DTU Library

Semantic Contours in Tracks Based on Emotional Tags

Petersen, Michael Kai; Hansen, Lars Kai; Butkus, Andrius

Link back to DTU Orbit

Citation (APA):
Petersen, M. K., Hansen, L. K., & Butkus, A. (2009). Semantic Contours in Tracks Based on Emotional Tags. In *Computer Music Modeling and Retrieval.: Genesis of Meaning in Sound and Music* (1st ed., Vol. Volume 5493/2010, pp. 45-66). Springer. Lecture Notes in Computer Science Vol. 5493 https://doi.org/10.1007/978-3-642-02518-1_3

Semantic contours in tracks
based on emotional tags

Michael Kai Petersen, Lars Kai Hansen, and Andrius Butkus

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

Abstract. Outlining a high level cognitive approach to how we select media based on affective user preferences, we model the latent semantics of lyrics as patterns of emotional components. Using a selection of affective *last.fm* tags as top-down emotional buoys, we apply LSA latent semantic analysis to bottom-up represent the correlation of terms and song lyrics in a vector space that reflects the emotional context. Analyzing the resulting patterns of affective components, by comparing them against *last.fm* tag clouds describing the corresponding songs, we propose that it might be feasible to automatically generate affective user preferences based on song lyrics.

Key words: Pattern recognition, emotions, text processing

1   Introduction

Both words and music move in time, and as T.S. Elliot phrased it "only by the form, the pattern, can words or music reach". A panoply of sensations and emotions are elicited when we listen to a song, which in turn reflect cognitive aspects of the underlying structure in both sound and lyrics. Over the past half century these 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 structures forming our expectations [1-3], to how we consciously experience these patterns empathetically as contours of tensions and release [4]. Basic feelings of happiness, being sad or getting angry are not just perceived but materialize as changes in heart rate, skin conductance, respiration or blood pressure, as has been documented in numerous cognitive studies of music and emotions [5]. Applying biosensors to measure the features that underlie the various affective states, the resulting patterns appear sufficiently consistent to determine what emotions are being triggered based on the physiological changes alone [6]. But listening to songs involves not only basic elements of affect, but also higher level structures reflected in the lyrics which provide the basis for a song. To a large extent language allows us to share and describe distinct affective aspects that we extract from the continuous affective ebb and flow of emotions shaping our frame of mind. Despite the often idiosyncratic character of tags defined by

hundred thousands of users in social networks like *last.fm*, a number of studies within the music information retrieval community indicate that users often tend to agree on the affective terms they attach to music, which can be interpreted as a simplified mood ground-truth reflecting the perceived emotional context of the music [7-8].

During the past decade advances in neuroimaging technologies enabling studies of brain activity have established that musical structure to a larger extent than previously thought is being processed in "language" areas of the brain [9], and specifically related to lyrical music some fundamental aspects appear essentially identical to those of language [10]. Neural resources between music and language appear to be shared both in syntactic sequencing and also semantic processing of patterns reflecting tension and resolution [11-13], adding support for findings of linguistic and melodic components of songs being processed in interaction [14]. Similarly there appears to be an overlap between language regions in the brain and so-called mirror neurons, which transfer sensory information of what we perceive by re-enacting them on a motor level. Mediating the inputs across audiovisual modalities, the resulting sensory-motor integrations are represented in a similar form, whether they originate from actions we observe in others, only imagine or actually enact ourselves [15-16]. This has led to the suggestion that our empathetic comprehension of underlying intentions behind actions, or the emotional states reflected in sentences and melodic phrases, are based on an imitative re-enactment of the perceived motion [17].

So if both low-level features of media and our emotional responses can be encoded in words, we hypothesize that this might allow us to define a high level cognitive model emulating how we select media based on affective user preferences. In such a model the bottom-up part would resemble cognitive component analysis [18]. Coined as a term to describe aspects of unsupervised clustering of data, the underlying algorithms approximate how our brain discovers self-organizing patterns when assembling images from lines and edges of visual objects [19], reconstructs words from the statistical regularities of phonemes in speech [20] or learn the meaning of words based on their co-occurrence within multiple contexts [21-23]. But equally important: cognitive processes involve a large amount of top-down feedback which sculpts the receptive responses of neurons on every level and vastly outnumbers the sensory inputs [24-26]. That is, the brain applies an analysis-by-synthesis approach, which combines a top-down capability to infer structure from bottom-up processing of statistical regularities in what we perceive. Our emotions are in this sense essential for maintaining a balance between cognition and perception, as core affect is an integral element in what attracts us to objects and turn what we sense into meaningful representations that can be categorized in words [27-29].

A way to emulate this approach of the human brain in relation to search of media, could be to apply unsupervised learning of features based on latent semantics, extracted from lyrics associated with songs. And combine the bottom-up extracted representation with top-down aspects of attention reflecting preferred emotional structures, similar to the combinations of user generated affec-

tive terms found in tag clouds in social networks like last.fm. Selecting a number of frequently used emotional *last.fm* tags as buoys to define a semantic plane of psychological valence and arousal dimensions, we project a number of song lyrics into this space and apply LSA latent semantic analysis [21-23], to model the correlation of texts and affective terms as vectors reflecting the emotional context of the songs. We outline in the following sections: the affective plane used for modeling emotional structure, the extraction of latent semantics from texts associated with media, an analysis of the emotional patterns, followed by a discussion of the potential in combining latent semantics and emotional components to enable personalized search of media.

2   Affective plane

Drawing on standard psychological parameters for emotional assessment, affective terms are often mapped out along the two psychological dimensions of valence and arousal [30-32]. Within this 2D emotional plane 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 terms like angry or happy. This approach to represent emotions within an affective space framed by valence and arousal dimensions goes beyond earlier attempts to define distinct categories like Hevner's circle of adjectives (1935). Based on responses from participants listening to musical excerpts, clusters of words were grouped into eight segments of similar adjectives covering emotions like happy, lyrical, calm, dreamy, sad, serious, tragic, angry and exciting [33]. How many different parameters are required to capture the various components in an affective space has since then been the subject of a number of studies. The results indicate that a model which provides a good fit to how people describe emotional states, can be defined based on five underlying latent variables: anger, sadness, disgust, fear and happiness [34]. In such a model these factors are not necessarily correlated with whether they are perceived as pleasant or unpleasant, as opposing aspects might often occur together even if they represent contrasting positive and negative aspects of valence. Empirical results for rating of emotional words, also indicate that certain terms e.g. synonyms for happy or anger seem to be based on one category only and are defined as either positive or negative along a single dimension. Whereas other affective terms appear more complex and appear to be combinations of more emotional categories, like despair being perceived as a mixture of sadness and anxiety, or excitement involving aspects of both happiness and surprise [35]. In any linguistic description we perceive not only the lexical meaning of words but infuse them with feelings of positive or negative valence [30], which might serve to filter what is essential and determine what becomes part of our memories [28]. Experiments using MDS multidimesional scaling to group musical excerpts according to similarity instead of word categories, indicated that the model which provides the best fit to the emotional responses seems again to be based on the

two psychological dimensions of valence and arousal, but here combined with an additional third dimension capturing aspects of shape related to the degree of melodic continuity or rhythmical fragmentation in the music [36]. In a further cluster analysis of the data, the participants were found to separate the responses into two segments along the arousal dimension, meaning how energetic or laid back the music was perceived as being. The similarity judgements of musical excerpts within the two clusters in turn appeared to be divided along the fault lines of valence, essentially separating the emotional reactions into an affective plane consisting of happy, mellow, angry and sad quadrants.

If we attempt to model top-down cognitive attentional aspects reflecting affective structure, tag-clouds in music social networks like *last.fm* provide an interesting case. The affective terms which are frequently chosen as tags by users to describe music seem to form clusters around primary moods like mellow, sad, or more agitated feelings like angry and happy. This correlation between social network tags and the specific music tracks they are associated with, has been used in the music information retrieval community to define a simplified mood ground-truth, reflecting not just the words people frequently use when describing the perceived emotional context, but also which tracks they agree on attaching these tags to [7-8]. Selecting twelve of these frequently used tags:

happy, funny, sexy, romantic
soft, mellow, cool
angry, aggressive
dark, melancholy, sad

makes it possible to define an affective plane reflecting the above cognitive models of emotional responses to music, as a basis for extracting latent semantics.

3   Semantic space

To generate the bottom-up part of how we cognitively extract meaning from strings of texts, LSA latent semantic analysis models comprehension from word occurrences in multiple contexts, analogous to human language acquisition [21-23]. Words rarely come shrink-wrapped with a definitive meaning but are continuously modified by the context in which they are set. No matter how many examples of word usage for a verb are listed in a dictionary they remain just that: case stories which illustrate how a predicate will map onto a certain value given a specific argument. Replacing any of the surrounding words in the sentence will create yet another instantiation of the proposition, which we might again interpret differently depending on what phrases come before or after in the text. Instead of attempting to define the specific meaning of a word based on how it fits within a particular grammatical phrase structure, LSA latent semantic analysis, models the plethora of meanings a word might have by concatenating all the situations in which it appears and represent them as a single vector within

in a high dimensional semantic space. Squeezing as many of the syntactic relations and senses of word usage into a single vector, makes it possible to extract statistical properties based on how often a term appears in a large number of paragraphs. And subsequently condense this representation into meaningful semantic relations constructed from an average of the different contexts in which the word is used.

Initially a text corpus is constructed which allows for modeling terms as linear combinations of the multiple paragraphs and the sentences in which they occur, assembled from tens of thousands of pages of literature, poetry, wikipedia and news articles. The underlying text corpora can be thought of as resembling human memory where numerous episodes combined with lexical knowledge are encoded into strings of text. Spanned by rows of words and columns of documents, the cells of this huge term-document matrix sum up how frequently each word appears in a corresponding paragraph of text. However in a simple co-occurrence matrix any similarities between words like car and vehicle will be lost as each individual term appears only within its own horizontal row. Nor will it be obvious that a word like rock might mean something completely different depending on which of the contextual columns it appears in. The raw matrix counts of how many times a word occurs in different contexts does therefore not by itself provide a model of comprehension, as we would normally expect texts that describe the same topic to share many of the terms that are used, or imagine that words that resemble each other are also applied in a similar fashion. Most of these relations remain hidden within the matrix, because there are tens of thousands of redundant variables in the original term-document matrix obscuring the underlying semantic structure. Reducing the dimensionality of the original matrix using SVD singular value decomposition [28], the number of parameters can be diminished so we can fit synonymous words or group similar documents into a much smaller number of factors that can be represented within a semantic space.

Geometrically speaking, the terms and documents in the condensed matrix derived from the SVD dimensionality reduction, can be interpreted as points in a k dimensional subspace, which enables us to calculate the degree of similarity between texts based on the dot product of their corresponding vectors. But before comparing terms or documents, the entries in the cells of the matrix need to be adjusted so they reflect how we cognitively perceive associative processes. First by replacing the raw count of how often a word appears in a text by the logarithm of that number. This will smooth the word frequency so it resembles the shape of learning curves typically found in empirical psychological conditioning experiments. Likewise the degree of association of two words both occurring in two documents will be higher than if they each appear twice separately in a text. Here a local weighting function defines how salient the word occurrence is in the corresponding document, and a global weighting function how significant its appearance is among all the contexts [37]. As a next step the word count is divided by the entropy of the term, to ensure that the term frequency will be modified by how much information the word actually adds about the context it

appears in. This log-entropy weighting significantly improves the results when compared to a raw word frequency count [23]. Another way to interpret the relations between words forming a semantic neighborhood would be to think of them as nodes constituting a neural network. In such a network of nodes, resembling populations of neurons in our brains, LSA could model the strength of the links connecting one word to another. When we come across a word like 'sad' in a phrase, it will create a node in our short term episodic memory, which will in turn trigger neighboring nodes representing words or events that invoke similar connotations in our past memories. The strength of the connections initially based on word co-occurrence are gradually transformed into semantic relations as the links between nodes are being constrained by the limitations of our memory. As a result only those nodes which remain sufficiently activated when our attention shifts towards the next phrase will be integrated into the patterns forming our working memory. And whether these connections grow sufficiently strong for the nodes to reach a threshold level of activation necessary for being integrated in working memory, can be seen as a function of the cosine between the word vectors [38].

When we compare two terms in the LSA semantic space based on the the cosine of the angle between their vectors, values in-between 0.05 and 1 will indicate increasingly significant degrees of similarity between the words, while a negative or low value around 0 will indicate a random lack of correlation. If we for instance select the affective term sad and calculate the cosine between the angle of its vector representation and any other word in the text corpus, we can determine which other term vectors are semantically close, and in decreasing order list to what degree they share aspects reflecting the meaning of that word:

1.00 sad
0.74 grief
0.73 sorrow
0.63 mourn
0.62 sigh
0.58 weep
0.53 tear
0.51 griev
0.50 piti
0.49 ala

Looking at these nearest neighbors it would seem that instead of interpreting sad isolated as a single vector made from the various documents in which it appears, we might rather think of the meaning of that word as a semantic neighborhood of vectors. In this part of our LSA semantic space these nearest neighbors form a network of nodes, where each word add different aspects to the meaning depending on the strength of their associative links to sad. So if we imagine text comprehension as a process that combines the words which shape a sentence with the associations they trigger we can model this as a bottom-up spreading activation process. In this network the strength of links between

nodes will be defined by their weights and consequently the connections among all nodes can be mapped out in a connectivity matrix. Being exposed to an incoming word the stimulus will spread from the node generated in episodic memory, to its semantically nearest neighbors stored in long term working memory. How many of these connections grow sufficiently strong for the nodes to be integrated in long term working memory, determines whether our comprehension is reduced to an assembly line where separate words are merely glued together based on the incoming text alone. Or it will instead provide a blueprint for reconstructing a situation model, resembling an animated pin-ball machine where the associations triggered by the words bounce off walls forming an intricate maze of memories. And once reality kicks in, in terms of the constraints posed by the limited capacity of our working memory, what nodes will remain activated could be understood as proportional to the LSA cosine similarity of vectors, triggered by the words being parsed and their nearest neighbors already residing in our memories [38].

4   Results

In the next subsections we outline the structure of the LSA patterns in regards to the distribution of emotional components, compare the LSA analyses of lyrics against their corresponding *last.fm* tag clouds, and finally explore correlations between LSA patterns and the underlying musical structure of the songs.

4.1   Distribution of LSA components

Projecting the lyrics of twenty-four songs selected from the weekly top track charts at *last.fm*, we compute the correlation between lyrics and the previously selected twelve affective tags used as markers in the LSA space, while discarding cosine values below a threshold of 0.09. Whereas the user-defined tags at *last.fm* describe a song as a whole, we aim to model the shifting contours of tension and release which evoke emotions, and therefore project each of the individual lines of the lyrics into the semantic space. Analyzing individual lines on a timescale of seconds also reflects the cognitive temporal constraints applied by our brains in general when we bind successive events into perceptual units [39]. We perceive words 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. We thus assume that lines of lyrics consisting of a few words each correspond to one of these high level perceptual units.

The outputs are matrixes consisting of columns of LSA values triggered by each line in the lyrics in response to the twelve emotional tags making up the rows. Similar to an emotional space, the columns of the matrix reflect a vertical span from positive to negative valence. The upper rows in the columns correspond to active positive emotions like happy and funny followed by more passive aspects like mellow and cool towards the center of the columns. Further down the values in the columns correspond to active negative aspects like angry

while the bottom rows in the matrix reflect passive negative emotions such as melancholic and sad. When the individual lines of lyrics are projected against the selected *last.fm* tags, it results in twelve dimensional vector values signifying affective components that are activated simultaneously rather than discrete emotions. Initially adding up the LSA values in the matrices along each row, we can plot the activation of each emotional component over the entire lyrics. Analyzing a small sample of twenty-four songs, the summed up values of LSA correlation between the individual and the *last.fm* affective terms appears to divide the lyrics into roughly three groups:

Balanced distribution of emotions where the lyrics simultaneously trigger affective components from the outer extremes of both happy and sad. Combined with more passive positive aspects like soft it results in types of patterns as found in the songs: "21 Things i want in a lover" (Alanis Morissette), "Bleeding love" (Leona Lewis), "The Scientist" (Coldplay), "Mad world" (Gary Jules), "Nothing else matters" (Metallica), "Starlight" (Muse) and "Come away with me" (Norah Jones). (Fig.1)
- or alternatively the patterns juxtapose active positive and negative elements of happy and versus angry against each other, with relatively less contribution from passive positive aspects like soft, as in the songs: "Everybody hurts" (R.E.M), "Iris" (Goo Goo Dolls), "Wonderwall" (Oasis), "Time to pretend" "Rehab" (Amy Winehouse). (Fig.2)

Centered distribution of emotional components emphasizing passive positive aspects like soft mellow or cool combined with passive negative emotions close to sad, with relatively less significant contributions from active positive affective components such as happy as in the songs: "Now at last" (Feist), "My immortal" (Evanescence), "Creep" (Radiohead) and "Colorblind" (Counting Crows). (Fig.3)

Uniform distribution of emotional components activated across the entire affective spectrum of valence and arousal as in the songs:"Always where i need to be" (The Kooks), "San Quentin" (Johnny Cash), "Clocks" (Coldplay), "What I've done" (Linkin Park), "Falling slowly" (Glenn Hansard), "Stairway to heaven" (Led Zeppelin), "Smells like teen spirit" (Nirvana) and "Such great heights" (The Postal Service)(Fig.4).

## 4.2   LSA emotions versus last.fm tags

The tag clouds at *last.fm* describe a song as a whole, so in order to assess to what degree the retrieved LSA correlation values of lyrics and affective terms approximate the user-defined tags, we use the accumulated LSA values summed up over the entire lyrics as outlined in previous section. To facilitate a comparison between lyrics and the tag clouds, which my only contain a few of the selected *last.fm* affective terms used in the LSA analysis, we subsequently group the LSA values of closely related tags into an emotional space consisting of four
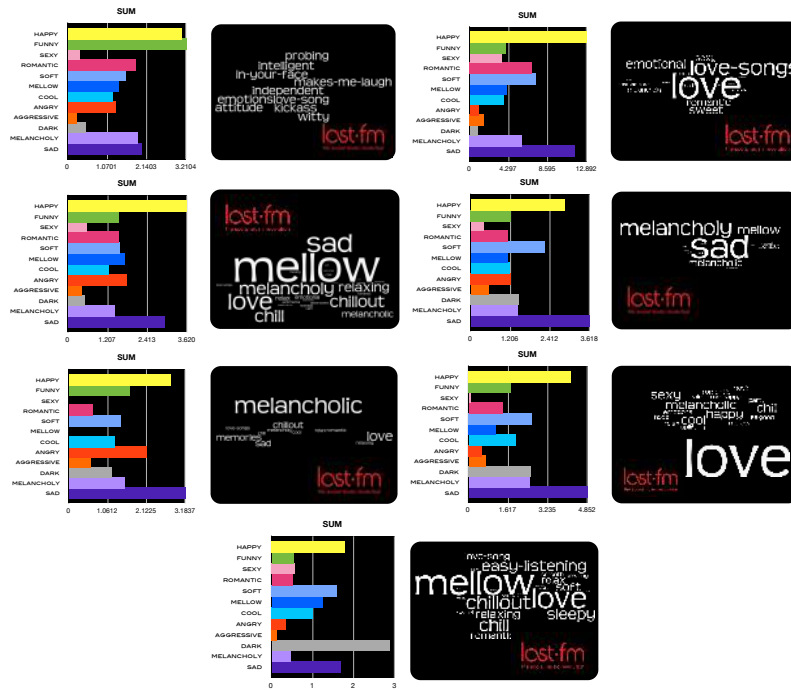
Fig. 1. Accumulated LSA values of emotional components triggered by the lyrics and their corresponding *last.fm* tag clouds - from top left and across: "21 Things i want in a lover" (Alanis Morissette), "Bleeding love" (Leona Lewis), "The Scientist" (Coldplay), "Mad world" (Gary Jules), "Nothing else matters" (Metallica), "Starlight" (Muse) and "Come away with me" (Norah Jones).
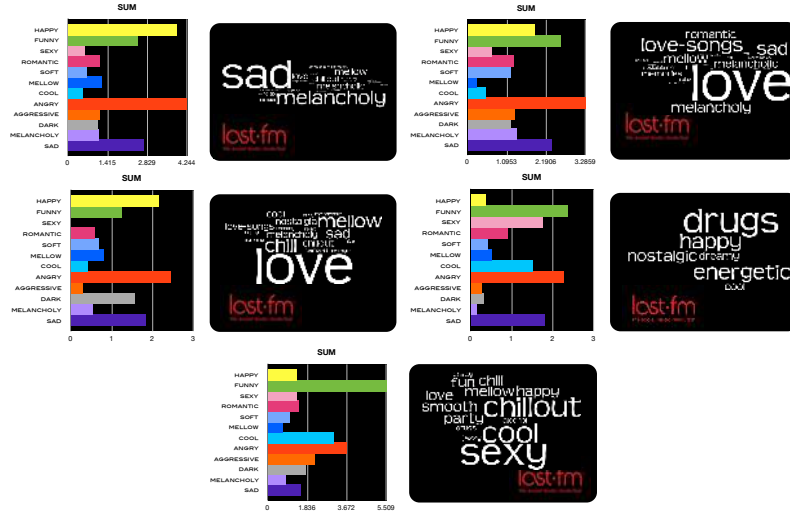
Fig. 2. Accumulated LSA values of emotional components triggered by the lyrics and their corresponding *last.fm* tag clouds - from top left and across: "Everybody hurts" (R.E.M), "Iris" (Goo Goo Dolls), "Wonderwall" (Oasis), "Time to pretend" "Rehab" (Amy Winehouse).
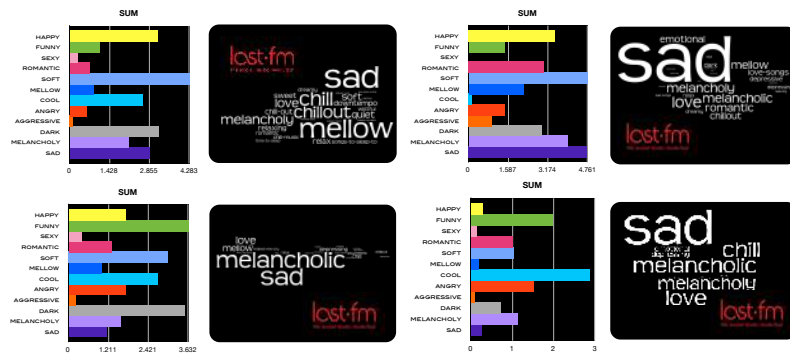
Fig. 3. Accumulated LSA values of emotional components triggered by the lyrics and their corresponding *last.fm* tag clouds - from top left and across: "Now at last" (Feist), "My immortal" (Evanescence), "Creep" (Radiohead) and "Colorblind" (Counting Crows)

Fig. 4. Accumulated LSA values of emotional components triggered by the lyrics and their corresponding *last.fm* tag clouds - from top left and across: "Always where i need to be" (The Kooks), "San Quentin" (Johnny Cash), "Clocks" (Coldplay), "What I've done" (Linkin Park), "Falling slowly" (Glenn Hansard), "Stairway to heaven" (Led Zeppelin), "Smells like teen spirit" (Nirvana) and "Such great heights" (The Postal Service).

segments of emotions framed by the dimensions of valence and arousal:

active positive - happy, funny, sexy, romantic
passive positive - soft, mellow, cool
active negative - angry, aggressive
passive negative - dark, melancholy, sad

Within the tag-clouds a number of more idiosyncratic expressions like "kickass" or "makes-me-laugh" will similarly have to be mapped onto one of the above four affective groups, in this case defined as active positive. Terms referring to complex emotions like "love" has similarly been assigned to this segment based on user-rated valence and arousal values [32] . To simplify the comparison against the tags, the emotional segment with the highest accumulated LSA values has been highlighted for each of the songs below:

"21 Things i want in a lover" (Alanis Morissette)
last.fm tags include: "attitude, in-your-face, kickass, makes-me-laugh" (Fig.1)
LSA values summed in groups: **active positive: 8,4** passive positive: 4,2 active negative: 1,5 passive negative: 4,4

"Bleeding love" (Leona Lewis)
last.fm tags include: " love, romantic, sweet, sexy, melancholy, sad" (Fig.1)
LSA values summed in groups: **active positive: 25,3** passive positive: 15,2 active negative: 2,6 passive negative: 18,2

"The Scientist" (Coldplay)
last.fm tags include: "mellow, sad, love, chill, melancholy" (Fig.1)
LSA: **active positive: 7,2** passive positive: 4,5 active negative: 2,2 passive negative: 4,9

"Mad world" (Gary Jules)
last.fm tags include: "sad, melancholy, mellow" (Fig.1)
LSA values summed in groups: active positive: 5.6 passive positive: 4,6 active negative: 1,8 **passive negative: 6,5**

"Nothing else matters" (Metallica)
last.fm tags include: "melancholic, love, chillout, sad" (Fig.1)
LSA values summed in groups: active positive: 5,2 passive positive 2,7: active negative: 2,7 **passive negative: 5,9**

"Starlight" (Muse)
last.fm tags include: "love, cool, chill, happy, melancholic, sexy" (Fig.1)
LSA values summed in groups: active positive: 7,4 passive positive: 5,6 active negative: 1,2 **passive negative: 10,0**

"Come away with me" (Norah Jones)
last.fm tags include: "mellow, love, chillout, sleepy" (Fig.1)
LSA values summed in groups: active positive: 3,4 passive positive: 3,8 active negative: 0,5 **passive negative: 5,1**

"Everybody hurts" (R.E.M)
last.fm tags include: "sad, melancholy, mellow, chillout" (Fig.2)
LSA values summed in groups: **active positive: 8,1** passive positive: 2,5 active negative: 5,4 passive negative: 4,9)

"Iris" (Goo Goo Dolls)
last.fm tags include: "love, sad, mellow, romantic, melancholy" (Fig.2)
LSA values summed in groups: **active positive: 6,5** passive positive: 1,9 active negative: 4,6 passive negative: 4,9

"Wonderwall" (Oasis)
last.fm tags include: "love, chill, mellow, sad" (Fig.2)
LSA values summed in groups: **active positive: 4,1** passive positive: 1,9 active negative: 2,7 passive negative: 3,9)

"Time to pretend" (MGMT)
last.fm tags include: "drugs, happy, energetic, nostalgic" (Fig.2)
LSA values summed in groups: **active positive: 5,4** passive positive: 2,4 active negative: 2,6 passive negative: 2,3

"Rehab" (Amy Winehouse)
last.fm tags include: "sexy, cool, chillout, fun, happy, party, smooth" (Fig.2)
LSA values summed in groups: **active positive: 9,6** passive positive: 4,8 active negative: 5,9 passive negative: 4,2

"Now at last" (Feist)
last.fm tags include: "sad, mellow, chill" (Fig.3)
LSA values summed in groups: active positive: 5,2 passive positive: 7,7 active negative: 0,7 **passive negative: 8,2**

"My immortal" (Evanescence)
last.fm tags include: "sad, love, melancholy" (Fig.3)
LSA values summed in groups: active positive: 8,0 passive positive: 7,1 active negative: 2,4 **passive negative: 11,6**

"Creep" (Radiohead)
last.fm tags include: "melancholic, sad, love, mellow" (Fig.3)
LSA values summed in groups: **active positive: 7,0** passive positive: 6,7 active negative: 1,9 passive negative: 6,3

"Colorblind" (Counting Crows)
last.fm tags include: "sad, chill, melancholic, love" (Fig.3)
LSA values summed in groups: active positive: 3,5 **passive positive: 4,2** active negative: 1,6 passive negative: 2,1

"Always where i need to be" (The Kooks)
last.fm tags include: "makes-me-happy, sounds-like-summer, party, cool" (Fig.4)
LSA values summed in groups: active positive: 2,6 passive positive: 1,6 active negative: 1,3 **passive negative: 2,9**

"San Quentin" (Johnny Cash)
last.fm tags include: "prison, angry, black, cynical" (Fig.4)
LSA values summed in groups: **active positive: 3,6** passive positive: 1,1 active negative: 1,9 passive negative: 1,7

"Clocks" (Coldplay)
last.fm tags include: "chill, mellow, cool" (Fig.4)
LSA values summed in groups: **active positive: 5,5** passive positive: 3,1 active negative: 2,5 passive negative: 2,6

"What I've done" (Linkin Park)
last.fm tags include: "love, energetic, intense, memories" (Fig.4)
LSA values summed in groups: active positive: 4,4 passive positive: 3,4 active negative: 1,1 **passive negative: 5,8**

"Falling slowly" (Glenn Hansard)
last.fm tags include: "mellow, love, feel-good, sad" (Fig.4)
LSA values summed in groups: **active positive: 6,6** passive positive: 4,3 active negative: 3,5 passive negative: 3,3

"Stairway to heaven" (Led Zeppelin)
last.fm tags include: "melancholic, cool, mellow, sad" (Fig.4)
LSA values summed in groups: active positive: 7,3 passive positive: 6,5 active negative: 4,7 **passive negative: 7,7**

"Smells like teen spirit" (Nirvana)
last.fm tags include: "love, cool, energetic, kick-ass, melancholic" (Fig.4)
LSA values summed in groups: **active positive: 8,4** passive positive: 4,6 active negative: 2,5 passive negative: 7,7

"Such great heights" (The Postal Service)
last.fm tags include: "love, chill, mellow, happy" (Fig.4)
LSA values summed in groups: **active positive: 8,2** passive positive: 7,6 active negative: 3,5 passive negative: 5,7

To summarize, in order to assess to what degree the retrieved LSA correlation values of lyrics and affective terms approximate the user-defined *last.fm* tags describing the songs, we compared the maximum accumulated LSA values against the twenty-four tag clouds. Mapping the highest accumulated LSA values onto one of the four generalized groups of emotions we retrieved the following results:

Thirteen lyrics were correctly identified to represent emotions related to active positive aspects of valence: "21 Things i want in a lover" (Alanis Morissette), "Bleeding love" (Leona Lewis), "Iris" (Goo Goo Dolls), "Wonderwall" (Oasis), "Time to pretend" (MGMT), "Rehab" (Amy Winehouse), "Smells like teen spirit" (Nirvana), "Such great heights" (The Postal Service), - or passive negative aspects of valence as in the songs: "Mad world" (Gary Jules), "Nothing else matters" (Metallica), "Now at last" (Feist), "My immortal" (Evanescence), "Stairway to heaven" (Led Zeppelin).

Five songs were wrongly identified to represent active positive aspects of valence instead of passive positive: "The Scientist" (Coldplay), "Clocks" (Coldplay), "Falling slowly" (Glenn Hansard) - or passive negative: "Everybody hurts" (R.E.M), "Creep" (Radiohead).

Three songs were wrongly identified to represent passive negative aspects of valence instead of active positive: "Starlight" (Muse), "Always where i need to be" (The Kooks), "What I've done" (Linkin Park)

One song was wrongly identified as active positive aspects of valence instead of active negative: "San Quentin" (Johnny Cash). One song was wrongly identified as representing passive positive aspects of valence instead of passive negative: "Colorblind" (Counting Crows). One song was wrongly identified as passive negative instead of passive positive: "Come away with me" (Norah Jones)

4.3   LSA emotions mapped over time

While the accumulated LSA values facilitate characterizing the patterns in terms of their distribution of emotional components, and simplify a comparison against the corresponding *last.fm* tag clouds, mapping eight examples of the retrieved LSA matrices over time, allows us to explore to what extent the triggered emotions reflect the underlying musical structure of the songs. The grayscale plots define the range of emotions that are triggered by each line in the lyrics over time. A third dimension is indicated by the amount of saturation, where black signifies higher cosine correlation between the affective terms and the lyrics. Separating the grayscale plots into sections corresponding to the structure of the song, makes it possible to compare the patterns of emotions against the formal divisions of the song. Adding up the LSA values for each line in the lyrics, provides an alternative view of the accumulated emotional peaks and valleys, plotted over time in colors based on the values from the grayscale matrices.

Taking as an example the Metallica song "Nothing else matters", LSA peak values of happy and sad are triggered simultaneously by the lyrics "Couldn't be much more from the heart" in line 2, 17 and 37 marking the beginning of the 1'st and 2'nd section, as well as the final lines of the coda. The persistent pattern of juxtaposed angry, cool and funny components is caused by lines of "Nothing else matters", interspersed between clusters of dark, melancholy and sad elements resulting from the "Never cared for .." sections in the lyrics. The three happy-sad peaks partition the overall structure of the song. In between these peaks, the texture consists of the pointed angry-cool-funny structures, connected to the declining slopes made out of the contrasting clusters. (Fig.5)
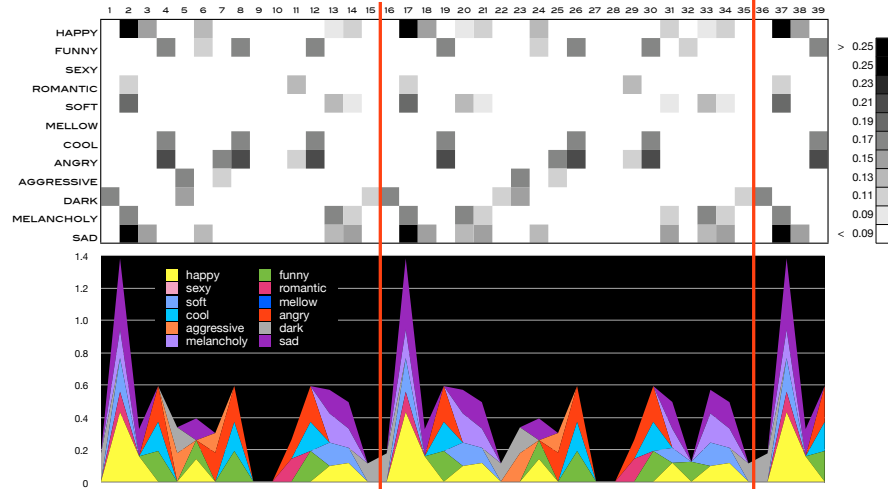
Fig. 5. Nothing else matters (Metallica): LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom).

The ABAB musical structure of the Coldplay song "The scientist" is marked by the vertical LSA columns triggering almost every emotion at line 3 and 24 followed by scattered activation of mellow and soft emotions., reflecting the lines "You don't know how lovely you are" and "Tell me you love me", in the beginning and middle of the two A sections respectively. In contrast the subsequent two B sections, commencing at line 13 and 30, are characterized by a sustained juxtaposition of happy and sad aspects largely devoid of more central soft and romantic components. Two affective peaks dominate the beginning of the first A section and the middle of the second A section. Following the peaks, the remaining parts of the A sections are characterized by scattered aspects of more central emotions, that lead into the shorter B sections with simultaneous balanced activation of happy and sad components (Fig.6)
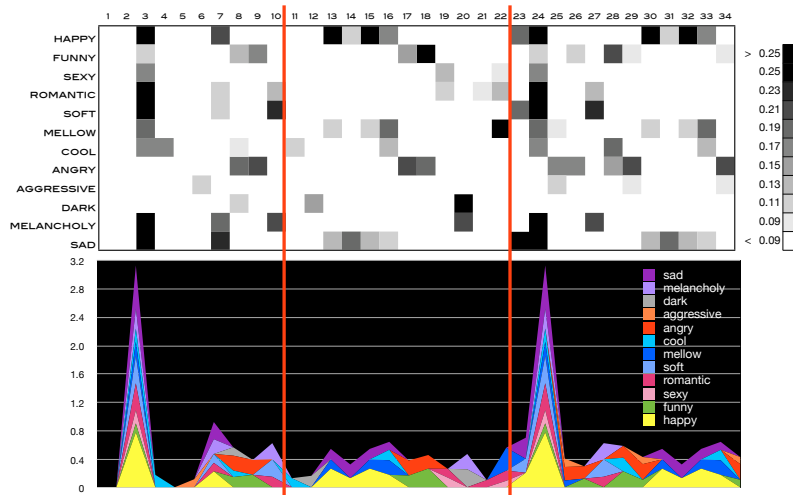
Fig. 6. The Scientist (Coldplay): LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom).

The layout of the song "Iris" by Goo Goo Dolls is marked by the LSA saturated clusters in the intro culminating in line 6 "And all i can breathe is your life" and a less strong activation in the second verse at line 15 "When everything feels like the movies", which are in both cases generated by simultaneous happy and sad elements combined with romantic and soft aspects. This is contrasted with the pattern concluding both the first and second verse, triggered by funny and alternating angry and aggressive elements in the lines "Cause I don't think that they'd understand" and "When everything's made to be broken", similarly sustained throughout the refrain "I just want you to know who I am" that brings the song to an end. Two connected peaks are generated from simultaneous happy and sad elements combined with romantic and soft aspects, culminating in the

beginning of the first verse followed by a slightly lower peak in the beginning of the second verse. In between a balanced texture is activated by funny and alternating angry and aggressive elements (Fig.7).
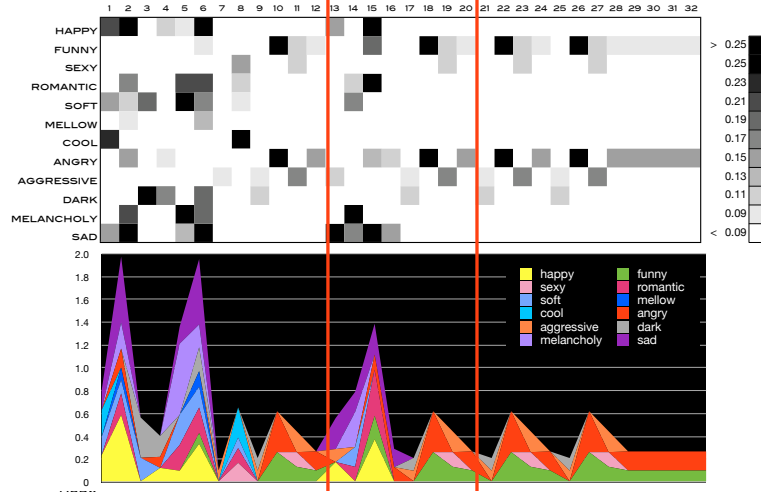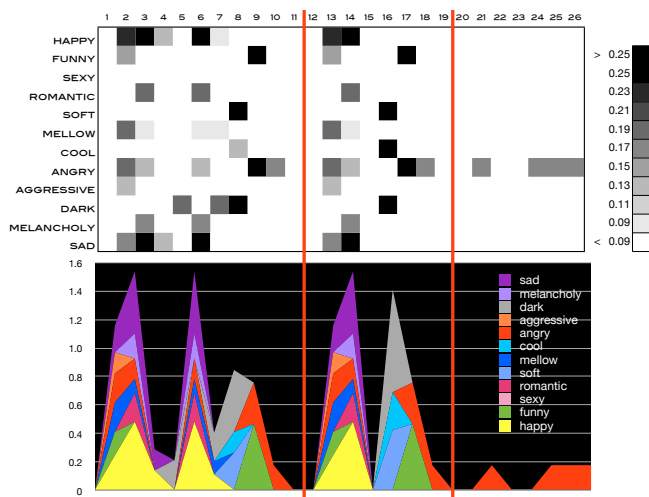

Fig. 7. Iris (Goo Goo Dolls): LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom).

A simultaneous activation of happy and sad elements in line 3, 6 and 14 characterize the first two sections of the Oasis song 'Wonderwall' corresponding to the lyrics "I don't believe that anybody feels the way I do about you now". It is followed in the second section by soft cool and dark components triggered by line 16 "And all the lights that light the way are blinding" contrasted with elements of anger , here caused by line 9 and 17 "There are many things that I would like to say to you but I don't know how", which are sustained through the end of the sparsely activated last section (Fig.8)

The very beginning of Feist's "Now at last" is marked by the cluster generated by the lyrics "For I've lost the last last love" in line 3 eliciting a wide range of emotions followed by a smaller peak emphasizing the soft aspects triggered by line 7 "To the joys before me", that make up the two following sections with the added aspect of cool caused by line 15 and 25 "When the spring is cold" (Fig.9)

Apart from the isolated spike in line 29 "Whatever makes you happy" marking the beginning of the third section in Radiohead's "Creep" in a centered distribution of emotional components, the song is throughout reflecting the pointed cool soft dark textured peaks caused by the text "But I'm a creep" in line 9, 21 and 33 of the lyrics. (Fig. 10)

The lyrics of Led Zeppelin's "Stairway to heaven" trigger components uniformly distributed across a range of emotions resulting in a coherent texture

Fig. 8. Wonderwall (Oasis):LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom).
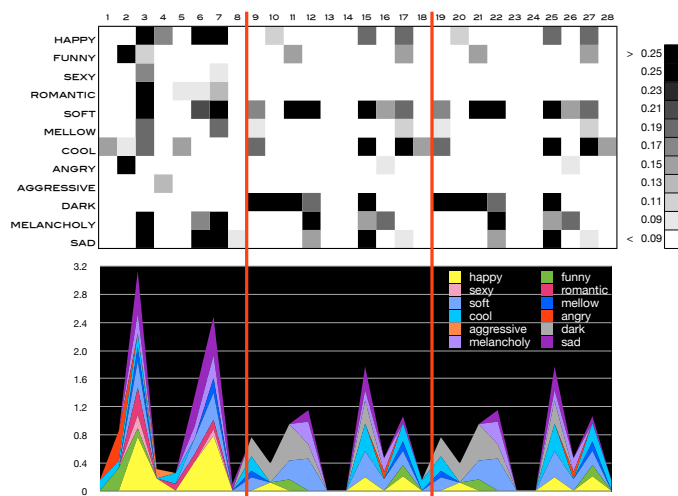
Fig. 9. Now at last (Feist): LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom).
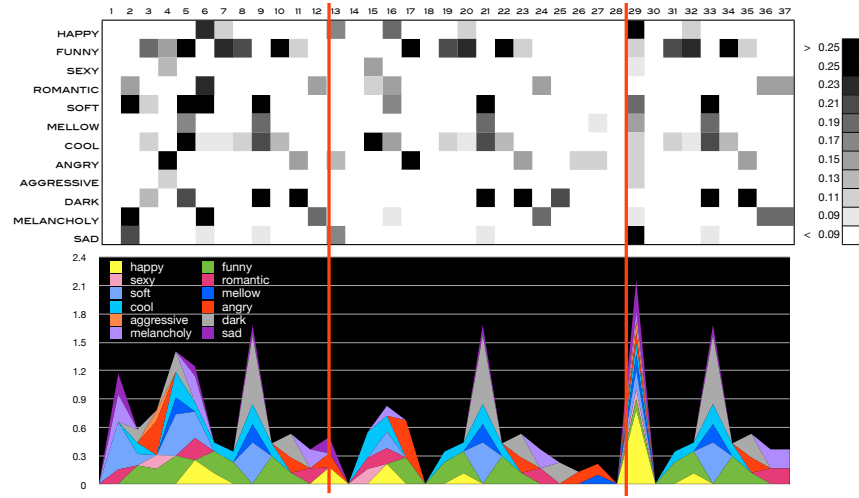
Fig. 10. Creep (Radiohead): LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom).

even though only a few lines are repeated, as in the text "makes me wonder" in lines 10-11, 16-17 and 26. A number of sad peaks establish the melancholy atmosphere at the end of the first section line 9 "Sometimes all of our thoughts are misgiven" as well as the beginnings of the second section in line12 "There's a feeling I get when i look to the west" as well as the section in line 24 "Yes there are two paths you can go by but in the long run". (Fig. 11)

The generally light atmosphere of "Such great heights" by The Postal Service, is generated from the very beginning of the song culminating in line 4 "And when we kiss they are perfectly aligned", that leads into the second and third sections marked by the repeated patterns of romantic happy and funny components triggered by the text "But everything looks perfect from far away" in line 19, 31, 35 and 39 of the song.(Fig.12)

5   Conclusion

While we have here only analyzed twenty-four songs our first results indicate that it is possible to describe the emotional context of songs by applying LSA latent semantic analysis to extract latent semantics from the lyrics using a selection of *last.fm* affective tags. Summing up the retrieved emotional values in the LSA analyses along each row, the matrices can be divided into roughly three groups characterized by a balanced, centered or uniform distribution of emotional components.

Assessing to what degree the summed up LSA correlation values of lyrics and affective terms approximate the user-defined *last.fm* tags within a choice of
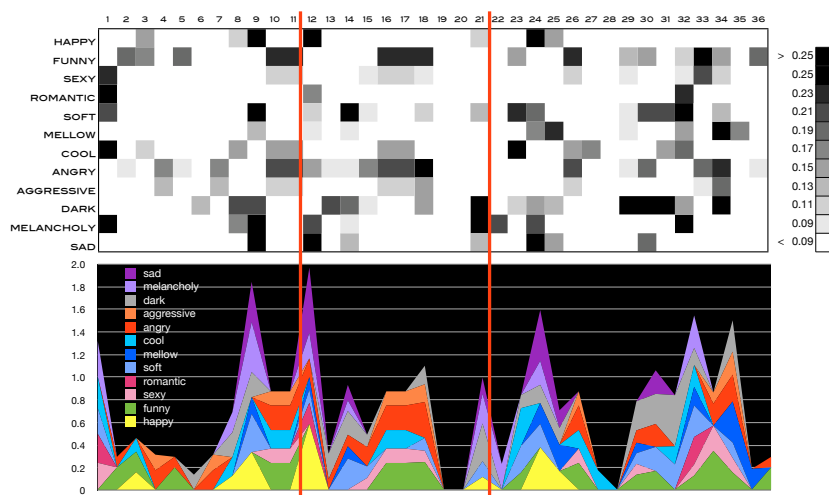
Fig. 11. Stairway to heaven (Led Zeppelin): LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom).
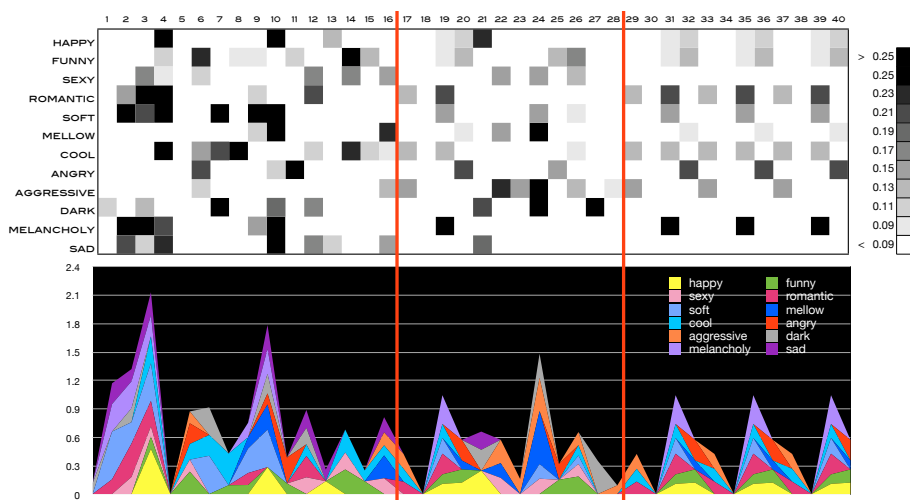
Fig. 12. Such great heights (The Postal Service): LSA patterns of emotions triggered by the lyrics, separated into sections corresponding to the musical structure of the song (top). Accumulated LSA values for each line of the lyrics (bottom)

four emotional segments, thirteen of the lyrics were correctly identified. Three of the eleven lyrics that were wrongly identified due to LSA failing to distinguish more subtle differences between happy versus mellow, might be improved by adjusting the weights in the additive model which is initially simply summing up al values equally. Similarly other lyrics mistakenly identified as sad instead of happy or the other way around, could be due to a few peaks influencing the balance between positive and negative valence too strongly.

Considering how the emotional peaks and valleys of the lyrics, seem to align with the underlying musical structure of the songs, we speculate that extracting emotional components over time might provide a simplified approach to model how we perceive media. We hypothesize that these emotional components reflect compositional forms perceived as patterns of tension and release, which form the dramatic undercurrents of an unfolding structure. As exemplified in the plots of song lyrics each matrix column corresponds to a time window of a few seconds, which is also the approximate length of the high level units from which we mentally construct our perception of continuity within time. Interpreted in that context we suggest that the LSA analysis of textual components within a similar size of time window is able to capture a high level representation of the shifting emotions triggered when we listen to a song. Or from a cognitive perspective the dimensionality reduction enforced by LSA might be interpreted as a simplified model of how mental concepts are constrained by the strengths of links connecting nodes in our working memory. It seems that even if we turn off the sound, the emotional context as well as the overall formal structural elements can still be extracted from media based on latent semantics.

References

1. L. B. Meyer: *Meaning in music and information theory*, Journal of Aesthetics and Art Criticism, Vol.15, pp. 412-424, 1957
2. D. Temperlie: *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, pp. 33-72, 2006
5. C.L. Krumhansl, *Music: A link between cognition and emotion*, Current Directions in Psychological Science pp.35-55, 2002
6. K. Jongwha and E. André, *Emotion recognition based on physiological changes in music listening*, IEEE Transactions on pattern anlysis and machine intelligence:12 pp.2067-2083, 2008
7. 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
8. 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
9. 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

10. C. Calln, V. Tsytsarev, T. Hanakawa, A. Calln, M. Katsuhara, H. Fukuyama and R. Turner: *Song and speech: Brain regions involved with perception and covert production*, NeuroImage, Vol. 31(3) pp. 1327-1342, 2006

11. S. Koelsch and W.A. Siebel: *Towards a neural basis of music perception*, Trends in Cognitive Sciences, Vol.9(12) pp. 578-584, 2005

12. 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

13. L.R. Slevc, 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

14. 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

15. V. Gallese: *Embodied simulation: From neurons to phenomenal experience*, Phenomenology and the Cognitive Sciences, Vol.4 pp. 23-48, 2005

16. V. Gallese and G. Lakoff: *The brain's concepts: the role of the sensory motor system in conceptual knowledge*, Cognitive Neuropsychology, Vol.22 pp. 455-479, 2005

17. I. Molnar-Szakacs and K. Overie: *Music and mirror neurons: from motion to 'e' motion*, Social cognitive and affective neuroscience, Vol.1(33) pp. 235-241, 2006

18. L.K. Hansen and L. Feng: *Cogito componentiter - ergo sum*, In: J. Rocha et al.(eds): Independent Component Analysis and Blind Signal Separation, LNCS 3886, pp.446-453, 2006

19. A.J.Bell and T.T.Sejnowski: *The independent components of natural scenes are edge filters*, Vision Research (37:23), pp.3327-3338, 1997

20. L.Feng and L.K.Hansen: *On phonemes as cognitive components of speech*, Proceedings of IAPR workshop on cognitive information processing, 2008

21. G.W. Furnas, S. Deerwester, S.T. Dumais, T.K. Landauer, R. Harshman, L.A. Streeter and K.E. Lochbaum: *Information retrieval using a singular value decomposition model of latent semantic structure*, In: 11th annual international SIGIR conference, pp.465-480, 1988

22. S. Derwester, S.T. Dumais, G.W. Furnas, W. George, T. Landauer and R. Harshman: *Indexing by latent semantic analysis*, Journal of the American Society for Information Science (41:6), pp.391-407, 1990

23. 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, pp. 211-240, 1997

24. J. Cudeiro and A.M. Sillito: *Looking back: corticothalamic feedback and early visual processing*, Trends in Neurosciences (29:6), pp.298-306, 2006

25. A.M. Sillito, J. Cudeiro and H.E. Jones: *Always returning: feedback and sensory processing in visual cortex and thalamus*, Trends in Neurosciences (29:6), pp.307-316, 2006

26. J.H.R Maunsell and S. Treue: *Feature-based attention in visual cortex*, Trends in Neurosciences (29:6), pp.317-322, 2006

27. J. Storbeck and G.L. Clore: *On the interdependence of cognition and emotion*, Cognition & Emotion (21:6) pp.1212-1237, 2007

28. S. Duncan and L.F. Barret: *Affect is a form of cognition: a neurobiological analysis*, Cognition & Emotion (21:6) pp.1184-1211, 2007

29. L.F. Barret: *Solving the emotion paradox: categorization and the experience of emotion*, Personality and social psychology review (10:1) pp.20-46, 2006

30. C. E. Osgood, G.J. Suci and P.H. Tannenbaum: *The measurement of meaning*, University of Illinois Press, 1957
31. J.A. Russel: *A circumplex model of affect*, Journal of personality and social psychology (39:6), pp. 1161-1178, 1980
32. M. M. Bradley and P.J. Lang: *Affective norms for English words (ANEW), Stimuli, instruction manual and affective ratings*, The Center for Research in Psychophysiology, University of Florida, 1999
33. E. Schubert: *Update of the Hevner adjective checklist*, Perceptual and Motor Skills (96), pp.1117-1122, 2003
34. M. J. Power: *The structure of emotion: an empirical comparison of six models*, Cognition & Emotion (20:5) pp.694-713, 2006
35. G. P. Strauss and D.N.Allen: *Emotional intensity and categorisation ratings for emotional nonemotional words*, Cognition & Emotion (22:1), pp. 114-133, 2008
36. E. Bigand, S. Vieillard, F. Madurell, J.Marozeau and A.Dacquet: *Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts*, Cognition & Emotion (19:8), 2005
37. D.I. Martin and M.W. Berry: *Mathematical foundations behind latent semantic analysis*, In: Handbook of latent semantic analysis, Erlbaum, 2007
38. W. Kintsch: *Comprehension - a paradigm for cognition*, Cambridge University Press, 1998
39. E. Pöppel: *A hierarchical model of temporal perception*, Trends in Cognitive Sciences, (1:2) pp. 56-61, 1997