

Automatic Generation of Music for Inducing Physiological Response

Kristine Monteith (kristine.perry@gmail.com)

Department of Computer Science

Bruce Brown(bruce.brown@byu.edu)

Department of Psychology

Dan Ventura(ventura@cs.byu.edu) and Tony Martinez (martinez@cs.byu.edu)

Department of Computer Science

Brigham Young University

Provo, UT 84602 USA

Abstract

Music is known to have a profound impact on human cognitive and emotional response, which in turn are strongly correlated with physiological mechanisms. This paper presents a system that is designed to create original musical compositions that elicit particular physiological responses. The experiments described below demonstrate that the music generated by this system is as effective as human-composed music in effecting changes in skin resistance, skin temperature, breathing rate, and heart rate. The system is particularly adept at composing pieces that elicit target responses in individuals who demonstrated predictable responses to training selections.

Keywords: music; emotion; perception; cognition; physiological response; targeted response

Introduction

Music can have a profound impact on human physiology. It affects how we think, how we feel, and how we relate to others. It captivates and holds our attention, stimulating many areas of the brain. From movie scenes to dance floors, the added sensory input of music makes activities and situations more enjoyable and compelling. One study found that pleasurable music activated the same areas of the brain activated by other euphoric stimuli such as food, sex, or drugs. They highlight the significance of the fact that music would have a similar effect on the brain as “biologically relevant, survival-related stimuli” (Blood & Zatorre, 2001).

Music’s impact on human physiology may help explain its long-recognized ability to sway human emotion. It provides not only a medium for expressing a particular emotion, but also the accompanying physiological change to add significance and depth to that emotion. According to the Schachter-Singer theory, emotion is a function of both physiological arousal and cognitive interpretation of that response. The degree of arousal is associated with the degree of emotional response, but it is up to the individual to label that response according to past experience (Schachter & Singer, 1962).

Music can also have significant power to calm the body and mind. While relaxation responses such as lowered breathing and heart rate may not be as closely tied with emotional perception and cognition, their elicitation can often have significant therapeutic benefits. Numerous studies have demonstrated the ability of music to induce a relaxation response (e.g. White, 1999; Lepage et al., 2001; Khalfa et al., 2002).

Both speed and accuracy of task performance can be enhanced with relaxing music (Allen & Blascovich, 1994).

While there is little question about whether or not music has an effect on humans, predicting the precise effect is challenging. A few effects, however, do seem to be relatively consistent. For example, one study found that more complex rhythms tended to increase the rate of autonomic functions such as breathing and cardiovascular activity. Silence tended to have the opposite effect—lowering breathing rates and heart rates (Bernardi et al., 2006). White (1999) found that heart rate, respiratory rate, and myocardial oxygen demands were lower among patients recovering from myocardial infarctions; Khalfa et al. (2002) found that arousal responses were more likely with pieces that the subjects found to communicate happiness or fear, while pieces described as sad or peaceful tended to decrease arousal. However, even these results only hold true for a majority of individuals. Finding a piece of music that would reliably effect a desired physiological response in a given individual remains a considerable challenge.

Computer-generated music (Chuan & Chew, 2007; Cope, 2006) may provide some advantages in addressing this challenge. Computers are well-suited to sifting through a large number of both large-scale and fine-grained musical features and to keeping track of which features will most likely have a particular effect. Indeed, some work has been done in generating music to target a listener emotion or mood (Delgado et al., 2009; Rutherford & Wiggins, 2003; Oliveira & Cardoso, 2007). In addition, a human composer might be more biased towards features that would effect his or her own physiology when producing compositions. While a reliance on one’s own physiological experiences may be inspiring and helpful in the creative process, when it comes to eliciting physiological responses from others, it may also sometimes result in pieces that are less generalizable. Additionally, once they have “learned” how to do so, computers can generate large quantities of music at virtually no cost in terms of time or effort. A computer would have a much easier time generating a number of different potential compositions to effect a desired result in a given individual until it happened upon the right one. Therefore, the ability of a computer to compose music that elicits a target response could have significant benefits.

This paper presents a system capable of generating selections designed to elicit desired physiological responses. Data collected in biofeedback experiments with 96 different subjects show that the system is able to generate selections that elicit an average change in a target physiological response with roughly the same ability level as a human performing the same task. The system is particularly effective at eliciting such a response if an individual's response to other musical selections is known.

Methodology

Our approach can be decomposed into three major components: selection of musical pieces to use as training data for our generative models, construction of those models using the training data and evaluating the effectiveness of the models in eliciting the target response when compared to human-generated music designed for the same targeting task.

Training Data Selection

Seventy-two MIDI files were downloaded from the Free MIDI File Database.¹ Themes from movie soundtracks were used due to the wider variety of emotional content available in this genre. The first forty-five seconds of each piece was isolated for use in experiments.

Biofeedback experiments were conducted to determine effective candidate training pieces. In our preliminary experiments, forty-eight subjects were asked to listen to a number of different training pieces while their heart rate, breathing rate, skin resistance, and skin temperature were monitored. Physiological responses were recorded using the I-330-C2+ biofeedback machine manufactured by J&J Engineering. All were university-enrolled students or professors. Subjects ranged in age from 18 to 52, with the average age being 22. Thirty-four males and 14 females participated.

The seventy-two MIDI selections were split into six groups of twelve selections, and each group of songs was played for eight people.² At the beginning of experiments, forty-five seconds of baseline readings were collected. (Subjects were asked to sit quietly and count upwards in their minds during this time in order to achieve neutral results.) Measurements were sampled at one second intervals. For each of the physiological measures, responses were averaged for the duration of baseline readings and for the duration of each of the forty-five second song samples. Then, a z -score was calculated for each of the selections, indicating how many standard deviations the average for a given song varied from the baseline.

Responses were then analyzed to determine which selections were most likely to affect a given physiological response. A corpus of training songs comprised of the selections that elicited the largest average change in response was then created for each of the measures studied.

¹<http://themes.mididb.com/movies/>

²While the song grouping could likely have been randomly assigned without significantly affecting the results, an attempt was made to make the groupings as similar as possible.

Automatic Music Generation

Each generative model (one for each targeted response) is composed of four separate modules, for producing rhythm, pitch, harmony and accompaniment.

Rhythm Generator The rhythm for the selection is generated simply by selecting phrases from randomly chosen selections in the training set and stochastically perturbing them. Each new rhythmic phrase is evaluated by two decision tree Rhythm Evaluators (described below). Generated phrases are only used if they are classified positively by both classifiers.

Pitch Generator Once the rhythm is determined, pitches are selected for the melodic line using a probabilistic n -gram model of melodic progression built from the training corpus. The system generates one hundred possible series of pitches for each rhythmic phrase, and each of the melodies is evaluated by two decision tree Pitch Evaluators (see below). Generated melodies are only selected if they are classified positively by both classifiers.

Harmony Generator The underlying harmony is determined using a hidden Markov model, with melody notes considered as observed events and the chord progression as the latent state sequence. The probability distributions for populating the model are estimated using statistics gathered from the corpus of music representing the target response.

Accompaniment Planner To generate accompaniment, the system takes as input a measure from a song in the training corpus outlining a characteristic baseline, percussion track, and instrumentation. These act as style files for the computer-generated selections – each measure is transposed according to the generated chord pattern, producing accompaniments in much the same manner as a pianist selecting a given style on an electronic keyboard.

Decision Tree Evaluators A set of two evaluators is developed for interacting with the rhythm module and another set of two evaluators is developed for interaction with the pitch module. The first classifier in each set is trained using analyzed selections in the target corpus (e.g. raise heart rate) as positive training instances and analyzed selections from the other corpora (e.g. the other seven responses, lower heart rate, raise breathing rate, etc.) as negative instances. This is intended to help the system distinguish selections that elicit specific physiological response. The second classifier in each set is trained with melodies from all corpora versus thirty-two unevaluated melodies previously generated by the algorithm. In this way, the system learns to distinguish melodies which have already been accepted by human audiences. An example decision tree (identifying features for eliciting raised heart rate response) developed for evaluating the pitch assignment model is shown in Figure 1.

Evaluation

A second round of biofeedback experiments was conducted to evaluate the generated musical selections. Forty-eight ad-

Raise Heart Rate

ClimaxPosition ≤ 0.67
 — Dissonance ≤ 0.01
 — — PitchMovementByTonalStep ≤ 0.63 : No
 — — PitchMovementByTonalStep > 0.63 : Yes
 — Dissonance > 0.01 : Yes
 ClimaxPosition > 0.67 : No

Figure 1: Decision tree model of musical characteristics contributing to raised heart rate

ditional subjects participated in this evaluation phase. Again, all were university-enrolled students or professors. Subjects ranged in age from 17 to 46, with the average age being 22. Twenty males and 28 females participated.

Physiological responses were recorded for twenty-four songs (eight computer-generated selections, eight training selections for reference, and eight human-composed selections). To prevent subject fatigue, selections were divided into two groups, one group consisting of pieces targeted to affect breathing and heart rate and one group consisting of pieces targeted to affect skin resistance and skin temperature, and subjects were only asked to listen to one of the groups. Each subject listened to twelve selections; each piece was played for twenty-four people. A Cronbach’s alpha coefficient (Cronbach, 1951) was calculated on the responses of subjects in each group to test for inter-rater reliability. Coefficients for the two groups were both $\alpha = 0.99$. (Values over 0.80 are generally considered indicative of a reasonable level of reliability and consequently, a sufficient number of subjects for testing purposes.)

Baseline readings were collected at the beginning of each recording session. Responses were averaged for the duration of baseline readings and for the duration of each of the selections. Since some individuals were more reactive than others, z-scores are used in analysis instead of absolute changes in measurement.³

After listening to each selection, subjects were asked to respond to the following questions (on a scale from 1 to 9):

1. Did you like the selection?
2. How familiar were you with the selection?
3. How musical was the selection?
4. How original was the selection?

Results

This section provides tables reporting the average z-scores for selections designed to elicit the various target physiolog-

³Recall that z-scores calculate the number of standard deviations an average varies from a given baseline. They are calculated by the formula $z = (x - \mu) / \sigma$, where x is the average for a given selection, μ is the average for baseline, and σ is the standard deviation for readings taken over the duration of the session. Please note that, while z-scores are sometimes used to calculate statistical significance, in this case, these measures are only being used to normalize scores from one individual to the next. A high Cronbach’s alpha value for a low average z-score indicates that, while a given selection did not tend to elicit a high magnitude change in a response, it was consistent in eliciting a given change for a significant number of subjects.

Table 1: Average z-scores of computer and human-generated selections designed to affect breathing rate

Lower Breathing Rate			
	Overall	Adjusted	
		Average	Included
Computer-Generated	-0.27	-1.33	29%
Human-Composed	0.13	-0.90	29%

Raise Breathing Rate			
	Overall	Adjusted	
		Average	Included
Computer-Generated	0.71	1.18	46%
Human-Composed	0.06	0.36	46%

ical responses. In most cases, both the computer-generated and human-composed selections were effective at eliciting arousal responses. However, they were less effective at eliciting relaxation responses. This is not surprising considering findings suggesting that music is often more effective than silence at eliciting an arousal response (Bernardi et al., 2006).

Many of the more conclusive studies on the relaxing effects of music deal with subject-selected pieces. Since both the computer-generated and human-composed selections being evaluated are unique to these experiments, subjects would not associate any of them with previous relaxing experiences and consequently experience a relaxation response due to classical conditioning. It would also be difficult for any of the subjects to identify ahead of time which pieces they would find most relaxing. Instead, we look at how subjects responded to the training selections. Each table also reports an adjusted score, calculated by averaging only measurements for individuals for whom the training selections also had the target effect for the measure being considered (reported in the tables as a percentage of the 24 total people that listened to the selection). While a computer-generated piece may not be able to elicit a particular physiological response in all subjects, this adjusted score allows us to measure whether it elicits a response in a specific group of subjects. (e.g. If it is known that a group of individuals react with a lowered breathing rate to a given song or set of songs, the adjusted score reveals how effective the computer might be in using those training pieces to generate a song that also lowers breathing rate.)

Breathing Rate

Breathing rate responses tended to vary by up to one breath per minute. (Considering that normal human breathing rates tend to range from 12 to 18 breaths per minute, an average increase of one breath per minute is non-negligible.) The most significant changes tended towards an increase in breathing rates as compared to baseline.

As shown in Table 1, only the computer-generated selection was able to successfully lower breathing rate on the average for all subjects. However, the magnitude of the change was small enough that the average change was not significantly different from the human-composed selections. With the adjusted scores, both computer-generated and human-

composed songs were able to successfully lower breathing rates. Seven individuals—29% of subjects in this group—responded as expected to the top training selection for lower breathing rate; four responded similarly to the computer-generated selection.

The computer-generated song designed to raise breathing rate was able to accomplish this task more effectively than the human-composed song. The 0.71 z -score for the computer-generated song corresponds to an average increase of over one breath per minute, and the difference in average z -scores between this and the human-generated song was significant at the $p < 0.05$ level. A similar pattern is seen with the adjusted scores. The average difference between the computer-generated selection and the human-composed song was also significant. Nine of the eleven individuals who responded with elevated breathing rate to training selections targeted to raise breathing rate responded similarly to the computer-generated selections.

Note that the computer-generated selections designed to lower breathing rate are as effective at doing so as the human-composed selections. The computer-generated selections designed to raise breathing rate are performing this task at a level that exceeds that of human performance.

Heart Rate

Changes in average heart rate were not quite as pronounced. While individual heart rates could vary by up to fifty beats per minute over the course of a session, the average range for a given individual was only ten beats per minutes. When averaged over all subjects, reactions to songs only varied by a couple of beats per minute.

As shown in Table 2, only the human-composed selection was able to reduce average heart rate, although the difference in mean heart rate variation was not significant at the $p < 0.05$ level. With the adjusted scores, the computer-generated selection proved more effective at lowering heart rate. For five of the eight individuals whose heart rate lowered for the top training selection, heart rates were also lowered for the computer-generated songs in these categories.

The computer-generated song was the most effective at raising average heart rate for all subjects, though the difference was not statistically significant. The computer-generated song was also more effective at raising heart rate using the adjusted score, but not significantly so. Ten of the thirteen individuals who responded as expected to the training selection for raising heart rate also had their heart rates raised by the computer-generated selection.

As with breathing rate, the computer appears to be addressing the task of composing music that lowers or raises heart rate at a level comparable to that of human performance.

Skin Temperature

Skin temperature tended to rise, on average, by two degrees during the course of the session for most subjects, regardless of the piece of music being played. Not surprisingly, all selections were better at raising average skin temperature for all

Table 2: Average z -scores of computer and human-generated selections designed to affect heart rate

Lower Heart Rate			
	Overall	Adjusted	
		Average	Included
Computer-Generated	0.40	-0.40	33%
Human-Composed	-0.20	-0.61	33%
Raise Heart Rate			
	Overall	Adjusted	
		Average	Included
Computer-Generated	0.72	1.09	54%
Human-Composed	0.12	0.53	54%

Table 3: Average z -scores of computer and human-generated selections designed to affect skin temperature

Lower Skin Temperature			
	Overall	Adjusted	
		Average	Included
Computer-Generated	2.18	-1.22	17%
Human-Composed	1.23	-1.84	17%
Raise Skin Temperature			
	Overall	Adjusted	
		Average	Included
Computer-Generated	2.22	3.03	83%
Human-Composed	1.75	2.49	83%

subjects than they were at lowering it.

However, when individual subjects did have their skin temperature lowered by a training set selection, they also tended to have their skin temperature lowered by pieces generated from those selections. This was true for all four of the individuals whose temperature was lowered by the training selection targeting lower skin temperature. The adjusted score for the human-composed selection designed to lower skin temperature was lower than the adjusted score for the computer-generated piece, but the difference was not statistically significant at the $p < 0.05$ level.

The computer-generated piece was significantly more effective at raising skin temperature than the human-composed pieces when considering both the regular and the adjusted averages. However, this is almost certainly an artifact of the order in which the pieces were played. (The software used in these experiments did not allow for a randomized order of selection presentation that was unique to each subject.)

While it appears that an effective method of raising skin temperature would simply be composing a piece with sufficient duration, the computer seems as competent at the task as a human. Composing music that lowers skin temperature appears to be a much harder task, but again, these experiments show no statistically significant difference between the performance of the computer and the human.

Table 4: Average z-scores of computer and human-generated selections designed to affect skin resistance

Lower Skin Resistance			
	Overall	Adjusted	
		Average	Included
Computer-Generated	-0.87	-2.48	63%
Human-Composed	-1.06	-2.00	63%

Raise Skin Resistance			
	Overall	Adjusted	
		Average	Included
Computer-Generated	-1.06	2.27	33%
Human-Composed	-1.03	0.21	33%

Skin Resistance

Most of the selections were likely to elicit an arousal response (lower skin resistance). However, unlike skin temperature, the effect was not cumulative over the course of the session.

For compositions designed to lower skin resistance, there was no significant difference between the computer-generated selection and the human-generated selection. The training selections lowered skin resistance in fifteen individuals. With the adjusted scores, computer-generated selections were more successful at lowering skin resistance than the human-composed song, though the difference was not statistically significant.

There was also no significant difference between the computer-generated selection designed to raise skin resistance and the human-composed selection. The training selection raised skin resistance in eight individuals and those subjects for whom it did have the target effect also reacted strongly to the selection generated from all the training soundtracks, with the improvement over the human-composed selection being significant at the $p < 0.05$ level.

As with the other measures, the computer is able to generate music that elicits change in skin resistance as effectively or more effectively than a human composition.

Subjective Responses

Average responses to the subjective questions asked after each selection are shown in Table 5. Not surprisingly, the initial training selections and the human-composed selections received higher rating for likability and musicality. However, the computer-generated selections received slightly higher ratings for originality and significantly lower ratings for familiarity than the training selections and human-composed selections—evidence to suggest that the computer is producing genuinely original compositions and not borrowing too heavily from training data.

As shown in Table 6, there was no correlation between subjective responses and physiological changes. While for some individuals, liking a song might result in a more dramatic increase or decrease in a given physiological response, this does not appear to be the case overall.

Table 5: Average results to subjective questions (Responses were measured on a scale of 1 to 9)

Did you like the selection?	
Training Selections	5.83
Computer-Generated Selections	3.97
Human-Composed Selections	5.56

How familiar was the selection?	
Training Selections	5.53
Computer-Generated Selections	2.17
Human-Composed Selections	3.01

How musical was the selection?	
Training Selections	5.35
Computer-Generated Selections	3.88
Human-Composed Selections	5.12

How original was the selection?	
Training Selections	6.36
Computer-Generated Selections	6.97
Human-Composed Selections	6.70

Table 6: Correlations between subjective responses and physiological measures

	Like	Familiar	Musical	Original
Breathing Rate	0.02	0.03	0.03	-0.04
Heart Rate	0.03	-0.01	0.04	0.09
Skin Temperature	0.04	0.04	0.00	-0.06
Skin Resistance	-0.03	-0.05	-0.02	-0.03

Musical Features

Musical characteristics identified by the evaluating decision trees as being responsible for various physiological responses may be only briefly touched on here. Pieces that raised heart rates tended to have more dissonance and more scale-wise movement. Pieces that lowered heart rate, on the other hand, tended to have less rhythmic variety (perhaps contributing to more flowing rhythms) and a stronger climax.

Melodies that tended to raise breathing rate tended to higher rhythmic variety and either a non-tonal climax note or lower climax strength. Somewhat surprisingly, melodies that lowered breathing rate also tended to have higher rhythmic variety, but also some syncopation and a tendency to upward pitch direction.

Features contributing to a lowered skin temperature response included stability of melodic direction and a non-tonal climax. In other words, upward movement towards a climax that involved a non-tonal suspension note were arousing. A greater pitch range also contributed to lowered skin temperature. Pitch movement by minor tonal step leading to a strong climax tended to contribute to raised skin temperature.

Melodies that tended to lower skin resistance had lower pitch variety and less stability of melodic direction; some of these arousing melodies tended to bounce back and forth between notes. Melodies that raised skin resistance had a greater stability of melodic direction, as well as less rhythmic

Table 7: Ability to elicit arousal response via musical stimuli (RBR = raise breathing rate; RHR = raise heart rate; LST = lower skin temperature; LSR = lower skin resistance)

	RBR	RHR	LST	LSR
Computer-Generated	✓*	✓		✓
Human-Composed	✓	✓		✓
Computer-Generated (Adjusted)	✓*	✓	✓	✓
Human-Composed (Adjusted)	✓	✓	✓	✓

mic variety and range.

Conclusion

Tables 7 and 8 summarize how effective we were at eliciting a change in physiological responses in various situations.

Neither the computer-generated nor the human-composed selections were able to lower average skin temperature, but both computer-generated and human-composed selections designed to elicit the other arousal responses (raised breathing rate, raised heart rate, and lowered skin resistance) were, on average, able to do so successfully. In the case of breathing rate, the computer generated song was able to raise breathing rate more effectively than the human-composed song at a level that was significant (marked with asterisk).

When considering only subjects who responded as expected to the training selections, both the computer-generated and human composed songs were successful at eliciting an average arousal response for all of the measures studied. For breathing rate and skin resistance, the difference between the computer-generated selection and the human-composed selection was significant, the computer-generated one again being more effective at eliciting the target response.

Eliciting relaxation responses proved more challenging for both the computer-generated and human-composed selections. Both were able to raise skin temperature, but neither was able to raise skin resistance. Only the computer-generated selection was able to lower heart rate, and only the human-composed selection was able to lower breathing rate. The difference between the computer-generated and human-composed songs was not statistically significant.

When considering adjusted scores, both the computer-generated and human-composed selections were able to elicit all target relaxation responses. In the case of skin resistance, the computer-generated song was significantly better at raising average response.

Overall, the system proves itself able to generate songs that elicit target physiological responses with similar effectiveness to songs generated by a human composer. Both still require information about a given individual's physiological responses in order to generate a new piece that also reliably elicits those responses in many categories. However, given the variability of human biofeedback responses, the ability to consistently effect targeted physiological responses under any conditions can be viewed as fairly impressive.

Table 8: Ability to elicit relaxation response via musical stimuli (LBR = lower breathing rate; LHR = lower heart rate; RST = raise skin temperature; RSR = raise skin resistance)

	LBR	LHR	RST	RSR
Computer-Generated	✓		✓	
Human-Composed		✓	✓	
Computer-Generated (Adjusted)	✓	✓	✓	✓*
Human-Composed (Adjusted)	✓	✓	✓	✓

References

- Allen, K., & Blascovich, J. (1994). Effects of music on cardiovascular reactivity among surgeons. *Journal of the American Medical Association*, 272(11), 882-884.
- Bernardi, L., Porta, C., & Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart*, 92, 445-452.
- Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, 98(20), 11818-11823.
- Chuan, C., & Chew, E. (2007). A hybrid system for automatic generation of style-specific accompaniment. In *Proceedings of the International Joint Workshop on Computational Creativity* (p. 57-64).
- Cope, D. (2006). *Computer Models of Musical Creativity*. Cambridge, Massachusetts: The MIT Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Delgado, M., Fajardo, W., & Molina-Solana, M. (2009). In-mamusys: Intelligent multi-agent music system. *Expert Systems with Applications*, 36(3-1), 4574-4580.
- Khalifa, S., Peretz, I., Blondin, J., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, 328(2), 145-149.
- Lepage, C., Drolet, P., Girard, M., Grenier, Y., & DeGagne, R. (2001). Music decreases sedative requirements during spinal anesthesia. *Anesthesia-Analgesia*, 93, 912-916.
- Oliveira, A., & Cardoso, A. (2007). Towards affective-physiological foundations for music production. In *Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction* (p. 511-522).
- Rutherford, J., & Wiggins, G. (2003). An experiment in the automatic creation of music which has specific emotional content. In *Proceedings of MOSART Workshop on Current Research Directions in Computer Music* (p. 35-40).
- Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69, 379-399.
- White, J. M. (1999). Effects of relaxing music on cardiac autonomic balance and anxiety after acute myocardial infarction. *American Journal of Critical Care*, 8(4), 220-230.