Emotion - Is it measurable ?

Shlomo Dubnov

Abstract

In this work a texture paradigm is introduced as a means for modeling certain types of music organization. This approach allows quantitatively to assess differences between musical materials which are not captured by musical "rules" or schemes. In the context of emotion it is assumed that relations exist between deviations from a given, existing 'normative' musical organisation and expressive inflections. Excitement and calmness are considered as deviations from a normative state that a given texture state induces. Different types of textures are considered and the degree of change is measured by statistical distance between their probabilistic models. This method is suggested to be a way to measure the amount of change (degree of inflection) that occurs in the musical material. An implementation of this work in a simple midi real time system is given.

A short Prelude

- "Can you imagine a system that evaluates the amount of emotional affect present in a music piece ?" I ask.
- "But this is completely absurd!" one might immediately respond. "Emotion is completely subjective, its sense being impounded by one's cultural and social backgrounds, aesthetic preferences and even momentary moods, predispositions and the most seemingless, fleeting sensations ..."
- "All right", I insist. "But let us assume a more limited system, let's say just something like a set of dials that dance according to the music as it changes towards more tense vs. more calm, becoming more soothing or more arousing ? Does this still require so much prior knowledge ?"
- "May be not, but then how can you define a measure of such changes ? Music is such a complex phenomena, with the interplay between its parameters being completely un-graspable !"
- "So let us consider just a very simple music - let's say an aleatoric texture or an ambient sound - may be now we can ?"

...

1 Introduction

To my experience, when listening to a contemporary piece of music, one often takes the first few seconds just to observe and learn the music (its orchestration, the opening music material, the first electronic timbres) by listening analytically, or may be critically, but without getting emotionally or otherwise attached to any "non musical" content of the first few bars. Only as the piece unfolds, one may allow himself (if not completely bored by now), to get carried away with the sounds, and the piece that succeeds to "take you away" will be probably the piece that you'll remember after the concert.

So, if you did eventually manage to follow the music, it means that you somehow succeeded to learn its properties and follow its course of changes. Exactly this issue of "learning" the music material and finding its trends, as they occur during the course of the piece, is the subject of this work. In the present paper an attempt is made to define a computational model that would allow us to quantitatively measure such changes, based on the notion of a normative state and deviation from the normative. Using simple mathematical measures of probability distance/separability, a statistical modeling approach is applied for the case of limited music material that can be broadly termed as "texture".

Although still lacking a good definition of "texture", this approach is based on the concept of texture/gesture as developed in (Cohen and Dubnov)[1]. In this earlier work the texture was viewed very generally as a statistical method of description of principles of musical organization, which is specifically suited for phenomena that can not be otherwise described by musical rules (syntactic schemes). This concept of texture includes diverse musical aspects such as melodic contours, patterns of ornamentation/figuration (Dubnov, Cohen, Wagner)[3], intonation and expressivity in performance, spectromorphological characterization of timbre (Dubnov and Tishby)[4],[5] and etc. Gesture is viewed as a change of texture over time, or in other words a control curve that modifies the textural (i.e. stationary statistical) organization properties.

The basic assumption underlying the current work in relation to emotion is that expressive inflections of the music material are the result of (or at least closely related to) a change in musical organization, and the more the new material deviates from earlier established normative, the more is the expressive load conveyed to the new material.

Assessing the changes in texture is done by measuring the statistical distances between the textures. A very limited test case, which was already suggested in (Cohen and Dubnov)[1] as a way to measure the expressive trends (inflections) of the musical material, is presented in the current paper. It is implemented in a simple midi program that learns first order statistics of the music material, and then compares the relative entropy between new material and the original model.

Musical theoretic aspects and the mathematical basics are discussed in the paper. The realized system with several music examples will be also presented in the demonstration concert.

2 Definition of the basic terms: gesture and texture

Texture, as we define it, can be viewed as a super-parameter that is related to various combinations of other parameters. In the most general sense, textural phenomena are familiar to us from nature as psychoacoustic factors that evoke various emotional associations. Music in a given style contains two kinds of information: one having to do with learned schemes and the other with textural information. A major property of texture is that it is not well defined quantitatively. By contrast, in most pre-twentieth-century styles the parameters (intervals, rhythms, and meters) are not only defined quantitatively but also belong to systems of learned schemes that serve as building blocks for the musical organization of the particular style. Difference in syntactic schemes refers to quantitative difference (for example, the major scale differs from the minor scale in the sizes of three intervals); whereas in texture the important factor is the relative relationship (greater than, less than, or equal to): for instance, the direction of change and not its exact magnitude. We have termed such a relationship a type.

Thus, in the most general sense we can define texture as the way of distributing the sound (of defined or undefined pitches) in the dimensions of frequency, time, and intensity. (This definition also holds true for timbre, but texture is divisible and refers to longer durations than does timbre.)

The change over time of the texture parameters, due to external control or, on a larger time scale, due to changes in musical structure, can be seen as an organization of the texture distribution on a higher hierarchical level. We use the term "gesture" to denote at large these processes of change.

A particular static texture can be considered only if the parameters of the process that generates the texture had been held constant for a certain amount of time. Having a normative state defined by a stationary distribution (texture) and having control functions to modify the distribution parameters (gestures), a basic hierarchy is established[1].

The hierarchical nature of the texture units is represented schematically in figure 1.

Hierarchical Texture Units

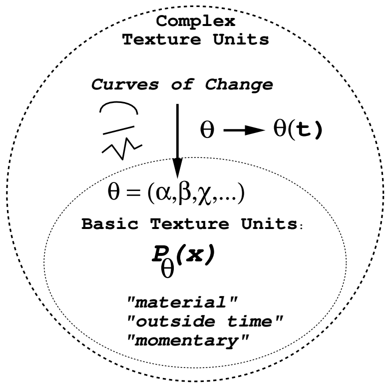

Figure 1: Control of texture parameters creates hierarchy of units.

3 Description of the experimental system

The purpose of the system is to serve as a test bed for principles of textural organization, in research and

[1] Considering the temporal structure in texture, one must be careful to distinguish between stationary textures having temporal structure, such as texture constructed by repetitive or random placement of basic elements or figures (with memory), and the complementary process of applying gestures which actually change the probability function and thus break the stationarity assumption.

composition. The system consists of a simple midi program realized in Max, which runs in real time. It operates in three modes: a "learning phase" during which the program builds a probability function estimate for a given material - this probability describes the "normative" state. A "measuring" phase when the probabilistic deviations of new material relative to the existing models are calculated and displayed, and a "generation" mode where new musical material is generated according to specified distributions.

As defined above, each musical realization of texture is a manifestation of a certain "type," and conversely, each "type" has many realizations, with the typical ones having high probability of occurrence. Thus, one might speak about texture either as a probability distribution P that underlies a certain type, or look at a typical musical example as a representative of that type. We can measure the difference/similarity by using probability separability measures $D[P||Q]$, which measure the "distance" between distribution functions P and Q. Such measures are widely used in the pattern recognition literature.

The system design includes a preprocessing step where the proper music representation is obtained. (In terms of generation, the choice of the music parameters defines the type of texture to be produced. For analysis, "feature extraction" is necessary for correct statistical modeling.) A learning module that constructs the probability tables of each feature, and finally a distance measure function that accepts pairs of probability tables and outputs a distance number according to the relative entropy between two probabilities.

Two versions of the program exist at the current stage:

1. Analysis tool: given segments of music materials (input as midi sequences), the probability distributions are estimated and the statistical distance is measured.

2. Generation Aid: Using a simple graphical interface (the table object), the user is allowed to design probability functions of different musical parameters. The amount of change between existing and candidate materials can be evaluated prior to execution (i.e. generation of the new material) and is graphically represented by a set of dials dials that give the amount of difference between the existing and the proposed distributions.

   Once the new probability is chosen, the probability tables are switched and a change in material (texture) occurs.

In both cases the statistics are limited to zero order (memoryless) distributions of the parameters.

The interface with the Generation program is demonstrated in figure 2.

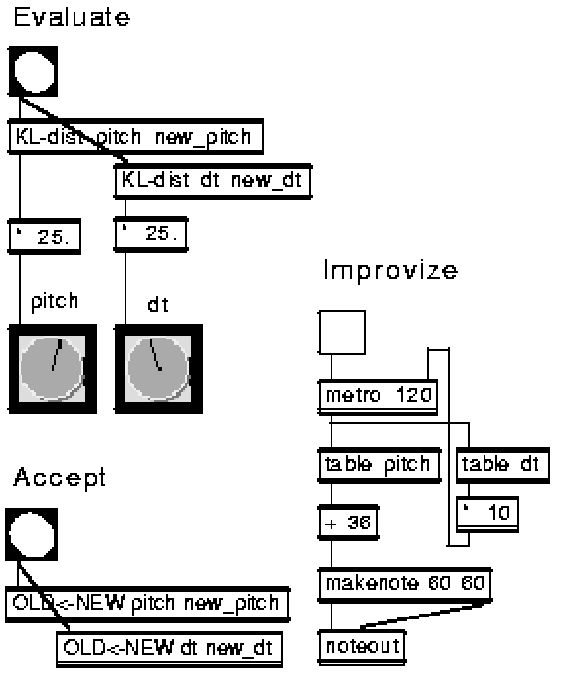

Figure 2: Generation of Aleatoric Texture. Note the dials that give the amount of difference between the existing pitch and delay time (time difference) distributions and the proposed new probabilities.

3.1 Calculation of the cross entropy

Let us turn now to the separability measure itself. The idea of looking at signals as random processes and replacing the signal by its stochastic model is a very powerful one. When one receives a new sequence of symbols, the model enables to judge how probable the observed sample is within a given model. Low probability would mean that the new sequence is "far" from the model. Musically, distortion measure says how probable it would be for instance for a Phrygian ornamentation figure to appear on a dominant chord or for a sound of slow strings to emerge from a bell resonance.

It is important to note that this is not a symmetric quantity, i.e. it is not the same as the probability of having a Ionian or mixolydian scale appear on minor chord[2]

Calculation of the distortion between probabilities $P(Y)$ and $Q(Y)$ is achieved by means of the Kullback-Liebler (KL) divergence (Kullback 1968) [7].

$$D[P||Q] = \left\langle \ln \frac{P}{Q} \right\rangle_P \tag{1}$$

Given a typical sample $Y = \{y_0, .., y_N\}$ drawn from probability P one might use the KL distance to evaluate the probability that the same sample appears in Q. It can be shown that for typical sequences (or if N is big enough) that

$$Q(Y) \propto e^{-ND[P||Q]} \tag{2}$$

In other words, close probabilities generate the same high probability (typical) sequences.

In our implementation the zero order probability is easily estimated using the Histo object. Since zero probabilities are not allowed in the Q distribution a small probability is given to zero probability elements.

4 Types of textures considered

The representation of the music material consists of a choice of the basic parameters, their scope and distribution, and is extremely important for the definition of the texture type. The examples at this stage are limited to very basic Pitch/Interval textures of aleatoric and repetitive types. It is important to note that in no manner the following discussion of the correspondences that occur (or not) between statistical description and their audible counterpart are suggestive of the KL distance being proper (or improper) measure in psycho-acoustical terms. It is rather ment to show the applicability of the methods for music research and composition.

Aleatoric texture, as demonstrated earlier in figure 2 is based on random pitch and delay time selection. The pitch selection can be done from either a 12 tone or a limited pitch set. The delay times can be either quantized uniformly (around 30 msec.) or logarithmically.

Since all parameters are generated independently in time, the statistical model for such a type is a first order, memoryless probability, or a histogram of the distribution of parameters of pitch, duration (and intensity).

Subjective results of listening to the computer generating sequences indicate that:

- a surprising (positive) correspondence exists between statistical distances and the perceived amount of change for pitch distributions. The following points of correspondence can be indicated:

  1. The convexity property of $D[P||Q]$: distance from a given state to a combination of two or more states is smaller then the distance to each one of the the components separately. This property seems to be partially supported by the tests.
  2. The amount of statistical distance among distributions is sensitive to the margins of the distributions. A similar sensitivity seems to audibly exist. (For instance, slightly varying the width of the probability or adding narrow but high probability “outliers”.)

- a smaller correspondence is found for time delay distributions. Using logarithmic scale increases the amount of change but also creates a metric illusion which is not intentional in the probabilistic description.

- using intervals (the difference representation) creates “random walk” pattern of the melody. The statistics seem related to factors such as smoothness/roughness of the melody[3]

- an interesting property is the concurrence/simultaneity of change among the different parameters. Situations where the change in entropies of the different parameters occur in the same direction seems preferable (more coherent but then maybe less tense) to a situation when changes are in opposite or unrelated trends. By comparing the actual entropy values we do not anymore consider the KL distance among two probabilities, but rather consider whether the new distribution moves closer or further from a uniform distribution [4]

[2] This type of “fuzzy belonging” was considered in (Dubnov et al.)[3]. It is too complex and unnecessary for our current needs. It is suggested that the relations between prior knowledge, and specifically musical schemes and texture is a whole new area of research.

[3] Characterization of the actual melodic curve would require pairwise of higher order statistics and is not considered in this work. Additional transformations, such as inversion, retrograde, modal substitution and etc., will not be discussed also.

[4] The entropy is proportional, up to a constant factor, to KL distance of a distribution from a uniform distribution.

Repetitive texture is based on probabilistic mixture of predetermined patterns (rhythmic and/or pitch). Although the patterns, which are discrete distributions of several notes, are not subject to change or control by the user, the important part of this method is in the mixing of patterns (morphing). This mixing is done by random switching between the original distributions, which is equivalent to weighted summation of original distributions. It can be shown (El-Yaniv et al.) that this method creates a distribution which is closest (i.e. minimal statistical distance) to the original components[5]

5 Conclusion

The proposed system allows for an easy implementation of many contemporary texture procedures, such as aleatoric, repetitive, hetero-phonic ant etc. The greater value of the setup is not only in testing the audible relevance of the different parameters, but also in suggesting statistical analysis tools for its quantification.

The following points are goal for further research:

- Implementation of more types of musical textures. This requires a computational definition of each texture type.
- Research of gesture. Display, recording and design of curves of change of the distribution parameters.
- Probability functions with memory. It is suggested to use a Markovian model for modeling of more complex textures. The evaluation of the distance function can be achieved by using existing "universal" sequence classification algorithms (ZL'78)[12].
- Design of an online segmentation algorithm. Such an algorithm would be useful for interactive performances.
- Research into hybrid, rule based and probabilistic systems for the purpose of understanding the relations between musical schemes and textural organization.

[5] Having two original distributions (patterns) P and Q and a mixture parameter $0 \leq \alpha \leq 1$ one looks for a distribution R which is minimal in terms of the combined distance $\alpha D[P||R] + (1-\alpha)D[Q||R]$. Due to the convexity property of $D[\cdot||\cdot]$, this minimum is the mixture $R = \alpha P + (1-\alpha)Q$.

Afterword

(instead of Musical Discussion)

The topic of the current workshop is emotion. But have we ever said a word about emotions ? Not as such.

Descartes[2] (1649) viewed emotions as passions of the soul and mentioned six of them. Modern psychology (Shalif)[11] added one more component, establishing a list of at least seven basic emotions (namely happiness, fear, anger, surprise, sorrow, interest and disgust). These emotions, claimed to be inborn properties present already in the infantile stage, manifest themselves in reactions to voice or facial expressions. As the personality of the individual develops, more complex relations among emotions are developed, and shape the individuals personality profile.

Although these basic emotional categories can not been applied directly to music, emotion in music is sometimes attributed the meaning of a "meaning", i.e. semantic component in the musical discourse. Musicologists used emotional categories to analyse *classical* repertoire (Meyer)[8] and even classify complete pieces (Petrushin)[10].

In the twenty century, when the non-dialectical principles in music prevail, the importance of the music organization principles on one hand, and of the basic feelings (which I intentionally avoid calling emotions) only increased. Works, in the post-serialist (including the aleatoric) approach, which are no more then realisation of the process that determines them, produce their form simply through the unfolding of the material. These must be judged/listened to according to the changes in material and understanding of the work as a process. In the repetitive music (what is sometimes called "American minimalism"), the emphasis on rhythmic pulse and the choice of well recognisable sound images can be viewed as an attempt to increase the importance of the basic feelings, devoid of dramatic or dialectic elements. Here again the unfolding of the works is done through a process of repetition and change, lacking directionality and predictability of the tonal schemes.

The need for a better understanding of texture is apparent in the recent compositional practice. Modern methods such as the spectral composition, introduce texture through paying closer attention to the dynamic aspect of timbre. Exploring the textural / counterpointal relations additionally to the more familiar timbral / harmonic correspondences seems a promising direction. In repetitive music, the attempt to escape from complete stasis calls also for more textural variations within and in transition between repetitive units.

In parallel, one might consider two approaches to the question of emotions in its "technological" aspect. Adopting the structuralist approach, one might want to look at the music shapes, much as we have done here (without any musical or aesthetic predisposition.)

On the other hand, a whole new approach of interactive art, which attempts to increase the level of communication among artist-artwork-audiences, strives towards a new understanding of the nonverbal communications. The appearance of computerized "emotional agents" (Nakatsu)[9] which are designed to recognize, react and thus communicate on the level of several basic emotions (such as happiness, fear, anger), might be representative of a new artistic concept, still to be pursued.

But then, from the computational point of view, the distinction between the two methods almost does not exist. It simply lies in different approaches to data analysis. In the first approach there are no assumptions about the nature of the data which is performed in an "un-supervised" manner. The second approach takes advantage of the psychological knowledge in order to guide the analysis towards predefined categories.

I would like thus to conclude this paper not by answering to the original question - Is it (emotion) measurable ? - but by posing a new question :
Emotion - should it be measured ?. Basically, we need to establish a better communication between us and the machine, in an artistic concept. Now it is up to the users intent to decide - does one want to have the data analysed in terms of emotional categories or he prefers to let the data "speak of itself" according to its various modes of existence? In both cases, the evaluation of the results and/or providing the right labels for the emotional contents of the data is a question for the artists / psychologists.

References

[1] D. Cohen and S. Dubnov, *Gestalt Phenomena in Musical Texture* to appear in "Studies in Systematic and Cognitive Musicology", ed. M. Leman, Springer-Verlag.

[2] R. Descartes, *Treatise On the Passions Of the Soul.*, 1649, (In The Philosophical Works Of Descartes: E.S. Halden and G.R.T. Ross, Trans., Cambridge University Press, 1967).

[3] S. Dubnov, D. Cohen and N. Wagner, *Uncertainty in Distinguishing between Structure and Ornamentation in Music*, in Proceedings of the International Fuzzy Systems Association Conference (IFSA), Brazil, 1995.

[4] S. Dubnov and N. Tishby, *Clustering of Musical Sounds using Polyspectral Distance Measures*, in Proceedings of the International Computer Music Conference (ICMC), Banff, 1995

[5] S. Dubnov and N. Tishby, *Analysis of sound textures in musical and machine sounds by means of higher order statistical features*, in Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP), Munich, 1997.

[6] Ran El-Yaniv, Shai Fine and N. Tishby, *Agnostic Clustering of Markovian Sequences*, unpublished manuscript.

[7] S. Kullback, *Information Theory and Statistics*, John Wiley and Sons, New-York, 1959.

[8] L.B. Meyer, *Emotion and Meaning in Music*, Chicago, University of Chicago Press, 1956.

[9] R.Nakatsu, *Image/Speech Processing that adopts an artistic approach*, in Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP), Munich, 1997.

[10] V.Petrushin, *Music Psychology*, Passim, Moscow 1994 (in Russian).

[11] I. Shalif, *The Emotions and the Dimensions of Discrimination among them in daily life* ,Ph.D. Thesis, Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel, 1991.

[12] J.Ziv and A.Lempel, *Compression of Individual Sequences by Variable Rate Coding*, IEEE Transactions on Information Theory, 24, September 1978.