Modeling Emotional Content of Music Using System Identification

Mark D. Korhonen, David A. Clausi, *Senior Member, IEEE*, and M. Ed Jernigan, *Member, IEEE*

***Abstract*—Research was conducted to develop a methodology to model the emotional content of music as a function of time and musical features. Emotion is quantified using the dimensions valence and arousal, and system-identification techniques are used to create the models. Results demonstrate that system identification provides a means to generalize the emotional content for a genre of music. The average R^2 statistic of a valid linear model structure is 21.9% for valence and 78.4% for arousal. The proposed method of constructing models of emotional content generalizes previous time-series models and removes ambiguity from classifiers of emotion.**

***Index Terms*—Appraisals, emotion, information retrieval, model, mood, music, perception, system identification.**

I. Introduction

THERE is a growing interest in analyzing the emotional content of music in the fields of music information retrieval and music psychology. Music information can be stored and retrieved using emotional content in addition to other musical characteristics such as artist, title, style, genre, or similarity [1]. Music psychologists are interested in studying how music communicates emotion [2]. Both of these fields require a method to measure and analyze the emotional content of music. Currently, no standardized methodology exists.

Feng *et al.* [3], Li and Ogihara [4], and Liu *et al.* [5] classify musical selections from various genres into 4, 6, or 13 different emotions. All of these studies rely on measuring musical features representing musical properties such as tempo, articulation, intensity, timbre, and rhythm to train a classifier. A comparison of these studies reveals that treating emotion as a discrete variable involves ambiguously selecting the number of emotions. To resolve this ambiguity, Schubert recommends representing an emotion as a continuous multidimensional variable [2].

Because music changes with time, the emotion communicated by the music can also change with time [6]. Because the emotion can vary throughout a musical selection, a time-varying method of measuring emotion is more appealing than describing music with a single emotion. To allow varying emotional content of a musical selection, Liu *et al.* [5] analyze emotion as piecewise constant over musical selections, whereas Schubert [2] analyzes emotion as a continuous function of time.

For reasons given in the preceding paragraphs, the emotional content of music should be quantified as a time-varying continuous variable. Schubert has expressed the time-varying emotional content of particular musical selections as a function of five time-varying musical features through a time-series analysis [2]. By generalizing Schubert's models to many different musical selections, it is possible to construct a mathematical model of the time-varying emotional content of music as a function of features in the music.

The goal of this paper is to develop a methodology to create valid models of time-varying continuous emotional content for a genre of music. The emotional content of various musical selections will be measured by representing the perceived emotional content of the music made by a population of listeners.

These models can be used to determine the regions of a musical selection that communicates a particular emotion or measures how much the emotional content deviates from a "base" emotion, as a function of time. The models may aid music information retrieval by enhancing classification and retrieval algorithms. Also, the models may provide a means to evaluate how various musical features affect the emotional content of music.

This paper is organized to present and evaluate a methodology to create valid models for emotional content of music. Section II provides the background necessary to quantify emotion as a multidimensional signal. Section III discusses a general methodology that can be used to create a model. Section IV describes the authors' implementation of the methodology and their results. This paper concludes in Section V with a discussion of the model and possible applications, as well as directions for future research.

II. Background

When presented with emotional stimuli, a person may experience the autonomic reactions and expressive behaviors associated with an emotion. In this paper, the term emotional response is used to indicate the person's experience of emotion. A person may also recognize emotion in the stimuli without experiencing the reactions and behaviors associated with emotion [7]. The process of recognizing emotions in the stimuli is referred to as perceiving emotion, and the term emotional appraisal is used to indicate the emotion recognized in the stimuli.

Manuscript received November 3, 2004; revised May 25, 2005. This work was supported in part by Naxos, by the Natural Sciences and Engineering Research Council of Canada (NSERC), and by the Ontario Graduate Scholarship (OGS). This paper was recommended by Associate Editor T. Takagi.

M. D. Korhonen is with the University of Waterloo, Waterloo, ON N2L 3G1 Canada, and also with CIMTEK, Burlington, ON L7L 6A6, Canada (e-mail: mdkorhon@alumni.uwaterloo.ca).

D. A. Clausi and M. E. Jernigan are with the Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: dclausi@engmail.uwaterloo.ca; jernigan@engmail.uwaterloo.ca).

Digital Object Identifier 10.1109/TSMCB.2005.862491

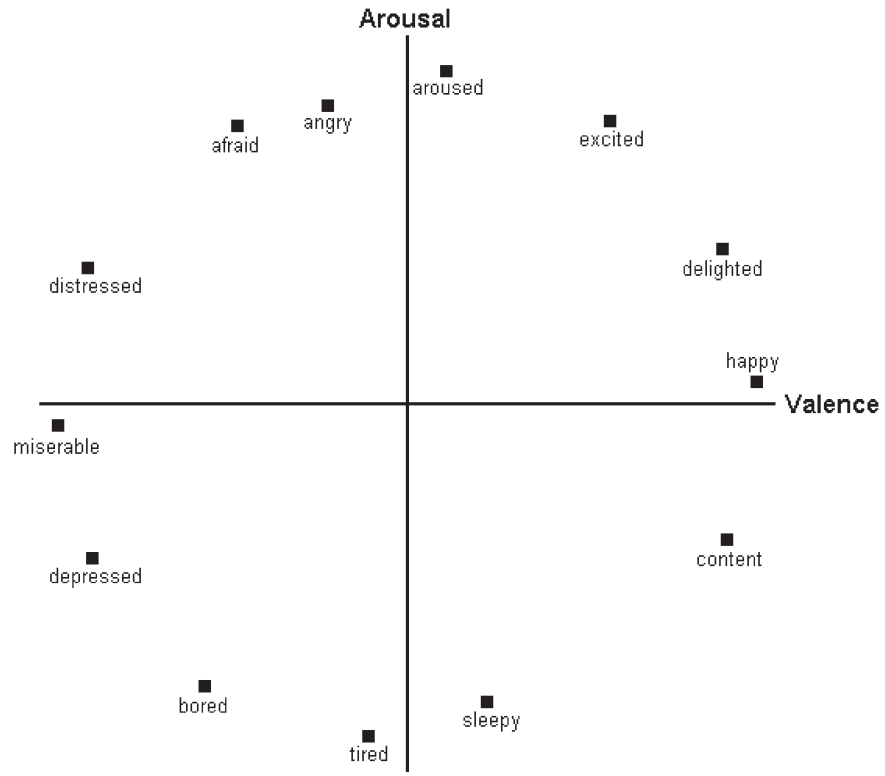

Fig. 1. Example of emotions plotted in a two-dimensional emotion space (2-DES) [9, p. 86].

If one considers music to be a medium for communicating emotions, emotional appraisals are more intuitive to investigate than emotional responses [8]. A person's emotional appraisal of music can be interpreted as the emotional content they perceive in the music.

The measurement of emotional appraisals of stimuli is accomplished by having the person report the emotions they perceive in the stimuli. This can be done in several different ways such as verbal descriptions, choosing emotional terms from a list, or rating how well several different emotional terms describe the appraisal [2], [7].

By rating emotional terms, emotions can be continuous quantities that are described using a vector. Results from multivariate analysis studies have "... suggested that many, perhaps most, emotions recognized in music may be represented in a two-dimensional (2-D) space with valence (positive versus negative feelings) and arousal (high–low) as principal axes ..." [7, p. 126]. These are the dimensions suggested by Russell to describe emotion [9]. These dimensions are also similar to those proposed by Thayer [10] and used by Liu *et al.* [5].

Fig. 1 is an adaptation of Russell's figure showing how several different emotions can be described using the dimensions valence and arousal[1] [9]. Valence refers to the happiness or sadness of the emotion, and arousal is the activeness or passiveness of the emotion [2]. Each component can be quantified by limiting the range of each dimension to $[-100\%, 100\%]$ and rating each component on this scale.

A person can describe an emotional appraisal on a computer by using a mouse (or similar input device) to move a cursor in the two-dimensional emotion space (2-DES), and the cursor position would correspond to the emotional appraisal. By recording how the cursor position changes with time, the person can easily describe how their emotional appraisals change with time as the stimulus changes. FEELTRACE [11] and EmotionSpace Lab [12] are examples of software that are able to collect reliable time-varying emotional appraisals using a 2-DES to emotionally appraise stimuli (e.g., words, faces, music, and video).

When people perceive emotion in music, there are some emotions that are reliably perceived and other emotions that are confused with different emotions [7]. The emotions that are reliably perceived, such as happiness and sadness, each appear to have distinctive arousal and/or valence values. Generally, the emotions that are confused (e.g., calm versus sorrow, anger versus fear) appear to have similar arousals and valences. This may mean that while emotion may consist of components other than arousal and valence, these two components may be the ones that are most clearly communicated through music. These reasons provide additional motivation for using the 2-DES to emotionally appraise music.

To summarize, the emotional content of music can be quantified by measuring emotional appraisals of music collected using a software such as EmotionSpace Lab. For example, many different people could appraise the same musical selections using this software, and their appraisals could be combined to

[1]The coordinates and relative positions of the labeled emotions in this space have been selected for illustrative purposes. The authors make no attempt to describe the exact coordinates of particular emotions in this space.

generate an emotional appraisal representative of the population. The representative emotional appraisal of a musical selection can be interpreted as the emotional content of the music.

III. Proposed Method

The goal of this project is to develop a methodology to model the emotional content of music. A model should meet the following criteria.

1) The measured emotional content (emotional appraisals of a population of listeners) needs to be time varying.
2) The musical features that are inputs to the model need to represent many musical properties that communicate emotion and also need to be time varying.
3) The model needs to be estimated/trained using emotional appraisals to musical selections representing a genre of music.
4) The model needs to accurately simulate emotional appraisals to any musical selection from the genre of music.

Initially, only one genre of music should be represented per model. Although multiple genres of music could be modeled, Li and Ogihara [4] suggest that limiting a model to one genre of music can result in improved performance.

Once a model is obtained that meets these four criteria, it can be used to estimate the emotional content of all musical selections in the genre. To create a model meeting these criteria, the system-identification procedure described by Ljung will be used [13]. Through model construction, the first three criteria can be met. To evaluate how well a model meets the fourth criterion, the model can be evaluated to measure how well it generalizes emotional appraisals.

The system-identification process consists of multiple stages that can be performed iteratively [13]. These stages form the basis of the methodology discussed in the following sections.

1) Design the experiment[2] to collect input and output signals.
2) Select the input signals (musical features) to be used in the study.
3) Perform the study to collect output signals (emotional appraisals).
4) Select model structures for evaluation.
5) Select the algorithm used to estimate the parameters of the models.
6) Estimate the parameters of the models using the input and output signals and evaluate the models to determine their validity.

A. Experiment Design

To be able to measure emotional appraisals of a population of listeners and create models from these measured appraisals meeting the four criteria, several variables need to be selected. These variables include the genre of music to model, the number of musical selections to be appraised, the duration of music listened to by the volunteers, the number of volunteers, and the sampling rate of the cursor in the 2-DES.

After selecting the genre of music, the genre can be represented using multiple musical selections. To avoid biasing the model performance to longer songs, it may be desirable to modify the musical selections to be approximately the same duration.

To ensure that each listener is able to concentrate throughout the study, the duration of the session with each listener should be limited [2]. Thus, it is impractical to have each listener appraise a large number of pieces. To overcome this limitation, many listeners could appraise a random subset of the musical selections, where A is the total number of musical selections. If not enough listeners are available, another alternative is to use a limited amount of data (a small value for A) that are as informative as possible. To be maximally informative, the A musical selections need to differ and vary considerably. This can be accomplished by using as many musical selections as possible in the time period that have possibly been duration modified.

The sampling rate of the cursor in the 2-DES needs to be selected. Ideally, the sampling rate should approximately equal the time constants of the system [13]. Since these are not exactly known, one can sample as fast as possible, and digitally prefilter and decimate the signals to obtain a desired sampling rate [13].

B. Feature Measurement

To satisfy the second model criterion and to use the musical selections as input signals in the model, the music needs to be represented by m time-varying musical features. These m features are measured every second and treated as an m-dimensional vector $\underline{u}_a(t)$, where t is the time in seconds when the features are calculated for musical selection a ($a = 1, 2, \ldots, A$).

For the features to represent the emotional content of the music, the m features should represent musical properties that communicate emotion. Schubert has performed a comprehensive review of studies that determine which musical properties cause listeners to perceive an emotion [2]. There are 16 properties identified by Schubert: dynamics, mean pitch, pitch range, variation in pitch, melodic contour, register, mode, timbre, harmony, texture, tempo, articulation, note onset, vibrato, rhythm, and meter. Features representing these properties can be measured using algorithms, such as those found in PsySound [14] and in Musical Research System for Analysis and Synthesis (MARSYAS) [15]. Once the features have been measured, it may be necessary to resample the features in order to have the same sampling rate as the appraisal measurements or to perform other preprocessing as discussed by Ljung [13].

C. Appraisal Measurement

By sampling the cursor position of a listener's emotional appraisal in a 2-DES, the first model criterion can be met. For example, Schubert's EmotionSpace Lab can be used to measure emotional appraisals [2], [12].

[2]The term "experiment design" is used to be consistent with system-identification literature. The term "study" will be used interchangeably with "experiment" in the remainder of this paper.

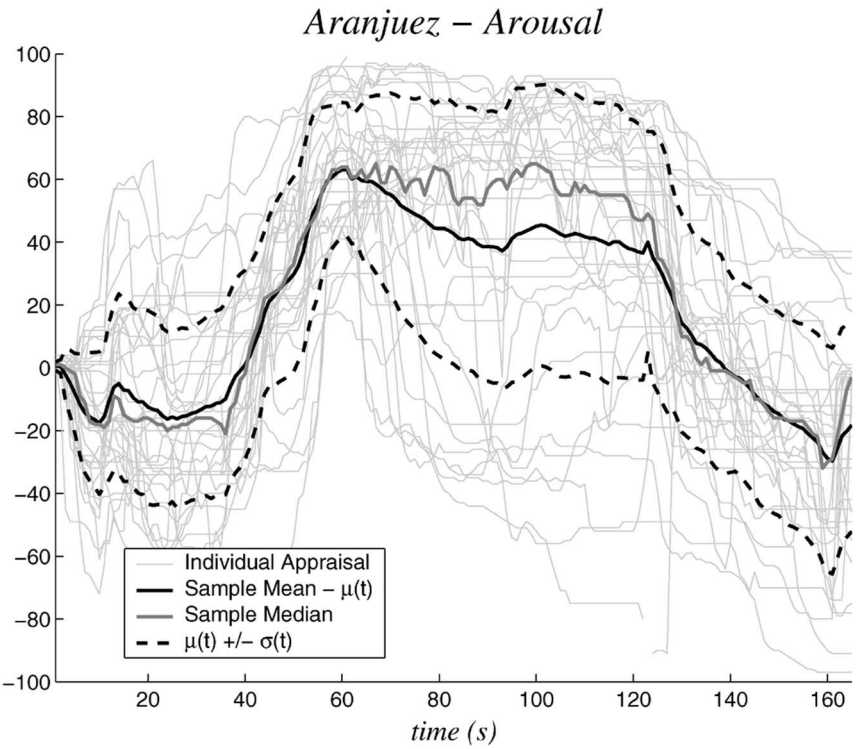

Fig. 2. Comparison of median and mean to calculate a representative appraisal (for the arousal dimension of a selection from Rodrigo's Concierto de Aranjuez) [8, p. 110].

To collect emotional appraisal data from a population of B listeners, each listener must appraise the music. Let $b = 1, 2, \ldots, B$ represent each listener, and let A_b represent the number of musical selections heard by listener b. For each listener, the A_b musical selections should be randomly selected from the A musical selections for evaluation in the session. To ensure that many people have evaluated each musical selection, the following expression should be met.

$$BA_b \gg A, \qquad b = 1, 2, \ldots, B \tag{1}$$

Once the emotional appraisals of B listeners have been collected, each of the A musical selections in the database will have been appraised multiple times. To create a model, it is advantageous to combine the multiple emotional appraisals for each musical selection into a single emotional appraisal. The advantages of creating a representative appraisal include: data reduction, improvements in signal to noise ratios, and improvements in the conditioning of the model-estimation algorithm [13]. This single emotional appraisal should be representative of all of the listeners. Creating a representative emotional appraisal for each musical selection assumes that emotional appraisals are consistent across cultures, music training, and other variables of the sample population. For the discussion that follows, the following 2-D time-varying vectors are defined:

- $\underline{\gamma}_{ab}(t)$ emotional appraisal of person b to musical selection a at time t, $b = 1, \ldots, B$, $a = 1, \ldots, A$;
- $\underline{Y}_a(t)$ random vector describing the population's emotion appraisal of musical selection a at time t;
- $\underline{y}_a(t)$ emotional appraisal representative of the population for musical selection a at time t.

The probability distribution function (pdf) of $\underline{Y}_a(t)$ is a function of musical features and emotional appraisals prior to time t. However, by considering the marginal pdf of the emotional appraisal as a function of time only, it is possible to calculate an emotional appraisal representative of the population at a particular time t by considering only the observed emotional appraisals at t. This is acceptable because the models that will be identified determine how the musical features and emotional appraisals affect $\underline{Y}_a(t)$.

The vector $\underline{\gamma}_{ab}(t)$ can be interpreted as the bth observation of $\underline{Y}_a(t)$. Because each person appraises a subset of the A musical selections, $\underline{\gamma}_{ab}(t)$ will not have data for some of the musical selections.

There are several ways to obtain $\underline{y}_a(t)$ from $\underline{\gamma}_{ab}(t)$. Fig. 2 illustrates a comparison of using the sample median and sample mean to calculate $\underline{y}_a(t)$ for a particular piece (the sample standard deviation as a function of time is labeled $\sigma(t)$ in this figure) [8]. Korhonen notes that the median and mean appraisals are similar except when the distribution of appraisals appears to be bimodal or skewed; in these cases, the median is a more robust measure of centrality [8]. Usage of the sample median also allows handling missing data and outliers by omitting them from the calculation of the representative appraisal. For these reasons, the sample median is a reasonable method to calculate $\underline{y}_a(t)$.

Once a representative emotional appraisal $\underline{y}_a(t)$ has been calculated for all A songs in the database, it should be preprocessed to improve model estimation. Two methods of preprocessing the data are lowpass filtering to remove high-frequency noise and highpass filtering to remove drifts and offsets [13].

D. Model Structure

Once the musical features and emotional appraisals are collected, the next step is to select the model structures to use. Each model structure is parameterized using a d-dimensional vector $\underline{\theta}$ consisting of all of the parameters needed to describe the model. Each model can be described using the following expression:

$$\underline{\hat{y}}_a(t|\underline{\theta}) = f\left(\underline{u}_a(t), \underline{u}_a(t-1), \ldots, \underline{e}(t), \underline{e}(t-1), \ldots\right) \quad (2)$$

where

$\underline{\hat{y}}_a(t\|\underline{\theta})$	simulated output for musical selection a;
$\underline{u}_a(t)$	feature vector for musical selection a;
$\underline{e}(t)$	2-D white noise process with zero mean;
$f()$	function representing the model structure;
$\underline{\theta}$	d-dimensional vector containing all of the parameters needed to describe $f()$.

Because $f()$ is not a function of $\underline{y}_a(t-1), \underline{y}_a(t-2), \ldots$, this model structure is a simulation model (as opposed to a prediction model) [13]. Although prediction models are used in many system-identification problems, simulation models are required to meet the fourth model criterion. Also, $f()$ is the same for all musical selections to satisfy the third model criterion. To simplify the discussion in the following sections, only linear models will be considered; the methodology described in the following two sections can be extended to nonlinear model structures as well.

After selecting the model structures, the number of parameters needs to be chosen. For example, in an artificial neural network, the parameters are the weights and biases that depend on the number of layers and neurons. For another example, in a state-space model, choosing the order of the model determines the number of parameters.

E. Model Estimation

To ensure that the third model criterion is met, the parameters ($\underline{\theta}$) of a model need to be estimated using the musical features and the representative emotional appraisals. A subset of the musical selections, referred to as the training set, is used to estimate the parameters in a model. The remaining set of musical selections is referred to as the testing set and is used to validate the model.[3]

Before estimating the parameters in the linear models, data fusion needs to be used to combine the input and output data from all of the musical selections in the training set. The A musical selections [represented by $\underline{u}_a(t), \underline{y}_a(t)$] are treated as one continuous musical selection [represented by $\underline{u}(t), \underline{y}(t)$], but the initial conditions are reset at the beginning of each musical selection. Using a similar notation, let $\underline{\hat{y}}(t|\underline{\theta})$ represent the simulation of $\underline{y}(t)$.

Once the structure of the model is selected, the parameters of the model can be estimated using various algorithms. For example, if the model is nonlinear, methods such as a gradient descent can be used. If the model's structures are linear, the prediction error method (PEM) is suggested because it will generate unbiased estimates of the parameters regardless if the "true" system can be represented using the model structure [13]. To use PEM, a norm must be selected such as the determinant of the estimated error covariance, $\widehat{\Lambda}_N(\underline{\theta})$. This choice of norm for PEM is described using the following equations [13]:

$$\underline{\hat{\theta}} = \arg\min_{\underline{\theta}} V_N(\underline{\theta}) \quad (3)$$

$$V_N(\underline{\theta}) = \left|\widehat{\Lambda}_N(\underline{\theta})\right| \quad (4)$$

$$\widehat{\Lambda}_N(\underline{\theta}) = \frac{1}{N-d}\sum_{t=1}^{N} \underline{\epsilon}(t|\underline{\theta})\underline{\epsilon}^T(t|\underline{\theta}) \quad (5)$$

$$\underline{\epsilon}(t|\underline{\theta}) = \underline{y}(t) - \underline{\breve{y}}(t|\underline{\theta}) \quad (6)$$

where

$V_N(\underline{\theta})$	loss function;
$\underline{\hat{\theta}}$	estimate of $\underline{\theta}$;
$\underline{\breve{y}}(t\|\underline{\theta})$	one-step-ahead prediction for $\underline{y}(t)$ for the model structure;
N	total number of samples in the training set;
d	number of parameters in $\underline{\theta}$.

If the model structure is linear, it is straightforward to relate the one-step-ahead prediction ($\underline{\breve{y}}(t|\underline{\theta})$) to the simulated output ($\underline{\hat{y}}(t|\underline{\theta})$) [13]. For nonlinear models, other estimation methods may be more appropriate.

F. Validation

If a model can accurately simulate emotional appraisals to any musical selection from the genre of music, the fourth model criterion will be satisfied. To measure the accuracy of a model, the bias and variance errors will be estimated. To verify assumptions made by a given model structure, a residual analysis will be performed.

Evaluating the bias error of a model can be done using K-fold cross validation [16]. For each model structure, use K different training sets and measure the mean-squared error (mse) for each of the K different testing sets. Because there are two outputs, the mse should be calculated separately for each of the outputs. The mse for testing set k is described by the following equation:

$$\text{mse}_{\alpha_k,w} = \frac{1}{N_{\alpha_k}}\sum_{t=1}^{N_{\alpha_k}} \left|y_{\alpha_k,w}(t) - \hat{y}_{\alpha_k,w}(t|\underline{\theta})\right|^2 \quad (7)$$

where

w	dimensions valence and arousal;
k	testing set ($k = 1, 2, \ldots, K$);
α_k	subset of the A musical selections in testing set k;
$\text{mse}_{\alpha_k,w}$	mse for dimension w of testing set k;
N_{α_k}	total number of samples in testing set k;

[3]To perform a cross validation as described in Section III-F, several different training sets will be used for a given model architecture. However, the method of model estimation will remain the same for all training sets.

$\underline{y}_{\alpha_k}(t)$ representative emotional appraisal of the musical selections in α_k that have been combined using data fusion, as discussed in Section III-E;

$y_{\alpha_k,w}(t)$ wth dimension of $\underline{y}_{\alpha_k}(t)$ in testing set k (i.e., $\underline{y}_{\alpha_k}(t) = [y^T_{\alpha_k,\text{valence}}(t), y^T_{\alpha_k,\text{arousal}}(t)]^T$);

$\hat{y}_{\alpha_k,w}(t|\underline{\theta})$ wth dimension of the simulated output of testing set k.

By constructing the K testing sets so that the data for each musical selection are found in exactly one of the K testing sets, a simulated output exists for all A musical selections. The resultant mse for output w can then be calculated using the following weighted average:

$$\text{mse}_w = \frac{\sum_{k=1}^{K} N_{\alpha_k} \text{mse}_{\alpha_k,w}}{\sum_{k=1}^{K} N_{\alpha_k}}. \tag{8}$$

Because the mse is a function of the energy of the signal, it is desirable to normalize the mse using the squared-multiple-correlation coefficient (R^2). By using the R^2 measure, it is possible to compare the bias of models estimated using any dataset. The R^2 statistic is sometimes referred to as the "fit" and should be as close to one as possible.[4] The mse for output w can be related to R^2 for output w using the following expression [13]:

$$R^2_w = 1 - \frac{\text{mse}_w}{\frac{1}{N}\sum_{a=1}^{A}\sum_{t=1}^{N_a} |y_{a,w}(t)|^2} \tag{9}$$

where $y_{a,w}(t)$ is the wth dimension of $\underline{y}_a(t)$.

To measure the variance error of the model structures, two techniques will be used. First, the variance of the parameters can be estimated to calculate 98%-confidence intervals. For a linear model, this corresponds to ± 2.33 standard deviations (σ), since the parameters estimated with PEM converge to a normal distribution as the number of data samples increases [13]. Parameters that reflect design decisions (such as model order or time delay) should be statistically significant from zero to be included in the model. Also, if the confidence intervals of many parameters are large, then this implies that there are too many parameters [13]. For linear models, the covariance of the parameters $\widehat{P}_{\alpha_k,\theta}$ can be estimated for each of the K testing sets using the following equations:

$$\widehat{P}_{\alpha_k,\theta} = \left[\frac{1}{N_{\alpha_k}} \sum_{t=1}^{N_{\alpha_k}} \underline{\psi}_{\alpha_k}(t,\hat{\underline{\theta}}) \widehat{\Lambda}_{N_{\alpha_k}}(\hat{\underline{\theta}}) \underline{\psi}^T_{\alpha_k}(t,\hat{\underline{\theta}})\right]^{-1} \tag{10}$$

$$\underline{\psi}_{\alpha_k}(t,\underline{\theta}) = \frac{d\underline{\hat{y}}_{\alpha_k}(t|\underline{\theta})}{d\underline{\theta}} \tag{11}$$

[4]If the R^2 statistic is negative, the energy of the error is greater than the energy of the true emotional appraisals. This implies that the simulated emotional appraisal is extremely different from the true emotional appraisal. For reference, a constant simulated output results in the R^2 statistic equal to zero.

where

$\underline{\psi}_{\alpha_k}(t,\underline{\theta})$ $d \times 2$ matrix representing the gradients (sensitivity) of the simulated output of testing set k with respect to each parameter at time t;

$\widehat{\Lambda}_{N_{\alpha_k}}(\hat{\underline{\theta}})$ estimated error covariance for testing set k.

The second measure used to analyze the variance of the model is the estimated variance of the output signals. Ideally, the variance of the output signals is small so that the output is known with some certainty. To analyze the variance of the output signals, 98% confidence intervals of the simulated output can be graphically compared to emotional appraisals.

If the model structures are linear, the output is a linear function of $\hat{\underline{\theta}}$. This implies that $\underline{\hat{y}}_{\alpha_k}(t|\hat{\underline{\theta}})$ can be expressed as follows:

$$\underline{\hat{y}}_{\alpha_k}(t|\hat{\underline{\theta}}) = \underline{\psi}^T_{\alpha_k}(t,\hat{\underline{\theta}})\hat{\underline{\theta}} + \underline{e}(t). \tag{12}$$

Since $\hat{\underline{\theta}}$ is approximately normally distributed and $\underline{e}(t)$ is a white noise process, (12) illustrates that $\underline{\hat{y}}_{\alpha_k}(t|\hat{\underline{\theta}})$ is approximately normally distributed as well. The variance of $\underline{\hat{y}}_{\alpha_k}(t|\hat{\underline{\theta}})$ can be calculated on the validation data using the following equation, since $\underline{e}(t)$ will be uncorrelated with $\hat{\underline{\theta}}$:

$$\text{Var}\left(\underline{\hat{y}}_{\alpha_k}(t|\hat{\underline{\theta}})\right) = \underline{\psi}^T_{\alpha_k}(t,\hat{\underline{\theta}}) \widehat{P}_{\alpha_k,\theta} \underline{\psi}_{\alpha_k}(t,\hat{\underline{\theta}}) + \widehat{\Lambda}_{N_{\alpha_k}}(\hat{\underline{\theta}}). \tag{13}$$

Assumptions made during the creation of the models need to be verified using a residual analysis. To verify that the inputs are independent of the noise process, the cross-correlation function between each input and the model residuals will be examined to ensure no negative lags are significantly different than zero. The autocorrelation function (ACF) of the output residuals will also be calculated to ensure only the zeroth lag is significantly different than zero. This test will be done to ensure that the noise is white.

Once all of the model structures have been evaluated, a resultant model can be created for the best model structures. The resultant models should be estimated using all of the musical selections. These models can be compared using Akaike's final prediction error (FPE) criterion to assess the tradeoff between minimizing the mse while minimizing the variance error by limiting the number of parameters in the model. The expression to calculate the FPE is given by Ljung [13] as

$$\text{FPE} = \frac{(N+d)}{(N-d)} V_N(\underline{\theta}). \tag{14}$$

IV. Implementation and Results

The methodology described in Section III was used to create linear models of emotional appraisals. MATLAB's System Identification Toolbox was used. The dataset used to generate the following models can be found at http://www.sauna.org/kiulu/emotion.html.

TABLE I
MUSICAL SELECTIONS USED IN THIS STUDY

Musical Selection	Composer	Times	Duration
Concierto de Aranjuez (Adagio)	Rodrigo	7:05 – 9:45 & 5s silence	2:45
Fanfare for the Common Man	Copland	0:00 – 2:50	2:50
Moonlight Sonata (Adagio Sostenuto)	Beethoven	0:00 – 0:22 3:08 – 5:19	2:33
Peer Gynt (Morning)	Grieg	0:00 – 2:39 & 5s silence	2:44
Piano Concerto No. 1 (Allegro Maestoso)	Liszt	0:00 – 5:15	5:15
Pizzicato Polka	J. Strauss	0:00 – 2:31	2:31

A. Experiment Design

Emotional appraisals for six musical selections were measured using EmotionSpace Lab to quantify emotions using the dimensions valence and arousal [2]. Because EmotionSpace Lab collects emotional appraisal data at 1 Hz, it is assumed that emotional appraisals contain information only at frequencies below 0.5 Hz. It would be worthwhile to sample much faster in future studies and then resample the signal to a desired frequency to ensure all frequencies of interest are collected.

The genre of music was selected was the Western art musical style. Table I lists the musical selections from Naxos' "Discover the Classics" compact disk [(CD) 8.550035–36] that are used in this study. Only six musical selections are used to limit the scope, and the total duration of the songs was limited to 20 min. To minimize bias towards a particular musical selection, the musical selections were modified to be approximately of the same length. Although Liszt's Piano Concerto is longer than the other songs, it was not modified for duration because preliminary testing showed that this song appeared to be more informative than the other songs.

B. Feature Measurement

To achieve the second model criterion, time-varying musical features need to be measured from the musical selections. The features were extracted using PsySound [14] or the fast Fourier transform (FFT) extractor from MARSYAS [15] (tempo was extracted manually using the method described in Schubert's Ph.D. thesis[5] [2]). Features are extracted using established algorithms to minimize the subjectivity in the features. PsySound is used because it extracts psychoacoustic features that represent many musical properties that communicate emotion. MARSYAS is used for feature extraction because it has successfully been used in music-information-retrieval applications (e.g., [17]).

The diffuse field was used for PsySound analysis because music is the auditory stimulus and the music may be interpreted as originating around the listener since they are wearing headphones [14]. The features extracted by MARSYAS were resampled from 86 (17/128) to 1 Hz using a polyphase antialiasing filter to eliminate a high-frequency noise [18]. After the extraction of the musical features, the mean was subtracted (i.e., dc removal).

[5]Ideally, a reliable method of extracting a tempo programmatically should be used. However, to the best of the authors' knowledge, there is no reliable algorithm to estimate a tempo.

TABLE II
MUSICAL FEATURES USED IN THIS STUDY

No.	Musical Property	Musical Feature	Extraction Method
1	Dynamics	Loudness Level	PsySound
2		Short Term Max. Loudness	PsySound
3	Mean Pitch	Power Spectrum Centroid	PsySound
4		Mean STFT Centroid	MARSYAS
5	Pitch Variation	Mean STFT Flux	MARSYAS
6		Std. Dev. STFT Flux	MARSYAS
7		Std. Dev. STFT Centroid	MARSYAS
8	Timbre	Timbral Width	PsySound
9		Mean STFT Rolloff	MARSYAS
10		Std. Dev. STFT Rolloff	MARSYAS
11		Sharpness (Zwicker and Fastl)	PsySound
12	Harmony	Spectral Dissonance (Hutchinson and Knopoff)	PsySound
13		Spectral Dissonance (Sethares)	PsySound
14		Tonal Dissonance (Hutchinson and Knopoff)	PsySound
15		Tonal Dissonance (Sethares)	PsySound
16		Complex Tonalness	PsySound
17	Tempo	Beats per Minute	Schubert's method
18	Texture	Multiplicity	PsySound

Eighteen musical features used in this project are summarized in Table II. These features are selected to represent the 16 musical properties identified by Schubert (see Section III-B) [2]. Seven of these properties are directly represented by features and six others may be indirectly represented by the same features. The remaining three properties are either difficult to quantify using a continuous variable (rhythm, meter) or difficult to quantify using a time-varying variable (pitch range). For a detailed description of how the musical features relate to musical properties, consult Korhonen [8]. The portion of the emotional appraisals influenced by omitted musical properties is assumed to be accounted for by the stochastic component of the models.

C. Appraisal Measurement

Emotional appraisal data were collected from 35 volunteers—21 male (60%) and 14 female (40%). Each volunteer listened to all six musical selections in a random order. Because $A_b = A = 6,\ b = 1, \ldots, 35$, (1) is satisfied since BA_b is 35 times greater than $A \forall b$.

To calculate an emotional appraisal representative of the population, the median emotional appraisal was used. After the representative emotional appraisal for each musical selection $\underline{y}_a(t)$ was calculated, the mean was subtracted (i.e., dc removal) to remove any offsets.

D. Model Structure

For this study, only two linear models are investigated. The two linear models considered are the autoregression with extra inputs (ARX) and state-space model structures. From the work of Tillman and Bigand, it appears that fewer than 6 s of musical

stimuli is needed to represent emotion so the maximum order considered will be five [19].

Given m-dimensional input data $\underline{u}(t)$ and 2-D output data $\underline{y}(t)$, the ARX model structure can be described using the following expression:

$$\underline{y}(t) + A_1(\underline{\theta})\underline{y}(t-1) + \cdots + A_{n_a}(\underline{\theta})\underline{y}(t-n_a) = B_0(\underline{\theta})\underline{u}(t) + \cdots + B_{n_b}(\underline{\theta})\underline{u}(t-n_b) + \underline{e}(t) \quad (15)$$

where

- $A_k(\underline{\theta})$ 2×2 matrix;
- $B_k(\underline{\theta})$ $2 \times m$ matrix;
- $\underline{e}(t)$ 2-D white noise process with zero mean;
- n_a maximum number of auto-regressive terms in the model;
- n_b maximum number of lagged inputs in the model;
- $\underline{\theta}$ d-dimensional vector containing all of the nonzero elements of $A_k(\underline{\theta})$ and $B_k(\underline{\theta})$.

Given the same input and output data as in the ARX model structure, the state-space model structure can be described using the following expressions:

$$\underline{x}(t+1) = A(\underline{\theta})\underline{x}(t) + B(\underline{\theta})\underline{u}(t) + K(\underline{\theta})\underline{e}(t) \quad (16)$$

$$\underline{y}(t) = C(\underline{\theta})\underline{x}(t) + D(\underline{\theta})\underline{u}(t) + \underline{e}(t) \quad (17)$$

- $\underline{x}(t)$ n-dimensional state vector;
- $A(\underline{\theta})$ $n \times n$ matrix representing the dynamics of the state vector;
- $B(\underline{\theta})$ $n \times m$ matrix describing how the inputs affect the state variables;
- $C(\underline{\theta})$ $2 \times n$ matrix describing how the state vector affects the outputs;
- $D(\underline{\theta})$ $2 \times m$ matrix describing how the current inputs affect the current outputs;
- $K(\underline{\theta})$ $n \times 2$ matrix used to model the noise in the state vector.

The initial state $\underline{x}(0)$ can be set to zero or estimated from the data by including it in $\underline{\theta}$. Also, all nonzero elements of the matrices are represented using $\underline{\theta}$.

See Ljung [13] for expressions describing the simulation model $\hat{\underline{y}}(t|\hat{\underline{\theta}})$ and the one-step-ahead prediction model $\hat{\underline{y}}_p(t|\hat{\underline{\theta}})$ for these two model structures.

E. Model Estimation

PEM was used to estimate the parameters of the models, and the determinant of the estimated error covariance was used as the norm. Because the means of the input and output signals were removed, the initial value of the emotional appraisal for each musical selection was estimated for the calculation of mse and R^2 measures.

F. Resultant Model

Twelve different state-space models and 45 different ARX models were estimated and evaluated. For a detailed description of some of the models used in this study, consult Korhonen [8] and Korhonen *et al.* [20]. The best model structure was an ARX model using 16 of the 18 musical features and 38 parameters, as shown in (18)–(24) at the bottom of the page, where

- $\underline{y}(t)$ vector consisting of valence and arousal at time t;
- $\underline{u}(t)$ vector consisting of the following features from Table II measured at time t: loudness level (LN), power spectrum centroid (Centroid), short term maximum (Max.) loudness (NMax), sharpness (Zwicker & Fastl) [S(Z&F)], timbral width (TW), spectral dissonance (Hutchinson & Knopoff) [SDiss(H&K)], spectral dissonance (Sethares) [SDiss(S)], tonal dissonance (Hutchinson & Knopoff) [TDiss(H&K)], tonal dissonance (Sethares) [TDiss(S)], complex tonalness (CTonal), multiplicity (Mult), mean short time Fourier transform (STFT) centroid (MeanCentroid), mean STFT rolloff (MeanRolloff), mean STFT flux (MeanFlux), standard deviation (Std. Dev.) STFT centroid

---

$$\underline{y}(t) + A_1(\underline{\theta})\underline{y}(t-1) + A_2(\underline{\theta})\underline{y}(t-2) = B_0(\underline{\theta})\underline{u}(t) + B_1(\underline{\theta})\underline{u}(t-1) + B_2(\underline{\theta})\underline{u}(t-2) + \underline{e}(t) \quad (18)$$

$$\hat{\underline{y}}(t|\underline{\theta}) = \left[I + A_1(\underline{\theta})q^{-1} + A_2(\underline{\theta})q^{-2}\right]^{-1}\left[B_0(\underline{\theta}) + B_1(\underline{\theta})q^{-1} + B_2(\underline{\theta})q^{-2}\right]\underline{u}(t) \quad (19)$$

$$A_1(\underline{\theta}) = \begin{bmatrix} \theta_1 & \theta_2 \\ 0 & \theta_3 \end{bmatrix} \quad (20)$$

$$A_2(\underline{\theta}) = \begin{bmatrix} \theta_4 & \theta_5 \\ 0 & \theta_6 \end{bmatrix} \quad (21)$$

$$B_0(\underline{\theta}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_7 & \theta_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_9 & \theta_{10} & \theta_{11} & 0 & \theta_{12} & 0 & 0 & 0 & \theta_{13} & 0 & 0 & \theta_{14} & \theta_{15} & \theta_{16} & \theta_{17} & \theta_{18} \end{bmatrix} \quad (22)$$

$$B_1(\underline{\theta}) = \begin{bmatrix} 0 & \theta_{19} & \theta_{20} & \theta_{21} & 0 & 0 & \theta_{22} & 0 & \theta_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{24} & 0 & \theta_{25} & \theta_{26} & \theta_{27} & 0 & \theta_{28} & 0 & 0 & 0 & \theta_{29} & \theta_{30} & \theta_{31} & 0 & 0 & 0 & 0 & \theta_{32} \end{bmatrix} \quad (23)$$

$$B_2(\underline{\theta}) = \begin{bmatrix} 0 & 0 & 0 & \theta_{33} & 0 & 0 & 0 & 0 & \theta_{34} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{35} & 0 & \theta_{36} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_{37} & 0 & 0 & 0 & 0 & 0 & \theta_{38} \end{bmatrix} \quad (24)$$

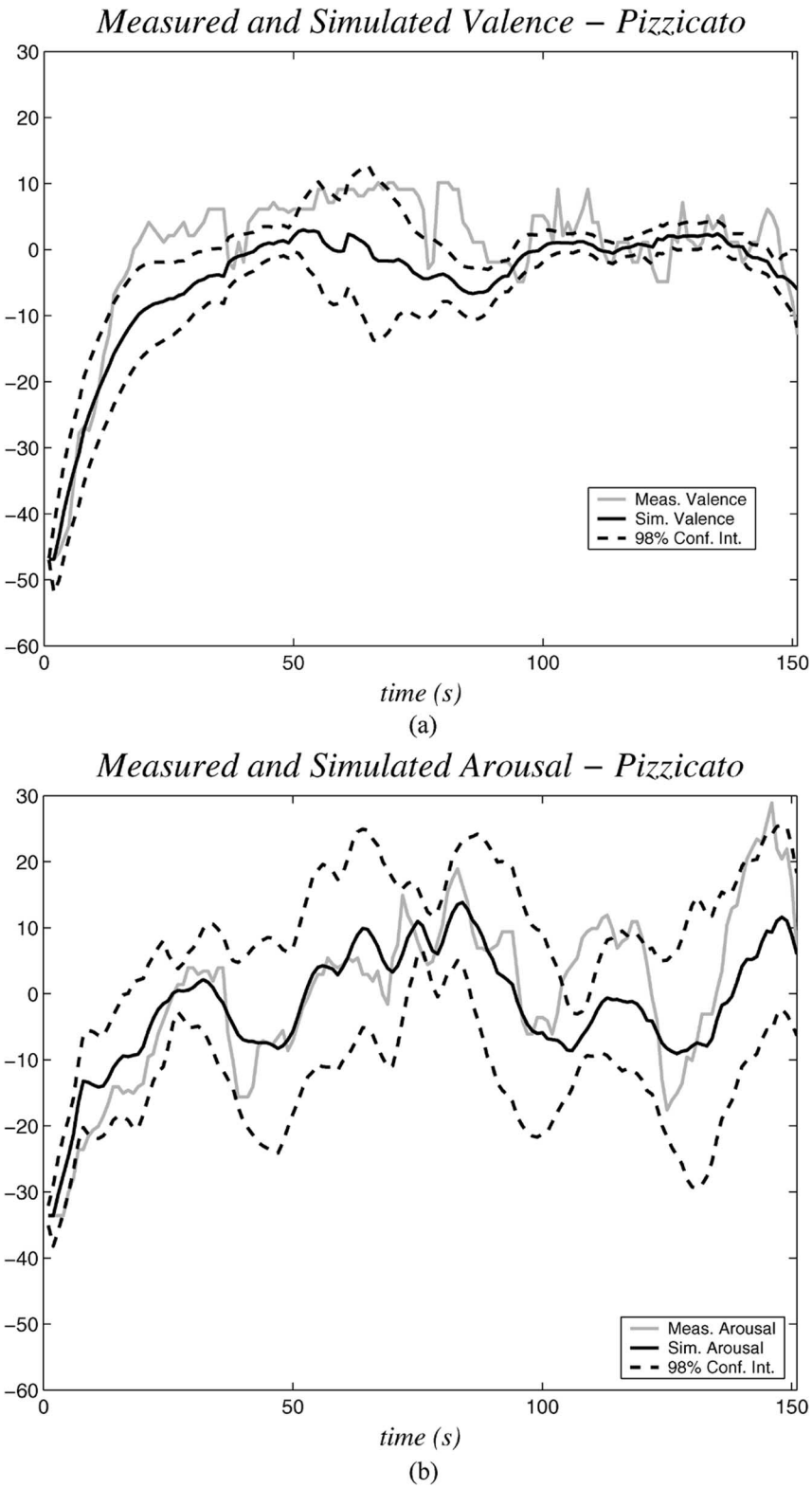

Fig. 3. Simulation of Pizzicato Polka. (a) Valence and (b) arousal.

(StdCentroid), Std. Dev. STFT rolloff (StdRolloff), Std. Dev. STFT flux (StdFlux), beats per minute (BPM);

q unit-shift operator: $q^{-k}\underline{y}(t) = \underline{y}(t-k)$;

$\underline{\theta}$ 38-dimensional vector of all the parameters used to describe $A_1(\underline{\theta})$, $A_2(\underline{\theta})$, $B_0(\underline{\theta})$, $B_1(\underline{\theta})$, and $B_2(\underline{\theta})$.

This model structure had an R^2 value of 21.9% for valence, an R^2 value of 78.4% for arousal, and an Akaike's FPE value of 131.5. The estimated variance errors and the residual analysis were similar for all of the best models.

To illustrate a typical simulation from the best model structure, the simulation for Pizzicato Polka is shown in Fig. 3. For

the model used to generate this simulation, Pizzicato Polka was in the testing set and the other five musical selections in Table I were in the training set.

V. Discussion

By following the proposed methodology, the model structure described by (18) meets the first three model criteria: the measured emotional appraisals of the listeners are time varying, the musical features used in the model are time varying and represent musical properties that communicate emotion, and the model is estimated using emotional appraisals to musical selections representing a genre of music. To satisfy the fourth model criterion, a model needs to accurately simulate emotional appraisals to any musical selection from the genre of music. Because the average R^2 statistic for the best model structure is 78.4% for arousal and 21.9% for valence, this criterion is met for arousals but not for valences. Because there is potential to improve the R^2 statistic for valences by using different model structures, it appears that using the proposed methodology allows valid models to be created that satisfy the four model criteria.

There are several comments to be made about the parameters in (18)–(24). First, the parameters in matrices $B_0, B_1,$ and B_2 correspond to the contribution of each input (the columns) to each output (the rows). Because the sixth and eighth columns of these matrices are zero, the features SDiss (H&K) and TDiss (H&K) are not used in this model. Second, the structure of $A_1(\underline{\theta})$ and $A_2(\underline{\theta})$ indicates how previous emotional appraisals affect the current emotional appraisal. This model structure implies that a valence may be a function of an arousal, but an arousal can be calculated independently of a valence. This finding supports the hierarchical methodology of Liu *et al.* [5]. Finally, the number of parameters (38) has been chosen based on the performance of this model compared to other models with a different number of parameters. In this study, reducing the number of parameters typically increased the bias error (reducing the R^2 statistic), and increasing the number of parameters typically increased the variance error (increasing the FPE or the size of the confidence intervals).

A. Comparison With Other Research

It is difficult to quantitatively compare the models of emotional content in this study with the models created by Feng *et al.* [3], Li and Ogihara [4], and Liu *et al.* [5] because emotion is considered to be a discrete variable in these studies as opposed to a time-varying continuous variable. However, the model created in this paper can be considered an improvement over the models by Feng *et al.* [3] and Li and Ogihara [4] because there is no longer a need to ambiguously select a discrete number of emotions and the emotional content can vary with time. Similarly, this paper can be considered an extension of the paper by Liu *et al.* [5] because the four emotions used in their paper are analogous to the four quadrants of the 2-DES and thus can be further quantified using a continuous variable.

TABLE III
Comparison With Schubert's Models [2]

Musical Selection	Valence R^2 (%)		Arousal R^2 (%)	
	Schubert	Korhonen et al.	Schubert	Korhonen et al.
Pizzicato Polka	38	64	36	70
Peer Gynt (Morning)	40	26	67	71
Concierto de Aranjuez (Adagio)	33	-88	57	93

Schubert treats emotion as a time-varying continuous variable [2]. In Schubert's study, time-series models of emotional appraisals were created for Pizzicato Polka and longer versions of Peer Gynt (Morning) and Concierto de Aranjuez (Adagio). The R^2 values calculated for the individual musical selections modeled in both of theses studies are shown in Table III.

According to Table III, it appears that the arousal component of the model for the genre of classical music developed in Section IV is an improvement over Schubert's model for each individual musical selection. However, the valence component of the model has lower R^2 values than Schubert's models for Peer Gynt (Morning) and Concierto de Aranjuez (Adagio). There are several possible reasons for these lower values: shorter versions of these songs are used in this study so the R^2 statistic can only be used subjectively, Schubert evaluates the R^2 statistic using the training set so larger values are expected for his models, and the data are filtered differently in the two studies so different frequencies of the emotional content are emphasized. For these reasons, definite conclusions about the model "fit" cannot be made by comparing the R^2 statistic.

However, despite the differences in the two studies, one can conclude that principles of system identification afford mathematical models of emotion content of music that generalize to a genre of music. Valid models can be constructed, by applying the systematic method used in system identification for designing experiments, selecting model structures, and evaluating the models.

B. Applications

To apply this research to the field of music information retrieval, at least two possible approaches can be taken. First, the resultant model could be used to determine a distance between the emotion communicated by a musical selection and a given emotion (point) in the 2-DES. This would allow a person to search for music by a given emotion and be able to sort the results by distance. Second, by analyzing the simulated emotional appraisal of a musical selection, the variation in emotion can be measured to determine if the music constantly expresses one emotion or changes to express different emotions. Both methods of analysis would aid storing the emotional information content of music.

At least three methods exist to apply this research to the field of music psychology. First, the structure of the models could be analyzed to determine how particular musical features

communicate an emotion. Second, the assumption that emotional appraisals are consistent across cultures, music training, and other variables could be investigated. By applying this methodology to create models for subsets of the population (different cultures, musical training, music exposure, etc.), the differences between the models could be compared to determine if they are significantly different. Third, determining if an emotional content varies with a genre can be investigated. This could be done by comparing the performance of one model constructed for several genres of music to several different models that each represent a single genre.

C. Future Work

There are several suggested areas to investigate in future works. First, because the total duration of the music was limited to 20 min in this study, it is unlikely that an entire genre of music was represented. It would be worthwhile to evaluate this methodology with a larger selection of music and a greater number of subjects.

Second, modeling music containing lyrics has not been considered in this study. While measuring the emotional content of music with lyrics is possible using software such as EmotionSpace Lab, more features may be needed to create valid models for this music.

Third, in the models studied in this study, the dc value of all inputs and outputs was removed. Because there are some applications that would benefit from including the dc values in the model, it would be worthwhile to either 1) create a model without removing the dc values or 2) find a method to estimate the dc value of the outputs.

Also, the preliminary models could be improved using several different techniques. Sampling the emotional appraisals at a frequency higher than 1 Hz could improve the model performance. Including other features representing other musical properties, such as articulation, could also be included. Replacing the manually extracted tempo measurements with a reliable algorithm to measure tempos would make the application of models an automated process.

Other model structures could be used to get improved results. For example, this methodology could be applied to nonlinear models such as an artificial neural network. Alternatively, a separate model could be created for the arousal and the valence to allow treating an arousal as an input to a valence. Also, analyzing the differences between individual emotional appraisals collected with EmotionSpace Lab could lead to an improved noise model.

ACKNOWLEDGMENT

The authors would like to thank the members of the Vision and Image Processing (VIP) Group in Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada; the people who volunteered for this study; Laurier Centre for Music Therapy Research (LCMTR), Wilfrid Laurier University, Waterloo; the Office of Research Ethics, University of Waterloo; and E. Schubert for allowing the authors to use EmotionSpace Lab.

REFERENCES

[1] D. Huron, "Perceptual and cognitive applications in music information retrieval," in *Proc. Int. Symp. Music Information Retrieval (ISMIR)*, Plymouth, MA, 2000.

[2] E. Schubert, "Measurement and time series analysis of emotion in music," Ph.D. dissertation, School of Music & Music Education, Univ. New South Wales, Sydney, Australia, 1999.

[3] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR)*, Toronto, ON, Canada, Jul. 2003, pp. 375–376.

[4] T. Li and M. Ogihara, "Detecting emotion in music," in *Proc. 5th Int. Symp. Music Information Retrieval*, Baltimore, MD, 2003, pp. 239–240.

[5] D. Liu, L. Lu, and H. Zhang, "Automatic mood detection from acoustic music data," in *Proc. 5th Int. Symp. Music Information Retrieval*, Baltimore, MD, 2003, pp. 81–87.

[6] R. Kamien, *Music: An Appreciation*, 5th ed. New York: McGraw-Hill, 1992.

[7] A. Gabrielsson, "Perceived emotion and felt emotion: Same or different?," *Music. Sci.*, vol. Special Issue 2001–2002, no. Special Issue, 2001/2002, pp. 123–147, 2002.

[8] M. D. Korhonen, "Modeling continuous emotional appraisals of music using system identification," M.S. thesis, Syst. Des. Eng., Univ. Waterloo, ON, Canada, 2004.

[9] J. A. Russell, "Measures of emotion," in *Emotion: Theory Research and Experience*, vol. 4, R. Plutchik and H. Kellerman, Eds. New York: Academic, 1989, pp. 81–111.

[10] R. E. Thayer, *The Biopsychology of Mood and Arousal*. New York: Oxford Univ. Press, 1989.

[11] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "'FEELTRACE': An instrument for recording perceived emotion in real time," in *Proc. Speech and Emotion, ISCA Tutorial and Research Workshop (ITRW)*, Newcastle, U.K., Sep. 2000, pp. 19–24.

[12] E. Schubert, "Measuring emotion continuously: Validity and reliability of the two-dimensional emotion space," *Aust. J. Psychol.*, vol. 51, no. 3, pp. 154–165, Dec. 1999.

[13] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[14] D. Cabrera, *PsySound2: Psychoacoustical Software for Macintosh PPC*, Jul. 2000.

[15] G. Tzanetakis and P. Cook, "MARSYAS: A framework for audio analysis," *Organ. Sound*, vol. 4, no. 3, pp. 169–175, 2000.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[17] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, Jul. 2002.

[18] B. Porat, *A Course in Digital Signal Processing*. New York: Wiley, 1997.

[19] B. Tillman and E. Bigand, "Does formal musical structure affect perceptions of musical expression?," *Psychol. Music*, vol. 24, no. 1, pp. 3–17, 1996.

[20] M. D. Korhonen, D. A. Clausi, and M. E. Jernigan, "Modeling continuous emotion appraisals using system identification," in *Proc. 8th Int. Conf. Music Perception and Cognition*, Evanston, IL, Aug. 2004.

Mark D. Korhonen received the B.A.Sc. and M.A.Sc. degrees from the University of Waterloo, Waterloo, ON, Canada, in 2002 and 2004, respectively, all in systems design engineering.

He is currently employed as a Computer Engineer, programming automated testing equipment and performing digital signal processing at CIMTEK, Burlington, ON, Canada. His research interests include signal and image processing, machine intelligence, and pattern recognition.

David A. Clausi (S'93–M'96–SM'03) received the B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Waterloo, Waterloo, ON, Canada, in 1990, 1992, and 1996, respectively, all in systems design engineering.

In 1996, he worked in the medical-imaging field at Mitra Imaging Inc., Waterloo. He started his academic career in 1997 as an Assistant Professor in geomatics engineering at the University of Calgary, Calgary, AB, Canada. In 1999, he returned to the University of Waterloo and was awarded tenure and promotion to Associate Professor in 2003. He is an active Interdisciplinary and Multidisciplinary Researcher. He has an extensive publication record, publishing refereed journal and conference papers in the diverse fields of remote sensing, image processing, pattern recognition, algorithm design, and biomechanics. The research results have led to successful commercial implementations.

Dr. Clausi has received numerous graduate scholarships, conference paper awards, and a Teaching Excellence Award.

M. Ed Jernigan (M'76) received the B.S., M.S., and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1969, 1971, and 1975, respectively, all in electrical engineering.

In 1976, he joined the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada, where he is currently a Professor and the past Chair. He is a Distinguished Teacher of the University of Waterloo. His research interests include nonlinear and adaptive systems for signal and image processing, vision and machine perception, and pattern recognition, particularly with applications in medical imaging and remote sensing.