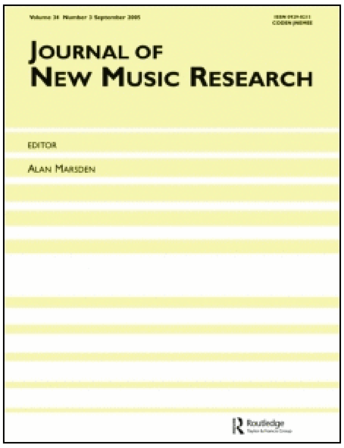

Journal of New Music Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713817838

Automated Music Emotion Recognition: A Systematic Evaluation

Arefin Huq[a]; Juan Pablo Bello[b]; Robert Rowe[b]
[a] Sourcetone, USA [b] Music and Audio Research Laboratory (MARL), NYU, USA

Online publication date: 09 November 2010

Journal of New Music Research
2010, Vol. 39, No. 3, pp. 227–244

Automated Music Emotion Recognition: A Systematic Evaluation

Arefin Huq[1], Juan Pablo Bello[2], and Robert Rowe[2]

[1]Sourcetone, USA; [2]Music and Audio Research Laboratory (MARL), NYU, USA

Abstract

Automated music emotion recognition (MER) is a challenging task in Music Information Retrieval (MIR) with wide-ranging applications. Some recent studies pose MER as a continuous regression problem in the Arousal-Valence (AV) plane. These consist of variations on a common architecture having a universal model of emotional response, a common repertoire of low-level audio features, a bag-of-frames approach to audio analysis, and relatively small data sets. These approaches achieve some success at MER and suggest that further improvements are possible with current technology.

Our contribution to the state of the art is to examine just how far one can go within this framework, and to investigate what the limitations of this framework are. We present the results of a systematic study conducted in an attempt to maximize the prediction performance of an automated MER system using the architecture described. We begin with a carefully constructed data set, emphasizing quality over quantity. We address affect induction rather than affect attribution. We consider a variety of algorithms at each stage of the training process, from preprocessing to feature selection and model selection, and we report the results of extensive testing.

We found that: (1) none of the variations we considered leads to a substantial improvement in performance, which we present as evidence of a limit on what is achievable under this architecture, and (2) the size of the small data sets that are commonly used in the MER literature limits the possibility of improving the set of features used in MER due to the phenomenon of Subset Selection Bias. We conclude with some proposals for advancing the state of the art.

1. Introduction

The capacity of music to induce an emotional response in its listeners is one of its most appreciated and least well-understood properties. In this paper we discuss the design of an automated system to recognize emotion in music and the issues that arise in implementing such a system. The potential uses of a music emotion recognition (MER) system are manifold, from traditional applications such as managing personal music collections and enhancing music recommendation systems, to promising new applications in music therapy for treating emotional disorders and improving the performance and well-being of healthy individuals. This study was undertaken in the context of the latter interest, and was specifically targeted to improving the quality of emotion predictions made on an online streaming music service devoted to music and health, www.sourcetone.com.

Therefore, our study is motivated by the following use case: *Given an emotion, automatically retrieve a recording from a music library that will induce that emotion in a listener*.

Recordings often vary in their emotion content over their duration, and detecting such variation is an area of active research. However since we want to retrieve entire tracks and not segments of tracks, our system will assign an overall emotion to each track and ignore variation within the track.

We require our system to associate an emotion to each recording in the library using only the audio signal. In particular our system does not use track metadata or song lyrics. There are several reasons for this constraint. We want our system to be self-contained so that it can operate offline, for example embedded in a portable music player. We want our system to be able to handle new or original music for which there is no metadata. Furthermore, from

Correspondence: Arefin Huq, Sourcetone, 5525 N Winthrop Ave, Chicago, 60640-1487, USA. E-mail: fig@arefin.net

DOI: 10.1080/09298215.2010.513733

a research perspective we are interested in exploring the connection between information in the audio stream and the corresponding subjective response to the music.

2. General approach and related research

A summary and comparison of this research to related research appears in Table 1. In this paper the terms *mood* and *emotion* are used interchangeably. The term *track* refers to a particular audio recording of a piece of music. A *clip* is an excerpt from a track. In our study only one clip is selected from each track, however some of the studies cited use multiple clips from a given track.

2.1 Emotion space

Several basic choices must be made in designing an MER system. Perhaps the most significant is how to model the space of emotions. One approach enumerates the emotions of interest and treats MER as a multi-class classification problem (Lu, Liu, & Zhang, 2006; Hu, Downie, Laurier, Bay, & Ehmann, 2008). Another approach treats MER as a multi-label classification problem (Li & Ogihara, 2003; Skowronek, McKinney, & van de Par, 2007; Trohidis, Tsoumakas, Kalliris, & Vlahavas, 2008), where a given track can belong to more than one emotion category. We instead follow the practice of studies such as Yang, Lin, Su, and Chen (2008), by adopting Thayer's arousal-valence (AV) emotion plane (Thayer, 1989) and treating MER as a continuous regression problem in two dimensions. The AV plane is shown along with our data set in Figure 1. Both Valence and Arousal lie in the range $[-1,1]$. The upper left quadrant contains high energy negative emotions like anger, while the lower right quadrant contains low energy positive emotions like contentment. Similarly the upper right and lower left quadrants contain emotions like ecstasy and sadness, respectively. More details about the data set can be found in Section 4.1.

The regression approach is desirable for several reasons. It avoids ambiguities with the words used to label emotions and avoids the artificial division of emotion into discrete categories. It naturally handles gradations of emotion: two emotion points can be 'close' or 'far' whereas two emotion classes are either the same or different. Regressors also compute the desired numeric output directly; binary classifiers must be adapted for multi-class classification, for example by adding a separate 1-vs.-1 or 1-vs.-all layer. An added benefit of the continuous approach is in the application domain: it makes it possible to generate a playlist that transitions smoothly from one emotion to another by following a path in the AV plane.

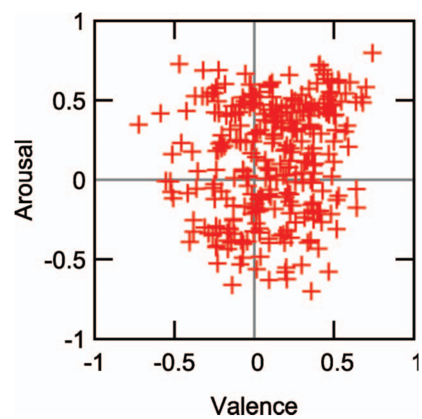

Fig. 1. Thayer's emotion plane along with the 'ground truth' emotion labels for the 288 clips in our data set. Each point is the average of the overall subject responses for the corresponding clip as described in Section 4.1.

Table 1. Summary and comparison of related research. 'Artificial Set' means that tracks were chosen manually to have clear or dominant emotions. 'Artificial Clips' means the extracted segment of each track was chosen manually to have a clear or dominant emotion. 'Total Ratings' = (Num. Clips) × (Raters Per Clip).

	Emotion Space	Artificial		Num. Clips	Raters Per Clip	Total Ratings	Learner
		Set	Clips				
2007 MIREX[1]	multi-class	Yes	No	600	< 8	≪4800	n/a
Leman et al. (2005)	continuous[2]	Yes	Yes	60	8	480	Linear
Li & Ogihara (2003)	multi-label	No	No	499	1	499	SVM
Lu et al. (2006)	multi-class	Yes	Yes[3]	800	3 experts	2400	GMM
MacDorman & Ho (2007)	continuous	No	No	100	85	8500	Various
Schubert (2004)[4]	continuous	Yes	n/a	4	67	268	Linear Autoregressive
Skowronek et al. (2007)	multi-label	Yes	Yes	1059	6	6354	Quadratic Discriminant Analysis
Trohidis et al. (2008)	multi-label	Yes	No	593	3 experts	1779	Various
Yang et al. (2008)	continuous	Yes	Yes	195	> 10	> 1950	Linear, SVR, AdaBoost.RT
This paper	continuous	No	No	288	50	14,400	Various

Notes: [1]Hu et al. (2008). [2]The original emotion label space is also transformed. [3]An automatic segmentation algorithm is also presented. [4]Treats emotion *variation* detection more than MER.

Other possibilities exist for modelling the space of emotions in music. In Zentner, Grandjean, and Scherer (2008) the authors present evidence that emotion models such as the AV model that are drawn from non-musical domains may not be well-suited to representing the nuances of music-induced emotion. In that paper the authors propose an alternative domain-specific model of emotion in music. For this study we have nonetheless favoured the regression approach because it allows us to extend the research cited above, and because (as described earlier) for our intended application we are interested in a framework that allows us to relate music-induced emotion to emotion and well-being in everyday life.

2.2 Affect induction

Another choice is whether to model affect *attribution* or *induction* (Leman, Vermeulen, De Voogdt, Moelants, & Lesaffre, 2005)—does the system predict the emotion that the music is thought to convey or does the system predict the emotion that is actually induced in the listener? As observed in Leman et al. (2005), Schubert (2004), and Skowronek et al. (2007), there is generally more agreement and less noise for affect attribution, and most of the research cited deals (implicitly or explicitly) with this task. However the difference between attributed and induced emotion is not merely a difference in variance but may include more complex relationships, as is discussed in detail in Evans and Schubert (2008) and Schubert (2007).

Our data set measures affect induction. Subjects in the study were specifically asked to indicate what they were feeling as they were listening to the music. We made this choice because we are particularly interested in the effects of music on health through affect induction; if the emotion is not induced, the effect we are interested in is not present.

2.3 Universal model

Another choice is how to treat the 'subjectivity issue': the evaluation of mood in music is a subjective matter, and one listener's rating will differ from another's. Li and Ogihara (2003) sidestep this issue and use the ratings of a single person. Other studies, for example Lu et al. (2006), Trohidis et al. (2008), and Skowronek et al. (2007), restrict their analysis to tracks or clips for which there is general agreement. Leman et al. (2005) instead creates user-specific models on top of a general model.

We assume, along with Lu et al. (2006) and Yang et al. (2008), that underlying the individual responses is a *universal* model of emotional response that characterizes how a typical listener will respond, and we attempt to determine this model. We reduce the effect of subjectivity (and measurement noise) by collecting and averaging the responses of approximately 50 subjects for each clip.

2.4 Data set quality

Data set quality comprises three factors: the number of ratings per clip, the number of clips in the data set, and the setting in which the data is collected.

Schubert (2004) uses 67 ratings per recording, however only four recordings are considered, and the models constructed are designed to predict the *variation* of emotion within a given recording. MacDorman and Ho (2007) have 85 ratings per clip, however the data set consists of only 100 clips and the data was collected through a self-administered online survey. All the other studies cited use fewer than a dozen ratings per clip.

Our data set consists of 50 ratings per clip for 288 clips, which as shown in Table 1 comprises the largest total number of ratings of the studies surveyed. Our data was collected in person in a controlled experimental setting as described in Bachorik et al. (2009). Collecting data in this manner is a time-consuming and labour-intensive process, which limits the feasibility of growing the number of tracks and reflects our decision to emphasize quality over quantity in constructing the data set.

2.5 Real world scenario

Some cited research, for example Lu et al. (2006) and Hu et al. (2008), attempts to simplify the MER task by selecting only tracks that contain a clear emotion for which there is subjective agreement. Other studies, for example Yang et al. (2008), manually extract excerpts or segments from tracks so that those excerpts contain a single dominant emotion. Table 1 indicates whether the set of tracks or the clips from each track were selected artificially for each study.

We avoid simplifications such as these because we want to construct a system that can robustly and automatically handle the real-world scenario of a music library that contains a variety of tracks, some with ambiguous or varying emotion, and for which there could be substantial subjective disagreement. Our track library was chosen to cover a variety of genres and styles, and was constructed without an artificial attempt to select for particular emotions. Each clip consists of a 1 min segment taken from the 30 s mark to the 90 s mark in each track.

2.6 Feature selection and model space

One of the primary goals of this study is to identify features and models that are most useful for MER within the framework just described.

We do not know *a priori* which audio features will be useful for MER so our strategy is to construct a large superset of potentially useful audio features and then rely on (1) dimensionality reduction (in this case PCA), (2) feature selection algorithms, and (3) in some cases the model itself to reduce the number of features and select the most useful. In particular we focus on feature selection and consider four feature selection algorithms (FSAs) in addition to no feature selection. None of the other research surveyed considers FSAs, except for Yang et al. (2008) which uses only one method.

We attempt to cover the space of possible models by applying five algorithms that represent fundamentally different ways of approaching regression. Each model is considered in combination with each FSA.

2.7 Summary

In this paper we seek to maximize the performance of an MER system based on the regression approach and a universal model of emotional response to music, using track excerpts representative of a real-world music library. Our study differs from prior research in the following respects:

- explicit focus on affect induction;
- quality of the data set, as measured by the total number of track ratings and the setting in which they were collected;
- emphasis on feature selection and consideration of several feature selection algorithms;
- systematic exploration of the space of possible architectures by considering combinations of preprocessing algorithms, FSAs, and models.

Our system is meant to represent the current state of the art in MER. Our goal is to determine the best performance achievable with this system, along with the choices of preprocessing algorithms, features, and models that maximize performance. As shown later, our findings suggest limits as to what this architecture can achieve. Furthermore, the phenomenon of Subset Selection Bias prevents us from identifying individual features that are useful for MER—a problem that applies not just to our data set but in principle to any comparably sized data set.

3. System description

The training process for our system is shown in the top of Figure 2. It takes as input a set of digital audio files and subjective AV labels for the corresponding clips, and produces as output a regressor. During Model Independent Audio Analysis each clip is reduced to a vector of real-valued features which are the input to Regressor Training. The bottom of Figure 2 shows how the resulting system would be used ‘in the field’ to label new tracks.

The correlation of Valence and Arousal in our training data was 0.1602, which we felt was low enough to justify treating them separately based on similar findings in Yang et al. (2008). Thus the process is run twice to produce separate regressors for Valence and

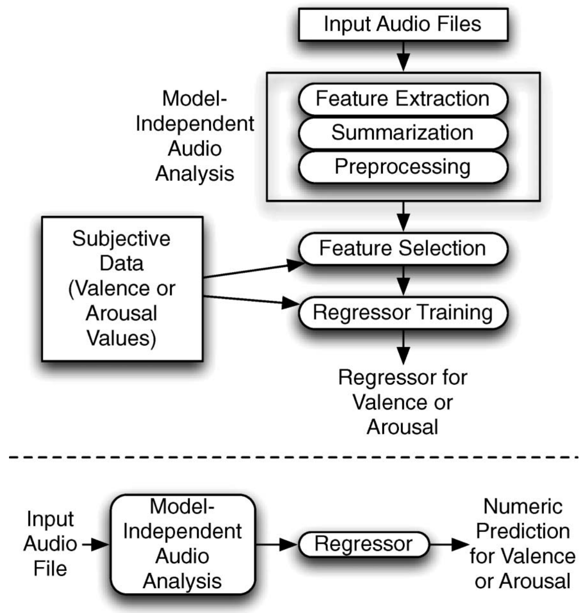

Fig. 2. *Top*: regressor training process. *Bottom*: use of a regressor in deployment.

Arousal. The Model-Independent Audio Analysis is identical for both and only run once.

3.1 Model independent audio analysis

Peeters (2004) documents a large set of audio features for Music Information Retrieval (MIR). Table 2 lists the audio features we used, which are taken as representative of the current repertoire of features in the MIR literature. The Timbral, Loudness, and Complex-Domain Onset Detection Function (*CDF*) features are computed over frames of approximately 23 ms. The Harmonic features are computed over frames of approximately 0.75 s. *Beat Histogram* (*BH*) and *Onset Rate* (*OR*) are computed for the entire clip from the values of *CDF*.

We employ the *MeanVar* approach as for example in Meng, Ahrendt, Larsen, and Hansen (2007) to reduce frame-level features to clip-level features: each feature is summarized by its mean and variance over all the analysis frames. *BH* and *OR* are already clip-level features and are not summarized. *TriadSeq* is discrete-valued so it is summarized by its statistical mode and variance. Although *MeanVar* is a simple ‘bag-of-frames’ summarization technique that loses a lot of signal information, Joder, Essid, and Richard (2009) find that it performs as well as or better than more complex summarization techniques when applied to the task of musical instrument recognition.

After summarization each clip is represented by a 160-dimensional vector. The range of these dimensions varies widely, and many features (for example adjacent MFCC coefficients) are correlated. Thus in addition to the raw (unscaled) features we consider: the effect of standardizing the features to have zero mean and unit variance along

Table 2. Features extracted from the audio signal.

Feature Name	Description
Timbral	
MFCC (20)	20 Mel Frequency Cepstral Coeffs.
HFC	High Frequency Content
SC, SS	Spectral Centroid, Spectral Spread
ZCR	Zero Crossing Rate
Loudness	
Sones (24)	Perceptual loudness in 24 bands
OL, RMS	Overall Loudness, Root Mean Square
Harmonic	
Chroma (12)	Pitch Class Profiles
TS (6)	Mapping onto 6-d Tonal Space (Harte, Sandler, & Gasser, 2006)
CDTS	Cosine Distance between consecutive feature vectors in Tonal Space
TriadSeq	Estimated major/minor Triad Sequence (Bello & Pickens, 2005)
TriadInt	Distance between consecutive triads as unitary steps in the perimeter of the circle of fifths
Rhythmic	
CDF	Complex-Domain Onset Detection Function (Bello, Duxbury, Davies, & Sandler, 2004)
BH (15)	Beat Histogram (15 BPM ranges) (Tzanetakis & Cook, 2002)
OR	Onset Rate

each dimension, the effect of performing Principal Components Analysis (PCA) on the inputs to produce an input space of de-correlated features, and PCA with Whitening, where the PCA components are again rescaled to unit variance.

3.2 Feature selection algorithms (FSAs)

In our system, Feature Selection is used to select a subset of the 160 features to improve regressor performance.

Filter methods select features using a model-independent heuristic. We investigate two such methods: Regressional Relief-F (RReliefF) (Robnik-Šikonja & Kononenko, 2003) and Correlation-based Feature Selection (CFS) (Hall, 2000). RReliefF ranks each feature based on how well it separates points of varying output values. CFS attempts to discover sets of features that have low correlation with each other but high correlation with the output.

Wrapper methods invoke the target learning algorithm to evaluate a feature set. We try two simple wrapper methods: Forward Selection, which starts with no features and successively adds features as long as the performance improves, and Backward Selection, which starts with all features and successively removes features as long as performance improves. Note that when a wrapper approach is used, the Feature Selection block in Figure 2 actually contains within it a module for Regressor Training and performance evaluation.

3.3 Regression

We survey a variety of algorithms that embody different approaches to numeric prediction.

- *Linear Regression (LR)*—a standard statistical technique that assumes a linear relationship between input and output variables and minimizes the least squares error.
- *Regression Tree (RT)*—an adaptation of decision trees where each leaf node is a numeric value rather than a class label.
- *Locally Weighted Regression (LWR-SLR)*—a 'lazy' algorithm that, when presented with a test sample, constructs a one factor linear model (Simple Linear Regression) on-the-fly based on nearby training points.
- *Model Tree (M5P)*—like Regression Tree except each leaf node is a linear model instead of a numeric value.
- *Support Vector Regression (SVR-RBF)*—Support Vector Machine for regression as implemented in LIBSVM (Chang & Lin, 2001) using the Radial Basis Function (RBF) kernel. This model contains a number of parameters that must be optimized during training with a parameter search.
- *Support Vector Regression with No Parameter Search (SVR-RBF-NP)*—same as SVR-RBF except the parameter values are hardcoded to sensible defaults. SVR-RBF-NP was used in place of SVR-RBF in the case of Forward and Backward Selection, where the nested parameter search inside the feature selection loop would not have been computationally practical.

4. Experimental setup

4.1 Data set

The training data (see Figure 1) is from a series of experiments described in Bachorik et al. (2009). 150 participants (84 females and 66 males) were recruited from the greater Boston metropolitan area via advertisements in daily newspapers. Participants ranged from 18 to 83 years of age (median age = 31). All participants completed a questionnaire regarding their musical background (including whether or not they had previously studied a musical instrument, for how long, how intensely they practiced, and whether they had taken music theory courses) and musical preferences (what music they liked to listen to now and what music they grew up with, etc.). Participants were compensated for their participation.

The music database consists of 60 s clips taken from 288 commercially available tracks in a variety of genres.

A complete listing of the tracks is given in the Appendix. Approximately 50 participants listened to each clip. Participants gave an overall rating to each clip after listening to it by indicating a point in the AV plane corresponding to how they felt. The 'ground truth' for each clip is the average of this overall rating for the 50 or so subjects that listened to that clip.

The mean and standard deviation in each dimension are listed in Table 3. The distribution of values for Valence and Arousal are shown in Figures 3 and 4.

4.2 Performance measure

Some earlier studies (Schubert, 2004; Yang et al., 2008) use R^2 for measuring MER performance. We use Mean Absolute Error (MAE) rather than R^2 for several reasons: Absolute Error has a natural interpretation in terms of the AV plane, as the distance in the relevant dimension between the desired and predicted values, Absolute Error can be computed for a single point whereas R^2 only has meaning for a collection of points, the implementation of the SVR we used minimizes Absolute Error by default, and finally the calculation of R^2 that is included in some reference software packages is numerically unstable and can produce spurious results (Chan, Golub, & Leveque, 1983).

Table 3. Summary statistics of the 288-clip AV training set and performance of the Baseline predictors. The standard deviations of both Baseline predictors were negligible.

	Valence	Arousal
Mean	0.111	0.133
Standard deviation	0.284	0.351
Mean deviation	0.235	0.302
Baseline MAE	0.236	0.305

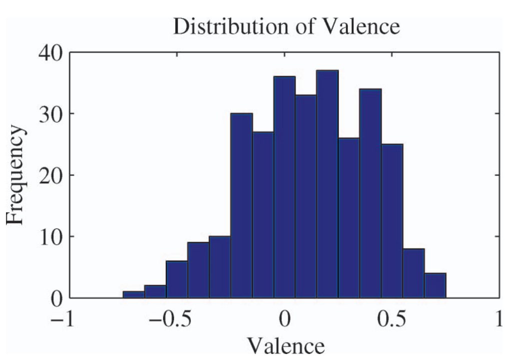

Fig. 3. Distribution of ground-truth Valence in the data set.

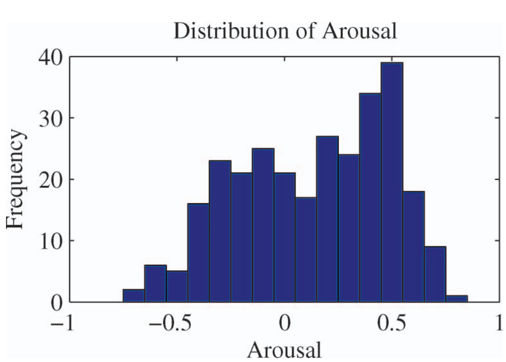

Fig. 4. Distribution of ground-truth Arousal in the data set.

4.3 Validation

We found that our data set was too small to use separate training and test sets, both because of the negative bias of using significantly less data in training and because of the variance of the performance measure across multiple runs. To counteract both of these problems we use 10 × 10-fold Cross-Validation (CV) as in Bouckaert (2003). 10 × 10 CV is used not just for reporting the final performance but internally for selecting feature subsets using a wrapper method and selecting parameters for SVR-RBF during parameter search. The fact that 10 × 10 CV is embedded inside loops contributes dramatically to the computational requirements of our experiments.

We also report the standard deviation (SD) of the errors across the 10 runs, where the error for each run is averaged over its 10 constituent folds. We do not report confidence intervals because the predictors constructed during 10 × 10 CV are not independent: on average, any two of these will have training sets with approximately 80% of the training instances in common. This phenomenon is discussed in Dietterich (1998), and McNemar's test is proposed as an alternative when the target variable is nominal. Although some techniques have been proposed to handle the continuous case (see e.g. Bouckaert, 2003), we know of no widely accepted method that can be used to report confidence intervals when the independence assumption is violated and the target variable is continuous. Further discussion of this issue can be found in Micheals and Boult (2001).

4.4 Baseline

Our baseline predictor simply predicts the mean value of the training labels. Table 3 lists the baseline performance.

There is a positive bias in the training set both for Valence and Arousal. The baseline predictor, if trained on the entire training set, would predict 0.111 for Valence for every clip and 0.133 for Arousal for every clip, and the MAE would then be the Mean Deviation, or 0.235 for Valence and 0.302 for Arousal. The reported baseline MAE is slightly higher than this because it was computed using 10 × 10 CV.

Notice how the choice of baseline predictor is dependent on the choice of performance measure. Our baseline, which predicts a constant value regardless of the input, has an R^2 of 0 (the worst possible). Using MAE, the baseline turns out to be a competitive predictor, particularly for Valence.

Since our baseline predictor returns the same value regardless of the input clip, our baseline MAE is indicative of having no predictive power. At the other end of the spectrum, a perfect predictor would have an MAE of 0. It follows that an MAE near the baseline suggests that the corresponding model has little ability to predict emotion whereas an MAE close to zero would be considered strong evidence for the model having learned the target concept.

4.5 Implementation

Much of our experimental setup, including preprocessing, feature selection, model training, and validation, was implemented in Weka (Witten & Frank, 2005) and RapidMiner (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006).

5. Results

In this section we discuss our experimental results and the implications for preprocessing strategy, feature selection algorithm, and model selection.

5.1 Preprocessing

Table 4 shows the results of using all features with no preprocessing (Raw), with Standardization (Std), with PCA, and PCA with Whitening (PCA-W).

Table 4. Mean Absolute Error (MAE) from 10 × 10 CV using all features and different preprocessing methods. All standard deviations were below 0.010.

	Raw	Std	PCA	PCA-W
Valence				
Baseline	0.236	–	–	–
LR	0.296	0.298	0.329	0.324
RT	0.217	0.217	0.235	0.235
LWR-SLR	0.222	0.222	0.230	0.230
M5P	0.225	0.224	0.231	0.231
SVR-RBF	0.236	0.198	0.198	0.217
Arousal				
Baseline	0.305	–	–	–
LR	0.201	0.203	0.221	0.224
RT	0.208	0.208	0.216	0.216
LWR-SLR	0.221	0.221	0.207	0.207
M5P	0.177	0.177	0.190	0.190
SVR-RBF	0.304	0.156	0.157	0.206

Without Standardization SVR performs at baseline, and with Standardization SVR is the best performer. The other algorithms are not affected by Standardization. We conclude that Standardization is a requirement for SVR, and does not hurt other regressors (and may improve numerical stability). We experimented separately with Normalization and found the effect to be indistinguishable from Standardization.

PCA-W produces results identical to PCA for RT, LWR-SLR, and M5P, since those algorithms are scale-invariant and PCA-W is simply a rescaling of the PCA features. In only one case (LWR-SLR for Arousal) did PCA or PCA-W improve the performance over the Std features. In most cases performance was degraded noticeably. We ran the same comparison using Std, PCA, and PCA-W on the features computed from RReliefF and found similar results.

We conclude that PCA or PCA-W does not improve performance on this feature set. The remaining sections of this paper use only the Standardized features.

5.2 FSAs

We begin our analysis of FSAs by running the algorithms on the entire data set to select a single set of features that is then used by the regressors in training. Using the entire data set in this way introduces the possibility of a type of bias known as Subset Selection Bias (SSB) (Miller, 2002). The bias is a form of overfitting that occurs because the FSA gets to examine all of the data before it is split into training and testing folds.

The reason for running these experiments in this way, with the risk of bias, is that with 10 × 10 CV the biased experiments run about 100 times faster than the unbiased experiments, where the FSA must be run once for each fold. In our implementation a typical unbiased experiment takes on the order of 1200 CPU-hours to run (compared to 12 CPU-hours for a typical biased experiment), so this is a serious practical consideration. The results of the biased experiments can be used to identify cases where the FSA may lead to improvement, and the corresponding unbiased experiment can be run to verify this hypothesis.

We discuss the effect of SSB later in this section. The biased FSA experiments are labelled with the prefix 'B-' (B-RReliefF, B-CFS, etc.) to indicate the possible bias. A summary of the results using the biased FSAs on the Standardized features is given in Table 5.

5.2.1 All features

The first column of Table 5, 'All', refers to using all the features, i.e. no feature selection. With all features, LR performs considerably worse than Baseline for Valence. This is indicative of overfitting: using all 160 features leads to a linear model with 161 parameters for only 288 points.

Table 5. Summary of results using standardized features with varying regressors and *Biased* FSAs. Values are estimates of Mean Absolute Error from 10 × 10 CV. Standard deviations were below 0.010 in all cases, and were below 0.005 in all cases except for LR (and LWR-SLR with B-RReliefF for Arousal).

	All	B-RReliefF	B-CFS	B-Forward	B-Backward
Valence					
Baseline	0.236	–	–	–	–
Linear Regression (LR)	0.298	0.223	0.206	0.189	0.295
Regression Tree (RT)	0.217	0.222	0.219	0.201	0.216
LWR-SLR	0.222	0.227	0.217	0.214	0.217
Model Tree (M5P)	0.224	0.216	0.210	0.201	0.223
SVR-RBF	0.198	–	–	–	–
SVR-RBF-NP	0.201	0.209	0.216	0.191	0.201
Arousal					
Baseline	0.305	–	–	–	–
Linear Regression (LR)	0.203	0.188	0.156	0.147	0.195
Regression Tree (RT)	0.208	0.197	0.199	0.191	0.208
LWR-SLR	0.221	0.227	0.212	0.210	0.217
Model Tree (M5P)	0.177	0.175	0.156	0.148	0.178
SVR-RBF	0.156	–	–	–	–
SVR-RBF-NP	0.157	0.179	0.164	0.142	0.155

In all other cases the models do somewhat better than Baseline for Valence and considerably better for Arousal (even LR).

5.2.2 Filter methods

We ran B-RReliefF using 10 nearest neighbours and without weighting by distance. We eliminated all features that were assigned a negative weight by B-RReliefF, resulting in 63 (out of 160) features being selected for Valence and 74 being selected for Arousal. As seen in Table 5, using B-RReliefF did not help in general. The notable exception is LR for Valence, which was improved by 0.075, presumably by alleviating the model overfitting issue. Even though B-RReliefF did not help performance in general, it did not seem to hurt it either. Throwing away about half the features made little difference in the performance of our regressors (this holds true even after correcting for SSB).

We ran B-CFS using Forward Selection as the subset search algorithm, though we found that the choice of search algorithm for B-CFS made little difference. B-CFS selected only 10 features for Valence and only 15 for Arousal. In the case of Valence, B-CFS improved the performance of LR dramatically (presumably by alleviating the model overfitting issue even further), produced a noticeable improvement in M5P, and made little difference for the other algorithms. In the case of Arousal, B-CFS again improves LR substantially, and appears to improve every other algorithm as well. The remarkable fact is that B-CFS produces dramatic improvements (or at worst only a slight degradation in performance) while using less than 10% of the original features. This fact also holds after correcting for SSB.

5.2.3 Wrapper methods

B-Forward added between four and 24 features depending on the regressor. The number of features added was roughly related to the improvement achieved over using all features: LR added 15 features and achieved considerable improvement and LWR-SLR added four features with little improvement, with RT and M5P in between. B-Forward did not improve SVR-RBF-NP as much since SVR-RBF-NP already had good performance using all features, but SVR-RBF-NP with B-Forward did produce the best (or tied for best) results of all the models.

B-Backward only removed between one and seven features and did not improve performance compared to using all features. Backward Selection seems prone to early termination: removing features generally only produces incremental improvements that are likely to be washed out by the variance in the performance estimates. Given the poor performance of B-Backward we did not run experiments to correct for SSB in Backward Selection.

Wrapper FSAs are extremely computationally intensive: just to select the first feature for addition or removal involves trying all 160 features, which with 10 × 10 CV means 16,000 invocations of the target algorithm.

5.2.4 SSB

Based on the biased FSAs, Forward Selection paired with LR and SVR-RBF-NP appear to be strong candidates. On this evidence we opted to run the corresponding unbiased experiments. We also ran unbiased versions of the RReliefF and CFS experiments, since these are not wrapper methods and therefore incur less overhead. The versions of these experiments that are corrected for SSB are prefixed with 'C-' and their results are shown in Table 6.

The results for C-RReliefF and C-CFS are not substantially worse than their corresponding biased versions, and in some cases appear slightly better due to random fluctuations. On the other hand, the results for C-Forward for LR and SVR-RBF-NP are considerably worse than the biased versions. It seems that most of the apparent improvement in performance from these FSAs was the result of SSB.

None of the other research surveyed used feature selection, except for Yang et al. (2008) which reports a slight improvement from using RReliefF. Trohidis et al. (2008) observes however that this improvement may have been subject to SSB.

We conclude that feature selection does not improve the performance of our system, due to SSB. The occurrence of SSB is a consequence of having a relatively small data set. There are more sophisticated feature selection algorithms that we did not consider, however it is even more likely that these algorithms would be subject to SSB. Although feature selection does not improve performance, a method such as CFS could possibly be used to *match* the performance of our system while using many fewer features and thus reducing the computational requirements.

5.2.5 *Best features*

An intended outcome of this study was to identify candidate audio features or sets of features that are most useful for MER. Due to the phenomenon of SSB, we found that FSAs tend to overfit the data and therefore the particular set of features that are selected in a given fold of the experiment cannot be generalized to an overall useful set of features.

Table 6. Summary of results using standardized features with varying regressors and *Corrected* FSAs. Values are estimates of Mean Absolute Error from 10 × 10 CV.

	All	C-RRF	C-CFS	C-Fwd
Valence				
Baseline	0.236	–	–	–
LR	0.298	0.223	0.216	0.219
RT	0.217	0.225	0.220	–
LWR-SLR	0.222	0.223	0.215	–
M5P	0.224	0.221	0.220	–
SVR-RBF	0.198	–	–	–
SVR-RBF-NP	0.201	0.214	0.219	0.215
Arousal				
Baseline	0.305	–	–	–
LR	0.203	0.191	0.161	0.168
RT	0.208	0.199	0.204	–
LWR-SLR	0.221	0.227	0.213	–
M5P	0.177	0.177	0.162	–
SVR-RBF	0.156	–	–	–
SVR-RBF-NP	0.157	0.184	0.169	0.162

5.3 Models

Once we correct for SSB, we find that the best performance overall for both Valence and Arousal is achieved by SVR-RBF using all features with standardization, as shown in Table 7. In this table we also present R^2 to facilitate comparison with other research. It is worth noting that although we used MAE and not R^2 in developing our system, we found that ranking performance by R^2 (descending) gave a nearly identical ordering to the top systems as ranking by MAE (ascending).

SVR-RBF appears remarkably robust in the presence of irrelevant and redundant features. Even with 160 features and only 288 points, SVR does not seem prone to overfit.

5.4 Summary of results

We presented an extensive evaluation of the different stages of training an automated music emotion recognition system with the regression approach. We found that standardization or normalization of the input is necessary and that PCA or PCA with Whitening does not help. We examined several FSAs and found that Subset Selection Bias is a serious risk on a data set of this size. Therefore the data set is too small to allow effective reduction of the feature space using standard feature selection methods.

SVR with the RBF kernel, using all standardized features and a coarse-grid search for the best parameters, was the best performing regressor for both Valence and Arousal. The performance for Valence is better than Baseline, however the R^2 of about 25% shows that much of the variation in Valence is not accounted for by our model. The situation with Arousal is much better: almost 70% of the variance is accounted for by the best model. The fact that Arousal is predicted better than Valence is in accordance with previous results, such as Yang et al. (2008) and Schubert (2004).

Table 7. MAE and R^2 achieved by the best regressors (SVR-RBF with parameter search and using all standardized features).

	MAE	R^2
Valence	0.198 ± 0.001	25.8% ± 5.6%
Arousal	0.156 ± 0.002	69.7% ± 7.1%

6. Conclusion and future work

The best performance achieved by our system, shown in Table 7, is in our opinion comparable to the level of performance reported elsewhere in the literature, to the extent that such a comparison can be made across data sets. The fact that we were not able to improve performance noticeably, even across a wide range of regression algorithms, feature sets, and preprocessing strategies, leads us to believe that these numbers are indicative of a limit of what is achievable under this general architecture.

We enumerate the limitations of this architecture as: (1) the small data set relative to the requirements of the model, (2) the set of fairly low-level musical features used, (3) the *MeanVar* bag-of-frames model which ignores the temporal aspects of the signal, and (4) the assumption of a universal model of emotional response to music.

The choice of the continuous AV plane as the representation space for emotion in music may itself also impose a limit on the system, as suggested in Zentner et al. (2008), however we do not have evidence from our study to either support or refute this possibility.

The relatively small data set is problematic for several reasons, all of which are related to overfitting: the data set is too small relative to the capacity of the learning algorithms (i.e. larger data sets lead to improved performance), the variance of performance estimates is high, requiring 10×10 CV which dramatically expands the computational requirements, and finally because it leads to Subset Selection Bias. SSB in turn is problematic because it makes it difficult to improve on the set or repertoire of features, for example by eliminating irrelevant or redundant features, or by allowing new candidate features to be evaluated effectively. This is the case even though the evidence from using CFS and RReliefF suggests that most of the features we used are not helpful in predicting emotion: SSB prevents us from separating the ones that are useful from the ones that are not.

Our experiment design favoured quality of data over quantity, and so makes it prohibitively expensive to expand the data set to the order of magnitude that we believe would be required to resolve these problems. Put simply, our emphasis on quality over quantity did not pay off.

Regarding the first issue (small data set), future research will likely depend on alternative ways of collecting large sets of mood labels for tracks in a scalable manner. One such approach using online games is described in Kim, Schmidt, and Emelle (2008). For our part, we intend to collect much larger quantities of data from user interactions with the Sourcetone streaming radio service. A discussion of alternative approaches to collecting tags for music in general can be found in Turnbull, Barrington, and Lanckriet (2008).

Regarding the second and third issues (improving musical features and using temporal information), we found in this study that automated methods are not sufficient to separate useful features from irrelevant ones due to Subset Selection Bias. Even if larger data sets are available we suspect that this 'bottom-up' approach is not powerful enough to lead to discovery of new features, especially for Valence. We expect that if progress is made in this area, it will be guided by research in human music cognition that can identify higher level features and musical structures that both relate to the induction of musical emotion and that can be computed automatically.

To illustrate the fourth issue (universal model of emotional response), we include a scatter plot of individual subject responses for a typical clip in Figure 5 (here 'typical' means close to the median in terms of standard deviation of subject responses for both Valence and Arousal). As the figure shows, individual subjects vary widely in their response. The 'ground truth' for the universal model is marked as a star, and as can be seen, its meaning in the context of the individual responses is questionable. In light of this issue we are interested in exploring the possibility of personalized music emotion recognition: to what extent can the variation of response to a given clip be accounted for by a system that models individual differences by considering characteristics of that person or by considering the response of that individual to other clips?

In summary, we have presented evidence that the current state of automated music emotion recognition as investigated in this paper is near the limit of what it can achieve, and that improvements will come not from variations on the current design but from advances in each of the following: (1) techniques for efficiently gathering large data sets of emotion labels for tracks, (2) higher-level music features—particularly those aimed at the harmonic structure and guided by new or existing

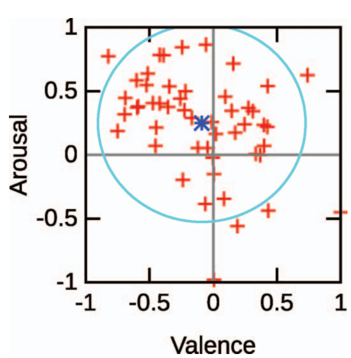

Fig. 5. Distribution of individual subject responses for the clip *Wir – Ticking Mouth*, illustrating the typical level of dispersion. The mean response, which is the value that would be used in training our universal model, is marked with a blue star. The ellipse indicates two standard deviations from the mean in each dimension.

research on human music cognition, (3) models of temporal evolution of music, and (4) personalization—systems that are trained to predict the different emotional responses of individuals or groups with varying backgrounds, tastes and characteristics.

Acknowledgements

This research was funded by Sourcetone, LLC. We wish to thank in particular Malcolm Goodman, President, and Dr Jeff Berger, Executive Vice President. The data was collected by our colleagues at Beth Israel Deaconess Medical Center (a teaching hospital of Harvard University), led by Dr Gottfried Schlaug. Their research was also funded by Sourcetone.

References

Bachorik, J.P., Bangert, M., Loui, P., Larke, K., Berger, J., Rowe, R., & Schlaug, G. (2009). Emotion in motion: Investigating the time-course of emotional judgments of musical stimuli. *Music Perception*, *26*(4), 355–364.

Bello, J.P., Duxbury, C., Davies, M.E., & Sandler, M.B. (2004). On the use of phase and energy for musical onset detection in the complex domain. *IEEE Signal Processing Letters*, *11*(6), 533–556.

Bello, J.P., & Pickens, J. (2005, 11–15 September). A robust mid-level representation for harmonic content in music signals. In *Proceedings of the 6th International Conference on Music Information Retrieval ISMIR 2005*, London, UK, pp. 304–311.

Bouckaert, R.R. (2003, 21–24 August). Choosing between two learning algorithms based on calibrated tests. In *Proceedings of the 20th International Conference on Machine Learning (ICML 2003)*, Washington DC, USA, pp. 51–58.

Chan, T.F., Golub, G.H., & Leveque, R.L. (1983). Algorithms for computing the sample variance. *The American Statistician*, *37*(3), 242–247.

Chang, C.C., & Lin, C.J. (2001). *LIBSVM: A library for support vector machines*. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*, 1895–1923.

Evans, P., & Schubert, E. (2008). Relationships between expressed and felt emotions in music. *Musicae Scientiae*, *12*(1), 75–100.

Hall, M. (2000, 29 June–2 July). Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, Stanford, CA, USA, pp. 359–366.

Harte, C., Sandler, M., & Gasser, M. (2006, 27 October). Detecting harmonic change in musical audio. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, Santa Barbara, CA, pp. 21–26.

Hu, X., Downie, S.J., Laurier, C., Bay, M., & Ehmann, A.F. (2008, 14–18 September). The 2007 MIREX audio mood classification task: lessons learned. In *Proceedings of the 9th International Conference on Music Information Retrieval ISMIR 2008*, Philadelphia, PA, USA, pp. 462–467.

Joder, C., Essid, S., & Richard, G. (2009). Temporal integration for audio classification with application to musical instrument classification. *IEEE Transactions on Audio, Speech and Language Processing*, *17*(1), 174–186.

Kim, Y., Schmidt, E., & Emelle, L. (2008, 14–18 September). Moodswings: A collaborative game for music mood label collection. In *Proceedings of the 9th International Conference on Music Information Retrieval ISMIR 2008*, Philadelphia, PA, USA, pp. 231–236.

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesaffre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, *34*(1), 39–67.

Li, T., & Ogihara, M. (2003, 27–30 October). Detecting emotion in music. In *Proceedings of the International Symposium on Music Information Retrieval ISMIR 2003*, Baltimore, MD, USA, pp. 239–240.

Lu, L., Liu, D., & Zhang, H.J. (2006). Automatic mood detection and tracking of music audio signals. *IEEE Transactions on Audio, Speech and Language Processing*, *14*(1), 5–18.

MacDorman, S., & Ho, S. (2007). Automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison. *Journal of New Music Research*, *36*(4), 281–299.

Meng, A., Ahrendt, P., Larsen, J., & Hansen, L.K. (2007). Temporal feature integration for music genre classification. *IEEE Transactions on Audio, Speech and Language Processing*, *15*(5), 1654–1664.

Micheals, R.J., & Boult, T.E. (2001). Efficient evaluation of classification and recognition systems. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*, 50–57.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006, 20–23 August). YALE: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, USA, pp. 935–940.

Miller, A. (2002). *Subset Selection in Regression* (2nd ed.). London: Chapman & Hall.

Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification) in the CUIDADO project* (Technical Report). Paris: IRCAM.

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of relieff and rrelieff. *Machine Learning*, *53*(1), 23–69.

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

Schubert, E. (2007). The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music*, *35*(3), 499–515.

Skowronek, J., Mckinney, M., & van de Par, S. (2007, 23–27 September). A demonstrator for automatic music mood estimation. In *Proceedings of the International Symposium on Music Information Retrieval*, Vienna, Austria, pp. 345–346.

Thayer, R.E. (1989). *The Biopsychology of Mood and Arousal*. Oxford: Oxford University Press.

Trohidis, K., Tsoumakas, G., Kalliris, G., & Vlahavas, I. (2008, 14–18 September). Multilabel classification of music into emotions. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, Philadelphia, PA, USA, pp. 325–330.

Turnbull, D., Barrington, L., & Lanckriet, G. (2008, 14–18 September). Five approaches to collecting tags for music. In *Proceedings of the 9th International Conference on Music Information Retrieval ISMIR 2008*, Philadelphia, PA, USA, pp. 225–230.

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293–302.

Witten, I., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.). New York: Morgan Kaufman.

Yang, Y.H., Lin, Y.C., Su, Y.F., & Chen, H.H. (2008). A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech and Language Processing*, *16*(2), 448–457.

Zentner, M., Grandjean, D., & Scherer, K.R. (2008). Emotions evoked by the sound of music: Differentiation, classification, and measurement. *Emotion*, *8*, 494–521.

Appendix. Listing of clips in data set

Artist	Title	Year	Genre
213	*Keep It Gangsta*	2004	Rap & Hip-Hop
Ad Vanz vs. Gescom	*Viral*	1998	Electronic
Adham Shaikh	*Ethereal Ion*	1995	Ambient
Adham Shaikh & Tim Floyd	*Embers*	1996	Ambient
Aerosmith	*What It Takes*	1989	Rock & Pop
Alanis Morrisette	*You Oughta Know*	1995	Rock & Pop
Alan Jackson	*Dont Rock The Jukebox*	1991	Country
Al Di Meola	*Splendido Sundance*	1980	Jazz
Alexander Kowalski	*Voices*	2001	Electronic
Alicia Keys	*Troubles*	2001	Rock & Pop
Amon Tobin	*Nightlife*	1998	Electronic
Amon Tobin	*Switch*	1998	Jazz
Andrew Thomas	*Fearsome Jewel*	2003	Ambient
Anindo Chatterjee	*Tal Nasruk*	1998	World
Anthony Manning	*Untitled 6*	1995	Ambient
Anthony Rother	*Human Made*	1997	Electronic
Aphex Twin	*Blue Calx*	1994	Ambient
Aquanote	*Nowhere (Speakeasy Remix)*	2002	Electronic
Aril Brikha	*Setting Sun*	1999	Electronic
Arovane	*Torn*	2002	Electronic
Autechre	*Eutow*	1995	Electronic
Bach	*Cello Suite No. 1 in G Major: Prelude*		Classical
Bach	*Prelude No. 21 in B-flat Major*		Classical
Bach	*Prelude No. 22 in B-flat minor*		Classical
Bandulu	*Run Run*	1994	Electronic
Barber	*Adagio for Strings*	1936	Classical
Barbra Streisand	*The Way We Were*	1973	Rock & Pop
Bartok	*The Miraculous Mandarin*	1964	Classical
Bebel Gilberto	*Tanto Tempo*	2000	World
Beethoven	*Piano Sonata No. 7 in D Major: Largo e mesto*		Classical
Beethoven	*Piano Sonata No. 7 in D Major: Presto*		Classical
Beethoven	*Symphony No. 9: Adagio molto e cantabile*		Classical
Benjamin Iobst	*Sound Meditation 1*	1999	Electronic
Bette Midler	*Wind Beneath My Wings*	1988	Rock & Pop
Biosphere	*Iberia Eterea*	2000	Ambient
Biosphere	*Poa Alpina*	1997	Ambient
Black Sabbath	*Planet Caravan*	1970	Rock & Pop
Boards Of Canada	*Everything You Do Is A Balloon*	1996	Electronic
Bob Marley	*Burnin' and Lootin'*	1997	Electronic
Brenton Wood	*The Oogum Boogum Song*	1967	Funk/Soul
Brian Eno & Harold Budd	*An Arc Of Doves*	1992	Ambient
Brian Eno & Harold Budd	*First Light*	1992	Electronic
Brooks & Dunn	*My Maria*	1996	Country
Bruckner	*Symphony No. 5: Finale*		Classical
Bryan Adams	*Everything I Do*	1991	Rock & Pop
B.T. Express	*Do It*	1974	Funk/Soul
Buddy Rich	*Standing Up In A Hammock*	1967	Jazz
Burger/Ink	*Flesh & Bleed*	1996	Electronic
Caia	*Jericho*	2003	Ambient
Camp Lo	*Black Nostaljack (Instrumental)*	1997	Rap & Hip-Hop
C + C Music Factory	*Gonna Make You Sweat*	1990	Rock & Pop
Capella Gregoriana	*Graduale: Miserere Passion (2)*		Classical
Carl Craig	*At Les*	1997	Electronic
Carl Craig	*Dreamland*	1997	Electronic
Carl Craig	*Red Lights*	1997	Electronic

(*continued*)

Appendix. (*Continued*).

Artist	Title	Year	Genre
Carlos Nakai	*Cleft In The Sky*	1988	World
Collective Soul	*December*	1995	Rock & Pop
Common	*Real People*	2005	Rap & Hip-Hop
Coolio feat. E-40,KAM, & 40 Thevz	*Exercise Yo' Game*	1995	Rap & Hip-Hop
Craig Pruess & Ananda	*Ganesh Invocation*	2002	World
Curtis Mayfield	*Make Me Believe In You*	1974	Funk/Soul
Curve	*Falling Free (Aphex Twin Remix)*	1992	Electronic
Das EFX	*They Want EFX*	1992	Rap & Hip-Hop
Dave Brubeck	*Take Five*	1961	Jazz
Dazz Band	*Keep It Live*	1982	Funk/Soul
Debussy	*Reflets dans l'eau*	1905	Classical
Depeche Mode	*Higher Love*	1993	Rock & Pop
Detroit Escalator Co.	*Theta*	1996	Ambient
Diana Krall	*When I Look In Your Eyes*	1999	Jazz
Digable Planets	*Black Ego*	1994	Rap & Hip-Hop
Dionne Warwick	*Thats What Friends Are For*	1985	Rock & Pop
Dire Straits	*Money For Nothing*	1984	Rock & Pop
DJ Sneak	*Play It Again*	1996	Electronic
Donna Summer	*Last Dance*	1978	Rock & Pop
Dots	*Dense*	1994	Electronic
Dvorak	*Slavonic Dance No. 1 in C Major (Furiant)*		Classical
Dynamo	*Außen Vor (Arovane Amx)*	1997	Electronic
Eddie Kendricks	*Goin' Up In Smoke*	1998	Funk/Soul
El Chicano	*Viva Tirado*	1970	Funk/Soul
Ella Fitzgerald &Louis Armstrong	*Under A Blanket Of Blue*	1957	Jazz
Enigma	*Mea Culpa*	1990	Ambient
Enya	*Book of Days*	1992	Rock & Pop
Fatback	*Up Against The Wall*	1983	Funk/Soul
Fluxion	*Influx*	1999	Electronic
Fluxion	*Plain*	2001	Electronic
Fluxion	*Prospect II*	2000	Ambient
Gescom	*Self Impersonation (Pharoid Remix)*	1998	Electronic
Giazotto	*Adagio in G minor*	1958	Classical
Glen Velez	*Bodhran*	2003	Electronic
Global Communication	*8 07 (Maiden Voyage)*	1994	Ambient
Global Communication	*9 39*	1994	Ambient
Gloria Gaynor	*I Will Survive*	1978	Rock & Pop
Gosub	*Waking Up*	2005	Electronic
Grieg	*Peer Gynt: Morning Mood*		Classical
Group Home	*Beefin' For Rap*	1999	Rap & Hip-Hop
Gwen Guthrie	*Ain't Nothin' Goin' On But The Rent*	1950	Jazz
Hariprasad Chaurasia	*Alap, Jod, Jhala*	2004	World
Harold Budd	*The Gunfighter*	1986	Ambient
Harold Budd & John Foxx	*Subtext*	2003	Ambient
Harold Budd	*The Room Alight*	2000	Ambient
Harold Budd & Brian Eno	*The Pearl*	1984	Ambient
Harry Morse Project	*So Rest*	2001	Electronic
Harry Connick, Jr.	*Let's Call The Whole Thing Off*	1989	Rock & Pop
Herbie Hancock	*Vein Melter*	1973	Jazz
Herbie Hancock	*Sun Touch*	1975	Jazz
Holst	*The Planets: Mars*	1916	Classical
Hugg & Pepp	*Tratt*	2003	Electronic
Human Mesh Dance	*Wet Moon*	1994	Ambient
I-Cube	*Le Dub*	1999	Electronic
Jeff Mills	*Mosainga*	1996	Electronic

(*continued*)

Appendix. (*Continued*).

Artist	Title	Year	Genre
Jersome Sydenham & Dennis Ferrer	*Sandcastles*	2004	Electronic
Jethro Tull	*Inside*	1970	Rock & Pop
Joe Satriani	*Circles*	1987	Rock & Pop
John Beltran	*Collage Of Dreams*	1996	Electronic
John Lee Hooker	*Onions*	1970	Blues
Kandis	*Bloop*	2002	Electronic
Kansas	*Dust In The Wind*	1977	Rock & Pop
Kazi feat. Dudley Perkins & Cornbread	*U.N.I.T.Y.*	2004	Rap & Hip-Hop
Kinesthesia	*Kobal*	1998	Electronic
King Tubby	*Dub Fever*	1982	World
Klaus Schulze	*Bayreuth Return*	1975	Electronic
Kool & The Gang	*Cherish*	1979	Funk/Soul
Kool & The Gang	*Winter Sadness*	1975	Funk/Soul
Kraftwerk	*Pocket Calculator*	1981	Electronic
Ladysmith Black Mambazo	*At Golgotha*	1987	World
La Funk Mob	*Motor Bass Get Phunked Up (Remix)*	1996	Electronic
Larry Heard	*Riverside Drive*	2002	Electronic
Lee Ann Womack	*I Hope You Dance*	2000	Country
Leon Ware	*Why I Came To California*	1982	Funk/Soul
Leon Haywood	*Wanna Do Something Freaky To You*	1975	Funk/Soul
Lifehouse	*Hanging By A Moment*	2001	Rock & Pop
Linval Thompson	*Jamaican Colley Version*	1976	World
Lionel Richie & Diana Ross	*Endless Love*	1981	Rock & Pop
Lit	*My Own Worst Enemy*	1999	Rock & Pop
Lonnie Liston Smith & The Cosmic Echoes	*Quiet Dawn*	1976	Jazz
Lonnie Liston Smith	*Night Flower*	1978	Jazz
Lonnie Liston Smith & The Cosmic Echoes	*Meditations*	1976	Jazz
Los Ladrones	*Las Lucas Del Norte*	2000	Electronic
Loscil	*Ampere*	2001	Electronic
Louis Jordan	*Beans and Corn Bread*	1949	Jazz
Low Res	*Minty*	1998	Ambient
Luomo	*Talk In Danger*	2003	Electronic
Lusine	*Falling In*	2004	Electronic
Madonna	*Like A Virgin*	1984	Rock & Pop
Man With No Name	*Pipeline*	2003	Electronic
Marco Di Marco feat. Nathan Haines	*Winding Dance*	2004	Jazz
Mariah Carey	*Always Be My Baby*	1995	Rock & Pop
Mathew Jonson	*Rewire*	2003	Electronic
Mercan Dede	*Dream Of Lover*	1998	World
Michael Andrews	*Manipulated Living*	2004	Soundtrack
Microboss	*Smoking*	2005	Electronic
Mikkel Metal	*Rollin*	2005	Electronic
Moby	*Lean On Me*	1993	Ambient
Moby	*My Beautiful Sky*	1993	Ambient
Monolake	*Abundance*	1999	Electronic
Monolake	*Fragile*	2001	Electronic
Monolake	*Gecko*	1999	Electronic
Monolake	*Occam*	1997	Electronic
Monsieur Charles	*The Flute Plays On (Basil Remix)*	2002	Electronic
Monteverdi	*Se tu mi lassi, perfida*		Classical
Moxie	*Undisputable Dub*	2004	Electronic
Mozart	*Don Giovannni: Overture*		Classical
Mozart	*Eine Kleine Nachtmusik: Allegro*		Classical
Mozart	*Le Nozze di Figaro: Deh, vieni, non tardar*		Classical
Mozart	*Le Nozze di Figaro: Overture*		Classical

(*continued*)

Appendix. (*Continued*).

Artist	Title	Year	Genre
Mr. Mister	*Broken Wings*	1985	Rock & Pop
Mr. Short Khop	*Dollaz Drank Dank (Instrumental)*	2000	Rap & Hip-Hop
Mussorgsky	*Night on Bald Mountain*		Classical
Nina Simone	*Blue Prelude*	1956	Jazz
Nine Inch Nails	*A Warm Place*	1994	Rock & Pop
Norah Jones	*Dont Know Why*	2003	Rock & Pop
Numina	*Awaken Within A Deeper Realm*	2004	Ambient
Omicron	*Caterpillar To Butterfly*	1994	Electronic
Omicron	*Earth Rider*	1994	Electronic
Opus III	*Guru Mother*	1994	Electronic
Pandit Sharda Sahai	*Rupak*	1986	World
Pat Metheny	*Spring Ain't Here*	1989	Jazz
Patti LaBelle & Michael McDonald	*On My Own*	1986	Rock & Pop
Peter Gabriel	*In Doubt*	2002	Soundtrack
Peter Gabriel	*Of These, Hope (Reprise)*	2002	Soundtrack
Peter Gabriel	*In Your Eyes*	1986	Rock & Pop
Phil Collins	*Do You Remember*	1989	Rock & Pop
Pink Floyd	*Any Colour You Like*	1973	Rock & Pop
Pink Floyd	*Eclipse*	1973	Rock & Pop
Pink Floyd	*On The Run*	1973	Rock & Pop
Plaid	*Sincetta*	2001	Electronic
Pole	*Raum 1 Variation (Kit Clayton)*	1998	Electronic
Pole	*Raum 2 Variation (Burnt Friedman)*	1998	Electronic
Polygon Window	*Quino-Phec*	1993	Electronic
Porn Sword Tobbaco	*Pinkys*	2004	Ambient
Prince	*Something In The Water Does Not Compute*	1982	Rock & Pop
Prince And The Revolution	*I Would Die 4 U*	1984	Rock & Pop
Prince And The Revolution	*The Beautiful Ones*	1984	Rock & Pop
Proem	*Place Gun To Head*	2004	Electronic
Rachmaninoff	*Piano Concerto No. 3: Intermezzo*	1939	Classical
R.E.M.	*Losing My Religion*	1991	Rock & Pop
Richard Devine	*Step Focus*	2001	Electronic
Richard Groove Holmes	*Blue Moon*	1961	Jazz
Richard Groove Holmes	*Speak Low*	1968	Jazz
Rimsky-Korsakov	*Scheherazade: The Young Prince & Princess*		Classical
RJD2	*Cornbread, Eddie & Me (Instrumental)*	2006	Electronic
R. Kelly	*Fiesta*	2001	Rock & Pop
Robert Leiner	*Northern Dark*	1994	Electronic
Rodrigo	*Concierto de Aranjuez*	1939	World
Roseanne Cash	*Never Be You*	1985	Country
Rude Boys	*It's Written All Over Your Face*	2006	Rock & Pop
Sade	*Kiss Of Life*	1992	Rock & Pop
Sade	*The Sweetest Taboo*	1985	Rock & Pop
Sandy B	*Make The World Go Round*	1996	Electronic
Santana	*La Fuente Del Ritmo*	1972	Rock & Pop
Santana	*Stone Flower*	1972	Rock & Pop
Schoenberg	*Five Pieces for Orchestra, No. 1*	1909	Classical
Schoenberg	*Five Pieces for Orchestra, No. 3*	1909	Classical
Schubert	*Symphony No. 5: Allegro*		Classical
Shapeshifter	*Etheric Stardust*		Electronic
Shapeshifter	*Inside The Atom, Outside The Universe*		Electronic
Shapeshifter	*Riders Of The Dawn*		Electronic
Shapeshifter	*Sourcetone*		Electronic
Shapeshifter	*The Return*		Electronic
Shapeshifter	*Tranquil Vapor*	2005	Electronic

(*continued*)

Appendix. (*Continued*).

Artist	Title	Year	Genre
Shiv Kumar Sharma	*The Spaces Of Silence*	2006	World
Shuttle 358	*Finch*	2002	Electronic
Simon & Garfunkel	*Scarborough Fair / Canticle*	1966	Rock & Pop
Simon & Garfunkel	*The Sound Of Silence*	1965	Rock & Pop
Simple Minds	*Don't You Forget About Me*	1985	Rock & Pop
Smashing Pumpkins	*Mayonaise*	1993	Rock & Pop
Sonar Circle	*Homebound*	2002	Electronic
Sonny & Cher	*I Got You Babe*	1965	Rock & Pop
Sono	*Keep Control*	2000	Electronic
Squarepusher	*Tundra*	1996	Electronic
St. Germain	*Sure Thing*	2000	Jazz
Stan Getz & Joao Gilberto	*Desafinado (Off Key)*	1963	Jazz
Stars Of The Lid & John McCafferty	*Anchor States Part 3*	1998	Ambient
Starseed Transmission	*Alkaline Paradygm*	1994	Ambient
Starseed Transmission	*Etherea*	1994	Ambient
Stephanie Mills	*Never Knew Love Like This Before*	1980	Rock & Pop
Strauss, J.	*Pizzicato Polka*		Classical
Strauss, R.	*Salome: Dance Of The Seven Veils*	1905	Classical
Stravinsky	*The Rite Of Spring: Ritual Action*	1913	Classical
Sun Electric	*The Return Of Zindato*	1993	Electronic
Sven Van Hees	*Too Many Captain Beefhearts*	2000	Ambient
Tangerine Dream	*The Dream Is Always The Same*	1983	Electronic
The Beginning Of The End	*Come Down Baby*	1971	Funk/Soul
The Cure	*Apart*	1992	Rock & Pop
The Cure	*Fascination Street*	1989	Rock & Pop
The Cure	*Plainsong*	1989	Rock & Pop
The Doors	*Blue Sunday*	1970	Rock & Pop
The Doors	*Break On Through (To The Other Side)*	1967	Rock & Pop
The Infamous ... Bobb Deep	*Shook Ones, Pt. 2*	2005	Rap & Hip-Hop
The Stone Roses	*Fools Gold*	1989	Rock & Pop
The Beatles	*Cant Buy Me Love*	1964	Rock & Pop
The Carpenters	*Close To You*	1970	Rock & Pop
The Doors	*Take It As It Comes*	1967	Rock & Pop
The Mariachi Garage	*Rewriting The History Of Elevator Music*	2006	Jazz
The Police	*Every Breath You Take*	1983	Rock & Pop
The Righteous Brothers	*Unchained Melody*	1965	Rock & Pop
The Rotating Assembly	*Good Question*	2004	Jazz
The Undisputed Truth	*Smiling Faces Sometimes*	1971	Funk/Soul
Tim McGraw	*Just To See You Smile*	1997	Country
Tipper	*Dissolve (Out) (Phoenecia Remix)*	2000	Electronic
Toni Braxton	*You're Makin' Me High*	1996	Rock & Pop
Tony Bennett and K.D. Lang	*What A Wonderful World*	2002	Rock & Pop
Tracy Chapman	*Give Me One Reason*	1995	Rock & Pop
Triola	*Ral*	2004	Electronic
Twerk	*From Green To Brown*	2003	Electronic
U2	*4th Of July*	1984	Rock & Pop
U2	*Heartland*	1988	Rock & Pop
U2	*Love Is Blindness*	1991	Rock & Pop
Urban Soul	*What Do I Gotta Do (Eric Kupper Urban Dub)*	1997	Electronic
Vainqueur	*Solanus (Extracted)*	1996	Electronic
Vangelis	*Blade Runner Blues*	1994	Ambient
Vangelis	*Movement X (Epilogue)*	1991	Ambient
Van Halen	*Ain't Talkin' 'Bout Love*	1978	Rock & Pop
Virus	*Sun (Oakenfold/Osborne Mix)*	1995	Electronic
Vivaldi	*The Four Seasons – Spring: Allegro Pastorale*		Classical

(*continued*)

Appendix. (*Continued*).

Artist	Title	Year	Genre
Vontel	*If You Want To Be A Playa*	1998	Rap & Hip-Hop
Wagner	*Die Meistersinger von Nürnberg: Overture*		Classical
Wagner	*Lohengrin: Act 1 Prelude*		Classical
Wagner	*Die Walküre: Ride of the Valkyries*		Classical
When In Rome	*The Promise*	1988	Rock & Pop
Whitney Houston	*I Will Always Love You*	1992	Rock & Pop
Willie Hutch	*Slick*	1996	Funk/Soul
Wir	*Ticking Mouth*	1991	Electronic
www.jz-arkh.co.uk	*ddrhodes*	2001	Electronic
X-Ile	*Fantasee*	1999	Electronic
Zero 7	*I Have Seen*	2001	Funk/Soul
Zorn	*Early Shift*	2001	Electronic
Zorn	*Schatten & Staub*	2001	Electronic