16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SEPARATION OF PERIODIC AND APERIODIC SOUND COMPONENTS BY EMPLOYING FREQUENCY ESTIMATION

Kristóf Aczél, István Vajk

Department of Automation and Applied Informatics, Budapest University of Technology and Economics
3-9. Műegyetem rkp. , H-1111, Budapest, Hungary
aczelkri@aut.bme.hu, vajk@aut.bme.hu

ABSTRACT

Separation of a sound to its two main components, periodic and aperiodic ones, is an area that has gained focus in the past years. Instrument classification, speech recognition and many other systems can benefit from such a representation of the digital sound. Unfortunately the separation of the two components is usually not a straightforward process, regarding that the sinusoidal and the noise-like parts of a waveform are not orthogonal, and can easily overlap in frequency. This article presents an efficient and elegant way of separating the two components. The algorithm operates in frequency domain and does the separation on the grounds of the frequency estimation method proposed by Brown. Being simple to implement, in the same time efficient in the task, it can be employed in many current DSP systems.

1. INTRODUCTION

Digital signal processing applications have become quite common and accessible for the public in the past decade. The list of applications employing DSP algorithms is already huge, and is growing each day. Audio processing is no exception. However, most audio algorithms have one thing in common: they usually do not work directly on the signal in time-domain, but transform the input signal into a representation that better suits the task.

Due to the nature of real-world signals, most tasks have to cope with the dual-component structure of sound: it is the composition of a periodic (or harmonic, deterministic) part and an aperiodic (or noise-like, stochastic) part. The recognition of the fact that the two components usually need to be handled in a completely different manner makes us look for solutions for isolating them from each other.

Our long term interest in periodic/aperiodic decomposition is motivated by the problem of correcting existing musical recordings, adjusting volumes of instruments separately, fixing misplayed notes etc. A system for such purposes was proposed in [2]. The system separates polyphonic music to isolated notes, thereby making the replacement of misplayed notes possible. However, it was found that the system had difficulties with the separation of notes in cases where one of the notes contained a strong noise-like component (e.g. drums) while other notes comprised mainly harmonic components. Under these circumstances the system generated incorrect output signals, resulting in strong noise component appearing in the output signal of a violin or harmonic component appearing in separated drum note. To overcome this issue it was inevitable to perform aperiodic/periodic decomposition on the polyphonic input signal before the actual separation takes place. Other areas that can also benefit from such a decomposition include sound compression, beat detection, speech recognition, formant adjustment, noise reduction – just to mention a few.

Speech signal decomposition has attracted a lot of research efforts in the recent past. In [5], [6] and [7] important contributions in the field are presented. The decomposition proposed here is performed on an approximation to the excitation signal, instead of decomposing the source signal directly. The linear prediction residual signal is used as an approximation to the excitation signal of the vocal tract system. Decomposition into periodic and aperiodic components is accomplished by first identifying the frequency regions of harmonic and noise components in the spectral domain. The signal corresponding to the noise regions is used as a first approximation to the aperiodic component. An iterative algorithm is proposed which reconstructs the aperiodic component in the harmonic regions. The periodic component is obtained by subtracting the reconstructed aperiodic component signal from the residual signal. The individual components of the residual are then used to excite the derived all-pole model of the vocal tract system to obtain the corresponding components of the speech signal.

A slightly different approach is introduced in [3]. The work discusses an analysis/synthesis method designed to obtain musically useful intermediate representations for sound transformations. The proposed method approximates the harmonic component by a series of sinusoids that are described by amplitude and frequency functions. These parameters are detected from STFT spectral peak trajectories. The stochastic component is represented by a series of magnitude-spectrum envelopes that function as a time-varying filter excited by white noise. The envelopes are calculated by subtracting the spectra of the harmonic component from the spectra of the original sound. These representations together make it possible for a resynthesized sound to attain all the perceptual characteristics of the original sound.

A similar approach was also proposed in [4]. The most important difference here is the estimation technique used for the approximation of the harmonic component. The approach makes use of a time-domain pitch-detector based on a normalized cross-correlation function.

Most contributions are based on parameter estimation of the harmonic part of the input signal. In our research we take a different approach. We still consider the signal to be the composition of a harmonic part and a residual, but our approach is not model-based in the sense that independent harmonic sources in the signal are not located. Instead of searching for harmonic components, we make the periodic/aperiodic decision on a lower level. STFT phases are used in the decision of each Fourier bin's periodicity status. The spectrum of the signal is then split into two spectra, one consisting of the bins previously considered periodic, the other consisting of bins previously considered aperiodic. This approach has the advantage of being reversible: it is possible to get back the exact same input signal by remixing the separated components.

The following sections provide detailed information on the basics of the decomposition algorithm. Section 2 introduces the conversion from time domain to frequency domain by using the Brown frequency estimator. Section 3 shows the details of the actual decomposition step. Then, section 4 deals with the fine-tuning of the parameters of the algorithm also showing synthetic test results, and finally, section 5 concludes.

2. TRANSFORMATION TO FREQUENCY DOMAIN

This section proposes an easy, yet powerful algorithm that is able to generate a spectrogram of the recording that is much more precise for musical analysis than the conventional STFT spectrogram.

Earlier literature [9], [10] covered different transformation methods in order to determine the best possible means for analysis of audio signals using STFT. Current research has examined the analysis of polyphonic musical signals in particular. STFT is known to be limited by the uncertainty principle: either we get fine time and poor frequency resolution or the other way around. For this reason many researches seek alternative methods to build a spectrogram-like representation of signals. However, with some modifications the original STFT algorithm can provide very precise results for the analysis of musical signals.

In [1] a frequency estimation method is shown that calculates true frequencies present in the original signal from subsequent phase values. For a frame starting at time t the FFT coefficients and phases are $c_{k,t}$ and $\varphi_{k,t}$, respectively. In this document the time index will be omitted in some of the equations for understandability. Two subsequent frames are needed by the algorithm for the calculation. Assuming that the frame starts at t_1 and ends at t_2 , a true frequency f_{k,t_2}^{true} can be computed for each bin. The frequency of the k[th] bin is

$$f_k = k\frac{samplerate}{framesize}. \tag{1}$$

The true frequency of each bin will deviate from this value:

$$f_{k,t_2}^{true} = f_k + \frac{\varphi_{k,t_2}^{dev}}{2\pi\cdot(t_2 - t_1)} \tag{2}$$

where

$$\varphi_{k,t_2}^{expt} = \varphi_{k,t_1} + (t_2 - t_1)\cdot 2\pi f_k \tag{3}$$

$$\varphi_{k,t_2}^{dev} = \varphi_{k,t_2} - \varphi_{k,t_2}^{expt} + l\cdot 2\pi\,, \tag{4}$$

where φ_{k,t_2} is the phase of bin k in time t_2; φ_{k,t_2}^{expt} is the expected phase; φ_{k,t_2}^{dev} is the deviance between the expected and measured phase; f_{k,t_2}^{true} is the estimated true frequency of bin k in time t_2 and $l \in Z : -\pi < \varphi_{k,t_2}^{dev} \le +\pi$. The greater the time difference (step length) between the start of the frames the more precise the estimated value of f_{k,t_2}^{true} . On the other hand, big time differences limit the maximum detectable distance between f_{k,t_2}^{true} and f_k. The selection of the right step length is covered in section 4.

3. PERIODIC/APERIODIC DECISION

Our approach decomposes the STFT spectra of the original signal on the level of the Fourier bins. Each Fourier bin will be considered periodic or aperiodic based on how the bin's current true frequency relates to its "frequency history" and "frequency future", that is, the true frequencies in previous and future frames. A measure is needed that reliably tells apart aperiodic, noise-like components from periodic components, even in cases when the frequency of the periodic signal is slowly but continuously changing (e.g. vibrato of a violin note). Measurements proved that observing quadratic deviances between subsequent past true bin frequencies provides a good hint about a bin's periodicity.

Let $f_{k,r\tau}^{true}$ denote the true frequency of the k^{th} bin for the frame starting at time $r\tau$ and $f_{k,(r-p)\tau}^{true}$ denote the true frequency of the same bin in the previous frames starting at $(r-p)\cdot\tau$. Let $d_{k,r\tau}$ denote the sum of the quadratic deviances between subsequent true frequencies in the last P frames:

$$d_{k,r\tau} = \sum_{p=-P}^{P}\left(f_{k,(r+p)\tau}^{true} - f_{k,(r+p-1)\tau}^{true}\right)^2. \tag{5}$$

The decision on the periodicity of one specific bin at time $r\tau$ can be based on the corresponding $d_{k,r\tau}$ value. If $d_{k,r\tau}$ is under a certain threshold, the bin is considered to be periodic, otherwise aperiodic. The periodicity detector function can be defined as:

$$\vartheta_{k,r\tau} = \begin{cases} 1 & d_{k,r\tau} < M \\ 0 & d_{k,r\tau} \ge M \end{cases} \tag{6}$$

where $\vartheta_{k,r\tau}$ denotes the periodicity status of bin k in time $r\tau$ and M is an experimental value defining the maximum sum quadratic deviation that is considered as periodic. Using this definition the input signal can be expressed as

$$\underline{\mathbf{c}} = \underline{\mathbf{c}}^{per} + \underline{\mathbf{c}}^{aper} \tag{7}$$

with

$$c_{k,r\tau}^{per} = \vartheta_{k,r\tau} \cdot c_{k,r\tau} \quad (8)$$

$$c_{k,r\tau}^{aper} = (1 - \vartheta_{k,r\tau}) \cdot c_{k,r\tau} \quad (9)$$

where $c_{k,r\tau}^{per}$ and $c_{k,r\tau}^{aper}$ denote the separated periodic and aperiodic components of the input signal.

The presented measure for periodicity provides a binary decision for each bin in each time frame. Although the measure is reliable for synthetic tests, real-life signals often pose a problem, especially for signals of high polyphony. Binary decision assumes that there are no components in the signal that overlap in frequency. However, real-life signals usually contain overlapping components, some of them periodic, others aperiodic. For a bin that holds energy originated from both periodic and aperiodic sources the periodicity detector will not be able to provide 100% accurate results. There may be cases when the periodicity status of one bin will oscillate between *periodic* and *aperiodic*. This causes audible artifacts in the two separated components.

This issue is overcome by extending the latent range of $\vartheta_{k,r\tau}$ decision function from {0,1} to [0,1], allowing a middle range in the decision between *periodic* and *aperiodic*. This range may be imagined as a certainty of periodicity between 0.0 and 1.0. The periodicity detector $\vartheta_{k,r\tau}(d_{k,r\tau})$ can now be expressed as a function with the following properties:

- Bins with constant true frequencies are periodic:
$$\vartheta(0) = 1$$
- Bins with periodicity above a certain threshold are considered aperiodic, under another threshold they are periodic ($M_2 \geq M_1 > 0$):
$$M_1 > x \quad \Rightarrow \quad \vartheta(x) = 1$$
$$M_2 < x \quad \Rightarrow \quad \vartheta(x) = 0$$
- ϑ is a monotonically decreasing function:
$$x_1 \leq x_2 \;\Rightarrow\; \vartheta(x_1) \geq \vartheta(x_2)$$

The original algorithm can also be further improved by emphasizing frequency differences from closer times. Using the weighted sum of squares of frequency deviations (5) can be expressed now as

$$d_{k,r\tau} = \sum_{p=-\infty}^{+\infty} \varsigma(p) \cdot \left(f_{k,(r+p)\tau}^{true} - f_{k,(r+p-1)\tau}^{true} \right)^2, \quad (10)$$

where ς is a function with similar recommended properties to the window-functions used at STFT calculation:

- The frequency deviation closest in time should have the largest coefficient
$$\varsigma'(0) = 0$$
- Frequency deviations farther in time than a certain *L* threshold are not considered in the calculation:
$$L < |x| \;\Rightarrow\; \varsigma(x) = 0$$
- ς monotonically increases for past deviations and decreases for future ones:
$$|x_1| < |x_2| \;\Rightarrow\; \varsigma(x_1) > \varsigma(x_2)$$

Although the above criteria are usually recommended for the periodicity detector, some cases may require a different ς function. One example is a signal with harmonic components and short noise bursts or sudden onsets. In this case past frequency deviations are of little use in the detection of the current frame's periodicity. Situations like this may require considering the use of a ς weighting function that neglects past deviations or at least favors future ones. The best type of weighting function should be chosen for each application.

4. FINE-TUNING OF THE ALGORITHM AND TEST RESULTS

To effectively employ the periodicity detector its parameters have to be optimized. This section covers some practical details of the algorithm. The separation is sensitive to the following parameters:

- Step length and frame size
- $\varsigma(x)$ weighting function
- $\vartheta(x)$ periodicity decision function

In our tests we used a CD-quality signal (44100 samples/sec) with a frame size of 2048 samples. Previous researches showed that shorter frames do not provide enough resolution in frequency for music analysis, while longer frames cannot reasonably follow fast spectral changes in our test signals.

We found smaller step sizes to be more suitable for periodicity detection. In our research a big overlap of 128 was used, which means that the delay between the starting time of two subsequent windows were 1/128 the length of the frame. Greater overlaps did not seem to produce audibly better results, while smaller values did decrease the decomposition quality significantly. Latter is due to the fact that the maximum detectable deviation of the f_k^{true} true frequency from the f_k bin frequency was too low, in other words the phase change $\varphi_{k,t_2} - \varphi_{k,t_1}$ on the bins was too fast.

In order to analyze the effect of the weighting function on the quality we built a set of synthetic test samples. Each sample consisted of a sine signal as the periodic component, to which we added white-noise with different spectral properties (different spectral energy distribution and/or magnitude) as the aperiodic component. Figure 1 shows the waveform of nine different noise components, while Figure 2 depicts the full test signal (sine + noises). Latter waveform was fed into the decomposition algorithm which was then carried out using weighting functions of different types and sizes. Finally the separated components were subtracted from the original ones in time-domain, resulting in a signal that contains only the separation error. Table 1 presents the RMS of the separation errors for different weighting functions.

Figure 3 shows the spectrogram of the previous nine test signals. Figure 4 and 5 show the separated aperiodic and periodic components, respectively. Note that the algorithm was able to perform the decomposition with acceptable result even for the last three signals. Although the image shows some grain in these cases (Figure 5), we must keep in mind

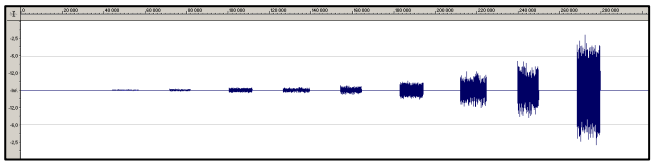

Figure 1: Waveform of test noise component

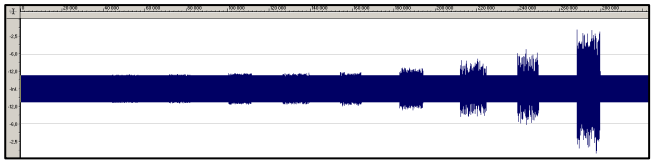

Figure 2: Waveform of full test signal: noise component added to 500Hz sine wave

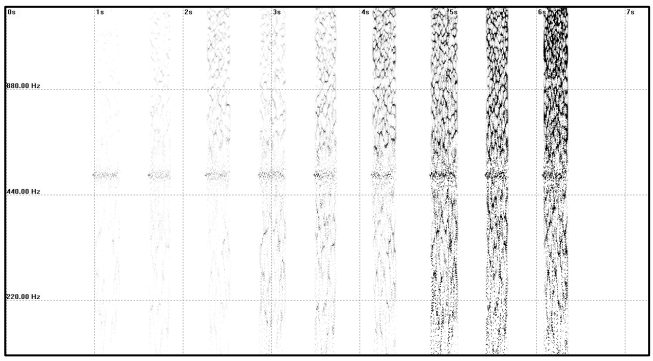

Figure 3: STFT spectrum of separated aperiodic component of the synthetic test signal

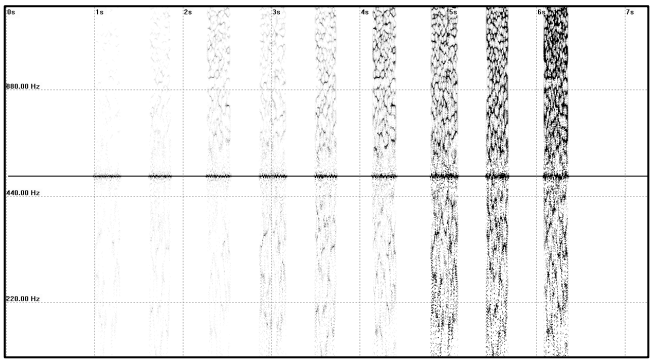

Figure 4: STFT spectrum of original synthetic test signals

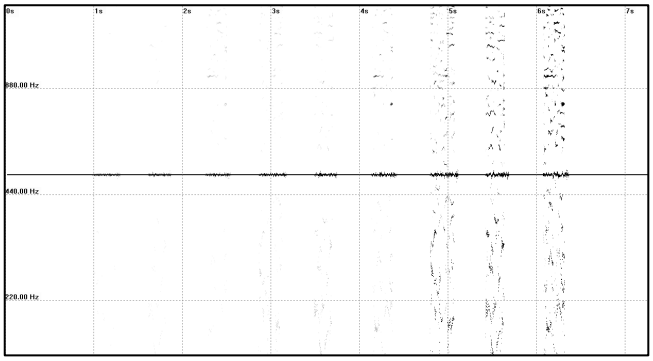

Figure 5: STFT spectrum of separated periodic component of the synthetic test signal

that the noise/harmonic ratio was 10dB, which means the original sine was hardly audible.

Table 1: Separation error (dB (RMS))

size/shape of $\varsigma(x)$	6ms	12ms	23ms	46ms
$\varsigma(x)=\begin{cases}1 & \text{if } \lvert x\rvert \le L/2\\ 0 & otherwise\end{cases}$	-24	-24,7	-26,1	-28,3
$\varsigma(x)=\begin{cases}1 & \text{if } 0\le x\le L\\ 0 & otherwise\end{cases}$	-24,2	-25,3	-27,1	-28,9
$\varsigma(x)=\begin{cases}x+L/2 & \text{if } \lvert x\rvert \le L/2\\ 0 & otherwise\end{cases}$	-25,4	-26,1	-27,2	-28,7
$\varsigma(x)=\begin{cases}x & \text{if } 0\le x\le L\\ 0 & otherwise\end{cases}$	-25,6	-26,7	-28,3	-29,8
$\varsigma(x)=\begin{cases}\lvert -x+L/2\rvert & \text{if } \lvert x\rvert \le L/2\\ 0 & otherwise\end{cases}$	-22,4	-23,3	-25,6	-26,6

The synthetic test results show that the quality of the periodicity detector mostly depends on the support range of the weighting function. Raw RMS values suggest that the larger the support range the better the decomposition quality will be. However, this is only true for signals where the frequency of the harmonic component does not change often. If the location of the harmonic component changes, new harmonic components appear or existing ones end, it will take some time till the periodicity detector reflects the change. This is shown on Figure 6-8. As our long term interest in separation mainly involves digital processing of musical signals, the algorithm was also tested on polyphonic music material. Of course in these cases the original harmonic and stochastic components are not available for comparison with the output. Human listeners were asked to evaluate the output results of the algorithm with different weighting functions. It was found that the best result was achieved by using a weighting function with a support range of 23 ms.

The optimal shape of the $\vartheta(x)$ periodicity decision function was also investigated. Using M for the periodicity threshold and $2m$ for denoting the range of uncertain decision the following types of decision functions were tested:

$$\vartheta(x)=\begin{cases}1 & \text{if } x\le M\\ 0 & \text{if } x>M\end{cases} \tag{11}$$

$$\vartheta(x)=\begin{cases}1 & \text{if } x\le M-m\\ \dfrac{M+m-x}{2m} & \text{if } M-m<x<M+m\\ 0 & \text{if } x\ge M+m\end{cases} \tag{12}$$

$$\vartheta(x)=\begin{cases}1 & \text{if } x\le M-m\\ 0{,}5\cdot\left(1+\cos\left(\dfrac{x-M+m}{2m}\cdot\pi\right)\right) & \text{if } M-m<x<M+m\\ 0 & \text{if } x\ge M+m\end{cases} \tag{13}$$

Human listeners were asked to evaluate the decomposition results. While test subjects reported inferior quality with au-

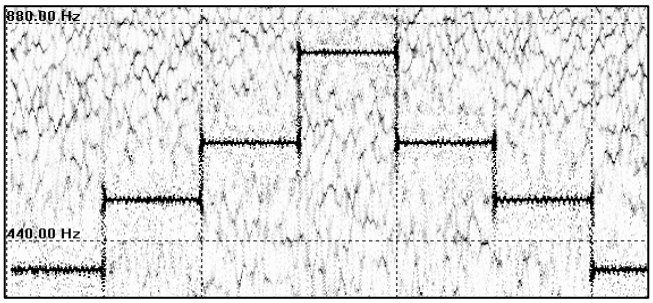

Figure 6: STFT spectrum of the original synthetic test signal 2

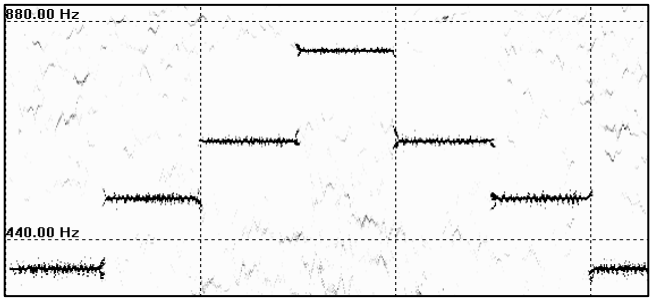

Figure 7: STFT spectrum of separated periodic component of the synthetic test signal 2

dible artifacts for (11), they were unable to differentiate between the output of (12) and (13). The quality was reported to be dependant mostly on M and $2m$. However, subjects could not always agree on optimal values, which is in accordance with the fact that periodicity here is in fact a very subjective term. For example, the onset of a sine wave that one describes as periodic can be considered aperiodic by others. The most favored values were

$$M \approx \frac{3{,}125 \cdot \sum_{x=-\infty}^{\infty} \varsigma(x)}{t_{frame} / \tau}, \qquad (14)$$

and

$$m \approx 0{,}2 \cdot M \ . \qquad (15)$$

Finding the ultimate parameters is for now out of the scope.

5. SUMMARY

The article proposed a simple, yet elegant and powerful algorithm for the decomposition of audio signals to periodic and aperiodic components. Test cases were presented that validated the basic idea of the approach. The synthetic test results showed that the separation was of relatively good quality even in cases when there was a considerably big amplitude difference between the harmonic and noise components.

Real-life test results can be downloaded from http://avalon.aut.bme.hu/~aczelkri/separation.

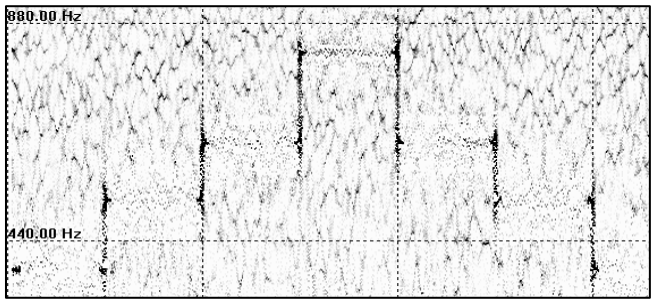

Figure 8: STFT spectrum of separated aperiodic component of the synthetic test signal 2

ACKNOWLEDGEMENTS

This work has been supported by the fund of the Hungarian Research Fund (grant number T68370).

REFERENCES

[1] Judith C. Brown, M.S. Puckett: "A high resolution fundamental frequency determination based on phase changes of the Fourier Transform", *J. Acoust. Soc. Am* 94 (2), pp. 662-667, 1993

[2] K. Aczél, Sz. Iváncsy, "Sound separation of polyphonic music using instrument prints", *15th European Signal Processing Conference (EUSIPCO 2007)*, Poznan, Poland, 2007

[3] Serra, X., J. O. Smith. 1990. "A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition". *Signal Processing V: Theories and Applications*, L. Torres, E. Masgrau, and M.A. Lagunas (eds.). Elsevier Science Publishers B.V., 1990.

[4] Mangui Liang, Qi Hu, Shouzeng Huang, "Decomposition of speech signal into a periodic and an aperiodic part", *Proceedings of the First International Conference on Innovative Computing, Information and Control 2006 (ICICIC 2006)*, Vol. 2, pp. 746 – 750

[5] d'Alessandro, C. Darsinos, V. Yegnanarayana, B., "Effectiveness of a periodic and aperiodic decomposition method for analysis of voice sources", *IEEE Transactions on Speech and Audio Processing*, 1998, pp. 12–23

[6] B. Yegnanarayana, Christophe d'Alessandro, Vassilis Darsinos, "An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components", *IEEE Transactions on Speech and Audio Processing*, 6 (1), 1998, pp. 1 – 11

[7] S. Ben Jebara, „Periodic/aperiodic decomposition for improving coherence based multi-channel speech denoising", *International Symposium on Signal Processing and Applications,* Sharjah, Emirates, 2007.

[8] Y. Stylianou, "Decomposition of Speech Signals into a Deterministic and a Stochastic Part", *Proceedings of ICSLP '96*, 1996, vol.2, pp. 1213 – 1216

[9] R. Pintelon, J. Schoukens, *System Identification, A frequency domain approach*, ISBN 0-7803-6000-1, Wiley-IEEE Press, May 2001, pp. 33-44

[10] S. Gade, H. Herlufsen, Use Of Weighting Functions in DFT/FFT Analysis (Part I), *Brüel & Kjær Technical Review,* No. 3., 1987