David Rector
COSC 174
Project

Local Musical Features and their Affect

Milestone Report

I. Summary

The proposed project has aimed to ambitiously delve into the "mushy" field of the relationship between music and emotional affect by first, and critically, collecting data the experimenter *trusts* as an accurate representation of his own physiological response to a piece of music, and then applying to the data an input-output Hidden Markov Model (IOHMM) with local features derived from spectral and music theory analysis and latent variable chains derived primarily from music theory concepts. A milestone goal of completing data collection was established; as expected, data collection – and more importantly, perfecting the methods of data collection - has been a substantial hurdle, but a worthy one that has indeed yielded data the experimenter trusts. This writeup will first outline the steps taken thus far to achieve collection of accurate and meaningful data, and then address the issues and steps that remain.

II. Methodology & achievements thus far

Listening set change

Because the training data will necessarily be limited, it is critical that the stimulus contain as few degrees of freedom as possible. The listening set was thus changed from the Bach chorales to Bach's "Goldberg Variations" - chosen for four reasons: a) as a theme and variations, each of the 32 component pieces is very similar structurally and harmonically, reducing the possibility that significant key/harmony changes cause an observed physiological response; b) the work was written for the harpsichord, an instrument incapable of dynamic variability; thus broad dynamic changes can also be eliminated as a degree of freedom; c) because of the similar motivic content in the works, it is difficult to develop, despite repeated listening, accurate near-term expectations of future musical events, somewhat preserving expectation/surprise effects over the course of multiple listenings and allowing data over these listenings to be reasonably compared; d) despite the described similarities and superficial homogeneity, the work is celebrated as one of Bach's most profoundly nuanced and affective works.

The nuance and affect, then, can be assumed to be located primarily in the sequences of notes (as opposed to in instrumentation/loudness/spectral shifts, as in orchestral and electronic music, or in lyrical content, as in pop music). These sequences of notes can be entirely represented as MIDI data; thus, simple MIDI representations, or simple numerical analyses of MIDI data, become suitable, low-cost feature vectors, and, even better, can be used in the design of simple latent-variable chains in the IOHMM to model the musical context in which each "listening slice" takes place.

Creation of the listening set

To even further eliminate degrees of freedom, the researcher decided to listen not to human performances of the Goldberg Variations, but to MIDI-generated performances played on a sampled

piano. MIDI performances were obtained from the internet, processed minimally to eliminate possible sources of variability (key velocity was standardized, then randomized within a 5-10% window to preserve a minimal degree of human verisimilitude) and fed into a high quality, 6 GB sampled piano. (Critically, slight reverb was added, as a "dry" piano recording was perceived as less pleasant than a "wet" one – the cause of this additional pleasantness cannot be modeled through MIDI, but can through spectral analyses of the resulting audio file; it is for this reason that the feature vector will retain spectral components instead of pure MIDI data.[1])

The spectrum of the audio file was gently lowered in the 100-500 Hz range with linear equalization and slight linear multi-band compression ("linear" so as to avoid harmonic distortion caused by non-linear EQ's), as unpleasant resonances seemed to accumulate there (due to the imperfect recording of the sampled piano) that would, left alone, have a fatiguing effect upon prolonged listening. Lastly, ever so slight full-range compression and limiting were applied to reduce the "jabbing" of the ear through repeated excessively percussive piano keystrokes; this is a common technique in mastering of piano recordings.

In sum, this audio processing was undertaken so as to permit the listener to uniformly enjoy the listening experience over each ~70 minute data collection session, rather than become increasingly negatively aroused not because of musical stimulation but because of ear fatigue – with this processing, the 70th minute should, in theory, be comparable to the tenth minute.

Sensor assembly

The researcher acquired and assembled as large and as diverse an array of sensors as would accurately report physiological data *without* impinging upon the listening experience. The final array: a Neurosky MindWave EEG Headset, a light-based pulse meter that outputs data akin to that of the clip-on-the-finger Heart Rate Variability sensors used ubiquitously in clinical settings; and a home-made skin conductance sensor. Regarding the skin conductance sensor: three different circuits were assembled and compared; the researcher found that the simplest one – consisting of just a resistor to put the signal in the proper range and a capacitor to filter out high frequency noise, performed with the best signal-to-noise ratio (i.e. provided a relatively smooth signal that still responded quickly and profoundly to a test stressor – pinching the skin). Skin conductance permitted very effective and nuanced measurement of emotional arousal *when the leads were attached securely and yet such that the skin could "breathe,"* allowing moisture to quickly dissipate and skin conductance to return to baseline. A breathe-able wrap designed for use under athletic tape performed well.

The researcher also experimented with including a camera that could measure pupil dilation, but found this to interfere excessively with the listening experience. (Furthermore; as both pupil dilation and skin moisture content are regulated by the sympathetic nervous system, the two measures have been found to have high covariance; thus, discarding this measure should not discard much useful data.)

The EEG headset sent its data wirelessly to a USB port; the researcher modified its settings so as to receive a raw EEG stream with a sample rate of approximately 250 Hz. The pulse meter and skin conductance sensor were routed to an Arduino Uno, which projected its measurements to a Java program adapted by the researcher to consolidate all three signals with occasional corresponding timestamps to a CSV file.

---

1 Interesting future work would involve comparing biometric data gathered while listening to dry and wet recordings.

Data collection & interpretation

The researcher has completed several sessions of data acquisition – listening to the MIDI-generated version of the Goldberg Variations while collecting data from the aforementioned sensors. Several measures speculated or confirmed to be associated with emotional arousal or valence can be derived from the sensors: heart rate variability, EEG alpha band frequency fluctuation, and the rate of change of skin conductance. This data processing will be the next step performed.

The purpose of multiple listening sessions is primarily to ensure repeated data is weighted highly. A critical question entering the experiment was the degree to which data would be repeated – would a skin conductance measurement taken over the course of several listenings to the same music be similar in its large-scale and/or small-scale contours? The answer seems to be, mostly, yes, particularly with small-scale contours – though with some time variation that requires analysis. A few more data collection sessions will aid in delineating valid data from outliers.

However, the variance of the collected instances of observed biometric data at each point in the music may also prove an interesting observation – low variance may be an indicator of consistent arousal; high variance may indicate consistently waning attention at that point in the music. This will require some analysis.

III. Issues & next steps

The first issue requiring resolution after the biometric data is processed is to reduce its dimensionality. An eigenvalue method - either principal component analysis or singular value decomposition, whichever performs best - will be used to achieve a synthesis of the observed data. Since the training set will be small, reduction to a single scalar value representing simple emotional arousal may be the best course – i.e., the direction of greatest variance after dimensionality direction. However, a 2-D arousal-valence space would be the ideal representation of emotion; at a future time, given more time and data, this may be attempted.

Varying the time slice size will be an interesting means of deducing the broad- and small-scale contours of music-driven affect. One possibility the researcher has considered is whether combining analyses using standard Hidden Markov Models using many different orders of time-slicings – 1 s, 5 s, 30 s, 5 min, etc. - may reduce the error rate to a comparable extent as introducing latent variable chains. In other words, the standard HMM using a time slicing of 1 s is a poor because it assumes the listener's predicted state will only be dependent on his/her state one second earlier, combined with transition probabilities based upon the current time slice's feature vector - plainly inadequate to capturing the complexities of human emotion. But if we synthesize the one-second HMM prediction with many predictions accumulated by using longer and longer time slices, weighted in some combination (exponential decay?), perhaps the model *can* accurately model complex sequential data, given enough training instances.

The IOHMM will nevertheless be the default model, and the means of modeling the “context” of each listening slice will be through the development of multiple latent variable chains. Simplified, these latent variables will be, in effect, the contribution of past musical features, reduced through music theory analysis. Multiple possibilities – e.g. harmonic distance from the tonic chord, relationship to the motivic content of the theme (via Euclidean distance between a combination of features, or some already-established scalar metric), etc. will be attempted and the error rate of predictions (through cross-validation) will be measured; the best-performing latent variable chains will be combined for the final analysis.