Extracting patterns from tracks using emotional tags

Michael Kai Petersen and Andrius Butkus

DTU Informatics
Technical University of Denmark, Building 321, 2800 Lyngby, Denmark
{mkp,ab}@imm.dtu.dk
http://www.imm.dtu.dk

Abstract. Despite the idiosyncratic character of tags applied to songs in social networks like *last.fm*, recent studies have revealed that users often tend to agree on the affective terms they attach to music. Using some of these frequently occurring words as emotional buoys to form a semantic plane of psychological valence and arousal dimensions, we project lyrics into this space and apply LSA latent semantic analysis to model the affective context of a number of songs. We compare the components retrieved from the lyrics with the user defined affective terms that constitute the tag clouds of the corresponding songs at *last.fm*, and propose that LSA could be applied to extract structural patterns as a basis for automatically generating emotional playlists.

1 Introduction

Musical meaning evolves dynamically as we discover self-organizing patterns transforming our memories of what just faded away into prior probabilities for what is likely to emerge. In the case of sounds by transforming the incoming temporal waveforms from a single dimensional representation of the world, based on the Newtonian mechanics of our ear drums, into an auditory scene unfolding over time as it ascends through the neuroaxis towards the auditory cortex. Over the past half century aspects of musical affect have been the focus of a wide field of research, ranging from how emotions arise based on the underlying harmonic and rhythmical hierarchical structures forming our expectations [1][2][3], to how we consciously experience these patterns empathetically as contours of tensions and release [4], in turn triggering physiological changes in heart rate or blood pressure as has been documented in numerous cognitive studies of the links between music and emotions [5]. Neuroimaging studies have established that musical structure to a larger extent than previously thought is being processed in “language” areas of the brain, and that shared neural resources between music and language indicate semantic processing of patterns reflecting tension and resolution [6-9]. Specifically related to songs both fMRI magnetic resonance imaging and ERP evoked response potential results point to linguistic and musical dimensions as being processed by similar overlapping brain areas, which seems to support the hypothesis that the linguistic and melodic components of songs are processed in interaction [10]. Additional studies on retrieving songs from memory indicate that lyrics and melody appear to be recalled from two separate versions: one storing the melody and another containing only the text [11], while further priming experiments indicate that song memory is not organized in strict temporal order, but rather that text and tune intertwine based on reciprocal connections of higher order structures [12].

As it has been suggested that the human mirror neuron system may provide a mechanism for cross-modal processing of language, action and music as a basis for conveying affect [13], we speculate that it might be possible to model the emotions elicited by songs by extracting latent semantics from the lyrics. Such an approach to retrieve high level representations of songs based on the lyrics alone would be in line with neuroimaging results, showing that retrieval of perceptual knowledge related to sound is associated with increased activity in areas inferior and posterior to the primary auditory cortex [14]. It would additionally be substantiated by recent fMRI experiments focused on predicting human brain activity reflecting the meaning of nouns, which have demonstrated a direct relationship between the observed patterns and the statistics of word co-occurrence in large collections of documents, meaning that brain activity can be modeled as linear sums of contributions associated with semantic features of input stimulus words [15]. Applying a similar linear dimensionality reduction method to extract meaning in texts based on LSA latent semantic analysis [16], which resembles cognitive text comprehension by capturing patterns of word usage in multiple contexts, we hypothesize it might be possible to model a high-level emotional representation of the underlying structure which cause the shifting contours of tension and release in songs.

Outlining our approach we describe in the following sections: the methodology used for extracting latent semantics within an emotional space, the early results retrieved when mapping the constituent emotional components over time, and conclude with a discussion formulating our hypothesis.

2 Emotional tag space

Despite the often idiosyncratic character of tags defined by hundred thousands of users in social networks like *last.fm*, several studies done within the music information retrieval community have confirmed that users tend to agree on the affective terms they attach as tags to describe music [17-18]. The affective terms which are frequently chosen as tags by *last.fm* users seem to form clusters around primary moods like mellow, sad, or more agitated feelings like angry and happy. Drawing on Osgood, Suci and Tannenbaums earlier findings, which have established that emotional assessment can be represented based on the two dimensions of *valence* and *arousal* [19], we use these two parameters to outline an emotional plane containing four groups of frequently used *last.fm* tags:

happy, funny, sexy
romantic, soft, mellow, cool

angry, aggressive
dark, melancholy, sad

The dimension of *valence* describes how pleasant something is along an axis going from positive to negative associated with words like happy or sad, whereas *arousal* captures the amount of involvement ranging from passive states like mellow and sad to active aspects of excitation as reflected in tags like angry or happy. Applying these twelve frequently used emotional *last.fm* tags, as buoys to define a semantic plane of psychological valence and arousal dimensions, we apply LSA latent semantic analysis to assess the correlation between the lyrics and each of the selected affective terms. Selecting these affective terms simultaneously enables us to implement them as a reference to compare the LSA retrieved values of correlation against how frequently users actually apply these words in *last.fm* tag clouds to describe the emotional context of songs. But whereas users at *last.fm* attach emotional tags to a song as a whole, we aim to model the shifting contours of tension and release which trigger emotions based on each of the individual lines making up the lyrics. Taking into consideration that words are hierarchically perceived as successive phonemes and vowels on a scale of roughly 30 milliseconds, which are in turn integrated into larger segments with a length of approximately 3 seconds [20], we assume that lines of lyrics consisting of around 5 words each approximately corresponds to one of these higher level perceptual units. Projecting the lyrics into a semantic LSA space line by line, could in a cognitive sense be interpreted as similar to how mental concepts are constrained by the amount of activation among neural nodes in working memory reflecting events and our associations, as proposed by Kintsch in his construction-integration theory of comprehension [21]. In that respect the co-occurrence matrix formed by the word frequencies of *last.fm* tags and song lyrics, might be understood as corresponding to the strengths of links connecting nodes in a mental model of semantic and episodic memory.

As a machine learning technique LSA extracts meaning from paragraphs by modeling the usage patterns of words in multiple documents and represent the terms and their contexts as vectors in a high-dimensional space. The basis for assessing the correlations between lyrics and emotional words vectors in LSA is an underlying text corpus consisting of a large collection of documents which provides the statistical basis for determining the co-occurrence of words in multiple contexts. For this experiment we chose the frequently implemented standard *TASA* text corpus, consisting of the 92409 words found in 37651 texts, novels, news articles and other general knowledge reading material that American students are exposed to up to the level of their 1st year in college.The frequency at which terms appear and the phrases wherein they occur are defined in a matrix with rows made up of words and columns of documents. Many of the cells made up by rows and columns contain only zeroes, so in order to retain only the most essential features the dimensionality of the original sparse matrix is reduced to around 300 dimensions. This makes it possible to model the semantic relatedness of song lyrics and affective terms as vectors, with values towards 1 signifying degrees of similarity between the items and low or minus values typically around 0.02 signifying a random lack of correlation. In this semantic space lines of lyrics or emotional words which express the same meaning will be represented as vectors that are closely aligned, even if they they do not literally share any terms. Instead these terms may co-occur in other documents describing the same topic, and when reducing the dimensionality of the original matrix the relative strength of these associations can be represented as the cosine of the angle between the vectors.

3 Results

Selecting the lyrics of thirty songs as input, we compute the cosine values between vectors representing each of the individual lines constituting the lyrics of a given song against each of the twelve affective terms used as markers in the LSA space, and discard cosine values of correlation between lyrics and tags below a threshold of 0.09.

3.1 Accumulated distribution of emotional components

In order to compare the retrieved LSA correlation between lyrics and affective terms against the actual tag cloud attached to the song at *last.fm*, we first sum up the LSA values retrieved from each line of the lyrics.

Taking the song "Nothing else matters" as an example, the actual affective terms attached to the song at *last.fm* also include less frequently used tags like *love, love songs, chill, chillout, relaxing, relax, memories* and *melancholic*, so we subsequently combine these tags into groups of associated tags in order to facilitate a direct comparison with the LSA retrieved values:

happy, funny, sexy, romantic + (love, love songs)
soft, mellow, cool + (chill, chillout, relaxing, relax)
angry, aggressive
dark, melancholy, sad + (melancholic, memories)

Comparing the accumulated LSA values of emotional components against the reference tags at *last.fm*, the terms *melancholy* and *melancholic* which describe the most dominant emotions in the tag cloud seem to be captured by the affective term *sad* in the LSA analysis. When interpreting *love* from the *last.fm* tag-cloud as associated with the term *happy* (based on a cosine correlation of 0.56 between the words love and happy) the LSA analysis appears similarly to retrieve aspects of this emotion. Likewise if *chill* in the *last.fm* tag cloud is understood as associated with *soft* and *mellow* (based on cosine correlations of 0.36 and 0.35 respectively) the LSA analysis also here appears to capture that mood. On the other hand there seems to be no tags in the *last.fm* tag-cloud supporting the affective terms *angry* and *aggressive* resulting from the LSA analysis of the lyrics.

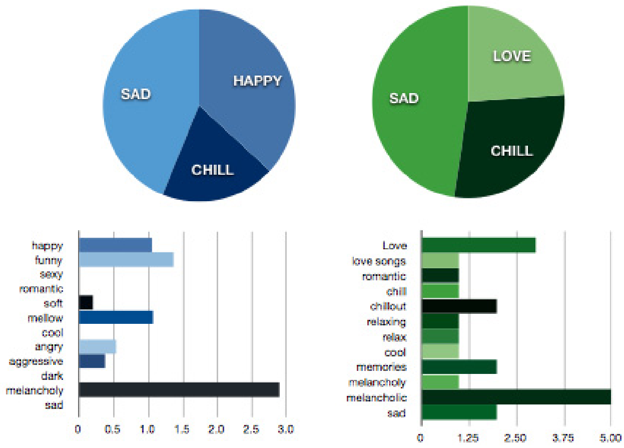

Fig. 1. Summed up values of LSA correlation between the lyrics of the song "Nothing else matters" and 12 affective terms (top), compared to the actual distribution of user-defined emotional tags attached to the song at *last.fm* (bottom).

Applying a similar approach to a set of 25 songs we grouped *last.fm* tags into larger segments consisting of *sad*, *happy*, *love* and *chill* aspects to facilitate a comparison with the LSA derived correlations between song lyrics and the selected affective terms. Though we found a significant overlap between LSA values and user defined *last.fm* tags in most of the songs,we were not able to define a general correlation between LSA retrieved values and the distribution of *last.fm* tags for each song based on the small sample in this study.

3.2 Distribution of emotional components over time

Maintaining the LSA values retrieved from each of the individual lines in the lyrics, we proceeded by plotting the values over time to provide a view of the distribution of emotional components. The plots can be interpreted as mirroring the structure of patterns of changing tension in the songs along the horizontal axis. Vertically the color groupings indicate which of the aspects of valence and arousal are triggered by the lyrics as well as their general distribution in relation to each other. Any color will signify an activation beyond the cosine similarity threshold level of 0.09, and the amount of saturation from light to dark signifies the degree of correlation between the song lyrics and each of the affective terms.

The contribution of each emotional component apparent in the overall LSA values of the lyrics, can be made out when considering their distribution as single pixels over time triggered by the individual lines in each of the songs. Analyzing which components are predominant and their overall contribution in the lyrics,

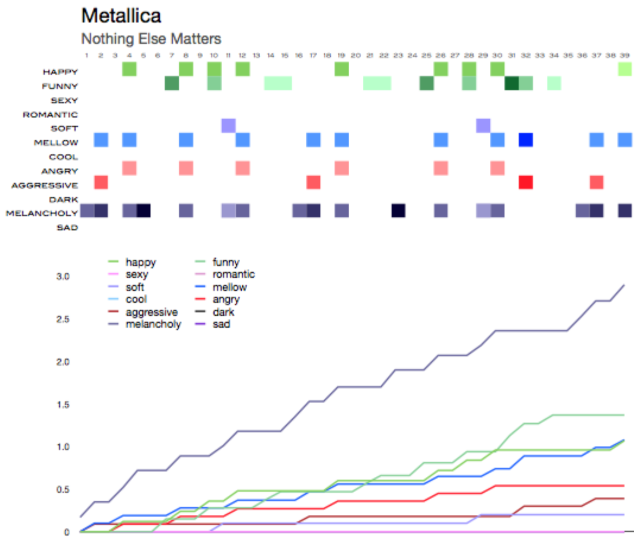

Fig. 2. LSA correlation between each of the lines in the lyrics of "Nothing else matters" and 12 affective terms plotted over the entire duration of the song - unbalanced sparse distribution with an increasing sustained bias towards the extreme emotions of *melancholy* juxtaposed against *happy* and *funny* while the more central *mellow* and *soft* elements appear subdued.

the LSA plots can roughly be grouped into the three categories of *unbalanced distributions*, *centered distributions* and *uniform distributions*.

The song "Nothing else matters", Fig.2, exemplifies the first category by having a bottom-heavy distribution of emotional components biased towards melancholy. The below curve indicates the contribution of each component over the duration of the song, where the dominant aspect of melancholy can again be made out. A centered distribution as found in "Now at last" Fig.3, shows a lack of the more extreme emotions like *happy* or *sad* and instead the main contribution comes from mellow and soft aspects. In the below diagram the curves of emotional contributions appear sustained combined with more steep increases between each level. A uniform distribution of a wide range of simultaneous emotional components is exemplified by "Mad world", Fig.4. The sudden steep rise in the below curve diagram also illustrates the overall structure in the song.

In the following pairwise comparison of song samples within categories of distributions, Fig 5 - 9, we find it remarkable that the overall saturation defining the amount of correlation between lyrics and emotional markers, as well as the distributional patterns of emotional components throughout the songs seem

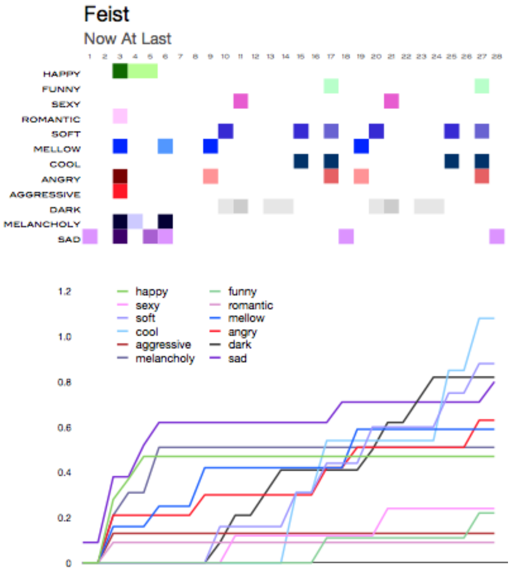

Fig. 3. LSA correlation between each of the lines in the lyrics of "Now at last" and 12 affective terms plotted over the entire duration of the song - centered distribution stressing *soft* and *mellow* aspects periodically clustered with additional *cool* and *angry* elements - the *last.fm* tag cloud includes "mellow, sad, chillout, melancholy, love, quiet, dreamy, relax, slow, soft, sweet, wistful"

consistent. Lyrics that appear more or less saturated in relation to the emotional markers used for the LSA analysis remain so over the entire song. The distributional patterns of emotional elements seem throughout the songs to form consistent schemas of contrasting elements,which appear to form sustained lines or clusters that are preserved as pattern once initiated. The balanced distribution of emotional components in the songs "Wonderwall" and "My Immortal", Fig.5, both lack central aspects like *soft*, instead emphasizing the outer edges by juxtaposing elements around *happy* against *sad*. This sparsity can also be made out in the separation of curves in the diagrams below. Otherwise so in the central distributions exemplified by the songs "Falling slowly" and "Stairway to heaven", Fig.6, which emphasize the aspects of *soft* and *mellow* at the expense of *happy* and *sad*. In the uniform distributions of the songs "Everybody hurts" and "Smells like teen spirit", Fig.7, both of the patterns display sustained lines while grouping elements into clusters, reflecting a strong continuous activation of emotional components. The highl levels of saturation also apparent in the curve diagrams, similarly signify an overall high correlation of the lyrics against the

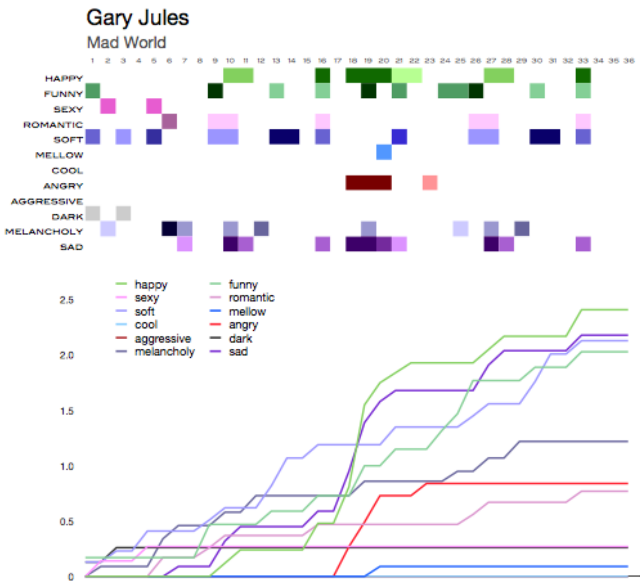

Fig. 4. LSA correlation between each of the lines in the lyrics of "Mad world" and 12 affective terms plotted over the entire duration of the song - uniform distribution juxtaposing sutained *happy* and *funny* elements against *sad* and *melancholy* elements coupled with clusters of *soft* - the *last.fm* tag cloud includes "sad, melancholy, mellow, chillout, calm, dark, depressing, emotional, love, relax, slow, soft, touching".

affective terms throughout the entire songs. In contrast the patterns "Rehab" , "Time to pretend", "21 Things" and "Creep", Fig.8 & 9, come out as much more periodic, visible in the steep increases of curves while simultaneously maintaining high levels of saturation. We speculate that these apparent consistencies reflect underlying compositional structures which reflect patterns of tension and release. Each column in the plots of lyrics corresponds to a time window of roughly 3 seconds, which as before mentioned is the approximate length of the high level units from which we mentally construct our perception of time [20]. Understood in that context we hypothesize that the LSA analysis of lyrics within a similar size of time window is able to capture a high level representation of the shifting emotions triggered by the lyrics when listening to songs. And that the LSA analysis when projecting the lyrics against the affective terms, approximates a very simplified model of how each of the lines in eral time would trigger emotions, which depending on their strengths of associations would be activated in our working memory.

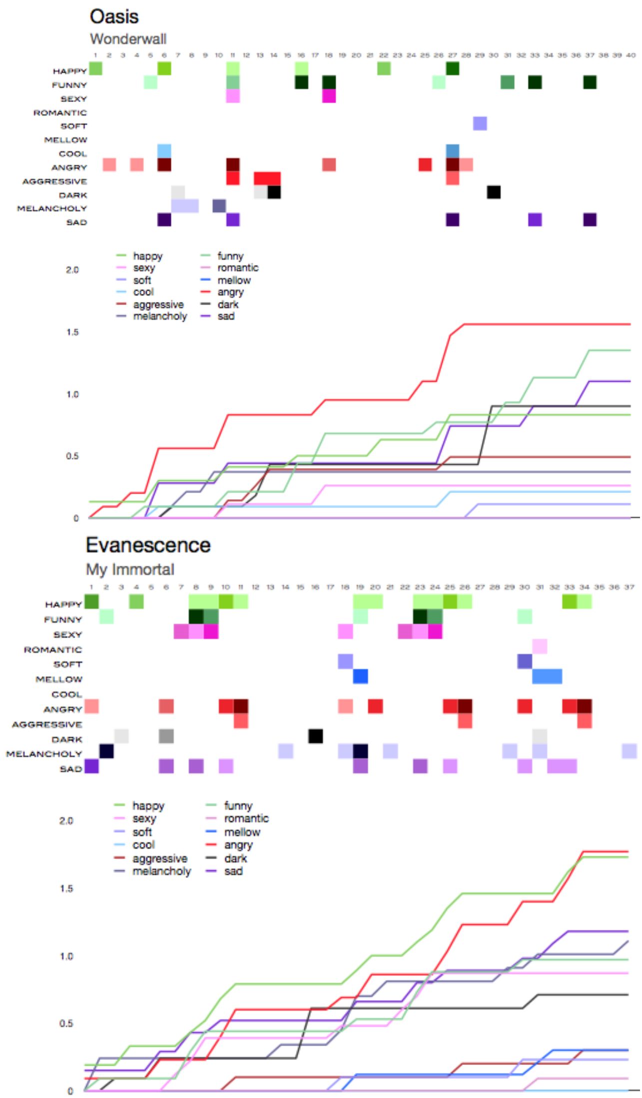

Fig. 5. LSA cosine similarity of the lyrics of "Wonderwall" (top) and "My Immortal" (bottom) mapped against 12 affective terms line by line over the entire songs

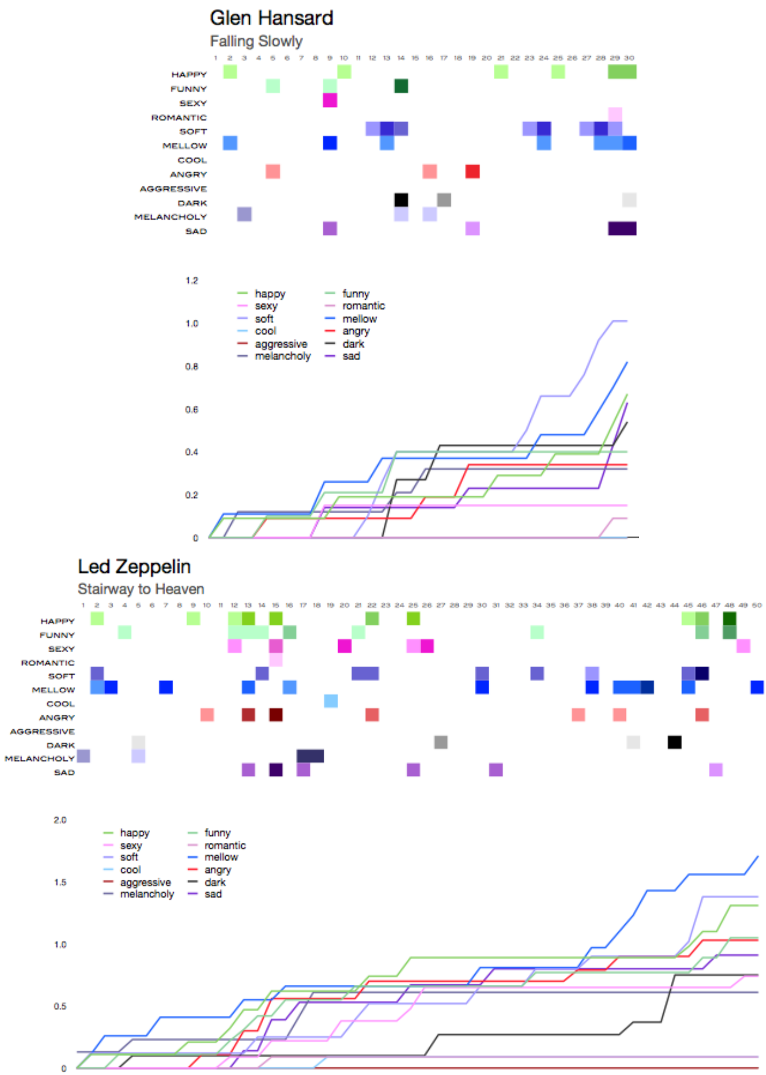

Fig. 6. LSA cosine similarity of the lyrics of "Falling slowly" (top) and "Stairway to heaven" (bottom) mapped against 12 affective terms line by line over the entire songs

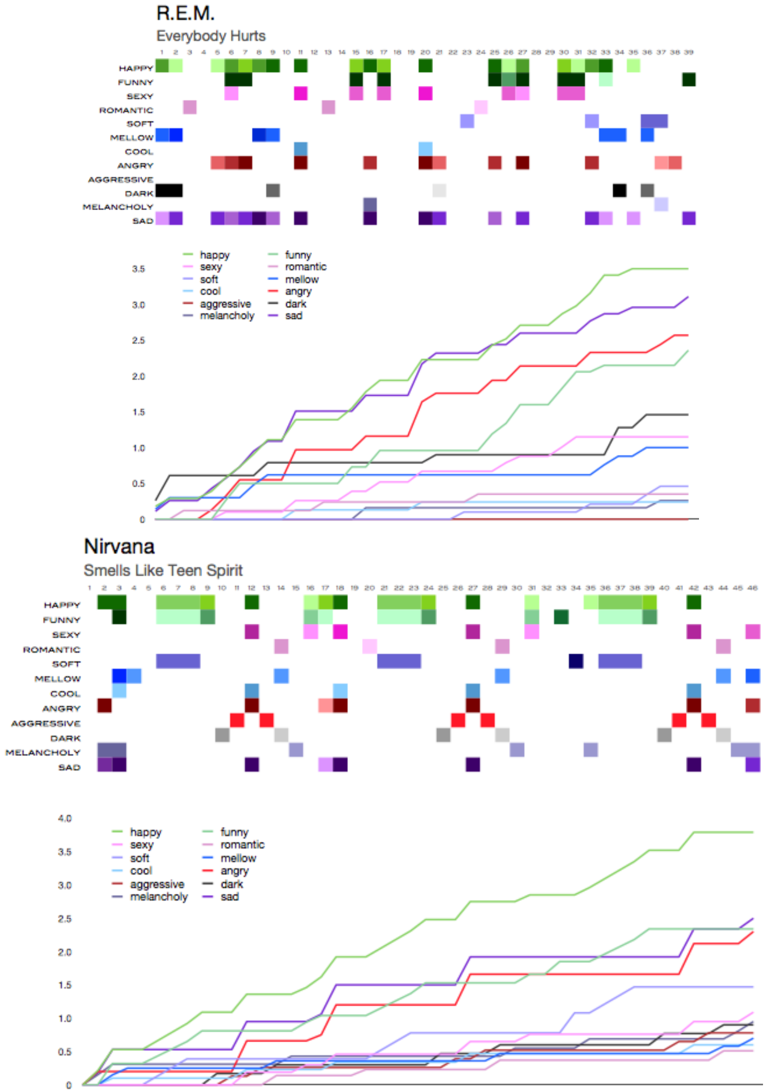

Fig. 7. LSA cosine similarity of the lyrics of "Everybody hurts" (top) and "Smells like teen spirit" (bottom) mapped against 12 affective terms line by line over the entire songs

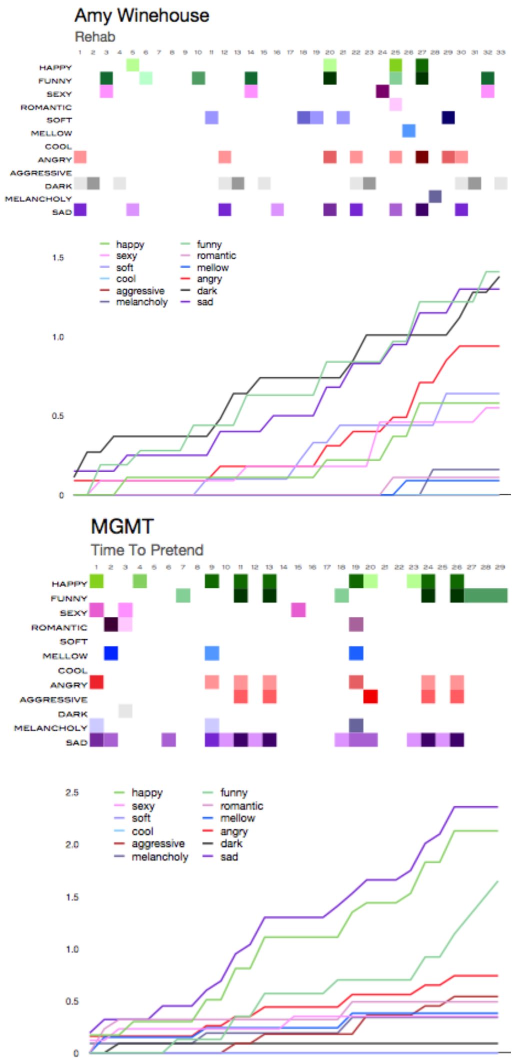

Fig. 8. LSA cosine similarity of the lyrics of "Rehab" (top) and "Time to pretend" (bottom) mapped against 12 affective terms line by line over the entire songs

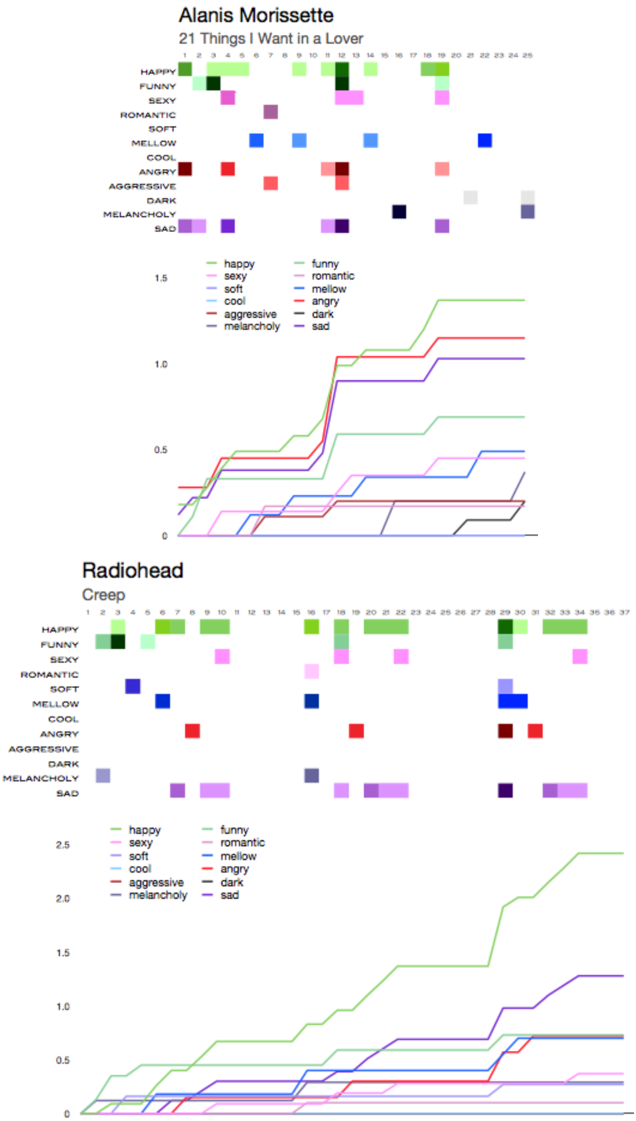

Fig. 9. LSA cosine similarity of the lyrics of "21 Things i want in a lover" (top) and "Creep" (bottom) mapped against 12 affective terms line by line over the entire songs

4 Discussion

While we have here only analyzed a small number of songs our first results indicate that it appears possible to retrieve the emotional elements contributing to user descriptions of songs, by applying LSA to extract latent semantics from the lyrics using a selection of frequently occurring *last.fm* affective tags. When comparing the LSA accumulated emotional components extracted from the lyrics with the actual user defined tag clouds of the corresponding songs at *last.fm* they appear to a large extent overlapping or complementary related to the affective terms when grouped into larger segments consisting of *sad*, *happy*, *love* and *chill* aspects to facilitate a comparison with the LSA derived correlations. However this type of generalized classification is not capable of capturing the affective components reflected in the lyrics over the entire duration of a song as illustrated by the plots of emotions triggered over time.

Considering the consistency of the patterns in the LSA analysis of song lyrics in the above matrix examples, combined with recent neuropsychological findings indicating that music and language appear both syntactically and semantically to be processed in interaction, we speculate that our approach might provide a basis for modeling a high level representation of how emotions arise based on the underlying structures in songs. We propose that these components can be retrieved as latent semantics by using affective terms as sensors in a semantic space, and hypothesize that LSA might be applied to extract structural patterns from song lyrics as a basis for automatically generating emotional playlists. It seems that even if we turn off the sound both emotional context as well as overall formal structural elements can be extracted from songs based on latent semantics.

References

1. L. B. Meyer: *Meaning in music and information theory*, Journal of Aesthetics and Art Criticism, Vol.15(7) pp. 412-424, 1957
2. D. Temperley: *Music and probability*, MIT Press, 2007
3. D. Huron: *Sweet anticipation: Music and the psychology of expectation*, MIT Press, 2006
4. R. Jackendoff and F. Lerdahl, *The capacity for music: what is it, and what's special about it?*, Cognition, Vol.100(1) pp. 33-72, 2006
5. C.L. Krumhansl, *Music: A link between cognition and emotion*, Current Directions in Psychological Science, Vol.11(2) pp.35-55, 2002
6. D.J. Levitin and V. Menod: *Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence*, NeuroImage, Vol.20(4) pp. 2142-2152, 2003
7. S. Koelsch and W.A. Siebel: *Towards a neural basis of music perception*, Trends in Cognitive Sciences, Vol.9(12) pp. 578-584, 2005
8. N. Steinbeis and S. Koelsch: *Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns*, Cerebral Cortex, Vol.18(5) pp. 1169-1178, 2008

9. L.R. Sleve, J.C. Rosenberg and A.D. Patel: *Language, music and modularity*, Evidence for shared processing of linguistic and musical syntax, Proceedings of the 10th International Conference on Music Perception & Cognition, 2008
10. D Schön, R.L. Gordon, and M. Besson: *Musical and linguistic processing in song perception*, Annals of the New York Academy of Sciences, Vol.1060 pp. 71-81, 2005
11. I. Peretz, R. Gagnon, and S. Hbert: *Singing in the brain: Insights from cognitive neuropsychology*, Music Perception, Vol.21(3) pp. 71-81, 2004
12. I. Peretz, M. Radeau, and M. Arguin: *Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs*, Memory & Cognition, Vol.32(1) pp. 142-552, 2004
13. I. Molnar-Szakacs and K. Overie: *The brain's concepts: the role of the sensory motor system in conceptual knowledge*, Social cognitive and affective neuroscience, Vol.1(33) pp. 235-241, 2006
14. R.F. Goldberg, C.A. Perfetti and W. Schneider: *Perceptual knowledge retrieval activates sensory brain regions*, Journal of Neuroscience, Vol.26(18) pp. 4916-4921, 2006
15. T. Mitchell, S. Shikareva, A. Carlson, K. Chang, V.L. Malave, R.A. Mason and M.A. Just: *Predicting human brain activity associated with the meanings of nouns*, Science, Vol.320(5880) pp. 1191-1195, 2008
16. L.K. Landauer and S.T. Dumais: *A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge*, Psychological Review, Vol.104(2) pp. 211-240, 1997
17. M. Levy and M. Sandler, *A semantic space for music derived from social tags*, Proceedings of the 8th International Conference on Music Information Retrieval, pp. 411-416, 2007
18. X. Hu, M. Bay and S.J. Downie: *Creating a simplied music mood classication ground-truth set*, Proceedings of the 8th International Conference on Music Information Retrieval, pp. 309-310, 2007
19. C.E. Osgood, G.J. Suci and H.P. Tannenbaum: *The measurement of Meaning*, University of Illinois Press, 1957
20. E. Pöppel: *A hierarchical model of temporal perception*, Trends in Cognitive Sciences, Vol.1(2) pp. 56-61, 1997
21. W. Kintsch: *Comprehension - a paradigm for cognition*, Cambridge University Press, 1998