UDC 517.5

WAVELET-FUNCTION FORMATION IN THE PROBLEM OF MUSIC SIGNAL IDENTIFICATION

A.S. Fadeev

Tomsk Polytechnic University
E-mail: fas@aics.ru

Approach allowing forming wavelet-functions on the basis of periodic signals and signal fragments of musical instruments has been suggested. The required and sufficient conditions made to the formed wavelet-functions were considered. The experiment allowing identifying some harmonic components of a signal localized in time was described. The possibility of applying the developed approach in the tasks of identifying complex musical signals was shown.

Growth of information technologies in recent years made possible software and hardware implementation of many mathematical problems connected with pattern recognition. Owing to the development of modern high-performance computing a number of problems in the range of recognition of speech, graphics objects, processing of seismic dada, cardiograms and others is solved. The problems in the range of musical information recognition and possibility of their application in the systems of information automated processing are less studied [1]. One of such problems is a problem of separate note identification in one-voice and poly-phonic melodies performed on a certain musical instruments.

The base of computing experiment is mathematical model of a music signal $F(t)$ obtained at sound recording of a melody of a certain musical instrument:

$$F(t)=A_1n_1(t-\theta_1)+A_2n_2(t-\theta_2)+...+A_Nn_N(t-\theta_N)+h(t),$$

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

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

$$n_i(t)=n_0\left(\frac{t}{m_i}\right),$$

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 $m_i=2^{-i/12}$, $i=0,\pm1,\pm2...$.

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. Time function of note «C» of the second octave $n_2(t)$ is compressed two times relative to function $n_1(t)$:

$$n_2(t)=n_1\left(\frac{t}{2^{\frac{-12}{12}}}\right)=n_1(2t).$$

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

This property is applied in all modern musical synthesizers using the method of wave table [2]. Such synthesizers use data bank of signals 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 a value m_i. Each note is shifted in time by a value θ_i and scaled by amplitude by a value A_i according to this instrument part. All formed functions for a musical instrument are summarized. As a result the model of a musical signal formed by a synthesizer for a concrete musical instrument has the 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 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 function of basic note signal for временная k musical instrument; θ_i is the time interval of note $n_0^k(t)$ shift; m_j 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 amplitude of note $n_i^k(t)$ sounding.

Function $f(t)$ represents an idealized model of a musical signal obtained similarly to recording several musical instruments (orchestra, ensemble).

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 a certain scale m_i and time shift θ_i for all k musical instruments presenting in the examined signal $f(t)$. The task of identification of a tune of a singly recorded musical instrument or a tune generated by a musical synthesizer may be a particular case of this task.

Selecting mathematical apparatus for studying time-frequency properties of signal $f(t)$ which is nonstationary and nonperiodic one it was determined to apply continuous wavelet transform (CWT) [3]:

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

where $w(t)$ is the wavelet basis 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 the basis wavelet $w(t)$. Wavelet family formation is similar to a 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 use in this task is rather restricted.

The procedure of selecting basis wavelet function is empirical for each concrete task and reduced to a search of functions of maternal wavelets in CWT till the achievement of a desired result. The investigations of wavelet-function properties [4] showed that the best graphic presentations of CWT results are obtained at conformity of frequency spectra of a signal $f(t)$ and wavelet $w(t)$.

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

$$Wf_s(\tau) = \int_{-\infty}^{\infty} f(t) w_s(t-\tau) dt, \quad w_s(t) = w(\frac{t}{s}).$$

It is known that values of cross-correlation function are maximal at function coincidence [6]. In this case values $Wf_s(\tau)$ are maximal for such τ at which 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.

Wavelet-function used in CWT should meet a number of required conditions [5]:

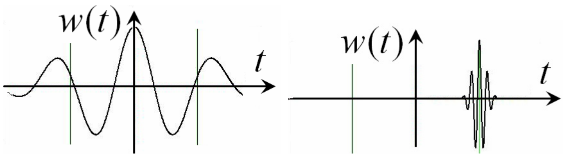
1. Limitation in time:

$$w(t)>0, \text{ at } t\to\infty;$$

2. Sectional continuity of function $w(t)$.
3. Integrability with zero equality:

$$\int_{-\infty}^{\infty} w(t) dt = 0.$$

The example of basis wavelet-function satisfying all given conditions is Morlet wavelet (Fig. 1).



a b

Fig. 1. *Wavelets of Morlet family: a) basis Morlet wavelet obtained by scaling s=0,2; б) basis Morlet wavelet obtained by a shift by τ=5 and scaling s=2*

Owing to the conditions of time localization imposed to basis wavelet function the value of function $Wf_s(\tau)$ is more when wavelet of scale s $w_s(t-\tau)$ coincides more precise with local section of a signal $f(t)$. Functions of wavelet and signal should be equal for exact coincidence at local time cells.

Function of basis wavelet formed of local section of the examined musical signal is suggested to be used in the given paper. For a more narrow task – identification of musical instrument notes determined before the maternal wavelet may be formed of local section of basic note signal function of this musical instrument $n_0(t)$ (Fig. 2).

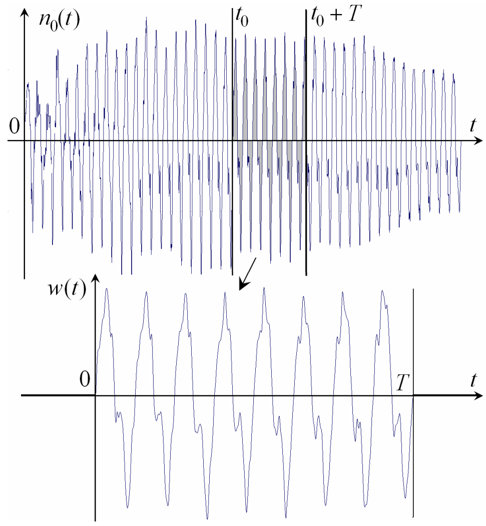


Fig. 2. *Function $n_0(t)$ and wavelet $w(t)$ formed on the basis its fragment*

For the formed wavelet the following conditions 1–3 should be fulfilled.

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 values of wavelet function $w(t)$ equal to values of functions of signal $n_0(t)$ and value T equals to fragment duration of signal $n_0(t)$, coinciding with wavelet $w(t)$. Values t_0 and T are selected in such way that the rest required conditions for wavelet-function were supported [5].

2. *Sectional continuity*

Function of basic note $n_0(t)$ is continuous on the whole interval of existence as it describes oscillations of physical body with finite mass in time and can not have breaks by definition.

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

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

3. *Integrability with zero equality*

One of the properties of musical instruments is absence of harmonic components with frequency lower than frequency of note pitch. Zero harmonic is absent as well in musical instrument signals [1]. This property allows supporting zero mean for $n_0(t)$ on interval $[t_0, T+t_0]$ at integer of periods in function $n_0(t)$ on the given interval.

Thus, to form wavelet possessing the highest selectivity to signal $n_0(t)$ it is necessary to use periodic section of signal $n_0(t)$ with zero initial $n_0(t_0)$ and finite $n_0(T+t_0)$ moments such that $\int_{t_0}^{T+t_0} n_0(t)dt = 0.$

Identification of musical composition notes consists in determination of frequencies of note pitch themselves and 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 domains [5]. At CWT each wavelet of one family (obtained of one basis wavelet) forms time-frequency window of finite size. Window area for one family wavelets on the plane shift-scale is always constant [3] (Fig. 3).

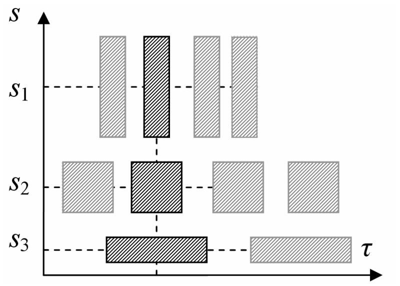


Fig. 3. *Window of wavelet time-frequency localization at different values of parameters of shift τ and scale s*

However, carrying out a number of experiments it was found out that changing the wavelet itself window geometric characteristics are changed. So, increasing a number of periods in the maternal wavelet (and, therefore, in all wavelets of the family) the window is extended along time axis narrowing relative to scale axis and, on the contrary, decreasing a number of periods in the maternal wavelet the window is extended along scale axis *s,* narrowing along time axis (Fig. 4).

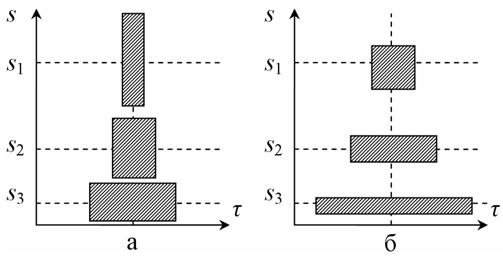


Fig. 4. *Window of time-frequency localization of wavelets with a number of periods a) N=2 and б) N=8*

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

1. frequency resolution (window width along the scale axis Δs) should support such resolution that frequencies of pitches of two adjacent notes are different;
2. time resolution (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 resolution of artificial basis wavelets. The aim of the experiment was to determine a number of periods in wavelet allowing identifying frequency scales m_i of all notes being in the signal simultaneously (at CWT wavelet scales *s* relative to the basis one are the equivalent to m_i).

Triad «C major» of the first octave was used in the experiment. Frequencies of note pitches of this triad corresponds to harmonic signals with frequencies 261,6, 329,6 and 392 Hz. Signal duration is chosen conventionally equal to 0,2 s:

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

The basis wavelets $fw_i(t)$ were developed for studying test signal. Quantity of periods of harmonic signal in wavelets was 1, 2, 4, 8 and 16 periods respectively (Fig. 5).

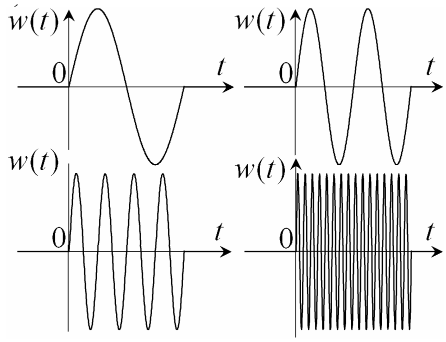


Fig. 5. *Maternal wavelets with a number of periods of harmonic signal 1, 2, 4 and 16*

CWT was carried out with all wavelet families $w_i(t)$ for test signal $f(t)$. For each transformation the results graphical interpretations were constructed in the form of three-dimensional models and isoline projection maps. Isoline projection maps of CWT results for the families of wavelets $w_i(t)$ are given in Fig. 6. Coordinate axis of each isoline projection map represents an axis of wavelet *s* scale, Abscissa axis – the axis of time shift τ, τ_0, τ_1 – time of start and termination of signal $f(t)$. Amplitude of CWT results is presented by shadows of gray: darker sections correspond to higher magnitude of CWT results (Fig. 6).

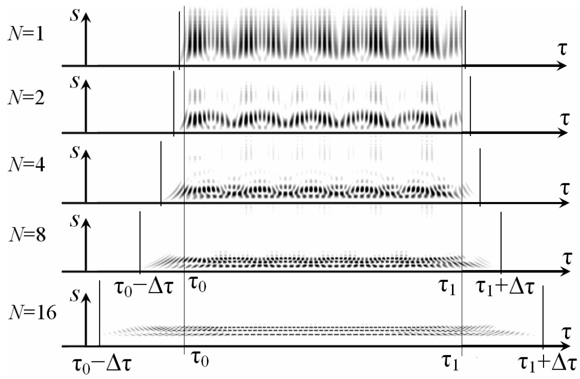


Fig. 6. *Graphical interpretations of CWT results for wavelets with different number of periods N of harmonic signal*

As it is seen from Fig. 6 using wavelet with one period a domain of uncertain results in the region of time shifts τ, with duration $\Delta\tau$ is rather low, therefore, the error of signal time estimate is low. In this case frequency resolution is so low that for the majority of scales *s* CWT results have high values indicating relatively even presence of frequency components on the whole frequency range of investigation. Though there only three harmonics in test signal $f(t)$.

It is seen from Fig. 6 that for wavelet with 16 periods over the 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 of the signal $f(t)$. For the rest values s, CWT results are practically equal to zero. Low time resolution forms uncertain results at the moments of occurrence and attenuation of signal with amplitude intermediate values on intervals with duration $\Delta\tau$. It does not allow judging the changes of test signal magnitude on this interval.

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

2. *Time resolution*

Note recording supposes using note and pause symbols for noting tune elements and gaps at which notes do not sound. Note duration as well as pause duration multiple to sounding duration t_1 of «whole» note – note of maximum possible duration. System of note recording consisting of alternating notes and pauses imposes strict constraints on the moments of note sounding start and pause start. Moments of note sounding start are sampled with a 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 note with duration $t_{64}=1/64\cdot t_1$. In practice notes with duration, t_{64} occur rather seldom due to the technical complicity of performance. Practically duration $t_{32}=1/32\cdot t_1$ may be considered the shortest note. Fragment of two-voice melody is given in Fig. 7. At each time moment not more than two notes simultaneously sound. Note 2 is the shortest one. Notes 1, 3, 4 equal in duration and two times longer than note 2 and a pause. If $t_d=t_{32}$ then note 2 and pause have the duration t_{32} and notes 1, 3, 4 – t_{16}.

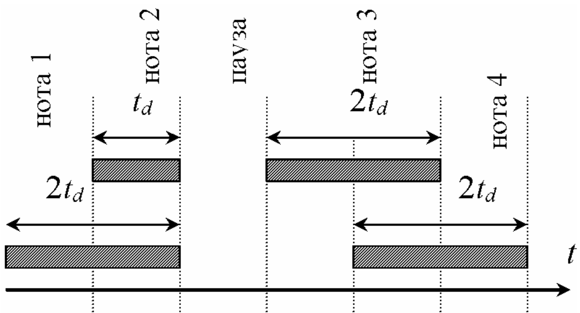


Fig. 7. *Time discretization of start of sounding and note and pause duration in a polyphonic tune*

At wavelet duration $T=t_d$ an envelope of joint correlation function $Wf_s(t)$ of wavelet and signal of one note coinciding with it by form (for concrete value s) regenerating into autocorrelation function has the form of equilateral triangle with maximum in the centre of note sounding [6] with the width $2t_d$ (Fig. 8, *a*). If all values $Wf_s(t)$ smaller than $Wf_{ош}(t)$ are ejected then note identification time is $t_и$. At $Wf_{ош}(t)=0,5Wf_{max}(t)$, $t_и=0,5t_d$. It means that note identification time equals to a half of its length.

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

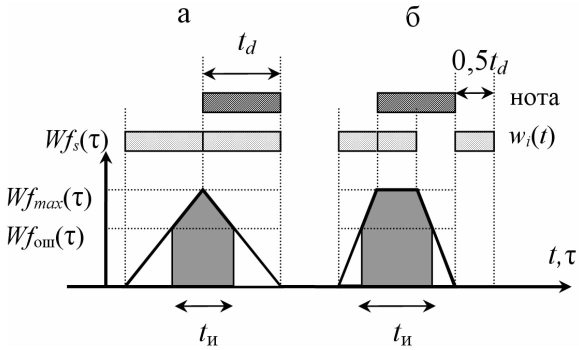


Fig. 8. *Envelops of correlation function of note and wavelet w(t) corresponding to the frequency of scale note pitch*

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

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

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

1. $T\geq 16/\upsilon$, where υ is the frequency of identified note pitch;
2. $T\leq t_d$.

Let us calculate the boundary values of note pitch frequency υ the reliable identification of which is possible both in time and frequency: $T=16/\upsilon$, $T=t_d$, therefore, $t_d=16/\upsilon$ or $\upsilon=16/t_d$. Dependence $\upsilon(t_d)$ determines minimal (boundary) frequency υ of note pitch with duration t_d which is definitely identified by a wavelet with 16 periods.

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

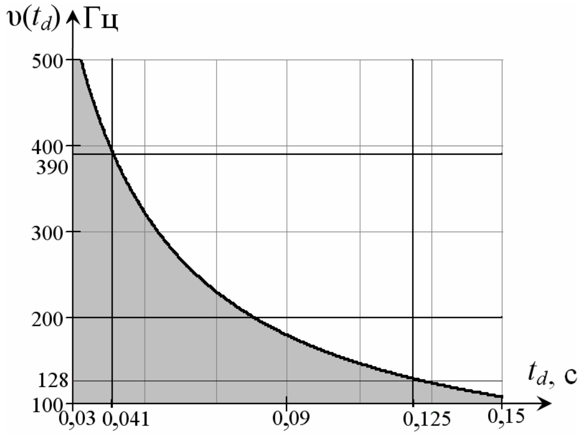


Fig. 9. *Dependence of boundary frequency of note pitch identification on its duration*

For slower tempo of composition performance or for tasks of detecting note with duration longer than t_{32} boundary frequency decreases. So, for example, for modern club dancing compositions reproducing tempo varies near the value 120 beats/min and the shortest note is $t_d=t_{16}=0{,}125$ s. Value of boundary frequency of note pitch identification $\upsilon(t_d)=16/0{,}125=128$ Hz. Just identification of time of note start and final sounding is possible only for notes higher than «C» of larger octave with pitch frequency 130,8 Hz. The majority of musical compositions use notes of the range of the smallest, first and second octaves that is higher than the range of large octave.

Conclusion

The suggested approach allows developing basis wavelet families with defined selectivity. Wavelet developed from a fragment of a certain basic signal allows detecting not separate time-frequency characteristics of the investigated signal but time-frequency behavior of basic signal in the investigated one.

Basis wavelets for instruments of various orchestra groups: piano, organ, violin, bell and trumpet were obtained experimentally. All wavelets contain 16 periods of a signal of a proper musical instrument with pitch frequency 55 Hz. Calculating CWT the maternal wavelet is scaled so that the next wavelet $w_i(t)$ coincides by pitch frequency with note pitch frequency $n_i(t)$ of a musical instrument.

Investigations of maternal wavelet families developed on the basis of musical instrument notes showed the possibilities of detecting frequency and time parameters of a certain instrument notes in one-voice and polyphonic tunes. Besides, it was managed to identify a tune of a certain musical instrument on the background of sounding of another one in a number of experiments [7].

This technique application in tasks requiring detecting in a signal the fragments of a certain families with finite length on the background of other signals or disturbances was suggested to be further investigated.

REFERENCES

1. Kudinov A.A. Use of pattern recognition for processing and restoration of musical signals: Thesis ... of a candidate of technical science. – Moscow, 2003. – 179 p.
2. Simanenkov I.O. Digital synthesizers of musical sounds // Kompyuterra. – 1998. – № 32. – P. 21–24.
3. Astafieva N.M. Wavelet-analysis: theory and examples of application // Uspekhi Fizitcheskikh Nauk. – 1996. – V. 166. – № 11. – P. 1145–1170.
4. Fadeev A.S., Kochegurova E.A. Selection of wavelet functions for analyzing musical information // Sovremennye Tekhnika i Tekhnologii: Proc. of the XII Mezhdunarodnoy Naucho-Prakticheskoy Konferentsii Studentov i Molodykh Uchenykh. – Tomsk: TPU, 2006 – V. 2. – P. 194–196.
5. Novikov L.V. Basis of signal wavelet-analysis. – St.-Petersburg: MODUS+, 1999. – 152 p.
6. Course of lectures on signals and systems of signal processing. [Electronic resource] / Author of the course is A.V. Davydov. – Electronic text data. (32 files). – [Ekaterinburg] Access mode: http://prodav.narod.ru/signals/index.html, free. – Russian, English languages.
7. Kochegurova E.A., Fadeev A.S. Wavelet analysis in the task of musical information identification // Molodezh I Sovremennye Informatsionnye Tekhnologii: Proc. of the IV Vserossiyskoy Nauchno-Prakticheskoy Konferentsii Studentov, Aspirantov i molodykh uchenykh. – Tomsk: TPU Press, 2006. – P. 149–151.
8. Smolentsev N.K. Wavelet theory. Wavelets in MatLab. – Moscow: DMK Press, 2005. – 304 p.
9. Fadeev A.S., Kochegurova E.A. Preparation of the results of continuous wavelet-transform to automated treatment // Bulletin of the Tomsk Polytechnic University. – 2006. – V. 309. – № 7. – P. 32–35.
10. Fujinaga I. Machine recognition of Timbre Using Steady-State tone of Acoustic musical Instruments // International Computer Music Conference. – Singapore, San Francisco, CA: ICMA, 2003. – P. 89–96.
11. Hebert S., Peretz I. Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? // Memory & Cognition. – 1997. – V. 25. – № 4. – P. 518–533.

Received on 19.10.2007