

Music Tune Restoration Based on a Mother Wavelet Construction

A S Fadeev, V I Konovalov, T I Butakova, A V Sobetsky

Tomsk Polytechnic University, 30, Lenina Ave., Tomsk, 634050, Russia

E-mail: fas@tpu.ru

Abstract. It is offered to use the mother wavelet function obtained from the local part of an analyzed music signal. Requirements for the constructed function are proposed and the implementation technique and its properties are described. The suggested approach allows construction of mother wavelet families with specified identifying properties. Consequently, this makes possible to identify the basic signal variations of complex music signals including local time-frequency characteristics of the basic one.

1. Introduction

Discrete and continuous wavelet transforms are becoming an indispensable part of modern mathematics and other human activities. Furthermore, owing to the growth of computing technologies in recent years, hardware and software contributed to solution of many mathematical problems connected with pattern recognition. So, modern high-performance computing allows solving a number of problems including speech recognition, graphical objects, processing of seismic data, cardiograms, etc. A lot of them require using wavelet transforms. However, some problems in musical pattern recognition and their implementation in the automated information systems have been insufficiently studied [**Ошибка! Источник ссылки не найден.**, 10]. One of such problems is a problem of identifying a separate note in one-voice and polyphonic melodies performed on certain musical instruments.

2. Music signal model

A simple mathematical model of any music melody consists of a set of notes played in different times on a certain musical instrument [**Ошибка! Источник ссылки не найден.**]:

$$F(t) = A_1 n_1(t - \theta_1) + A_2 n_2(t - \theta_2) + \dots + A_N n_N(t - \theta_N) + h(t),$$

where $n_i(t)$ is the amplitude-time characteristics of a single note voice; θ_i is the temporal shift determining initial time of each note sounding; A_i is the sound volume of the separate note; $h(t)$ is the signal of disturbance introduced by sound-recording equipment; t is time.

The majority of musical instruments possesses the property of self-similarity which allows obtaining the temporal function of any note $n_i(t)$ from the same musical instrument from the temporal function of one note $n_0(t)$ of a certain musical instrument by scaling function $n_0(t)$ along time axis:

$$\tilde{n}_i(t) = n_0\left(\frac{t}{m_i}\right),$$



where m_i is the scale factor, i is the position of note $n_i(t)$ by height relative to note $n_0(t)$. For uniformly tempered pitch of European music, m_i is represented as $m_i = 2^{\frac{-i}{12}}$, $i = 0, \pm 1, \pm 2 \dots$ [**Ошибка! Источник ссылки не найден.**].

For example, note «C» of the second octave is situated 12 semitones higher than note «C» of the first octave, and has a pitch frequency two times higher than the pitch frequency of note «C» of the first octave [**Ошибка! Источник ссылки не найден.**]. The time function of note «C» of second octave $n_2(t)$ is two times compressed relatively function $n_1(t)$:

$$n_2(t) = n_1 \left(\frac{t}{2} \right) = n_1(2t).$$

Any note in the range of a certain instrument may be selected as base note $n_0(t)$ regardless of being used in concrete musical signal $F(t)$. For example, the temporal function of note «A» of the first octave with pitch frequency $\nu = 440$ Hz may be taken as $n_0(t)$.

This property is implemented in all modern musical synthesizers using the method of wave table [**Ошибка! Источник ссылки не найден.**]. Such synthesizers use the data bank of voices of one (basic) note for each musical instrument. When forming a tune of one instrument, the basic note signal is always scaled by a pitch by value m_i . Each note is shifted in time by value θ_i and scaled by amplitude by value A_i according to this instrument part [**Ошибка! Источник ссылки не найден.**]. All formed functions for musical instrument are summarized. As a result, the model of a musical signal formed by a synthesizer for a definite musical instrument has the following form:

$$F(t) = \sum_i A_i \cdot n_0 \left(\frac{t - \theta_i}{m_i} \right).$$

Similarly, for k -different musical instruments simulated by the synthesizer, a musical signal may be represented by a sum of signals of all notes played at different moments of time with different amplitude:

$$f(t) = \sum_k \sum_i A_i \cdot n_0^k \left(\frac{t - \theta_i}{m_i} \right),$$

where $n_0^k(t)$ is the time-dependent function of a basic note signal for a k -musical instrument; θ_i is the time interval shift of note $n_0^k(t)$; m_i is the scale of note $n_i^k(t)$ in regard to basic note $n_0^k(t)$ specifying frequency of the main tone; A_i is the magnitude of note $n_i^k(t)$.

The above-mentioned function $f(t)$ represents an idealized model of a musical signal obtained similarly to recording several musical instruments in an orchestra or an ensemble [**Ошибка! Источник ссылки не найден.**].

According to the model of musical signal $f(t)$, the task of musical signal identification may be presented as the task of identification of note amplitudes with certain scale m_i and time shift θ_i for all k musical instruments presented in analyzed signal $f(t)$. The task of identification of a singly recorded musical instrument melody or a melody generated by a musical synthesizer may be considered as a particular case of this task [**Ошибка! Источник ссылки не найден.**].

3. Continuous wavelet transform

Selecting the mathematical apparatus for time-frequency analysis of signal $f(t)$ which is nonstationary and nonperiodic one, it was determined to apply a continuous wavelet transform (CWT):

$$Wf(\tau, s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{s}} f(t) w \left(\frac{t - \tau}{s} \right) dt,$$

where $w(t)$ is the mother wavelet function; s is the coefficient of wavelet scaling; τ is the coefficient of wavelet shift.

One of the features of CWT is the formation of wavelet family $w_{s,\tau}(t)$ by shifts τ and scaling s of mother wavelet $w(t)$. The wavelet family formation is similar to the system of formation of note family $n_i^k(t)$ of one basic note $n_0^k(t)$ by shifts θ_i and scaling m_i . Therefore, CWT usage in this task is rather restricted.

The procedure of selecting a mother wavelet function is empirical for each definite task and reduced to searching functions of mother wavelets in CWT until the achievement of a desired result. The research of wavelet function properties [Ошибка! Источник ссылки не найден.] showed that the best graphic presentations of CWT results are obtained in case of conformity of frequency spectra of signal $f(t)$ and wavelet $w(t)$.

For each scale s , function $Wf_s(\tau)$ is similar to the cross-correlation function of signals $w_s(t)$ and $f(t)$, and describes them both as the similarity measure of two signal form and their positional relationship to each other on the time axis.

It is known that values of the cross-correlation function are maximal in case of function coincidence [Ошибка! Источник ссылки не найден.]. In this case, values $Wf_s(\tau)$ are maximal for such τ as functions $f(t)$ and $w_s(t - \tau)$ are equal: $f(t) = w_s(t - \tau)$. It is obvious that besides shift τ , the equality of two functions in each point t is required for fulfillment of this condition.

The wavelet function used in CWT should meet a number of required conditions [Ошибка! Источник ссылки не найден.]:

1. limitation (localization) in time: $w(t) > 0$, at $t \rightarrow \infty$;
2. sectional continuity of function $w(t)$;
3. integrability with zero equality.

The example of the mother wavelet-function satisfying all given conditions is Morlet wavelet (Figure 1).

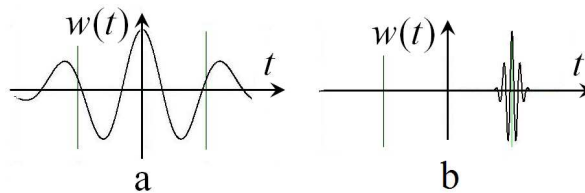


Figure 1. Wavelets of Morlet family: a) a mother Morlet wavelet obtained by scaling $s=0,2$; b) a mother Morlet wavelet obtained by a shift by $\tau=5$ and scaling $s=2$.

4. Mother wavelet construction

Owing to the conditions of time localization imposed to the basic wavelet function, function $Wf_s(\tau)$ reaches a maximum value when wavelet $w_s(t-\tau)$ of scale s coincides more precisely with the local section of signal $f(t)$ [Ошибка! Источник ссылки не найден.]. Functions of a wavelet and a signal should be equal for exact coincidence at the local time interval.

We suggested using the function of the mother wavelet formed of the local section of the analyzed musical signal in the given paper. For a more restricted task, identification of musical instrument notes, the mother wavelet may be formed of the local section of the basic note signal function of this musical instrument $n_0(t)$ (Figure 2).

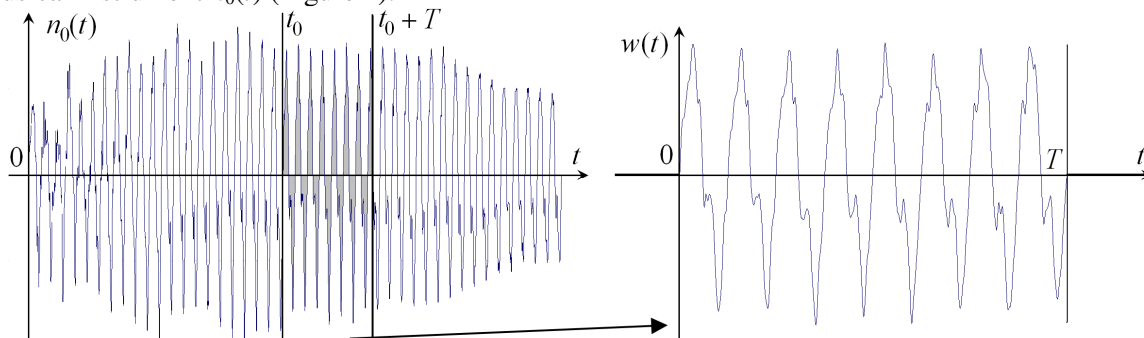


Figure 2. Function $n_0(t)$ and wavelet $w(t)$ formed on the basic function fragment.

For the formed wavelet, conditions 1, 2, 3 should be fulfilled as follows.

1. Limitation in time

$$w(t)=0, t \notin [0, T];$$

$$w(t)= n_0(t+t_0), \text{ at } t \in [0, T].$$

Here, t_0 characterizes time moment from which the values of wavelet function $w(t)$ are equal to values of the function of signal $n_0(t)$, and value T equals to the fragment duration of signal $n_0(t)$, coinciding with wavelet $w(t)$. Values t_0 and T are selected to satisfy the rest conditions required for wavelet functions [Ошибка! Источник ссылки не найден.].

2. Sectional continuity

The function of basic note $n_0(t)$ is continuous on the whole interval of existence as it describes oscillations of a physical body with finite mass in time and cannot have any breaks.

To support sectional continuity of the basic wavelet function, the condition of zero equality of initial and finite values of function $n_0(t)$ on interval $[t_0, T+t_0]$ should be fulfilled:

$$n_0(t_0)=0 \text{ and } n_0(T+t_0)=0.$$

3. Integrability with zero equality

One of the properties of musical instruments is absence of harmonic components with frequency lower than frequency of the note pitch. Zero harmonic is absent in musical instrument signals as well [Ошибка! Источник ссылки не найден.]. This property allows supporting zero mean for $n_0(t)$ on interval $[t_0, T+t_0]$ at the integer number of periods in function $n_0(t)$ on a given interval [Ошибка! Источник ссылки не найден.].

Therefore, to form the wavelet possessing the highest selectivity to signal $n_0(t)$, it is necessary to use a periodic section of signal $n_0(t)$ with zero initial $n_0(t_0)$ and zero end $n_0(T+t_0)$ moments such as

$$\int_{t_0}^{T+t_0} n_0(t)dt = 0.$$

Identification of musical notes comprises the determination of frequencies of notes pitch themselves, start time, and duration of their sounding. One of the properties of wavelet function, determining the main time-frequency selectivity of CWT, is wavelet localization in time and frequency regions [Ошибка! Источник ссылки не найден.]. During CWT, each wavelet of one family, obtained from one basic wavelet, forms a time-frequency window of the limited size. The window area for one family wavelet on the plane shift-scale is always constant [Ошибка! Источник ссылки не найден.] (Figure 3).

After a number of experiments, it was found out that changing the wavelet itself makes window geometric characteristics change [Ошибка! Источник ссылки не найден.]. So, increasing a number of periods in the basic wavelet (and, therefore, in all wavelets of the family), the window is extended along the time axis, narrowing relative to the scale axis and, on the contrary, decreasing a number of periods in the basic wavelet, the window is extended along scale axis s , narrowing along the time axis (Figure 4).

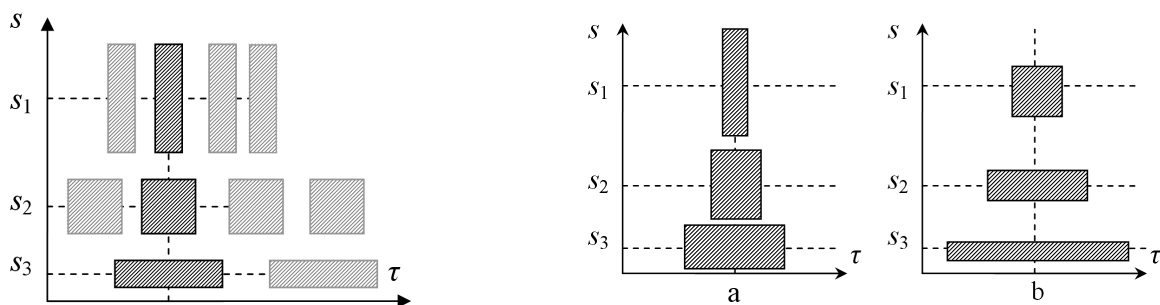


Figure 3. The window of wavelet time-frequency localization at different values of parameters of shift τ and scale s .

Figure 4. The window of time-frequency localization of wavelets with a number of periods equal to a) $N=2$ and b) $N=8$.

For the tasks of elementary components (notes) detection in a musical signal, two conditions should be followed for the wavelet family:

1. frequency resolution (the window height along scale axis Δs) should identify different frequencies of two adjacent notes;
2. time resolution (the window width along time axis $\Delta \tau$) should allow identifying all notes of minimum possible duration.

1. Frequency resolution

The experiment was carried out to estimate the resolution of artificial mother wavelets. The aim of the experiment was to determine a number of periods in the wavelet that allows identifying frequency scales m_i of all notes being in the signal simultaneously (at CWT wavelet scales s relative to the basic one are equivalent to m_i).

Chord «C major» of the first octave was used in the experiment. Frequencies of note pitches of this chord corresponds to harmonic signals with frequencies 261,6; 329,6; and 392 Hz [Ошибка! Источник ссылки не найден.]. The signal duration is chosen to be equal to 0,2 s:

$$f(t) = \sin(2\pi 261,6t) + \sin(2\pi 329,6t) + \sin(2\pi 392t).$$

Mother wavelets $f w_i(t)$ were constructed from the harmonic signal to study the test signal. The amount of harmonic signal periods in wavelets was 1, 2, 4, 8, and 16 periods, respectively (Figure 5).

CWT was carried out with all wavelet families $w_i(t)$ for test signal $f(t)$. For each transform, the results of graphical interpretations were interpreted in three-dimensional models [Ошибка! Источник ссылки не найден.]. 3-D models of CWT results for the families of wavelets $w_i(t)$ are presented in Figure 6. The ordinate axis of each illustration represents an axis of the wavelet s scale, the abscissa axis is the axis of time shift τ ; τ_0, τ_1 – time of signal $f(t)$ beginning and ending. The magnitude of CWT results is presented by the gray scale where darker sections correspond to a higher magnitude of CWT results [Ошибка! Источник ссылки не найден.] (Figure 6).

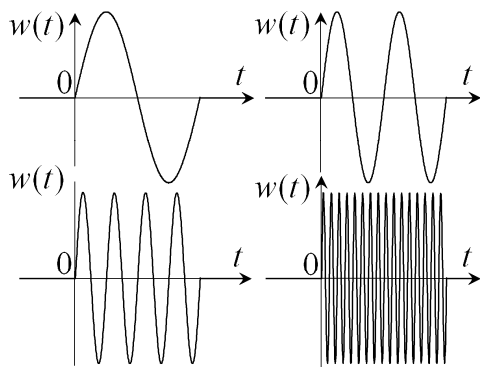


Figure 5. Mother wavelets with a number of periods of harmonic signal 1, 2, 4, and 16.

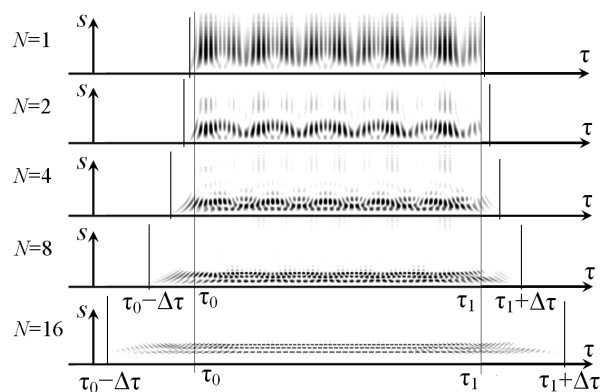


Figure 6. Graphical interpretations of CWT results for wavelets with different numbers of periods N of the harmonic signal.

Figure 6 shows that when the wavelet was used with one period, a region of uncertain results in the region of time shifts τ , with duration $\Delta \tau$ is rather low; therefore, the error of the signal time estimate is low. In this case, frequency resolution is so low that for the majority of scales s , CWT results have the same values indicating the presence of frequency components in the whole frequency range of research. However, there are three localized harmonics in test signal $f(t)$.

It is seen that for the wavelet with 16 periods over signal $f(t)$, on the interval from τ_0 to τ_1 , there are high values only for CWT results for three scales s equivalent to frequencies 261,6; 329,6; and 392 Hz of signal $f(t)$. For the rest values of s , CWT results are almost equal to zero. Low time resolution produced uncertain results at the beginning and ending moments of the signal with magnitude intermediate values on intervals with duration $\Delta\tau$. It does not allow judging the changes of the test signal magnitude on this interval.

Thus, the wavelet family with one period of the sine signal during CWT gives high time resolution, but rather low frequency resolution. However, the wavelet family with sixteen periods gives high frequency resolution (all harmonics in the signal are identified definitely), but low time resolution.

2. Time resolution

Musical notation implies using notes and the rest symbols to mark out melody elements and gaps at which voices do not sound. The note duration as well as the rest duration are multiplied by duration t_1 of semibreve («whole note») – a note of maximum possible duration. The system of musical notation consists of alternating notes and the rest imposes strict constraints on the moments of the note sounding start, and the rest start. Moments of the note sounding start are sampled with sampling period t_d equal to duration of the shortest note. Both in classical and modern musical compositions, the shortest note by duration is the hemidemisemiquaver note with duration $t_{64}=1/64 \cdot t_1$ [Ошибка! Источник ссылки не найден.]. In practice, notes with duration t_{64} occur rather seldom due to the technical complexity of performance. In fact, duration $t_{32}=1/32 \cdot t_1$ (demisemiquaver note) may be considered as the shortest note. The fragment of a two-voice melody is shown in Figure 7. Each time, not more than two notes sound simultaneously. Note 2 is the shortest one. Notes 1, 3, and 4 are equal in duration and two times longer than note 2 and the rest. If $t_d=t_{32}$, then note 2 and the rest have duration t_{32} , and notes 1, 3, 4 have duration t_{16} .

At wavelet duration $T=t_d$, an envelope function of joint-correlation function $Wf_s(\tau)$ of the wavelet and the signal of one note coinciding with it by form (for concrete value s) regenerating into the autocorrelation function has the form of an equilateral triangle with maximum in the centre of note sounding [Ошибка! Источник ссылки не найден.] and width $2t_d$ (Figure 8, a). If all values of $Wf_s(t)$ are smaller than $Wf_{mist}(t)$, it would be ejected, and the note identification time is t_i . If $Wf_{mist}(t)=0,5Wf_{max}(t)$, then $t_i=0,5t_d$. This means that the note identification time equals to a half of its length.

If the wavelet duration is $T=0,5t_d$, then the envelope of correlation function $Wf_s(\tau)$ of the wavelet and one note signal, coincided with it in the form, has the form of an isosceles trapezium the width of the upper boundary of which is equal to $0,5t_d$ (Figure 8, b). If all values $Wf_s(t)$ are smaller than $Wf_{mist}(t)$, they are rejected, and then the note identification time is t_i . If $Wf_{mist}(t)=0,5Wf_{max}(t)$, then $t_i=0,5t_d$. It means that note identification time equals to its length.

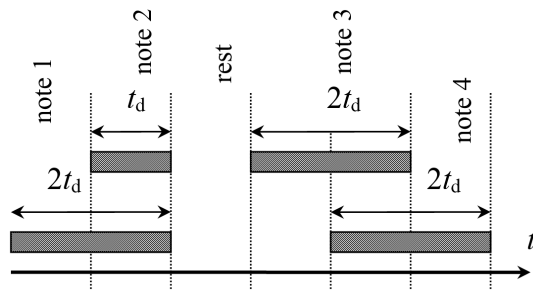


Figure 7. Time discretization of beginning and duration of notes and rests in a polyphonic melody.

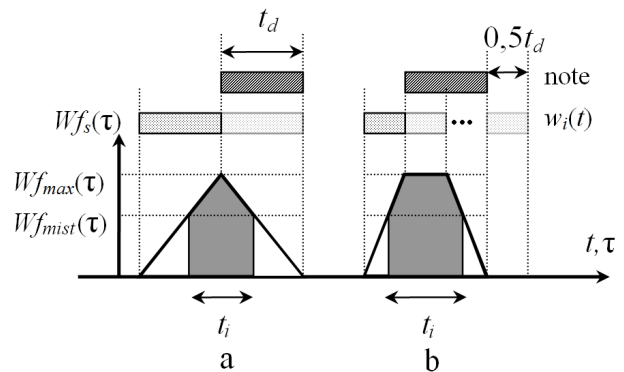


Figure 8. Envelopes of the correlation function of the note and wavelet $w(t)$ corresponding to the frequency of the scale note pitch.

Upon further decrease of wavelet width T and condition $Wf_{mist}(t)=0,5Wf_{max}(t)$, note identification time t_i is constant and equal to t_d .

Conclusion: the wavelet of length T is not more than t_d , and should be used for time identification of the note with smallest length t_d .

Taking into account two conditions to duration T of wavelet in time:

1. $T \geq \frac{16}{\nu}$, where ν is the frequency of the identified note pitch;
2. $T \leq t_d$.

Let us calculate the boundary values of note pitch frequency ν , the reliable identification of which is possible both in time and frequency: $T = \frac{16}{\nu}$, $T = t_d$, therefore, $t_d = \frac{16}{\nu}$ or $\nu = \frac{16}{t_d}$. Dependence

$\nu(t_d)$ determines minimal (boundary) frequency ν of the note pitch with duration t_d which is definitely identified by the wavelet with 16 periods.

Tempo of the majority compositions performance varies as a rule in the range of 60–180 beats per minute (BPM) that corresponds to the time of sounding for one semibreve (whole) note $t_1=1,3...4,0$ s. Therefore, note sounding with duration t_{32} at quick tempo (180 BPM) amounts to $t_{32}=1,3/32=0,041$ s. Therefore, the length of the wavelet capable of identifying the time interval of a note with the shortest duration at high tempo should amount to not more than $t_d=t_{32}=0,041$ s. The value of boundary frequency of the note is pitch identification equal to $\nu(t_d)=16/0,041=390$ Hz. This means that just identification of the beginning and ending time of note sounding is possible only for notes with pitch frequency higher than 390 Hz (Figure 9) (starting with note «G» of the Middle octave, the pitch frequency of which amounts to 392 Hz).

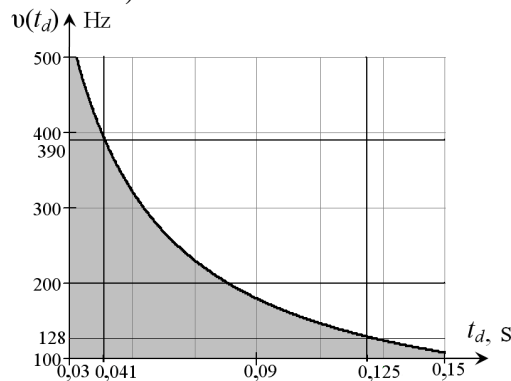


Figure 9. Dependence of boundary frequency of note pitch identification on its duration.

For slower tempo of composition performance or for tasks of detecting the note with the duration longer than t_{32} , the boundary frequency decreases. So, for example, for modern club dancing compositions, the reproducing tempo varies near the value of 120 BPM. And the shortest note is $t_d=t_{16}=0,125$ s. The value of boundary frequency of note pitch identification is $\nu(t_d)=16/0,125=128$ Hz. Identification of sounding time of note beginning and ending is possible only for notes higher than «C» of Bass octave with pitch frequency 130,8 Hz. The majority of musical compositions use notes at the range of the Middle, under Middle and above Middle octaves, which is higher than the range of Bass octave.

5. Conclusion

The suggested approach allows making mother wavelet families with defined selectivity. The wavelet developed from a fragment of a certain basic signal allows detecting the time-frequency behavior of the basic signal under study instead of separate time-frequency characteristics of the studied signal.

Mother wavelets for voices of various orchestra groups such as piano, organ, violin, bells, and trumpet were obtained experimentally. All wavelets contain 16 periods of a signal of a proper musical instrument with pitch frequency 55 Hz. While calculating CWT, the mother wavelet is scaled in such a manner that next wavelet $w_i(t)$ coincides by pitch frequency with note pitch frequency $n_i(t)$ of a musical instrument.

Construction of mother wavelet families, developed on the basis of musical instrument notes, showed the possibilities of detecting frequency and time parameters of certain instrument notes in one-voice and polyphonic tunes. Besides, it made it possible to identify a tune of a certain musical instrument on the background of sounding of another one in a number of experiments [**Ошибка! Источник ссылки не найден.**].

The authors suggested that the implementation of this technique in tasks, requiring detecting the fragments of certain families with the limited length in a signal against the background of other signals or disturbances, is to be researched further.

Acknowledgments

The reported study was funded by the Russian Foundation for Basic Research according to research project No. 16-37-00402 mol_a.

References

- [1] Barthelet M Kronland-Martinet R and Ystad S 2006 Consistency of Timbre Patterns in Expressive Music Performance Proc. 9th Int. Conf. on Digital Audio Effects DAFx-06 (Montreal, Canada) pp 19–24
- [2] Devyatikh D Gerget O and Berestneva O 2014 Sleep Apnea Detection Based on Dynamic Neural Networks *Communications in Computer and Information Science* pp 556-67
- [3] Elert G 2016 The Physics Hypertextbook (<http://physics.info/music/>)
- [4] Fadeev A and Kochegurova E 2006 Preparation of the results of continuous wavelet-transform to automated treatment *Bulletin of the Tomsk Polytechnic University* **309** 32–35
- [5] Hankinson A Burgoyne J A Vigliani G and Fujinaga I 2012 Creating a large-scale searchable digital collection from printed music materials *Proc. Advances in Music Information Research* (Lyon, France) pp 903–8
- [6] Holopainen R 2012 Self-organised Sound with Autonomous Instruments: Aesthetics and experiments (University of Oslo) 404
- [7] Janelle K Hammond and Kelly S 2011 Mathematics of Music *UW-L Journal of Undergraduate Research* **14** 11
- [8] Kochegurova E and Gorokhova E 2015 Current derivative estimation of non-stationary processes based on metrical information *Lecture notes in artificial intelligence* **9330** (Springer) 512–19
- [9] Kochegurova E and Fadeev A 2006 Wavelet analysis in the task of musical information identification *Proc. Vseros. Conf. Molodezh I Sovremennye Informatsionnye Tekhnologii* (Tomsk: TPU Press) pp 149–151
- [10] 1. Makhijani R Shrawankar U and Thakare V M 2010 Opportunities and challenges in automatic speech recognition *Proc. Int. Conf. Biomedical Engineering and Assistive Technologies* (NIT Jalandhar, India)
- [11] Mallat S 1998 *A Wavelet Tour of Signal Processing* (Academic Press Inc) p. 704
- [12] McKay C and Fujinaga I 2010 Improving automatic music classification performance by extracting features from different types of data *Proc. ACM Int. Conf. on Multimedia Information Retrieval* 257–66
- [13] Phinyomark A Limsakul C and Phukpattaranont P 2011 Application of Wavelet Analysis in EMG Feature Extraction for Pattern Classification *Measurement Sci. Rev.* **11** 45–52
- [14] Sadowsky J 1996 Investigation of signal characteristics using the continuous wavelet transforms *Johns Hopkins APL Technical Digest* **17** 258–69

- [15] Skirnevskiy I and Korovin A 2016 Optimal methods of segmentation of tomographic volume searching system: a preliminary review *Key Engineering Materials* **685** 857–62
- [16] Wright D 2009 Mathematics and Music *Washington. Mathematical World* **28** 176