VOICE RANGE PROFILE AND PROBLEM OF FUNDAMENTAL FREQUENCY

FRANTIŠEK VÁVRA, PAVEL NOVÝ, HANA MAŠKOVÁ

Abstract. This paper deals with some problems of detection of fundamental frequency. An essential approach is to employ methodologies in time domain. The autocorrelation signal function of a modified signal is a basic operational apparatus. Both theoretical analysis and experimental modelling of audio signal are proposed in this contribution. Real signal records were used.

Keywords: fundamental vocal frequency, autocorrelation function, Fourier representation of signal
AMS classification: 62M10, 68T50, 92C50

Motivation

The method of examination and representation of voice range profile is described in paper [1]. During the realization of individual procedures, the authors of the contribution encountered with some problems of detection and definition of the fundamental "glottal" frequency. This paper presents their solution and optionally also summary of topics resulting from them. The method used here is more than fifty years old [2]. Therefore it has its tradition with all the positives and negatives joined to this concept. A considerable goal is also to clarify some of the data and procedures that are handed down. On that account, resources used for this paper were mainly out of the acoustic field.

Fundamental frequency - the method of solution

In audiologic literature, the following terms are often used: fundamental vocal frequency, fundamental formant frequency, fundamental "glottal" frequency. From this literature, however, it is not self-evident how such frequency is defined in a signal domain. It is not obvious, whether it is the first harmonic of a periodic signal, or it is a repeating frequency. The problem is represented by the following signal:

$$u(t) = 0.3 \sin\left(\frac{2\pi}{3}t + \varphi_1\right) + 1 \sin\left(\frac{2\pi}{7}t + \varphi_2\right). \tag{1}$$

The paper was supported by the Research Plan of Ministry of Education: "Information Systems and Technologies", No. MSM-235200005.

By this signal (see Fig. 1), it is possible to represent the essence of the given problem exactly. Is the searched fundamental frequency the component with period 7 (i.e. the one with the maximal amplitude), or the period of the complete signal $7 \cdot 3 = 21$? Naturally, if the real signal looks like our abstraction, the problem will not arise - it is enough to set the complete spectrum. In reality, the problem is more complicated:

$$u(t) = 0.3 \sin\left(\frac{2\pi}{3}t + \varphi_1\right) + 1 \sin\left(\frac{2\pi}{7}t + \varphi_2\right) + \epsilon(t), \tag{2}$$

where $\epsilon(t)$ is a random component modelled by MA-process. An extreme situation (for low vocal frequencies) is shown in Fig. 2.

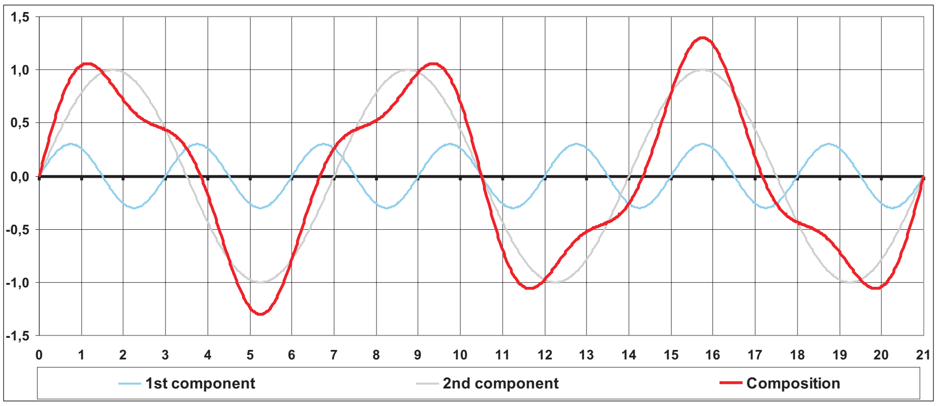


Fig. 1

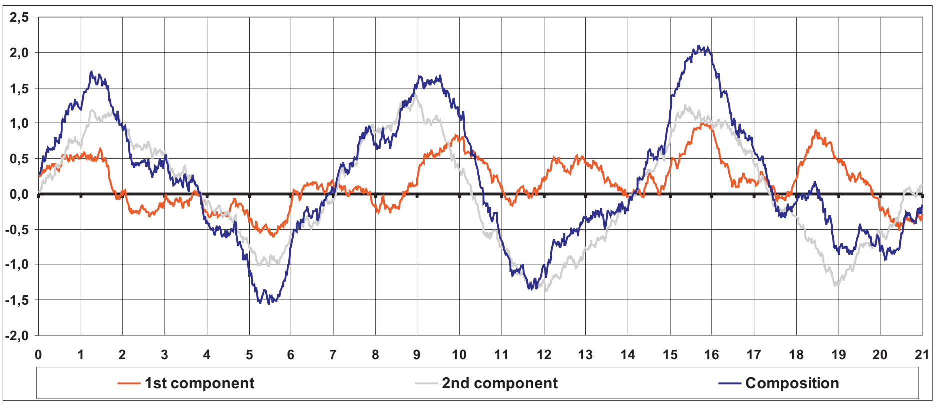


Fig. 2

In the course of processing a vocal signal for the purposes of "the measuring of voice range profile", we are limited by some assumptions. The fundamental ones are:

(1) Rather short data segmentation (max. tens of periods).
(2) Requirement of fast processing.

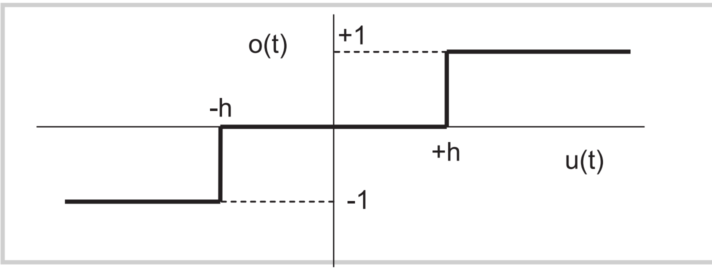
Therefore the classical correlation analysis is used for the detection of the fundamental vocal frequency. The signal segment is sampled, if need be is also normed or simultaneously symmetrised into the interval $[-1, 1]$. Then it is transformed by the relay non-linearity with restriction around zero. For such $-1, 0, +1$ signal, the autocorrelation function of the random process realization in a finite interval is calculated:

$$R(T, \tau) = \frac{1}{T} \int_0^T u(t)u(t+\tau)dt \tag{3}$$

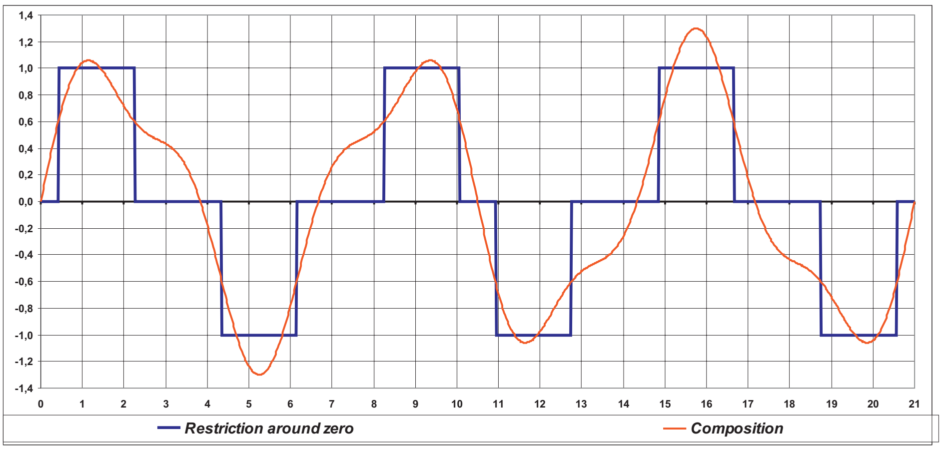
and for the numeric solution, where $\tau = k\Delta$ and $T = n\Delta$, and Δ is a sample interval:

$$r(T, \tau) = \frac{\Delta}{n-k} \sum_{i=1}^{n-k} u_i \, u_{i+k}. \tag{4}$$

Firstly, the signal is transformed by the relay non-linearity with restriction around zero according to the following figure:



After such transformation, the signal has the following course:



The transformed signal can be described by:

$$u(t) = \textstyle\sum_{i=0}^{m} A_i \sin(\varpi_i t + \varphi_i) + \epsilon(t), \tag{5}$$

$$\varpi = \tfrac{2\pi}{T}; \quad \varpi_0 = 0; \quad T = NSN\left(\tfrac{2\pi}{\varpi_1}, \tfrac{2\pi}{\varpi_2}, \ldots, \tfrac{2\pi}{\varpi_m}\right),$$

where ϖ is the circular frequency of all the composition.

$$o(t) = sgn_h\left(\sum_{i=0}^{m} A_i \sin(\varpi_i t + \varphi_i) + \epsilon(t)\right), \tag{6}$$

$$\begin{aligned} \text{where } sgn_h(x) &= -1 \Leftrightarrow x < -h, \\ sgn_h(x) &= 0 \Leftrightarrow -h \leq x \leq +h, \\ sgn_h(x) &= +1 \Leftrightarrow x > +h. \end{aligned}$$

For all other derivations, we will suppose $\epsilon(t) = 0$; i.e. the situation without any noise.

With respect to the above mentioned presumptions, we get:

$$o(t) = \sum_{i=0}^{\infty} a_i \sin(i\varpi t + \psi_i), \tag{7}$$

where a_i is the Fourier amplitude and ψ_i is the phase of the signal $o(t)$ on the interval T. For the signal represented in such a way, it is necessary to mention several notes: Due to the relay non-linearity with restriction around zero, it comes about the spectrum transformation that keeps the primary period ϖ of the original signal $u(t)$. In extreme cases this can lead to the situation that even partial period occurs in the signal (the half one - see the figure above). We should know that just the sampling causes a frequency transformation (Nyquist frequency, ...). Obviously the same holds true also for the sampled signal $o(t)$, it cannot contain frequencies higher than double of the sampler frequency (which affects the smoothing of additive noise).

The autocorrelation function of realization $o(t)$ is then for T_{poz} big enough:

$$R_o(T_{poz}, \tau) = a_0^2 + \frac{1}{2} \sum_{i=1}^{m_{max}} a_i^2 \cos(i\varpi\tau), \tag{8}$$

where T_{poz} is the period of sampled segment. It is obvious that such correlation function does not depend on the individual phases of harmonical components. Further it is evident that the correlation function gains the local maxima:

$$a_0^2 + \frac{1}{2} \sum_{i=1}^{m_{max}} a_i^2, \quad \text{for } \varpi\tau = k2\pi;\ k = 0, 1, \ldots\ i.e.\ \frac{\tau}{T} = k;\ \tau_k = kT. \tag{9}$$

Evidently these maxima are total maxima. The correlation function gains its zero points for:

$$\varpi\tau = \frac{\pi}{2} + 2\pi; \quad k = 0, 1, \ldots \quad \tau_k = T\left(\frac{k}{2} + \frac{1}{4}\right). \tag{10}$$

The correlation function can have (and also often has) other local maxima and zero points.

Discussion

Calculation of autocorrelation function: The enumeration of the sum

$$r(T,\tau) = \frac{\Delta}{n-k}\sum_{i=1}^{n-k} u_i u_{i+k} = \frac{\Delta}{n-k}\sum_{i=1}^{n-k} j_{ik}$$

is very fast and simple. It is just addition or substraction of ones (in the case of pre-processing by the relay non-linearity).

The accuracy: The following process can be used as a model for accuracy examining:

$$z_k = \tfrac{1}{n-k}\textstyle\sum_{i=1}^{n-k} j_{ik}, \text{ where } j_{ik} \in \{-1,0,+1\} \text{ and}$$
$$P(j_{ik}=-1) = q;\; P(j_{ik}=+1) = p;\; P(j_{ik}=0) = 1-p-q.$$

On the assumption that j_{ik} is independent, it holds:

$$\begin{aligned} E\{z_k\} &= p - q = r_k, \\ \sigma\{z_k\} &= \frac{1}{\sqrt{n-k}}\sqrt{(p+q)-(p-q)^2} = \frac{1}{\sqrt{n-k}}\sqrt{(2p-r_k)-r_k^2}\,, \end{aligned} \tag{11}$$

where r_k is a value of the correlation function in the kth point. Using more detailed analysis, it is possible to determine that the standard deviation reaches its maximum for $r_k = 0$. Thus the zeroes of the correlation function will be probably rather inaccurate. On the contrary with increasing of r_k from zero, the standard deviation decreases. The maxima will be the most accurate identifiable. The extremes are searched on the rhomboid bounded by points: (r_k, p) $[(-1,0),(0,0),(1,1),(0,1)]$.

It is evident from the relation of standard deviation that the larger argument $\tau(\equiv k)$ the smaller precision will be obtained. For equal values r_k and r_l, the ratio of the standard deviation is:

$$\frac{\sigma\{r_k\}}{\sigma\{r_l\}} = \sqrt{\frac{n-l}{n-k}}. \tag{12}$$

From this we can derive that moving on the time axes to the right, inaccuracy (measured by the ratio) increases with the square. Also some qualitative inferences about the precision of estimation of the correlation function can be expressed: the more closely to the zero value and the more far from the beginning the less reliable correlation value. The correlation function is, according to the theoretical relation, a periodical function:

$$R_o(T_{poz},\tau) = a_0^2 + \frac{1}{2}\sum_{i=1}^{m_{max}} a_i^2 \cos(i\varpi\tau).$$

In reality, however, it was not observed except only a few cases. It is caused by the frequency unstability of the signal or the sample frequency. As a theoretical model, the "small frequency modulation" of the signal can be used.

$$R_o(\varpi) = a_0^2 + \frac{1}{2}\sum_{i=1}^{m_{max}} a_i^2 \cos(i\varpi\tau), \tag{13}$$

$$R_o(\varpi+\delta) \cong R_o(\varpi) + \delta \cdot R_o'(\varpi) + \frac{\delta^2}{2} R_o''(\varpi).$$

Using the above formula we get for the value of the kth maximum M_k:

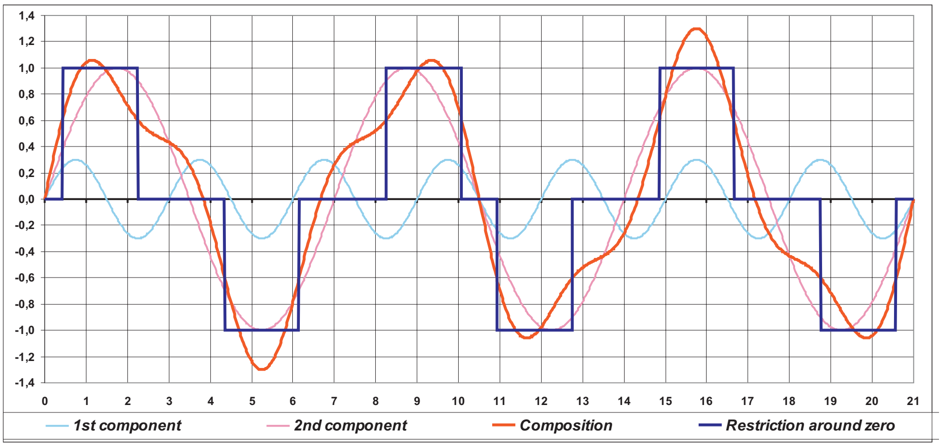
$$M_k = a_0^2 + \frac{1}{2}\sum_{i=1}^{m_{max}} a_i^2 - k^2 \left[\frac{\delta^2}{4}\sum_{i=1}^{m_{max}} (iT)^2 a_i^2\right]. \tag{14}$$

The values of maxima (corresponding with the "repeating" frequency) will decrease with the square (the smallest decrease will occur on the assumption that there is no additive noise). The conclusion of the previous discussion is that the position of the first non-zero maximum for $\tau > 0$ has a substantial identification property.

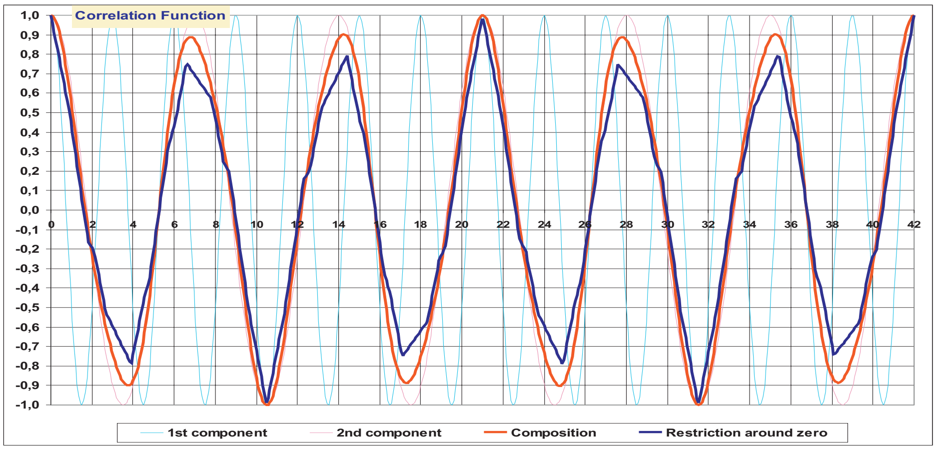
Experimental modelling

A generated mixture of two harmonical:

The signal:

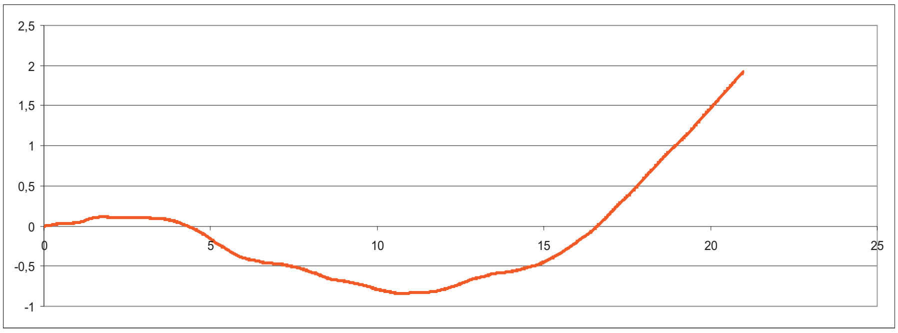


The autocorrelation function:

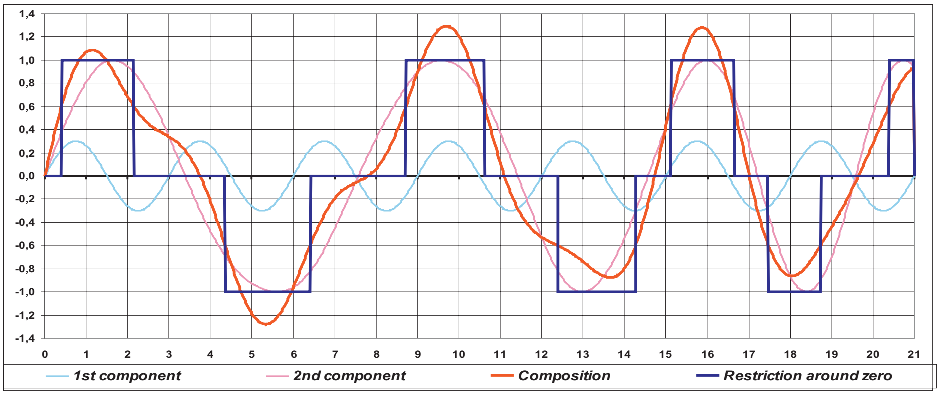


A generated mixture of two harmonical with a phase error:

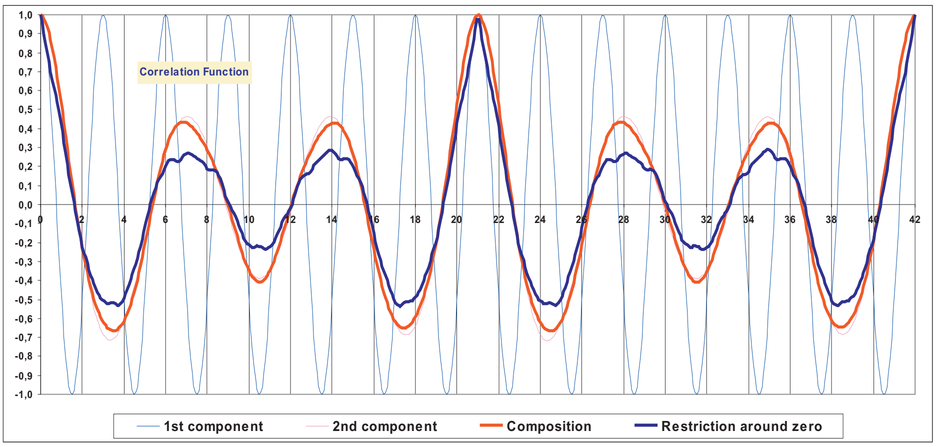
The model phase error:



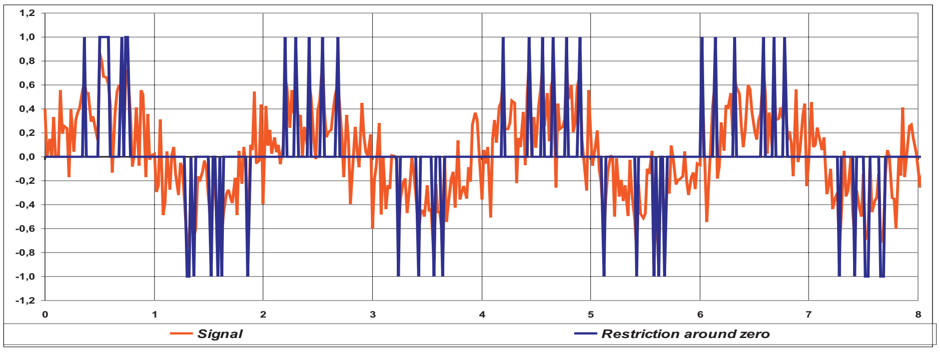
The signal:



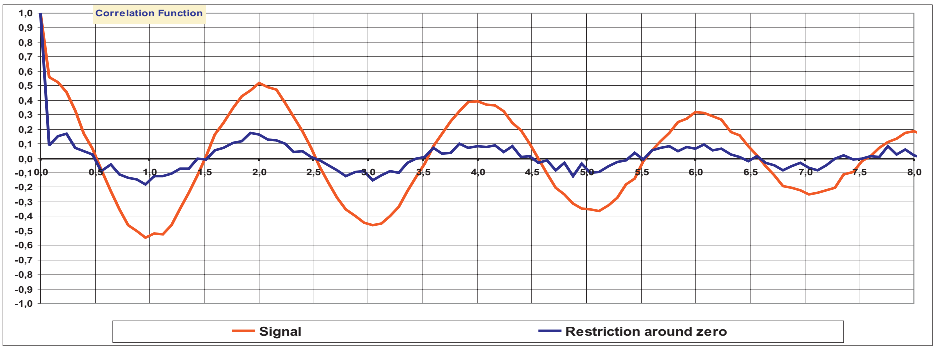
The autocorrelation function:



Generated harmonical signal + MA noise 60 % + 40 %:
The signal:

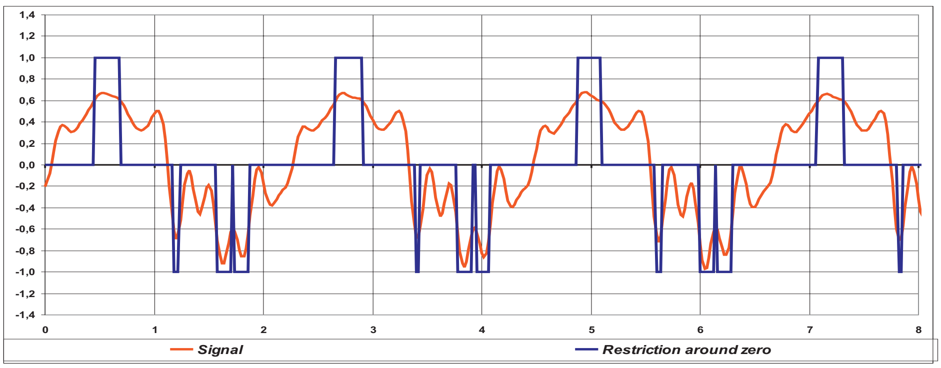


The autocorrelation function:

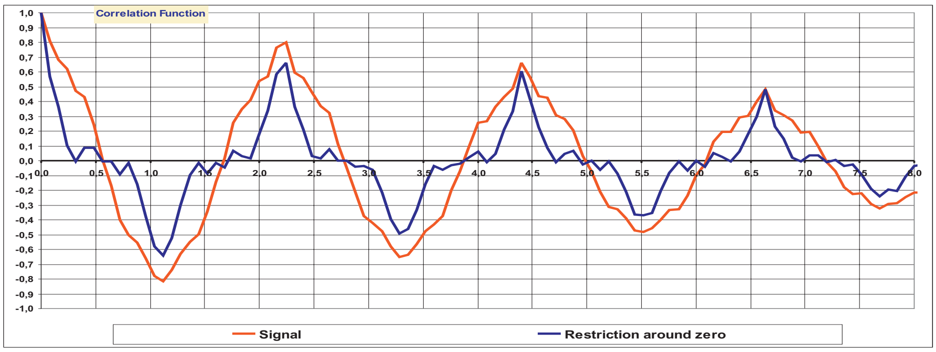


Real examples

Flute C-dur a1
The signal:

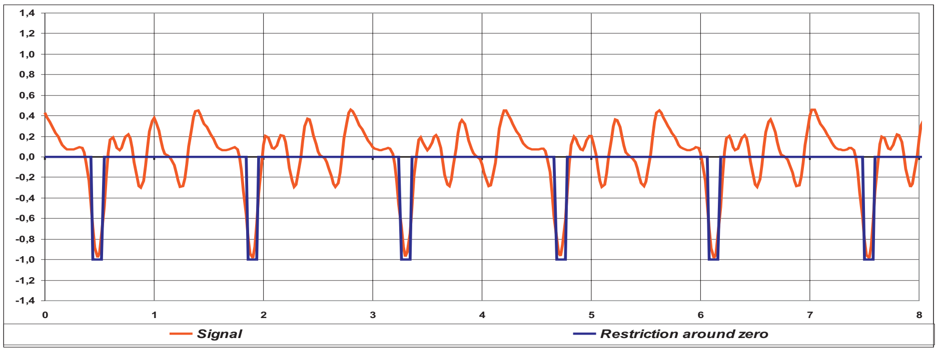


The autocorrelation function:

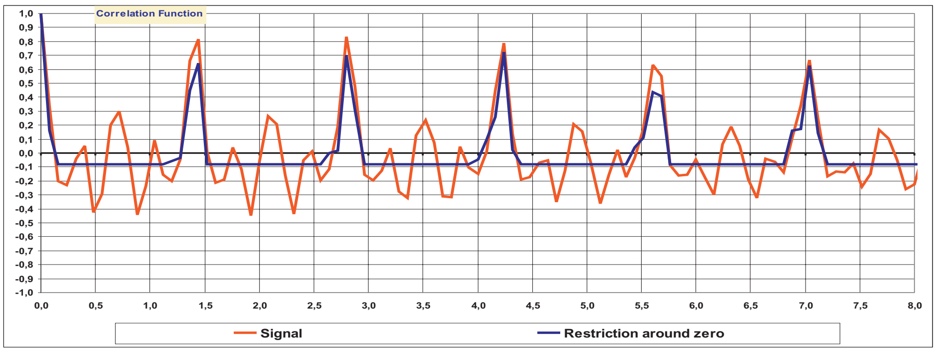


Flute C-dur f2 (an example of a false maximum)

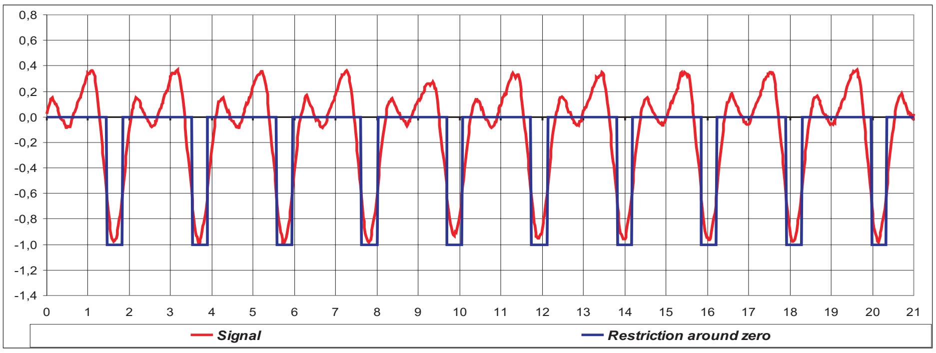
The signal:

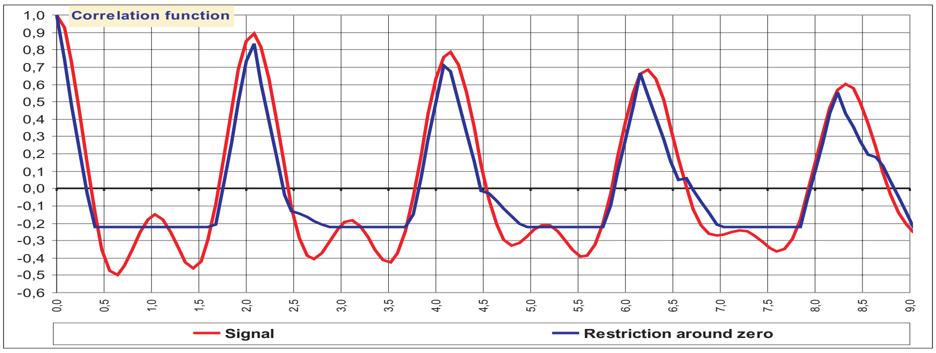


The autocorrelation function:



The real microsegment - one of the authors, sound ”a”, 440 Hz
The example of the ”small” frequency modulation:





Conclusions

(1) The correlation analysis detects mainly the repeating frequency. The detection of harmonic components is possible, but with some problems.
(2) The transformation by the relay non-linearity simplifies and speeds up the calculation, it can, however, give on detection of false frequencies.
(3) The estimation of value of the correlation function is the most accurate toward the direction of zero value and farther time.
(4) Decrease of the envelope of the correlation function is determined by the frequency unstability of the signal or the sample frequency (and not only by it).
(5) The selection of threshold ”h” of the restriction around zero is quite robust, the threshold need not be even symmetrical, it is possible also one-sided thresholding (it will be the aim of our next research).

Further research

(1) What frequency (the acuteness of a tone) can we hear?
A mixture of harmonical (sine wave) tones will be generated and ”send” to

one head phone. In the latter one, there will be a "pure" harmonical tone. A listener will adjust the acuteness (frequency) of the pure tone in such a way that, according to his subjective feeling, the acuteness of the tones are the same. Parameters of the generated mixture and the subjectively determined acuteness of the pure tone will be recorded.

(2) How do we repeat a "heard" pure tone?
A pure harmonical tone will be send to receivers. A listener will repeat it to a microphone. This signal will be sampled and its spectrum and correlation detection of contained frequencies will be analyzed.

References

[1] Nový, P., Vávra, F., Kotlíková, M.: Voice range profile examination method and its applications. Datastat-03, Svratka, 2003.

[2] Beneš, J.: Statistická dynamika regulačních obvodů, SNTL Praha, 1961.

[3] Levin, B.R.: Teorie náhodných procesů a její aplikace v radiotechnice, SNTL Praha, 1965.

[4] Kufner, A., Kadlec, J.: Fourierovy řady, Academia Praha, 1969.

[5] Boersma, P.: Accurate Short Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound, Institute of Phonetic Sciences, University of Amsterdam, Proceedings 17 (1993), pp. 97-110.

[6] Artl, J.: Moderní metody modelování ekonomických časových řad, GRADA Publishing, Praha, 1999.

[7] Groth, A.: Estimation of Periodicity in Time Series by Ordinal Analysis with an Application to Speech, Ernst Moritz Arndt University of Greifswald.

[8] Cuadras, C.M.: On the Covariance between Functions, Journal of Multivariate Analysis 81, 19-27 (2002).

Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia in Pilsen

E-mail address: novyp@kiv.zcu.cz, vavra@kiv.zcu.cz, maskova@kiv.zcu.cz