AN ANALYSIS OF ITERATIVE ALGORITHM FOR ESTIMATION OF HARMONICS-TO-NOISE RATIO IN SPEECH

A. Stráník, R. Čmejla

Department of Circuit Theory, Faculty of Electrical Engineering, CTU in Prague

Abstract

Acoustic analysis of speech is a noninvasive technique that has been proven to be an effective tool for the objective speech assessment. In pathological speech (for example hoarseness) a harmonic-to-noise ratio is one of the most frequently used parameter because it can reveal an additive noise in voiced parts of speech. Additive noise is a result of leak of a glottal closure during phonation which can be a consequence of vocal edema or vocal polyps for example. This paper deals with an analysis of an iterative algorithm for the estimation of the noise component in speech.

1 Introduction

Pathological speech signals are commonly corrupted with additive noise and the energy of additive noise can be used as a parameter for determination of the level of speech pathology [1, 2]. Generally, the speech signal can be described as

$$x(k) = s(k) + w(k), \tag{1}$$

where $x(k)$ is a speech signal, $s(k)$ is a periodic part of speech generated by vocal folds and $w(k)$ is a noise part of speech generated by airflow from lungs. In normal (healthy) speech the component $w(k)$ is low and almost negligible compared to $s(k)$. In a pathological speech the energy of $w(k)$ increases due to an imperfect glottal closure which can be caused by, for example, vocal fold edema, polyp etc.

Well known and often used parameter *harmonics-to-noise ratio* (HNR) is defined as a ratio between $s(k)$ and $w(k)$

$$\mathrm{HNR} = 20 \log \left(\frac{En_{s(k)}}{En_{w(k)}} \right) \quad [\mathrm{dB}], \tag{2}$$

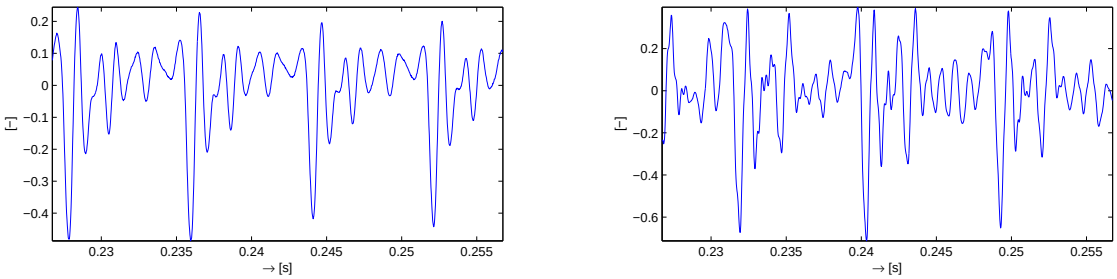
where $En_{s(k)}$ is the energy of the periodic component of speech and $En_{w(k)}$ is the energy of the noise component of speech.

There is no consensus on how to separate speech signal $x(k)$ to periodic and noise component. There are several ways: analysis in the time domain [1], frequency domain [2, 3], using wavelets [4] or cepstral analysis [5].

This article deals with an analysis of iterative algorithm for a noise component estimation in frequency domain published by [3] and its implementation in MATLAB.

2 Data

For testing purposes two signals were used – the first is a record of a healthy male and the second is a record of a male with functional dysphonia. Both signals contain a sustained phonation of vowel /a/ for cca 0.4 s.



(a) (b)

Figure 1: Example of test signals: (a) healthy, (b) functional dysphonia.

3 Iterative algorithm description

As mentioned above, this algorithm has been developed by YEGNANARAYANA et al. [3] and operates in the frequency domain. An input speech signal is segmented into microsegments the length of M samples and weighted by the Hamming window of the same length. The N-point DFT ($N > M$) is applied to every microsegment and spectrum $X(k)$ is obtained. In the amplitude spectrum $|X(k)|$ two types of regions are found, see Fig. 2:

- P_i – harmonic part of spectrum; contains both the periodic and the noise components of the input speech signal; the width of these regions corresponds to the length of DFT (N) and the length of Hamming window used for weighting of the microsegment (M): $2N/M$
- D_i – dip between harmonic parts; it is assumed that this part contains only the noise component of the input speech signal; to obtain non-empty dip region D_i with d points, the Hamming window length M should satisfy

$$M \geq \frac{4N}{f_0 NT - (d+1)}, \tag{3}$$

where M is the Hamming window length, N is the DFT length, f_0 is the fundamental frequency detected in the analysed microsegment, T is the sampling period ($1/f_s$) and d is the demanded number of points in dip region D_i.

Regions P_i and D_i can be identified as

$$P_i = \left\{ k | k_i - \frac{2N}{M} \leq k \leq k_i + \frac{2N}{M} \right\}, \tag{4}$$

$$D_i = \left\{ k | k_{i-1} + \frac{2N}{M} \leq k \leq k_i - \frac{2N}{M} \right\}, \tag{5}$$

where k is spectral line order and k_i is a position of and i-th harmonic region P_i.

After locating the regions P_i and D_i the iterative algorithm computes IDFT from spectrum with zeros at harmonic regions P_i and actual values at noise regions D_i. Then the N-point DFT is computed again, harmonic regions P_i are zeroed and so on, see Fig 3. After a few iterations (8 to 10 iterations according to [3]) the noise component is reconstructed with sufficient precision. To get the harmonic component in the time domain the reconstructed noise component has to be subtracted from original signal in the time domain.

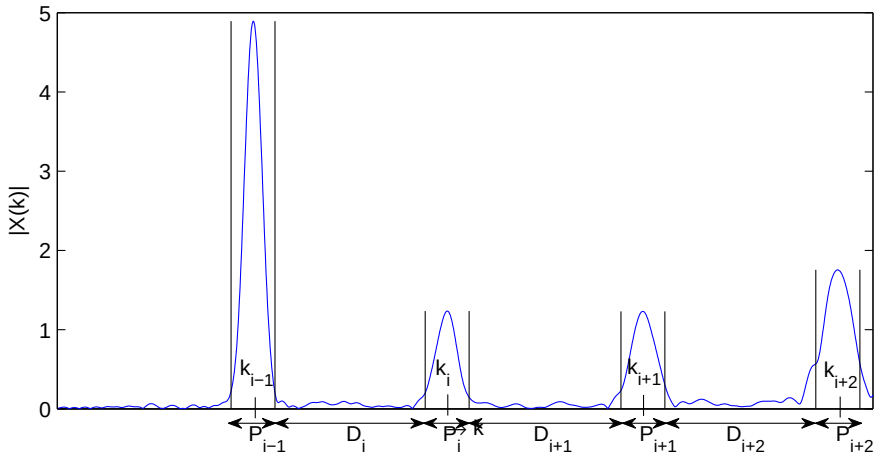


Figure 2: Description of harmonic part P_i and noise part D_i of the spectrum of a windowed voice speech segment.

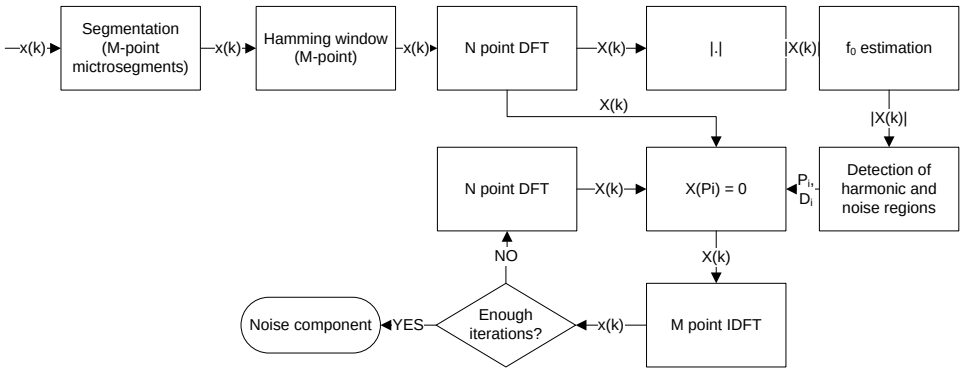


Figure 3: Block scheme of the iterative algorithm for noise component estimation.

4 Iterative algorithm analysis

An analysis of the algorithm focuses on the two main areas:

- f_0 detection and determination of harmonic and noise component in the frequency domain,
- the choice of M, N, d.

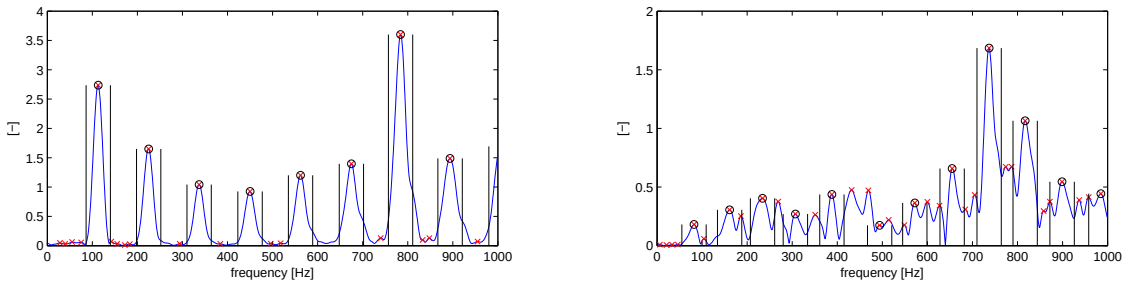
4.1 Harmonic and noise regions detection

The first step in the detection of the harmonic and the noise regions P_i and D_i is a f_0 detection – f_0 is supposed to be the main harmonic component in the speech signal. For this purpose, an amplitude spectrum is used and the first dominant peak is assumed to be the fundamental frequency f_0. The position k of f_0 is then used to determine P_i and D_i according to (4) and (5), see Fig. 4. Positions of the first harmonic regions in every microsegment are shown in Fig. 4.1.

4.2 Choice of M, N, d

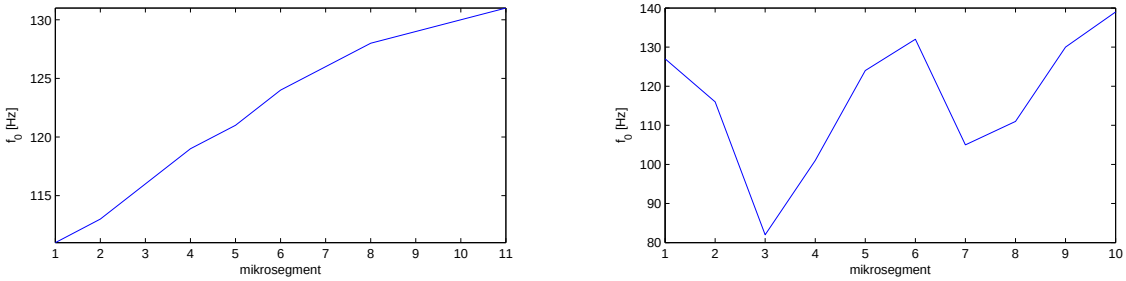
Practically, the window length M is fixed for the whole signal and cannot be changed at runtime, f_0 can be different in every microsegment, the only requirement on the parameter d is the non-zero size. The only parameter that can be changed during the calculation is the DFT length by zero-padding the input microsegment. Equation (3) has to be transformed to the following form

$$N \geq \frac{M(d+1)}{M f_0 T - 4}. \tag{6}$$



(a) (b)

Figure 4: Determination of harmonic regions P_i in amplitude spectrum for 4(a) healthy voice and 4(b) voice with functional dysphonia.



(a) (b)

Figure 5: Position of the first harmonic regions P_i in records with 4(a) healthy voice and 4(b) voice with functional dysphonia.

Equation (6) is not defined for

$$f_0 = \frac{4}{M_{\text{samples}} T} = \frac{4000}{M_{\text{ms}}} \tag{7}$$

which restricts the choice of the microsegment length. Fig. 6 shows the dependence of a critical f_0 on the microsegment length according to (7).

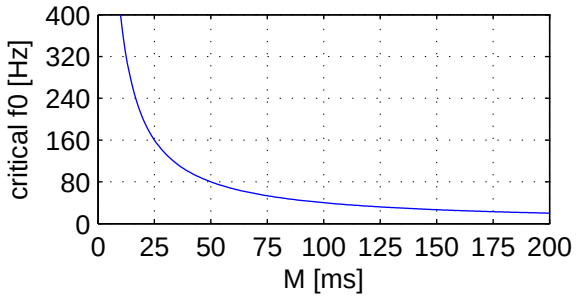


Figure 6: Dependence of critical f_0 on the microsegment length.

Fig. 7 shows dependence of DFT length N on detected f_0 while $d = 20$, M=100 ms and $f_s \in \{8, 16, 22.05, 44.1\}$ kHz. For $f_0 > 50$ Hz the required N is in acceptable range for all f_s.

Fig. 8 shows a block scheme of a modified algorithm which respects a different DFT length N for different f_0. First, f_0 with default N is estimated and if noise regions D_i in spectrum are empty due to the f_0 being too low, the smallest suitable N is computed and used for iterative noise component estimation. This modification lets the algorithm use smaller N as default and in case of high pitched voices or in case of pathological voices with unexpected voice breaks the required DFT length is adapted.

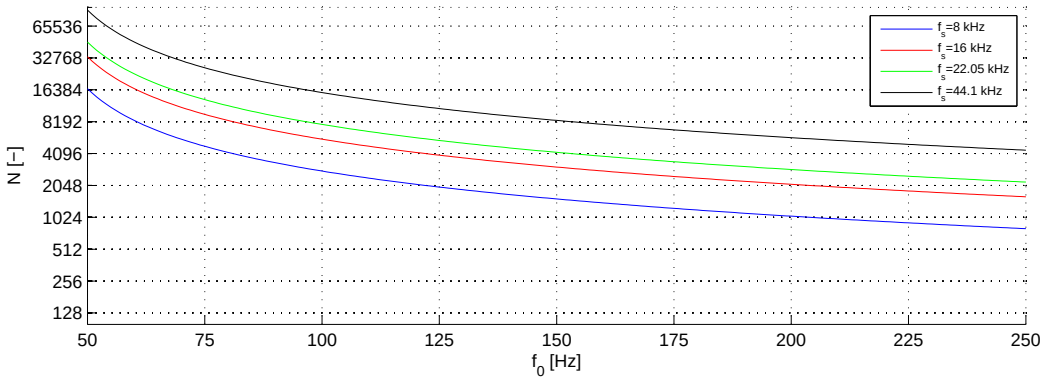


Figure 7: Dependence of DFT length on f_0.

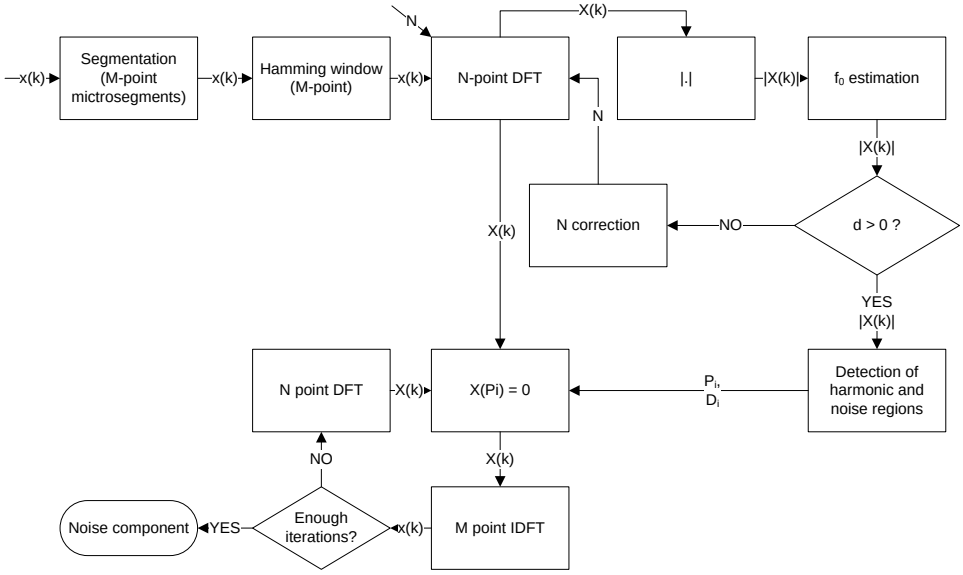


Figure 8: Block scheme of the modified iterative algorithm for noise component estimation.

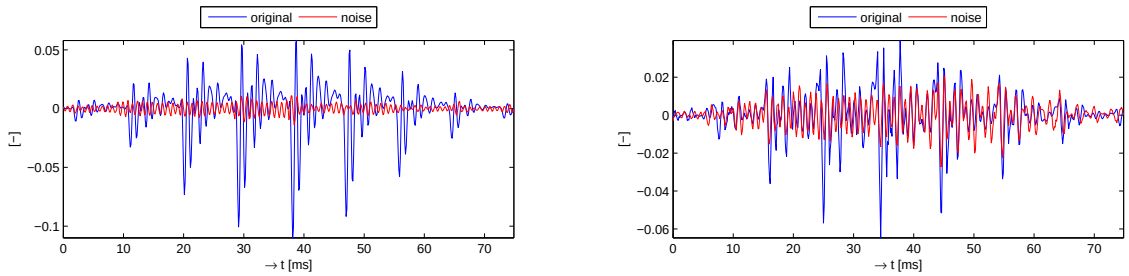
5 Results

Examples of noise components estimated in the test signals are shown in Fig. 9; input algorithm parameters are the following: f_s=8 kHz, M=80 ms (640 samples), d=20, default N=8192 samples.

Is is obvious that noise component energy of a healthy voice shown in Fig. 9(a) is smaller relative to overall energy than for a pathological voice depicted on Fig. 9(b). Also HNR is higher for the healthy voice, which is expected. A summary of results for both test records is shown in Tab 1.

Table 1: ESTIMATED HARMONICS-TO-NOISE RATIO IN TEST RECORDS.

	HNR [dB]
healthy	22.84 ± 4.29
functional dysphonia	2.88 ± 5.05



(a) (b)

Figure 9: Examples of estimated noise components for (a) healthy and (b) functional dysphonia in one microsegment.

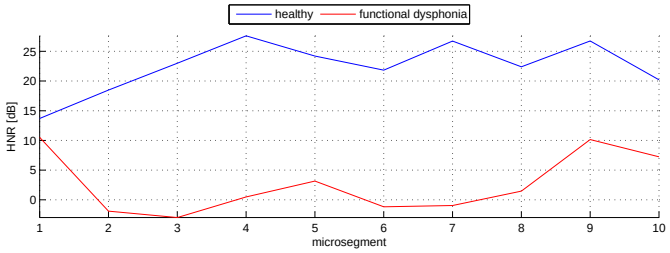


Figure 10: Estimated HNR in test records.

6 Conclusion

An implementation of modified iterative estimation of noise component in voiced parts of speech was introduced. The modification reflects various settings of DFT length for various fundamental frequency. Two records of sustained vowel /a/ were used for testing purposes. The first record contains a healthy voice and the second record contains a voice with functional dysphonia. In accordance with the assumption the noise component in the healthy voice is smaller than in the pathological one.

Acknowledgements

This work has been supported by: **GACR102/08/H008** *Biological and Speech Signal Modelling*, **SGS10/180/OHK3/2T/13** *Assessment of voice and speech impairment*, **MSM6840770012** *Transdisciplinary Research in Biomedical Engineering.*

References

[1] Eiji YUMOTO, SASAKI Yumi, and Hiroshi OKAMURA. Harmoics-to-noise ratio and physiological measurement of the degree of hoarseness. *JSHLR*, 27:2–6, 1984.

[2] Kumara SHAMA, Anantha KRISHNA, and Miranjan U. CHOLAYYA. Study of harmonics-to-noise ratio and critical-band energy specrtrum of speech as acoustic oindicators of laryngeal and voice pathology. *EURASIP J. Appl. Signal Process.*, pages 50–50, 2007.

[3] B. YEGNANARAYANA, Christophe d'ALESSANDRO, and Vassilis DARSINOS. An iterative algorithm for decomposition of speech signals into periodic and aperiodic components. *IEEE Transactions on Speech and Audio Processing*, 6(1):1–11, 1998.

[4] Claudia MANFREDI. Adaptive noise energy estimation in pathological speech signals. *Biomedical Engineering, IEEE Transactions on*, 47(11):1538 –1543, 2000. doi: 10.1109/10.880107.

[5] Peter J. MURPHY and Olatunji O. AKANDE. Quantification of glottal and voiced speech harmonics-to-noise ratios using cepstral-based estimation. In *NOLISP*, volume 3817, pages 150–160, 2005. doi: http://dx.doi.org/10.1007/11613107_13.

Adam Stráník
stranada@fel.cvut.cz

Roman Čmejla
cmejla@fel.cvut.cz