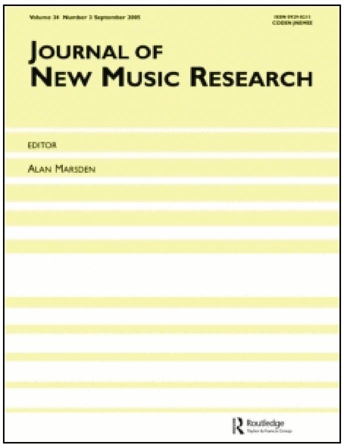
This article was downloaded by: *[Aalborg University]*
On: *11 May 2011*
Access details: *Access Details: [subscription number 912902580]*
Publisher *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of New Music Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713817838

Segmentation of Hungarian Folk Songs Using an Entropy-Based Learning System

Zoltán Juhász

Online publication date: 09 August 2010

To cite this Article Juhász, Zoltán(2004) 'Segmentation of Hungarian Folk Songs Using an Entropy-Based Learning System', Journal of New Music Research, 33: 1, 5 — 15
To link to this Article: DOI: 10.1076/jnmr.33.1.5.35395
URL: http://dx.doi.org/10.1076/jnmr.33.1.5.35395

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Journal of New Music Research
2004, Vol. 33, No. 1, pp. 5–15

Routledge
Taylor & Francis Group

Segmentation of Hungarian Folk Songs Using an Entropy-Based Learning System

Zoltán Juhász

Research Institute for Technical Physics and Materials Science, Budapest, Hungary

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

Abstract

A memory-based maximum entropy model has been developed to simulate the learning in an oral musical tradition, with the aim to find the optimal segment streams in melody sections. The model operated on a representative set of 2323 Hungarian folk songs. The pentatonic motive boundaries became more and more preferred during convergence, which shows the self-supporting feature of pentatonality in certain melodic systems. A pronounced correlation between typical segment contours and complete section contours poses the existence of certain typical schemata at macro- and micro levels in the melodies.

1. Introduction

"He plays wonderfully, but I don't understand any word of this music," said an old traditional Hungarian bagpipe player, when for the first time in his life he heard, an old Serb piper play. With all certainty, he expressed the basic problem that he was not able to separate characteristic parts of the Serb piece on the basis of his preliminary knowledge about motive structures in melodies. At the same time, he "understands" Hungarian tunes at the first hearing; moreover, he can compose melodies which fit to any Hungarian melody type. People, living in a well-defined oral musical tradition, obtain this ability during a long, unconscious, and mostly passive learning process which begins "nine month before the birth of their mother" (expression of Kodály) and continues their whole lives. However, the most effective learning period is in childhood, when the musical communication system evolves in a way which is very similar to that of learning to speak their native language. This learning is based on an oral musical tradition, therefore, the resulting musical communication system can be well applied in the native musical culture to perceive music and performing variation and improvisations (Hood, 1971; Nettl, 1983; Wiora, 1941). However, problems may more or less emerge when hearing melodies of a different tradition. This phenomenon was studied in a psychological test, comparing the uncertainty of parsing experiments when segmenting melodies of familiar, as well as of unfamiliar cultures (Oura & Haatano, 1988). In the present paper, we describe a computer model with the aim of simulating the above-mentioned learning process in a special musical tradition.

According to the quoted sentence, our old master tried to "understand" the heard melody by subdividing it into musical "words." Parallel to this conception, we defined the elementary operation of our learning model as a melody-segmenting experiment, using an adaptively adjusted optimality-criterion. The most frequently applied melody segmentation techniques can be divided into two main groups. In the first group, segmenting is based on pre-defined and data-independent rules. Using such rules, the so-called Local Boundary Detection Model (LBDM) determines a boundary strength value between each couple of notes and determines the segment boundaries at the maximal strength values (Cambouropoulis, 1996, 1998). Due to their pre-defined rules, such rule-based methods are not available to model any learning process. The second group of segmenting techniques, having been elaborated originally for text parsing, is based on a search for the most probable segmentation of the melodies. The so-called Treebank grammar defines this probability as the product of the frequencies of the phrases constructing the given segmentation, while the Markov grammar defines a

Accepted: 16 February, 2003

Correspondence: Zoltán Juhász, Research Institute for Technical Physics and Materials Science, P.O. Box 49, Budapest H-1525, Hungary. E-mail: juhász@mfa.kfki.hu

0929-8215/04/3301-005$16.00 © 2004 Taylor & Francis Ltd.

product of conditional probabilities (Charniak 1996, 2000; Seneff, 1992). The Markov technique has been developed by incorporating global contexts of the melody into probability conditions, in search of segmenting folk songs (Bod, 2001). The probabilities of these methods are calculated on the basis of a large amount of input melodies, thus, they are called data-based techniques. We elaborated a new data-based method which determines the conditional entropy of the phrases and defines an average entropy increment value for a given segmentation. The optimal segmentation is defined by the maximum of the entropy increment. Maximum-entropy based parsers have also been elaborated for text parsing (Charniak, 2000; Ratnaparkhi, 1999).

The operation of data-based methods is determined by their input melody collection. Therefore, these methods are suitable to study certain specific features of a given culture, supposing that the data base is built up by a representative collection of melodies. The database of our model is a digitised collection of 2323 Hungarian folk songs, well representing the most important Hungarian melody types. In order to identify historical layers and structural classes of Hungarian folk music, the research goes back a century (Bartók, 1981; Kodály, 1971; Járdányi, 1974; Kovács 1982; Corpus Musicae Popularis Hungaricae, 1951–1997). These works have shown that a significant part of archaic melodies is pentatonic, but the complete scale of Church keys is also well represented, without any dominance of major and minor. The influence of harmonisation rules on the melodies is negligible, which shows the unison habit of this musical culture. The overwhelming part of pentatonic tunes is taken by la-pentatonic melodies. (The la-pentatonic scale can be derived whether from dorian or minor scales, by deleting semitone intervals. For instance, a melody with the tonic G is called la-pentatonic if it is constructed by the notes G – B flat -c-d-f-g.) The section structure of Hungarian folk songs has been characterised in the fundamental works of Bartók and Kodály, using a complex analysis of line ending notes, rhythmic regularities, syllabic structure, and rhyme schemes (Bartók, 1981; Kodály, 1971). This study showed that the melodies are constructed mainly as four-section structures, but strophic systems of 2–8 sections, as well as motivic structures, are also found. Due to the dominance of four-section structures, we speak usually about "first, second, third, and fourth sections," instead of the more precise notation of "first, second, before-last and last sections." As a result of this research, the above-mentioned 2323 vocal melodies, called the "Old Hungarian folk song types," have been selected from the complete collection of nearly 200000 melodies to represent this musical tradition from the very beginning to the middle of the 19th century (Szendrey, 1978; Dobszay, 1978, 1992). Based on this corpus, we constructed a "basis collection," including all interval sequences appearing in the 2323 melodies. The "bases" were characterised by their appearance number and the conditional entropy, calculated from the probabilities of their continuations.

The optimal segment stream of a given melody is determined by the above "basis collection." The initial values of the appearance numbers and conditional entropies can easily be determined, and the resulting initial state of the basis collection can be regarded as the most unprejudiced "knowledge" about the musical phrases in our melody collection. However, when segmenting a melody, we pick out one possible segment stream as the valid solution, and the remaining possible segment streams are considered invalid. This decision modifies our knowledge about the actual appearance of motives, therefore, the basis collection has to be modified. It follows from this consideration that any segmentation experiment is essentially a phase of a learning process which results in a more and more prejudiced but at the same times a more and more valid picture about typical and less important motives in the studied musical system. The learning is modelled by an adaptive algorithm, modifying the appearance numbers and the conditional entropies step by step. In order to understand some features of traditional musical cultures, we investigated the convergence of this learning system from several different points of view. Chapters 2, 3, and 4 give a mathematical description of our learning model. The results of the simulations are discussed and interpreted from a musical point of view in Chapters 5 and 6.

2. The basis collection

As we have described in the previous chapter, the database of our system was a representative set of 2323 Hungarian folk songs, called the "Old Hungarian folk song types." The sub-division of these melodies to strophes and the strophes to sections is very clearly determined by the textual and musical characteristics in these songs. Therefore, our aim was to find shorter sequences – sequences of 2, 3, . . . , 6 notes – which are more "frequent" and "stable" in a certain sense, and thus can be considered as characteristic sub-sectional "motives."

Using the above melody collection, we first constructed the complete collection of "bases," which contains all intervals and interval sequences appearing in the melody corpus, from 1 to 5 intervals. Figure 1 shows an example of how the intervals (that are called "1 dimensional bases") and the sequences of two intervals ("2 D bases") are determined in a given melody section. The intervals are expressed with the number of semitones constructing them. For example, a minor third interval can be 3 or −3, depending on the direction of the movement. The number denoting how many times the given "basis" appears in the entire corpus is also stored in the collection. The appearance number of the k-th basis of dimensionality d is denoted by $n_d(k)$.

These data allow us to determine the conditional entropy of each "basis," seen as follows. Let us select the k-th basis of dimensionality d: $basis_d(k)$. Any sequence can be continued by a further interval, usually not larger than one octave. Using the appearance numbers of the $d + 1$ dimensional bases, the probabilities of the continuation can be determined: $\pi_{-12}, \pi_{-11} \ldots \pi_{11}, \pi_{12}$. Here, π_i is the conditional prob-

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

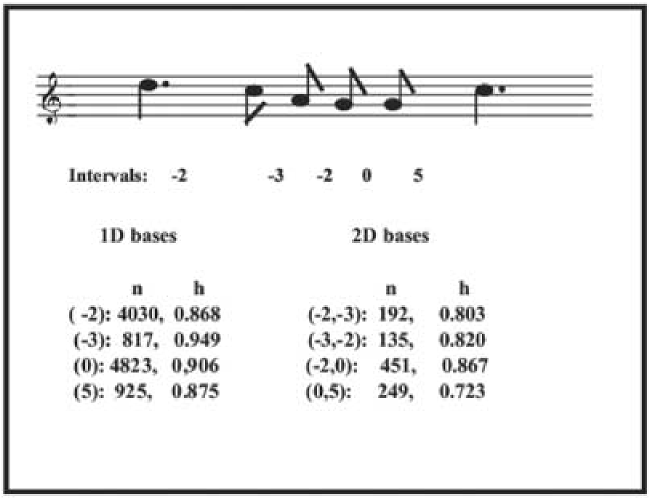


Fig. 1. Generation of 1D and 2D bases in a melody section. Additional data characterise the appearance number (n) and the conditional entropy (h) in the whole basis collection.

ability of the event that the sequence $basis_d(k)$ is followed by the interval i:

$$\pi_i = prob(basis_d(k), i \,|\, basis_d(k)). \tag{1}$$

Since the overwhelming part of the intervals is in the range of ± one octave, the set of conditional continuation probabilities is complete in the region of ±12 semitones $\left(\sum_{i=-12}^{12} \pi_i = 1\right)$, so the conditional entropy of $basis_d(k)$ is

$$h_d(k) = -\sum_{i=-12}^{12} \pi_i \log(\pi_i). \tag{2}$$

This expression becomes low, if one of the continuations has a dominating probability, and reaches its maximum value if all the continuations are uniformly probable (Shannon, 1949). In other words, the conditional entropy expresses the predictability of $basis_d(k)$. Low conditional entropy of a $basis_d(k)$ shows that an expected continuation can easily be predicted, therefore, such a sequence is considered as a part of a longer segment. On the other hand, high conditional entropy of a $basis_d(k)$ shows that the continuation cannot be predicted, consequently, such a $basis_d(k)$ should be considered as a self-dependent "motive." Based on the above consideration, it is straightforward to characterise each $basis_d(k)$ by its conditional entropy $h_d(k)$. Additionally, the appearance number $n_d(k)$ is also stored for each $basis_d(k)$. Figure 1 shows a few examples for 1D and 2D bases, as well as the corresponding appearance numbers and conditional entropies.

3. The total entropy of the basis collection

Using the conception of information theory, the above basis collection can be considered as a statistical description of an information source which emits sequences of intervals (Shannon, 1949). Such information sources can be characterised by their total entropy H. It is clear that the sequences described in our basis collection are not independent, since each $basis_d(k)$ ($d > 1$) consists of a few other bases of lower dimensionality. Therefore, the whole entropy of the source must be calculated from the conditional entropy collection as follows.

$$H = -\sum_{k=-12}^{12} p_1(k)\log(p_1(k)) + \sum_{k=1}^{N_1} p_1(k)h_1(k) + \ldots + \sum_{k=1}^{N_5} p_5(k)h_5(k) = H_0 + H_1 + \ldots H_5 \tag{3}$$

where H_d is the expected value of the conditional entropy of the d dimensional bases (A derivation of the above expression is given in Appendix 1). The probabilities $p_d(k)$ are defined as the appearance frequencies of the bases on the basis of their own dimensionality:

$$p_d(k) = \frac{n_d(k)}{\sum_{i=1}^{N_d} n_d(i)}, \tag{4}$$

where N_d denotes the number of kinds of d dimensional bases (the total number of d dimensional bases is given in the sum of the denominator).

4. The principle of segmentation

Speaking about segmentation, we suppose that each melody section is constructed as a sequence of 2, 3, or 4 "motives." It follows from the construction of the basis collection that all possible "motives" are registered here. Segmentation of a melody section means that we search for a set of $basis_d(k)$-s which fills the section completely, without overlapping, and fulfils our optimality requirement. We call the optimal set of such bases "motives" of the given melody section. The concept of optimality is defined using the following consideration.

The most unprejudiced knowledge about the possible motives is embodied in the basis collection, since this stores all interval sequences, appearing in the melody collection. When decomposing a given melody section, our decision is based on this knowledge. We select 2, 3, or 4 sequences as "motives," while the remaining sequences in the section are considered superfluous (redundant). Therefore, each superfluous basis can be removed from the basis collection by reducing the corresponding appearance number by one. These modifications result in the change of the appearance number and conditional entropy of certain bases, and consequently, the variation of the total entropy, H, too. The next segmentation can be based on the new, modified basis collection which expresses a more realistic picture about the really "important" motives. It follows from the above consideration that any segmentation experiment can be considered as a phase of a learning process approaching the optimal basis collection with less and less redundant bases,

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

and therefore more and more knowledge about the more typical motives.

There remains only one question: how can the concept of "optimality" be defined in a mathematical form? To define optimality exactly, let us summarise the steps of a segmentation experiment. First of all, we delete all the interval sequences of the selected section from the basis collection. (As a consequence of this, conditional entropy of certain bases changes. However, supposing that the appearance numbers $n_d(k)$ are high, the variation of the conditional entropy can be neglected). After that, we accomplish the segmentation of the selected section and add the resulting motives to the basis collection. These modifications result in the change of the total entropy, H. We defined the optimal segmentation as the choice of motives, which results in the maximal increment of the total entropy H.

The entropy increment, resulted by the segmentation of a melody section, can be determined as follows. Adding a $basis_d(k)$ to the basis collection means that the appearance number $n_d(k)$ is increased by one. This leads to a new value of the total entropy, H', which can be easily calculated when the variation of the conditional entropies is neglected. It can be shown that the total entropy increment, caused by adding $basis_d(k)$ to the basis collection, $\Delta H = H' - H$ takes the form of

$$\Delta H = H' - H = \frac{1}{\sum_{j=1}^{Nd} n_d(j) + 1}(h_d(k) - H_d). \qquad (5)$$

(A more detailed derivation of the above form is given in Appendix 2. Although we supposed that the slow variation of the conditional entropies is negligible in the derivation, the algorithm systematically re-calculates all conditional entropies several times during the learning process.) Now, the average increment of the total entropy, due to a subdivision of a section to m parts, becomes

$$\Phi_m(\Delta H_1 + \ldots \Delta H_m)/m, \qquad (6)$$

and we have to chose the segment stream resulting in the maximal Φ_m. This value expresses the "clearness" of the segmentation. A high value of Φ_m shows that the motives, identified by the segmentation, are really uncontinuable, therefore the segmentation is clear. On the other hand, a low Φ_m value refers to the fact that a relatively predictable continuation can be guessed for some of the motives, so the segmentation produced a rather uncertain result.

The average conditional entropy H_d increases with decreasing dimension d, therefore a simple algorithm, searching for segments with the highest conditional entropies, would prefer low dimensional bases. However, the above form of ΔH (Eq. 5) prefers those $basis_d(k)$-s that have a relatively high conditional entropy compared to the average entropy in the group of their own dimension, H_d. This feature assures equal chances for segments of a different length. At the same time, the algorithm tends to equalise the total number of bases of different dimensions, since the total entropy increment is inversely proportional to the total number of bases with a given dimension d. This means a further guarantee to avoid the dominance of low dimensional bases.

As we have shown, sub-dividing is always based on the current state of the basis collection, which represents the actual knowledge about the motives. On the other hand, the main goal of segmentation is to increase the total entropy of the basis collection by deleting predictable sequences and saving those of highly uncertain continuation. Therefore, our definition of optimal segmentation must necessarily be interpreted in the frame of a learning process, where the aim of learning is to approach the maximum entropy state of the basis collection. As we have shown, this aim, formulated in the maximal entropy increment ΔH, produces motives with the possibly highest conditional entropy; in other words, motives of the possibly less predictability.

The learning process searches for the maximum of the entropy function in the space of the appearance numbers. In a given step, the algorithm moves the vector of appearance numbers in the direction of the highest entropy increment, which is estimated from the current state of the appearance numbers. This is essentially a discrete form of the well-known adaptive gradient search technique (Treicler, Johnson, & Larimore, 1987). Thus, the segmentation based on the requirement of maximal entropy increment (formulated in Eq.-s 5 and 6), is equivalent to a step of an adaptive gradient search process approaching the maximal total entropy of the system in the discrete space of appearance numbers. Since the total entropy is formulated as a complicated logarithmic function, it may have more local maximums, and the goal of our learning algorithm is necessarily limited to the search for one of them. (Furthermore, the absolute maximum is at the trivial point of $n_d(k) = 1$, since all continuation probabilities are equal and consequently, all conditional entropies are at maximum here.) However, there are some considerations, which control the convergence more rigorously. Firstly, the local extreme value found by the gradient search algorithm usually depends on the initial point of the search. In our case, this point is clearly defined as the total basis collection of the corpus. Secondly, as we have seen, the basis appearance numbers are reduced during the learning process. Therefore, at a given phase of learning, $n_d(k)$ expresses the "fitness" of a basis sequence much rather than its real number of appearance. However, segmenting the total melody collection, the "motive appearance number," $m_d(k)$ of any $basis_d(k)$ can also be determined, expressing how many times the given $basis_d(k)$ as an actual motive appeared. (These numbers do not necessarily agree with the basis appearance numbers, since the aim of learning is maximising the total entropy, and not to approach the two parameters to each other.) It is easy to see that $n_d(k)$ may not be less than the corresponding motive appearance number, since a sequence may not appear less timely as a basis than a motive (since the definition of "basis" is "motive or superfluous

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

sequence”). Therefore, the minimum value of $n_d(k)$ is limited by the current motive appearance number in any phase of learning: $n_d(k) \geq m_d(k)$.

Running the learning algorithm several times with randomly selected sequence of melody sections, the above additional constraint determined very close convergence points of the system. At the convergence point, the reduction of the appearance numbers stopped, which shows that the basis appearance numbers approached the motive appearance numbers. As a consequence of the stabilisation of the basis collection, the increase of the entropy also stopped. A typical learning process consumes 1–2 hours, depending on the accuracy requirement, which is defined as a maximum value of the variation of appearance numbers.

5. Results

We saved the basis collection at the convergence point (which represents the system state after learning) and we accomplished several segmenting experiments to compare the results to subjective human cognition. (This means nothing more than the subjective opinion of the author.) We found that the results can be accepted by a human in most cases, although the rhythmic characteristics were not taken into consideration in the model. However, a complex model, including rhythmic information would give more familiar results and probably more important lessons about motives in a melodic culture. A systematic human test would also be necessary. To illustrate the effectiveness of the method, a few examples of automatic subdividing are shown in Figure 2.

Fig. 2. A few examples of automatic segmentation. The number of segments (2, 3 and 4) was determined also automatically. Bar lines represent segment boundaries.

In order to find more comprehensive features of motives, we should define a general characteristic, independent of their rhythm and length (dimension). Therefore, we introduced the concept of a musical “displacement” which is the interval between the first and last notes of the sequence. The displacement can also be calculated as a sum of the intervals constructing the sequence. It follows from this definition that the displacement can be determined for any sequence, independently of its rhythm and length. The results below verified that displacement is a rather relevant musical characteristic of the motives and bases.

Firstly, we determined the basis appearance number distribution as a function of the displacement. Figure 3a shows the number of possible intervals in la-Pentatonic, Dorian, and Mixolydian scales in a domain of one octave, plus-minus one second below the tonic and over the supertonic. (F–G – B flat -c-d-f-g-a, F–G–A – B flat -c-d-e-f-g-a, and F–G–A–B-c-d-e-f-g-a, with the tonic G, respectively.) The curves show that second, quart and fifth, as well as recurrence, are generally the most dominant intervals in all of the studied modes. The less frequent intervals are the semitone, as well as augmented fourth and fifth. The main difference between diatonic and pentatonic modes formulates in the fact that the above-mentioned, less frequent intervals are completely prohibited in the la-pentatonic mode. Figure 3b shows the number of bases in our melody collection as a function of the displacement. Here, the appearance numbers of bases with a defined displacement were summed up, regardless of the dimension. The resulting sums were normalised to the number of bases with zero displacement. These basis number distributions are represented by a thin curve in the initial state and a thick curve after learning. Figure 3b shows that the maximum and minimum values of the basis number densities correspond to the theoretical diatonic curves. Comparing density functions before and after learning, one can see that quart and fifth displacements became relatively more frequent after learning. Therefore, the basis number distribution becomes more similar to the la-pentatonic curve after learning than it was before. A question arises from this result, does our learning model prefer the pentatonic scale when it determines motive boundaries?

In order to answer this question, we studied the correlation between the clearness of the motive structure and the frequency of pentatonic motive beginning and ending notes. As we have mentioned in the previous chapter, the sections can be ordered into a sequence according to their optimal Φ_m values, since these characterise the clearness of their motive structure. We selected the sections over a critical value of

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

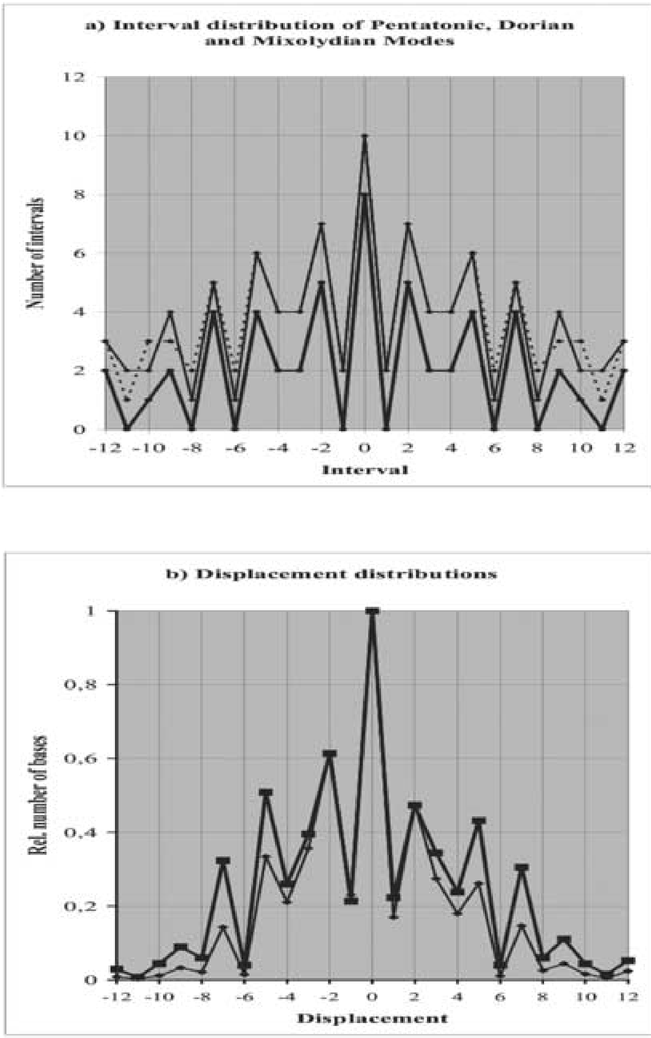


Fig. 3a. Interval distribution in la-Pentatonic, Dorian and Mixolydian modes (thick, thin and dashed lines, respectively). The intervals were summed up in the regions beginning one tone below the tonic and ending one tone over the supertonic.

Fig. 3b. Displacement distributions of the basis collection before and after learning (thick and thin lines). The data are normalised to the number of bases with zero displacement, for both cases.

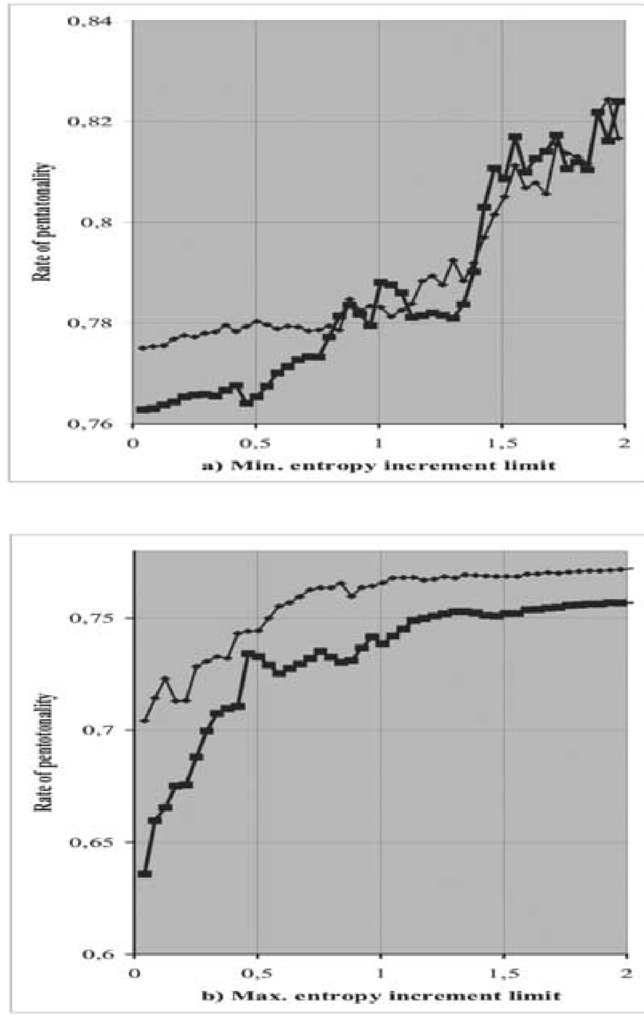
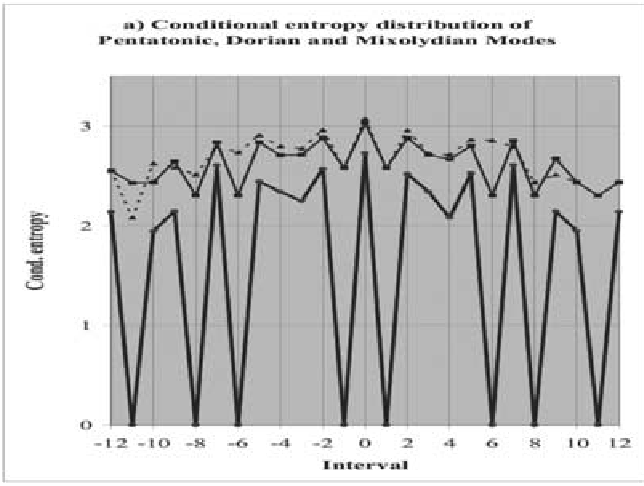


Fig. 4. Rate of pentatonality of motive beginning and ending notes (thick and thin lines), as a function of the clearness-limit.
a) Melody sections, where the clearness of segmenting (see Eq. 6) exceeded a critical value were collected and the ratio of pentatonic motive boundaries to the total boundary number was determined in the selected group. The curves show the resulting rates as a function of the critical limit.
b) Melody sections, where the clearness of segmenting became lower than a critical value were collected and the ratio of pentatonic motive boundaries to the total boundary number was determined in the selected group.

clearness ($\Phi_m > \lambda$), and determined the absolute pitch values of their motive beginning and ending notes. (Zero level of the pitch being at C. All melodies were transposed to the final tone G.) Now, we summed up the motive beginning (as well as ending) notes fitting to the la-pentatonic scale and determined the ratio of this number to the total number of the motives in the selected group. This rate, called the "rate of pentatonality" is represented in Figure 4a as a function of the limit of clearness λ. The figures show that the rate of pentatonality really increases with increasing clearness for both motive beginning and ending notes. Similarly, we determined the rate of pentatonality for sections beyond a critical value of clearness ($\Phi_m < \lambda$). Figure 4b shows that the smallest rate of pentatonality characterises melody sections with the most uncertain motive structure. These results verify our expectation that pentatonic motive boundaries are preferred by the segmentation algorithm. Moreover, the improved importance of quart and fifth intervals in the final state verifies that this phenomenon becomes more and more pronounced during learning (See Fig. 3b).

What is the reason of the dominance of la-pentatonic motive boundaries in sections with relatively clear motive structure? To understand this, we grouped the basis collection into subsets according to their displacements and determined the average entropy increment (ΔH in Eq. 5) values

Downloaded By: [Aalborg University] At: 14:08 11 May 2011



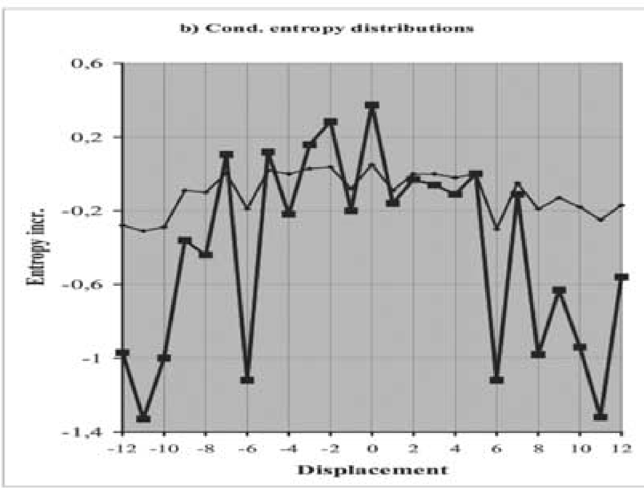


Fig. 5a. Theoretical conditional entropy of different intervals in la-Pentatonic, Dorian and Mixolydian modes. The conditions are identical to those of in Figure 3.

Fig. 5b. Average entropy increment (see Eq. 5) values of 1D-5D bases as a function of the displacement, before and after learning.

for each subset. The expected entropy increments are demonstrated as a function of the displacement for both the initial and final states of the basis collection in Figure 5b. The thin and thick curves (corresponding to initial, as well as final state) show that the addition of a motive with a zero, second, minor third, fourth and fifth displacement to the basis collection may cause the highest entropy increment, while the addition of motives with semitone, augmented fourth and augmented fifth displacement may cause the largest entropy reduction. This behaviour becomes more pronounced after learning, but it is seen in the initial state too. According to Figure 3a, intervals (displacements) of low entropy increment (large entropy reduction) are completely prohibited, while those of high entropy increment are the most frequent in la-pentatonic mode. Therefore, the clearness value Φ_m of a pentatonic melody probably becomes higher than that of a non-pentatonic one (see Eq. 6). The dominance of pentatonic motive boundaries in the clearly segmented melodies is explained by this consideration itself, but the study of non-pentatonic modes also supports the result. Figure 3a also shows that intervals (1D displacements) with low ΔH values have a non-zero frequency in Dorian and Mixolydian modes, therefore, motives of the lowest entropy increment may also be identified in non-pentatonic melodies. However, a segmentation with pentatonic motive boundaries, resulting in a higher clearness value of Φ_m, may also be accomplished in certain non-pentatonic tunes. Since the correlation between extreme values of entropy increment distribution and interval frequency distribution is also valid for the non-pentatonic curves, there is a good statistical chance to find pentatonic motive boundaries in such modes, too. Moreover, Figure 3b shows that the real basis number distribution also prefers displacements fitting to the pentatonic system, thus, a significant part of the melodies can really be segmented to motives with pentatonic boundaries, independently of their actual mode. The advanced frequency of fifth and quart displacements in the final state refers to the fact that pentatonic motive boundaries become more and more important during the learning process. Summing up the above results, the driving force of pentatonality has been identified as the correspondence between extreme values of expected entropy increment and basis number distribution functions, exactly at the prohibited and most frequent displacements of la-pentatonic mode.

Is the above-discussed driving force of pentatonality a special feature of the musical system in our study, or is it more general? In order to answer this question, the conditional entropy of different intervals was determined in the Pentatonic, Dorian, and Mixolydian modes. Instead of an abstract description of the method determining the conditional entropies, we illustrate it using a simple example. Let us consider the Dorian scale with the base of g, between notes f and a′. There are two semitone intervals in this scale, between a and b flat, as well as between e′ and f′. Both "bases" can be followed by a major third, (b flat -d′, and f-a), so the number of continuations of a semitone interval by a major third interval is 2. However, the number of augmented quart continuations is only one (b flat -e), and an octave continuation is completely impossible in the given region (the b flat -b′ flat continuation is prohibited by the upper limit a'.). All the possible continuations can be counted by a similar way, and the conditional entropy, expressing the predictability of a semitone interval can be determined from these numbers. We collected all the possible intervals of 0, $+-1 \ldots +-12$ semitones and determined the corresponding conditional entropies. The curves in Figure 5a show that second, quart and fifth intervals have the largest conditional entropy, while local minima exist at semitone, as well as augmented quart and fifth intervals. (Augmented quart has a rather high conditional entropy in Mixolydian mode.) It is reasonable to say that the entropy and interval distributions in Figures 3a and 5a can be regarded as characteristics of a

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

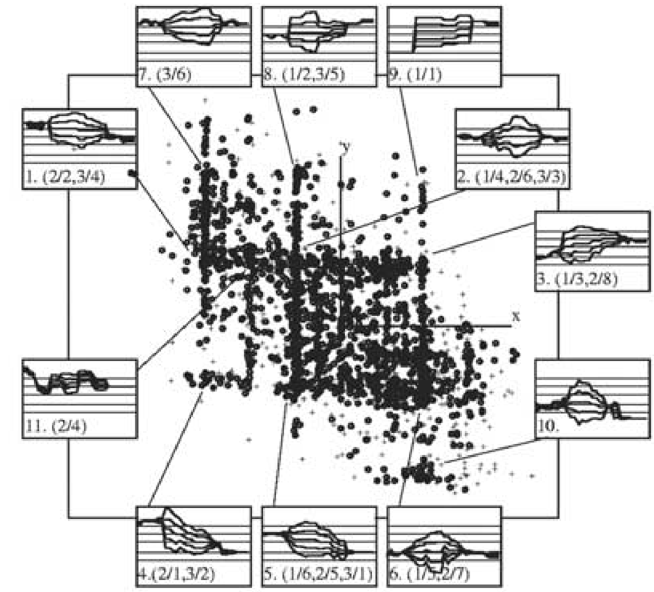
set of random melodies with equally probable pitches. Since the above-mentioned correspondence between maximal and minimal conditional entropies as well as basis number distributions can also be found in such a random system, it is straightforward to suppose that the driving force of pentatonality is a more general characteristic of melodic cultures.

In a previous work, we have shown that musical relations of melody contours can be well represented and interpreted as spatial relations of points in a multidimensional space (Juhász, 2001). Using a principal component analysis, we determined three orthogonal vectors in the multidimensional space, and we have shown that the points can be well approximated in the resulting 3D subspace. The basis vectors of the subspace describe the characteristic pitch at the beginning, ending, and central parts of the sections independently. A projection of the point system representing first melody sections is shown in Figure 6a (All melodies are transposed to the final note G.) The plane of projection is the first main plane of the 3D basis. X and Y co-ordinates determine the pitch level at the beginning and the end parts of the sections, respectively, while co-ordinates on the Z axis, which is perpendicular to the projecting plane, refer to the amplitude of the contour in the central part of the sections. With knowledge of the musical meaning of the three axes, the clustered point system can easily be interpreted. In point of fact, the figure shows an unbroken system of large flat clusters, being perpendicular to the projecting plane. The intersections of these flat clusters, called "axes," are also perpendicular to the plane of projection. The possible melody contours along the axes are also represented in Figure 6a. These schematic melody contours show that the beginning and end parts of the sections are determined along an axis, while central parts can vary depending on the third co-ordinates. The notes beginning and ending a section are tonic, dominant, and supertonic.

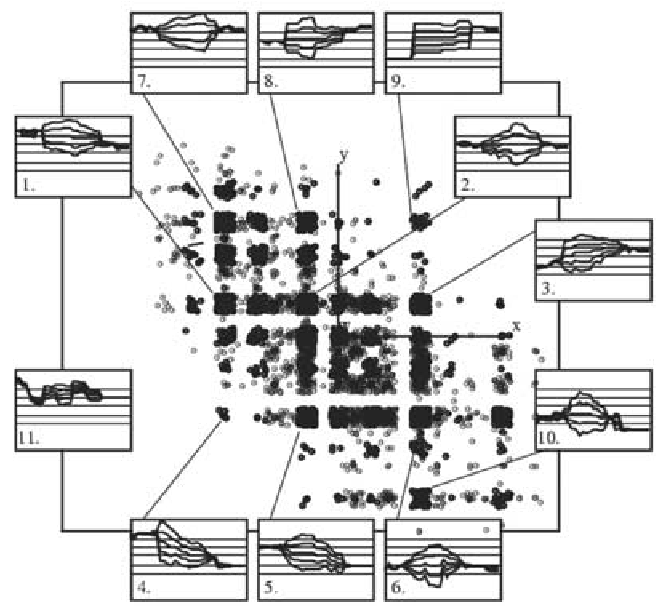
Since a contour can also be attributed to any motive constructing a section, it is obvious to map all motives in our melody collection to the same subspace. Although our basis collection is constructed by intervals, a segmentation determines concrete notes in the studied section, so each motive can be represented as a sequence of pitch values, too. Since the basis of the 3D subspace was determined by 32 dimensional vectors of the original "melody space," we augmented our motives to sequences of 32 pitch values. (For example, let the motive be G, c, B, G. The resulting augmented pitch vector is constructed by four sequences, each of them consisting of 8 elements, corresponding to G, c, B and G, as follows:

$$(7\ldots 7, 12\ldots 12, 10\ldots 10, 7\ldots 7).$$

As we have mentioned, the zero level is defined at C, and the pitch value increases one by one semitone.) The co-ordinates in the 3D subspace are determined as the scalar product of the above vector with the basis vectors. We determined the co-ordinates of all motives by a method, shown in this short example and projected the resulting point system to the plane



A



B

Fig. 6a. Map of first melody sections. Each point corresponds to the contour of a melody section. The multidimensional point system is projected to a plain. X and Y co-ordinates refer to the pitch at the beginning and ending parts, while the Z co-ordinate describes the amplitude in the central part of the section. The possible melody contours along the intersections of the large flat clusters ("axes") are also represented. (Upper map.)

Fig. 6b. Map of motives. The motive contours are augmented and represented in the above space. Full circles represent pentatonic motive boundaries, while empty ones represent motives where at least one of the boundaries is non-pentatonic. The possible motive contours along the axes show that the most important motives are shortened versions of characteristic melody sections. (Lower map.).

Fig. 7. Examples representing the connection between first sections and motives at the axes. The bar lines show motive boundaries.

of the basis vectors, determining the pitch at motive beginning and ending. The resulting point system is shown in Figure 6b. The well-marked similarity between point systems of motives as well as first sections refers to the fact that the typical contours of first sections and motives are identical. The principal component analysis of the largest clusters of motives verified that the resulting 2D subspaces are essentially identical to those of the sections. In other words, the most important motives are shortened versions of typical sections. Figure 7 shows a few examples of typical melody sections and corresponding motives, situated on the "axes." A particularly interesting example is the so-called "Peacock" section contour, which is very typical in old Hungarian folk melodies. This section contour, starting at the supertonic, descending to the sub-dominant and then ascending to the seventh, was identified as a motive in several melody sections too. These examples verify that typical contours of first sections are also encoded at the level of motives and an unconscious method of segmenting may be used to search for typical section contours in the heard melody, or melody part. It is worth mentioning here that point systems of second, last-but-one, and last sections are not identical to the first one. Therefore, first sections *may* have particular importance in the motive identification, and thus in traditional cognition and interpretation of Hungarian folk melodies. The idea mentioned above should be validated by further studies, including psychological tests.

6. Conclusions

The aim of our work was to simulate the natural learning process, which leads to the ability to conceive melody structures immediately at first hearing, which people who have in a traditional oral musical culture, pick up naturally. We modelled this learning process as an adaptive gradient search algorithm, maximising the total entropy of an information source, which "emits" 2323 typical Hungarian folk songs. We have shown that a decomposition of any melody section to the less predictable motives is equivalent to a step of the gradient search. At the same time, the "clearness" of the motive structure in any melody section can be characterised by the resulting entropy increment of the entire source. Although a systematic human test of the results exceeded our opportunities, some results show that this entropy-based model may be an appropriate tool for finding the motive structure of melodies. The most hopeful result in this respect is that the contours of the most characteristic motives have also been identified as well-defined contours of typical first melody sections. It seems realistic to consider that first sections function as archetypes of motives and that they are unconsciously fitted to the segments when conceiving the motive structure of melodies.

In the present work, we focused on sequences of intervals, neglecting the role of the rhythm. It seems trivial that rhythm plays an important role when people sub-divide melody parts to motives. In this way, the performance of our automatic system would also be improved by utilising rhythmic information. However, our simplification allowed us to introduce the concept of musical displacement, which highlighted the importance of pure pitch and interval information in some aspects of music perception. As a continuation of the present work, stable and typical rhythmic sequences could also be determined. After that, the interdependence between melodic and rhythmic structures could be investigated in a complex model. The main lessons of the current status quo, focusing on intervals, can be summarised as follows.

Although pentatonic melodies take the minority of our collection, the results show that motive beginning and ending notes, fitting to the la-pentatonic scale, are more and more preferred during the learning process. As a consequence of this, the motive structure of pentatonic melodies becomes usually clearer than that of diatonic tunes. Bartók and Kodály directed the attention to the relations between old pentatonic Hungarian tunes to certain melodies in the Volga-region, and they interpreted the similarities as a proof of the long-term memory of oral musical traditions (Kodály, 1971; Bartók, 1981). Our model shows that the requirement of entropy

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

maximisation (which formulates the requirement of well-arranged melody structure in a musical sense) operates as an inherent driving force of pentatonality. Since the motive structure of pentatonic melodies often proves clearer, and motive boundaries of non-pentatonic melodies often become also pentatonic, pentatonality seems a self-supporting feature in Hungarian musical tradition. Thus, pentatonic melodies in Hungarian folk music can be regarded much rather as products of a continuously effecting aesthetic principle, than as fossized records of an ancient musical culture.

We also investigated a hypothetical musical system of random melodies with a relatively large ambit (typically one octave ore more) and different modes. We have shown that the correlation between extremal values of conditional entropy and interval density functions also prefer pentatonality in this artificial system. Consequently, the driving force of pentatonality is not a special feature of our melody collection, and it operates in any musical system as an aesthetic requirement of clearly arranged melody structures. However, it is worth mentioning here that other aesthetic requirements, such as the predominance of harmonisation rules, may suppress the above effect. Therefore, the validity of our idea concerning the interdependence between segmentation and pentatonality must be restricted to musical cultures adhering to unison melodies with large ambit typical, such as Irish, Breton, Hungarian, Central-Asian, South-American, and Chinese. All in all, the validity of the model described here should be verified by studying further musical systems, including unison and polyphonic cultures.

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

Acknowledgement

The author is grateful to Dr. László Vikár, for proposing the idea of the present work.

Appendix 1. The Total Entropy of the Source Emitting Interval Sequences of Different Dimension

Let us consider a source emitting purely the 1 and 2 dimensional bases. The appearance probability of a one-dimensional $basis_1(k)$ is

$$p_1(k) = \frac{n_1(k)}{\sum_{i=1}^{N_1} n_1(i)},$$

where N_1 denotes the number of kinds of 1D bases (the total number of 1D bases is given in the sum of the denominator). Furthermore, we know the conditional entropy collection $h_1(k)$ which characterises the entropy of the continuation, supposing that the first interval of the 2D basis is $basis_1(k)$. It is well-known in information theory that the total entropy of the 2D system is

$$H = -\sum_{k=-12}^{12} p_1(k)\log(p_1(k)) + \sum_{k=1}^{N_1} p_1(k)h_1(k) = H_0 + H_1,$$

where H_0 is the entropy of the 1D bases and H_1 is the expected entropy of the continuation. Considering now the 3D case, the total entropy of the source becomes the sum of the 2D system and the expected entropy of the continuation, which can be calculated from the conditional entropy collection of the 2D bases.

$$H = -\sum_{k=-12}^{12} p_1(k)\log(p_1(k)) + \sum_{k=1}^{N_1} p_1(k)h_1(k) + \sum_{k=1}^{N_2} p_2(k)h_2(k) = H_0 + H_1 + H_2$$

Finally, the total entropy of a D + 1 dimensional source is

$$H = -\sum_{k=-12}^{12} p_1(k)\log(p_1(k)) + \sum_{k=1}^{N_1} p_1(k)h_1(k) + \dots + \sum_{k=1}^{N_D} p_D(k)h_D(k) = H_0 + H_1 + \dots H_D$$

Appendix 2. Variation of the Total Entropy Due to Adding $basis_d(k)$ to the Basis Collection

Adding a $basis_d(k)$ to the basis collection means that the appearance number $n_d(k)$ is increased by one. Supposing that the variation of the conditional entropy is negligible, the new total entropy can be written as

$$H' = H_0 + \dots + H_{d-1} + \sum_{i=1}^{N_d} \frac{n_d(i)}{\sum_{j=1}^{N_d} n_d(j) + 1} h_d(i) + \frac{1}{\sum_{j=1}^{N_d} n_d(j) + 1} h_d(k) + H_{d+1} \dots + H_5.$$

Using the above form and the expression of $p_d(k)$ (Eq. 4), the total entropy increment caused by adding $basis_d(k)$ to the basis collection, $\Delta H = H' - H$ becomes

$$\Delta H = \sum_{i=1}^{N_d} \left(\frac{n_d(i)}{\sum_{j=1}^{N_d} n_d(j) + 1} h_d(i) - \frac{n_d(i)}{\sum_{j=1}^{N_d} n_d(j)} h_d(i) \right) + \frac{n_d(i)}{\sum_{j=1}^{N_d} n_d(j) + 1} h_d(k)$$

After algebraic transformations, the above expression gets the simple form

$$\Delta H = H' - H = \frac{1}{\sum_{j=1}^{N_d} n_d(j) + 1}(h_d(k) - H_d).$$

References

Bartók, B. (1981). The Hungarian Folk Song by B. Bartók. B. Suchoff (Ed). Transl: M.D. Calvocoressi, Animations: Z. Kodály. State University of New York.

Bod, R. (2001). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research*, *31*, 27–37.

Bod, R (2001) *Probabilistic Grammars for Music Proceedings BNAIC 2001*, Amsterdam.

Cambouropoulos, E. (1996). A formal theory for the discovery of local boundaries in a melodic surface. *Proceedings of the Troisiémes Journées d'Informatique Musicale (JIM-96),* Caen, France.

Cambouropoulos, E. (1998). Musical parallelism and melodic segmentation, *Proceedings XII Colloquium on Musical Informatics*, Gorizia, Italy.

Charniak, E. (1996). Tree-bank grammars, *Proceedings AAAI-96*, Menlo Park, Ca, USA.

Charniak, E. (2000). A Maximum-entropy-inspired parser. *Proceedings ANLP-NAACL'2000,* Seattle, Washington

Corpus Musicae Popularis Hungaricae l-X. (1951–1997), Budapest.

Dobszay, L. (1978). Der Begriff Typus in der ungarischen Volksmusikforschung. In *Studia Musicologica XX*. (pp. 227–243). Budapest.

Dobszay, L., & Szendrei, J. (1992). Catalogue of Hungarian folk song Types. Budapest.

Hood M. (1971). *The ethnomusicologist*. New York: McGraw-Hill.

Járdányi, P. (1974). Experiences and Results in Systematizing Hungarian Folksongs. In *Studia Musicologica XXII* (pp. 17–20). Budapest.

Juhász, Z. (2000). Contour analysis of Hungarian folk music in a multidimensional music-space. *Journal of New Music Research*, *29*, 71–83.

Juhász, Z. (2000). A model of variation in the music of a Hungarian ethnic group. *Journal of New Music Research*.

Juhász, Z. (2001, October). Relations between structure, long-term stability and short-term variability of Hungarian folk music. *In: Stochastic Modelling of Music. Proceedings of the Fourteenth Meeting of the FWO Research Society on Foundations of Music Research. Ghent University*.

Kodály, Z. (1971). *Folk music of Hungary*. Budapest: Corvina.

Kovács, S. (1982). Über die Vorbereitung der Publikation von Bartóks grosser ungarischer Volksliedaugabe. In *Studia Musikologica XXIV* (pp. 133–155). Budapest.

Nettl, B. (1983). *The study of ethnomusicology. Twenty-nine issues and concepts*. Urbana: University of Illinois Press.

Oura, Y., & Hatano, G. Parsing and Memorising Melodies of Different Styles. *CMPC7 Book of Abstracts (pp. 322–325)*.

Pressing J. (1988). Improvisation: Methods and models. In J.A. Sloboda (Ed.), Generative Processing in Music (127–178).

Ratnaparkhi, A. (1999). Learning to parse natural language with maximum entropy models. (*pp. 1–28*) Boston: Kluwer Academic Publishers.

Seneff, S. (1992). TINA: A natural language system for spoken language applications. *Computational Linguistics*, *18*, 61–86.

Szendrei, J. (1978). Auf dem wege zu einer neuen Stilordnung der ungarischen Volksmusik. In *Studia Musicologica XX.* (pp. 361–379.) Budapest.

Shannon, C.E., & Weaver, W. (1949). *The Mathematical theory of communication.* Board of Trustrees of the University of Illinois, IL.

Treichler J.R, Johnson C.R., & Larimore, M.G. (1987). *Theory and design of adaptive filters.* New York: John Wiley and Sons.

Wiora, W. (1941). Systematik der Musikalischen Erscheinungen des Umsingens. *Jahrbuch für Volksliedforschung VII* (pp. 128–198.)

Downloaded By: [Aalborg University] At: 14:08 11 May 2011

Downloaded By: [Aalborg University] At: 14:08 11 May 2011