Biomedical Signal Processing and Control 10 (2014) 137–143

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

The harmonic and noise information of the glottal pulses in speech

Ricardo Sousa [a,*], Aníbal Ferreira [a,1], Paavo Alku [b,c,2]

[a] Department of Electrical and Computer Engineering, University of Porto, School of Engineering, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
[b] Department of Signal Processing and Acoustics, Aalto University, School of Science and Technology, Espoo, Finland
[c] Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, P.O. Box 3000, FIN-02015 TKK, Finland

ARTICLE INFO

Article history:
Received 8 March 2013
Received in revised form 6 November 2013
Accepted 12 December 2013
Available online 2 January 2014

Keywords:
Voice quality
Voice diagnosis
Glottal inverse filtering
Glottal excitation
Harmonic and noise components

ABSTRACT

This paper presents an algorithm, in the context of speech analysis and pathologic/dysphonic voices evaluation, which splits the signal of the glottal excitation into harmonic and noise components. The algorithm uses a harmonic and noise splitter and a glottal inverse filtering. The combination of these two functionalities leads to an improved estimation of the glottal excitation and its components. The results demonstrate this improvement of estimates of the glottal excitation in comparison to a known inverse filtering method (IAIF). These results comprise performance tests with synthetic voices and application to natural voices that show the waveforms of harmonic and noise components of the glottal excitation. This enhances the glottal information retrieval such as waveform patterns with physiological meaning.

1. Introduction

In producing speech sounds, humans are able to regulate the tension of laryngeal muscles in combination with the respiratory effort. These physiological settings change the time-varying flow of air through the vibrating vocal folds, i.e. the glottal volume velocity waveform. Since this signal serves as the source of (voiced) speech, it has an essential role in the production of several acoustical phenomena and cues that are used in everyday speech communication such as the regulation of vocal intensity [1–3], voice quality [4–7], and the production of different vocal emotions [8–10]. In addition, glottal pulse forms reveal physiological conditions and dynamics of the vocals folds, which might help detecting voice pathologies related to vocal fold changes [11–13]. Therefore, accurate analysis and parameterization of the glottal pulseform is beneficial in several areas of speech science including both healthy and disordered voices.

It is well known that even in the case of sustained vowels produced by healthy subjects, the vibration of the vocal folds is never completely periodic. Consequently, the glottal source is typically regarded to comprise two major components; the harmonic (periodic) and the noise (aperiodic) component. Several previous studies indicate that for certain voice types, such as breathy and hoarse voices, the amount of noise is increased in the glottal flow [14,15]. The perceptual effects of the aperiodic components of the glottal flow have been studied, for example, in relation to the hoarseness [16] and the breathiness [17] of the voice. The perceptual importance of aperiodic components of the voice source is also recognized in speech synthesis where increasing efforts are currently devoted toward a better understanding of aperiodicities in the voice source [18,19]. Moreover, the aperiodic behavior of the vocal apparatus has been studied by voice pathologists who have used perceptual parameters such as hoarseness and roughness in their voice diagnosis. The importance of these perceptual parameters is reflected on the RASAT and GRBAS scales of voice quality [20] and it has been shown that hoarseness and roughness are connected to the presence of noise and acoustic aperiodicities in speech [21]. In particular, it has been found that some physiological conditions of the vocal folds mucosa are connected to specific perceptual parameters. For instance, rigidity of the mucosa is related to rough voices while the corrugation is related to hoarse voices [20].

The separation of a voice signal into the harmonic and noise components, a concept named harmonic-noise splitting, has been widely studied in speech science during the past three decades. In most of the methods described in the literature, the (time-domain) signal to be processed is represented by the speech pressure waveform captured by a microphone but the processing can be also performed for the glottal flow waveform. Yegnanarayana et al. developed an algorithm based on a frequency domain approach [22]. In their method, harmonic regions of the speech pressure signal are defined by the harmonic peaks and the noise regions

* Corresponding author. Tel.: +351 22 508 1471.
E-mail addresses: sousa.ricardo@fe.up.pt, dee05004@fe.up.pt (R. Sousa), ajf@fe.up.pt (A. Ferreira), paavo.alku@tkk.fi (P. Alku).
[1] Tel.: +351 22 508 1471.
[2] Tel.: +358 9 47025680; fax: +358 9 460 224.

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.12.004

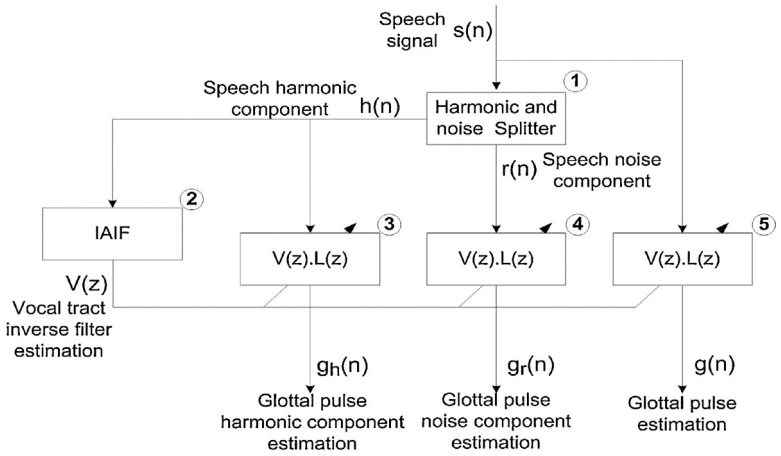

Fig. 1. Main block diagram of glottal harmonic-noise splitter. Signals $s(n)$, $h(n)$ and $r(n)$ denote, respectively, the speech signal and its harmonic and noise components. Signals $g(n)$, $g_h(n)$ and $g_r(n)$ denote, respectively, the glottal excitation, and its harmonic and noise components. $V(z)$ denotes the vocal tract transfer function. IAIF denotes the glottal inverse filtering algorithm described in Alku [24] and Alku et al. [25].

correspond to the inter-harmonic valleys and regions where harmonics are absent. A similar method was suggested by Jackson and Shadle [23] who used a comb filter structure to separate the harmonic and noise regions of the speech spectrum. In this method, the fundamental frequency of speech needs to be estimated prior to comb filtering in order to find the harmonic positions. Stylianou proposed a harmonic-noise splitting algorithm based on the assumption that there is a cut-off frequency that separates the speech spectrum into a low-frequency band and high-frequency band [19]. In his method, it is assumed that the low-frequency part contains mainly harmonic component information and the high-frequency contains noise information.

In this paper, two techniques are combined to yield an algorithm that estimates the harmonic and noise components of the glottal pulse. These techniques take advantage of the harmonic-noise splitting which decomposes the signal into a harmonic and noise component, and the inverse filtering which removes the contribution of the vocal tract. The application of the harmonic-noise splitting technique to the signal followed by inverse filtering gives rise to better glottal pulse estimations. This new algorithm was tested with synthetic voices in order to assess the accuracy of the method, and was also tested with natural voices in order to characterize the algorithm behavior against an acoustic diversity.

2. The splitting algorithm

2.1. Algorithm overview

The main goal of the study is to develop an algorithm that splits the waveform of the estimated glottal airflow velocity into a harmonic and a noise component. The block diagram of the method is shown in Fig. 1.

The algorithm consists of the following main phases. First (block 1), the speech pressure signal is divided into a harmonic and a noise component using a method that is described in detail in the following section. It is worth emphasizing that this harmonic noise splitting takes place prior to the estimation of the glottal airflow, a choice which is motivated by the fact that the estimation of the glottal airflow with inverse filtering deteriorates if the signal involves a significant amount of noise. Secondly (block 2), the obtained harmonic component of the speech signal, denoted by $h(n)$ in Fig. 1, is used as an input to the glottal inverse filtering which yields an estimate of the vocal tract inverse filter (an FIR filter), denoted by $V(z)$ in Fig. 1. Inverse filtering based on all-pole modeling is computed with a previously developed automatic algorithm, Iterative Adaptive Inverse Filtering (IAIF). For the detailed description of the IAIF method, the reader is referred to Alku [24] and Alku et al. [25]. Thirdly, this FIR filter is used in order to cancel the effects of the vocal tract from three signals: both from the harmonic and noise components obtained from the harmonic-noise splitter, and from the original speech pressure waveform. By further canceling the lip radiation effect using an integrator whose transfer function is simply given by $L(z) = 1/(1 - 0.99z^{-1})$, three glottal signals are finally obtained: the glottal pulse harmonic component, the glottal pulse noise component, and the glottal pulse, which are denoted in Fig. 1 by $g_h(n)$, $g_r(n)$, and $g(n)$, respectively. The underlying model of the complete method and the principles of the harmonic-noise splitter and inverse filtering methods will be described in the following sections. Both the harmonic-noise splitting and inverse filtering are linear operations. Eqs. (1)–(4) express the resulting signals in Fig. 1.

$$s(n) = h(n) + r(n) \tag{1}$$

$$g(n) = v(n) * \ell(n) * [h(n) + r(n)] \tag{2}$$

$$g(n) = v(n) * \ell(n) * h(n) + v(n) * \ell(n) * r(n) \tag{3}$$

$$g(n) = g_h(n) + g_r(n) \tag{4}$$

The parameters $v(n)$ and $\ell(n)$ denote the impulse response of the inverse model of the vocal tract and lip radiation effect, respectively. Eq. (1) represents the harmonic-noise model, which serves as the basis for the harmonic-noise splitter. Inverse filtering is represented by Eq. (2). Finally, Eqs. (3) and (4) show that the glottal excitation consists of harmonic and noise components.

The main advantage of the procedure depicted in Fig. 1 is the fact that it is very simple to be implemented once the harmonic and noise components of speech are split. However, it does not take into account non-linear phenomena in voice production.

2.2. Harmonic-noise splitter

The harmonic-noise splitter used in our study is based on a model of the harmonic structure of speech, which is parameterized in frequency, magnitude and phase. The block diagram of the harmonic-noise splitter is depicted in Fig. 2.

In the first stage (block no 1), the time domain input signal is transformed into the frequency domain using an Odd-Discrete Fourier Transform (ODFT) [27]. ODFT is obtained by shifting the