Predicting Human Reactions to Music on the Basis of Similarity Structure and Information Theoretic Measures of the Sound Signal

S. Dubnov[1], S. McAdams[2], R. Reynolds[1]
[1]Music Department, UCSD,
[2]STMS-IRCAM-CNRS

Abstract

Memory, repetition, and anticipatory structure are considered important characteristics of musical style. Both composers and listeners often refer to these parameters in describing the music. In this work we conducted audio analyses in an attempt to determine correlations between audio features and human responses of Familiarity and Emotional Force. The analyses were performed on two recordings of an extended contemporary musical composition by one of the authors. Signal properties were analyzed using statistical analyses of signal similarity structure over time and information theoretic measures of signal predictability. The analyses show strong evidence that signal properties and human reactions are related. Application to style analysis are discussed.

Introduction

The current work[1] is based on a project that attempts to explore structural and affective aspects of human experience over time when listening to a musical work. Relatively few empirical studies of complete musical works have addressed the reaction of listeners across time. These works mostly relate the experience of musical emotions to psycho-physiological responses when listening to tonal music (e.g., Krumhansl, 1997). In this paper we attempt to relate human experience to statistical properties measured directly on the acoustic signal. The two questions investigated in the work are whether signal similarity grouping and the predictability structure of signal features could be related to familiarity and emotional content of an audio signal, respectively. The basic assumptions were that global spectral similarity should be related to human familiarity judgments, while the local anticipation structure (i.e. the predictability of signal features on a short time scale) might be related to the emotional affect. The experiments were carried out on a contemporary musical piece, *The Angel of Death* by Roger Reynolds for piano, chamber orchestra and computer-processed sound, in a live concert setting. The experiment consisted of collecting continuous ratings on two scales: Familiarity and Emotional Force. For the Familiarity Rating (FR) scale, listeners were to continually estimate how familiar what they were currently hearing was to anything they had heard from the beginning of the piece on a scale from "Completely New" to "Very Familiar". For the Emotional Force (EF) scale, they were to continually rate the force of their emotional reaction to the piece at each moment on a scale from "Very Weak" to "Very Strong". As a result, the obtained audio recordings and listener responses were aligned in time. This allowed us, among other things, to test various signal information processing methods in relation to human reactions. A preliminary report of the project was presented in McAdams et al. (2002).

This paper is a short version of a paper to appear in Journal of the American Society for Information Science and Technology, special issue on Style.

Psychological Research

In the realm of tonal music, several approaches to the evolution of emotional experience have been used. Krumhansl (1997) related the experience of musical emotions to psychophysiological responses. Sloboda & Lehmann (2001) studied listeners' perceptions of emotionality in reaction to different interpretations of a Chopin Prelude. Schubert (1996) has developed techniques for two-dimensional, continuous response to emotional aspects of music. Fredericskon (1995) used a continuous response method to track the online evolution of perceived tension.

The study on which the present analysis is based (McAdams et al., 2002, submitted) recorded continuous responses by listeners in a live concert as they heard *The Angel of Death* for piano, chamber orchestra and computer-processed sound by Roger Reynolds. Two response scales were used: familiarity and emotional force. The first one concerned perceptual and cognitive aspects of musical structure

processing, and the second one concerned emotional response to the music. The main findings of the analysis were that, although the piece had never been heard before and the style was unfamiliar to many of the listeners, the temporal shapes of the emotional experience and of the sense of familiarity were clearly related to the formal structure of the piece. Moreover, the piece elicits an emotional experience that changes over time, passing through different emotional states of varying force, and without having overlearned the stylistic conventions of the particular work or style.

The music

The structure of the piece (Reynolds, 2002, submitted) was conceived to allow experimental exploration of the way in which musical materials and formal structure interact. The piece is conceived in two main parts, one sectional (S) and the other a more diffusely organized domain (D) structure. Certain musical materials occur at the same place in time and in nearly identical form in the two parts (sometimes changing between piano and orchestral versions, sometimes between instrumental and computer-processed versions). Further the two parts can be played in either order (S-D or D-S), but the computer-processed part (evoking the angel) always starts at the end of the first part and continues throughout the second. This structure allowed for the study of the perception of certain materials under different formal settings (embedded in the sectional or the domain part, played alone or in the presence of the computer part, heard first in the sectional version or in the domain version, etc.).

Signal Analysis

The signal similarity was evaluated in terms of groupings within a spectral similarity matrix across time (also called signal recurrence matrix) using matrix-partitioning methods. As appropriate features, we used spectral envelopes that were estimated from short signal segments in a time-varying manner, represented by low-order cepstral coefficients. The similarity was obtained using Euclidian distance or dot product between normalized cepstral feature vectors. Partitioning of the similarity matrix by singular value decomposition results in a vector that represents plausible similarity grouping structures. This method of grouping analysis, sometimes called spectral matrix clustering, or in general Spectral Clustering, recently emerged as an effective method for data clustering, image segmentation, Web ranking analysis, and dimension reduction. At the core of spectral clustering is the Laplacian of the graph adjacency (pairwise similarity) matrix, evolved from spectral graph partitioning. We begin by performing an eigenvector decomposition of our recurrence matrix $(D-W)v = \lambda Dv$, where $D_{ij} = d(i, j)$ is the recurrence matrix, $W_{ii} = \sum_j D_{ij}$ is the diagonal affinity matrix, λ are the eigenvalues of the system, and v are the eigenvectors of the system. For clustering purposes the first eigenvector is usually used and each value of the eigenvector is assigned to one of two signal groups by setting up appropriate threshold or decision boundaries. The eigenvector represents the main "direction" or pattern of behavior in time, according to which the similarity matrix is oriented. This vector was compared to mean FR profiles produced by the listeners.

The signal predictability was evaluated using the same cepstral feature vector sequences. The predictability was measured in terms of Information Rate (IR), a measure that represents the reduction of uncertainty that an information-processing system achieves when predicting future values of a stochastic process based on its past. Information Rate (IR) is defined as the difference between the information contained in the variables $x_1, x_2, ..., x_n$ and $x_1, x_2, ..., x_{n-1}$, i.e. the additional amount of information that is added when one more sample of the process is observed

$$\rho(x_1, x_2, ..., x_n) = \frac{1}{n}\{I(x_1, x_2, ..., x_n) - I(x_1, x_2, ..., x_{n-1})\}.$$

It can be shown that for large n, IR equals the difference between the marginal entropy and entropy rate of the signal x(t),

$$\rho(x) = \lim_{n\to\infty} \rho(x_1, ..., x_n) = H(x) - H_r(x).$$

In our experiments we have applied the IR analysis to a sequence of cepstral vectors that describe the evolution of the spectral envelope over time. Assuming independence of the coefficients *s* after an appropriate transformation, one can show that a generalization of the IR for sequences of vectors becomes

$$\rho_{IC}^{n}(X_1, X_2, ..., X_L) \triangleq I(X_1, X_2, ..., X_L) - \{I(X_1, X_2, ..., X_{L-1}) + I(X_L)\} = \sum_{i=1}^{n} \rho(s_i(i), ..., s_i(L)$$

Using a decorrelation procedure, the sequence of feature vectors is transformed into an alternative representation where it can be regarded as a sum of approximately independent, time-varying expansion coefficients in an appropriate feature basis. IR of a vector process may then be computed from the sum

of the IR's of the individual components, as will be described below.
An additional signal feature that was employed for the estimation of EF was signal Energy (E). Both IR and E were compared separately to mean listener EF profiles. Moreover, a combined estimate of the two features was obtained using non-negative least squares regression over one-minute-long time segments. The weights of the regression, being positive values, might be considered as an indication of the relative importance of IR and E for EF estimation.

Experimental Results

Figure 2 shows the mean profiles of the listener Familiarity Rating (FR) and the Emotional Force (EF) responses, aligned with the formal structural scheme of the composition.

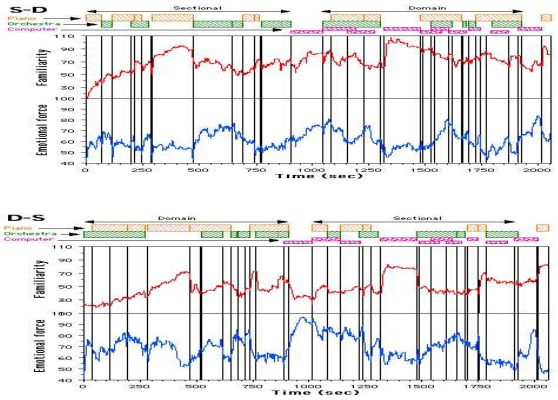
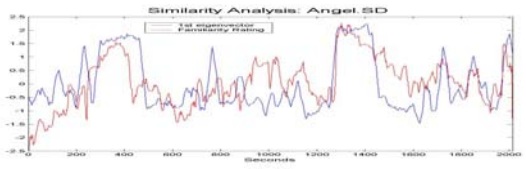


Figure 1
Average Familiarity Rating and the Emotional Force responses, aligned with the formal structural scheme of the composition of the S-D and D-S versions.

We compared the eigenvectors of the S-D and D-S versions of the piece to their corresponding FR profiles, as presented graphically in Figure 4. The y axis corresponds to normalized (zero mean and unit variance) values of the Familiarity Rating profile and the Similarity eigenvector.



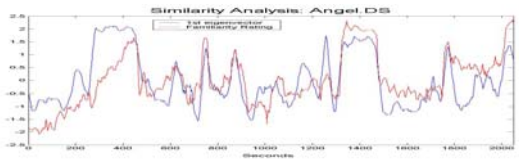


Figure 2
Familiarity Rating profiles and similarity eigenvector for the S-D and D-S versions of the piece.

The correlations between the similarity eigenvectors and FR for the S-D and D-S versions of the piece are summarized in Table 1.

Music Familiarity	**Similarity Eigenvector correlation**
S-D	0.54
D-S	0.65

Table 1: Correlation between Similarity Matrix Eigenvector (normalized version) and experimental Familiarity Ratings by human listeners.

The similarity eigenvector explains 29% (R^2) of the variance in the mean FR profile for the S-D version and 42% in the D-S version. These correlations are highly significant, but also demonstrate that other factors are involved in the familiarity ratings.

EF estimation using combined Energy and IR

In order to better approximate the EF from signal analysis, we have performed a least-squares fit of E and IR curves to the EF profile. In the following, we shall denote E and IR as predictors, in order not to confuse them with the cepstral *features* that are derived directly from the audio signal. E and IR predictors could be considered as higher-order features needed for the higher-level processing involved with emotional responses.
Using a combination of predictors for estimation of EF, the predictor weights might change slowly over time, depending on various factors related possibly both to the nature of the signal or to the listening process. A tradeoff exists when considering a time-varying regression: one should note that in principle a perfect fit is possible if the weight coefficients vary every sample. On the other hand, we cannot expect to have a single constant set of weights over the whole duration of the signal. As a reasonable compromise we have chosen a block of 1 minute in duration as the regression period over which the weight coefficients would remain constant.

Additionally, we should require a non-negative contribution of the predictors to the total EF response. This decision is justified by the claim that the various factors can contribute positively to the emotional response but they cannot cancel each other out or inhibit the total EF response. Accordingly, we employed a Non-Negative Least Squares (NNLS) regression for estimation of the EF match from E and IR. The NNLS algorithm was first introduced by (Lawson and Hanson 1974). NNLS solves the algebraic equations of the Least Squares problem subject to the added constraint that the fitting parameters contain no negative elements.

Figure 7 shows the results of NNLS regression of E and IR so as to match EF in a time-varying manner with regression weights varying every minute. The correlations between the NNLS fit of IR and E and EF are summarized in Table 3. Note that the gain in predictability obtained with the multiple correlation is quite large, resulting in 79% and 69% of the explained variance for the two versions.

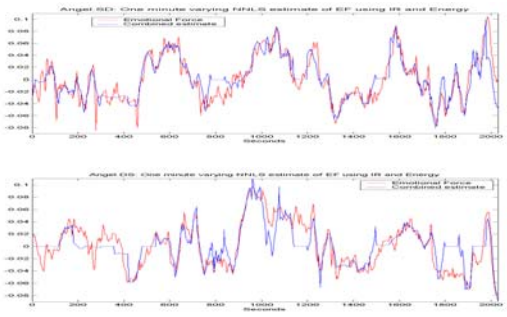


Figure 7
NNLS regression of IR and E onto Emotional Force profiles for the S-D and D-S versions of the piece.

Emotional Force	NNLS fit of IR and E
SD	**0.89**
DS	**0.83**

Table 3:
Correlations between the NNLS fit of IR and E and EF for the S-D and D-S versions.

References

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*, John Wiley & Sons, New-York.

Dubnov, S. (2003). Non-Gaussian Source-Filter and Independent Components Generalizations of Spectral Flatness Measure, *Proceedings of International Conference on Independent Components Analysis (ICA2003)*, Nara, Japan.

Frederickson, W. E. (1995). A comparison of perceived Musical tension and æsthetic response, *Psychology of Music*, 23:81-87.

Hayes, M. (1996) Statistical Signal Processing and Modeling, Wiley.

Jayant, N.S. and Noll, P. (1984). Digital Coding of Waveforms, Prentice-Hall Signal.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology, Canadian Journal of Experimental Psychology, 51:336-352.

Lawson, C. L. & Hanson, B. J. (1974), Solving Least Squares Problems, Prentice-Hall (Englewood Cliffs, NJ).

McAdams, S., Smith, B. K., Vieillard, S., Bigand, E., and Reynolds, R., (2002) Real-time perception of a contemporary musical work in a live concert setting, Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney, edited by C. Stevens, D. Burnham, G. McPherson et al. (Causal Productions, Adelaide [CD-ROM], Sydney).

McAdams, S., Vieillard, S., Vines, B., Smith, B., Bigand, E. & Reynolds, R. (submitted) Real-time perception of a contemporary piece in a live concert setting.

Oppenheim, A.V. and Schafer, R.W. (1989). Discrete-Time Signal Processing. Prentice-Hall. Englewood Cliffs.

Reynolds, R., (2002). Compositional strategies in The Angel of Death for piano, chamber orchestra and computer processed sound, Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney, edited by C. Stevens, D. Burnham, G. McPherson et al. (Causal Productions, Adelaide [CD-ROM], Sydney).

Reynolds, R. (submitted). Compositional strategies in The Angel of Death for piano, chamber orchestra and computer processed sound.

Schubert, E. (1996). Continuous response to music using a two-dimensional emotion space, Proceedings of the 4th International Conference on Music Perception and Cognition, Montreal, pp. 263 268.

Shi, J. and Malik J. (2000). Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888–905.

Sloboda, J. A. & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude, Music Perception, 19:87-120,.