Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 54 (2012) 611–621
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Disordered voice measurement and auditory analysis

David M. Howard [a,*], Evelyn Abberton [b], Adrian Fourcin [b]

[a] Audio Laboratory, Department of Electronics, University of York, UK
[b] Department of Phonetics and Linguistics, University College London, UK

Available online 5 April 2011

Abstract

Although voice disorder is ordinarily first detected by listening, hearing is little used in voice measurement. Auditory critical band approaches to the quantitative analysis of dysphonia are compared with the results of applying cycle-by-cycle time based methods and the results from a listening test. The comparisons show that quite large rough/smooth differences, that are readily perceptible, are not as robustly measurable using either peripheral human hearing based GammaTone spectrograms, or a cepstral prominence algorithm, as they may be when using cycle-by-cycle based computations that are linked to temporal criteria. The implications of these tentative observations are discussed for the development of clinically relevant analyses of pathological voice signals with special reference to the analytic advantages of employing appropriate auditory criteria.

Keywords: Hearing modelling; Pathological voice; Dysphonia; Temporal analysis; Spectral analysis; Laryngograph; Clinical voice analysis

1. Introduction

The evaluation of disordered speech, in the present instance, dysphonia, is a vital element in the clinical appraisal and treatment of the human voice. This evaluation typically relies on a subjective as well as an objective analysis carried out via critical listening and algorithmically-based speech processing respectively. The results of these approaches may not, however, always tally in terms of what they indicate but it is, nevertheless, often the case that the objective results are relied upon to a greater degree, for insurance purposes for example, than are those from subjective listening. The discrepancy is due perhaps, at least in part, to the difficulty of carrying out meaningful and useful listening tests set against the relative ease of obtaining data by running objective analyses on a speech signal. If the objective analysis does indeed provide a true and meaningful representation of the auditory processing of the speech signal then this should not be an issue. However, if compromises have been made in the design and implementation of the algorithm, perhaps through the use of mathematically convenient criteria that do not reflect the natures of either speech production or perception, or by making assumptions about the onset and offset of voiced speech segments that are not appropriate in the case of disordered speech, or solely by the use of sustained vowels when the sensitive data comes from running speech, then the resulting analyses can provide misleading data.

The COST 2103 Action on "Advanced Voice Function Assessment" that is supporting this Special Issue aims to "combine previously unexploited techniques with new theoretical developments to improve the assessment of voice for as many European languages as possible, while acquiring in parallel data with a view to elaborating better voice production models." and it is noted that "Progress in the clinical assessment and enhancement of voice quality requires the cooperation of speech processing engineers and laryngologists as well as phoniatricians." (WWW-1, 2011). This paper is directed squarely at this goal, in partic-

* Corresponding author. Address: Audio Laboratory, Department of Electronics, University of York, Heslington, York YO10 5DD, UK.
E-mail address: dh@ohm.york.ac.uk (D.M. Howard).

0167-6393/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2011.03.008

ular by investigating both hearing modelling analysis and temporal analyses for clinical assessment and voice quality rehabilitation. The analysis of disordered speech is considered in the context of combining subjective and objective analyses through the use of computational models of the peripheral hearing system alongside established speech analysis methods.

1.1. Human hearing modelling

Seventy years ago Fletcher (1940) introduced the concept of a "critical band width" in human auditory frequency analysis and proposed that this could be quantified in terms of "equivalent rectangular bands" — now referred to as "ERBs". The basic idea pervades modern work in psychoacoustics and this approach has been more recently developed practically by the use of the "GammaTone" (GT) and "gamma chirp" formulations (Section 2.2 below). The critical band representation, in these different new forms, has widespread application in speech communication systems: for speech recognition (e.g. Cooke et al., 2010; Shao et al., 2010; Watkins and Makin, 2007); speaker identification (e.g. Abdulla and Zhang, 2010; Falk and Chan, 2010; Li and Huang, 2010); and in the analysis of pathological voice signals (e.g. Caeiros et al., 2010; Wang et al., 2008; Malyska et al., 2005).

1.2. Pathological voice

Clinically oriented voice analysis work in common with that on holistic techniques of speech recognition, currently makes use of the critical band approach either directly by the use of the bark scale (Moore, 2004; Zwicker, 1961) or, most often, implicitly by the use of the MFCC (mel frequency cepstral coefficient (Abdulla and Zhang, 2010; Falk and Chan, 2010) and PLP (perceptual linear prediction) front end processing methods (Bridle and Brown, 1973; Mermelstein, 1976; Hermansky, 1998). All these frequency analytic scales are based on psychophysical techniques that involve the exploration of the steady state characteristics of the hearing mechanism. Normal speech signals are essentially dynamic at the syllabic, phone and sub-segmental levels with important linguistically significant contrasts capable of being associated with temporal intervals as small as 30 ms (Lisker and Abramson, 1964). For the normal voice (laryngeal) frequency range of $\sim$30 Hz to $\sim$1.2 kHz auditory critical band widths are not well able to respond to small temporal irregularities. The analysis of disordered voice is largely directed towards the measurement of voice irregularity and an important subset of analysis algorithms used is based on the use of time waveform data (Buder, 2000) (see Section 2.2 below). In consequence, it is of real interest and practical value to contrast frequency domain filter based analysis with time domain based cycle-by-cycle approaches and — for this initial tentative appraisal — to be linked to the results of a listening test.

Two hypotheses are proposed.

Hypothesis I. When Fx > $\sim$100 Hz the GammaTone filters in the low frequency region up to $\sim$1 kHz will not robustly analyse cycle duration irregularities as great as those associated with changes of a musical tone due to the temporal smoothing imposed by the corresponding GT impulse responses, where Fx, the instantaneous frequency, is determined from the duration of an individual cycle of vocal fold vibration.

Hypothesis II. The ordinary listener has two pitch perceptual mechanisms available and although classic auditory filtering, as exemplified by the GammaTone model, is not totally adequate for the detection of irregularity, auditory temporal processing may well be able to detect duration irregularities corresponding to as large as a musical tone in the band 0 to $\sim$1 kHz.

2. Material and methods

The key objective of this work is to make and compare frequency and temporal analyses of disordered voice in the context of salient properties of the human peripheral hearing system with special reference to current clinical voice evaluation. In order to achieve this, the following hearing-related and well-established analysis techniques are employed as well as a simple perceptual test, and these are described below. Where appropriate, the analytic outputs are compared with traditional wide and narrow-band spectrograms (Baken, 1991, 2000).

- Frequency analytic hearing modelling (Section 2.1).
- Time related waveform measurements (Section 2.2).
- A simple perceptual test linked to clinically relevant examples of objective measurements using common data (Section 2.3).

2.1. Frequency analytic hearing modelling

Hearing modelling spectrograms provide an indication of the nature of the output from the basilar membrane of the inner ear that is based on modelling it as a bank of band-pass filters operating in parallel (Howard and Angus, 2009; de Cheveigné, 2010). These filters do not exhibit symmetrical response curves in the frequency domain (de Boer and de Jongh, 1978; Moore, 2004) and therefore their outputs differ from those used in traditional wide-band and narrow-band spectrography. One commonly used model for the shape of an auditory filter is the GammaTone filter which was originally introduced to describe the shape of the impulse response function of the auditory system as estimated by the reverse correlation function of neural firing times (Patterson, 1976). In the time domain, the