DOI https://doi.org/10.15407/csc.2022.03.039
UDC 004.383.3

A.K. SIERIEBRIAKOV, Ph.D. student (Eng.), Junior Researcher, Department of Intellectual Management, International Research and Training Center for Information Technologies and Systems of the NAS and MES of Ukraine, ave. Acad. Glushkov, 40, Kyiv, 03187, Ukraine, ORCID: https://orcid.org/0000-0003-3189-7968, sier.artem1002@outlook.com

YU.P. BOGACHUK, Ph.D. (Eng.), Senior Researcher, Department of Intellectual Management, International Research and Training Center for Information Technologies and Systems of the NAS and MES of Ukraine, ave. Acad. Glushkov, 40, Kyiv, 03187, Ukraine, ORCID: https://orcid.org/0000-0002-3663-350X, bip47@ukr.net

S.O. BONDAR, Ph.D. student (Eng.), Department of Intellectual Management, International Research and Training Center for Information Technologies and Systems of the NAS and MES of Ukraine, ave. Acad. Glushkov, 40, Kyiv, 03187, Ukraine, ORCID: https://orcid.org/0000-0003-4140-7985, seriybrm@gmail.com

V.M. SIMAKHIN, PhD student (Eng.), Senior Research, Department of Intellectual Management, International Research and Training Center for Information Technologies and Systems of the NAS and MES of Ukraine, ave. Acad. Glushkov, 40, Kyiv, 03187, Ukraine, ORCID: https://orcid.org/0000-0003-4497-0925, Thevladsima@gmail.com

FUNDAMENTAL FREQUENCIES CONTOUR EXTRACTION BASED ON THE EXTENDED HARMONIC-PERCUSSIVE SOURCE SEPARATION

Introduction. *Amidst the multiplicity of audio signal processing tasks connected with the necessity of source separation, the problem of obtaining of its most prominent components occupies special place. One of the approaches, dedicated to solve such task, is based on melody extraction from musical audio recording. If generalized, such technique can be applied to a wider class of audio signals for extraction from them of the so-called fundamental frequencies contour. For its realization, an attempt was made to combine the method of median filtering with the salience estimation method, for their application at various stages of the analysis of the input audio signal. A combination of methods is used to obtain the* f_0*-representation of the melody, based on the processing of the filtered values obtained at the first step.*

The purpose *of the article is to obtain the trajectory of* f_0*-values of the input audio signal and filtering of the corresponding to this trajectory harmonics.*

Results. *The proposed method is effective for use in audio signal processing systems for fundamental frequencies contour of the most prominent tonal components extraction and its further reuse.*

Conclusions. *Realized spectrum decomposition technique, based on the tendency of the time-frequency distribution of its constituent sounds, allows to effectively extract melodic contours from non-melodic contours. Nevertheless, there is a necessity for further research regarding the distribution model of harmonic and percussive characteristics relative to each other. Such a model should be extended with heuristics for more accurate filtering of stable in time tonal components of complex audio signals.*

Keywords*: fundamental tone, fundamental frequency trajectory, melody, harmonic-percussive source separation, audio signal, median filtering, salience function, musical tone, note, harmonic, melodic line, melodic contour, short-time Fourier transform, logarithmic frequency scale, harmonic summation.*

Introduction

Among the various types of signals that can be obtained from the environment, sound is one of the main sources of information about it. This is due to the physical properties of the acoustic waves and the manner of their longitudinal propagation in the air space, a distinctive feature of which is well-manifested diffraction, that is, the ability to bypass obstacles, because of the long length of such waves.

Since sound waves arise as a result of the oscillation of the air medium directly, obtaining their frequency and intensity is based on measuring the excitation of the corresponding oscillating surface (for example, the sensor membrane of a microphone or the tympanic membrane of a human ear). All this leads to the relative ease of sound waves recording for their analysis and processing, which requires corresponding low-cost equipment and alleviates a number of technical difficulties.

In addition, unlike other types of signals, sound waves contain a richer spectrum of frequencies. Its analysis can provide a relatively larger amount of useful information about the sound source and the surrounding environment.

This indicates the advantage of acoustic methods of obtaining information as often more accessible and effective in contrast to other methods of observation and monitoring. With the constructive use of the listed properties of sound waves, digitally processing them in the form of audio signals, it is possible to solve a number of tasks related to their source, such as localization and classification of objects, comparison of audio signals from different objects, creation of a signal source signature for storage in the database, and so on.

All this testifies to the relevance of the tasks, which deal with processing of signals coming from audio sources, and the systems that implement it. Since such audio signals are complex objects, there are numerous multilevel ways of interacting with them [1]. One of such methods is based on extracting some useful information from the signal, which can be used to solve the problem of audio source separation.

In this case, an important characteristic is the data on what exactly are the time-frequency distributions of the components' energy, that make up the received audio signal, assuming that they are distinctly different from each other. These distributions acquire horizontal and vertical directions, respectively, when viewed on a spectrogram (as will be discussed later). At the same time, the so-called tonal (horizontally distributed) components can contain representative information for signal identification. On the basis of the data about the most prominent frequencies distribution, it is possible to effectively solve the problems of audio source separation and its classification.

The implementation of this approach can be based on the use of one of the melody extraction methods from a musical audio recording, which is based on the harmonic summation of prominent signal frequencies (as discussed further). This approach should be generalized to obtain the so-called fundamental frequencies contour, which will allow its application to a wider class of audio signals.

In the generally accepted form, the melody extraction consists in automatically obtaining a representation of the main melodic line from an audio recording of a musical work.

A melody can be defined as a linear sequence of musical tones perceived by the listener as a single whole. In the simplest case, a musical tone is a periodic sound of a certain pitch, which has a sinusoidal waveform. Such a tone is called pure. Its graphic notation is called a note. If the emphasis is on a specific sequence of tones, the melody is called a melodic line. In the context of other musical components, it can also be called a melodic contour (as opposed to other, non-melodic contours) [2].

At the same time, which exact sequence of musical tones composes the melody is not strictly established. Further uncertainty arises when it is necessary to distinguish the main melody among several melodic lines that can be played simultaneously.

As stated in [3], the term melody is a musicological concept based on the judgment of listeners, so it can have different meanings in different contexts. The Music Information Retrieval (MIR) community has adopted the following definition: ". . . the melody is the single (monophonic) pitch sequence that a listener might reproduce if asked to whistle

or hum a piece of polyphonic music, and that a listener would recognize as being the 'essence' of that music when heard in comparison".

As mentioned above, melody can be represented by the trajectory of the fundamental frequencies, which corresponds to the pitches of the musical tones of the considered melodic line [4]. In the simplest case the fundamental frequency is the lowest frequency of some musical tone. In the case of a more complex aperiodic sound, it is still possible to define it as the lowest frequency among all others, which make up the sound at each moment of time. The fundamental frequency trajectory is also denoted as f_0-trajectory. Since it represents the melodic line, these terms will be used further interchangeably.

In addition to the f_0-trajectory, the harmonics of its musical tones can be extracted. As commonly known, harmonics are the additional tones of some sound, frequencies of which are integer multiples of its fundamental (that is, greater in 2, 3, 4, and so on times). Harmonics with other resonant frequencies above the fundamental are called overtones. Together with their fundamental, overtones form so called partials of a sound. Extracting harmonics in addition to the f_0-trajectory allows to get a richer timbral portrait of the melody [5, 6].

It is important that the elements of a musical composition are hierarchically interconnected. In it, the melody tries to express the "main message" of the work's author, due to which it is brought to the foreground with the help of a special musical means called accompaniment. Having information about the melodic progress over time, it is possible to draw conclusions about the remaining components of the audio signal. This results from the fact that musical events follow a certain dynamic logic, which guides when they play relative to each other [7]. Also, the extracted melody makes it possible to reuse it in a different context (for example, to play it to a different accompaniment, to modify it, to convert it to symbolic notation, to mix it with other musical tracks, and other).

Many different methods for melody extraction have been proposed [8]. One approach attempts to identify a melody by separating it from other sounds using timbral source separation techniques. Such systems use two different timbre models, one for the melody (which can be sung by voice) and the other for the accompaniment. Some systems incorporate grouping principles inspired by the analysis of auditory scenes, most commonly based on frequency proximity.

Algorithms using spatial information were also proposed. They use stereo information to estimate the panning of each source, and use a generative model (source-filter) to identify and extract melody.

Moreover, approaches based exclusively on data were studied. According to them, the entire short-term amplitude spectrum is used as a training data for setting up the vector machine classifier.

A separate group of methods for melody extraction is represented by various methods of estimating the tonal energy of a signal. Most of them are based on obtaining a salience function, which gives an estimate of the pitch prominence over time. Such systems follow a single general structure: first, a spectral representation of the signal is obtained; next, the time-frequency representation of pitches salience is calculated. The peaks of the salience function are considered as potential candidates for extraction as melody.

A similar group of methods, related to the task of melody extraction, is known as harmonic-percussive source separation [9]. These methods are based on the fact that tonal, or harmonic sounds form stable horizontal ridges on spectrograms over time, while percussive sounds are distributed on them vertically due to their noise-based broadband nature [10]. Algorithms that exploit this property could be based on anisotropic diffusion, Bayesian models, and others.

The simple and effective approach distinguished among them is based on median filtering [11]. According to it, harmonic features of the signal are filtered by a horizontal median filter relative to its percussive counterparts, which are filtered by a vertical one. Thus, the spectral values, which potentially contain melody are filtered into separate dataset for further analysis.

In the current paper, an attempt was made to combine the median filtering method with the salience estimation method, with their application at

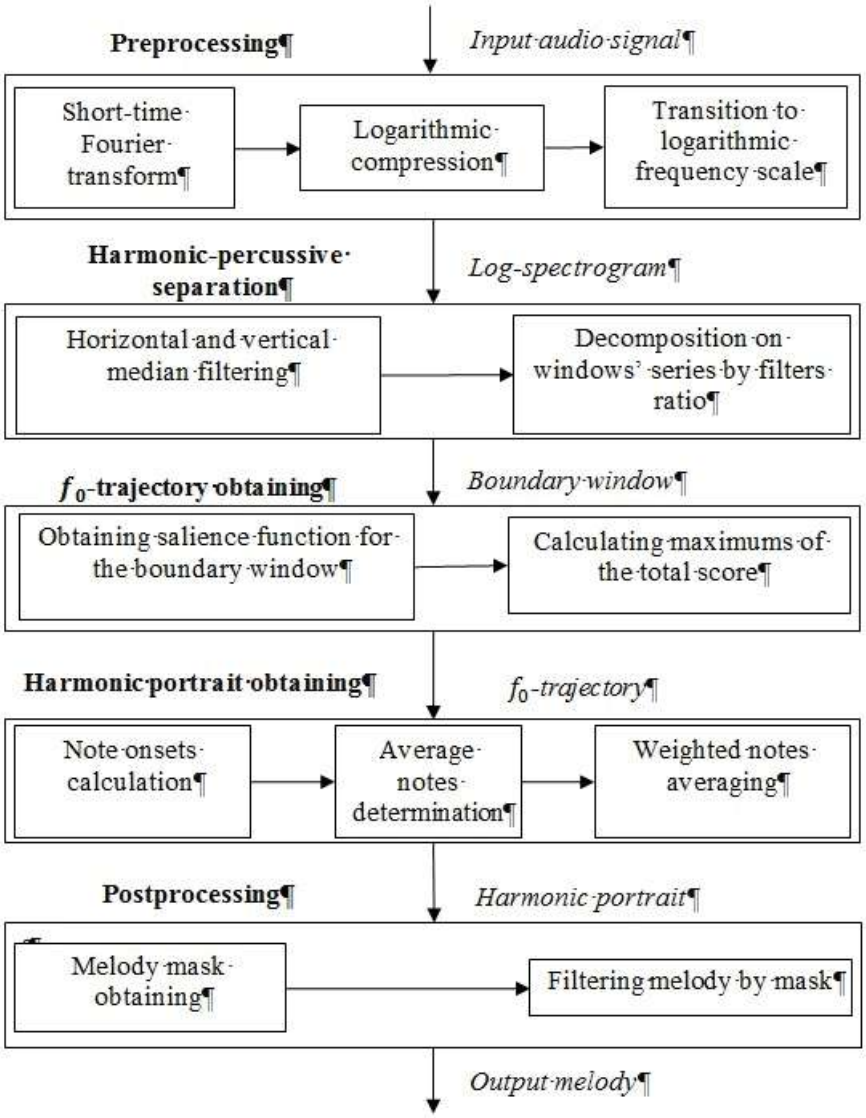

Fig. 1. The proposed algorithm for melody extraction

different stages of the analysis of the input audio signal [12, 13].

Problem Statement

The idea of the proposed algorithm is to expand the well-known harmonic-percussive source separation technique, which uses median filtering of the signal spectrum. It consists in decomposing of the received filtering result into a series of windows that characterize the gradual transition from percussive to tonal sounds [14]. Next, a window of values is selected from the obtained series, for which the salience function is calculated, as well as the attack

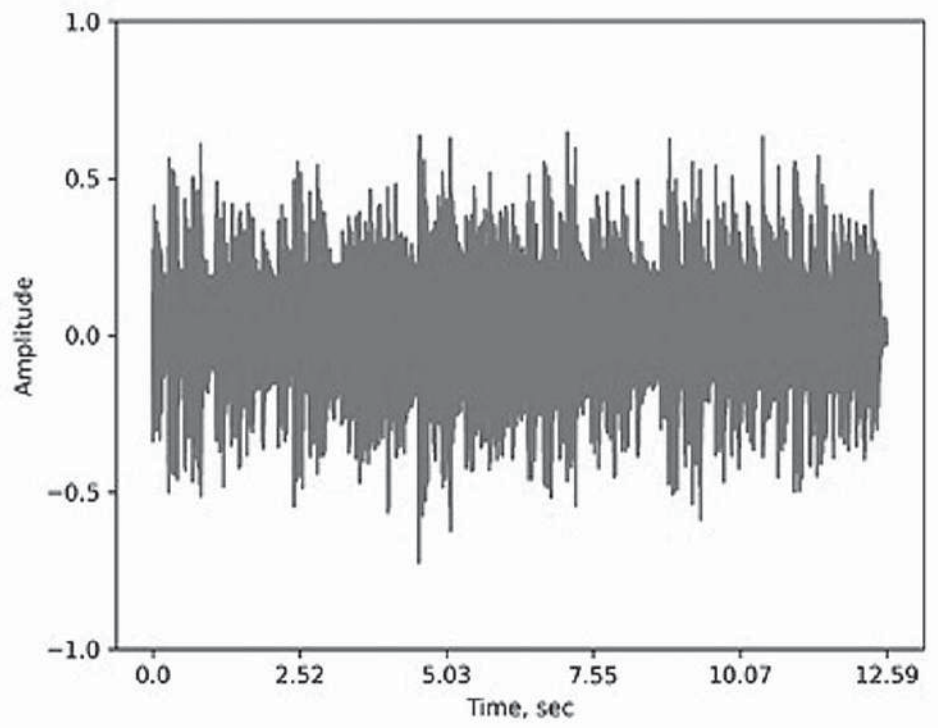

Fig. 2. The input signal waveform

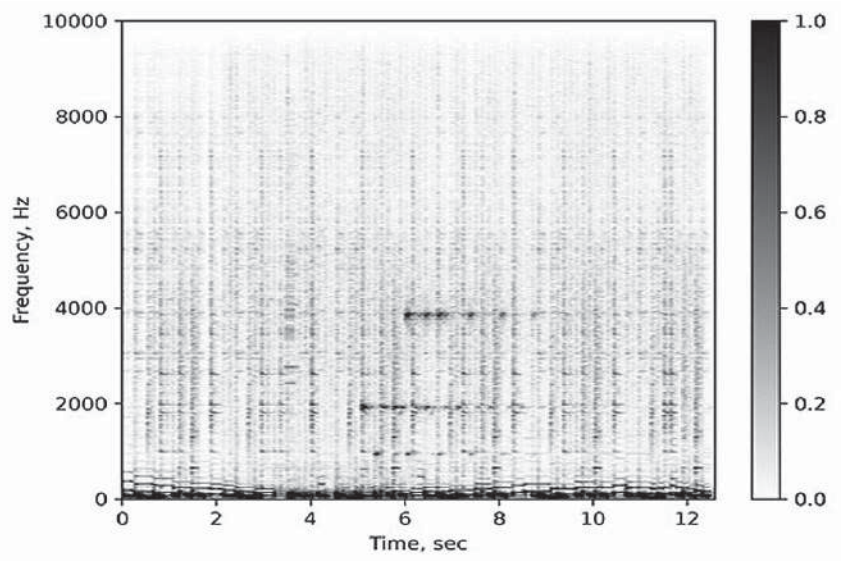

Fig. 3. Logarithmically compressed spectrogram Γ_γ

onsets for played notes are determined. Based on obtained data, averaged notes are processed from the obtained series of windows, which are then additionally summed up together. The received harmonic portrait allows to filter tonal musical events from the input signal.

For goal realization several restrictions have been adopted. First, algorithm receives a sound file as input, which is a musical audio recording with some melody present in it. Melody is considered as a sequence of the most salient fundamental frequencies. It is assumed that the energy of the melo-

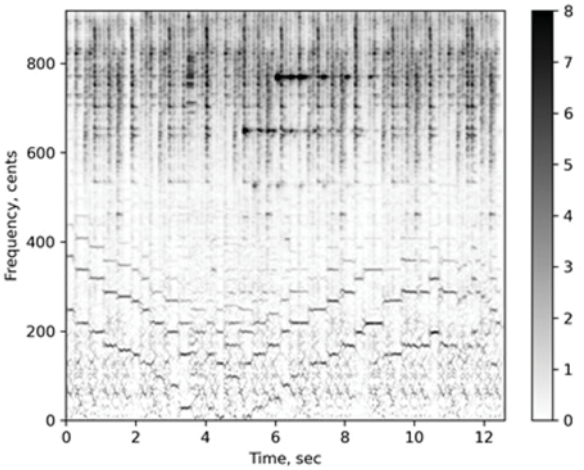

Fig. 4. The resulting spectrogram with a logarithmic frequency scale

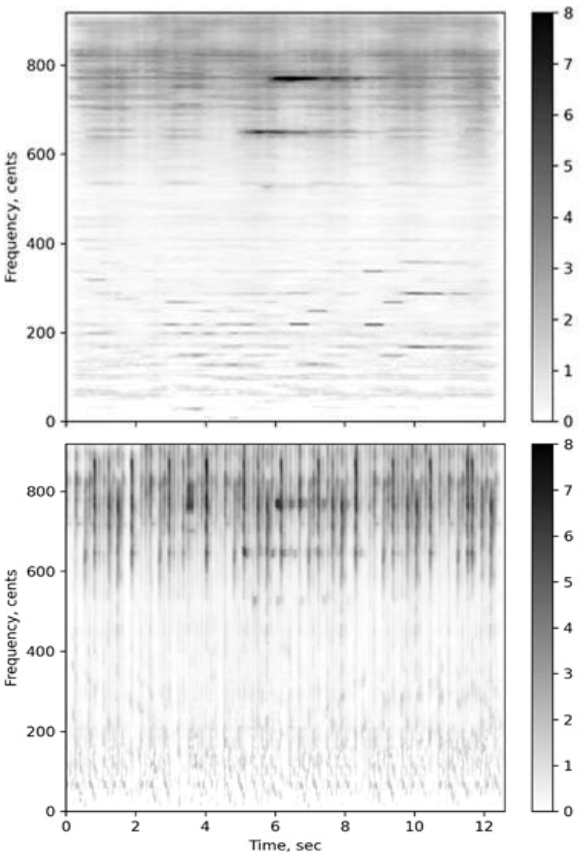

Fig. 5. Values after horizontal (top) and vertical (bottom) filtering

dy in the audio signal prevails over its other components. That is, there is some prominent melodic line in the signal, which can be associated with one or several sound sources (instruments).

Thus, the purpose of the algorithm and research is to extract the f_0-trajectory of the input audio signal and its corresponding harmonics.

The algorithm consists of 5 main subprocedures, each containing several stages, as represented by the block diagram in Fig. 1.

Let's consider each stage of the algorithm in greater detail.

Preprocessing

A recording of the audio signal is given for further processing using the short-time Fourier transform. As input data an excerpt from a musical composition called "Voltaic Algebraic" by Brian Ales is taken, with duration of approximately 12,5 seconds. In it, a simple sequence of notes is played with a certain rhythm, which is clearly established by percussion instruments. This piece is taken as an example for illustrative purposes.

Let's get a discrete signal x from the input recording with a sample rate of $F_s = 22050$ Hertz. The resulting waveform x is shown in Fig. 2. Horizontal axis represents time given in seconds, and the relative amplitude is represented on the vertical axis.

Let's apply the short-time Fourier transform to x. Let $w:[0:N-1]$ be a window function of length N. The last value determines the duration of one sample, which is equal to N / F_s seconds. Also, let's introduce the hop size H, equal to the number of samples by which the window should be shifted along the signal x. According to the above, the discrete Fourier transform of the signal x is given in the form

$$\chi(m,k) := \sum_{n=0}^{N-1} x(n+mH)w(n)e^{-\frac{2\pi ikn}{N}},$$

where $m \in [0:V]$ and $k \in [0:K]$, V is the number of samples, $K = N/2$ is the index of the Nyquist frequency.

In order to balance the difference between the large and small values of calculated χ, logarithmic compression should be carried out. Let's denote this operation as Γ_γ and apply it to each of these values: $\Gamma_\gamma(m,k) := \log_{10}\left(1+\gamma\cdot|\chi(m,k)|\right)$.

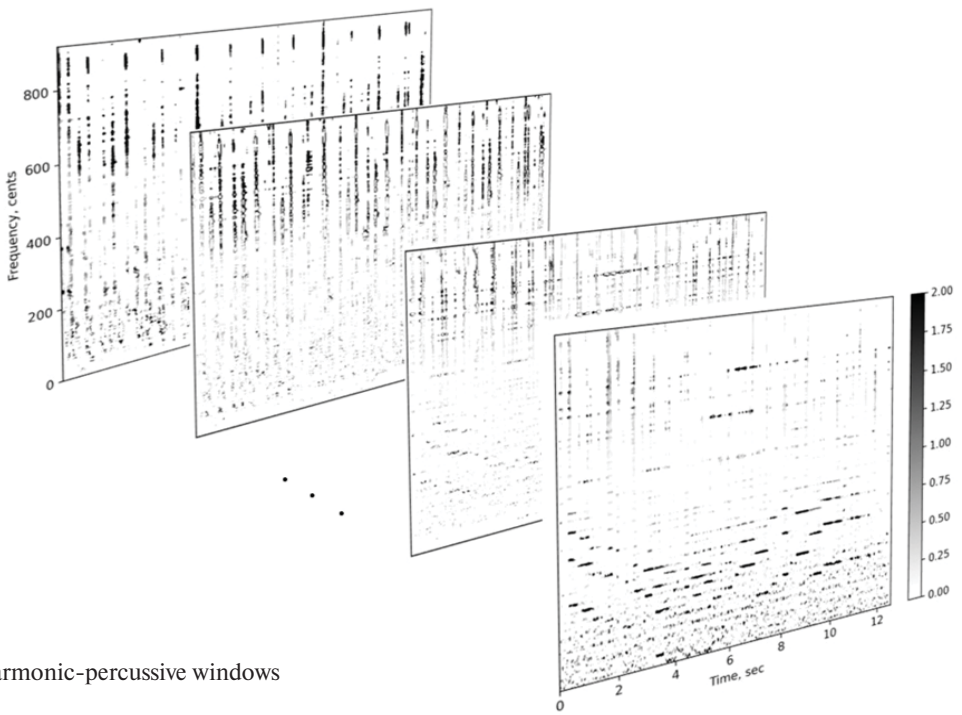

Fig. 6. Received harmonic-percussive windows

Where is the scaling factor. The obtained values are displayed on a spectrogram, as shown in Fig. 3. For practical purposes, only frequencies up to 10,000 Hertz are considered. The relative intensity of the values is shown by the color saturation (specified by the colorbar on the right).

The next step is the transition to a logarithmic frequency scale from a linear one [15]. This is motivated by the fact that the harmonics of several musical sounds with the same timbre, but played at different pitch, are located on a linear frequency scale relative to each other logarithmically. Transition to logarithmic scale makes the intervals between harmonics to appear as linear, which will simplify the comparison of the several sounds' partials with different fundamental frequencies.

Let ω_{ref} be the reference frequency given to the first interval. Also, let R be the desired resolution of the logarithmic frequency axis (given in cents). Then for the frequency $\omega \in R$ (specified in hertz) the index of the interval $\mathrm{Bin}(\omega)$ is defined as

$$\mathrm{Bin}(\omega) := \frac{1200}{R} \cdot \log_2\left(\frac{\omega}{\omega_{ref}}\right) + 1{,}5.$$

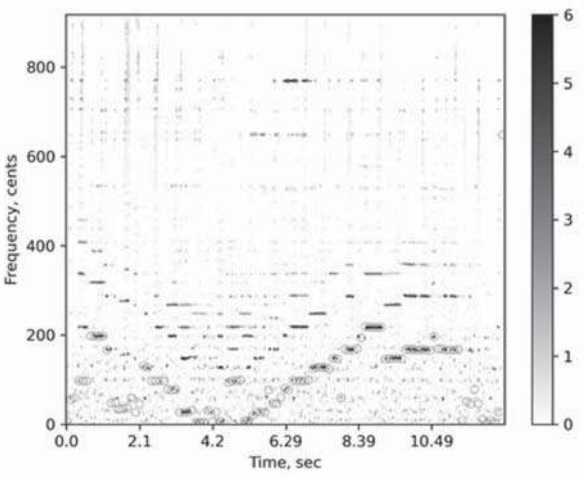

Fig. 7. Boundary harmonic window and highlighted fundamental frequency trajectory

For each interval $b \in [1:B]$, where B is the number of logarithmic intervals, the set is defined

$$P(b) := \left\{k : \mathrm{Bin}\left(F_c(k)\right) = b\right\},$$

where $F_c(k) = k\frac{F_s}{2K}$ is the formula for converting the index $k \in [1:K]$ to the corresponding frequency.

Further calculate the sum:

$$Y_b(n,b) = \sum_{k \in P(b)} \Gamma_\gamma(n,k),$$

where n is the time interval index and k is the frequency interval index. The resulting spectrogram with a logarithmic vertical scale is shown in Fig. 4.

Extended Harmonic-Percussive Source Separation

Taking into account property of the musical sounds, which make up the melody, to distribute on the spectrogram horizontally, let's apply median filtering to obtained values Y_b in horizontal and vertical direction.

Let's formalize the procedure in the context of the considered task. Let $A = (a_1, a_2, \ldots, a_L)$ be a list of length L consisting of numbers $a_\ell, \ell \in [1:L]$. At first, all elements of A are sorted in ascending order. As a result list A is obtained where $a_\ell \le a_m$ for $\ell < m$ and $\ell, m \in [1:L]$. Next, the median $\mu_{1/2}(A)$ of the list A is defined as

$$\mu_{1/2}(A) = \begin{cases} a_{(L+1)/2}, & \text{for } L/2 \notin \mathbb{Z}, \\ (a_{L/2} + a_{L/2+1})/2, & \text{for } L/2 \in \mathbb{Z}. \end{cases}$$

The local application of the median to some sequence of numbers is also called a median filter of length L. Let's apply the median filter to Y_b horizontally and vertically to get Y_h and Y_p respectively:

$$Y_h(n,k) := \mu_{1/2}\big(Y_b(n-(L_h-1)/2,k),\ldots$$
$$\ldots, Y_b(n+(L_h-1)/2,k)\big),$$
$$Y_p(n,k) := \mu_{1/2}\big(Y_b(k,n-(L_p-1)/2),\ldots$$
$$\ldots, Y_b(k,n+(L_p-1)/2)\big),$$

where L_h and L_p are the lengths of the respective filters. The results of filtering are shown in Fig. 5.

Let's calculate the ratio of the harmonic component to the percussive component:

$$Y_r(n,k) = \frac{Y_h(n,k)}{Y_p(n,k)}.$$

Values Y_r for timbrally similar musical sounds are more likely to be in the same range. Taking this into account, let's decompose ratio Y_r on certain sets of values and call them a series of harmonic-percussive windows, each of which will contain values lying in a given range, as follows:

$$Y_{K_i}(n,k) = \begin{cases} Y_r(n,k), & \text{for } k_1 \ge Y_r(n,k) > k_2 \\ 0, & \text{otherwise} \end{cases},$$

where $K_i = \{k_1(i), k_2(i)\}$ is a pair of values that define the values range, $i \in [1,M]$, M is the number of windows. $k_1(i)$ and $k_2(i)$ are successively chosen so that the number of non-zero values Y_r, which are located between them, is equal to the arbitrarily given constant m. The obtained harmonic-percussive windows are shown in Fig. 6.

Fig. 6 demonstrates, that the values, which are more related to percussive musical sounds, are located at the initial intervals K_i.

With a gradual increase of i, more harmonic values are taken into the windows.

For further consideration, let's take the values of the boundary window, for which $i=M$, since it contains the most salient harmonics of the signal (shown separately in Fig. 7): $Y_f = Y_{K_M}(n,k)$.

Obtaining f_0-trajectory

Let's carry out the harmonic summation for the obtained result Y_f. Let H be the number of harmonics to be summed. Then, for the values Y_f obtain

$$Z(n,b) := \sum_{h=1}^{H} \alpha^{h-1} \cdot Y_f\left(n, b + \frac{1200}{R}\log_2(h)\right),$$

where $\alpha \in [0,1]$ is a weighting factor that allows to exponentially scale (decrease) harmonics' magnitude.

Because Z demonstrates the relative energy of the tonal components, these values are also called the salience function. They can be considered as a measure that expresses the probability that frequency at $b \in [1:B]$ corresponds to the dominant tonal component at the time $n \in [1:N]$.

Next, let $T \in R^{B \times B}$ be the transition matrix, where the value $T(b,c)$ expresses the probability of

the transition from $b \in [1:B]$ and $n \in [1:N-1]$ to $c \in [1:B]$ and $n+1$.

Considering the salience function Z and the transition matrix T, the total score $\sigma(\eta)$ can be associated with the trajectory $\eta : [1:N] \rightarrow [1:B]$:

$$\sigma^{\max}(b,n) = \max_{b \in [1,K]} \left(\sigma^{\max}(b,n-1) + Z(b,n) + T(b) \right)$$

where K is the Nyquist frequency index.

The requested trajectory η^{DP} in this case is defined as the maximum of the total score:

$$\eta^{DP}(n) := \underset{\eta}{\operatorname{argmax}}\, \sigma^{\max}(\eta).$$

The maximum total score can be calculated using dynamic programming (similar to the Viterbi algorithm). Moreover, the trajectory η^{DP} is obtained using suitable backtracking procedure. The resulting f_0-trajectory for Z is shown in Fig. 7.

Obtainingthe Harmonic Portrait

As commonly known, there is often a sudden increase in energy at the beginning of a musical tone, which is called the attack of the note. According to this, the note onset is the moment in time (not the interval) that marks the beginning of the transient process, or the early point in time at which this process can be reliably detected, as shown in Fig. 8.

Based on this, it is possible to detect the onsets for the obtained values Y_f, identifying the intervals of sudden energy changes, which indicate the beginning of the transient regions.

Let's extract the values along the trajectory η^{DP} as follows:

$$B_{K_i}(n,b) = Y_{K_i}\left(n, \eta^{DP}(n) + b\right),$$

where $b \in [0, H_H]$ is a frequency shift upwards, H_H is the frequency interval containing harmonics to be considered. Actually, the trajectory of the fundamental frequencies T_f is obtained when b=0:

$$T_f(n) = B_{K_i}(n,0).$$

Obtained values for the boundary window B_{K_M} are shown in Fig. 10.

The next step is to calculate the discrete time derivative of B_{K_M}. To do this, let's consider only positive differences (increase in intensity) and discard negative ones. Summing up obtained differences

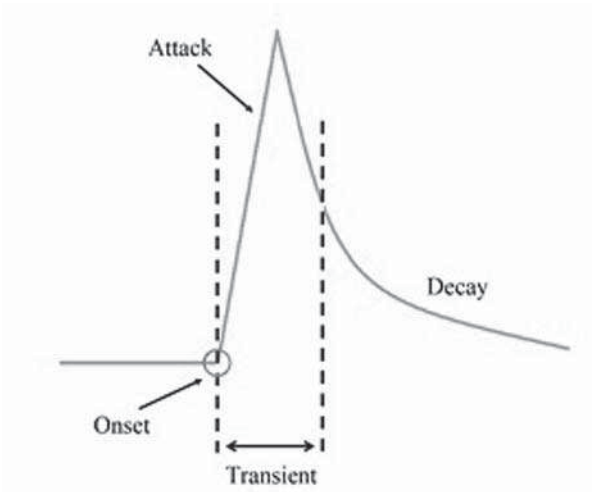

Fig. 8. Illustration of the note onset

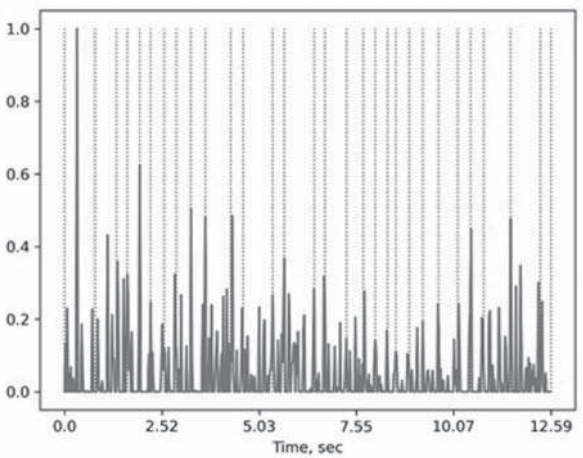

Fig. 9. Calculated novelty function and onsets (indicated by vertical lines)

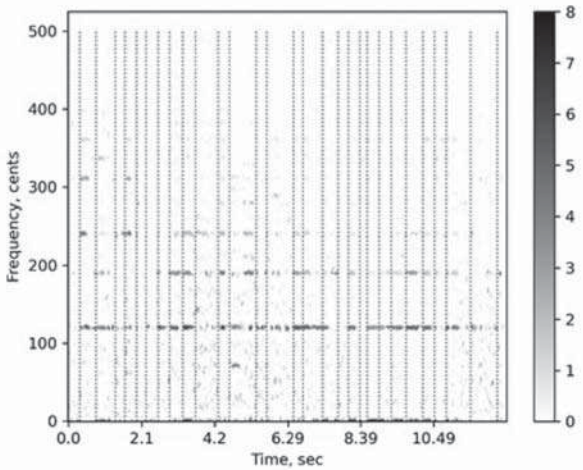

Fig. 10. The values of the boundary harmonic window taken along the obtained trajectory, with designated onsets

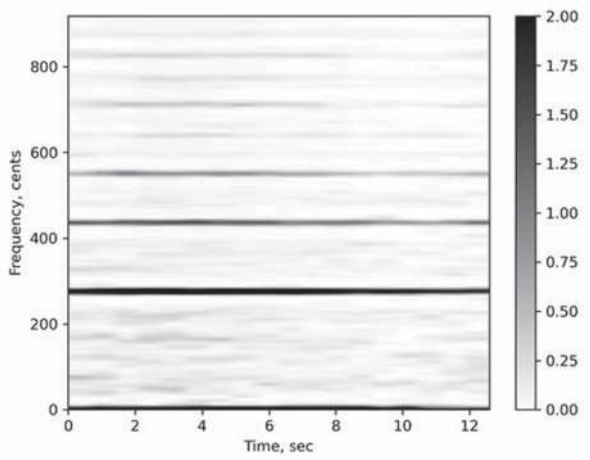

Fig. 11. The received harmonic portrait of the melody

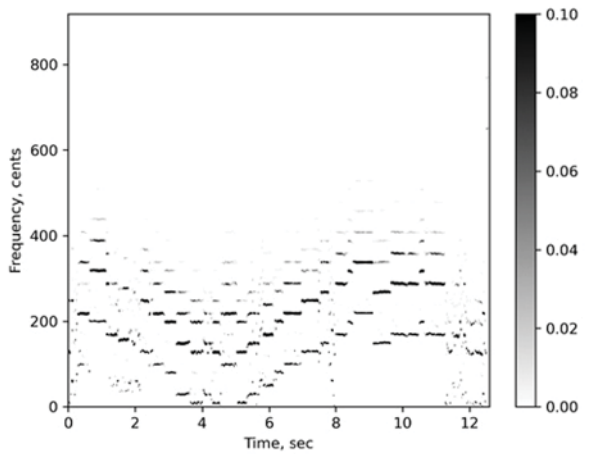

Fig. 12. Melody mask on a logarithmic frequency scale

along the frequency axis gives the spectral novelty function Δ_S:

$$\Delta_S(n) := \sum_{k=0}^{K} \left| B_{K_i}(n+1,k) - B_{K_i}(n,k) \right|_{\geq 0}.$$

Let's process the function: strengthen the peak structure while suppressing small deviations. For this, introduce the local average function μ:

$$\mu(n) := \frac{1}{2M_l + 1} \sum_{m=-M_l}^{M_l} \Delta_S(n+m),$$

where M_l is a parameter that determines the size of the averaging window. Enhanced novelty function $\overline{\Delta}_S$ is obtained by subtracting the local average from Δ_S and leaving only the positive part:

$$\overline{\Delta}_S(n) := | \Delta_S(n) - \mu(n) |_{\geq 0}$$

additionally, the obtained function is normalized by its maximum value.

To reduce the impact of noise and other distortions, a smoothing filter should be applied to the novelty function. Moreover, instead of considering a global threshold for rejecting small noisy peaks, let's apply an adaptive thresholding technique to select peak only if its value exceeds the local average value of the novelty function. Let's denote this operation as $p_{K_i} = MSAF\left(\overline{\Delta}_S(n)\right)$, where p are the peak indices of the input novelty function.

To include all values of the spectrum from the beginning to the end, add the initial and final indices to the set of peaks:

$$o = \{1, p, N_l\}.$$

Where N_l is the number of time samples (the length of the novelty function). The obtained set contains the requested onsets. The results are shown in Fig. 9 and 10.

Let's take the values that are between two consecutive onsets, which correspond to the time intervals when individual notes are played:

$$N_{K_i}(j) = B_{K_i}(n_o, b), \qquad (1)$$

where $n_o \in \left[o(j) : o(j+1) \right]$ is the interval between adjacent onsets, $j \in \left[1 : O-1 \right]$, where O is the number of onsets.

Since the notes may have different duration (have different lengths), for further operations they should be reduced to one, average note length, which is defined as

$$L_a = \frac{\sum_{j=1}^{O-1} \left| o(j+1) - o(j) \right|}{O-1}.$$

Let 's assume that $R_s = f\left(Y_n, L\right)$ is the function which stretches or squeezes input data to length L_a. Then calculate the arithmetic mean of the received notes for each previously obtained window:

$$N_{K_i}^a = \frac{\sum_{j=1}^{O-1} R_s\left(N_{K_i}(j), L(j)\right)}{O-1},$$

where $L(j) = \left| o_{K_i}(j+1) - o_{K_i}(j) \right|$ is the length of the interval between adjacent onsets.

The procedure described above is applied to all K_i windows, after which the obtained average notes are summed up using a suitable weighting coefficient $\alpha \in [0,1]$:

$$N_p = \sum_{i=1}^{M} N_{K_i}^{a} \cdot \alpha^{M-i}.$$

Here, the first window will have the smallest coefficient α, meanwhile the average note of the last window will be taken without reduction, since with a gradual transition to the last window, the requested harmonics will become more salient.

Let's calculate the harmonic portrait of the melody (Fig. 11), normalizing the obtained average note by its maximum value, reducing it to the interval [0,1]:

$$H_p = \frac{N_p}{\max N_p}.$$

The figure clearly demonstrates both the extracted fundamental, as well as its accompanying harmonics.

Post-Processing

Let's take the notes between the found onsets from the values Y_b similarly to how it was done in (1). Let's designate the obtained notes as $Y_{nb}(j)$.

Then, multiply each note by the obtained portrait (while changing the length of the portrait according to the length of the note being multiplied):

$$Y_{nf}(j) = Y_{nb}(j) \cdot R_s\left(H_p, L(j)\right).$$

Place received notes along the trajectory η^{DP} on respective height. Let's denote these values as $Y_{bf}(n,k)$. It should be noted that the obtained spectrum contains only information in the filtered intervals $\eta^{DP}(n): \eta^{DP}(n) + H_H$. Other frequencies are not taken into account (equated to zero).

Let's calculate the melody mask by normalizing notes' sequence (Fig. 12):

$$M_B = \frac{Y_{bf}(n,k)}{\max\left(Y_{bf}(n,k)\right)}.$$

Comparing the resulting figure with the original data in Fig. 4 gives actual representation of exactly which values were filtered.

Since the above calculations were performed for logarithmically compressed values, it is necessary to apply inverse operation to the M_B:

$$M_L = \frac{e^{M_B} - 1}{\gamma}.$$

To filter out values with the help of received notes, multiply the initial values χ on M_L, by mapping them to the linear frequency scale:

$$X_f(n,k) = \chi(n,k) \cdot \sum_{k \in P(b)} M_L(n,b).$$

Finally, let's transform the filtered values into a reproducible signal waveform using the inverse Fourier transform.

Thus, it is possible to obtain the melody from an input audio recording in the form of a sound signal shown in Fig. 13.

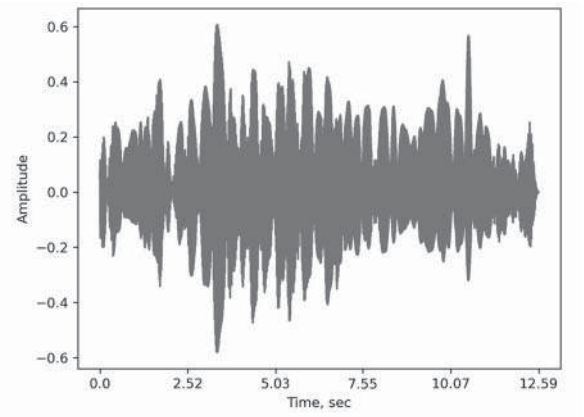

Fig. 13. Melody output waveform of the filtered melody

Conclusions

This article presents an approach for automatically extracting the fundamental frequencies contour from the input audio signal. Several stages of signal processing were described, namely: its filtering into a series of harmonic-percussive windows; obtaining the salience function for the purpose of evaluating the f_0-trajectory; and the procedure of weighted windows averaging to obtain a harmonic portrait for further melody filtering.

It was shown that the used decomposition of the spectrum of the audio signal, based on the tendency of the time-frequency distribution of its constit-

uent sounds, allows to effectively extract melodic contours from non-melodic contours. Next, it was shown how these contours together affect the filtered result. All this makes it possible to identify when and at what pitch the main melody is played [16]. This leads to the minimization of errors associated with the determination of the exact pitches of the notes of the melodic contour being played, as well as the formation of a set of all frequencies that are included in these complex tones.

The novelty of the algorithm lies in taking into account the relationship between the percussive features and the melody. The gradual transition from vertical to horizontal components allows to take into account the influence of the rhythmic pattern of the accompaniment on the process of the melody development over time. By changing the weight of the windows, it is possible to control the intensity of this transition. It also allows to enrich the filtered melodic contour with sounds that would otherwise be erroneously identified as more percussive. This can refer to sharp and abrupt sounds, which are characterized by short notes attack intervals or their abrupt changes. Directly obtaining a harmonic portrait of the most salient tonal components provides an opportunity for a more rigorous formalization of what constitutes a melodic contour.

Despite these advantages, certain difficulties arise when applying the approach. One of them is connected with informal rules of transition between windows, in addition to choosing the optimal window for further analysis. A smoother and more flexible transition is needed based on some model of harmonic-percussive distribution, as well as a generalization of the rules for applying weight to windows, which affects their averaging. Another problem is the assumption that a single melody is dominant in an audio recording. While this may be correct in some cases, the algorithm will perform worse for polyphonic compositions in which multiple musical lines, or even accompaniment, are played with the same or similar timbre.

The basis for further research should be the development of a model of the distribution of harmonic-percussive characteristics relative to each other. It should be supplemented with heuristics in the necessity of more accurate filtering of various melodic lines of polyphonic music. f_0-trajectories to be obtained for each filtered window should iteratively influence the next steps of the filtering procedure. For further work, the development of structural analysis is also valuable, namely the formalization of relationships between individual tonal sounds and percussion events that emphasize them.

In general, the proposed method is promising for its application in audio signal processing systems for obtaining the fundamental frequencies contour and its further reuse. It is an effective approach that is conceptually simple and computationally fast at the same time.

REFERENCES

1. *Sargent, G., Bimbot, F., & Vincent, E.* (2016). "Estimating the Structural Segmentation of Popular Music Pieces Under Regularity Constraints". IEEE/ACM Transactions on Audio, Speech, and Language Processing. pp. 344 –358. DOI: 10.1109/TASLP.2016.2635031.
2. *Canadas-Quesada, F. J., Vera-Candeas, P., Ruiz-Reyes, N., Carabias-Orti, J., & Cabanas-Molero, P.* (2014). "Percussive/ harmonic sound separation by non-negative matrix factorization with smoothness/sparseness constraints". EURASIP Journal on Audio, Speech, and Music Processing. 10.1186/s13636-014-0026-5.
3. *Bosch, J. J., Marxer, R., & G mez, E.* (2016). "Evaluation and combination of pitch estimation methods for melody extraction in symphonic classical music", Journal of New Music Research, 45 (2), pp. 101-117, DOI: 1080/09298215.2016.1182191.
4. *Salamon, E., Gomez, D.P., Ellis, W., Richard, G.,* 2014. "Melody Extraction from Polyphonic Music Signals: Approaches, applications, and challenges," IEEE Signal Processing Magazine, vol. 31, no. 2, pp. 118-134, March 2014, doi: 10.1109/MSP.2013.2271648.
5. *Raczy ski, S. A., Ono, N., & Sagayama, S.* (2007). "Multipitch Analysis with Harmonic Nonnegative Matrix Approximation". 8th International Conference on Music Information Retrieval, pp. 381-386.

6. *Cella, C.-E.* (2010). "Harmonic Components Extraction in Recorded Piano Tones". 128th Audio Engineering Society Convention. 2010, May 22–25 London, UK. http://www.aes.org/e-lib/browse.cfm?elib=15408.
7. *Prince, J.* (2014). "Contributions of Pitch Contour, Tonality, Rhythm, and Meter to Melodic Similarity". Journal of Experimental Psychology Human Perception & Performance. In press. 10.1037/a0038010.
8. *Salamon, J., Peeters, G., & R bel, A.* (2012). "Statistical characterisation of melodic pitch contours and its application for melody extraction". Proceedings – 13th International Society for Music Information Retrieval Conference, pp. 187-192.
9. *Canadas-Quesada, F. J., Vera-Candeas, P., Ruiz-Reyes, N., Munoz-Montoro, A., & Bris-Penalver, F. J.* (2016). "A method to separate musical percussive sounds using chroma spectral flatness". SIGNAL, 51.
10. *Fitzgerald, D., Liutkus, A., Rafii, Z., Pardo, B., Daudet, L.,* (2014). "Harmonic/Percussive Separation Using Kernel Additive Modelling". In 25th IET Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies, pp. 35-40. DOI: 10.1049/cp.2014.0655.
11. *Fitzgerald, D.*, (2010). "Harmonic/Percussive Separation using Median Filtering". 13th International Conference on Digital Audio Effects (DAFx-10).
12. *Driedger, J., M ller, M., Disch, S.*, (2014). "Extending Harmonic-Percussive Separation of Audio Signals". In ISMIR, pp. 611-616.
13. *F g, R., Niedermeier, A., Driedger, J., Disch, S., & M ller, M.* (2016). "Harmonic-percussive-residual sound separation using the structure tensor on spectrograms," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 445-449, doi: 10.1109/ICASSP.2016.7471714.
14. *Ono, N., Miyamoto, K., Le Roux, J., Kameoka, H., & Sagayama, S.* (2008). "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram". In 2008 16th European Signal Processing Conference, IEEE., pp. 1-4.
15. *Degani, A., Leonardi, R., Migliorati, P., Peeters, G.* (2014). A Pitch Salience Function Derived from Harmonic Frequency Deviations for Polyphonic Music Analysis. In DAFx, pp. 195-201. 13140/2.1.4409.7922.
16. *Zhang, W., Chen, Z., Yin, F.* (2018). "Melody Extraction Using Chroma-Level Note Tracking and Pitch Mapping". Applied Sciences. 8 (9), 1618, 10.3390/app8091618.

Received 01.10.2022

А.К. Сєребряков, аспірант, відділ інтелектуального управління, Міжнародний науково-навчальний центр інформаційних технологій і систем НАН та МОН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, ORCID: https://orcid.org/0000-0003-3189-7968,
sier.artem1002@outlook.com
Ю.П. Богачук, к.т.н., провідний науковий співробітник, відділ інтелектуального управління, Міжнародний науково-навчальний центр інформаційних технологій і систем НАН та МОН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, ORCID: https://orcid.org/0000-0002-3663-350X ,
bip47@ukr.net
С.О. Бондар, аспірант, відділ інтелектуального управління, Міжнародний науково-навчальний центр інформаційних технологій і систем НАН та МОН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, ORCID: https://orcid.org/0000-0003-4140-7985,
seriybrm@gmail.com
В.М. Сімахін, аспірант, відділ інтелектуального управління, Міжнародний науково-навчальний центр інформаційних технологій і систем НАН та МОН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, ORCID: https://orcid.org/0000-0003-4497-0925,
Thevladsima@gmail.com

ОТРИМАННЯ КОНТУРА ОСНОВНИХ ТОНІВ НА ОСНОВІ РОЗШИРЕНОГО ГАРМОНІЙНО-ПЕРКУСІЙНОГО РОЗДІЛЕННЯ ДЖЕРЕЛА

Вступ. Серед задач обробки сигналів, які виникають при необхідності розділення аудіоджерела, особливе місце посідає проблема отримання його характерних, найбільш виражених компонент. Один з підходів, який вирішує поставлене завдання, базується на виділенні мелодії з музичного аудіозапису. Такий підхід може бути використано для більш широкого класу аудіосигналів при його узагальненні для отримання з них так званого контура основних тонів. Для його реалізації зроблено спробу об'єднання методу медіанної фільтрації з методом оцінки значущості,

для їх застосування на різних етапах аналізу вхідного аудіосигналу. Комбінація підходів використовується для отримання f_0-представлення найбільш виражених тональних компонент, на основі обробки отриманих на першому етапі відфільтрованих значень.

Метою статті є отримання траєкторії f_0-значень аудіосигналу і фільтрація відповідних цій траєкторії гармонік.

Результати. Запропонований метод є перспективним для застосування в системах обробки аудіосигналів для отримання основних тонів найбільш виражених тональних компонент та їхнього подальшого перевикористання.

Висновки. Використане у підході розкладання спектра сигналу, засноване на тенденції частотно-часового розподілення різних за характеристиками звуків, дозволяє ефективно відбирати мелодійні контури від немелодійних контурів. Незважаючи на це, існує потреба в подальших дослідженнях стосовно моделі розподілу гармонійно-перкусійних характеристик відносно одне одного. Така модель має бути розширена евристиками для більш точної фільтрації стійких у часі тональних компонент комплексних аудіосигналів.

Ключові слова: *основний тон, траєкторія частот основних тонів, мелодія, гармонійно-перкусійний поділ джерела, аудіосигнал, медіанна фільтрація, оцінка значущості, музичний тон, нота, гармоніка, мелодійна лінія, мелодійний контур, віконне перетворення Фур'є, логарифмічна шкала частот, гармонійна сумація.*