Biomedical Signal Processing and Control 36 (2017) 183–193

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Comparison of parametrization methods of electroglottographic and inverse filtered acoustic speech pressure signals in distinguishing between phonation types

Dong Liu [a,b], Elina Kankare [c], Anne-Maria Laukkanen [d,*], Paavo Alku [e]

[a] *CAS Key Laboratory of Microscale Magnetic Resonance and Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China*
[b] *Synergetic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China*
[c] *Ear and Oral Diseases, Department of Phoniatrics, Tampere University Hospital, Tampere, Finland*
[d] *Speech and Voice Research Laboratory, University of Tampere, Tampere, Finland*
[e] *Department of Signal Processing and Acoustics, Aalto University, Helsinki, Finland*

ARTICLE INFO

Article history:
Received 18 July 2016
Received in revised form 5 April 2017
Accepted 10 April 2017

Keywords:
Contact quotient (CQ)
Normalized amplitude quotient (NAQ)
Perceived firmness
Voice training and therapy

ABSTRACT

This study compared for the electroglottographic (EGG) signal how well six earlier presented and two new parameters distinguish between normal, breathy and pressed phonation and how well they correlate with perceptual evaluation. The results were compared with those obtained for nine parameters describing the glottal flow waveform obtained through inverse filtering of the acoustic speech pressure signal. Acoustic and dual-channel EGG signals were recorded for twenty female and twenty male subjects with healthy voices phonating sustained samples of the vowel [a:] in their habitual normal voice and in simulated breathy (hypofunctional) and pressed (hyperfunctional) phonation. The samples were perceptually evaluated by five voice specialists and rated for firmness of phonation. The best examples from 12 females and 12 males were used for the analyses. Few earlier studies have ranked the behavior of this many EGG and glottal flow parameters from this large speech data.

Although the parameters differed in their ranking order, contact quotient calculated with a criterion level at 50% both from the EGG and the inverse filtered signal was strong in correlating with perception and in distinguishing phonation types in cases where fundamental frequency and sound pressure level also varied. When this variation was taken into account, the normalized amplitude quotient NAQ still had an effect in predicting voice quality. The results will have applicability in voice training and therapy and in development of machine learning -based classification methods.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

How the voice is produced – i.e., phonation quality – is important, as it essentially affects how the voice sounds and functions in communication and how it resists vocal loading. Thus, phonation quality also has an important role in the prevention of potential traumas related to vocal overloading. Of special interest are non-invasive methods of studying phonation quality. Electroglottography (EGG) is a non-invasive method where a high frequency, low voltage current is fed through the larynx to study changes in the vocal fold contact area during phonation, based on the electrical impedance changes the varying contact causes [1]. EGG has been used to study phonation types [2–6], vocal differences between genders [7–10], ages [11–13], vowels [5,14–17], emotional expressions [18–21], vocal registers [e.g., [22–27]] and vibrato or tremor [28,29], and the effects of voice training [e.g., [30,31]], vocal exercises [32–38], and voice disorders [16,39–43]. The correspondence between the EGG signal waveform and physiologic and acoustic measures has been investigated and found to be relatively good [44–50]. Due to difficulties in pointing to the exact beginning and ending of the opening and closing times of the glottis when using the EGG, the method has been regarded as more suitable for analyzing the duty cycle [51]. The relative contact time (contact quotient, CQ, i.e., contact time divided by period time) has been extensively focused on. It has been found to distinguish between registers [26,27] and vocal expressions of emotions [19], to differentiate healthy and disordered voices at least in some cases [43], and to correlate with perception of voice quality [6,52] and even to some

* Corresponding author.
E-mail address: Anne-Maria.Laukkanen@uta.fi (A.-M. Laukkanen).

http://dx.doi.org/10.1016/j.bspc.2017.04.001
1746-8094/© 2017 Elsevier Ltd. All rights reserved.

extent with the impact stress (force per unit area) in vocal fold vibration [53]. The impact stress (IS) is regarded as the main loading factor during phonation [54]. Since IS is difficult to measure in humans [55,56], methods for non-invasive estimation of IS are important.

CQ has been measured using different peak-to-peak amplitude-based criterion levels from 10 to 80% [4,6,8,11,51,57,58] because it is problematic to place the exact beginning and ending of the glottal closing events in the EGG signal. The choice of criterion level has been made on the basis of the sample and signal types or based on another method used for comparison with the EGG signal (such as stroboscopy, high-speed filming, inverse filtering, videokymography, or modeling of vocal fold vibration). Hacki suggests the use of area-based contact quotient (CQA) to study disordered voices [41]. The results of Higgins and Schulte showed that gender effects become visible for criterion levels from 55% upward [8]. In male singers, a CQ with a criterion level at 25% (CQ25%) seems to fit best with videokymographic images [47]. According to Kania et al. [59], the criterion level of a CQ higher than 25% is more affected by F0 and intensity than phonation type in male voices. Furthermore, Kankare et al. found that in the female speaking voice, a CQ with criterion levels at 25% and 35% (CQ35%) correlates best with perceived phonation type, and CQ25% is least affected by F0 and sound pressure level (SPL) but seems to reflect phonation type best [52].

To avoid the difficulty of choosing the glottal opening instant (GOI) and glottal closing instant (GCI) in the EGG signal [49,60], the first derivative of the EGG signal (DEGG) has been used to detect the GOI and GCI, as shown in Fig. 1. One problem related to the use of DEGG is that the signal is vulnerable to noise, and it may be very difficult to pinpoint its opening instant in particular. Therefore, a hybrid parameter for the calculation of the CQ has been proposed [30,61]. The opening instant is obtained by using a criterion level of about 42% (three-sevenths) of the peak-to-peak amplitude of the EGG, and the closing instant is defined by the maximum peak of the DEGG.

The amplitude of the maximum peak of the DEGG (MDEGG) reflects the glottal closing speed. Therefore, it should also correlate with SPL, F0, and phonation type. Results by Kankare et al. showed that MDEGG correlates with the perceived firmness of phonation in female speakers [62]. As far as the authors know, the parameter has not been systematically studied. For example, the parameter's behavior has not been tested for male voices.

Inverse filtering (of either flow or the speech pressure signal) is another non-invasive method for studying voice quality, and many methods have been applied to parametrize the resulting volume velocity waveform. Relative glottal closing speed (closing quotient; ClQ; i.e., closing time divided by period time) calculated from the volume velocity waveform is known to increase with SPL and a stronger adduction [63]. It thus seems to be well suited for estimating vocal loading, since IS also increases together with these factors [64]. The pulse asymmetry parameter speed quotient (SQ; i.e., glottal opening time divided by closing time) has also been found to increase with loudness, especially in males [63,65]. Furthermore, it has been reported to correlate with perceived effort of voice production [60]. On the other hand, both SQ and ClQ are most sensitive to abnormalities of the glottal flow, for example due to lesions of the vocal folds [66]. The normalized amplitude quotient (NAQ_{inv}) from the glottal volume waveform has been found to distinguish between phonation types [67]. In the present paper, NAQ is for the first time calculated for the EGG. Glottal spectrum-based parameters like the harmonic level difference between the first two harmonics (DH12), the harmonic richness factor (HRF), and the parabolic spectrum parameter (PSP) have also been shown to correlate with voice quality [68–70].

The glottal inverse filtering has several advantages, such as the method's ability to estimate the voice source non-invasively from the microphone signal, and the possibility to implement an analysis in an automatic manner for modern applications such as parametric speech synthesis [71]. Glottal inverse filtering, however, suffers from a few drawbacks, the most severe of which is poor estimation accuracy in the analysis of high-pitched speech due to the biasing of the formant estimates by the sparse harmonics in the spectra of high-pitched speech. In addition, most of the inverse filtering algorithms are not capable of modeling non-linearities in speech production because the methods are built on the assumption of linearity between the source and the tract. For more details on the pros and cons of glottal inverse filtering, see two recent review articles [72,73].

In summary, for the EGG signal, DEGG should be more accurate in reflecting the glottal opening and closing events than the threshold-based methods [49], and MDEGG should reflect perceived firmness of phonation well, at least for females voices [62]. So far MDEGG has not been tested for male voices. The hybrid parameter CQ3/7 combines threshold-based and derivative-based approaches to provide a more accurate and robust method [30]. However, derivative of the EGG signal is vulnerable to noise. $\text{CQ}\% < 55$ should be less affected by gender [8], and CQA should be robust enough to suit even for pathological voices [41], but CQ35% should correlate best with perceived voice quality and CQ25% should distinguish phonation type best in samples where F0 and SPL also have variation [52,59]. NAQ_{inv} and several spectrum-based parameters calculated for the inverse filtered acoustic speech pressure waveform have been found to correlate with voice quality [67–70]. However, their performance has not been extensively tested against the more traditional time-based parameters and against each other. Furthermore, NAQ has not been calculated for the EGG signal before. Additionally, in most of the previous studies, simulated phonation type has been investigated keeping F0 and SPL constant, which is an unnatural situation.

Due to the above mentioned reasons there seems to be a need to test the performance of various parametrization methods of EGG and inverse filtered signal for the same speech sample. The present study compared a set of eight EGG parametrization methods and nine glottal flow parameters (thus 17 parameters in total). The parameters chosen were for the EGG: CQ25%, CQ35%, CQ50%, CQA, CQDEGG, CQ3/7, NAQ and MDEGG. The inverse filtered signal parameters were CQ, $\text{CQ50\%}_{\text{inv}}$, CQA_{inv}, ClQ, SQ, NAQ_{inv}, DH12, HRF, and PSP. F0 and SPL were allowed to vary naturally in the samples presenting three phonation types. The questions of interest are: 1) Do the parameters differentiate between phonation types, and 2) Which of the parameters correlate best with perception of phonation quality. To the best of our knowledge, no study so far has extensively ranked the behavior of this many parametrization methods of the EGG and glottal flow signal in reflecting the phonation quality of the same sound samples.

2. Materials and methods

2.1. Subjects and recordings

Twenty females and twenty males with healthy voices volunteered as subjects. They phonated at their habitual conversational pitch and loudness on the vowel [a:] in three ways: habitual voice, breathy voice, and pressed voice. The duration of each vowel sample was approximately five seconds. The samples were recorded in a sound-treated studio using a dual-channel EGG (Glottal Enterprises; low frequency limit set to 20 Hz) and a headset microphone (AKG C477) at a distance of 6 cm from the corner of the subject's mouth. The samples were recorded on a PC through an external sound card (M-Audio, MobilePre USB) using SoundForge software. The sampling rate was 44.1 kHz and the amplitude quantization

Download English Version:

https://daneshyari.com/en/article/4973546

Download Persian Version:

https://daneshyari.com/article/4973546

Daneshyari.com