AUTOMATIC MANIPULATION OF MUSIC TO EXPRESS DESIRED EMOTIONS

António Pedro Oliveira
Centre for Informatics and Systems, University of Coimbra
apsimoes@student.dei.uc.pt

Amílcar Cardoso
Centre for Informatics and Systems, University of Coimbra
amilcar@dei.uc.pt

ABSTRACT

We are developing a computational system that produces music expressing desired emotions. This paper is focused on the automatic transformation of 2 emotional dimensions of music (valence and arousal) by changing musical features: tempo, pitch register, musical scales, instruments and articulation. Transformation is supported by 2 regression models, each with weighted mappings between an emotional dimension and music features. We also present 2 algorithms used to sequence segments.

We made an experiment with 37 listeners who were asked to label online 2 emotional dimensions of 132 musical segments. Data coming from this experiment was used to test the effectiveness of the transformation algorithms and to update the weights of features of the regression models. Tempo and pitch register proved to be relevant on both valence and arousal. Musical scales and instruments were also relevant for both emotional dimensions but with a lower impact. Staccato articulation influenced only valence.

1. INTRODUCTION

The automatic production of music that expresses desired emotions is a problem with a large spectrum for improvements. The importance of developing systems with such a capability is evident to the society. Every context with a particular emotional need can use systems of this kind to accomplish its objectives. However, only recently there has been a great improvement in this area. Scientists have tried to quantify and explain how music expresses certain emotions [3, 4, 11]. Engineers have developed systems with the capability of producing music conveying specific emotions [7, 16, 17] by using the knowledge acquired by scientists.

We are developing a computational system used to produce music expressing desired emotions (section 3), grounded on research of Music Psychology and Music Computing (section 2). In this work we are focused on the transformation of music and improvement of the weights of features of the regression models used in the control of the emotional content of music; we also present sequencing algorithms (section 4). We made an experiment with 37 listeners that emotionally labeled 132 musical segments: 63 of transformed and 69 of non-transformed music (section 5). The analysis of the data obtained from this experiment and the update of the regression models are present in section 6. Section 7 makes some final remarks.

2. RELATED WORK

Our work involves research done in Music Psychology and Music Computing. The comprehension of the influence of musical features in emotional states has contributed to bridge the semantic gap between emotions and music. We are interested in the effect of structural and performance features on the experienced emotion [10]. We analyzed several works [2, 3, 4, 6, 7, 11, 15] and made a systematization of the relevant characteristics to this work that are common to four types of music: happy, sad, activating and relaxing (Figure 1).

Musical Feature		Happy music	Sad music	Activating music	Relaxing music
Instruments	timbre	piano, strings instruments, few harmonics, bright, percussion instruments	timpani, violin, woodwind instruments, few harmonics, dull, harsh	brass, low register instruments, timpani, harsh, bright, percussion instruments	woodwind instruments, few harmonics, soft
Dynamics	loudness	high	low	high	low
	articulation	staccato	legato	staccato	legato
	articulation variab.	large	small	-	-
	sound variability	low	-	-	-
Rhythm	tempo	fast	slow	fast	slow
	note density	high	low	high	low
	note duration	small	large	small	large
	tempo variability	small	-	-	-
	duration contrast	sharp	soft	-	-
Melody	pitch register	high	low	-	-
	pitch repetition	high	low	high	low
	stable/ expect notes	accented	-	-	-
	unstab/ unexp notes	accented	-	-	-
Harmony	harmony	consonant	dissonant	complex, dissonant	-
	scale	major, pentatonic	minor, diminished	-	-

Figure 1. Characteristics of happy, sad, activating and relaxing music

This systematized knowledge is used by works aiming to transform the emotional content of music. These works have developed computational systems with a knowledge-based control of structural and performance features of pre-composed musical scores [2, 7, 16, 17]. Winter [17] built a real-time application to control structural factors of a composition. Pre-composed scores were manipulated through the application of rules with control values for different features: mode, instrumentation, rhythm and harmony. REMUPP [16] is a system that allows real-time manipulation of features like tonality, mode, tempo and instrumentation. Pre-composed music is given to a music player and specific music features are used to control the sequencer (e.g., tempo); to employ filters and effects (e.g., rhythmic complexity); and to control synthesizers (e.g., instrumentation). Livingstone and Brown [7] implemented a rule-based system to affect perceived emotions by modifying the musical structure. This system is grounded on a list of performance and structural features, and their emotional effect. The KTH rule-based system for music performance [2] relates performance features to emotional expression. This system is grounded on studies of music psychology.

3. COMPUTATIONAL SYSTEM

The work presented in this paper is part of a project that intends to develop a system that produces music expressing a desired emotion. This objective is accomplished in 3 main stages: segmentation, selection and transformation; and 3 secondary stages: features extraction, sequencing and synthesis. We are using 2 auxiliary structures: a music base and a knowledge base. The music base has pre-composed MIDI music tagged with music features. The knowledge base is implemented as 2 regression models that consist of relations between each emotional dimension and music features.

Aided by Figure 2 we will describe with more detail each of these stages. Pre-composed music of the music base is input to a segmentation module that produces fragments. These fragments must as much as possible have a musical sense of its own and express a single emotion. Segmentation consists in a process of discovery of fragments. This process occurs from the beginning of the piece by looking to each note onset with the higher weights. An adaptation of LBDM [1] is used to attribute these weights according to the importance and degree of variation of five features: pitch, rhythm, silence, loudness and instrumentation). Resulting fragments are input to the module of features extraction that obtains music features used to label these fragments which are then stored in the music base.

Selection and transformation are supported by the same knowledge base. Selection module intends to obtain musical pieces with an emotional content similar to the desired emotion. These pieces are obtained from the music base, according to similarity metrics between desired emotion and music emotional content. This emotional content is calculated through a weighted sum of the music features, with the help of a vector of weights defined in the knowledge base for each emotional dimension. Selected pieces can then be transformed to come even closer to the desired emotion. Transformation is applied in 6 features (section 4). The knowledge base has weights that control the degree of transformation for each feature. Produced pieces from the transformation module are sequenced in the sequencing module. This module changes musical features with the objective of obtaining a smooth sequence of segments with similar emotional content. This sequence is given to a synthesis module, which uses information about the MIDI instruments and timbral features to guide the selection of sounds from a library of sounds.

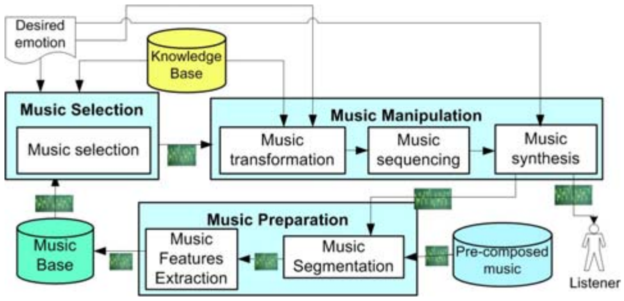

Figure 2. Diagram of our computational system

4. METHODS

This section presents the methods being used to transform music, sequence music and improve regression models.

4.1. Transformation of musical segments

Music transformation algorithms have the objective to approximate the emotional content of selected music to the desired emotion. By knowing the characteristics common to different types of music (section 2) we developed six algorithms that transform different features: tempo, pitch register, musical scale, instruments, articulation and contrast of the duration of notes. These algorithms start by calculating the emotional distance between the emotional content of the selected music and the desired emotion. The value of this distance is divided by the value of the weight of the feature being transformed. The value that results from this division corresponds to the amount of increase/decrease we need to make on the feature to approximate music emotional content to the desired emotion. Next paragraphs explain how this increase/decrease is made by each algorithm on the MIDI file.

The algorithm used to transform tempo obtains the original tempo of the music (in beats per minute) and then increases/decreases the note onsets and duration of notes.

The algorithm that transforms pitch register transposes up/down music by a specific number of octaves to increase/decrease valence/arousal. We choose octaves, because they are the intervallic transformation more consonant [14] with audible repercussion in the frequency spectrum. This is done by adding positive/negative multiples of 12 to the pitch of all the notes.

The algorithm that transforms musical scales finds the original scale of the MIDI file and selects a target scale according to emotional tags to be defined for each scale. Then, it finds the pitch distance relative to the tonic for each note in the original scale. If this distance is not found in the target scale, it finds the closer pitch distance to this distance that is present in the target scale and changes the pitch of the note, according to the new distance. For instance, suppose we want to transform from a ragha madhuri (pitch distances of 4, 5, 7, 9, 10 and 11 semitones to the tonic) to a minor gipsy scale (pitch distances of 2, 3, 6, 7, 8 and 11 semitones to the tonic). A note distant 4 semitones from the tonic in the ragha madhuri scale would have its pitch decreased by one semitone to be distant 3 semitones from the tonic in the gipsy scale. This happens because the interval of 4 semitones is not present in the minor gipsy scale. We used a group of 27 twelve-tone scales[1]. We chose this group and not others because it has a higher variety of number of notes and intervals: scales have between 2 and 7 notes and intervals vary from 1 to 7 semitones.

The algorithm used to transform instruments obtains original MIDI instruments and selects new instruments according to the emotional tags of each timbre. These tags are calculated through a weighted sum of audio features (e.g., spectral dissonance and spectral sharpness), with the help of a vector of weights defined in the knowledge base for each emotional dimension.

The algorithm that transforms normal to staccato articulation decreases the duration of all notes by a specific percentage. If we consider 75%, notes with a duration of X would have a new duration of X-X*0.75.

We also have an algorithm that increases the contrast between the duration of notes. It increases/decreases the duration of longer/shorter notes according to a degree of transformation (k) and the duration of notes expressed in beats:

```
if (beat < 1 && beat > 0)
   adjustment = -(11/170) * (170 - beat*170)
else if (beat < 2 && beat > 1)
   adjustment = (4/200) * (beat*200)
else if (beat > 2)
   adjustment = (2/200) * ((beat-1)*200) +  6
newDuration = oldDuration + k*adjustment /400
```

This algorithm is based on an equivalent algorithm of KTH [2]. It was not yet tested.

4.2. Sequencing of musical segments

Music sequencing algorithms have the objective to obtain a smooth sequence of segments with similar emotional content. To date we only have algorithms for rhythmic matching and to do fade in and fade out of volume.

The algorithm of rhythmic matching intends to match the rhythm of previous segment(s) to the rhythm of next segment. This objective is accomplished by matching the mean of the interonset intervals (IOI) of the notes of the N^{th} segment (IOI_N) with the mean of the IOI of the notes of the N-1 (IOI_N-1) segments according to the pseudo-code:

```
adjustment = IOI_N-1 / IOI_N
if (IOI_N-1 > IOI_N)
adjustment = -IOI_N / IOI_N-1
change = 0
for firstNote to pnultimateNote
  change = change + (onsetNextNote – onsetThisNote) –
  (onsetNextNote – onsetThisNote)*adjustment
  onsetNextNote = onsetNextNote – change
end for
```

The algorithms of fade in and fade out are used to smooth the transitions between segments, respectively, by gradually increasing the volume of the starting segment and decreasing the volume of the finishing segment.

4.3. Improvement of the regression models

We intend to improve the regression models in order to control transformation algorithms, as well as to allow a better control of selection algorithms. Updated weights take into account new weights obtained in this experiment and weights obtained in previous experiments [8, 9] and are calculated according to the following formula: 0.1*Weights [8] + 0.6*Weights [9] + 0.3*NewWeights. The values of 0.1, 0.6 and 0.3 for each experiment were obtained according to the number of music files and listeners used in each of the experiments.

5. EXPERIMENT

In our experiment we started by using our system to randomly select 14 MIDI files of film and pop/rock music expressing different emotions. These MIDI files were subject to a segmentation process that produced a set of 746 segments, from which a set of features were automatically extracted. From the set of segments and with the help of the set of features, our system selected 132 segments that were expected to express different emotions and last between 10 and 15 seconds. We used small segments to try to have only one emotion expressed in each segment and to allow a fine-grained correlation between musical features and emotions. 63 of the 132 segments were changed in only one of the following features: 12 in tempo, 10 in pitch register, 27 in musical scale and 14 in articulation. The other 69 segments were not subject to transformations.

We made an online questionnaire[2] to allow anonymous people to emotionally classify the 132 segments. 37 different listeners labeled 2 emotional dimensions for each

[1] http://papersao.googlepages.com/musicalscales

[2] http://student.dei.uc.pt/~apsimoes/PhD/Music/smc09/

segment with values selected from the integer interval between 0 and 10. We obtained standard deviations of 1.76 and 1.85, respectively, for the answers for valence and arousal. These deviations were computed first between listeners, and then averaged over segments. Obtained labels were related with the extracted features for both transformed and non-transformed music. Feature selection algorithms were used (best-first and genetic search [18]) over the 476 extracted features from non-transformed music to select the features emotionally more relevant.

6. RESULTS

This section presents the results of the emotional impact of transformed (tempo, pitch register, musical scales, instruments and articulation) and non-transformed features in order to update the relations and their weights for each feature of the regression models

6.1. Tempo

We transformed 6 segments by accelerating in 50% and slowing down in 30% their tempo, obtaining 3 versions for each one: fast, normal and slow tempo. For each of the resulting 6 groups of 3 segments, we correlated the tempo of each version with the emotional data obtained in the experiment. Table 1 presents the correlation coefficients.

Group	1	2	3	4	5	6	Mean
Valence(%)	94	96	91	-7	62	100	75
Arousal(%)	100	98	98	-16	97	-36	74

Table 1. Correlation coefficients between tempo and valence and arousal for the 6 groups of segments

The expected high positive coefficients were confirmed by most of the results. However, the fourth group of segments obtained small negative coefficients for both valence and arousal, and the sixth group for arousal. This may be explained by the presence of an imperceptible transformation, because of the presence of very long notes (> 4seconds) on the original segment. A higher percentage of acceleration and slowing down of the original segment would be needed. The result of 100% for valence in the sixth group is not very reliable because the answers were very close: 3.5, 3.3 and 3.2. Emotional transformations contributed to an increase of 0.4/0.2 in valence/arousal with changes from low to normal tempo, and an increase of 1/0.8 in valence/arousal with changes from normal to high tempo.

6.2. Pitch register

We transformed 5 segments by transposing them up and down two octaves, obtaining 3 versions for each one: high, normal and low register. For each of the resulting 5 groups of 3 segments, we correlated the register of each version with the emotional data obtained in the experiment. Table 2 presents the correlation coefficients.

Group	1	2	3	4	5	Mean
Valence(%)	79	100	15	60	33	57
Arousal(%)	-39	-63	-98	-92	-62	71

Table 2. Correlation coefficients between pitch register and valence and arousal for the 5 groups of segments

Generally speaking, the increase of register correlates positively with valence and negatively with arousal. A more detailed analysis of the results in groups 3 and 5 showed lower correlation for valence, which revealed that the change from normal to high register contributes to a decrease in valence. From the analysis of the mean pitch of the segments, we can observe that the increase in register affects valence positively only till we have mean values of MIDI pitch around 80, whilst higher values contribute to a decrease in valence. We assisted to a similar situation in this first group for arousal: values of MIDI pitch higher than 80 seem to do not affect the arousal of music. Emotional transformations contributed to an increase of 2/-0.6 in valence/arousal with changes from low to normal register, and an increase of 0.7/-0.4 in valence/arousal with changes from normal to high register.

6.3. Musical scales

We transformed 1 segment by changing the original major scale to other 27 musical scales (subsection 4.2). We used feature selection algorithms in the process of finding the features that best characterize the emotional variation when changing the scale. The number of semitones in scale, the difference between successive intervals of the scale, the spectral dissonance and spectral sharpness with, respectively, the weights: -0.17,-0.15, 0.18 and -0.14 were important for valence. The number of semitones in scale, the difference between successive intervals of the scale, the spectral dissonance and stepwise motion with, respectively, the weights:-0.19, -0.07, 0.14 and 0.24 were important for arousal. Table 3 presents the correlation coefficients between the most discriminant features and emotional dimensions for the considered 27 versions of the segment.

Musical feature	**Valence(%)**	**Arousal(%)**
Spectral dissonance	46	31
Tonal dissonance	28	-
Timbral width	-32	-
Sharpness	-34	-20
Stepwise motion	24	33
Melodic thirds	-34	-18
Number of semitones in scale	-40	-23
Difference of successive intervals in scale	-28	-16
Correlation coefficient	**61**	**45**

Table 3. Correlation coefficients between musical features and valence and arousal for the 27 versions of the segment

6.4. Instruments

We changed the instruments of the 69 segments not subject to other types of transformations: tempo, register, scales and articulation. We tried to have each of the General Midi 1 (GM1) instruments [12] present in, at least, 1 of these 69 segments, in order to analyze the emotional impact of every GM1 instrument.

Table 4 presents the correlation coefficients between audio features and the valence and arousal for the 69 segments. From Table 4, we can infer that instruments are essentially relevant to the arousal, being spectral dissonance, timbral width and spectral sharpness relevant features in the emotional analysis of the sound/timbre of instruments. We found that violin, string ensembles, choirs and piccolo contribute to low valence; and percussion instruments contribute to high valence/arousal.

Musical feature	Valence(%)	Arousal(%)
Spectral dissonance	28	72
Timbral width	-	54
Tonal dissonance	19	27
Sharpness	-	44

Table 4. Correlation coefficients between musical features and valence and arousal for the 69 segments

6.5. Articulation

We transformed 14 segments by changing their articulation to staccato, obtaining 2 versions for each one: normal and staccato. We correlated the articulation of the 28 versions with the emotional data obtained in our experiment and found that the change from normal to staccato articulation is 40% correlated with the increase of valence and has no impact in arousal.

6.6. Emotional impact of non-transformed features

We used experimental data from the 69 segments not subject to transformations to analyze the emotional impact of non-transformed features. Tables 5 and 6 present the features emotionally more discriminant for each emotional dimension. We obtained correlation coefficients of 79% and 85% (tables 7 and 8), respectively, for valence and arousal, using these features.

6.7. Update of the regression models

After the analysis of the emotional effect of tempo, pitch register, musical scales, instruments and articulation on the transformed music and of the emotional effect of the more important features on the non-transformed music we updated the weights of features of the regression models according to the formula present in subsection 4.3.

In tables 7 and 8, we compare the weights as well as the correlation coefficients of previous experiments [8, 9] with the weights and correlation coefficient obtained in this experiment. The fifth features emotionally more discriminant are present with a bold font with higher size.

7. CONCLUSION

We successfully tested the effectiveness of algorithms of music transformation. Change of tempo was positively related to both valence and arousal. Change of pitch register was positively related to valence and negatively related to arousal. The presence of semitones in musical scales was found to be an important feature negatively related to valence. Spectral dissonance, timbral width and spectral sharpness were found to be important features for instruments and are positively related to arousal. Staccato articulation was found to be positively related to valence.

These results and correlation coefficients of features emotionally more relevant served to update the weights of features of the regression models that have been used to control the emotional changes made by the transformation algorithms. This experiment was grounded on previous experiments [8, 9].

Musical feature	Corr. Coeff.(%)
Staccato incidence	57
Number of unpitched instruments	53
Note density	52
Average note duration	-50
Average time between attacks	-50
Overall dynamic range	48
Variability of note duration	-46
Melodic fifths	45
Pitch variety	43
Note prevalence of closed hi-hat	42
Rhythmic looseness	41
Percussion prevalence	40

Table 5. Features emotionally more discriminant for valence

Musical feature	Corr. Coeff.(%)
Variability of note prevalence of unpitched instruments	70
Percussion prevalence	69
Note density	66
Number of unpitched instruments	58
Staccato incidence	56
Importance of loudest voice	55
Variation of dynamics	48
Note prevalence of snare drum	47
Overall dynamic range	46
Variability of note prevalence of pitched instruments	45
Note prevalence of bass drum	45
Note prevalence of closed hi-hat	43

Table 6. Features emotionally more discriminant for arousal

Musical feature	*W* [7] 16files 53list.	*W* [8] 96files 80list.	*NewW* 69files 37list.	*Update W*
Average note duration	**0**	**-0.30**	**0**	**-0.18**
Chromatic motion	0	-0.16	-0.12	-0.13
Importance bass register	**-0.35**	**-0.17**	**-0.24**	**-0.20**
Initial tempo	**0**	**0.60**	**0.4**	**0.48**
Muted guitar fraction	**0**	**0.27**	**0**	**0.16**
Note density	0	0.09	0	0.05
Note prevalence of marimba	0	0.17	0	0.10
Num. unpitched instruments	0.20	0.11	0	0.08
Orchestral strings fraction	**0**	**-0.43**	**-0.14**	**-0.30**
Polyrhythms	-0.26	0	-0.16	-0.08
Saxophone fraction	0.26	0.16	0	0.13
Staccato incidence	0	0	0.17	0.05
String ensemble fraction	-0.46	0	-0.19	-0.10
Variability of note duration	-0.26	0	-0.37	-0.14
Key mode	0	-0.13	0	-0.08
Spectral sharpness	0	0	0.24	0.07
Spectral volume	0	0	0.28	0.08
Correlation coefficient	**97%**	**81%**	**79%**	**-**

Table 7. Weights and correlation coefficients for features emotionally more discriminant for valence

Musical feature	*W* [7] 16files 53list.	*W* [8] 96files 80list.	*NewW* 69files 37list.	*Update W*
Average note duration	**-0.57**	**-0.44**	**0**	**-0.32**
Avg. time between attacks	0	-0.12	0	-0.07
Importance bass register	-0.22	0	-0.21	-0.08
Importance of high register	-0.57	0	0	-0.06
Initial tempo	**0**	**0.31**	**0.4**	**0.31**
Note density	**0.21**	**0.47**	**0.22**	**0.36**
Number of common pitches	0	0	0.19	0.06
Percussion prevalence	0.27	-0.26	0	-0.12
Primary register	0	-0.12	-0.25	-0.16
Repeated notes	0.16	0.11	0	0.08
Strength strong. rhythm. pulse	0	0	0.22	0.07
Variability of note duration	-0.26	-0.17	0	-0.13
Variability note prevalence of unpitched instruments	0	0.20	0.12	0.15
Spectral dissonance	0	0	0.27	0.08
Spectral sharpness	**0**	**0.30**	**0.22**	**0.25**
Spectral similarity	**0**	**-0.35**	**-0.21**	**-0.27**
Average dynamics	0	0.08	0	0.05
Correlation coefficient	**99%**	**88%**	**85%**	**-**

Table 8. Weights and correlation coefficients for features emotionally more discriminant for arousal

8. REFERENCES

[1] Cambouropoulos, E.. "The local boundary detection model (LBDM) and its application in the study of expressive timing", *International Computer Music Conference*, 17-22, 2001.

[2] Friberg, A., Bresin, R. and Sundberg, J., "Overview Of The KTH Rule System For Musical Performance", *Advances in Cognitive Psychology*, 2:145-161, 2006.

[3] Gabrielsson, A. and Lindstrom, E., "The Influence Of Musical Structure On Emotional Expression", *Music and emotion: Theory and research*. Oxford University Press, 223–248, 2001.

[4] Juslin, P., "Communicating emotion in music performance: A review and a theoretical framework" *Music and Emotion: Theory and Research,* 309-337, 2001.

[5] Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D. and Lesaffre, M., "Prediction of musical affect using a combination of acoustic structural cues", *Journal of New Music Research*, 34(1):39-67, 2005.

[6] Lindstrom, E., *A Dynamic View of Melodic Organization and Performance*, PhD thesis, Acta Universitatis Upsaliensis Uppsala, 2004.

[7] Livingstone, S. and Brown, A., "Dynamic response: real-time adaptation for music emotion." *Australasian Conf. On Interactive Entertainment*, 105–111, 2005.

[8] Oliveira, A. and Cardoso, A., "Towards bi-dimensional classification of symbolic music by affective content", *Int. Computer Music Conf.*, 2008.

[9] Oliveira, A. and Cardoso, A., "Modeling Affective Content of Music: A Knowledge Base Approach", *Sound and Music Computing Conference*, 2008.

[10] Scherer, K. and Zentner, M., "Emotional effects of music: Production rules", *Music and emotion: Theory and research*, 361–392, 2001.

[11] Schubert, E., *Measurement and Time Series Analysis of Emotion in Music*, PhD thesis, University of New South Wales, 1999.

[12] Selfridge-Field, E., *Beyond MIDI: the handbook of musical codes*, MIT Press, 1997.

[13] Sethares, W., *Tuning, timbre, spectrum, scale*, Springer, 2005.

[14] Vassilakis, P., Auditory Roughness as a Means of Musical Expression, *Perspectives in Systematic Musicology*, 12:119-144, 2005.

[15] Wassermann, K., Eng, K., Verschure, P and Manzolli, J., "Live soundscape composition based on synthetic emotions", *IEEE Multimedia,* 10:82-90, 2003.

[16] Wingstedt, J., Liljedahl, M., Lindberg, S. and Berg, J. "Remupp: An interactive tool for investigating musical properties and relations", *New Interfaces For Musical Expression*, University of British Columbia, 232–235, 2005.

[17] Winter, R., *Interactive music: Compositional techniques for communicating different emotional qualities*, Master's thesis, University of York, 2005.

[18] Witten, I., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S. "Weka: Practical machine learning tools and techniques with java implementations.", *Int. Conf. on Neural Info. Processing*, 192-196, 1999.