# Musical Composition by Regressional Mapping of Physiological Responses to Acoustic Features

D. J. Valtteri Wikström
Cognitive Brain Research Unit
University of Helsinki
and
Media Lab Helsinki
Aalto University
valtteri.wikstrom@aalto.fi

## ABSTRACT

In this paper an emotionally justified approach for controlling sound with physiology is presented. Measurements of listeners' physiology, while they are listening to recorded music of their own choosing, are used to create a regression model that predicts features extracted from music with the help of the listeners' physiological response patterns. This information can be used as a control signal to drive musical composition and synthesis of new sounds – an approach involving concatenative sound synthesis is suggested. An evaluation study was conducted to test the feasibility of the model. A multiple linear regression model and an artificial neural network model were evaluated against a constant regressor, or dummy model. The dummy model outperformed the other models in prediction accuracy, but the artificial neural network model achieved significant correlations between predictions and target values for many acoustic features.

## Keywords

music composition, biosignal, physiology, music information retrieval, concatenative synthesis, psychoacoustics

## 1. INTRODUCTION

The goal of this study is to explore the link between music and emotion through the effect that music has on the body of the listener, and to use that as a basis for a justifiable representation of emotions through sound. By creating a statistical model between the listeners physiological responses and the features of the sound, reconstruction of new sounds can be achieved directly, without the need of explicitly assigning emotional meaning to sound features. Earlier attempts to generate emotionally relevant music with physiological measurements have used a somewhat arbitrary mapping between the biosignals and the generative composition [6, 10]. Other approaches for music generation from biosignals, such as BioMuse [26], have treated the biosignals as control data for synthesis.

In this study, a regression model to predict acoustically analyzable musical information of songs from measurements of the listeners physiology during listening, on a second-by-second basis, is created and evaluated in a small experimental study. The current system works offline, but special attention has been given to using tools and methods that make real-time processing possible.

The generation of new sounds is achieved by using corpus-based concatenative sound synthesis, which is a method for matching a certain sequence of small sound units to create a new sound [25]. In concatenative sound synthesis, the corpus is a database of sound units that includes the descriptors for each unit. The descriptors can be extracted from the source signal, or external descriptors can be attributed to the sounds. The selection for the synthesised sequence of units is made by composing with the descriptors. The unit selection algorithm is used to find the best match according to certain criteria, such as distance and continuity. Using the regression output of this study as a composition source for driving the unit selection algorithm, with a corpus of sounds analysed for the same musical information, new sounds can be reconstructed as a representation of the users physiology. The resulting sound is a recreation of the physiological state of the subject, based on how sound has affected the subject before.

In a musical context this creates an elegant feedback loop – essentially the sound generated by the system expresses the users physiological state, which in turn is affected by the sound properties. Assuming that the model is imperfect and that factors external to the music can influence the user, the non-stationary sound generated by this system is an extension of the subjects involuntary expression as well as a self-reinforcing loop that could have interesting properties for example for music therapy. The reconstruction of sound and music can be approached in various ways in the future by using different kinds of data to create the regression model with physiology. For example higher level data, such as chords, harmonies and lyrics could be used, as well as synthesis parameters of a specific synthesiser or a sound-generating system.

## 2. BACKGROUND

William James argued in an essay from 1884 that emotions are separate from cognitive processes, and that the body acts as a mediator between the cognitive and emotional systems [11]. This theory was disputed, and the origins of emotions have been a hot topic in affective science ever since [4, 22, 17]. In any case, an idea first proposed by James, that emotions exist in the human physiology as distinguishable patterns, has garnered substantial evidence [1, 9, 18]. Music is considered to be capable of influencing the listener's emotions [19, 12]. Because music influences emotions, and emotions are recognisable from physiology, the assumption that a link exists between the content of music

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*NIME'14,* June 30 – July 03, 2014, Goldsmiths, University of London, UK.
Copyright remains with the author(s).

and the physiological responses of the listener is justifiable.

The range and causes of musical emotions are still under debate, and musical emotions have been described both in dimensional and categorical ways [23, 24, 27]. The field of music emotion recognition is concerned with the automatic detection of emotional content in music. A common way to approach this problem is to predict the emotion perceived and reported by the listener based on the content of a song or segment of a song. Studies have been able to predict emotional ratings of music from acoustic descriptors with regression models [24, 7]. Other studies have successfully predicted emotional content reported by the listener based on physiological measurements during the listening [13, 21].

## 3. RECREATING MUSIC WITH THE BODY

The purpose of the model explained in this section is to reconstruct sounds in a perceptually and emotionally meaningful way. The approach suggested here involves creating a regression model between musical content and physiology, so that new musical content can be predicted on the basis of learned relationships between music and the users physiology (Figure 1). The prediction result can then be used further to create a musical display of the psychophysiological state of the user.

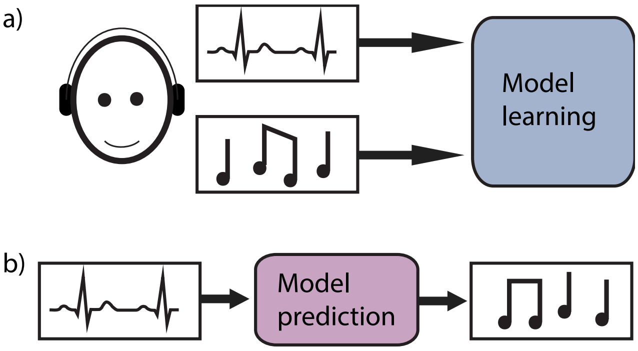

**Figure 1: The regression model functions in two phases. a) The model learns the relationships between the listeners physiology and the musical information. b) The model reconstructs musical information based on the users physiology.**

To approach the task of modelling the relationship between music and a listener's physiological responses, finding solutions for the analysis of audio and physiological data was necessary. The system was developed with the goal that all the components can be run in real time. At this point, the analysis itself was done offline, but the analysis methods afford for creating a fully online system. This required certain qualities from the signal processing algorithms, especially the ability to analyse subsequent chunks of data as they are measured, instead of the whole data set at once. Physiological variables have typically a several-second lag from the stimulus, and certain physiological components need a longer time window to be analysable [24, 14]. A compromise was made to analyse all the data with a sliding 10-second time window with a 1-second hop between subsequent data points, and 0, 1, 2, 3 and 4 second lags for the physiological variables. Continuous control data can still be created by interpolating between the regression result of subsequent seconds.

### 3.1 Real-time analysis of physiological data

The physiological measurements that were used in this study are electrocardiogram (ECG), electrodermal activation (EDA) and respiration inductance plethysmography (RIP). ECG measures the electrical activity of the heart muscle, making accurate heart beat detection possible. EDA, or skin conductance measures the activity of the sweat glands on the palm, representing sympathetic nervous system activity and the "fight-or-flight" response. RIP is measured with a band across the chest to quantify the subject's breathing.

An open source tool for the real-time analysis of physiological data could not be identified, so a decision was made to begin the development of a new software tool[1] for this purpose, using verified analysis algorithms found in literature [5, 2]. This tool is based on a modular structure, where the data stream from the data gathering device driver is buffered and piped to different analysis modules. The result is either stored to disk, visualised or sent to another application as Open Sound Control (OSC) messages (Figure 2).

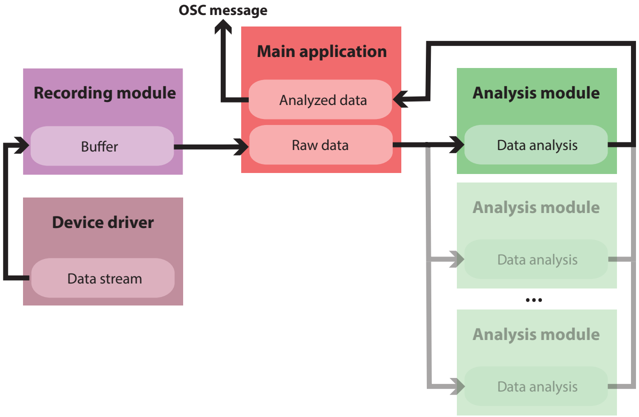

**Figure 2: The software tool for real-time analysis of physiological data has a modular architecture.**

In this study, a total of ten different variables were extracted from the physiological signals. From the EDA signal, the electrodermal level (EDL), the amplitude and the rise-time of the electrodermal response (EDR) were analysed. From the ECG signal, heart rate, the square root of the mean squared differences (RMSSD) of the heart-rate variability (HRV) time series, and the standard deviation of the HRV time series were calculated. The RIP signal was analysed for respiration rate and inhale and exhale durations.

### 3.2 Music information retrieval

The automatic analysis of music information retrieval (MIR) features was approached with an existing tool, MIRToolbox developed by Olivier Lartillot [16]. The current version of MIRToolbox does not support real-time analysis, but the variables chosen for this study are all analysable in a real-time setting as well. Other tools can be used in the future for real-time MIR analysis, such as Libxtract [3]. MIRToolbox was chosen at this point, as it runs in Matlab (The MathWorks Inc., Natick, MA), which was used also for regression modelling.

Eight variables were extracted from the audio signal. In MIRToolbox, root mean square (RMS) is calculated directly from the signal envelope. Fluctuation strength is calculated from the Mel-scale spectrum. Key strength and mode are estimated from the chromagram. Both the Mel-scale spectrum and chromagram are calculated from the frequency data which is derived from the fast Fourier transform (FFT).

[1] https://github.com/vatte/rt-bio

The FFT and signal envelope are used to calculate the spectral centroid. Tempo and pulse clarity are estimated in a complex way using a filterbank, onset detection and auto correlation. Structural novelty is assessed with a method that operates without the need of future events for determining the novelty value at a given time. [16, 15]

### 3.3 Regression model

The prediction of a MIR descriptor with the help of several physiological variables can be described as a many-to-one mapping problem. Because the underlying relationships in the data are for a large part unknown, it was decided that two very different models would be evaluated and compared.

Multiple linear regression (MLR) is a simple algorithm for regression with many input variables [20]. It solves a linear polynomial equation by minimising a loss function, such as the mean squared error. $\delta$-values for the physiological variables were used with the MLR model, representing the change between the current time window and the preceding time window of the same length. $\delta$-values for the MIR descriptors were evaluated as targets for both models.

MLR was chosen as a baseline model, to be compared with a more complex model, in this case a non-linear autoregressive exogenous (NARX) neural network model [8]. This type of model is especially suited for the prediction of longitudinal data, as it includes an autoregressive component. Both models were compared against a dummy model, a constant linear regressor created from the learning data.

## 4. EVALUATION STUDY

A small evaluation study was conducted to test the system and the feasibility of the regression model. Six test subjects participated in the experiment. Each participant brought four songs to the experiment with the instruction: ”Choose four songs that you would like to listen to right now”. Physiological data was recorded in a closed test room, where the test subject was sitting alone. The subject could choose to have the lights on or off, and whether they wanted to listen through loudspeakers or headphones. Physiological data was measured with the Nexus-10 MkII wireless bluetooth amplifier (MindMedia BV, Herten, Netherlands). Four songs chosen by another subject were added to the stack of stimuli, for a total of eight songs per subject. The justification for this was that when more data has been gathered, group wise comparisons of the effect on model performance of familiar versus unfamiliar songs can be made. Between the playback of each song, 2 minutes of silence was inserted, as well as in the beginning and the end of the measurement. The order of the songs was randomised.

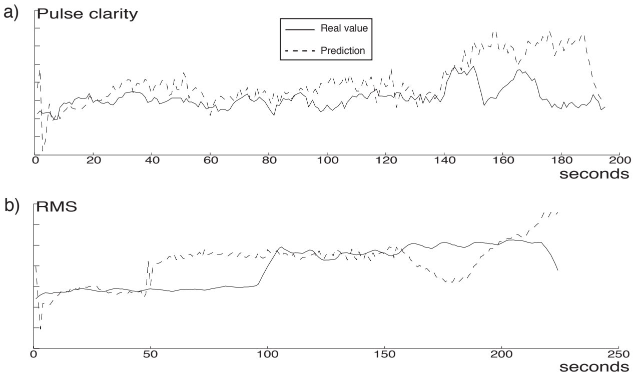

**Figure 3: Prediction results for the NARX model. a) Pulse clarity, subject 3 (song: Miguel - Adorn, $R^2$: 0.17, p-value:$1.5 \cdot 10^{-09}$). b) RMS, subject 7 (song: Placebo - Without You I’m Nothing, $R^2$: 0.26, p-value: $2.2 \cdot 10^{-16}$).**

The regression models were trained and tested individually for each subject. Feature selection was performed for the MLR model, with a forward selection paradigm. Both the MLR and the NARX models were trained with a ”leave-one-song-out” cross-validated paradigm, to avoid bias from using the same song in both training and testing. All the songs were truncated according to the length of the shortest song in the data set, so that each song would have an equal effect on the model.

**Table 1: Comparison of constant (const), multiple linear regression) MLR and NARX artificial neural network (ANN) models. The values that represents best performance in the different model evaluation tests are underlined for each MIR target. The MIR targets are divided into absolute values and $\delta$-values, which represent the change in subsequent time windows. The model evaluation parameters are mean square error (MSE), coefficient of determination ($R^2$) and the probability that there is no correlation between the prediction and the target value (p-value). Statistical significance (p-value $< 0.05$) is marked with a *.**

| MIR feature | MSE const | MSE MLR | MSE ANN | $R^2$ MLR | $R^2$ANN | p-value MLR | p-value ANN |
|---|---|---|---|---|---|---|---|
| RMS | $2.50 \cdot 10^{-5}$ | $2.61 \cdot 10^{-5}$ | $3.60 \cdot 10^{-5}$ | 0.0195 | 0.0293 | 0.051 | 0.010* |
| Fluctuation | $1.98 \cdot 10^{4}$ | $1.98 \cdot 10^{4}$ | $3.83 \cdot 10^{4}$ | 0.0138 | 0.0917 | 0.079 | 0.000* |
| Tempo | 175 | 191 | 440 | 0.0009 | 0.0399 | 0.677 | 0.003* |
| Pulse clarity | $4.53 \cdot 10^{-3}$ | $7.09 \cdot 10^{-3}$ | $3.47 \cdot 10^{-2}$ | 0.0164 | 0.1663 | 0.074 | 0.000* |
| Spectral centroid | $1.54 \cdot 10^{3}$ | $3.32 \cdot 10^{3}$ | $1.626 \cdot 10^{3}$ | 0.0471 | 0.1010 | 0.002* | 0.000* |
| Key strength | $8.87 \cdot 10^{-9}$ | $8.99 \cdot 10^{-9}$ | $4.23 \cdot 10^{-8}$ | 0.0014 | 0.0090 | 0.606 | 0.156 |
| Mode | $4.25 \cdot 10^{-4}$ | $4.51 \cdot 10^{-4}$ | $8.93 \cdot 10^{-3}$ | 0.0118 | 0.0191 | 0.188 | 0.052 |
| Novelty | 0.850 | 0.985 | 1.82 | 0.0166 | 0.0180 | 0.054 | 0.104 |
| $\delta$RMS | $3.07 \cdot 10^{-6}$ | $3.10 \cdot 10^{-6}$ | $8.40 \cdot 10^{-5}$ | 0.0147 | 0.0068 | 0.070 | 0.318 |
| $\delta$Fluctuation | $5.79 \cdot 10^{3}$ | $5.19 \cdot 10^{3}$ | $7.05 \cdot 10^{3}$ | 0.0455 | 0.0071 | 0.001* | 0.219 |
| $\delta$Tempo | 87.5 | 87.7 | 984 | 0.0048 | 0.0076 | 0.401 | 0.203 |
| $\delta$Pulse clarity | $2.01 \cdot 10^{-3}$ | $3.07 \cdot 10^{-3}$ | $3.29 \cdot 10^{-2}$ | 0.0034 | 0.0178 | 0.483 | 0.046* |
| $\delta$Spectral centroid | $3.79 \cdot 10^{2}$ | $7.21 \cdot 10^{2}$ | $2.15 \cdot 10^{3}$ | 0.0430 | 0.0120 | 0.011* | 0.184 |
| $\delta$Key strength | $1.94 \cdot 10^{-8}$ | $1.95 \cdot 10^{-8}$ | $7.02 \cdot 10^{-8}$ | 0.0004 | 0.0028 | 0.753 | 0.430 |
| $\delta$Mode | $4.24 \cdot 10^{-4}$ | – | $2.34 \cdot 10^{-2}$ | – | 0.0544 | – | 0.000* |
| $\delta$Novelty | 0.520 | 0.647 | 3.63 | 0.0028 | 0.0229 | 0.438 | 0.023* |

Median values for the mean squared prediction error and the p-value for a pairwise correlation test were used to evaluate the performance of the models. The evaluation results were mixed, as can be seen in Table 1. The dummy model outperformed both models in prediction accuracy, but the NARX model achieved significant correlations ($p < 0.05$) for fluctuation, tempo, pulse clarity, spectral centroid, $\delta$-pulse clarity, $\delta$-mode and $\delta$-novelty. The MLR model achieved significant correlations for spectral centroid, $\delta$-fluctuation and $\delta$-spectral centroid. Examples of the NARX models prediction can be seen in Figure 3.

## 5. CONCLUSIONS

A new approach for mapping physiological signals to acoustical variables has been presented in this paper. A regression model that learns in a music listening situation and predicts acoustical variables from physiological responses was constructed and evaluated. Currently, acoustical variables can be predicted with the regression model. The results of the evaluation study were promising, especially the neural network model achieved significant correlations for many of the acoustical variables. This can be seen as an indicator that non-linear models with autocorrelating capabilities are feasible for future development. To further assess the performance and feasibility of the regression model, a larger dataset needs to be collected.

Currently, individual models for each subject were created. In the future, the performance of individual models can be evaluated against an inter-subject-model. Also the effect of familiarity to the music listened to can be explored. Regarding the sound generating strategy, a user study can be made to collect the users' own impressions of the relevance of the generated sound to their emotions. Will the sounds and music created through regression be identifiable to the user as an auditory display of their emotions?

## 6. ACKNOWLEDGMENTS

I would like to thank Professor Mari Tervaniemi and the Finnish Centre of Excellence in Interdisciplinary Music Research for providing the opportunity and funding for carrying out this research. I would also like to thank Tommi Makkonen, Tommi Vatanen and Olivier Lartillot for their advise and support.

## 7. REFERENCES

[1] M. Arnold. Physiological differentiation of emotional states. *Psychological Review*, 52(1):35, 1945.

[2] W. Boucsein. *Electrodermal activity.* Springer, 2012.

[3] J. Bullock. Libxtract: A lightweight library for audio feature extraction. In *Proceedings of the International Computer Music Conference*, volume 43, 2007.

[4] W. Cannon. The James-Lange theory of emotions: A critical examination and an alternative theory. *The American Journal of Psychology*, 39(1):106–124, 1927.

[5] I. Christov. Real time electrocardiogram QRS detection using combined adaptive threshold. *BioMedical Engineering OnLine*, 3(1):28, 2004.

[6] N. Coghlan and R. B. Knapp. Sensory chairs: A system for biosignal research and performance. In *In Proceedings of 2008 Conference on New Instruments for Musical Expression (NIME08).*, 2008.

[7] E. Coutinho. Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion*, 11(4):921, 2011.

[8] E. Diaconescu. The use of NARX neural networks to predict chaotic time series. *WSEAS Transactions on Computer Research*, 3(3):182–191, 2008.

[9] P. Ekman, R. Levenson, and W. Friesen. Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210, 1983.

[10] S. Giraldo and R. Ramirez. Brain-activity-driven real-time music emotive control. In *Proceedings of the 3rd International Conference on Music & Emotion*, 2013.

[11] W. James. What is an emotion? *Mind*, 9(34):188–205, 1884.

[12] P. N. Juslin and P. Laukka. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33(3):217–238, 2004.

[13] J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067, 2008.

[14] S. Kreibig, G. Schaefer, and T. Brosch. Psychophysiological response patterning in emotion: Implications for affective computing. In K. Scherer, T. Bänziger, and E. Roesch, editors, *A Blueprint for Affective Computing: A sourcebook and manual.* Oxford University Press, 2010.

[15] O. Lartillot, D. Cereghetti, K. Eliard, and D. Grandjean. A simple, high-yield method for assessing structural novelty. In *Proceedings of the 3rd International Conference on Music & Emotion*, 2013.

[16] O. Lartillot and P. Toiviainen. A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects*, pages 237–244, 2007.

[17] R. Lazarus. *Emotion and adaptation.* Oxford University Press New York, 1991.

[18] R. Levenson. Autonomic nervous system differences among emotions. *Psychological science*, 3(1):23–27, 1992.

[19] L. Meyer. *Emotion and meaning in music.* University of Chicago Press, 1956.

[20] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning.* The MIT Press, 2012.

[21] F. Russo, N. Vempala, and G. Sandstrom. Predicting musically induced emotions from physiological inputs: linear and neural network models. *Frontiers in psychology*, 4, 2013.

[22] S. Schachter and J. Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, 69(5):379, 1962.

[23] K. Scherer. Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? *Journal of New Music Research*, 33(3):239–251, 2004.

[24] E. Schubert. Modeling perceived emotion with continuous musical features. *Music Perception*, 21(4):561–585, 2004.

[25] D. Schwarz. Corpus-based concatenative synthesis. *Signal Processing Magazine, IEEE*, 24(2):92–104, 2007.

[26] A. Tanaka. Musical performance practice on sensor-based instruments. *Trends in Gestural Control of Music*, 13:389–405, 2000.

[27] Y.-H. Yang and H. Chen. *Music Emotion Recognition.* CRC Press, 2011.