



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.

© 2013 Elsevier Ltd. All rights reserved.

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 vocal 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.

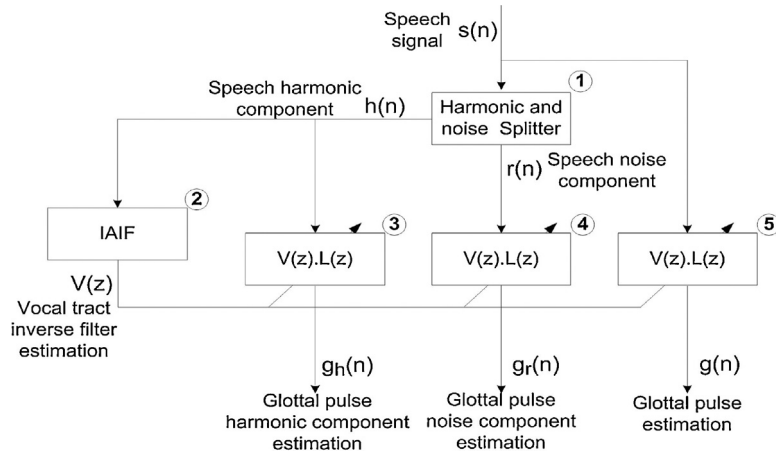


Fig. 1. Main block diagram of glottal harmonic-noise splitter. Signals $s(n)$, $h(n)$ and $r(n)$ denote, respectively, the speech signal and its harmonic and noise components. Signals $g(n)$, $g_h(n)$ and $g_r(n)$ denote, respectively, the glottal excitation, and its harmonic and noise components. $V(z)$ 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 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.

$$s(n) = h(n) + r(n) \quad (1)$$

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

$$g(n) = v(n) * \ell(n) * h(n) + v(n) * \ell(n) * r(n) \quad (3)$$

$$g(n) = g_h(n) + g_r(n) \quad (4)$$

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

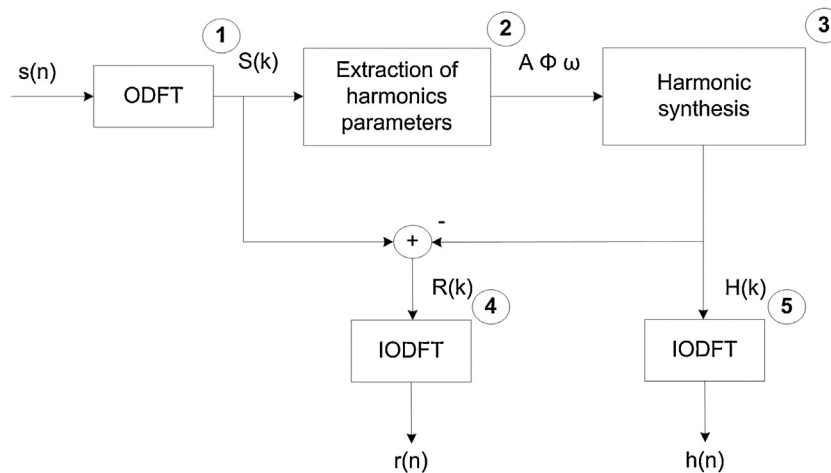


Fig. 2. Block diagram of the harmonic-noise splitter.

frequency index of the Discrete Fourier Transform (DFT) by half a bin [27]:

$$X_o(k) = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)(k+1/2)n}, \quad k = 0, 1, \dots, N-1 \quad (5)$$

where the time-domain input signal is denoted by $x(n)$ and the frame length is N . If $x(n)$ is real, this frequency shift makes the DFT samples above π a perfect mirror (in the complex conjugate sense) of the DFT samples below π which leads to more simple reconstruction equations. A peak picking algorithm is used to estimate the harmonics of the ODFT amplitude spectrum. This peak-picking algorithm was based on sliding window method over the spectral magnitude where the harmonics peaks are represented. The window has an odd number of points whose middle point is tested and if the middle point is largest value in the window, then a peak is identified. Note that the sliding window moves one point at a time along the spectral magnitude in order to test all points.

Next, the frequency, magnitude and phase of each harmonic are extracted (block 2) [27]. These parameters are then used to synthesize the spectrum of the harmonic structure of the input signal $s(n)$ (block 3). The spectrum of each individual sinusoid is synthesized using the parameters extracted from that harmonic. The harmonic synthesis equations are presented by Sousa [26].

Subsequently, the synthesized harmonic structure is subtracted from the signal $s(n)$ and the result is regarded as the noise component. Finally, the spectra of both components are inverse transformed in order to get time-domain representations for the harmonic and noise components (blocks 4 and 5).

An example of the harmonic-noise splitting algorithm is shown in the frequency domain in Fig. 3. The spectra are shifted vertically for visual clarity. The example demonstrates how the proposed method is able to estimate the harmonic structure of speech even in the inter-harmonic regions where the noise components prevail.

The main disadvantage is that some harmonics may be missing in the spectrum of the harmonic component due to errors caused by the peak picking algorithm. These occur typically for voices whose spectra show a significant amount of high-frequency noise or weak upper harmonics.

3. Performance tests

3.1. General overview

Experiments were conducted in order to assess the accuracy of the proposed algorithm and to compare its performance with a

previously known technique, the IAIF method. In order to achieve this, two major experiments were designed. The first one was based on synthetic voices and will be explained in the following section. In the second experiment, a database of natural voices was used in order to illustrate how the proposed algorithm performs with real speech. All the analyses were implemented in the Matlab environment. The IAIF algorithm was included by utilizing the TKK Aparat tool [28].

3.2. Synthetic voices

In order to generate synthetic speech material, a specific synthesizer was designed. The synthesizer is based on the source-filter model, but it also involves a harmonic-noise model. A block diagram of the synthesizer is shown in Fig. 4.

The synthesizer uses the Liljencrants–Fant (LF) model as the parametric representation to mimic the differentiated glottal flow. The LF model is among to most prevalent synthetic voice source models and it enables modeling a large variety of glottal pulse waveforms [29]. In order to create a glottal excitation with a desired F0, a single glottal pulse is first generated (block 1) and then

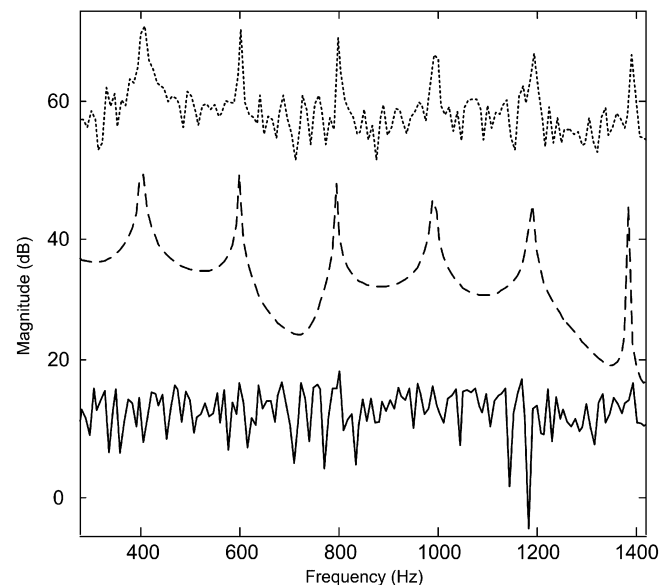


Fig. 3. Complete signal (dotted), harmonic (dashed) and noise spectra (solid). The signals are shifted along the magnitude axis to facilitate visualization.

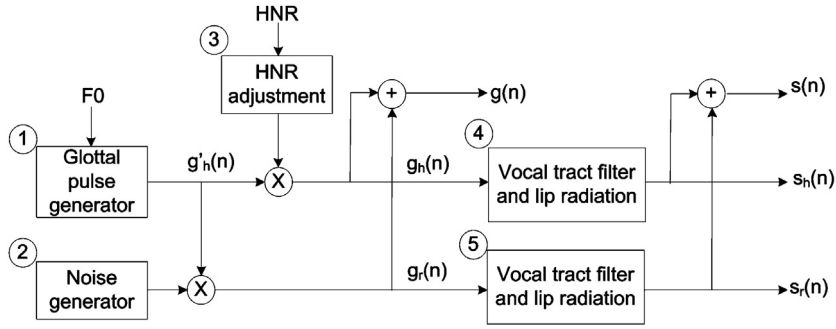


Fig. 4. Generation of synthetic voice signals. Signals $s(n)$, $s_h(n)$ and $s_r(n)$ denote, respectively, the speech signal and its harmonic and noise components. Signals $g(n)$, $g_h(n)$ and $g_r(n)$ denote, respectively, the glottal excitation and its harmonic and noise components. Signal $g'_h(n)$ is the harmonic component of the glottal excitation that is used to adjust the HNR.

concatenated to obtain a waveform with a train of glottal pulses. The resulting harmonic waveform is denoted by $g'_h(n)$ in Fig. 4. In order to produce a synthetic voice with the desired HNR, signal $g'_h(n)$ is multiplied by factor α (block 2). The value of α is determined from

$$HNR_{desired} = \frac{\sum_{n=1}^N [\alpha \cdot g'_h(n) * vr(n)]^2}{\sum_{n=1}^N [g_r(n) * vr(n)]^2} = \alpha^2 \cdot HNR_{current} \quad (6)$$

where $vr(n)$ is the impulse response of the filter that represents the vocal tract and lip radiation effect, N is the number of samples, $HNR_{desired}$ is the desired HNR value and $HNR_{current}$ is the value with no adjustment ($\alpha = 1$). To generate the noise component of the glottal excitation, a white Gaussian noise sequence is first produced (block 3). In order to modulate the noise component, the generated noise sequence is multiplied in the time domain by the LF modeled glottal pulse. A similar approach was utilized in a study by Yegnanarayana et al. [22], who used as a multiplier a rectangular pulse centered at the instant of glottal closure instead of the LF waveform. The resulting modulated noise sequence is denoted by $g_r(n)$ in Fig. 4. In the final stage, $g_r(n)$ and $g'_h(n)$ are summed up, resulting in a glottal excitation waveform denoted by $g(n)$. The two components are filtered (blocks 4 and 5) with the vocal tract and lip radiation filters in order to produce three outputs: the voice signal, and its harmonic and noise counterparts which are denoted, respectively, in Fig. 4 by $s(n)$, $s_h(n)$, and $s_r(n)$.

The synthesizer was used to generate a set of test vowels. The fundamental frequency F_0 was varied from 100 Hz up to 400 Hz with an increment of 10 Hz, in order to mimic both male and female speech. For each pitch, several vowel instances were generated by varying HNR from 9 dB up to 21 dB with an increment of 1 dB. The HNR is acquired as:

$$HNR = 10 \times \log_{10} \left(\frac{E_h}{E_r} \right) \quad (7)$$

E_h and E_r denote, respectively, the energy of the harmonic component, $s_h(n)$, and the noise component, $s_r(n)$, of synthetic speech. The values of the LF model were selected according to Gobl [9] in order to involve three different phonation types (breathy, normal and pressed). The vocal tract filter was adjusted to synthesize the vowel [a] ($F_1 = 664$ Hz, $F_2 = 1027$ Hz, $F_3 = 2612$ Hz). All the data were generated using the sampling frequency of 22.05 kHz.

3.3. Performance assessment

To assess the performance of both IAIF and the proposed algorithm, the estimated waveforms were compared objectively with the LF waveform using features extracted from the flow waveforms and their derivatives. By referring to Fig. 4, this implies that glottal excitations computed from signal $s(n)$ were compared with the

ideal excitation represented by signal $g'_h(n)$. The selected voice source parameterization methods were the Normalized Amplitude Quotient (NAQ) and the difference (in dB) between the amplitudes of the first and second harmonic (DH12). The NAQ parameter is a time-based parameter that is extracted for each glottal pulse and it measures the pressedness of phonation from the ratio of the peak-to-peak flow and the negative peak amplitude of the flow derivative [30]. The DH12 parameter is a frequency domain quantity and it measures the decay of the voice source spectrum [31,32]. Both parameters are independent of time and amplitude shifts. The relative error, defined in percentage according to Eq. (8), was used for NAQ since this parameter is a time-domain quantity that is typically measured on the linear scale.

$$NAQ_{Rel} = \left| \frac{NAQ_o - NAQ_m}{NAQ_o} \right| \times 100 \quad (8)$$

In Eq. (8), NAQ_o is the original NAQ value computed from the LF pulse used in the sound synthesis. NAQ_m parameter is the measured NAQ value computed as a mean of NAQ values extracted from individual glottal pulses of the estimated glottal excitation.

The absolute error, defined in dB according to Eq. (8), was used for DH12 because this parameter is typically expressed in the dB scale.

$$DH12_{Abs} = |DH12_o - DH12_m| \quad (9)$$

In Eq. (9), $DH12_o$ is the value of DH12 obtained from the original glottal source modeled by the LF waveform. $DH12_m$ is the value of DH12 measured from the estimated glottal excitation waveform.

3.4. Natural voices

The proposed algorithm was also tested with real speech. The database used included 39 sustained waveforms of the vowel [a] uttered by 13 subjects (7 males, 6 females) using breathy, normal and pressed phonation. The data were sampled with 22.050 kHz and a resolution of 16 bits. From these signals, the most stable segments with duration of 200 ms were selected for the voice source analysis.

4. Results

4.1. Experiments with synthetic voices

This section presents the results that were obtained for synthetic voices when the glottal source was estimated with IAIF and the proposed method. The NAQ error and DH12 error were determined separately for each phonation type. In order to compress the results, a set of ranges were defined for F_0 and HNR and the individual values obtained inside these ranges were pooled together. For F_0 , the following three ranges were used: 100–200 Hz, 210–300 Hz,

Table 1

NAQ mean relative error (%) for IAIF and the proposed method in the analysis of pressed synthetic voices.

F0 (Hz)	HNR (dB)					
	IAIF			Proposed method		
	9–15	16–21	22–27	9–15	16–21	22–27
100–200	27.8	14.8	22.6	13.0	11.2	15.5
210–300	52.8	27.5	75.6	21.2	38.4	60.4
310–400	64.7	68.9	131.3	55.9	101.1	151.0

Table 2

DH12 mean absolute error (dB) for IAIF and the proposed method in the analysis of pressed synthetic voices.

F0 (Hz)	HNR (dB)					
	IAIF			Proposed method		
	9–15	16–21	22–27	9–15	16–21	22–27
100–200	4.6	1.4	0.8	1.0	0.5	0.4
210–300	14.3	3.6	4.0	4.7	2.4	2.0
310–400	15.0	15.1	7.8	12.3	4.7	5.9

and 310–400 Hz. The first two ranges correspond to typical pitch used by males and females, respectively. The third range represents F0 values typical in voices produced by children. For HNR, the following three categories were used: 9–15 dB, 16–21 dB, and 22–27 dB. The first of these is typical for pathological voices while the second is characteristic to normal speech [33–35]. The last HNR range is related to voices which are highly periodic with a small amount of noise, such as the singing voice [36]. For each phonation type, the results are organized in tables that show the performance of NAQ or DH12 for the selected F0 and HNR ranges.

Tables 1 and 2 show that the proposed algorithm yields smaller DH12 errors for all the F0 and HNR combinations analyzed from pressed vowels.

The mean NAQ error was smaller with the proposed method also for all the F0 and HNR combinations except for three cases (F0 ranges 210–300 Hz and 310–400 Hz combined with HNR range of 16–21 dB; F0 range 310–400 Hz combined with HNR range 22–27 dB).

Tables 3 and 4 indicate that the proposed method yielded smaller errors for all the F0 and HNR ranges in the NAQ measurements in modal phonation.

For the DH12 error, the proposed method yielded larger distortion than IAIF only in two cases (F0 range of 100–200 Hz combined with the HNR range of 16–21 dB; F0 range of 310–400 Hz combined with HNR range of 9–15 dB).

Tables 5 and 6 show results from breathy voices that are in line with those observed for modal phonation: the mean NAQ error is smaller for the proposed method for all the F0 and HNR categories analyzed and the mean DH12 error was also smaller with the proposed algorithm in comparison to IAIF for

Table 3

NAQ mean relative error (%) for IAIF and the proposed method in the analysis of modal synthetic voices.

F0 (Hz)	HNR (dB)					
	IAIF			Proposed method		
	9–15	16–21	22–27	9–15	16–21	22–27
100–200	38.2	21.3	9.3	14.2	8.0	4.7
210–300	68.9	38.2	16.7	24.4	11.4	10.8
310–400	68.5	54.5	36.5	38.3	24.0	28.0

Table 4

DH12 mean absolute error (dB) for IAIF and the proposed method in the analysis of modal synthetic voices.

F0 (Hz)	HNR (dB)					
	IAIF			Proposed method		
	9–15	16–21	22–27	9–15	16–21	22–27
100–200	7.2	0.9	0.8	1.6	1.4	0.7
210–300	15.7	6.4	3.8	5.4	1.0	1.9
310–400	9.4	16.3	11.9	16.9	4.0	2.9

Table 5

NAQ mean relative error (%) for IAIF and the proposed method in the analysis of breathy synthetic voices.

F0 (Hz)	HNR (dB)					
	IAIF			Proposed method		
	9–15	16–21	22–27	9–15	16–21	22–27
100–200	56.9	37.0	16.5	25.8	11.6	12.0
210–300	77.9	68.2	23.9	46.9	17.9	13.3
310–400	83.8	80.7	45.8	54.4	31.6	18.9

all the F0 and HNR combinations except for few cases (F0 range of 100–200 Hz combined with the HNR ranges of 16–21 dB and 22–27 dB; F0 range of 210–300 Hz combined with HNR range of 22–27 dB).

In summary, the results obtained for the synthetic vowels show that the proposed method yields smaller mean NAQ and DH12 errors for the majority of the sounds analyzed. In particular, we highlight that the proposed method yields improved estimation accuracy in conditions with large amount of noise and for high-pitch voices. This accuracy improvement depends on the phonation type being more pronounced for modal voices.

4.2. Experiments with natural voices

In this section, representative examples of glottal excitations computed from natural speech by the proposed algorithm are shortly demonstrated. Results are shown in the form of time-domain waveforms by involving both the harmonic and the noise component yielded by the novel inverse filtering method.

Figs. 5 and 6 show waveforms computed from utterances produced by a male and female speaker, respectively. From both of these figures one can observe that the harmonic component is smoother than the glottal excitation waveform. In addition, low frequency fluctuations are not present in the harmonic component and the noise component indicates amplitude perturbations at the instants of glottal closure.

Table 6

DH12 mean absolute error (dB) for IAIF and the proposed method in the analysis of breathy synthetic voices.

F0 (Hz)	HNR (dB)					
	IAIF			Proposed method		
	9–15	16–21	22–27	9–15	16–21	22–27
100–200	9.8	4.6	2.4	5.3	5.1	4.3
210–300	32.8	24.3	4.5	15.7	6.7	5.5
310–400	21.0	28.2	13.3	20.8	9.1	5.7

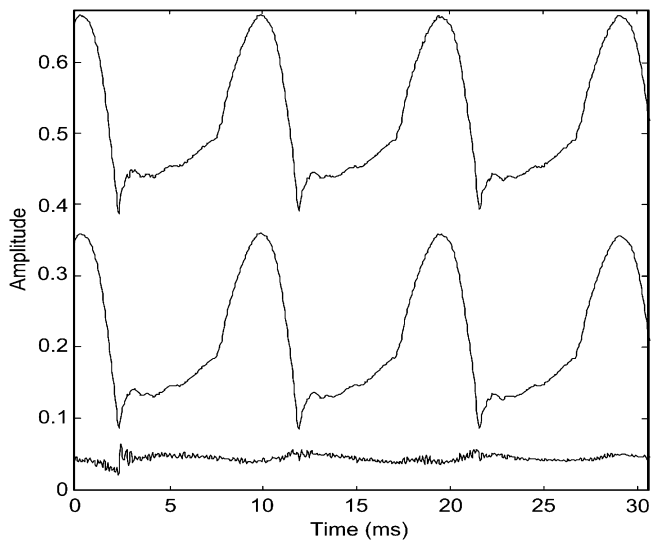


Fig. 5. Glottal excitation (top), its harmonic (middle) and noise (bottom) components estimated with the proposed method. A natural vowel [a] produced by a male speaker was used. The noise waveform is magnified 3 times for visual clarity.

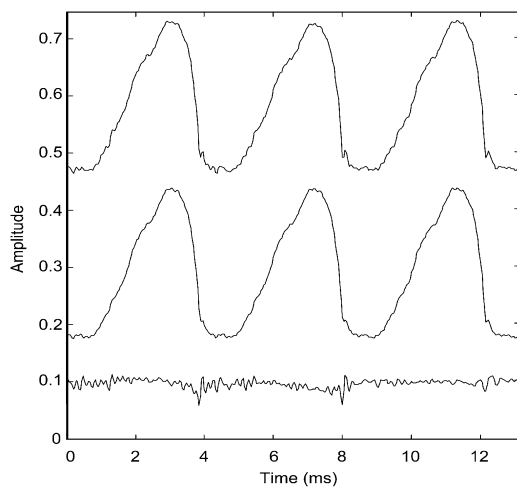


Fig. 6. Glottal excitation (top), its harmonic (middle) and noise (bottom) components estimated with the proposed method. A natural vowel [a] produced by a female speaker was used. The noise waveform is magnified 3 times for visual clarity.

5. Conclusions

In this article, a method to estimate the glottal excitation based on a known automatic inverse filtering method, IAIF, and a harmonic-noise splitter was proposed. The new method was compared with IAIF in the estimation of the glottal excitation using experiments with both synthetic and natural vowels.

Results obtained with synthetic voices show that the proposed method improves the estimation of the glottal waveform. The harmonic component given by the new algorithm is a more accurate estimate of the glottal source because the method is able to suppress the influence of noise which is always present in natural speech, particularly in pathological voices. For voices with low noise levels, this method may not be necessary and the performance of IAIF and the proposed algorithm are similar. The behavior of both algorithms was tested as a function of the noise level and fundamental frequency. It was found that both IAIF and the proposed algorithm show good accuracy for voices with low fundamental frequency and high HNR. Drawbacks of the proposed method are due to the harmonic-noise splitter, which may pass

noise to the harmonic component and itself is also sensitive to the noise level.

The proposed method also enables joint estimation of the harmonic and noise components of the glottal waveform. These components may be used in the evaluation of pathological voices since the separation enables characterizing the vocal folds dynamics as a function of noise produced in the speech production process. In addition, the noise component estimated by the proposed method can be used in speech technology in order to improve the naturalness of synthetic speech.

Acknowledgement

This work has been developed in the context of a doctoral program supported by the Portuguese Foundation for Science and Technology under grant SFRH/BD/24811/2005.

References

- [1] E.B. Holmberg, R.E. Hillman, J.S. Perkell, Glottal airflow and translottal air pressure measurements for male and female speakers in soft, normal, and loud voice, *J. Acoust. Soc. Am.* 84 (1988) 511–529.
- [2] P. Alku, M. Airas, E. Björkner, J. Sundberg, An amplitude quotient based method to analyze changes in the shape of the glottal pulse in the regularization of vocal intensity, *J. Acoust. Soc. Am.* 120 (2006) 1052–1062.
- [3] F.S. Hodge, R.H. Colton, R.T. Kelley, Vocal intensity characteristics in normal and elderly speakers, *J. Voice* 15 (2001) 503–511.
- [4] M. Airas, P. Alku, M. Vainio, Laryngeal voice quality changes in expression of prominence in continuous speech, in: *Proceedings of the 5th International Workshop on Models and Analysis of Vocal Emissions in Biomedical Applications (MADEBA 2007)*, 2007, pp. 135–138.
- [5] C. Gobl, A preliminary study of acoustic voice quality correlates, *STL-QPSR* (1989) 9–21.
- [6] C. Drioli, F. Avanzini, Hybrid parametric-physiological glottal modeling with application to voice quality assessment, *Med. Eng. Phys.* (2002) 453–460.
- [7] J. Laver, *The Phonetic Description of Voice Quality*, 1st ed., Cambridge University Press, Cambridge, 1980.
- [8] M. Airas, P. Alku, Emotions in vowels segments of continuous speech: analysis of the glottal flow using the normalizing amplitude quotient, *Phonetica* 63 (2006) 26–46.
- [9] C. Gobl, A. Chasaide, The role of voice quality in communicating emotion, mood and attitude, *Speech Commun.* 40 (2003) 189–212.
- [10] K. Scherer, Vocal communication of emotion: a review of research paradigms, *Speech Commun.* 40 (2003) 227–256.
- [11] P. Gomez-Vilda, R. Fernandez-Baillo, V. Rodellar-Biarge, V. Luis, A. Alvarez-Marquina, L. Mazaira-Fernandez, R. Martinez-Olalla, J.I. Godino-Llorente, Glottal source biometrical signature for voice pathology detection, *Speech Commun.* 51 (2009) 759–781.
- [12] J. Lee, M. Hahn, Automatic assessment of pathological voice quality using higher-order statistics in the LPC residual domain, *EURASIP J. Adv. Signal Process.* (2009) 1–9.
- [13] T. Drugman, T. Dubuisson, T. Dutoit, On the mutual information between source and filter contributions for voice pathology detection, *EURASIP J. Adv. Signal Process.* (2009) 1463–1466.
- [14] C. Ishi, H. Ishiguro, N. Hagita, Automatic extraction of paralinguistic information using prosodic features related to F0, duration and voice quality, *Speech Commun.* 50 (2008) 531–543.
- [15] S. Awan, N. Roy, Acoustic prediction of voice type in women with functional dysphonia, *J. Voice* 19 (2005) 268–282.
- [16] Y. El-Imam, On the assessment and evaluation of voice hoarseness, *Biomed. Signal Process. Control* (2008) 283–290.
- [17] G. De Krom, Spectral correlates of breathiness and roughness for different types of vowel fragments, in: *Third International Conference on Spoken Language Processing (ICSLP)*, 1994, pp. 1471–1474.
- [18] R. Van Dinther, R. Veldhuis, A. Kohlrausch, Perceptual aspects of glottal-pulse parameter variations, *Speech Commun.* 46 (2005) 95–112.
- [19] Y. Stylianou, Applying the harmonic plus noise model in concatenative speech synthesis, *IEEE Trans. Speech Audio Proc.* 9 (2001) 21–29.
- [20] S. Pinho, *Fundamentos em fonoaudiologia-tratando os distúrbios da voz*, 2nd ed., Guanabara, São Paulo, 2003.
- [21] W. Hanson, F. Emanuel, Spectral noise and vocal roughness relationships in adults with laryngeal pathology, *J. Commun. Disord.* 12 (1979) 113–124.
- [22] B. Yegnanarayana, C. d'Alessandro, V. Darsinos, An iterative algorithm for decomposition of speech signals into periodic and aperiodic components, *IEEE Trans. Speech Audio Proc.* 6 (1998) 1–11.
- [23] P. Jackson, C. Shadle, Performance of the pitch-scaled harmonic filter and applications in speech analysis, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* 3 (2000) 1311–1324.

- [24] P. Alku, Inverse filtering of the glottal waveform using the Itakura-Saito distortion measure, in: *Second International Conference on Spoken Language Processing (ICSLP)*, 1992, pp. 847–850.
- [25] P. Alku, J. Vintturi, E. Vilkman, On the linearity of the relationship between the sound pressure level and the negative peak amplitude of the differentiated glottal flow in vowel production, *Speech Commun.* 28 (1999) 269–281.
- [26] R. Sousa, A new accurate method of harmonic-to-noise ratio extraction, in: *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, 2009, pp. 351–356.
- [27] A. Ferreira, Accurate estimation in the ODFT domain of the frequency, phase and magnitude of stationary sinusoids, in: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 47–50.
- [28] M. Airas, TTK Aparat: environment for voice inverse filtering and parameterization, *Log. Phon. Vocol.* 33 (2008) 49–64.
- [29] G. Fant, J. Liljencrants, Q. Lin, A four parameter model of glottal flow, *STL-QPSR* 1 (1985) 1–13.
- [30] P. Alku, J. Vintturi, E. Vilkman, Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation, *Speech Commun.* 38 (2002) 321–334.
- [31] I. Titze, J. Sundberg, Vocal intensity in speakers and singers, *J. Acoust. Soc. Am.* 91 (1992) 2936–2946.
- [32] M. Swerts, R. Veldhuis, The effect of speech melody on voice quality, *Speech Commun.* 33 (2001) 297–303.
- [33] J. Godino-Llorente, P. Gomez-Vilda, F. Cruz-Roldan, M. Blanco-Velasco, R. Fraile, Pathological likelihood index as a measurement of the degree of voice normality and perceived hoarseness, *J. Voice* 15 (2009) 1–11.
- [34] A. Dehqan, H. Ansari, M. Bakhtiar, Objective voice analysis of Iranian speakers with normal voices, *J. Voice* 24 (2010) 161–167.
- [35] H. Leeper, V. Parsa, D. Jamieson, H. Heeneman, Acoustical aspects of vocal function following, *J. Voice* 16 (2002) 289–302.
- [36] J. Selby, H. Gilbert, J. Lerman, Perceptual and acoustic evaluation of individuals with laryngopharyngeal reflux pre- and post-treatment, *J. Voice* 17 (2003) 557–570.