COMPUTER SEARCH FOR INFORMATIVE FEATURES IN THE PROBLEM OF DIFFERENTIATION OF MELODIES BY THEIR "NATIONALITY"

Irina V. Bakhmutova, Vladimir D. Gusev, Tatiana N. Titkova

Institute of Mathematics, Siberian Branch of Russian Academy of Sciences,
Russia, Novosibirsk

Abstract

The problem of differentiation of song melodies by their «national» belonging with the use of interval-metric characteristics was considered. An easily interpreted system of quantitative characteristics, which enables the formulation in terms of repetitions of national-specific features intrinsic to samples of the Russian, French, and American melodies is proposed. The most informative (in terms of classification) fragments of melodies were found.

Introduction

The essential structural components of any kind of a text are repetitions. The role of repetitions (melodic, rhythmic, metric) in musical compositions is of special significance. The repetition of separate fragments (intonations) in a melody facilitates its better learning and the variation of repetitions enriches the melody. Repetitions are very important for the classification of melodies by the genre, style, composer, etc. The representation of musical texts in terms of repetitions proposed by authors seems to be very convenient for the purpose (Bakhmutova, Gusev, Titkova, 1990).

The aim of the paper is to illustrate the classification capability of this method of representation in the examples revealing the characteristics allowing to differentiate melodies by their «national» belonging. This problem has already drawn attention of musicologists and was studied at the level of rhythmical repetitions (Boroda, 1990). The base of our work is interval-metric characteristics of melodies.

1. The system of Melody Representation

The musical texts by their nature are multidimensional as every sound is characterized by the pitch, duration, and metric accent. Following R.Kh.Zaripov (Zaripov, 1983) we represent the musical texts in the form of interval-metric characteristics. The text

T_i consisting of N_i notes is replaced by the sequence of $(N_i - 1)$ $IS-$ codes. The $IS-$code in $k-$ th place $\{1 \le k \le N-1\}$ characterizes the transition from the $k-$th to the $(k+1)^{th}$ tone of melody and is represented by triplet: $|I_k|$ is the absolute value of interval (the number of degrees between k^{th} and $(k+1)^{th}$ tone in melody); the sign of I_k ("+" corresponds to an ascending motion of melody pitch line, "–" to descending one, if $|I_k|$, then "+" sign is formally put); S_k is a metric accent of sound ("+" corresponds to the transition from the metrically stronger tone to the metrically weaker one and "–" on the contrary). For example, the code $3+-$ is interpreted as a jump of the fourth up with simultaneous increase in the metric accent.

2. Representation of Texts in Terms of Repetitions[1]

Let $T = a_1 a_2 a_3 \ldots a_{N-1}$ be the $IS-$representation of the melody where a_k $(1 \le k \le N-1)$ is the k^{th} position is the triplet $\{|I_k|, sign\ of\ I_k, sign\ of\ S_k\}$. The $l-$long fragment of text will be called $l-$gram. In the text of length $N-1$ there are exactly $(N-l)$ $l-$grams separated by the sliding frame of the width l. The number of different $l-$grams will be denoted by M_l (it is obvious that $M_l \le N-l+1$). Let us call the set $\Phi_l(T) = \{\varphi_{1_1} \varphi_{l_2} \ldots \varphi_{l_{M_l}}\}$, where φ_{l_i} $(1 \le i \le M_l)$ is a pair $<i^{th}$ $l-$gram [2], F - frequency of its occurrence in the text> the *frequency characteristics of the* l^{th} *order of the text* T. The elements φ_{l_i} are convenient to be ordered by decreasing in F. *Full frequency spectrum of text* T is defined as a set of frequency characteristics $\Phi(T) = \{\Phi_1(T), \Phi_2(T), \ldots, \Phi_{l_{\max}}(T)\}$ where $l_{\max}$ is the length of maximum repetition in text T.

Thus, the frequency characteristic of the l^{th} order is just a set of all possible repetitions of the length *l* in the text, which is added by uniquely occurred $l-$grams and the full frequency spectrum contains information on all the repetitions in the text.

When we deal with a set of texts, the concatenation is formed as $\overline{T} = T_1 * T_2 * \ldots * T_m$, where * is the separation sign between different texts, and $\Phi(\overline{T})$ is calculated. *l*–

---

[1] The representation suggested is suitable for texts of any kind. In this paper, the investigation is concerned with the note texts of melodies.

[2] i^{th} $l-$gram is the *l*-gram in i^{th} position of the text.

grams can be ordered by the number of texts, where these *l*–grams are presented. This is essential for the choice of the most informative features characterizing the given class of objects (texts).

In the case of a multi-class problem of recognition, each class is represented by its individual learning set of texts. For simplicity, the number of classes is taken to be *2*, $\overline{T}_1$ and $\overline{T}_2$ are learning samples for classes *1* and *2*, respectively, and $\Phi_l(\overline{T}_1)$, $\Phi_l(\overline{T}_2)$ are the frequency characteristics of the *l*–th order for each class. A set of *l*–grams common for $\overline{T}_1$ and $\overline{T}_2$ be denoted as $\Phi_l(\overline{T}_1, \overline{T}_2)$. *l*–grams presented in one sample only are the most interesting for the purpose of classification. These *l*–grams can be interpreted as set-theoretic complements $D_l(\overline{T}_1)$ and $D_l(\overline{T}_2)$, respectively, to the intersection $\Phi_l(\overline{T}_1, \overline{T}_2)$ of two sets: $\Phi_l(\overline{T}_1)$ and $\Phi_l(\overline{T}_2)$ (see Fig.1).

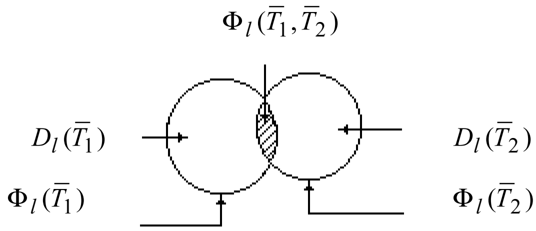

Fig. 1

Formally, $D_l(\overline{T}_1) = \Phi_l(\overline{T}_1) \setminus \Phi_l(\overline{T}_1, \overline{T}_2)$; $D_l(\overline{T}_2) = \Phi_l(\overline{T}_2) \setminus \Phi_l(\overline{T}_1, \overline{T}_2)$

If the number of classes $k > 2$, $\Phi_l(\overline{T}_1, \overline{T}_2, \ldots, \overline{T}_k)$ is defined as a totality of *l*–grams common for at least to a pair of samples from $\{\overline{T}_1, \overline{T}_2, \ldots, \overline{T}_k\}$ and complements are respectively $D_l(\overline{T}_i) = \Phi_l(\overline{T}_i) \setminus \Phi_l(\overline{T}_1, \overline{T}_2, \ldots, \overline{T}_k)$, $1 \le i \le k$.

If the classes of texts under consideration are close enough, the complements can occur to be weak (with a small number of *l*–grams). In this case, some *l*–grams with a «contrast» property taken from the intersection $\Phi_l(\overline{T}_1, \overline{T}_2, \ldots, \overline{T}_k)$ can be used for classification. The «contrast» property assumes that l-gram should have its maximum representation in one of samples (in different realizations) and its minimum representation in all the remaining samples. Algorithms for calculations of the frequency spectrum and intersections of two and more spectra and their complements are based on hashing procedures (Gusev, Kosarev, Titkova, 1975) (Gusev, Titkova, 1994).

3. Description of Experiment

The samples of the Russian - $\overline{T}_R$ (*219* melodies), $\overline{T}_F$ – French (*338* melodies), and $\overline{T}_A$ – American (*140* melodies) folk songs of different genres were analyzed. The total length of melodies in *IS*-–representation for the first sample is N_R = 9197, for the second sample, N_F = 18641, and for the third, N_A = 7779 symbols.

In the course of the experiment the full frequency spectra were obtained for each sample $\left(\Phi(\overline{T}_R), \Phi(\overline{T}_F), \Phi(\overline{T}_A)\right)$, their intersection $\Phi(\overline{T}_R, \overline{T}_F, \overline{T}_A)$, and complements $\left(D(\overline{T}_R), D(\overline{T}_F), D(\overline{T}_A)\right)$. On the base of frequency spectra some integral numerical values characterizing each of three samples on the whole were obtained. Some examples are given below.

(1) *Recitativity factor* k_r shows the occurrence frequency in sample $\overline{T}$ of the β *IS*–codes with $|I|=0$, which corresponds to the sound repetition at the same pitch ($\beta=(0++)$ or $\beta=(0+-)$). Formally, for the sample $\overline{T}$ with the total number of *IS*–codes $N(\overline{T})$ it is $k_r = \left(F(0++)+F(0+-)\right)/N$ where $F(\beta)$ is the occurrence frequency of the code β in $\overline{T}$.

(2). The coefficient of asymmetry of the pitch line k_{as} indicates an averaged difference in steepness of the growing and drop of individual peaks forming the melodic contour. Formally, $k_{as} = k_+ / k_-$, where k_+ is the mean value of the interval for ascending motion, k_- is for the descending motion. They are calculated in the following way: $k_+ = \sum_{I\geq 1} |I| \cdot n^+(I) \Big/ \sum_{I\geq 1} n^+(I)$, $k_- = \sum_{I\leq -1} |I| \cdot n^-(I) \Big/ \sum_{I\leq -1} n^-(I)$ where $n^+(I)$ is the number of *IS*–codes of the sample $\overline{T}$ with values $|I|++$ and $|I|+-$, respectively, $n^-(I)$ is the number of *IS*–codes with values $|I|-+$ and $|I|--$.

These and a number of other integral characteristics bear an important information on the differences of samples on the whole. Local characteristics allowing to judge of the «national» belonging of certain melodies will be considered in Section 5.

4. Comparative Analysis of Integral Characteristics of Samples $\overline{T}_R$, $\overline{T}_F$ and $\overline{T}_A$

On the average, Russian melodies are less recitative than French or American $\left(k_r(\overline{T}_R)=0.16; k_r(\overline{T}_F)=0.23; k_r(\overline{T}_A)=0.28\right)$. Some special experiments have shown that these differences are significant in the sense that they exceed substantially the spread admissible because of random factors. A significant difference is also observed for the mean lengths of maximum recitative chains: $\bar{l}_r(\overline{T}_R) = 1.74$, $\bar{l}_r(\overline{T}_F) = 2.54$, $\bar{l}_r(\overline{T}_A) = 3.33$. The effect of asymmetry of the pitch line displayed in Russian songs to a larger extent than that in the French or American: $k_{as}(\overline{T}_R) = 1.43$, $k_{as}(\overline{T}_F) = 1.11$, $k_{as}(\overline{T}_A) = 1.10$. This fact means that, on the average, Russian melodies are characterized by a steeper jump-like raise and by a smoother gamma-like drop of the pitch line. The relative fraction of large jumps with values $I \geq 4$ in the Russian samples is *2–3* times larger than that in French and American samples. For example, $F(7++)$ is equal to *28* (for $\overline{T}_R$), *22* $(\overline{T}_F)$ and *10*($\overline{T}_A$); similarly, $F(7+-)$ is equal to *29* (for $\overline{T}_R$), *19* $(\overline{T}_F)$, and *12* ($\overline{T}_A$). When comparing these values one should taking into account the difference in numbers of melodies. These specific features impart to Russian melodies their distinctive emotional color.

Among other features, we note the following. Absolutely symmetric *saw-like* melodic lines are untypical for Russian melodies, admissible for American and ordinary for French melodies as is seen from examples given below:

Chain α	$F_\alpha(\overline{T}_F)$	$F_\alpha(\overline{T}_A)$	$F_\alpha(\overline{T}_R)$
(1– +1+ –) (1– +1+ –)	16	4	0
(2+ –2– +) (2+ –2– +)	8	4	0
(2– –2++) (2– –2+ +)	11	2	0
(3– –3+ +) (3– –3+ +)	6	0	0

Repeating elements of the inner structure in Russian melodies are more *variative* than in French and American melodies. *An alternate meter* is rather typical for Russian melodies (occurred in approximately 20% of cases), is less characteristic for French (11.4%) and practically absent in American melodies. Significant differences between samples are also observed both in the initial and final stages of the melody. So, elements *3– –*and *4 + –* are occurred as the end stages (with the only exception) in Russian sample (*15* and *17* cases).

5. Nationally Specific Intonation Fragments

First of all, we mean the chains forming complements $D_1(\overline{T}_R)$, $D_1(\overline{T}_F)$, and $D_1(\overline{T}_A)$ i.e. the unique *l*–grams represented only in one of the samples ($\overline{T}_R, \overline{T}_F$, or $\overline{T}_A$). Second, we consider the chains from the intersection $\Phi_l(\overline{T}_R, \overline{T}_F, \overline{T}_A)$ having the «contrast» property (see section 2). According to the dominance in $\overline{T}_R, \overline{T}_F$, or $\overline{T}_A$, these chains are divided into subsets $K_l(\overline{T}_R)$, $K_l(\overline{T}_F)$ and $K_l(\overline{T}_A)$.

When forming complements we confined ourselves with the values *l =3* and *4* since the complements of the first and second orders are nearly empty and the complements of the 5^{th} and higher orders consist already of *l*–grams with frequency *F=1*. When selecting the contrast chains α two limitations were used:
1) $F_\alpha(\overline{T}_R) + F_\alpha(\overline{T}_A) + F_\alpha(\overline{T}_F) \geq 10$; 2) the relative occurrence frequency α in the sample, where it dominates, should exceed the sum of its relative frequencies in two other samples.

With an account for the above mentioned limitations the subsets D_3 are presented by *44* chains *(l=3)* for $\overline{T}_A$, *84* (for $\overline{T}_R$), and *106* (for $\overline{T}_F$). The *3*–grams selected identify unambiguously *56%* of all the melodies in $\overline{T}_A$, *70%* in $\overline{T}_R$, and *76%* in $\overline{T}_F$. Sets D_4 are larger: $|D_4(\overline{T}_A)| = 71$, $|D_4(\overline{T}_R)| = 99$ and $|D_4(\overline{T}_F)| = 141$ (chains). The learning samples are overlapped by the unique 4-grams as follows: *84%* (for $\overline{T}_A$), *86%* ($\overline{T}_R$), and *97%* ($\overline{T}_F$). Some examples of chains from D_3 and D_4 for different samples with the indication of their occurrence frequencies in ($F_\alpha(\overline{T})$) and the number of melodies where they were found ($m_\alpha(\overline{T})$), are given below.

l–gram (α)	$\overline{T}$	$F_\alpha(\overline{T})$	$m_\alpha(\overline{T})$	*l*–gram (α)	$\overline{T}$	$F_\alpha(\overline{T})$	$m_\alpha(\overline{T})$
7+ –1– +1– +	R	8	6	2+ +1– –1– +4+ –	R	9	8
1– –4+ +2– +	R	7	5	3– +2+ –1– +1– +	R	8	6
0+ –1+ +2– –	F	21	15	0+ +0+ –1+ +2– –	F	15	10
3+ –3– +0+ –	F	8	8	2+ +0+ –1– +1– +	F	15	9
0+ –2– –1– +	A	7	6	0+ –0+ –0+ +0+ –	A	8	6
0+ –0+ –0+ –	A	6	4	5– +3+ –0+ +1+ –	A	6	4

As to the contrast *l*–grams, it is worth mentioning that their maximum number is referred to *l=3* and they cover *95%* of «their» samples. Some examples of the contrast *l*–grams are given below ($K_l(\overline{T})$–type is determined by maximum $F_\alpha(\overline{T})$).

l–gram (α)	$F_\alpha(\overline{T}_R)$	$F_\alpha(\overline{T}_F)$	$F_\alpha(\overline{T}_A)$	*l*–gram (α)	$F_\alpha(\overline{T}_R)$	$F_\alpha(\overline{T}_F)$	$F_\alpha(\overline{T}_A)$
(R): 5+ – 1– +	58	12	5	*R):* 0+ + 5+ –1– +	20	1	0
(F): 3+ + 3– –	2	40	0	(F):1– +1– +1+ +	2	20	0
(A): 0+ – 0+ –	3	9	38	*(A):* 0+ +2– –1– +	1	7	18

An analysis of chains from D_l and K_l supports the regularities mentioned in the preceding Section.

Note that *4* types of sets of informative *l*–grams obtained (D_3, D_4 , K_3 , K_4) are quite independent. The independence of D_3 and D_4 of K_3 and K_4 follows from their definitions and their substantial difference in the threshold values of frequencies ($F \geq 2$ in the first case and $F \geq 10$ in the second case). At the same time, D_3 and D_4 (similarly to K_3 and K_4) are partially dependent since the unique or contrast *l*–gram can keep this property at the left-hand or right-hand expansion. However, the chains obtained in such a way are not dominant. A substantial fraction of sets at *l=4* consists of chains, which at *l=3* did not yet satisfy the required conditions, i.e. they were not neither unique nor contrast chains, and at *l=4*, they acquired this property.

Sets of specific *l*–grams formed for each sample can be used not only for *classification* of unknown melodies but also for finding the *most typical* and *untypical representatives* in each class. For example, the melody from $\overline{T}_R$ with the maximum number of chains from $D_1(\overline{T}_R)$ can naturally be called «typically Russian». At the same time, the melody from $\overline{T}_R$ with the maximum number of elements from $K_1(\overline{T}_F)$ can be called «the most French among Russian melodies», etc. Since sets D_3, D_4, K_3, K_4 are not closely correlated, we have some relatively independent possibilities for defining the «typical representative». The "*most typical representative*" should be considered as the representative which satisfies maximum number of criteria. Fig.2 shows examples of «the most typical representatives» for each sample with the indication of some informative fragments. Each melody satisfies at least two criteria.

Conclusion

The method of search for informative features in the problem of differentiation of melodies by their “nationality” is proposed. The *most and least typical representatives* of each class were found. The most typical representatives can be considered as some centers of clusterization of melodies. The presence of such a *clusterization* displayed in the form of the unconscious adaptation was mentioned earlier (Bakhmutova, Gusev, Titkova, 1997).

Acknowledgment

This work is supported by the Russian Fund of Basic Researches, under grant

N 00-06-80420

References

Bakhmutova I.V., Gusev V.D., Titkova T.N. (1990) Repetition and variation in Melody: Towards a Quantitative Study (on the material of the composed songs) //Quantitative linguistics, vol.43.--Musikometrica 2. Universitatsverlag Dr. N.Brockmeyer, Bochum, p.p.143-168.

Boroda M.G. (1990) Rhythmical repetition/Rhythmical variation in folk music and composed music: Towards a Quantitative Study //Quantitative linguistics, vol.43.--Musikometrica 2. Universitatsverlag Dr. N.Brockmeyer, Bochum,p.p.121-140.

Zaripov R.Kh. (1983) The Computer search for variants on the simulation of creative work, M: Nauka, 232 p (in Russian).

Gusev V.D., Kosarev Yu.G., Titkova T.N. (1975) On the problem of search for repeating fragments in text./Associative coding, Novosibirsk, issue 62,Computing Systems, pp.49-71 (in Russian).

Gusev V.D., Titkova T.N. (1994) The hashing of symbolic chains in mode of sliding window./Analysis of sequencies and data tables. -Novosibirsk, issue 150, Computing Systems, pp. 94-106(in Russian).

Bakhmutova I.V., Gusev V.D., Titkova T.N.(1997) The Search for Adaptations in Song Melodies //Computer Music Journal. Santa Barbara, Vol.21, No.1, p.p.58-67.