Sparse but emotional decomposition of lyrics

Michael Kai Petersen, Morten Mørup, and Lars Kai Hansen

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

Abstract. Both low-level semantics of song texts and our emotional responses can be encoded in words. In order to model how we might perceive the emotional context of songs, we propose a simplified cognitive approach to bottom-up define term vector distances between lyrics and affective adjectives, which top-down constrain the latent semantics according to the psychological dimensions of valence and arousal. Projecting the lyrics and adjectives as vectors into a semantic space using LSA latent semantic analysis, their cosine similarities can be mapped as emotions over time. Subsequently we apply a three-way Tucker tensor decomposition to the derived LSA matrices, combined with a hierarchical Bayesian automatic relevance determination to find similarities across a selection of songs, and as a result identify two time series dramatic curvatures and three mixtures of affective components, which might function as emotional building blocks for generating the structure in lyrics.

1 Introduction

Though one might think of media as an audiovisual stream of consciousness, we frequently encode frames of video or waves of sound into strings of text. Language allows us to both share the internal representations of what we perceive as mental concepts, as well as categorizing them as distinct states in the continuous ebb and flow of emotions [1]. Cognitively speaking our feelings can be thought of as labels that we consciously assign to the emotional responses triggered by the sensory inputs we are constantly exposed to [2]. That is, the brain applies an analysis-by-synthesis approach, which infers structure from bottom-up processing of statistical regularities, that are continuously compared against stored patterns of top-down labeled gestalts [3]. Whether we actually perceive something, coordinate our movements to grasp an object or only imagine that we do so by mentally simulating the action, the same neural populations are involved. The result is that circuits interconnecting the various areas in the brain will cause neurons to fire in synchrony regardless of their underlying modality [4]. Language builds on the very same sensory-motor mechanisms in the brain, and consequently fuses the various modalities of sound, sight and somatosensory sensations together in a semantic structure of action concepts. It has recently been shown that reading a word like 'smile' triggers the same motor resonances in our brains as a visual representation of the corresponding facial features, which means that also verbal emotional expressions are embodied [5]. As both low-level semantics of song texts and our affective responses can be encoded in words, we propose that it might be feasible to cognitively model the distances 'bottom-up' between lyrics and affective adjectives represented as vectors in a high-dimensional space based on LSA latent semantic analysis [6]. And 'top-down' constrain the latent semantics according to neurophysiological dimensions which allows us to model emotions within a continuos plane. Meaning, that the emotional adjectives used in the proposed LSA implementation function not only as markers for measuring the distances between affective labels and the lyrics of a song, but also represent points in a semantic plane framed by the psychological axes of valence and arousal, that cognitively appear to have actual neural correlates pertaining to two distinct networks in the brain [7].

2 Related work

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 [8]. Specifically related to songs, fMRI 'functional magnetic resonance imaging' experiments show that neural processes involved in perception and action when covertly humming the melody or rehearsing the song text activate overlapping areas in the brain. This indicates that core elements of lyrical music appear to be treated in a fashion similar to those of language [9], which is in turn supported by the electrophysical evidence of language and music competing for the same neural resources when processing syntax and semantics [10]. Looking specifically into the functional architecture of memory, it appears that both storage and representation of verbal and tonal information rely on the same neural networks. That is, processing and encoding of phonemes as well as pitch are largely based on the same sensorimotor mechanisms. And as a result our short term working memory, complemented by the phonological loop which can store words for a few seconds as well as rehearsal mechanisms for subvocally repeating syllables [11], appear not only to be used in speech but similarly involved when maintaining a sequence of tones in memory [12]. Further studies of the interaction between phonology and melody indicate that "vowels sing whereas consonants speak" [13], meaning that vowels and melodic intervals may have similar functionalities related to the generative structure of syntax in language and music, whereas consonants seem rather related to lexical distinction crucial for learning words. Likewise the affective adjectives used as semantic markers in the proposed model appear to have neural correlates, as intonation cues processed in the auditory cortices associated with the feelings of anger, sadness, relief or joy, have been identified as distinct patterns of activation in recent fMRI brain imaging studies [14]. Within the music information retrieval community a number of studies have similarly focused on retrieving the patterns in music underlying our emotional responses, by mapping audio features into a space framed by valence and arousal. Either modeling emotion continuously as a time varying function of spectral shape, pitch or rhythmic textures [15]. Or instead subtracting low level audio features from segments of songs, that are grouped into clusters associated with the basic emotions of happiness, calm, anger and sadness [16]. Similarly the foundation for how lyrics shape melody has recently been explored in a number of studies, detailing how lyrics influence rhythmical patterns [17], as well as finding correlations between syllabic stress and melodic peaks [18]. While the approach proposed in the present paper based on defining latent semantics based on term vector distances within a structured representation of affective adjectives may not previously have been implemented to model emotional context of media, a similar approach has recently been applied in fMRI neuroimaging studies [19]. Here instead the distances measured were between nouns and verbs, used to predict what patterns of voxels in the brain would be activated in response to words belonging to different semantic categories. These brain imaging studies were like in our approach inferred from the statistics of word co-occurrences in a large natural language corpus, although in the fMRI study using a Google data set containing one trillion n-gram counts of word sequences. While we in this paper apply LSA based on tens of thousands of samples of both fiction and non-fiction to find the underlying structure in the lyrics.

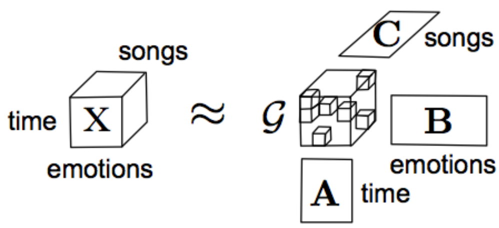
3 Emotional tensors

For our implementation of a simplified cognitive approach we project a selection of lyrics into a semantic space using LSA, in order to define the cosine similarity between individual lines of lyrics and each of the affective adjectives:

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

These adjectives are not only frequently used as emotional tags for describing music in social networks like *last.fm* [20], but also represent distinct points in four quadrants dividing a continuous plane framed by the psychological dimensions of valence and arousal, as has earlier been demonstrated in psychological experiments defining user rated affective values of words [21]. Hence the dimension valence describes how pleasant something is perceived, along an axis going from positive to negative associated with words like 'happy' or 'sad', while arousal captures the amount of involvement ranging from passive states like 'mellow' and 'sad' to active aspects of excitation as reflected in 'angry' or 'funny'. When projecting the lyrics into the semantic space, the outcome of the LSA are second-order matrices, with rows consisting of the lines making up the lyrics, and columns containing the cosine similarity values for the twelve affective adjectives. In essence the consecutive rows of lines making up the lyrics provide a temporal dimension, whereas the saturation represented by the cells in the matrix define the loadings of the twelve dimensional emotional vectors. The underlying matrix of terms and documents providing the foundation for LSA is based on a text corpus mixture of both fiction and non-fiction, consisting of 22829 terms found in 67380 contexts. These documents in turn consist of 500 word segments, made from 22072 literature and poetry samples of the Harvard Classics, 15340 segments of Wikipedia music articles, and 29968 general news items from the Reuters Corpus gathered over the period 1996-1997. Although there is no straight forward mathematical foundation for determining the optimal number of principal components to be retained when reducing the original term document matrix, other studies have demonstrated that LSA can attain an understanding of word meanings comparable to the results achieved by non-native english speakers taking a TOEFL 'test of english as a foreign language' [6]. An approach previously applied has therefore been to submit the LSA text corpus to a TOEFL synonymy test in order to compare the term vector distances between similar words in the vector space, while varying the number of dimensions until an optimal percentage of correct answers are returned. For our LSA configuration the best fit corresponding to 71,25% correctly identified synonyms is realized when reducing the matrix to 125 factors, providing a result above the 64.5 % average achieved by non-native college applicants, as well as the 64.5 % correct answers previously reported for LSA spaces [6].

Having previously established a measure of ground-truth, by comparing LSA emotional topics derived from 24798 lyrics against user-defined tags describing the corresponding songs at *last.fm* and finding correlations between 'happy-sad' emotions, as well as aspects defining 'soft, cool' and 'dark' textures [22-23], we in this article go beyond the initial second-order analysis. And subsequently apply a three-way factor analysis using a Tucker model [24], to derive similarities in emotional patterns over time in 50.274 lyrics selected from *LyricWiki* . To enable a comparison of the lyrics independent of their length the LSA matrices are resampled to a fixed length of 32 time points, corresponding to the average number of lines in the analyzed songs. Decomposing the selection of second-order LSA derived emotions over time matrices into a three dimensional tensor:



$$x_{ijk} \approx \mathcal{G} \times \mathbf{A} \times \mathbf{B} \times \mathbf{C} = \sum_{lmn}^{LMN} g_{lmn} a_{il} b_{jm} c_{kn} \qquad (1)$$

makes it possible to assess the strengths by which the vector loadings of the time and emotions matrices interact over a large number of songs. Here the $\mathcal{G}$ core array is positive and defines the strength by which the J_N columns or vector loadings of the $\mathbf{A}^{Time \times L}(\geq 0)$, $\mathbf{B}^{Emotions \times M}$ (unconstrained), and $\mathbf{C}^{Songs \times N}(\geq 0)$ matrices interact. Or in other words, the model will relate all potential linear interactions between vectors representing the three modes. And the variables L, M and N correspond to the number of components or columns in the factor matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, which could in turn be interpreted as principal components in each of the three modes [25].

To assure that the model retains only the minimal number of components necessary for representing the data, the Tucker tensor decomposition is fitted using a sparse regression algorithm to prune excess components. A hierarchical Bayesian automatic relevance determination ARD approach is applied to determine the amount of sparsity imposed on the core array as well as the loadings. Enabling that the relevance of different features can be determined based on hyperparameters, which define a range of variation for the underlying parameters. And to provide a sparse representation, these parameters are modeled as the width of exponential and Laplace prior distributions, which are non-negativity constrained and unconstrained respectively assigned to the loadings and core - for full details on the method please see [26]. For this implementation of the algorithm a signal to noise ratio of 0 dB was chosen, meaning that we do not assume a higher ratio of signal than noise to be present when applying the sparse Bayesian algorithm. In order to optimize the likelihood function, the combined Tucker and Bayesian ARD approach was applied ten times to the LSA matrices, and the decomposition achieving the highest logarithmic probability value based on 1000 iterations was selected to provide the best representation of the data.

4 Results

Starting out with ten components in each of the three modes of the Tucker tensor decomposition, representing emotions over time across fifty thousand songs, the number of excess components are gradually pruned by the Bayesian ARD algorithm until only significant interactions between emotions over time across thousands of songs remain. These are represented in the core arrays, related to two time-series components highlighted as dramatic curvatures in blue and green respectively. The two core arrays, where black signifies zero and white the highest correlation value, define in turn the interaction between the respective time-series components and nine retrieved mixtures of emotional topics that appear within five clusters of songs (Fig.**??**.) However, even though nine emotional topic groups and five clusters of songs are found to be interacting with the time-series components, the sparsity implied on the core array allows us to identify the components which are most significantly correlated across the three modes.

In the first core array, associated with the time-series curvature highlighted in blue (Fig.**??**), the outlined saturated squares indicate that primarily the second ('soft') and sixth ('happy-sad') emotional topics are the ones most strongly cor-

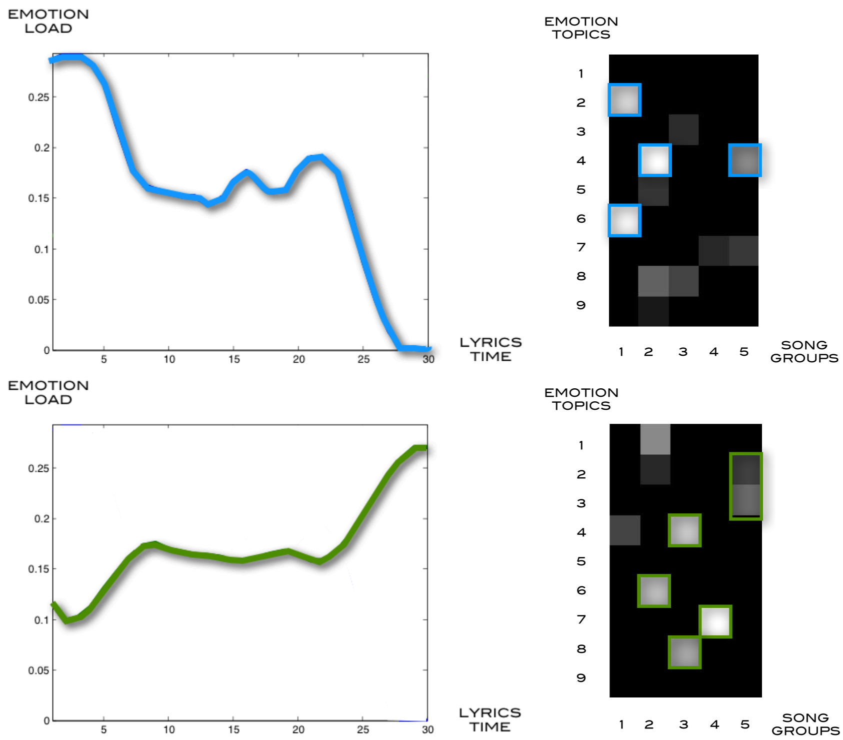


Fig. 1. Time-series components representing the variability of emotional load in lyrics derived from a Tucker 3-way tensor decomposition of emotions over time matrices compared across 50.000 songs, using a sparse hierarchical Bayesian ARD automatic relevance determination approach to prune excess components and identify the most significant interactions among components from the three modes. The most saturated light gradients in the core arrays indicate which of the vertically plotted emotional topics are most strongly correlated with the dramatic curvatures for time-series components 1 and 2, highlighted in blue and green respectively, and interacting with the five clusters of song groups mapped horizontally.

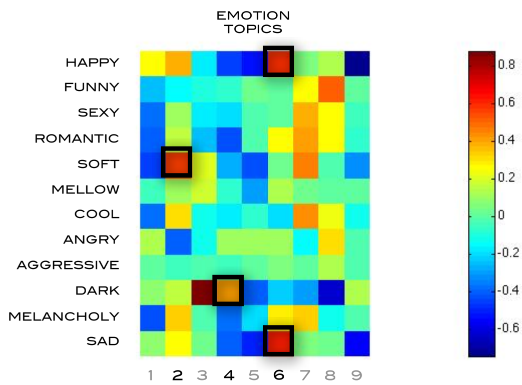


Fig. 2. First core array emotion topics constituted by mixtures of affective adjectives, where the groups 2 ('soft'), 4 ('dark') and 6 ('happy-sad') are most significantly correlated with the first time series component corresponding to the dramatic curvature highlighted in blue (Fig.**??**).

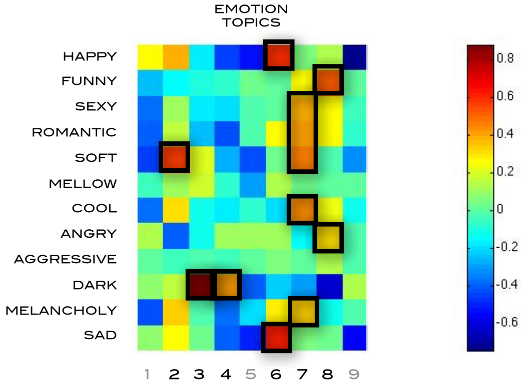


Fig. 3. Second core array emotion topics constituted by the mixtures of affective adjectives, where the groups 2 ('soft'), 3 and 4 ('dark'), 6 ('happy-sad'), 7 ('sexy, romantic, soft, cool, melancholy') and 8 ('funny-angry') are most significantly correlated with the second time series component corresponding to the dramatic curvature highlighted in green (Fig.**??**).

related with the first time-series component and most significantly found to be represented within the first group of songs. Whereas in the second group of songs the fourth ('dark') and to a lesser degree the not outlined eighth ('funny-angry') group of emotional topics are interacting with the first time-series component (Fig.**??**).

In the second core array, associated with the green time-series curvature (Fig.**??**), the seventh emotional topic ('sexy, romantic, soft, cool, melancholy') comes out as the mixture of affective adjectives most strongly interacting with the second time-series component. The other less strongly saturated emotional topics are a combination of the second ('soft'), third and fourth ('dark') plus the eighth ('funny-angry') components. The eighth topic also appears in the first core array while here it interacts with the second time-series component represented in the third group of songs. And similarly the previously encountered sixth ('happy-sad') juxtaposition of emotions are here found to be represented within the second cluster of songs (Fig.**??**).

Looking into samples of tracks maximally reflecting the load of emotional topics defining the first group of songs (Fig.**??**), here taking the Bon Jovi song "Not fade away" as an example, the sixth emotional component can clearly be made out (Fig.**??**). Correlated with the first time-series component this emotional topic is characterized by 'happy' activations that appear synchronized with 'sad' aspects sustained throughout the song in row 1 and 12 of the matrix, and culminating with an activation of 'soft-dark' textures at the very end. Also in The Mission's rendering of "Love" (Fig.**??**), the lyrics strongly trigger the simultaneous juxtaposition of 'happy-sad' contrasts, coupled with additional 'romantic-soft' components while being devoid of any 'angry' aspects.

Taking two of the top tracks most representative of the second group of songs as examples, Nirvana's "Love Buzz" and the Therapy? song "Stay Happy"

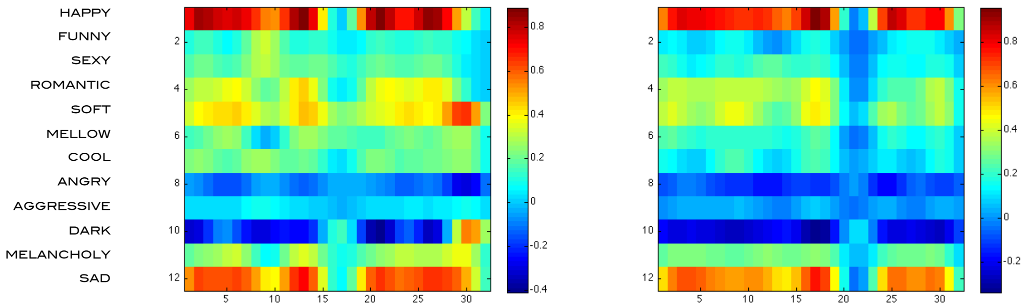


Fig. 4. Samples from the first group of songs, exemplified by the lyrics of the Bon Jovi's "Not fade away" (left), and The Mission's rendering of' "Love" (right).

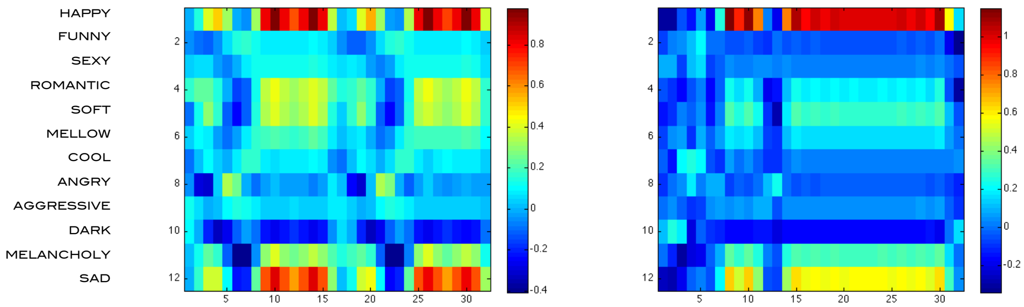


Fig. 5. Samples from the second group of songs, exemplified by the lyrics of Nirvana's "Love Buzz" (left), and the Therapy? song "Stay Happy" (right).

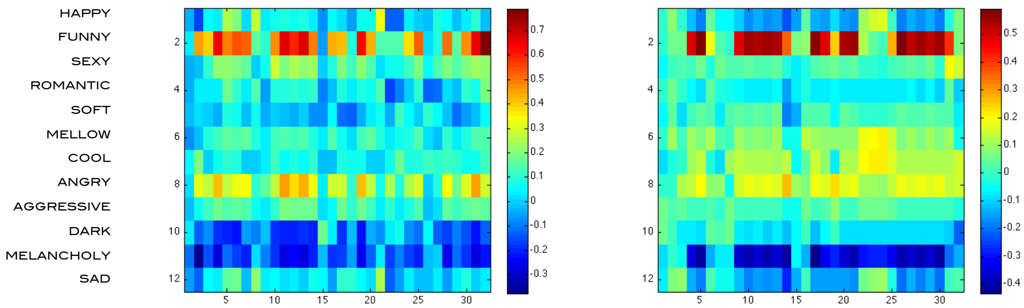


Fig. 6. Samples from the third group of songs, exemplified by the lyrics in The Sex Pistols' rendering of "No Fun" (left), and the Michael Jackson song "Jam" (right).

(Fig.**??**) both again reflect the simultaneous coupling of 'happy-sad' contrasts as found in the sixth emotion topic when correlated with the second time-series component. Although here appearing less sustained throughout the lyrics of "Love Buzz" which might also reflect the less significant saturation of the sixth emotion topic in the second core array (Fig.**??**). While in "Stay Happy" the 'happy-sad' contrasts are now strongly biased towards the top and most other emotions appear deactivated in the matrix

In the third group of songs, exemplified by two more problematic lyrics: the The Sex Pistols' version of "No Fun" and the Michael Jackson song "Jam" (Fig.**??**), the overall affective weighting of the matrices are strongly influenced by the eighth emotional topic capturing 'funny-angry' aspects in the lyrics, correlated with the second time-series component (Fig.**??**). However in the case of "No Fun" where the lyrics heavily trigger 'funny' despite the song lamenting the prospect of being alone, the lack of sequential syntactic order here highlights the challenge of retrieving the underlying meaning using a bag of words approach only. Whereas the frustration in the lyrics might come across more easily based on an activation of 'angry' aspects. Similarly the Michael Jackson song "Jam" could hardly be considered particularly 'funny' despite the strong triggering of these emotions based on the lyrics. Whereas the complementary 'cool' emotions triggered in the matrix seem more apt at capturing the atmosphere in the song. While also here the energetic aspects of the lyrics are much better channeled out through the 'angry' aspects.

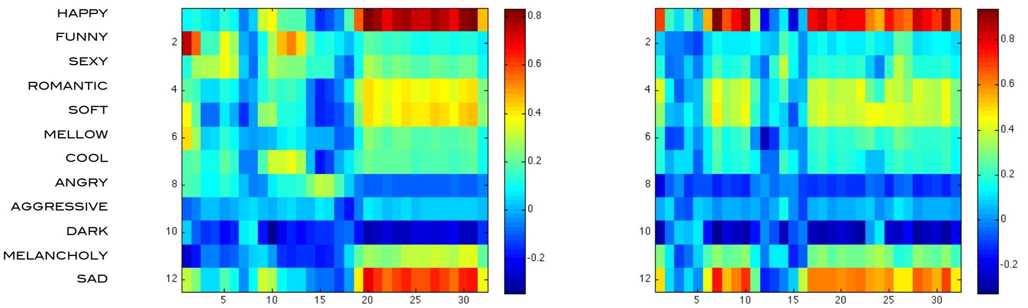


Fig. 7. Samples from the fourth group of songs, exemplified by the lyrics in Bo Didley's rendering of "Diddley Daddy" (left), and the Lou Reed song "The Blue Mask Women" (right).

.

The fourth cluster of songs correlated with the second time-series components is mainly representing the characteristics of the seventh emotional topic (Fig.**??**), which captures more 'romantic-soft' aspects visible in the saturated rows 5 and 6 in the matrix plots of Bo Didley's lyrics of "Diddley Daddy" and the Lou Reed song "The Blue Mask Women" (Fig.**??**). Even though the seventh emotional topic stands out clearly in the core array associated with the second time-series component (Fig.**??**), the lyrics in this cluster of songs also strongly trigger the'happy-sad' contrasts, making this emotional topic appear more like a complementary texture than a principal emotional component.

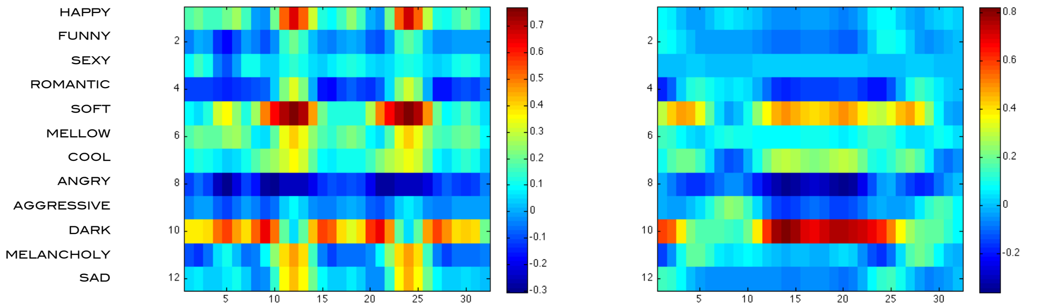


Fig. 8. Samples from the fifth group of songs, exemplified by the lyrics of The Doors' “End of the night” (left), and the Yeah Yeah Yeahs' song “Hello Tomorrow” (right).

The characteristics of the fifth cluster of songs appears to concatenate the interactions among the second, third and fourth emotional topics related to both of the time-series components. Or in other words, even though these emotional topics seem less saturated in the two core arrays (Fig.**??**), the combined effect appears distinctly in the fifth group of songs, as exemplified by the lyrics of The Doors' “End of the night” and the Yeah Yeah Yeahs' song “Hello Tomorrow” (Fig.**??**). As in both of these examples, the lyrics most representative of this cluster of songs are in general characterized by a strong activation of ‘soft-dark' aspects, while any other emotions appear subdued in the matrices. Meaning, that the emotional topics representing ‘soft-dark' aspects seem to capture textures that can metaphorically be interpreted as feelings. And together with the previously identified ‘happy-sad' and ‘funny-angry' mixtures of affective adjectives these textural elements appear to constitute principal components defining the structure of the lyrics.

Overall the significance of the affective mixtures that contrasts ‘happy' against ‘sad', as represented by the sixth emotional topic correlated with both the first and second time-series components, might reflect earlier findings that roughly half of the variance in emotional words can be captured by a ‘happy-sad' principal component [27]. It might also be interpreted in the way that the positive elements of ‘happy' which in terms of valence are contrasted against the negative feelings of ‘sad', here mainly constitute passive aspects of arousal. Whereas the more energetic aspects of arousal seem rather to be represented by the emotional mixture of ‘funny-angry' corresponding to the eighth emotional topic correlated with the second time-series component. So, these two pairs of contrasts might be interpreted as representative of the two principal dimensions framing a psychological space. And together the emotional topics could be understood to capture the dimensions of valence and arousal, and thus function as affective building blocks for generating the time-series dramatic curvatures in songs. Also the identified 'soft', 'cool' or 'dark' textures identified in the lyrics appear salient as they might not only be understood as abstract concepts, but in a larger context reflect somatosensori aspects of touch or timbre which are metaphorically mapped onto feelings, as has been documented based on cognitive semantics [28] as well as extensive neuroscientific evidence [2]. Taking into consideration that the interaction between lyrics and melody might reflect more general aspects of multi-modal processing, and knowing that language builds on the very same sensory-motor mechanisms in the brain, one might interpret the LSA patterns of emotions over time as capturing the various modalities of sound and somatosensory inputs together in a semantic conceptual structure [4].

5 Conclusion

Having previously established a measure of ground-truth based on correlations between LSA derived emotions and user-defined tags describing the corresponding songs at *last.fm*, we aimed to probe whether we could identify emotional topics and time-series curvatures that in general constitute affective components in lyrics. Applying a Tucker three-dimensional tensor decomposition combined with a hierarchical Bayesian automatic relevance determination approach we compared the LSA derived emotions over time matrices across 50.274 lyrics. Our findings indicate that two time-series components or dramatic curvatures characterize the shape of the lyrics in our sample. And that three mixtures of affective adjectives appear to capture the emotional topics, based on 'happy-sad' and 'funny-angry' contrasts combined with 'soft-dark' textures. The first time-series component is a descending curve reflecting a decreasing emotional load over the duration of the song, associated with a combination of primarily 'happy-sad' contrasts of valence. While the second time-series component is a curve that builds up to form an ascending line culminating at the very end of the song based on either mainly 'happy-sad' contrasts or combined with more aroused and energetic aspects reflected in a 'funny-angry' topic. The pairwise contrasting elements of affective components appear to represent statistically significant elements that might function as emotional building blocks for generating sequential patterns in songs. And these pairs of contrasts seem in turn correlated with the identified two time-series components, that could represent underlying dramatic structures generally found in lyrics. Meaning that the contrasts of 'happy-sad' or 'funny-angry' as well as the identified 'soft', 'cool' or 'dark' textures might be interpreted as not only principal emotional components that top-down constrain the latent semantics according to the neurophysiological dimensions of valence and arousal. But in a wider context that the emotional topics trigger our conscious responses to traces in memories capturing pleasure and pain of past experiences, that we as feelings conceptualize as bodily states when unwinding the phrases of words making up the lyrics.

References

1. Barrett, L.F. : *Solving the emotion paradox: categorization and the experience of emotion*, Personality and social psychology review,10:1, p.20-46, (2006)
2. Damasio, A. : *Feelings of emotion and the self*, Annals of the New York Academy of Sciences, 1001, p.253-261, (2003)
3. Borenstein, E., Ullman, S. : *Combined top-down / bottom-up segmentation*, IEEE Transactions on pattern analysis and machine intelligence, 30:12, p.2109-2125, (2008)
4. Gallese, V., Lakoff, G. : *The brain's concepts: the role of the sensory motor system in conceptual knowledge*, Cognitive Neuropsychology, 22, p.455-479, (2005)
5. Foroni, F., Semin, G.R. : *Language that puts you in touch with your bodily feelings*, Psychological Science, 20:8, p.974-980, (2009)
6. Landauer, T.K, S., Dumais, S.T.: *A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge*, Psychological Review, 104:2, p.211-240, (1997)
7. Posner, J., Russell, J.A., Gerber, A. et al. : *The neurophysiological bases of emotion: an fMRI study of the affective circumplex using emotion-denoting words*, Human Brain Mapping, 50, p.883-895, (2009)
8. Patel, A.D.: *Music, language and the brain*, Oxford University Press (2009)
9. Callan, D.E., Tsytsarev, V., Hanakawa, T. et al: *Song and speech: Brain regions involved with perception and covert production*, NeuroImage, 31:3, p. 1327-1342, (2006)
10. Koelsch,S. *Neural substrates of processing syntax and semantics in music*, Current Opinion in Neurobiology, 15, p. 207-212, (2005)
11. Baddeley,A.: *Working memory: looking back and looking forward*, Nature Neuroscience, 4, p. 829-839, (2003)
12. Koelsch,S., Schulze, K., Sammler,D., Fritz, T., Müller, K.: *Functional architecture of verbal and tonal working memory: an fMRI study*, Human Brain Mapping, 30:3, pp. 859-873, (2009)
13. Kolinsky,R., Lidji, P., Peretz, I., Besson, M. Morais, J:. *Processing interactions between phonology and melody: Vowels sing but consonants speak*, Cognition, 112, pp. 1–20, (2009)
14. Ethofer, T., Van De Ville, D. and Scherer, K. Vuilleumier, P. : *Decoding of emotional information in voice-sensitive cortices*, Current Biology, Volume 19:12, (2009)
15. Korhonen, M.D. et al.: *Modeling emotional content of music using system identification*, IEEE Transactions on systems, man and cybernetics, 36:3 , p. 588599, (2006)
16. Yang, Y. et al.: *Music emotion classification: a fuzzy approach*, Proceedings of 14th annual ACM international conference on multimedia, p. 8184, (2006)
17. Nichols E.: *Lyric-based rhythm suggestion*, Proceedings of the International Computer Music Conference (ICMC), (2009)
18. Nichols E., Morris, D., Basu, S., Christopher, S. : *Relationships between lyrics and melody in popular music*, Proceedings of the 10'th International Conference on Music Information Retrieval (ISMIR), (2009)
19. Mitchel T.M. Shikareva, S.V., Carlson, A. et al.: *Predicting human brain activity associated with the meanings of nouns*, Science, Volume 320:30, p.1191-1195, (2008)
20. Hu X., Bay, M., Downie, S..: *Creating a simplified music mood classification groundtruth set*, Proceedings of the 8'th Internaional Conference on Music Information Retrieval (ISMIR), p.309-310, (2007)

21. Bradley M.M., Lang, P.J.: *Affective norms for English words (ANEW): stimuli, instruction manual and affective ratings*, The Center of Research in Psychophysiology, University of Florida, (1999)
22. Petersen, M.K., Hansen, L.K., Butkus, A: *Semantic contours in tracks based on emotional tags*, In: Kronland-Martinet, R. et al. (Eds.): Genesis of meaning of sound and music, LNCS 5493, pp. 45–66, Springer (2009)
23. Petersen, M.K., Hansen, L.K.: *Modeling lyrics as emotional semantics*, (in press) (2009)
24. Tucker, L.R.: *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31:3, pp. 279-311, 1966
25. Kolda, T.C., Bader, B.W.: *Tensor decompositions and applications*, SIAM review, june, 2008
26. Mørup, M., Hansen, L.K.: *Automatic relevance determination for multi-way models*, Journal of Chemometrics,DOI 10.1002/cem.1223, 2009
27. Russell, J.A.: *A circumplex model of affect*, Journal of Personality and Social Psychology, 39:6, 1980
28. Lakoff, G., Johnson, M.: *Philosophy in the flesh: the embodied mind and its challenge to western thought*, Basic Books, 1999