Emotion detection in music, a survey

Bram van de Laar

b.l.a.vandelaar@student.utwente.nl

ABSTRACT

Mood and emotion detection can play an important role in (music) entertainment applications. There are a lot of methods for mood and emotion detection in music through acoustical feature analysis. Some perform better depending on the situation. Differences in either music or emotions or precision of a method play a role in the application of a method.

This paper presents a comparison of different methods of emotion detection. This comparison will make a distinction between the precision of a certain method, the granularity of emotions and the various possible applications with all their requirements.

KEYWORDS

Music, Emotion detection, Mood detection, Feature analysis

1. INTRODUCTION

Music is getting more and more into our everyday lives with the explosive growth of digital audio and media players. With the enormous space on such devices a selection to listen to has to be made. A way to do this is selecting the music on emotion or mood. Users often want to listen to music that is in a certain category of emotions or they want to listen to music that brings them in a certain mood. With the use of mood and emotion detection algorithms such a system can be devised. Such a system should crawl through all of the users music files and make a selection based on the users input (an emotion or category of emotions).

This paper presents a comparison between different methods of emotion detection in music. There are various aspects to a emotion detection method which can make it perform good or bad. In general methods perform better when fewer emotion categories are used, although a greater number of categories is often desirable. With lower granularity levels emotions like frustration and excitement could fade into each other. [LO2003]

To make a good comparison it is important to know how to distinguish between different acoustical features in music and which features are important for emotion. Secondly, knowing what emotions are relevant for detection and how they should be mapped to acoustic features is of great importance. The application in which an emotion detection system will be used also adds certain requirements or loosens some.

Almost all emotion detection methods make use of some sort of a learning algorithm. Often this is essential to getting good boundaries for the value of an acoustical feature that leads to a certain emotion. Some algorithms use a single label as a classification, so a musical piece represents happiness or sadness. Others use a multi-label classification. The result of this is that a musical piece can be, for example, happy and fanciful at the same time.

In most methods, the musical pieces are converted to a standard format (22050 Hz sampling frequency, 16 bit precision, mono and trimmed to 20 or 30 seconds) and then a great number of these clips are fed to the algorithm. These clips are annotated with a certain emotion and the result of the algorithm will be compared with the annotated value. Then the learning machine will do what it is supposed to do and will probably learn from its mistakes by trying different emotions for the values of acoustical features the algorithm found and ultimately only uses those with the best result. In this way the best values can be found and the algorithm gets a much higher precision.

The first thing discussed are the various applications of emotion detection in music. Then various features that are used for detecting emotions are explained. Different emotion models are reviewed and a comparison between six methods on different issues will be made.

2. APPLICATIONS OF EMOTION DETECTION IN MUSIC

Huron [Hur2000] gives some examples for different applications of emotion detection methods in music. The following list is based upon the Hurons list, but extra possible applications are and more explanation are given. To give an idea of how broad the applications could be:

- The owner of a trendy shop who wants to seek music that attracts s a certain clientele.
- An aerobics instructor who seeks a certain tempo for his or her workout.
- A film director who seeks music evoking a certain mood which matches the images on screen. In this way the person watching the movie will be totally grasped by the scene.
- An advertiser seeking a tune that is highly memorable or that evokes a positive emotion towards a certain product. Presumably happy and positively charged music.
- A physiotherapist who seeks music that will motivate a patient while doing recovery exercises. (e.g. Survivor - Eye of the Tiger)
- A call center that receives inbound calls that has to put callers on hold will want to give their users happy music. Often very old and typical music is used, this can be improved by an application that searches for happy music in recent music.
- Working personnel who seek music that will keep them alert. This will be mostly cheerful or arousing music.
- A DJ who seeks music that will has the same key as the current song or approximately the same beat so that the people on the dance floor notice as little as possible from mixing two songs.

4th Twente Student Conference on IT , Enschede 30 January, 2006

- People at home or listeners in general who feel a little down and want to listen to a bluesy song or lonely music. Also a very common phenomenon is that listeners don't want to listen to music with a certain mood but that they just want to listen to something new.
- Emotion detection possibly can be reversely used for music generation in games. Also game directors that want to use existing music can find the music they want for their game with the use of emotion detection. Emotions such as fear (danger), anger, victory and happiness are often used in games.

3. FEATURES IN MUSIC

What is music? Music is emotion. Music consists of notes, tones, rhythms, instruments, lyrics, vocals, timbre et cetera. But how can a computer make some sense out of all those high-level musical elements?

There are a lot of acoustical features that can be calculated from the raw audio waveform of a musical piece. In this way, an algorithm extracts numbers and figures out of music. There are different categories of these features differing from low to high level and in according complexity.

3.1 Musical surface

Firstly there are the musical surface or timbral texture features. Most of these features are based on a Short Time Fourier Transformation. (STFT)

- Centroid
  This is the mean of the short time Fourier amplitude spectrum. It gives an indication of how "bright" a musical piece is.
- Roll off
  This is the point where frequencies are getting smaller in amplitude and gives the shape of the spectrum. 95% of the total spectrum is within this range.
- Spectral Flux
  This indicates how much the spectral shape changes from frame to frame.
- ZeroCrossings
  This feature gives the number of times the signal crosses the zero line (thus when the signal changes its sign). It is a good indicator of the amount of noise.
- LowEnergy or Average Silence Ratio
  This represents the percentage of frames with a less than average energy.

3.2 Spectral Flatness Measure

This acoustical feature (also called tonality coefficient) quantifies how much tone-like a sound is. It is based on the resonant structure and the spiky nature of a tone which is quite different compared to the flat spectrum of a noise-like sound.

3.3 Spectral Crest Factor

The Spectral Crest Factor is the ratio between the highest peaks and the mean RMS value of the signal. This feature can be used in different frequency bands and quantifies how "spiky" the signal is.

3.4 Mel Frequency Cepstral Coefficients.

This group of features is based on perceptually based frequency bands. The conversion of Hertz's into the Mel scale is

$$MEL = c \cdot \log(\frac{f}{700} + 1)$$

With 'c' a constant equal to 1127.01048 and *f* for frequency. This is done for the whole amplitude spectrum based on a STFT. Then various frequency bands are defined according to the Mel scale. Means and standard deviations of these sub bands can be used to interpret. These features are also often used for speech recognition. [Wik2005-1]

3.5 DWCH

DWCH stands for Daubechies Wavelet Coefficient Histogram. This is a higher level and more complex feature than the above.

"In formal terms, this representation is a wavelet series, which is the coordinate representation of a square integrable function with respect to a complete, orthonormal set of basis functions for the Hilbert space of square integrable functions." [Wik2005-2]

Wavelets are a sort of histograms for sound. They divide up data, functions, or operators into different frequency components. There are various advantages to wavelets. Good time resolution at high frequencies and good frequency resolution at low frequencies. (Every octave lower down the scale normally has a lower resolution in terms of Hertz's.) Noise has less influence on the feature and there is less correlation between features so that a certain feature is less ambiguous to evoking an emotion thus improving accuracy.

3.6 Beat and tempo detection

Beat detection can give a lot of information on music. Fast music tends to be more happy than sad, but at the same time fast music is often more related to anger than fear. [FZP2003-2] Figure 1 shows an image of the distribution of tempi in a set of test songs. As can be seen most musical pieces are rated between 45 and 150 bpm.

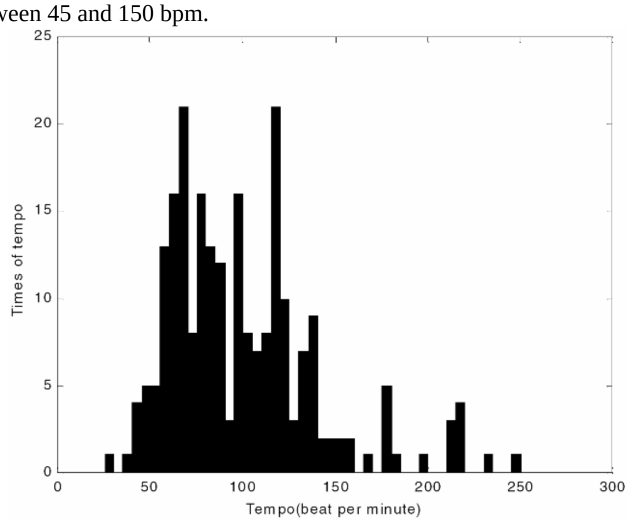

Figure 1 Distribution of tempi from [FZP2003-2]

There are a lot of beat detection mechanisms available for all sorts of applications. [Ove2006] gives a nice overview of which beat detection algorithm can be used in what application. The used beat detection algorithm for emotion detection doesn't have to be real-time but preferably it has to detect changes in the tempo so that this information can be used.

3.7 Genre information

There are two ways to get genre information to use in an emotion detection algorithm. Firstly there is the possibility to a form of annotation by hand. Artist and genre information is stored in a database and can be looked up. Secondly, there has been some research on how to automate genre recognition in music. Some methods work analogous to emotion detection methods. Others (e.g. [KPW2004]) use the Internet as a pool of information on artists to categorise them in a certain genre. [LO2003] shows that genre information can improve precision on emotion detection. A disadvantage of using genre information (or artist information) is that when an error is made in the genre detection algorithm it has a negative effect on the emotion detection algorithm as well.

3.8 Lyrics

Lyrics in music are of great importance in music and can evoke an emotion just by "saying things". This feature is not used in the reviewed articles (except [YL2004]) because it is fairly difficult to analyze the vocals in a musical piece. The voice (or even multiple voices) has to be separated from the music, a speech recognition algorithm has to convert the speech to text and finally an algorithm that interprets the words and maps the results to a certain emotion have to be implemented. This is a very cumbersome job with a lot of space for possible errors. In combination with [Dow2003] who says that the textual facet in music is more independent from melody than is generally believed, it isn't awkward that this feature isn't used frequently in emotion detection.

As said [YL2004] uses the lyrics for detecting the emotion but this is based on annotated lyrics. (The whole voice detecting and voice-to-text part is not needed). [YL2004] gives a decent overview of (a limited number of) emotions and their related words.

4. EMOTIONS IN MUSIC

An important question is: which emotions are relevant for detection in music? There has been a lot of research on emotions in psychology such as [Hev1936], [Laz1991], [RS1992] and [TWC1999]. When using an emotion model for the use of automated emotion detection in music one has to pay attention to three aspects:

1. The model has to be a good representation of reality (trivial)
2. The model should contain emotions which are often evoked by music
3. The model should contain one or more dimensions on which the emotions are measured.

Hevner [Hev1936] was the first one to do scientific research on the topic of emotional expression in music. The main conclusion was that music inherently carries an emotional meaning. Different recent papers on emotion detection mention this article and sometimes use the list of emotions Hevner proposed.

Thayers model of mood (figure 2) offers a simple but quite effective model for moods. Along the horizontal axis the amount of stress is measured and along the vertical axis the amount of energy. In music one can think of energy as the volume or the intensity of sound. Stress can be translated to "having to do too many things" so the difference in tonality and tempo would be a good mapping.

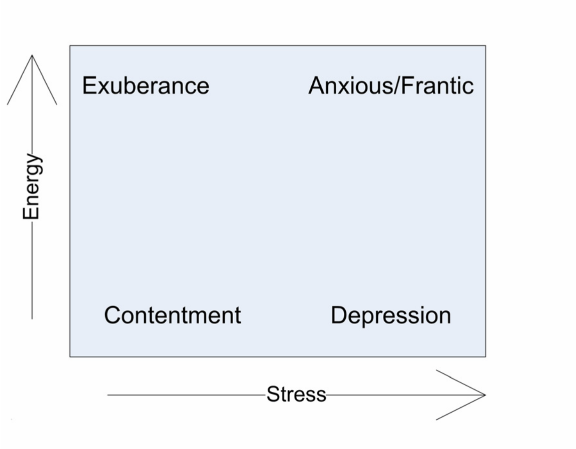

Figure 2 Thayers model of mood [Tha1989]

Another mood model is the Tellegen-Watson Clark model of mood [TWC1999]. This model contains a lot more emotions or moods and use the positive/negative affect as one dimension and the pleasantness/unpleasantness versus engagement/disengagement (45 degrees rotated) as the other. (Figure 3)

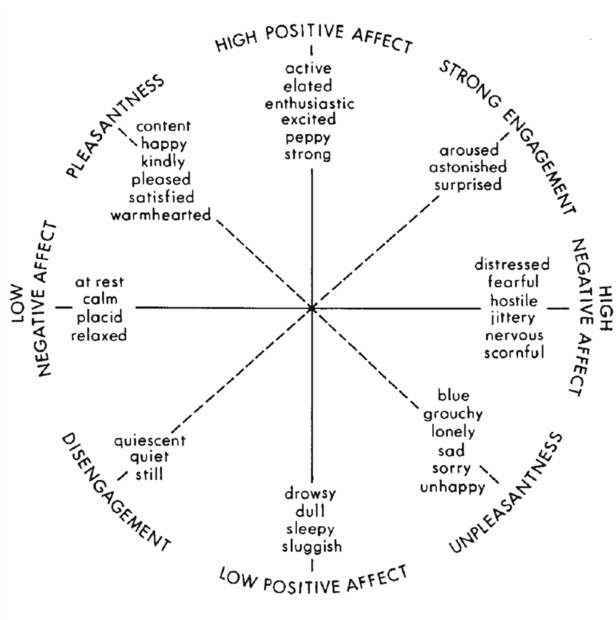

Figure 3 The Tellegen-Watson-Clark model of mood [TWC1999]

What the best emotion model is depends on the application it is placed in. The number of different emotions and their correlation has its impact on the precision of the method. For some applications (such as the physiotherapist) aren't a lot of emotions necessary as long as the emotion(s) that needs to be evoked is there and the algorithm has a very high recall percentage on that emotion. On the other hand a listener at home that wants to listen to warmhearted music but uses a program with just four emotions will get all happy music. In this case a higher granularity is needed.

5. ISSUES ON EMOTION DETECTION METHODS

There are a number of issues emotion detection methods. There always will be a certain compromise between various issues. These issues differentiate the various emotion detection methods.

5.1 Precision

One of the most obvious criteria for a good emotion or mood detection algorithm is the precision or accuracy that is achieved. An algorithm gives some sort of output. This output depends on the algorithm used. Some give just one emotion others a multi-label output with e.g. 70% emotion1 and 30% emotion2. This output will be compared with the annotated emotions evoked by a test subject. The percentage of "right answers" determines the accuracy.

5.2 Granularity

In a strong relationship with the above, granularity has a very big influence on the precision achieved. This is a quite logical phenomenon because when one has to choose between more options, there is a larger chance that a fault one in chosen. Low granularity (e.g. 4 emotions) can be useful but that will depend on the application.

5.3 Diversity

Some papers only use a limited number of songs or just one or two genres of music. This will have its influence on how much the algorithm can be optimized for that particular genre. However, this gives an unfair advantage when comparing the accuracy with other methods whose algorithm will be usable for more or even all sorts of music. These optimizations can be useful in certain applications.

5.4 Mobile use

Implementation of emotion detection on mobile devices gives a couple of restrictions to the methods. On a mobile device the resources are typically more limited than on a PC. Hard disk or flash space, computing power and total amount of RAM memory are limited because of cost and size.

None of the reviewed methods mentions how much computing power for instance is needed to work trough their test corpus. Adversely mobile devices are getting more advanced every few months. So if the resources on mobile devices aren't sufficient today, they probably will be within a few years from now.

Another facet to this issue is the nature of use of a mobile device. People will want to use the device and its "emotion drive selection search" right away and probably do not want train it for multiple hours on the music they just put on it.

5.5 Learning

The learning implemented in the algorithms is essential for getting good results. The algorithm has to learn what numbers of a feature links to a certain emotion. These numbers can be different for persons from different cultural backgrounds. Also other genres of music than tested can be used by a user. When a method is implemented into software that can be used on a PC or a mobile device one has to think at the trade off between giving the user software with a standard database and letting the user train the algorithm so that its accuracy will improve. Not all users will have the patience to train a program with hours of music. On the other hand, software that is too inaccurate will not be used.

5.6 Side effects of conversion and selection

In almost all review articles only a segment (often 20 or 30 seconds) of the total music is used. 30 seconds will be far too little to use especially for classical music. (The articulation and tempo can differ greatly in classical music). [LLZ2003] however proposes a mood tracking algorithm which tracks the mood trough the musical piece and gives an emotion for every 20 seconds of music.

All methods reviewed convert the stereo music to a mono signal with a sampling frequency of 22050 Hz with 16 bits precision. The extra channel in stereo encoded music however is not redundant. Very often the effect of (phase) differences between the two channels is used to generate a certain effect such as pinpointing instruments in an orchestra on a certain location in the stereo image or to set an ambient sound at the side of the stereo image. Removal of such effects has an impact on the emotion evoked and on the intensity of that emotion [Väs2003].

5.7 Cultural background

According to [EA2002] emotional expression is best recognized between members from the same ethnical group and that expression can loose their meaning when crossing cultural borders. In the next section [LO2004] uses two test subjects with different cultural backgrounds. Between these two there is some difference in accuracy.

6. EMOTION DETECTION METHODS

6.1 Carvalho and Chao [CC2005]

Sentiment Retrieval in Popular Music based on Sequential Learning

This article proposes a new taxonomy of sentiment classification based on how much "happiness" is present in a song. They rate a song on a 5 points based scale. Carvalho and Chao think that in this way the labels will me more appealing to users when used in a real-life application. On the other hand this will be a much easier job for their algorithm than when more categories (e.g. spooky, fear, passionate) would have been used.

Two hundred songs are used for testing. In advance, the songs are classified by two persons. The agreement between the two test subjects is 0,643 on a scale of -1 to +1. This is an agreement of ~82%, all in all not that high. The test songs are converted to a 22050Hz sampling frequency, 16bit mono. Four classes of features were used: Musical surface, Spectral Flatness Measure, Spectral Crest Factor and Mel Frequency Cepstral Coefficients, each with their own features. Furthermore the algorithm uses two different types of learning algorithms each with an adapted classifier. Results show that the difference between a binary problem and a more “fine-grained” problem of 5 labels has much more impact on precision (13.5% error rate vs. 63.45%) than the applied learning algorithm or classifier (all within 63.45% to 67%).

6.2 Li and Ogihara [LO2003]

Detecting emotion in music

The problem of emotion detection in music in this paper is presented as a multi-label classification problem. Musical pieces can belong to more than one emotion. The emotions used are the ten from Farnsworth and three added emotions. These three extra emotions were added according to a test subject who indexed the test songs. The test subject was also asked to group the emotions into groups. The classifiers used in this research are based on Support Vector Machines. The acoustic features that were used are timbral texture features, rhythmic content features (beat and tempo detection) and pitch content features. The test was performed with a corpus with a total of 499 musical pieces. 50% was used for training and the other 50% for testing. The accuracy is not so high (around 50%) but quite good compared to [CC2005] whilst this method has a higher granularity.

6.3 Li and Ogihara [LO2004]

Content-based music similarity search and emotion detection.

In this paper the researchers describe a system which actually does two things based on acoustic features in music. A similarity search which gives music which is (more or less) similar to a given piece and on the other hand emotion detection in music. The extracted music features used are the Mel-Frequency Cepstral Coefficients, the Musical surface features as mentioned in chapter 2 and the less frequently used (and more high tech) Daubechies Wavelet Filters (DWCH). The number of features DWCH gives is about 35 in different frequency bands but not all frequency bands are relevant according to the writers of this article. This method manages to yield a fairly high degree of precision over a test corpus of 235 Jazz musical pieces, at least 70% to a maximum of 83%. Then again, there are not so many categories to choose from ((Cheerful, Depressing), (Relaxing, Exciting), and (Comforting, Disturbing)), so the granularity is somewhat low. Interesting would be to see if this method would perform as good on other genres of music than Jazz. Pop, Classical and Rock music tend to be a whole lot different than Jazz. Between the two subjects used for comparison there is some difference in accuracy. This could be because of cultural differences. A suggestion of the researchers is to let the algorithm train for each subject.

6.4 Feng, Zhuang and Pan [FPZ2003-2]

Music Information Retrieval by Detecting Mood via Computational Media Aesthetics

The researchers in this paper present a system for mood detection in music. They state a simple table with tempo and articulation in music related to the emotions Happiness, Sadness, Fear and Anger. According to the article different tempo detection algorithms can be used. To determine the type of articulation (staccato or legato) in music the Average Silence Ratio is used. This feature represents what percentage of the sound in a given period of time (one frame) is below the average level in one second. A neural network with three layers is used to decide if a certain musical piece belongs to one of the four categories. The output of the neural network is placed in a vector (e.g. [1,0,0,0]) in which each component gives a score (0 to 1) on a certain mood. A sort of visualized music browser gives an overview of indexed music. The screenshots shown are not showing that this browser is more convenient than the songs categorized in (e.g. four) emotions though. The experiment carried out takes 330 musical pieces for training and only 23 for testing. This deteriorates the validity of the outcome of the experiment because the songs can be picked out on good performance. Another awkward aspect on the outcome of this experiment is that the precision percentages for the first three emotions (Happiness, Sadness and Anger) are all relatively high (75% - 86%) and that Fear only reaches 25%. This also shows that 23 songs are not enough because only three songs are marked as having a “fear-ish” mood.

6.5 Liu, Lu and Zang [LLZ2003]

Automatic Mood Detection from Acoustic Music Data

Mood detection for a specified part of music, namely classical music is the main subject in this paper. The algorithm is based on Thayer’s model of mood. Features of intensity, timbre and rhythm are used. Intensity is mapped to energy and both timbre and rhythm are mapped to the stress component (see figure 2). For the timbre spectral shape features such as Centroid, Roll off and Spectral Flux are used. The intensity is represented by the Root Mean Square value in decibels in each sub band and these values summed. The rhythm is extracted from the lower sub bands with a Canny estimator (this feature finds the sharp edges in the signal, the actual beat). This is a simple beat detection. From these beats, the strength, regularity and average tempo are devised. Two frameworks are given, a hierarchical and a non hierarchical framework. In the first framework musical pieces are firstly shifted on intensity into group 1 and group 2 (Depression/Contentment and Exuberance/Anxious respectively). The second step makes the distinction between the 2-tuples. In the second framework all features have their impact at the same time.

The interesting thing about this paper is that not only the mood or emotion in the 30 seconds after the initial 30 seconds is detected but there is also a so called mood tracking algorithm. Because the emotion or mood can differ from time to time in one musical piece especially in classical music this is a very relevant feature. For evaluation 250 pieces of music are used which are split up by music experts in clips of 20 seconds. This yields 800 clips of 20 seconds of which 75% is used for training and 25% for testing. This research manages to reach a high to very high accuracy ranging from 76.6% to 94.5% for the hierarchical framework and 64.7% tot 94.2% for the non-hierarchical framework. The accuracy is very high but the algorithm is trained on classical music only and with four moods it hasn’t got a very high granularity.

6.6 Yang and Lee [YL2004]

Disambiguating music emotion using software agents

The motivation behind this research is that music annotation poses too much pressure on listeners. The goal is to make music annotation (emotion in particular) easier but still provide a human input. As a fundament for emotion detection the Teller-Watson-Clark emotion model is used. [TWC1999] (See also section 4.) The researchers want to focus on the more negative emotions in this model because these emotions would be harder distinguishable than positive emotions. Wavelet tools, timbral features (section 3.1) and BPM detection methods were used to extract the necessary features. Also twelve features from the EDS system from Sony [PZ2004] were used. The resulting correlation between the annotated emotion and the results from the algorithm is almost 0.90. The best features according to the researchers are BPM detection and Sum of Absolute Values of the Normalized FFT. This acoustic based algorithm was used to categorize the songs into two classes of emotions corresponding to the TWC emotion model respectively Hostility, Sadness, Guilt and Love, Excitement, Pride. Secondly a quite unique approach is taken: musical pieces with annotated lyrics are analyzed. For each emotion a list of related key words of word classes (e.g. "political words" or "color words") is formulated. Of the total of 152 Alternative Rock musical pieces of 30 seconds 145 contained lyrics. The clips that contained lyrics were analyzed by the text mining algorithm which differentiates between songs which were in the same category of emotions. This yielded a high total accuracy of 82.8%.

6.7 Comparison of Methods

For all articles the most important and relevant features are summarized in table 1. Every article has a rating for each issue. The scale goes from -- (very bad) trough +/- (moderate) to ++ (excellent).

	CC2005	LO2003	LO2004	FZP2003/2	LLZ2003	YL2004
Precision	+	+/-	+	++	+/++	++
Granularity	-	++	+/-	-	-	+ *
Diversity	+/-	++	+/-	+/-	+/-	--
Selection**	-	-	-	-	++	-

Table 1 Comparison of methods

(* This method requires annotated lyrics to differentiate between emotions further than the two main categories of emotions. Without lyrics the granularity of this method would be much lower.

** This item only reflects the 20-30 seconds selection, not the conversion to mono sound.)

Between the reviewed methods there is a relatively strong negative correlation between the granularity in emotions and the precision that is achieved. So when the precision is high (e.g. FZP2003/2 and LLZ2003) the granularity has to be low (around 4 emotions). Otherwise, when precision is low (around 50% for LO2003) the granularity of the emotions is very high (13 emotions). In the case of the physiotherapist where only motivating music is wanted or the call center this low granularity and high precision is not a problem at all. Adversely, listeners at home or on the way have their own emotions they want to listen to. A higher granularity is needed here (such as LO2003 or YL2004).

The diversity stands for the diversity of music the algorithm is optimized for or tested on. Some methods (e.g. YL2004) only use one genre of music; others (e.g. LO2003) use a wider variety of music with all major genres in music present. Interesting to see is that the precision of [YL2004] is very high and the granularity is (with a trick) relatively high too. This can be explained by the fact that only alternative rock songs were used for testing. Optimization on one genre is obviously can be higher than when more genres are used such as in [LO2003]. In this research granularity and diversity are high but precision is lower.

The selection issue deals with the 20 to 30 seconds of selection in a song. This way, only one emotion for those 30 seconds will be found. Only [LLZ2003] proposes a mood tracking algorithm which returns an emotion for every 20 seconds of music.

The issue on mobile use is not so relevant. This is because all methods were in an experimental stage and were aimed at usage on a PC. Probably it is possible to convert these methods for implementation on mobile devices (advanced (multi)media players). To actually use this software would require a lot more computing power and memory capabilities than the current high-end devices have. Eventually this will be possible because the techniques for high-speed chips for mobile use are ready.

The learning aspect differs not so much between the reviewed methods. All require a large corpus of test songs and someone who evaluates it on being one emotion or the other.

That brings us to the last issue, the cultural background of the (future) user. There is only one research that evaluated their algorithm with more than one (actually two) test subject. There was a difference in achieved accuracy and the researchers thought that it was due to cultural background (no further details given). There still is a lot of research open for differences in emotion and the way emotions are expressed and evoked between different cultures and ethnic groups.

7. CONCLUSIONS AND DISCUSSION

There is no perfect method for emotion detection in music and probably there never will be. Emotion is something very personal and maybe it is for the better that computers will never completely understand a human.

However, when looking at a single application there are methods that will suffice. Possible applications of emotion detection methods, frequently used acoustical features and different emotion theories are reviewed. The important issues on emotion detection are compared between the reviewed methods. This yields a comparison from which the best method for a certain application can be deducted.

There is plenty of room for future research on this topic. A good start to make a quantitative evaluation of all the algorithms is making a standard corpus of songs that spans across all musical genres. The songs in this corpus should have the evoked emotions stored in a database. Also the set of emotions that can be appointed to a song should be standardized, preferably based on a good and complete psychological model of the human emotion.

REFERENCES

[CC2005] V. Carvalho and C. Chao, "Sentiment Retrieval in Popular Music based on Sequential Learning" 2005

[CL2004] B. Y. Chua and G. Lu, "Determination of Perceptual Tempo of Music" 2004 *CMMR 2004*

[Dow2003] J. Downie "Music Information Systems" 2003 *Annual Review of Information Science and Technology 37*: 295-340.

[EA2002] H. Elfenbein and N. Ambady, "On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis" 2002 *Psychological Bulletin Vol. 128, No. 2, 203–235*

[FZP2003] Y. Feng, Y. Zhuang and Y. Pan "Popular Music Retrieval by Detecting Mood" in *SIGIR'03, July 28-August 1, 2003, Toronto, Canada.*

[FZP2003-2] Y. Feng, Y. Zhuang and Y. Pan "Music Retrieval by Detecting Mood via Computational Media Aesthetics" 2003 in *Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI'03)*

[Hev1936] K. Hevner "Experimental studies of the elements of expression in music" 1936 *American Journal of Psychology*

[Hur2000] D. Huron "Perceptual and cognitive applications in music information retrieval." 2000 *International Symposium on Music Information Retrieval.*

[JL2003] P. Juslin, P. Laukka "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" 2003 Psychological Bulletin

[KPW2004] E. Knees, E. Pampalk and G. Widmer. "Artist classification with web-based data." 2004 *ISMIR2004*

[Lav2001] M. Lavy "Emotion and the Experience of Listening to Music A Framework for Empirical Research" April 2001

[Laz1991] R. Lazarus "Progress on a cognitive-motivational-relational theory of emotion." 1992 *American Psychologist*

[LLZ2003] D. Liu, L. Lu and H-J. Zang, "Automatic Mood Detection from Acoustic Music Data" 2003

[LO2003] T. Li and M. Ogihara, "Detecting Emotion in Music" 2003 *ISMIR2003*

[LO2004] T. Li and M. Ogihara "Content-based music similarity search and emotion detection" 2004

[LOL2003] T. Li, M. Ogihara and Q. Li "A Comparative Study on Content-Based Music Genre Classification" 2003 *SIGIR03*

[MB2003] M. F. McKinney, J. Breebaart, "Features for Audio and Music Classification", 2003 *Proc. of the InternationalConf. on Music Information Retrieval, 2003*

[Ove2006] J. M. Overgoor "An evaluation method for audio beat detectors", 2006 *4th Twente Student Conference on IT.*

[PGB1998] I. Peretz, L. Gagnon, B. Bouchard. "Music and emotion: perceptual determinants, immediacy and isolation after brain damage" Cognition 68 (1998) 111–141

[PZ2004] F. Pachet and A. Zils "Evolving Automatically High-Level Music Descriptors from Acoustic Signals" 2004 *Lecture Notes in Computer Science* Springer-Verlag GmbH

[RRSC1998] S. Rossignol, X. Rodet, J. Soumagne, J. L. Collette, P. Depalle, "Feature extraction and temporal segmentation of acoustic signals" 1998 *Proc. ICMC*

[RS1992] R. Reisenzein and W. Schonpflug "Stumpf 's Cognitive-Evaluative Theory of Emotion" 1992 *American Psychologist*

[Tha1989] R. Thayer. "The biopsychology of mood and arousal" 1989. *Oxford University Press.*

[TWC1999] A. Tellegen, D. Watson and L.Clark. "On the dimensional and hierarchical structure of affect." 1999 *Psychological Science*

[Väs2003] D. Västfjäll. "The Subjective Sense of Presence, Emotion Recognition, and Experienced Emotions in Auditory Virtual Environments" 2003 *CYBERPSYCHOLOGY&BEHAVIOR Volume 6, Number 2*

[Wik2005-1]

Wikipedia "Mel Frequency Bands"

http://en.wikipedia.org/wiki/Mel_Frequency_bands Wikipedia, last visited: 18-12-2005

[Wik2005-2]

Wikipedia "Wavelet"

http://en.wikipedia.org/wiki/Wavelet Wikipedia, last visited: 17-12-2005