The *Dynamics* of Listening: Time Series Regression of Bodily Responses with Performance Features

David Sears
MUMT 621
21 December 2009

Introduction

In considering the effect of music on the emotional experiences of listeners, two methodological obstacles have impeded progress in emotion research for decades: 1) the ability to provide a potentially objective measure for the study of emotions felt by listeners, and 2) an analytical technique that can account for how emotions change dynamically over time. Indeed, the lack of agreement among psychologists even about how to define emotion underscores the extraordinary methodological difficulties faced by those researchers involved in operationalizing and empirically testing the experience of emotions. A number of scholars studying music's effects on emotion have sidestepped this problem completely by suggesting that phenomenal experiences like music simply don't arouse emotions in listeners at all; instead, listeners merely recognize the emotions being expressed by the music. Known as the cognitivist position, this argument is at least partially the result of the prevalence in emotion research of subjective self-reports like adjective checklists, rating scales, and questionnaires (Sloboda & Juslin, 2001), none of which can provide any objective evidence for the position that listeners feel emotion (known as the emotivist position).

In an effort to provide such an objective measure, psychologists over the past century have frequently appealed to the sudden changes of psychophysiological mechanisms such as heart rate or respiration as necessary components of felt emotion.

Because these mechanisms are mediated by the sympathetic nervous system, a branch of the autonomic nervous system responsible for reacting to sudden changes in the external environment, psychologists as early as the 1880's attempted to link measurable bodily responses with their concomitant emotions, even suggesting that bodily responses cause emotions, and not the other way around.[1] Although scholars still disagree as to the actual role psychophysiological responses play in the experience of emotion, the application of these measures in emotion research remains a widespread method for inferring an objective emotional response in listeners.

While the difficulties relating to the objective measurement of emotions felt by listeners have a rich and well-documented history, the development of approaches to measure how emotions change dynamically over time, as well as to generate theories as to how these changes relate to a temporal phenomenal experience like music, has received little analytical treatment in the academic community until very recently. In the last twenty years, music scholars studying musical adopted statistical approaches related to forecasting in economics and meteorology as a way to correlate changes in musical/acoustic features with subjective measurements of tension and arousal. To this point, however, none of the studies adopting a time series analysis approach considered objective measures of emotion.

[1] Known as the James-Lange theory of emotion, William James claimed that physiological responses produce emotions (Sloboda & Juslin, 2001)

In the following study I attempt to fill the gap between objective measurement techniques and a time series analysis approach by applying a time series autoregressive model to the averaged psychophysiological responses of listeners as they listen to Romantic piano music. Using a time series model, I predict changes in psychophysiological responses by modeling changes in expressive performance features (dynamics, tempo variation) extracted from the acoustic stimuli.

Background

Galvanic skin response (GSR), an electrodermal measure of psychologically-induced sweat gland activity, indexes a number of processes: namely activation, attention, and the affective intensity of the stimulus (Dawson *et al*, 2007). Indeed, one author has noted a linear correlation between increasing arousal and increasing electrodermal activity (EDA), which suggests EDA is a more pure measure of emotional intensity than any other psychophysiological measure (Rickard, 2004). Psychophysiologists currently hold that GSRs represent "orienting responses" (ORs)—nearly automatic, defensive responses caused by a failure to predict change(s) in an external stimulus. ORs measured by GSR reflect the primary nature of emotions as action dispositions that are mediated by sympathetic activity preparing the organism for fight, flight, and other appropriate appetitive and defensive behaviors (Bradley & Lang, 2007).

GSR is measured by passing a small current through a pair of electrodes placed on the surface of the medial or distal phalanges of the fingers. Upon perceiving audible or visible change(s) in the external environment, the eccrine (palmar) sweat glands increase sweat secretion, which is theorized to promote grasping behavior. The electrodes measure this increase in sweat gland activity as a decrease in electrical resistance, or as an increase in electrical conductivity. Of all the psychophysiological measures, skin conductance responses are the most reliable at indicating an objective change in the arousal of the listening individual (Rickard, 2004). It should therefore be conceivable to model changes in the musical stimuli in order to predict changes in the GSR profiles.

Very few studies have actually considered analyzing GSR data over the period of an entire excerpt as a time series, instead preferring to use GSRs to record transient emotional events such as chills (Craig, 2005; Guhn *et al*, 2007, Rickard, 2004). Gomez & Danuser (2007) sought to determine the extent to which the relationships between musical features and experienced emotions correspond, using a number of physiological measures. They noted a linear correlation between tempo, accentuation, and rhythmic articulation with tonic arousal levels. They also reported a relationship between increases in sound intensity and increases in tonic arousal levels. However, they did not consider how musical features or physiological responses changed within excerpts, instead comparing statistical differences in analyzed musical features and physiological measures across multiple excerpts varying in pleasantness and arousal.

One possible reason for the dearth of time series studies of electrodermal activity lies in a fundamental disagreement as to whether the sympathetic nervous system attenuates arousal levels. Figure 1 depicts the principal components of electrodermal activity. While the response latency, rise time, and amplitude of the response are all considered primary components of electrodermal activity, physiologists disagree as to the importance of half-recovery time. Edelberg (1972) has argued that variations in electrodermal activity are not the result of an on-off switch in the sympathetic nervous system, whereby the sympathetic nervous system activates an increase in sweat gland activity as a response to a change in an auditory stimulus, after which it simply deactivates, resulting in a gradual decrease in arousal levels; the body therefore doesn't control the rate or amount of decrease in arousal levels. Instead, Edelberg has suggested that both increases and *decreases* in autonomic arousal levels are regulated by the sympathetic nervous system. Results from a time series model of galvanic skin response should therefore provide evidence as to whether or not decreases in autonomic arousal (from the maximum to the minimum of a waveform) reflect a response to change (or a lack thereof) in the auditory stimulus. Positive results will suggest that such decreases provide evidence of physiological regulation, while negative results will provide evidence that the body does not control the rate or amount of decrease in arousal levels.

Although no study to date has considered a time series approach to physiological measurements within a musical excerpt, a number of studies have considered continuous behavioral measurements of tension, resemblance or emotional force, in which subjects move a slider in response to their own experience (Nielsen, 1987; McAdams *et al*, 2004; Schubert, 2001, 2004). Thimm & Fischer (2003) extracted psychoacoustical features such as loudness and roughness from a number of musical stimuli, and compared each of these features independently with the real-time continuous behavioral ratings of arousal and valence provided by participants. Because both physiological and behavioral responses occur at a variable period of time after the onset of the musical features eliciting the reaction, the authors chose to compare the features to the behavioral ratings using cross-correlation to determine the appropriate time it takes to respond (referred to as *lag*). By using a cross-correlation technique, however, Thimm & Fischer could not provide an estimate for the relative contribution of all of the multiple features to the arousal and valence ratings within a single excerpt.

Farbood (2006) proposed a quantitative parametric model of musical tension, in which she regressed tension profiles using musical features like harmonic tension and melodic expectation as predictors.[2] Her multiple regression method indicated the relative salience of each measure over the entirety of each excerpt, but it could not

[2] Devised by Fred Lerdahl and Elizabeth Margulis, respectively, both harmonic Tension and melodic expectation attempt to quantify changes in perceived tension or expectation based on music-theoretical rules concerning tonality (i.e., effects of modulation and chromaticism on perceived tension). See Lerdahl (1988) & Margulis (2005) for more details.

account for effects of serial correlation among all of the variables, one of the most serious difficulties time series analysis attempts to address. Emery Schubert (1999, 2001, 2002, 2004) was the first scholar to apply time series methods to music in order to rectify the problems associated with the regression modeling of continuous data.

Known as ordinary least squares (OLS) linear regression, a common statistical approach for determining the relationship between a dependent variable (behavioral or physiological data) and a number of independent variables (musical features) is to find a straight line (or surface if more than one musical feature) that best predicts the relationship. The line or surface approximates a hypothesized true relationship between musical features and the emotional response and is positioned so as to minimize the variability between the predicted relationship and the actual data points (Schubert, 1999). OLS regression can be written as:

$$GSR = b_0 + b_1 \times MF_1 + b_2 \times MF_2 + \ldots + e_t$$

Each b is a coefficient that measures the amount of increase or decrease in GSR for a one-unit increase in the musical feature. Each regression coefficient is therefore determined in such a way as to best explain variability in GSR. In a successful regression model, the error term e_t will be small and will vary randomly. If the error term is too large and/or it doesn't vary randomly, the musical features implemented in the model don't adequately reflect changes in GSR, or there aren't enough musical features to represent the variations present in the musical stimulus. Unfortunately, these properties of the error term are often the case in time series analysis, a problem often exacerbated by the presence of serial correlation.

Perhaps the greatest difficulty in applying OLS regression to musical stimuli is that it is necessary that every event in the sample is independent (or context-free): each individual event can be compared with musical features at the same moment without any regard for the events preceding the event under examination. Because a time series is a set of data points that are ordered in time, the introduction of time necessarily violates the assumption of independence in the OLS regression method. Musical context is in fact music theory's equivalent to *serial correlation* (or *autocorrelation*). Statistical techniques like OLS regression attempt to remove effects of serial correlation from experimental design in order to determine the true relationship between the dependent variable and the independent variable(s). If left unexamined, a regression model can falsely attribute a greater relationship between these variables, when in fact it is also being affected by past values of itself.

Figure 2 provides a typical example of the degree of serial correlation in a musical feature such as loudness. Serial correlation can be assessed with an autocorrelation function (ACF), a measure of the degree of correlation when a musical feature is compared against itself moved forward in time at a pre-determined number of lags (1 lag=.5s). In this example, the first 3 lags (1.5s) are significantly serially correlated. Schubert explains that serial correlation may be thought of as memory.

When we reach a loud moment in the music, we won't simply respond to the absolute value of the event at that time, but to the degree of change between that event and the preceding events in time. Changes in musical features and emotional responses generally occur very gradually, providing a context in which listeners (and our model) predict(s) a specifically loud moment, not in terms of other features in the music, but in terms of the preceding few seconds (e.g., a crescendo). Autocorrelation acknowledges the effect of context, but OLS linear regression does not "know" this, and assumes that the high correlation of the features with the response is entirely related to the predictive power of the musical features, rather than to the preceding context (Schubert, 1999).

Given that serial correlation is a necessary effect of considering a temporal stimulus like music, Schubert employed two techniques for minimizing serial correlation without dismissing a time series approach altogether[3]. The first method applies a first order difference transformation to the time series of both the musical features and the emotion response (in our case GSR), where the difference between the current value of a series and the immediately preceding value is used to produce the current value of the series. The series 15, 8, 22, 16, for example, produces a difference series of -7, 14, -6. A time series like music that can be understood, at least in part, as an additive process (adding previous values of itself) is described as a first order

[3] Many scholars in music research often resolve the problem of serial correlation by simply extracting the events under examination from the music and submitting them to statistical tests without regard for the context in which those events took place. Schubert explains that this method isn't ecologically valid.

integrated process. By removing the integrated component of a time series through differencing, we end up with a series that fluctuates more randomly than the original series, thereby removing some of the effects of serial correlation (Schubert, 1999). The other significant advantage in comparing the differenced time series over the integrated time series is that we are comparing the event to event change in each of the variables, rather than in the absolute values themselves.

The second approach to removing serial correlation is an autoregressive adjustment. When the error term of an OLS regression model doesn't vary randomly, either the musical features haven't adequately reflected the musical stimulus, or the error term is serially correlated. If differencing the variables still hasn't removed the effects of serial correlation in the error term, a common approach is to treat the error term as having two components—a true error component and a serially correlated component. A first order autoregressive model writes the error term as:

$$e_t = a_1 e_{t-1} + v_t$$

where e_{t-1} is the error term from the previous point in time and v_t is a true random variable, as e_t was in the initial OLS regression model. The coefficient a_1 indicates the proportion of the previous error term that is carried forward in the next point in time. In effect, the a_1 coefficient indicates a type of serial correlation process known as first order autoregressive.

To account for effects of serial correlation in the musical features, the GSR profiles, and the error component of the model, I have applied a first order difference transformation to all of the variables and a first order autoregressive adjustment to the OLS multiple regression model. To best predict the GSR signal, I have modeled changes in expressive performance features (dynamics and tempo variation). Clarke (1999) has suggested that these features are important for conveying structural and emotional information to the listener, and several researchers have attempted to model the rule systems that govern expressive performance parameters (Todd, 1985; Palmer, 1996). Gomez and Danuser (2007) suggested that tempo and rhythmic articulation are the features most strongly correlated with psychophysiological measures, and a number of studies have noted a correlation between stimulus intensity and skin conductance response (Turpin & Siddle, 1979; Guhn *et al*, 2007).

Method

Participants

20 participants (10 male) with more than 8 years of musical training participated in the experiment. The average age of the participants was 22 $\pm$ 2.5 years and the average number of years of musical training was 11 $\pm$ 2.6. The participants were screened for at least 8 years of musical training, no hearing loss, no history of any emotional or anxiety disorder, and finally a maximum of 4 years of piano training, in order to minimize the potential effects of familiarity on their physiological responses.

Stimuli

To ensure that emotional responses not only varied within pieces, but just as crucially, between pieces, a pre-study was conducted to determine the degree of variation between pieces using the Russell valence/arousal space as a theoretical foundation. Russell's (1980) space is an example of a dimensional approach to the study of emotion, in which emotions are characterized by two orthogonal dimensions: pleasantness (valence), and arousal[4]. The dimensional approach provides a method by which to quantify differences between stimuli using behavioral responses. In the pre-study, 30 musicians (15 males), screened with the same conditions as the present experiment, listened to 40 Romantic piano excerpts that exhibited a small ternary structure (either ABA or AABAA) and lasted between 50 and 90 s. While listening, the participants indicated their valence of emotion and excitation/arousal on a seven point Likert-scale. Participants' responses for each scale were averaged over the four emotion quadrants and then the arousal and valence measures for each piece were analyzed using a k-means clustering solution. Shown in figure 3, 19 excerpts were finally chosen for the present experiment to optimally represent the four emotion quadrants and to minimize familiarity of the excerpts within each quadrant. For the time series analysis, 1 excerpt was selected from each of the 4 quadrants of the emotion space.

[4] This is not to suggest that researchers haven't considered a greater number of axes in the study of emotion. The first dimensional approach, adopted by Spencer in 1890, employed 3 dimensions: valence, arousal, and potency. (Sloboda & Juslin, 2001)

406 + valence, + arousal
108 + valence, - arousal
402 - valence, + arousal
304 - valence, - arousal

See table 1 for details about these stimuli. For details about the experimental procedures used in acquiring the physiological data, see Mattson (2009).

In order to maintain the ecological validity of the stimuli, recordings were taken from acoustic performances (rather than using mechanical midi performances). Although such a decision reflects a desire to study the effects of "real" music on listeners, it also poses the significant challenge of extracting performance features from the acoustic signal. In order to derive loudness information directly from the audio file, I employed the Moore & Glasberg (1997) model, an auditory filter model of the loudness of steady state sounds, measured in sones. As a psychoacoustic model, it attempts to account for the perceptual constraints of the human auditory system. Cabrera (1999) created a program implemented in *Matlab* that applies the Moore & Glasberg model to 16-bit, 44.1kHz audio files. It analyzes the files in a succession of overlapping 93 ms windows and records the results as time series data.

To obtain tempo variation information, I employed Simon Dixon's BeatRoot software (2001), which provides a means of manually extracting the inter-onset-interval (IOI) at an experimenter-selected metrical level. Although his beat extraction algorithms are generally very reliable at automatically extracting tempo information (91% accuracy across multiple styles), I manually extracted the IOI between beats, using his software to facilitate data acquisition and analysis. The equation to compute tempo as beats per minute (bpm) as a time series can be written as:

$$bpm = (60*b_i)/(IOI_t - IOI_{t-1})$$

where b_i represents the number of beats in a bar (for ex., common time would receive a score of 4), and $IOI_t - IOI_{t-1}$ refers to the IOI subtracted from the IOI that directly preceded it.

Analysis

The task of adequately comparing time series of independent features (GSR, loudness in sones, tempo in bpm) is admittedly complex, since each feature was sampled at a different rate, and a number of different filtering methods are common to both GSR and psychoacoustic features like loudness. To further complicate matters, it's essential in a time series approach to minimize the sampling rate, as too high a sampling rate results in a spuriously high p value.[5] In order to minimize this problem, a minimum alpha of .01 will be necessary to report significance, but the number of samples within a stimulus must still be carefully controlled.

To narrow the range of filtering possibilities, a common filtering method for GSR was applied to loudness. GSR signals measured at a sampling rate of 256 Hz were sent through a fourth-order Butterworth low-pass filter with a cutoff frequency of 0.3 Hz. The butterworth filter is ideal in a time series analysis because it is a maximally smooth

[5] In any parametric statistical measure, the larger the N, the higher the likelihood of obtaining a p value below .05.

filter, providing the possibility to downsample GSR and loudness (originally sampled at 51Hz) without losing much information (i.e., it doesn't violate the Nyquist theorem). In the case of tempo, a filtering method wasn't necessary, since the sampling rate for each stimulus was very close to the final sampling rate selected for the final analysis.

After filtering, each of the features was resampled to 2Hz and normalized to between 0-1. The normalized GSR time series for each participant were then averaged. As a last step in the pre-processing stage, each of the features were first-order differenced to minimize effects of serial correlation. Figure 4 provides an example of the processing stages for loudness in excerpt 108.

Model

Before implementing the 1st-order Autoregressive model with 1st order differenced features, one crucial obstacle still remains. The simplest regression model regresses a single musical feature to predict changes in the emotional response (GSR). However, changes in the musical features necessarily precede the onset of the physiological response. Galvanic skin response literature refers to the difference between the onset of the external stimulus and the onset of the physiological response *response latency*. Typically, the response latency lasts between 1-4 seconds (Dawson *et al*, 2007). Unfortunately, this latency, or *lag*, in the response, is variable rather than fixed. A more sudden increase in a salient external stimulus can decrease the latency between the initial event and the physiological response, and vice versa. Physiologists have yet to find a quantitative relationship to predict the variability in the lag in physiological response.

The solution in time series analysis to the problem of lag is to move the time series for each feature forward at regular intervals for the length of time the latency window is expected to last. For example, the time series for loudness is moved forward 1 second, 2 seconds, 3 seconds, and 4 seconds, resulting in a total of 5 time series (5 variables) for loudness. The ensuing model therefore regresses 10 features against GSR, and the equation can be written as:

$$\Delta GSR = b_0 + b_1 \times \Delta Loudness_0 + b_2 \times \Delta Loudness_1 \ldots + a_1 e_{t-1} + v_t$$

where Δ indicates that the variable has been first order differenced, and $Loudness_0$ indicates that the feature has not been moved forward. The coefficient a_1 represents the autoregressive adjustment of the error term to minimize the effect of serial correlation.

There are a number of techniques for determining the order of entry of each of the predictor variables into the model. Because this analysis involves the application of 10 features (loudness, tempo, and 4 time series of each moved forward at 1, 2, 3, and 4 s), many of which will be excluded in the final model, stepwise regression was adopted because it automates the criteria for variable selection. Stepwise regression uses a predetermined criterion for selecting whether each of a list of variables should be included in, or excluded from, the final model (qtd. in Schubert, 1999). This process ultimately maximizes the goodness of fit (R^2) such that the model only includes variables that significantly contribute to the fit of the model. The following model therefore attempts to predict changes in GSR by including all of the lagged time series for loudness and tempo, and allowing the algorithm to determine which of these features will make a significant contribution to explaining variability in GSR.

Results

406 (+ valence, + arousal)

For the mean GSR response across all participants to *Drei Klavierstücke*, D 946, No. 1 in E flat minor by Franz Schubert, the 1st order Autoregressive model (AR1) was:

$$\Delta GSR = -.301 \times \Delta Loud_0 + .289 \times \Delta Loud_6$$

The coefficient for the first order autoregressive component of the model (AR1) was .892, and all of the features listed in the model were statistically significant ($p<.001$). None of the other variables reached significance, and were excluded in the model by the stepwise regression algorithm. Figure 5 presents the model statistics generated by *SPSS*.

Approximately 81% of the variance was explained by the model. This unusually good fit suggests that the model successfully explains changing GSR in terms of loudness. The negative sign for $loudness_0$ indicates that as loudness decreased, GSR increased. Conversely, as $Loudness_6$ (sampled at 2Hz, a lag of 6 = 3 seconds), increased, GSR increased 3 seconds later.

To determine if the AR(1) model successfully removed effects of serial correlation from the model, an autocorrelation of the residual (the error term) was performed. Unfortunately, the AR(1) did not remove all effects of serial correlation, as is indicated in figure 6. Although the model explains changing GSR in terms of loudness (lag 0 and 3) and a first order autoregressive process, its validity is still compromised by the presence of serial correlation.

108 (+ valence, - arousal)

For the mean GSR response across all participants to *Lyriske Stykker*, Vol. 1, Op. 12, No. 1, Arietta by Edvard Grieg, the 1st order Autoregressive model (AR1) revealed no significant results for any of the entered features. According to the model, variations in tempo and loudness had no effect in explaining the within-excerpt variance in GSR.

402 (- valence, + arousal)

For the mean GSR response across all participants to *8 Piano Pieces*, Op.76, No. 5, Capriccio in C sharp minor, by Johannes Brahms, the 1st order Autoregressive model was:

$$\Delta GSR = .260 \times \Delta Loud_6$$

The coefficient for the first order autoregressive component of the model (AR1) was .824, and $Loud_6$ was statistically significant ($p = .007$). None of the other variables reached significance, and were excluded in the model by the stepwise regression algorithm. Figure 7 presents the model statistics generated by *SPSS*.

Approximately 79.7% of the variance was explained by the model. In this model, as loudness increased, GSR increased 3 seconds later. However, like the previous models, an ACF performed on the residual of the AR(1) model revealed significant effects of serial correlation at lags 1, 3, and 4, thereby compromising the validity of the model.

304 (- valence, - arousal)

For the mean GSR response across all participants to *Etudes symphoniques*, Op. 13, Thema, by Robert Schumann, the 1st order Autoregressive model was:

$$\Delta GSR = -.139 \times \Delta Tempo_2 + {*}{-.107} \times \Delta Tempo_8 + {*}{-.137} \times \Delta Loud_4$$[6]

The coefficient for the first order autoregressive component of the model (AR1) was .792, and $Tempo_2$ was statistically significant (p=.003). Two of the other variables reached marginal significance: $Tempo_8$ (p=.019) and $Loud_4$ (p=.043). None of the other variables reached significance, and were excluded in the model by the stepwise regression algorithm. Figure 9 presents the model statistics generated by *SPSS*.

Approximately 72.2% of the variance was explained by the model. In this model, as tempo decreased, GSR increased 1 second later and 4 seconds later. However, like the previous models, an ACF performed on the residual of the AR(1) model revealed significant effects of serial correlation at lags 1, 3, and 4, thereby compromising the validity of the model.

[6] * denotes marginal significance.

Discussion

For 3 of the 4 excerpts, a 1st order Autoregressive model found significant effects of loudness and tempo on GSR, a finding which supports prior work concerning the effect of performance variables on both behavioral (Schubert, 2001; Farbood, 2006) and physiological (Gomez & Danuser, 2007) responses. However, the failure to sufficiently remove effects of serial correlation reduces the model's validity.

Consider a plot of the residual for the AR(1) model of 108, shown in figure 11. In a successful regression model, the time series produced from the residual (proportion of the variance the model could not account for) should demonstrate no serial correlation; in effect, every data point in the series should be random. In the case of the residual in 108, and for all of the excerpts, these residual time series are still somewhat serial correlated (see figures 6, 8, and 10). The most likely explanation for the continued presence of serial correlation is that the variables of loudness and tempo do not adequately represent the amount of variation within the musical excerpts that could be eliciting a change in GSR. The time series shown in figure 11 could be explained by changes in a number of other features such as harmony, melody, onset density, etc. In order to remove serial correlation, a future model needs to quantify other aspects of the music to more adequately represent the multidimensional nature of the musical stimulus.

This criticism of the model should not diminish the significant role performances variables play in physiological responses, however. Although modeling the within-excerpt variance in GSR is an extraordinarily difficult task, particularly when only considering the variables of loudness and tempo, an analysis of the between-excerpt variance in GSR reveals the significance of these two features in physiological responses. Figure 12 provides a correlation matrix of the average loudness, tempo, GSR, and heart rate (HR) across all 19 excerpts. The matrix reveals significant effects of loudness with GSR ($r=.755$, $p<.0001$) and HR ($r=.706$, $p=.001$), and tempo with GSR ($r=.622$, $p=.004$) and HR ($r=.601$, $p=.006$). These data indicate that as GSR and HR increase between pieces, both loudness and tempo increase significantly.

From between-excerpt analyses, the effect of performance features on the physiological responses of listeners is unambiguous. However, the application of time series analysis on physiological response data is less transparent, as the number of factors both intrinsic and extrinsic to the music that can account for variation in physiological responses over the course of a short period of time is extraordinarily difficult to control, quantify, and model. Although the continued presence of serial correlation in this model should provide cause for concern, it most likely points to the failure of these 2 features to adequately represent variation in the musical stimulus, rather than to the failure of time series analysis as a tool for analyzing the temporal dynamics of musical listening.

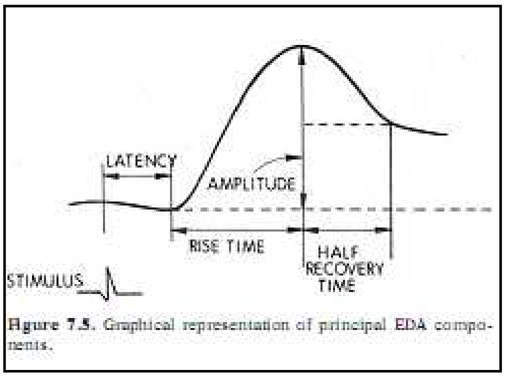
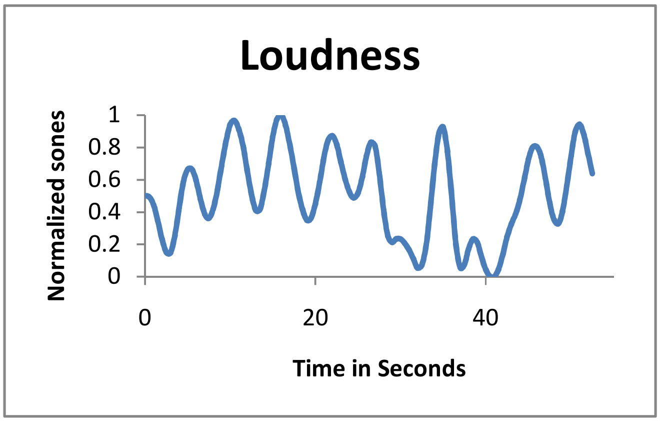


Figure 1—Dawson *et al*, 2007, pg. 165



Autocorrelations

Series:loud108

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.912	.096	90.738	1	.000
2	.682	.095	142.002	2	.000
3	.392	.095	159.078	3	.000
4	.133	.094	161.069	4	.000
5	-.027	.094	161.155	5	.000
6	-.070	.093	161.714	6	.000
7	-.019	.093	161.755	7	.000
8	.075	.093	162.413	8	.000
9	.159	.092	165.415	9	.000
10	.201	.092	170.223	10	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

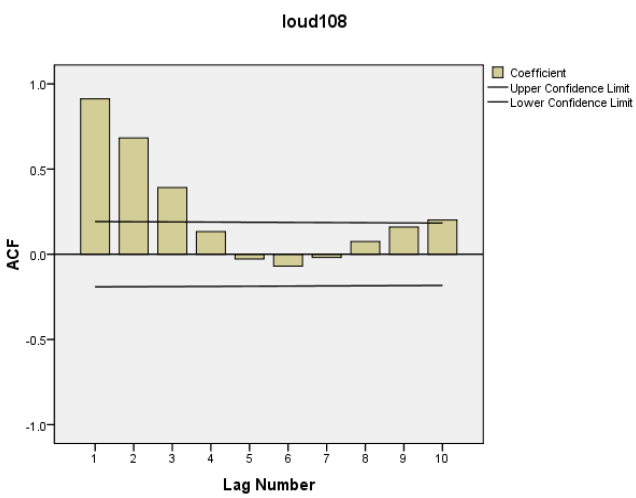


Figure 2—The top left graph is a plot of normalized loudness in sones for excerpt 108. The table to the right represents the degree of auto-correlation at each lag, where each lag=.5s. The figure in the bottom left is a plot of the auto-correlation function. In this example, the first 3 lags are significantly autocorrelated.

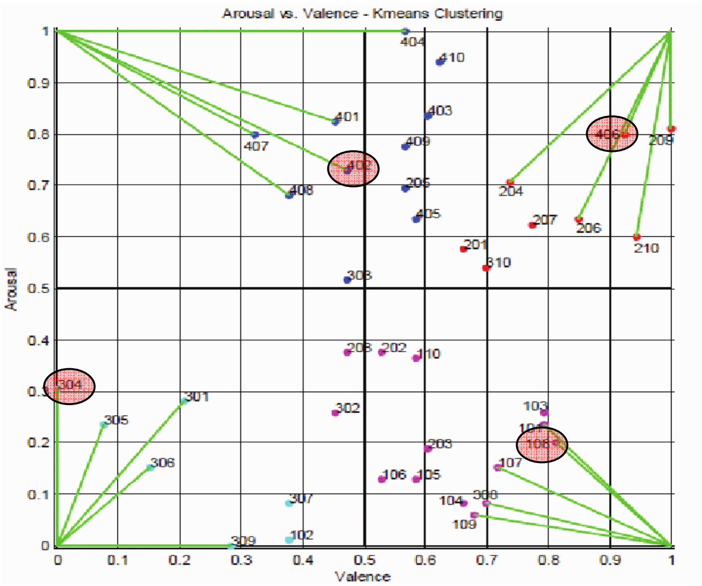


Figure 3. Kmeans clustering solution for the valence and arousal ratings of the excerpts in the pre-study. The excerpt points are color coded to indicate those that best represent each emotion quadrant as determined by the kmeans technique. The green bars indicate the center of each clustering per quadrant as determined by the kmeans technique. The red circles isolate each of the 4 stimuli examined in the present analysis.

Code	Composer	Piece & CD	Form	Duration
406 (+,+)	Franz Schubert	Drei Klavierstücke, D 946, No. 1 in E flat minor Harmonic Records, H/CD 8610 Paul Badura-Skoda, 1986	a a b a a	0:52 B: 0:11.49 A: 0:41.48
108 (+,-)	Edvard Grieg	Lyriske Stykker, Vol. 1, Op. 12, No. 1, Arietta Victoria VCD 19025 Geir Henning Braaten, 1990	a a b a a	0:52 fade out B: 0:11.75 A: 0:41.14
402 (-,+)	Johannes Brahms	8 Piano Pieces, Op.76, No. 5, Capriccio in C sharp minor, Agitato, ma non troppo presto London Records, 430 053-2 Julius Katchen, 1990	a a b a a	1:17 fade out B: 0:28.03 A: 0:53.79
304 (-,-)	Robert Schumann	Etudes symphoniques, Op. 13, Thema EMI Classics, 5 69521 2 Dimitri Alexeev, 1996	a b a	1:03 0:22-1:25 B: 0:18.72 A: 0:40.30

Table 1—The 4 excerpts used in the time series analysis.

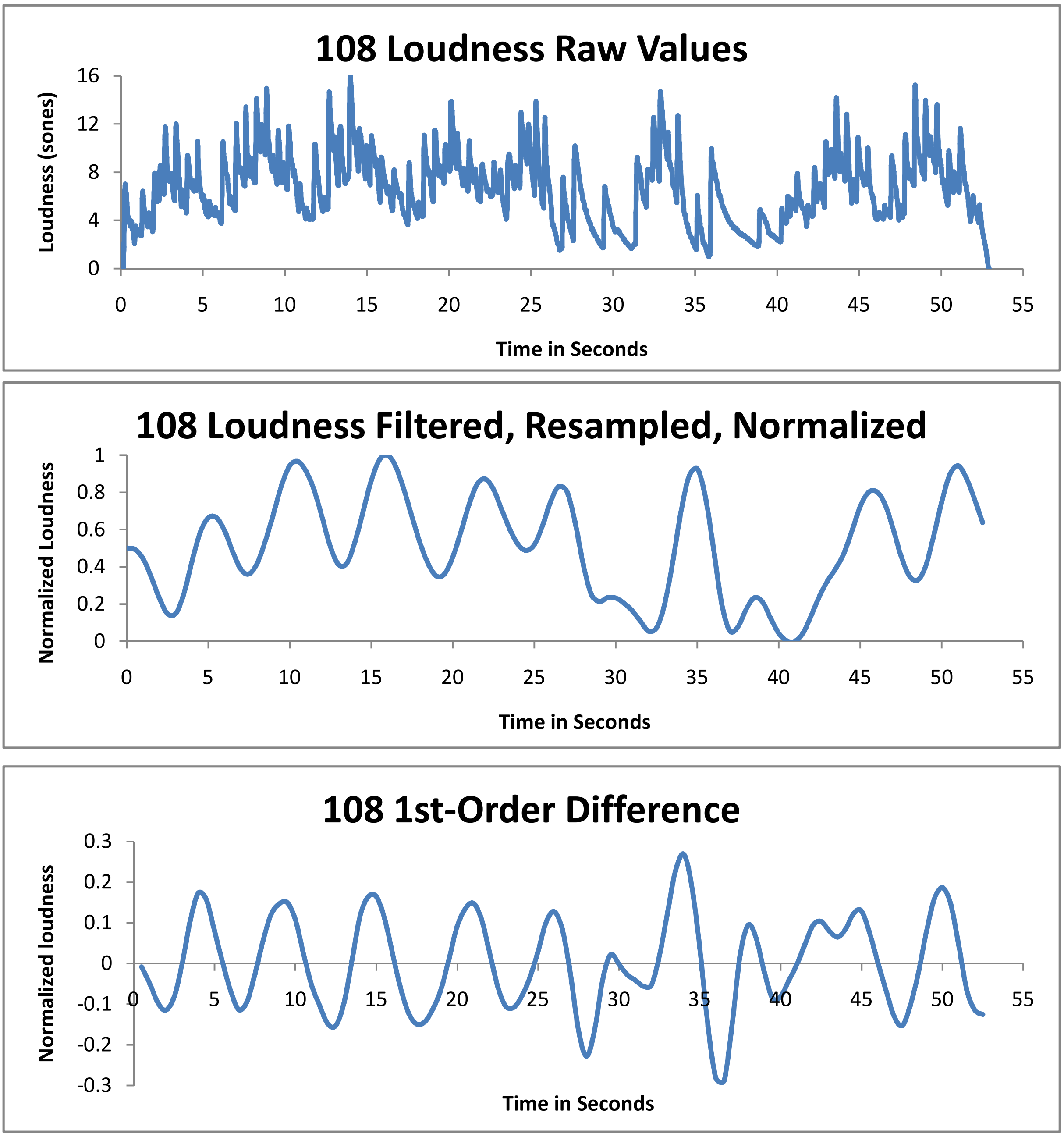


Figure 4—Comparison of loudness in raw scores with loudness with a Butterworth 4^{th} order low-pass filter, resampled from 51Hz to 2Hz, and range normalized (0-1), and loudness after applying a 1^{st}-order difference transformation.

Figure 5—406 AR(1) Model

Model Description

Model Name		406
Dependent Series		gsr
Independent Series	1	loud
	2	tempo
	3	LAGS(loud,2)
	4	LAGS(tempo,2)
	5	LAGS(loud,4)
	6	LAGS(loud,6)
	7	LAGS(loud,8)
	8	LAGS(tempo,4)
	9	LAGS(tempo,6)
	10	LAGS(tempo,8)
Constant		Included
AR		1

Applying the model specifications from MOD_16

This table indicates the order of entries determined by the stepwise regression algorithm.

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Rho (AR1)		.892	.050	17.936	.000
Regression Coefficients	loud	-.301	.088	-3.434	.001
	tempo	-.001	.018	-.059	.953
	LAGS(loud,2)	.102	.078	1.300	.197
	LAGS(tempo,2)	-.004	.022	-.189	.850
	LAGS(loud,4)	-.085	.096	-.881	.381
	LAGS(loud,6)	.289	.078	3.686	.000
	LAGS(loud,8)	-.004	.085	-.049	.961
	LAGS(tempo,4)	-.014	.024	-.599	.551
	LAGS(tempo,6)	-.012	.022	-.549	.585
	LAGS(tempo,8)	.000	.018	-.009	.993
Constant		.005	.022	.241	.810

Iteration History

	Rho (AR1)	Regression Coefficients										Constant	Adjusted Sum of Squares	Marquardt Constant
		loud	tempo	LAGS(loud,2)	LAGS(tempo,2)	LAGS(loud,4)	LAGS(loud,6)	LAGS(loud,8)	LAGS(tempo,4)	LAGS(tempo,6)	LAGS(tempo,8)			
0	.000	-.341	.053	.137	.026	-.084	.350	.018	-.054	-.076	-.041	.003	.201	.001
1	.870	-.301	.000	.103	-.004	-.084	.290	-.005	-.015	-.012	.000	.005	.048	.001
2	.891	-.301	-.001	.102	-.004	-.085	.289	-.004	-.014	-.012	.000	.005	.048[a]	.000

a. The estimation terminated at this iteration, because the sum of squares decreased by less than .001%.

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Fit for gsr from AREG, MOD 1[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: gsr

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.902[a]	.813	.811	.023975256565536

a. Predictors: (Constant), Fit for gsr from AREG, 406

The AR Model 1 was entered into an OLS linear regression model to determine the R^2, which indicates the AR 1 Model accounted for 81.1% of the variance in GSR.

Autocorrelations

Series:Residual

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	.603	.102	34.928	1	.000
2	.109	.101	36.089	2	.000
3	-.222	.101	40.922	3	.000
4	-.275	.100	48.442	4	.000
5	-.231	.100	53.817	5	.000
6	-.224	.099	58.901	6	.000
7	-.213	.099	63.543	7	.000
8	-.151	.098	65.913	8	.000
9	-.058	.098	66.272	9	.000
10	.010	.097	66.284	10	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

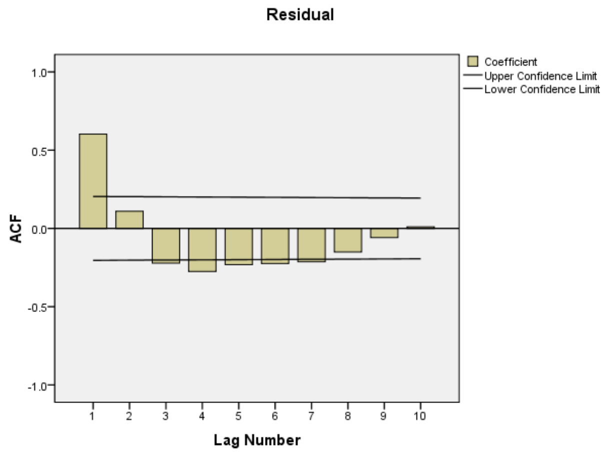


Figure 6—An ACF of the residual of the AR(1) model indicates that, though serial correlation has been dramatically reduced, effects of serial correlation are still present in the model at lag 1 and marginally at lag 4.

Figure 7—402 AR(1) Model

Model Description

Model Name		402
Dependent Series		gsr
Independent Series	1	loud
	2	tempo
	3	LAGS(loud,2)
	4	LAGS(tempo,2)
	5	LAGS(loud,4)
	6	LAGS(loud,6)
	7	LAGS(loud,8)
	8	LAGS(tempo,4)
	9	LAGS(tempo,6)
	10	LAGS(tempo,8)
Constant		Included
AR		1

The table above indicates the order of entry determined by the stepwise regression algorithm.

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Rho (AR1)		.825	.048	17.097	.000
Regression Coefficients	loud	-.106	.103	-1.033	.304
	tempo	-.027	.041	-.663	.509
	LAGS(loud,2)	-.145	.101	-1.427	.156
	LAGS(tempo,2)	-.024	.042	-.573	.568
	LAGS(loud,4)	.033	.122	.268	.789
	LAGS(loud,6)	.260	.095	2.737	.007
	LAGS(loud,8)	.233	.102	2.289	.024
	LAGS(tempo,4)	.002	.045	.040	.968
	LAGS(tempo,6)	-.054	.044	-1.224	.223
	LAGS(tempo,8)	.039	.043	.904	.368
Constant		-.003	.012	-.237	.813

Iteration History

	Rho (AR1)	Regression Coefficients										Constant	Adjusted Sum of Squares	Marquardt Constant
		loud	tempo	LAGS(loud,2)	LAGS(tempo,2)	LAGS(loud,4)	LAGS(loud,6)	LAGS(loud,8)	LAGS(tempo,4)	LAGS(tempo,6)	LAGS(tempo,8)			
0	.000	.028	-.001	-.214	-.044	.233	.197	.463	.004	-.102	.037	-.004	.240	.001
1	.790	-.096	-.027	-.141	-.025	.047	.266	.247	.003	-.055	.040	-.003	.082	.001
2	.821	-.105	-.027	-.144	-.024	.034	.261	.234	.002	-.054	.039	-.003	.082	.000
3	.824	-.106	-.027	-.145	-.024	.033	.260	.233	.002	-.054	.039	-.003	.082[a]	.000

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Fit for gsr from AREG, MOD_12[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: gsr

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.893[a]	.798	.797	.024560108059410

a. Predictors: (Constant), Fit for gsr from AREG, 402

AR 402 was entered into an OLS linear regression model to determine the R^2, which indicates the AR 1 Model accounted for 79.7% of the variance in GSR.

Autocorrelations

Series:Residual

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	.554	.084	43.260	1	.000
2	-.060	.084	43.764	2	.000
3	-.406	.084	67.403	3	.000
4	-.376	.083	87.799	4	.000
5	-.101	.083	89.280	5	.000
6	.106	.083	90.935	6	.000
7	.123	.082	93.149	7	.000
8	.027	.082	93.255	8	.000
9	-.100	.082	94.763	9	.000
10	-.164	.081	98.822	10	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

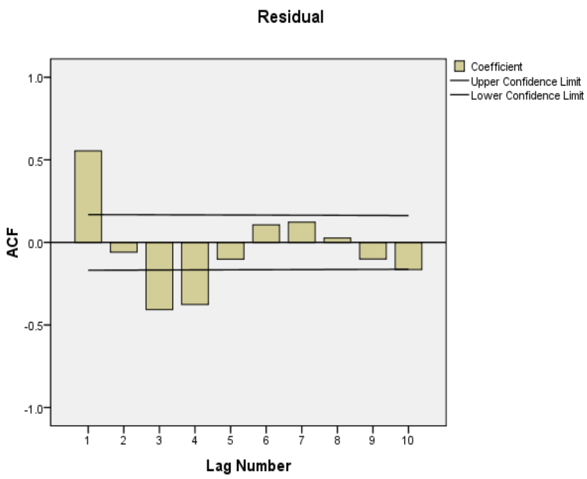


Figure 8—An ACF of the residual of the AR(1) model indicates that, though serial correlation has been dramatically reduced, effects of serial correlation are still present in the model at lag 1, 3 and 4.

Figure 9—304 AR(1) Model

Model Description

Model Name		304
Dependent Series		gsr
Independent Series	1	loud
	2	tempo
	3	LAGS(loud,2)
	4	LAGS(tempo,2)
	5	LAGS(loud,4)
	6	LAGS(loud,6)
	7	LAGS(loud,8)
	8	LAGS(tempo,4)
	9	LAGS(tempo,6)
	10	LAGS(tempo,8)
Constant		Included
AR		1

Applying the model specifications from MOD_8

The table above indicates the order of entry determined by the stepwise regression algorithm.

Parameter Estimates

		Estimates	Std Error	t	Approx Sig
Rho (AR1)		.792	.059	13.407	.000
Regression Coefficients	loud	-.051	.057	-.889	.376
	tempo	-.016	.044	-.376	.707
	LAGS(loud,2)	.060	.056	1.056	.293
	LAGS(tempo,2)	-.139	.046	-2.997	.003
	LAGS(loud,4)	-.137	.067	-2.046	.043
	LAGS(loud,6)	.074	.058	1.278	.204
	LAGS(loud,8)	.050	.059	.853	.395
	LAGS(tempo,4)	.017	.047	.365	.716
	LAGS(tempo,6)	.019	.047	.393	.695
	LAGS(tempo,8)	-.107	.045	-2.385	.019
Constant		.000	.009	-.105	.917

Iteration History

	Rho (AR1)	Regression Coefficients										Constant	Adjusted Sum of Squares	Marquardt Constant
		loud	tempo	LAGS(loud,2)	LAGS(tempo,2)	LAGS(loud,4)	LAGS(loud,6)	LAGS(loud,8)	LAGS(tempo,4)	LAGS(tempo,6)	LAGS(tempo,8)			
0	.000	-.051	.040	.091	-.166	-.123	.109	.074	-.011	-.014	-.187	.000	.105	.001
1	.777	-.051	-.016	.060	-.139	-.136	.075	.051	.016	.017	-.109	.000	.040	.001
2	.792	-.051	-.016	.060	-.139	-.137	.074	.050	.017	.019	-.107	.000	.040[a]	.000

a. The estimation terminated at this iteration, because the sum of squares decreased by less than .001%.

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Fit for gsr from AREG, 304[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: gsr

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.851[a]	.724	.722	.018812665828458

a. Predictors: (Constant), Fit for gsr from AREG, 304

AR 304 was entered into an OLS linear regression model to determine the R^2, which indicates the AR 1 Model accounted for 72.2% of the variance in GSR.

Autocorrelations

Series:Residual

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.531	.092	33.333	1	.000
2	-.080	.092	34.087	2	.000
3	-.436	.091	56.921	3	.000
4	-.423	.091	78.576	4	.000
5	-.201	.090	83.500	5	.000
6	.065	.090	84.016	6	.000
7	.222	.090	90.159	7	.000
8	.194	.089	94.874	8	.000
9	.048	.089	95.169	9	.000
10	-.104	.088	96.553	10	.000

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

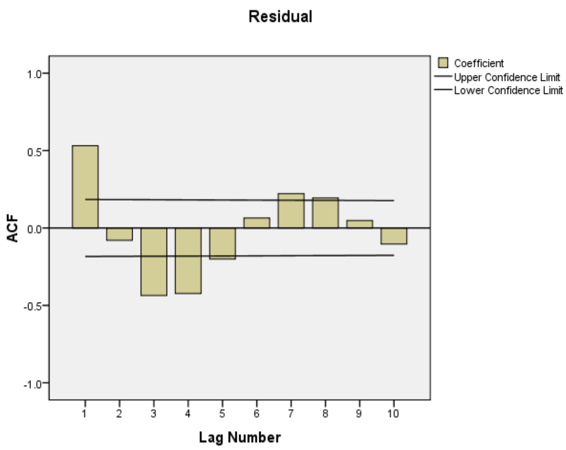


Figure 10—An ACF of the residual of the AR(1) model indicates that, though serial correlation has been dramatically reduced, effects of serial correlation are still present in the model at lag 1, 3 and 4.

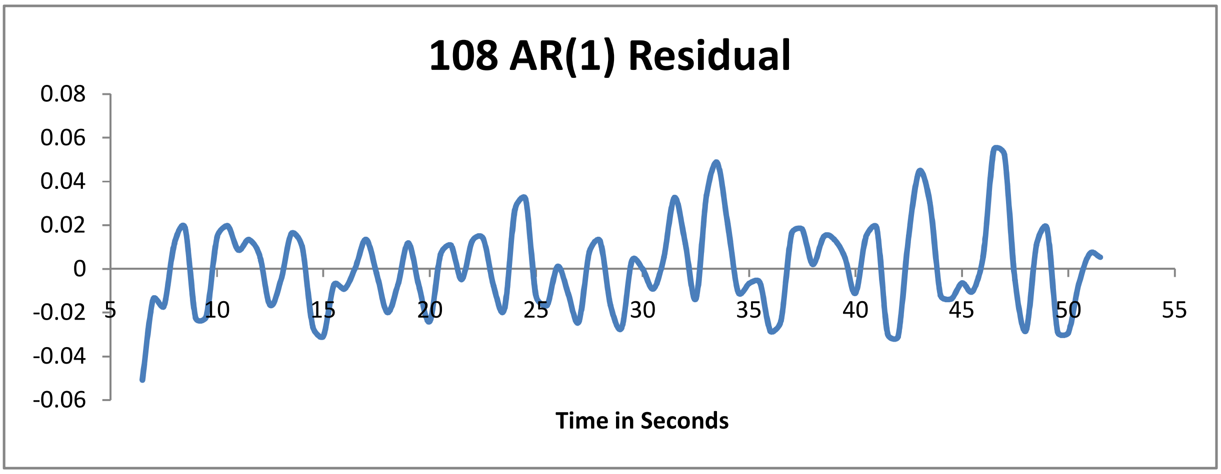


Figure 11—Plot of the residual for the AR(1) Model of 108. This time series represents the proportion of the variance not accounted for by the model. An ACF of this time series revealed autocorrelation at lags 1 and 2.

Correlations

		loudness	tempo	gsr	hr
loudness	Pearson Correlation	1	.502*	.755**	.706**
	Sig. (2-tailed)		.029	.000	.001
	N	19	19	19	19
tempo	Pearson Correlation	.502*	1	.622**	.601**
	Sig. (2-tailed)	.029		.004	.006
	N	19	19	19	19
gsr	Pearson Correlation	.755**	.622**	1	.767**
	Sig. (2-tailed)	.000	.004		.000
	N	19	19	19	19
hr	Pearson Correlation	.706**	.601**	.767**	1
	Sig. (2-tailed)	.001	.006	.000	
	N	19	19	19	19

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 12—A correlation matrix for the mean values of loudness, tempo, GSR, and HR across all 19 excerpts.

References

Bradley, M., and P. Lang. 2007. Emotion and motivation. In *The Handbook of Psychophysiology*, ed. Cacioppo, J. T., L. G. Tassinary, and G. Bernt, 581–608. Cambridge, UK: Cambridge University Press.

Cabrera, D. 1999. 'Psysound': a computer program for psychoacoustical analysis. In *Proceedings of the Australian Acoustical Society Conference*, 47–54.

Caplin, W. 1998. *Classical Form: A theory of formal functions for the instrumental music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press.

Clarke, E. 1999. Rhythm and timing in music. In *The psychology of music*, ed. D. Deutsch, 473–500. San Diego: Elsevier Academic Press.

Craig, D.G. 2005. An exploratory study of physiological changes during 'chills' induced by music. *Musicae Scientiae* 9 (2): 273–287.

Dawson, M., A. Schell, and D. Fillion. 2007. The electrodermal system. In *Handbook of Psychophysiology*, ed. Cacioppo, J. T., L. G. Tassinary, and G. Bernt, 159–81. Cambridge, UK: Cambridge University Press.

Dixon, S. 2003. On the analysis of musical expression in audio signals. Draft for *SPIE/IS&T Conference on Storage and Retrieval for Media Databases*. Online http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.5379&rep=rep1&type=pdf.

Dixon, S. 2007. Evaluation of the audio beat tracking system BeatRoot. Preprint for *Journal of New Music Research* 36. Online http://www.elec.qmul.ac.uk/people/simond/pub/2007/jnmr07.pdf.

Dubnov, S., S. McAdams, and R. Reynolds. 2006. Structural and affective aspects of music from statistical audio signal analysis. *Journal of the American Society for Information Science and Technology* 57 (11): 1526–36.

Edelberg, R. 1972. Electrodermal recovery rate, goal-orientation, and aversion. *Psychophysiology* 9: 512–520.

Farbood, M. 2006. A quantitative, parametric model of musical tension. PhD thesis, MIT Media Lab.

Gomez, P., and B. Danuser. 2007. Relationships between musical structure and psychophysiological measures of emotion. *Emotion* 7 (2): 377–87.

Guhn, M., A. Hamm, and M. Zenter. 2007. Physiological and musico-acoustic correlates of the chill response. *Music Perception* 24 (5): 473–83.

Krumhansl, C. L. 1997. An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology* 51 (4): 336–52.

Lerdahl, F. 1988. Tonal pitch space. *Music Perception* 5: 315–350.

Margulis, E. A model of melodic expectation. *Music Perception* 22 (4): 663–714.

McAdams, S., B. W. Vines., S. Vieillard., B. K. Smith., and R. Reynolds. 2004. Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception* 22 (2): 297–350.

McKay, C. 2010 (in preparation). Automatic music classification with jMIR. Ph.D. Dissertation. McGill University, Canada.

Moore, B., B. Glasberg, and T. Baer. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Scoeity* 45 (4): 224–40.

Nielsen, F. V. 1987. Musical tension and related concepts. In The *semiotic web '86: An international year-book*. Ed. T. Sebeok and J. Umiker-Seboek, 491–513. Berlin: Mouton de Gruyter.

Ogg, M. 2009. Physiological responses to music: measuring emotions. Undergraduate thesis. McGill University.

Palmer, C. 1996. Anatomy of a performance: Sources of musical expression. *Music Perception* 13 (3): 433–53.

Repp, B. H. 1994. On determining the basic tempo of an expressive music performance. *Psychology of Music*, 22: 157–67.

Rickard, N. 2004. Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of Music* 32 (4): 371–99.

Russell, J. A. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (6): 1161–78.

Schubert, E. 1999. Measurement and time series analysis of emotion in music. Ph.D. Dissertation. University of New South Wales.

Schubert, E. 2001. Continuous measurement of self-report emotional response to music. In *Music and Emotion*. Ed. P. Juslin & J. Sloboda, 393–414. New York: Oxford University Press.

Schubert, E. 2002. Correlation analysis of continuous emotional response to music: correcting for effects of serial correlation. *Musicae Scientiae* Special Issue 2001–2002: 213–36.

Schubert, E. 2004. Modeling perceived emotion with continuous musical features. *Music perception* 21 (4): 561–85.

Schubert, E. 2007. When is an event in a time-series significant? *International Conference on Music Communication Science*, 135–8.

Sloboda, J.A., and P.N. Juslin. 2001. Psychological perspectives on music and emotion. In P.N. Juslin & J.A. Sloboda Ed., *Music and Emotion*, 393–414. Oxford: Oxford University Press.

Thimm, K., and B. Fischer. 2003. Emotional responses to music: influence of psycho-acoustical features. In *Proceedings of the National Conference on Undergraduate Research*.

Turpin, G., and D. Siddle. 1979. Effects of stimulus intensity on electrodermal activity. *Psychophysiology* 16 (6): 582–91.

Widmer, G., S. Dixon, W. Goebl, E. Pampalk, and A. Tobudic. 2003. In search of the Horowitz factor. *AI Magazine* 24 (3): 111–29.

Todd, N. 1985. A model of expressive timing in tonal music. *Music Perception* 3 (1): 33–58.

Zwicker, E., and H. Fastle. 1999. *Psychoacoustics: Facts and Models*, 2nd Edition. Berlin: Springer.