false vocal folds or ary-epiglottal folds, as well as uncontrolled transients, such as voice breaks, register breaks, octave jumps and so forth.

External perturbations give rise to vocal jitter and shimmy, as well as vocal frequency and amplitude tremor. The main contribution to shimmy and amplitude tremor of the speech cycles is the modulation distortion by the vocal tract of phonatory jitter and phonatory frequency tremor. Other causes are the transfer of acoustic energy from cycle to cycle, as well as tremor of the speech articulators. Phonatory shimmy or phonatory amplitude tremor contribute only feebly to speech shimmy or speech amplitude tremor [4].

Correlation

More often than not, clinicians attempt to correlate observed acoustic features with data recorded at other levels of description. Data, correlations with which are sought for, are typically diagnostic, glottal, aerodynamic or perceptual, with a preference for the latter.

III. KNOWN PROBLEMS

Known problems with extant phonatory features are the following.

Signal processing

The most popular acoustic features are those that quantify the degree of irregularity of the vocal cycles. Typical examples are the period perturbation quotient, amplitude perturbation quotient, jitter in %, harmonics-to-noise ratio and so forth. More often than not, the signal processing involves methods that are based on assumptions of local stationarity and local periodicity that enable heuristics to detect and isolate vocal cycles or spectral harmonics. These heuristics may fail in the case of severely hoarse voices. As a consequence, insertion or omission errors are frequent in the case of highly irregular signals. These errors bias the values of the calculated features. As a consequence, phonatory features that describe vocal perturbations are thought to be reliable only when extracted from sustained speech sounds uttered by feebly or moderately hoarse speakers.

Another issue is measurement precision. Indeed, perturbations of cycle length may be small, e.g. less than one percent for speech cycle lengths, less than ten percent for speech cycle amplitudes. As a consequence, signal processing request precautions with regard to measurement precision. Otherwise, measurements may be biased by quantization noise, for instance.

Labelling

Labelling refers to the custom of giving phonatory features names that allude to vocal symptoms rather than to the measurements that are actually performed. For instance, features that summarize the dysperiodicity of glottal cycle lengths are often referred to as vocal jitter, although cycle length dysperiodicity may also be influenced by average phonatory frequency, additive noise, frequency tremor, non-modal dynamic regimes of the vocal folds and non-flat intonation, for instance.

Redundancy

The number of phonatory features that have been proposed in the literature is large. Software that is sold for clinical assessment of voice typically comprises tens of numerical cues that can be computed for a single sustained sound. Studies have shown that sub-sets of phonatory features are correlated with each other. Sub-sets of correlated features are roughly coextensive with the groups of vocal symptoms that are discussed above [5].

Interpretation

Even very simple measurements are influenced by multiple factors. Examples are given above for the perturbations of the speech cycle lengths, as well as for the perturbations of the speech cycle amplitudes, which are generated via modulation distortion. These observations suggest that phonatory features are difficult to interpret because they are determined by multiple causes that may be interdependent.

Stationary fragments of sustained speech sounds

One of the more frequently heard complaints is that many acoustic cues can only be obtained reliably for stationary fragments of sustained speech sounds. The reasons have been discussed in the section on signal processing. A consequence is that, at present, the effects of voice disorders on connected or natural speech are less well understood. Problems that are non-resolved are not only issues in signal processing, but also the choice of the phonatory features, the choice of speech material, as well as the perceptual assessment of connected speech fragments that are short or phonetically complex.

IV. SUMMARY

The following Table summarizes some of the factors that distinguish different approaches to voice assessment.

Etiology	organic, dysfunctional, and motor disorders, substitution voices
Transforms	Fourier, Hilbert, and Wavelet transforms (if applicable)
Signals	glottal source, speech signal
Material	connected speech, sustained speech sounds, stationary fragments of sustained speech sounds
Symptoms	dysperiodic, morphologic, supra-segmental, coordinative
Tasks	speaking, singing, loading, profiling
Performance	registers, phonation types, voicing, prosody, voice quality
Sensors	microphone, electroglottograph, photoglottograph, flow mask

REFERENCES

The synthetic description of feature-based vocal assessment that is given here rests on a large number of articles, which cannot be listed because of lack of space. References that are cited below are only pointers to further reading.

[1] Boone D., McFarlane S. (1996) The voice and voice therapy, Prentice Hall, NJ.

[2] Bettens F., Grenez F., Schoentgen J. (2005) Estimation of vocal dysperiodicities in disordered connected speech by means of distant-sample bidirectional linear predictive analysis, J. Acoust. Soc. Am., 117, 1, 328-337.

[3] Alku P., Vilkman E. (1996) A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers, Folia Phoniatrica et Logopaedica, 48, 240-254.

[4] Schoentgen J. (2003) Spectral models of additive and modulation noise in speech and phonatory excitation signals, J. Acoust. Soc. Am., 113, 1, 553-562.

[5] Dejonckere Ph., Remacle M., Fresnel-Elbaz E., Woisard V., Crevier-Buchman L., Millet B. (1996) Differential perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurement, Rev. Laryngol. Otol. Rhinol., 117, 3, 219-224.

[6] Till J., Yorkston K., Beukelman D. (1994) Motor speech disorders – Advances in assessment and treatment, Brookes Publ., Baltimore.

[7] Vilkman E., Lauri E., Alku P., Sala E., Sihvo M. (1999) Effects of prolonged oral reading on F0, SPL, sub-glottal pressure and amplitude characteristics, J. Voice, 13, 2, 303-315.

[8] Baken R., Daniloff R. (1991) Readings in clinical spectrography of speech, Singular Publ., San Diego.

[9] Heylen L., Wuyts F., Mertens F., De Bodt M., Van de Heyning P. (2002) Normative voice range profiles of male and female professional voice users, J. Voice, 16, 1, 1-7.

[10] Ziegler W., Deger K. (1998) Clinical Phonetics and Linguistics, Whurr Publ., London, 435-449.