O. M. Tkachenko, Cand. Sc. (Eng.), Associate Professor;
I. R. Arseniuk, Cand. Sc. (Eng.), Associate Professor;
S. V. Khrushchak, Cand. Sc. (Eng.)

METHOD OF RAPID SEARCH OF THE AUDIO RECORDING FRAGMENT

Taking into account large volumes of audio information, stored in musical compositions, rate and reliability of their search is of great importance. The paper describes the method of rapid search of the audio recording fragment with the improved assessment of the proximity degree between the unknown audio fragment and template, that enables to improve the authenticity of the decision making during the search.

For compact description of the signal parameters mel-frequency cepstral coefficients are chosen, on their base the body of music compositions parameters is formed as the set of centroids, obtained as a result of clusterization. The notion of the reduced proper distance as the assessment of the degree of the proximity of the unknown fragment of music composition and previously created templates of the audio recordings is introduced. Application of kd-trees for searching acceleration of the unknown fragment in the body of audio recordings is substantiated, basic stages of the search are presented. Different variants of the proximity degree calculation of the unknown audio fragment with audio recording are considered, namely: assessment of the degree of the proximity by the reduced proper distance, assessment of the degree of proximity by the number of hitting into the list k of the closest centroids, assessment of the degree of proximity by the weighted quantity of getting into the list k of the closest centroids. It is shown, that the execution of non accurate but approximate vectors search on the base of kd-tree enables to achieve considerable time saving, but it leads to the reduction of the validity of the searching results. That is why, to reduce the complexity of the calculations, saving the validity of the results it is suggested to perform combined search for large archives of audio recordings, this type of search combines rapid "inaccurate" search with the application of kd-tree of several nearest audio recordings of the body for the set audio fragment at the first stage of the search, among these audio recordings at the second stage by means of the complete oversampling one the nearest is determined. The suggested method enabled to increase the completeness and relevance of the searching results.

Key words: *audio recording, body of music compositions, degree of proximity, distance calculation, clusterization, kd-tree.*

Introduction

Growth of the volume of multimedia information, in particular, audio information, transferred and processed by the computer systems, stipulated the necessity of the automation of the processes of data analysis and search. That is why, in modern systems of audio information processing there appears the need of automatic searching of music compositions (MC) on the base of video content in the data bases. Taking into account large volumes of audio information, stored in the body of MC, rate and reliability of searching is very important. To improve the reliability of the searching results the method of rapid search of the audio recording fragment in the body of MC with the improved assessment of the proximity degree between the unknown audio fragment and the templates, this enables to enhance the validity of the decision making in the process of searching as a result of increasing the discrepancy degree of own MC with other audio recordings.

Objective and tasks of the paper

The objective of the given paper is the enhancement of the reliability of the decision making in the process of the own MC template searching (further in the paper – own template) for the unknown audio fragment due to the application of the improved assessment of the proximity degree between the fragment and templates, that enables to increase considerably the discrepancy level of own template with others.

For increasing the reliability of the decision making in the process of own template searching for the unknown fragment the following problems are to be solved:

1. Select the parameters which compactly describe it.
2. Suggest the assessment of the proximity degree of the unknown MC fragment and previously created templates of audio recordings to provide high level of the discrepancy of own template and others.
3. Perform experimental studies.

The results of the search will be the template of MC in the body with minimal difference with the output audio recording. Further it is stated in the research that the audio recording to be determined coincides with one of the templates of the body.

Selection of the mathematical model of audio signal

The scheme of determination the correspondence of MC on the base of the content is based on the usage of audio file for the construction of the audio signal model. It is not efficient to compare directly audio signals in the temporal area, that is why, audio recording is divided into frames (fragments of 10 – 30 ms), during which the characteristics of the signal remains relatively stable. For each frame spectral analysis is carried out, on the base of which the value of the parameters vector (parametrization) is calculated. Among different parameters Mel Frequency Cepstral Coefficients are selected, they became widely used in the systems of speech and speakers recognition [1], then they were used in the process of Music information retrieval, MIR [2]. As a result of parametrization the description of MC as a file, containing vectors MFCC is obtained.

Thus, if the frequency of the audio file discretization is – 44.1 kHz, the length of the frame is – 20 ms with the overlapping of 0.5 frame, then average MC, duration of 3 min, is characterized by approximately 8 mil counts or, taking into account frames overlapping, approximately 18 thousand frames, each of them is described by the vector of MFCC parameters of 13 dimensionality (Fig. 1).

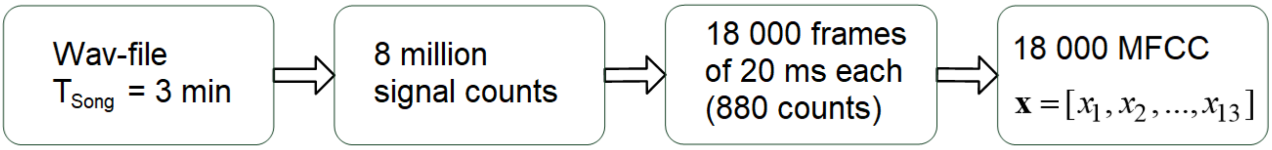

Fig. 1. Description of wav-file (3 min.)

As it is seen, parametrization allows to decrease the amount of information, needed for the description of MC dozens of times:

$$\frac{8000000\ counts}{18000\,MFCC\cdot 13}=\frac{8000000}{234000}\approx 34\,times$$

Hence, MFCC coefficients is compact presentation of spectral enveloping curve, in the process of MC recognition it allows to replace millions of counts of audio file.

Comparison of the unknown MC with the templates by the reduced distance

After the selection of the parameters for MC description it is necessary to pass to the comparison of the unknown MC with the audio recordings of the body and determination of the template the difference with which will be minimal. To identify the unknown audio recording, it is necessary to have a comparison criterio. As a rule, this criterion is distance. Comparison according to the selected parameters must provide high level of the divergence of the own template and templates of other compositions. That is, as a result of comparison of the unknown MC with own template the error (by the distance) between them must be minimal and as long as possible for the templates of other musical compositions:

$$\begin{cases} D(\boldsymbol{X},\tilde{\boldsymbol{X}})\to min \\ D(\boldsymbol{X},\tilde{\boldsymbol{Y}})\to max \end{cases} \Rightarrow D(\boldsymbol{X},\tilde{\boldsymbol{X}}) << D(\boldsymbol{X},\tilde{\boldsymbol{Y}}), \tag{1}$$

where $\tilde{\boldsymbol{X}}$ – is the set of parameter vectors of own template; $\tilde{\boldsymbol{Y}}$ – is the set of parameters vectors of the template of other MC.

General scheme of MC searching is presented in Fig. 2.

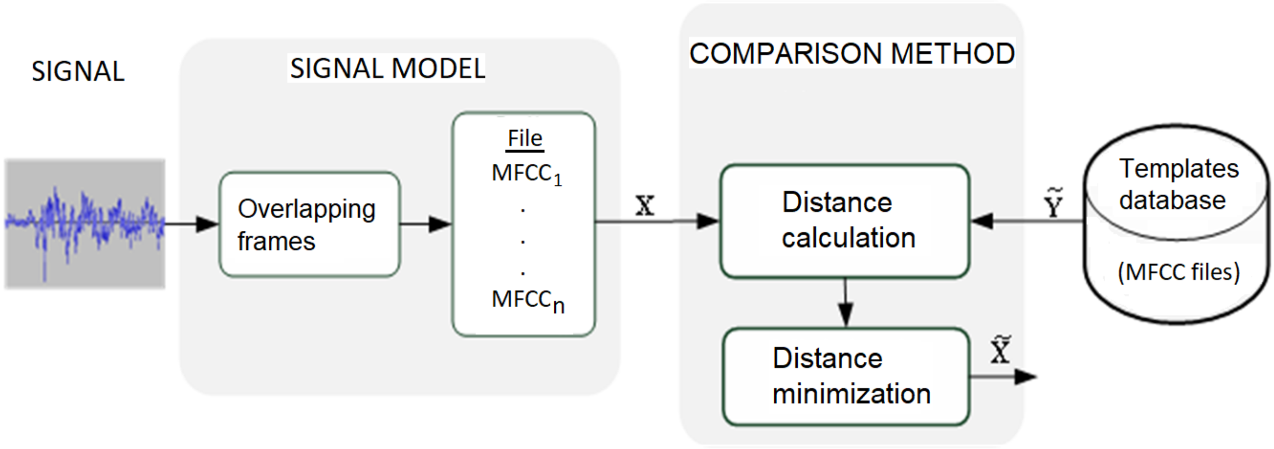

Fig. 2. General scheme of MC searching

The simplest and most obvious approach for the determination of the proximity between the sets of MFCC parameters of the unknown MC and audio recordings in the body is MFCC comparison on the base of the most widely spread Euclidian matrix, be more precise, square of the Euclidian distance. Formula of the non-weighted Euclidian distance between the vectors of parameters $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ MC, to be identified and the vector of the template $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_d)$:

$$D_{Eu}{}^2(\boldsymbol{x}, \tilde{\boldsymbol{y}}) = \sum_{i=1}^{d} (x_i - \tilde{y}_i)^2. \tag{2}$$

Correspondingly, the distance (error) between the files of the parameters of the unknown composition and template can be found as the sum of the distances:

$$D(\boldsymbol{X}, \tilde{\boldsymbol{Y}}) = \sum D_{Eu}{}^2(\boldsymbol{x}_j, \tilde{\boldsymbol{y}}_j). \tag{3}$$

In the ideal case, at such approach the distance between the files of MFCC parameters of one and the same MC will equal zero, for different MC – differ from zero:

$$\begin{cases} D(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = 0, \\ D(\boldsymbol{X}, \tilde{\boldsymbol{Y}}) > 0. \end{cases} \tag{4}$$

However, it should be noted, that even for one and the same MC audio recordings may differ, for instance, at the beginning of the recording there may be silence, melody of another MC, etc., recordings have different rhythm, duration. This means, that in case of frames displacement in time, the distance to own template $D(\boldsymbol{X}, \tilde{\boldsymbol{X}}) \neq 0$, i. e., the condition of differentiation of own MC (4) is not met, thus, direct comparison of parameters files by the Euclidian distance does not fit for searching audio fragment in the body of audio recordings.

It is obvious, that in greater part of cases MC is characterized by certain periodicity, it is the presence of the identical or very similar by the text and character of the melody fragments. Accordingly, we may speak about the redundancy of the data, the recording is described and reduce the number of parameters for its description. Taking this into account, it is expedient to apply the methods of cluster analysis for the formation of the body of audio recordings. Usage of clusterization for the formation of body models enables to reduce the volume of memory, needed for their storage.

Main steps of comparison of the files of unknown MC parameters with certain template in the body:

1. Searching of minimal Euclidian distance $Dmin^2$, between the current vector of $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ parameters from the set of parameters $\boldsymbol{X} = \{\boldsymbol{x}\}, |\boldsymbol{X}| = n$ MC, which must be identified and the set of vectors-centroids, $\tilde{\boldsymbol{Y}} = \{\tilde{\boldsymbol{y}}\}, |\tilde{\boldsymbol{Y}}| = m$, $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_d)$ of the template:

$$Dmin^2 = \underbrace{min}_{j}(D_{Eu}{}^2(\boldsymbol{x}, \tilde{\boldsymbol{y}}_j)) = \underbrace{min}_{j}(\sum_{i=1}^{d}(x_i - \tilde{y}_{ji})^2), j = \overline{1,m}. \quad (5)$$

2. Calculation of the evaluation of the distance on the whole to the template in the form of the sum of squares of minimal distances $Dmin^2$:

$$D(\boldsymbol{X}, \tilde{\boldsymbol{Y}}) = \sum_{l=1}^{n} Dmin^2 = \sum_{l=1}^{n} \underbrace{min}_{j}(\sum_{i=1}^{d}(x_i - \tilde{y}_{ji})^2), j = \overline{1,m}. \quad (6)$$

At the same time, as MC have different duration, each composition is characterized by own number of frames, presented by the vectors MFCC. During clusterization the same number of centroids of all the templates is calculated. As a result, these different compositions are in unequal conditions, i.e., for the compositions duration of which is greater, initial error (between the files of the same composition, before and after clusterization) also will be greater, as in this case, on each cluster more vector parameters will fall. We can avoid this situation, if we divide the distance to the template, determined in the formula (6), by the number of frames of MC. This value will be called reduced distance (RD) – D_{RD} of the audio recording:

$$D_{\Pi B} = \frac{D(\boldsymbol{X}, \tilde{\boldsymbol{Y}})}{n} = \frac{\sum_{l=1}^{n} Dmin^2}{n} = \frac{\sum_{l=1}^{n} min(\sum_{i=1}^{d}(x_i - \tilde{y}_{ji})^2)}{n}, j = \overline{1,m}. \quad (7)$$

As it is seen from the formula (7) characteristic D_{RD} does not depend on the number of frames (duration of the recording). It can be used as the criterion for the decision making both for recording on the whole and for separate audio fragment. The substantiation of this is given in [3], the duration of the MC fragment of 5 sec (500 frames or 500 vectors of MFCC parameters), sufficient for searching audio recording is also determined.

Creation of the body of MC templates on the base of clusterization

Methods of clusterization are widely used in the systems of audio signals recognition. The task of clusterization may be formulated in the following way: the set of ***n*** vectors of ***d*** dimensionality must be divided into the subsets according to the criterion of distortion minimization $e_i^2 \to min$. There exist different ways of the distortion assessment but in the majority of the applied realizations the sum of average square Euclidian distances between the center of the cluster (centroid) $\boldsymbol{c}_i$ and parameter vectors, which belong to it issued $\boldsymbol{X}_i = \{\boldsymbol{x}\}, \boldsymbol{X}_i \subset \boldsymbol{X}$ [4], that is:

$$e_i^2 = \{\boldsymbol{c}_i : \sum_{j=1}^{N_i} D_{Eu}{}^2(\boldsymbol{x}_j, \boldsymbol{c}_i) \mid \boldsymbol{x} \in \boldsymbol{X}_i \le \sum_{j=1}^{N_i} D_{Eu}{}^2(\boldsymbol{x}_j, \boldsymbol{c}) \mid \boldsymbol{x} \in \boldsymbol{X}_i\};$$

$$\boldsymbol{X}_i \subset \boldsymbol{X}, \forall \boldsymbol{c} \in \boldsymbol{X} \setminus \boldsymbol{X}_i, e_i^2 \to min,$$

where $\boldsymbol{N_i}$ – is the number of vectors, belonging to centroid $\mathbf{c}_i$.

Thus, MC templates can be described by the centroids of the clusters of MFCC parameters. As between each centroid $\boldsymbol{c}_i$ and vectors, belonging to it, there is an error e_i^2, then the distance between the files of parameters, describing the same MC prior the clusterization and after (even if audio recordings were identical), will be positive and equal to the value of the total error of clusterization $\boldsymbol{E^2}$.

Hence, the formed body of MC templates for the identification of the unknown fragment of the audio recording is actually the file, containing MFCC centroids of MC templates.

Application of kd-trees for searching MC fragment in the body of the templates

The simplest solution of the known problem of the nearest neighbor search and in our case – the nearest centroid in the body of audio recordings – it is calculation of the distance between the vector of the unknown audio fragment, which must id be identified and all the vectors-centroids of the templates. Time of the execution of the complete search is proportional ***O(dN)***, where ***N*** – is a number of the centroids in the body, ***d*** – is dimensionality of the vectors. However, it is practically impossible to use such an approach for searching in large data bases. To simplify the computations in the process of searching the fragment of audio recording simple method of space division – k-dimensional tree (kd-tree) was used. Centroids of audio recordings in the body were arranged, taking into account kd-tree [5]. Arrangement is performed very rapidly, as the division of the space takes place only along the data axes, that is why, there is no need to calculate d-dimensional distances.

To provide the obtaining of the nearest vector, the searching, besides the direct phase (descending along the tree), must have and reverse phase (searching of the elements of the tree in the set range). During the direct search all the distances to nodes d_i are fixed. Direct phase is completed by the calculation of the distance $D_{min} = D_k$ to the corresponding terminal peak, where the unknown vector is and also potentially the nearest element of the tree (centroid of the template in the body), that sets the searching radius to the reverse phase. During the reverse phase of the search distances d_i are calculated only to those nodes of the tree, which can provide the execution of $d_i < D_{min}$. If for corresponding page the condition $D_k < D_{min}$, is met the radius of search is corrected $D_{min} = D_k$ [6].

The considered searching procedure guarantees the finding of the nearest vector by the formula (1), but it requires greater amount of distance measurements than $log_2 n$.

It should be noted that the search on the base of kd-tree is very efficient only for small dimensions ($d < 20$). For $d > 20$ average searching time of the nearest vector in the body, arranged on the base of kd-tree on average grows proportionally $O(h) = O(d \cdot log(N+1))$, where h – is the height of the tree [6]. For larger dimensions computational complexity may reach $O(d \cdot N)$.

Thus, the sequence of steps in the process of searching of the corresponding for audio fragment audio recording in the body on the base of kd-tree is the following.

1. Rapid search in the base of parameters of the audio recording body, arranged on the base of kd-tree is performed, in the process of the search by Euclidian matrix, according to the formula (2), the set $\boldsymbol{C}$ of vectors-centroids of audio recordings (candidates for the nearest vector), arranged by the distance growth, is selected:

$$\boldsymbol{C} \subset \boldsymbol{T}, \boldsymbol{C} = \{\tilde{\boldsymbol{Y}}_1, \tilde{\boldsymbol{Y}}_2, \ldots, \tilde{\boldsymbol{Y}}_k\}, |\tilde{\boldsymbol{Y}}| = m, |\boldsymbol{C}| = k, |\boldsymbol{T}| = N = m \cdot Q, k \le N,$$

where k – is the number of vectors of audio recordings in the list of the nearest to the frame of the unknown audio fragment; m – is the number of centroids (vectors) of audio recording; Q – is total number of audio recordings in the body; N – is total number of centroids-vectors of the body of audio recording.

For each of k nearest centroids of the set $\boldsymbol{C}$ the following data are stored: distance to the frame of the unknown audio fragment $D_p{}^2(\boldsymbol{x}, \tilde{\boldsymbol{y}}_{qj})$ and index of the corresponding audio recording i_{pq} in the body:

$$\boldsymbol{I} = \{ i_{pq} \}, \boldsymbol{D} = \{ D_p{}^2(\boldsymbol{x}_l, \tilde{\boldsymbol{y}}_{qj}) \},$$

$$D_p{}^2(\boldsymbol{x}_l, \tilde{\boldsymbol{y}}_{qj}) = \sum_{i=1}^{d} (x_i - \tilde{y}_i)^2, p = \overline{1,k}, l = \overline{1,n}, j = \overline{1,m}, q = \overline{1,Q}.$$

2. Determination for the audio fragment its own audio recording in the body. Distance to each of audio recordings in the body is determined according to the rule:

$$Distance_q = \begin{cases} min(D_p{}^2), \forall q \in \boldsymbol{I}; \\ D_p{}^2, p = k, \forall q \notin \boldsymbol{I}. \end{cases} \quad (8)$$

In the process of realization of points 1 and 2 the sum of distances in all $\boldsymbol{n}$ frames of the unknown audio fragment for each audio recording $\tilde{\boldsymbol{Y}}_q$ in the body is calculated

$$D(\boldsymbol{X}, \tilde{\boldsymbol{Y}}_q) = \sum_{l=1}^{n} Distance_q. \quad (9)$$

At the last step minimum of the reduced distance, determined in the formula (7) is found:

$$D(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = \underbrace{min}_{q} D_{ПВ\,q} = \underbrace{min}_{q} \frac{D(\boldsymbol{X}, \tilde{\boldsymbol{Y}}_q)}{n}. \quad (10)$$

Accordingly, own audio recording $\tilde{\boldsymbol{X}}$ in the body as the nearest to the unknown audio fragment is considered to be determined.

The presented searching procedure provides the obtaining not only one nearest vector but a certain set of vectors arranged by the distance growth according to the formula (7). This is possible due to the fact that the data, regarding the distance to the covered terminal peaks is stored and simultaneously their arrangement by distance growth to the unknown vector is realized. That is why, additional finding of several nearest vectors does not require much time.

Method of rapid search of audio recording fragment with different variants of proximity measure assessment

In order to improve the reliability of the results of the search as a result of increasing the measure of difference between own audio recording and others possible variants of the proximity measure calculation of the unknown audio fragment with Audi recordings in the body, i. e. variants of the realization of the second step of searching on the base of kd-tree are suggested.

Assessment of the proximity measure by the reduced distance

1. For the current frame $\boldsymbol{x}_l$ of the unknown audio fragment, the distance to each audio recording $\tilde{\boldsymbol{Y}}_q$ in the body is determined in accordance with the rule, expressed by the formula (8).

That is, on this frame the smallest distance among its centroids of the list k nearest will be considered to be the distance to audio recording. The distance for audio recordings, centroids of which did not enter the list k of the nearest, is considered to be the distance that equals the distance to the last k^{th} centroid from the list of the nearest centroids. The latter means that we artificially decrease the distance to those audio recordings, which did not enter the list k of the nearest and correspondingly, decrease the level of differentiation of own audio recording and the rest of audio recordings in the body.

2. For each audio recording $\tilde{\boldsymbol{Y}}_q$ in the body the sum of distances on all n frames of the unknown audio fragment is calculated according to (9).

3. Own audio recording in the body is determined from the minimum of the reduced distance according to the formula (10).

In order to increase the difference of distances between own audio recording and the rest of audio recordings of the body other measures of audio recording proximity assessment are suggested.

Assessment of the measure of proximity by the number of hits in the list k of the nearest centroids (Hits)

1. For each frame $\boldsymbol{x}_l$ of the unknown audio fragment the number of hits of the centroids of q^{th} audio recording into k list of the nearest is fixed:

$$\boldsymbol{Hits}_q = \sum_{p=1}^{k} h_p \text{, де } h_p = \begin{cases} 1, \forall q \in \mathbf{I}; \\ 0, \forall q \notin \mathbf{I}. \end{cases} \quad (11)$$

2. For each audio recording $\tilde{\boldsymbol{Y}}_q$ total value of the number of hits in all n frames of the unknown audio fragment is found.

3. Own audio recording is considered to be the recording, number of hits in all n frames for which is the largest:

$$H(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = \underbrace{max}_{q} \sum_{l=1}^{n} \boldsymbol{Hits}_q. \quad (12)$$

At such approach the difference of the distances between the own audio recording and others grows, the complexity of calculations decreases.

Assessment of the proximity degree by the weighted number of hits in the list k of the nearest centroids (Weighted Hits)

This approach to the assessment of the degree of proximity between the unknown audio fragment and the base of parameters of audio recordings of the body is a combination of two previous ones, i. e., distances $D_p{}^2$, number of hits h_p of the audio- recording centroids in the list k of the nearest, and the position of the audio recording in the list of the nearest, as it is shown in the formula (13).

1. For each frame $\boldsymbol{x}_l$ of the unknown audio fragment the value, which was called weighted by the number of centroids hits of q^{th} Audi recording in the list k of the nearest by the formula:

$$\boldsymbol{Weighted\ Hits}_q = \sum_{p=1}^{k} (D_k{}^2 - D_p{}^2) \cdot h_p. \quad (13)$$

2. In the same way for each audio recording $\tilde{\boldsymbol{Y}}_q$ total value of the weighted number of hits in all n frames of the unknown audio fragment is found.

3. Own audio recording is considered to be audio recording, for which the value of weighted number of hits in all n frames is the largest, divided by the total number of frames:

$$WH(\boldsymbol{X}, \tilde{\boldsymbol{X}}) = \underbrace{max}_{q} \sum_{l=1}^{n} \frac{\boldsymbol{Weighted\ Hits}_q}{n}. \quad (14)$$

This improved approach to the assessment of the proximity degree between the unknown audio fragment and audio recordings of the body must balance the advantages and disadvantages of previous two approaches.

Thus, searching method of own audio recording with the improved assessment of proximity degree is suggested. The sequence of actions for its realization is proposed.

For each frame (vector) $\boldsymbol{x}_l$ of the unknown audio fragment the following steps are executed:

1. Rapid search of k nearest neighbors in the clustered body of audio recordings, arranged on the base of kd-tree.

2. Determination for the unknown audio fragment its own audio recording in the body by the weighted number of hits in the list k of the nearest centroids.

Experimental results

To carry out experimental studies the body of 1000 audio recordings of MC was formed. All MC had wav (mono) format with the discretization frequency 44.1 kHz. Initially the «silence» was removed from audio-files at the beginning and at the end of recordings. In the process of MC body formation audio recordings of the templates were divided into frames of 20 ms with overlapping of 10 ms, for each frame vector of MFCC parameters of 13 dimensionality was calculated. Successions of the vectors of MFCC parameters, describing MC, were clusterized, using the improved clusterization method [7]. After clusterization each template in the body was represented by 1000 centroids of MFCC (totally approximately 20 vectors per cluster). Fragments, duration of 5 sec for identification were selected randomly from the models of the formed MC body. Application of the search by kd-tree and assessment of the proximity degree to the templates, compared with the complete search leads to certain decrease of the distinguishing degree of the own audio recording and other templates of the body (shown in Fig. 3), however the difference remains sufficient for correct decision making.

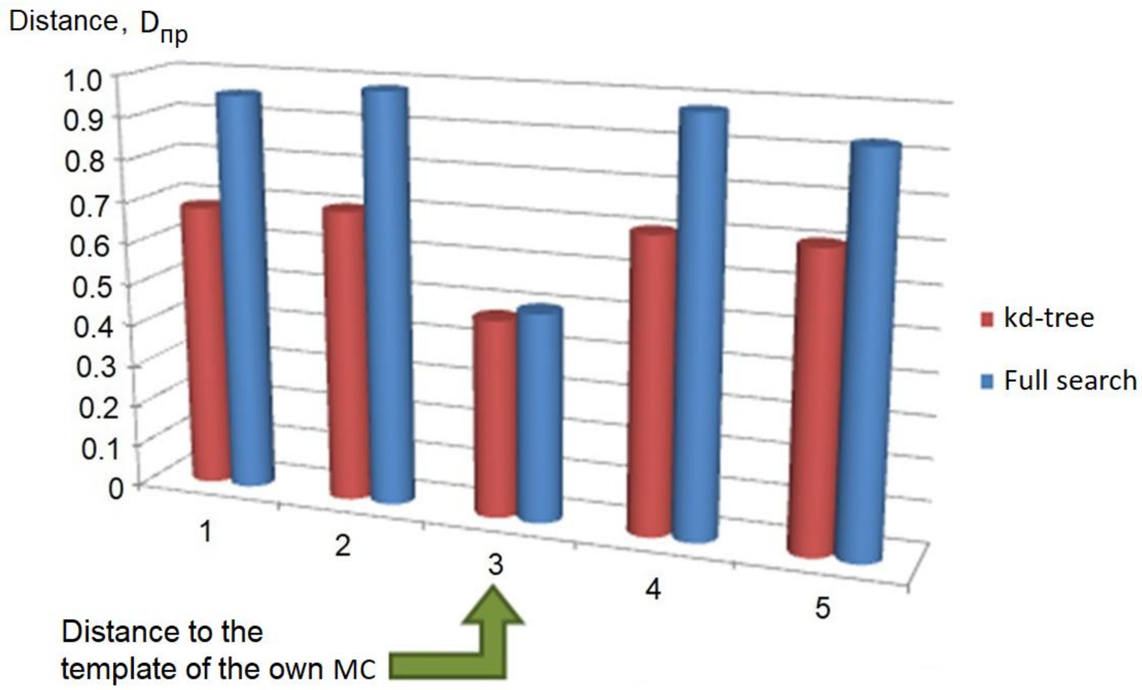

Fig. 3. Comparison of the results of complete search and kd-tree based search

Fig. 4 presents the examples of the unknown fragment comparison with templates (own template №3) for the assessment the degree of their proximity and consolidated histogram, where all the variants of proximity degree assessment and unknown fragment in relative units to achieve single scale are presented.

As it is seen from the given results, assessment of the proximity degree according to weighted number of hits of the template to the list of k nearest (Weighted Hits) provides high level of differentiation of own and strange templates.

The level of differentiation of the own template and others is influenced by the number of centroids, representing audio recordings in the body, duration of the fragment for search and number of the nearest centroids, obtained as a result of the searching by kd-tree. The last variant is the most suitable for increasing the level of differentiation as it does not lead to considerable growth of the searching time. The results for 20 and 50 nearest candidates were considered. In Table 1 for five fragments the results, obtained during the selection of own audio recording in the body by the assessments of the proximity degree of the unknown fragment of the templates are presented. For each assessment of the proximity degree values are given, obtained for own template (color indication), and the nearest template among others.

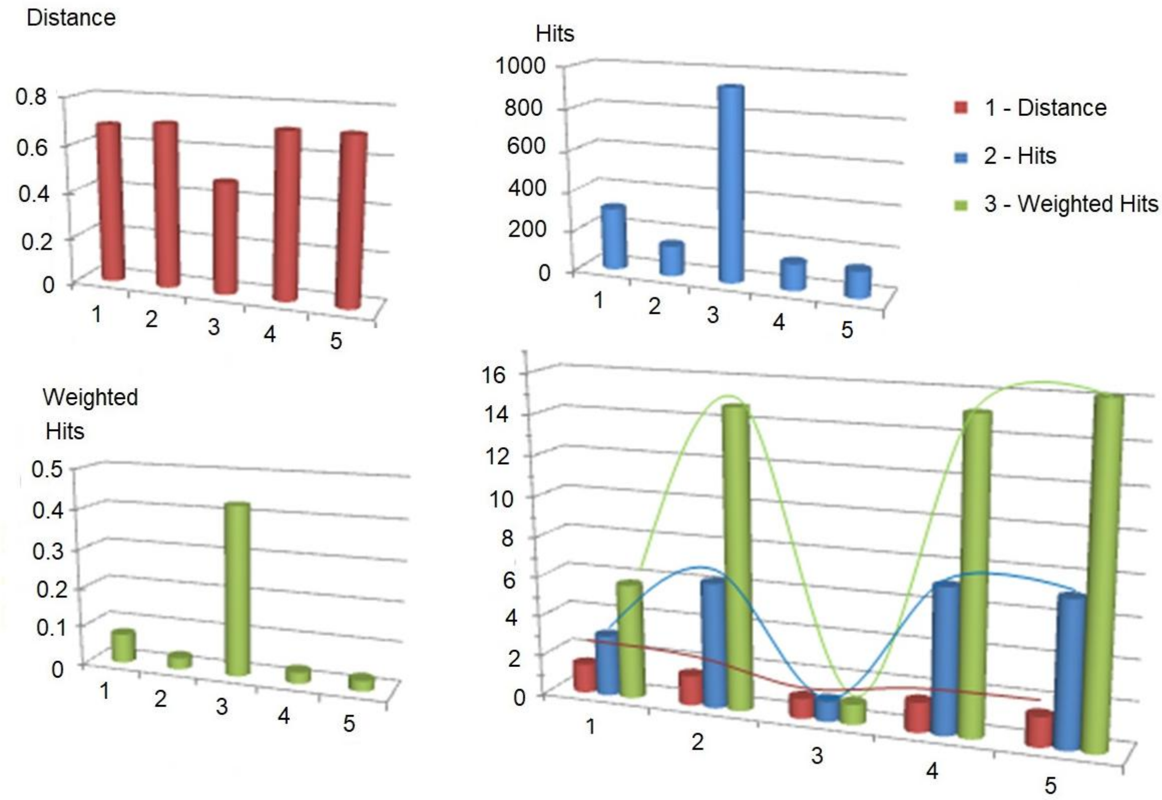

Fig. 4. Comparison of the assessments of the proximity degree of the unknown fragment with templates

Table 1

Results of the suggested assessment of the proximity degree for own template and the nearest among others for k = 20 and k = 50

k	20					50				
№ frag.	1	2	3	4	5	1	2	3	4	5
$D_{ПВ}$	0.412	0.454	0.304	0.364	0.454	0.425	0.490	0.324	0.386	0.475
	0.753	0.564	0.434	0.458	0.615	0.847	0.639	0.498	0.521	0.683
H	821	275	391	335	580	1238	412	535	479	931
	160	57	65	74	140	390	134	125	177	307
WH	0.567	0.152	0.174	0.118	0.260	0.900	0.251	0.261	0.191	0.426
	0.075	0.017	0.026	0.016	0.027	0.167	0.048	0.056	0.041	0.074

The obtained data show that with the increase of the number of the nearest neighbors to 50, the difference in assessments grows and the differentiation level between own templates and others also increases.

Assessment of the distance to own audio recording, obtained by the formula (14), at k = 20 on average 8 times exceeds the analog assessment of the distance to other templates. Thus, improved assessment of the proximity considerably enhances the validity of the decision-making in the process of searching own template of MC and provides the possibility for further enlargement of MC body.

At last, it is worth mentioning that for the search in real archives of audio recordings, containing hundreds of thousands – millions of audio recordings, critical requirement is high searching speed. Realization of not accurate but approximate search of vectors (when the found vector is not obligatory the nearest but rather close) on the base of kd-tree enables to achieve time economy, however it leads to the decrease of the validity of the searching results. That is why, for the reduction of the computations complexity and maintaining the validity of the results, it is expedient for large archives of audio recordings to perform the combined search, uniting rapid "non-accurate" search with the application of kd-tree of several nearest audio recordings of the body for the set audio fragment at the first stage of the search, at the second stage by means of the exhaustive search one the nearest is determined.

Method of searching based on kd-tree with the improved assessment of the proximity degree with weighted number of hits in the list of the nearest centroids was further developed. The suggested method enabled to enhance the completeness and relevance of the searching results, namely: for audio fragments of 5 sec. the completeness of the search is 99.8 %, relevance – 97.4 %, for audio fragment of 1 sec – 99.6 % and 94.3 % correspondingly, this exceeds the analog values of similar methods [8].

Conclusions

Method of rapid search of the audio fragments of 5 sec of duration on the base of kd-tree with the improved assessment of the proximity degree between the unknown fragment and audio recordings in the body is suggested in the paper, this enables to improve the validity of the searching results. Analytical relations for the assessment of the proximity degree of the unknown fragment with the templates ,based on calculation of he the distance from the fragment to the set number of the nearest centroids and number of hits to the list of the nearest neighbors are suggested . As a result of using the improved assessment of the proximity degree (by weighted number of hits to the list k of the nearest centroids) between the unknown audio fragment and templates the level of distinction of own audio recording in the body is considerably increased. Experimental results show that the assessment of the distance to own audio recording at k = 20 on average 8 times exceeds the same distance assessment to other audio recordings of the body.

REFERENCES

1. Ganchev T. Comparative evaluation of various mfcc implementations on the speaker verification task / T. Ganchev, N. Fakotakis, G. Kokkinakis // In Proceedings of 9th International Conference on Speech and Computer, SPECOM’05. – 2005. – P. 191 – 194.

2. Li T. Visual Classification of Music Style Transfer Based on PSO-BP Rating Prediction Model / T. Li // Complexity. – 2021. – Vol. 2021. – P. 1 – 9. – URL: https://doi.org/10.1155/2021/9959082.

3. Tkachenko O.M. Approach to the assessment of the fragment duration for the search of musical composition by the set template / O. M. Tkachenko, O. F. Griyo Tukalo // Information technologies and computer engineering . – 2014. – № 1. – P. 31 – 40. (Ukr).

4. Introduction to K-D Trees [Electronic resource] / H. Hristov // Baeldung. – Access mode : https://www.baeldung.com/cs/k-d-trees (date of access: 10.02.2024).

5. Moore A. W. Efficient memory based learning for robot control / A. W. Moore. – PhD thesis Nr. – University of Cambridge, Computer Laboratory UCAM-CL-TR-209. – 1990. – 248 p.

6. Jo J. A progressive k-d tree for approximate k-nearest neighbors / J. Jo, J. Seo, J. Fekete // 2017 IEEE Workshop on Data Systems for Interactive Analysis (DSIA), Phoenix, AZ, USA. – 2017. – P. 1 – 5. DOI: 10.1109/DSIA.2017.8339084.

7. Method of clusterization , based on successive activations of k-average and calculation of the distances to active centroids / O. M. Tkachenko, O. F. Griyo Tukalo, N. O. Bilichenko [et al.] // Registration, storage and processing of data. – 2012. – № 1, V. 14, B. 1. – P. 25 – 34. (Ukr).

8. Tkachenko O. M. Method of increasing the rate of audio recording fragment search ,using kd-trees / O.M. Tkachenko O.F. Griyo Tukalo // Information technologies and computer engineering. – 2014. – № 3. – P. 57 – 66. (Ukr).

Editorial office received the paper 25.02.2024.

The paper was reviewed 14.03.2024.

Tkachenko Olexandr – Cand. Sc. (Eng.), Associate Professor with the Department of Computer Science**.**

Arseniuk Igor – Cand. Sc. (Eng.), Associate Professor with the Department of Computer Science, e-mail: igrosars@gmail.com**.**
Vinnytsia National Technical University.

Khrushchak Sergiy – Cand. Sc. (Eng.), Senior Lecturer with the Department of Computer Science.
Vinnytsia National Agrarian University.