Cognitive nodes among lines of lyrics

Michael Kai Petersen, Lars Kai Hansen
Technical University of Denmark,
DTU Informatics, Cognitive Systems,, Building 321
DK-2800 Kgs.Lyngby, Denmark
{mkp, lkh}@imm.dtu.dk

ABSTRACT

The challenge of recommending songs when confronted with the vast number of music tracks and video streams available on the internet, might require new types of cognitive interfaces reflecting how we emotionally perceive media content. Both in music and language we rely on syntax for parsing sequences of symbols, which based on hierarchically nested structures allow us to express and share the meaning contained within a sentence or a melodic phrase. These structures become part of our memories when the 'bottom-up' sensory input raises above the background noise of core affect, and 'top-down' trigger distinct emotions reflecting a shift of our attention. As both low-level semantics of lyrics and our emotional responses can be encoded in words, we propose a simplified cognitive approach based on LSA latent semantic analysis. Modeling how we perceive the emotional content of song lyrics, the multiple contexts in which words occur are 'bottom-up' projected as vectors into a semantic space of reduced dimensionality. While patterns of emotional categorization emerge by defining term vector distances to affective adjectives, which as 'top-down' labels constrain the latent semantics, according to a psychological plane framed by the dimensions of valence and arousal.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications—*Text processing*

General Terms

Human factors

1. INTRODUCTION

Advances in neuroimaging 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 [1-4], and core elements of lyrical music in particular appear to be treated in a fashion similar to those of language [5]. Which in turn supports findings that language and music compete for the same neural resources when it comes to processing syntax and semantics [6]. Most likely due to similarities within the two domains in regards to integrating phrases into hierarchical structures, where the neural processing depends on the distances separating grammatical objects within a sentence or spatiotemporal patterns in a pitch space [7-8]. Whether encoding sonorous building blocks of phonemes or pitch they are largely based on the same sensorimotor mechanisms. Cognitively speaking, the interaction between lyrics and melody might reflect more general aspects of multi-modal processing, where mirror neurons in the brain respond to any audiovisual or somatosensory input we perceive or imagine by emulating them on a neural level [9-10]. Coming across action related verbs when reading a word like 'smile' trigger the same motor resonances in our brains as when we are exposed to the corresponding facial features of someone smiling at us, which establishes that even verbal emotional expressions are embodied [11]. Reflecting aspects of neural theory of language [12], we might think of grammar as premotor executing schemas, which are capable of metaphorically mapping small world semantic neighborhoods into action concepts, grounded in the same neural structures that allow us to infer meaning even when retrieving only the sound, a visual imprint or motor pieces of the jigsaw puzzle from memory. What aspects will form part of our memories, depend on whether the 'bottom-up' sensory input will raise above the threshold of background core affect, and 'top-down' trigger distinct emotions reflecting a shift of our attention [13]. Affective adjectives can in that sense be thought of as 'top-down' labels that we consciously assign to what is perceived. Brain imaging studies have shown that emotional words are modulating responses as early 100-140 ms after stimulus [14]. Suggesting that they represent patterns of learned associations that are processed even before a lexical lookup. As both low-level semantics of lyrics and our emotional responses can be encoded in words, we propose a simplified cognitive approach based on LSA latent semantic analysis [15]. Modeling how we perceive the emotional content of song lyrics, the multiple contexts in which words occur are 'bottom-up' represented as vectors in a semantic space of reduced dimensionality. While patterns of emotional categorization emerge by defining term vector distances to affective adjectives frequently used as tags in the *last.fm* social music network [16], which as 'top-down' labels constrain the latent semantics, according to a psychological plane framed by the dimensions of valence and arousal (Fig.1).

ACM Recommender Systems, September 26–30, 2010, Barcelona, Spain.

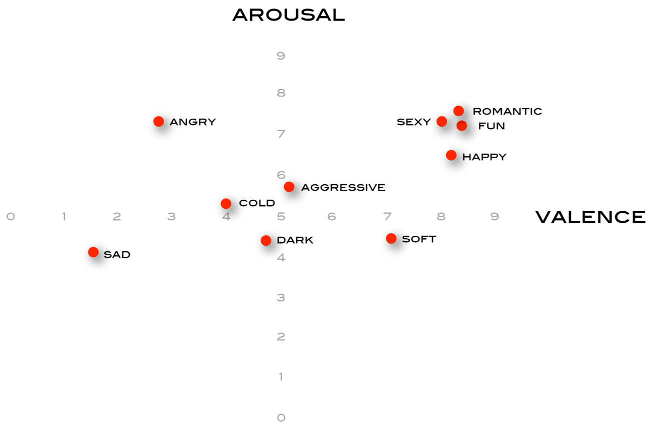

Figure 1: ANEW user rated emotional values of frequently applied *last.fm* tags defining how pleasant something is perceived along an axis horizontally going from negative to positive valence, and arousal vertically capturing the amount of intensity ranging from passive to active.

2. RELATED WORK

Experiments exploring how we perceive emotions have shown that while we often think of affective terms as describing widely different states, these can be represented as related components in a circumplex model framed by the two psychological primitives: valence and arousal [17]. Within this emotional plane the dimension of valence describes how pleasant something is along an axis going from positive to negative contrasting words like 'happy' against 'sad', whereas arousal captures the amount of intensity ranging from passive states like 'sad' to aspects of excitation reflected in terms like 'angry' or 'funny'. This mapping of feelings has actual neural correlates, as brain imaging studies using fMRI to trace which parts become involved when people read emotional words, indicate that activation is divided into two distinct neural networks which are linearly correlated with the values of valence or arousal [18]. And based on the ANEW data set which contains user rated values of valence and arousal for around thousand English words, it is possible to plot how a range of different terms are perceived on a psychological plane framed by these two psychological dimensions [19]. Lately the ANEW corpus has also been translated into Chinese in order to build an affective lexicon which has similarly been used to retrieve the emotions of lyrics and thereby assign a selection of songs into four general groups divided by the psychological axes of valence and arousal [20]. When aiming to retrieve the patterns in music underlying our emotional responses, a number of studies have focused on 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 [21]. 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 [22]. A regression approach to determine which audio features within a 114-dimensional feature space are most relevant for determining emotional context, has identified that arousal is highly influenced by changes in timbral texture and pitch, while it seems much harder to capture valence based on perceived roughness in spectral dissonance or from beat histograms. Auditory features capture a significant part of the variance related to spectral dynamics and energy in music, whereas components related to valence are much harder to pinpoint [23]. Alternatively tag-clouds generated by users in social networks like *last.fm* provide another way to capture the emotional context of songs, as users frequently agree on which emotional tags to attach to the tracks [16]. Even though emotions make up only 5 % of the words in clouds, tags can provide meaningful labels combined with auditory features. Similarity is here treated as a supervised multi-class labeling problem, where a model is trained to learn the joint probabilities of vectors consisting of audio features and semantic annotations [24]. Or alternatively audio features could be mapped onto social network tags, and thereby provide a framework for automatically predicting classes of *last.fm* features describing the genre, style, or mood of tracks not yet tagged by users [25]. However, while modeling audio features as a spectral 'bag of frames' works fine for naturally occurring urban soundscapes like kids playing in a park, this is not the case when considering the sound of somebody playing a solo violin partita by Bach. In music a minority of frames might be statistically insignificant outliers providing the underlying semantic structure [26], as features form patterns related to perceptually significant peaks or local maxima that generate the larger scale semantic structures [27]. Having in an earlier study limited to a small number of songs found correlations between the latent semantics of lyrics and tags describing the corresponding songs [28], we in the present analysis apply LSA to a large selection of lyrics and compare the retrieved emotions against the overall descriptions of the songs constituted by their tag-clouds in the *last.fm* social music network.

3. METHOD

If we think of meaning as being formed by the multiple contexts in which words occur, we can represent them as vectors in a latent semantic space, where the cosine of the angle between them will signify how similar they are to each other. Implemented as a matrix forming a structure of thousands of weakly linked elements, indirect higher order associations begin to emerge as similar features appearing in a large number of contexts are simultaneously squeezed into a reduced number of rows and columns. In order to define the relations between emotional tags and lyrics in LSA, a large general knowledge text corpus is initially constructed which allows for modeling terms as linear combinations of the multiple paragraphs and sentences they occur in. Assembled from tens of thousands of literature, poetry, wikipedia and news excerpts, this underlying corpus could be thought of as resembling our memory, where events combined with lexical knowledge are encoded into mental representations. The foundation here is a large text corpus constructed from 22829 terms found in 67380 contexts. These documents in turn consist of 500 word segments, selected from 22072 literature and poetry samples from the 50 volume Harvard Classics and 20 volume Shelf of fiction, 15340 segments of Wikipedia music articles, and 29968 general news items from the Reuters RVC1 corpora of Reuters stories of general news gathered over the period 1996-1997. Using SVD singular value decomposition [29] the number of parameters can be reduced so only the largest singular values are retained, and as a result the number of eigenvectors in the semantic space are reduced to what would correspond to the principal components containing the highest amount of variance in the matrix. Geometrically speaking, the terms and documents in the condensed matrix can be interpreted as points in a low-dimensional subspace, which enables us to calculate the degree of similarity between matrices based on the dot or inner product of their corresponding vectors. Or if instead interpreted using neural networks as a metaphor, coming across a word like ‘sad’ in a phrase could be thought of as forming a semantic node in our short term episodic memory. In turn triggering neighboring nodes representing words or events that invoke similar connotations. Understood in a cognitive perspective, the strength of the connections initially based on word co-occurrences are gradually transformed into semantic relations, as the links between nodes are being constrained by the limitations of our memory [30]. 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 that determine the latent semantic relations in our LSA space [31]. To decide what constitutes the optimal number of eigenvalues to be retained when reducing the original term-document matrix, an approach previously applied has been to submit the LSA model to a TOEFL ‘test of english as a foreign language’ while varying the number of dimensions until an optimal percentage of correct answers are returned [15]. For our LSA model the best fit corresponding to 71,25 % correctly identified synonyms in the TOEFL test, is realized when reducing the matrix to 125 factors, providing a result above the 64.5 % average achieved by non-native college applicants taking the test, and in line with the 64.5 % and 70.5 % correct answers previously reported for LSA and probabilistic LDA topic models respectively [32].

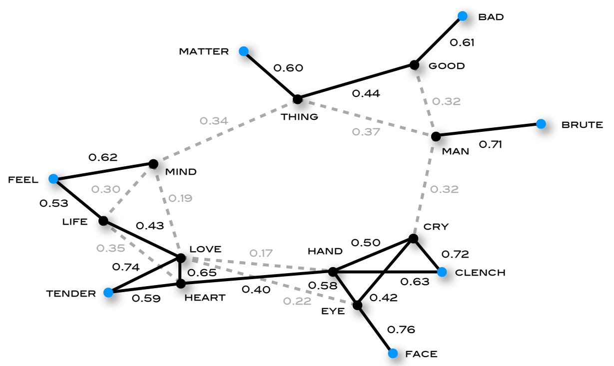

Figure 2: Force-directed graph of 3rd LSA component top words with neighbors. Connectivity and weights of edges are determined by the cosine similarity of the terms, where solid lines indicate correlation values above 0.40 and blue nodes are neighbors.

Looking into what constitutes the eigenvectors when reducing the dimensions in the latent semantic space, the 1’st and 2’nd LSA components capturing the highest variance represent terms mainly used for cataloguing music due to the encyclopedic structure of the Wikipedia articles describing artists and bands. Similarly the inclusion of Reuters news items can be made out from the top words in LSA components 4 and 5 which describe subjects related to economy and sports, while the highest correlated terms captured by LSA components 6 and 7 seem related to politics and business. However the top words in the 3’rd LSA component stand out as reflecting a much more poetic subject: [‘heart’, ‘life’, ‘good’, ‘love’, ‘cry’, ‘mind’, ‘hand’, ‘thing’, ‘eye’, ‘man’ ..]. Taking a closer look at the third LSA component, it can be plotted as an undirected graph in two dimensions (Fig.2) where the nodes are positioned using a Fruchterman-Reingold algorithm, which pulls nodes together or apart like charged particles while the edges are treated as if they were springs until reaching a mechanical equilibrium [33]. The weights of the edges are here determined by the LSA cosine similarity of the words, and the resulting graph of the 3’rd LSA component top words resembles a structure similar to those found in social or biological networks characterized by the number of other vertices a node is connected to [34]. This becomes even more evident when expanding each of the ten nodes in the network with an additional word selected from its top ten nearest neighbors. Three of these neighbors ‘feel’, ‘tender’ and ‘clench’, are shared among the nodes of ‘mind’, ‘life’, ‘love’, ‘heart’, ‘hand’ and ‘cry’ and thus reinforce the connectivity within these respective subgraphs. While the neighbors ‘matter’, ‘bad’ and ‘brute’ add new aspects to consolidate the separate subgraph of the nodes ‘good’, ‘thing’ and ‘man’. Expanding the graph like this emphasizes the hub-like characteristics of ‘hand’ and ‘love’ by increasing their link degrees to five and six, Whereas the overall transitivity of the network, that is the clustering coefficient fraction of closed triangles among the potential triads of two edges sharing a vertex is here 0.37, in line with the range 0.1 - 0.5 frequently found in real world networks.

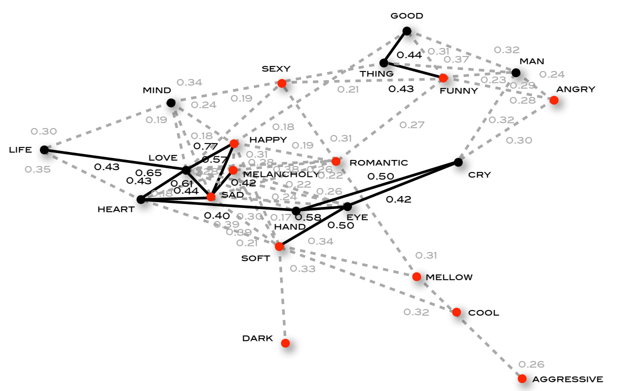

Figure 3: Force-directed graph of 3rd LSA component top words and the twelve affective adjectives. Connectivity and weights of edges are determined by the cosine similarity of the terms, where solid lines indicate correlation values above 0.40 and red nodes are affective adjectives.

Plotting the top words from the 3’rd LSA component together with the twelve affective adjectives, the overall layout of the network is still characterized by subgraphs with a few central nodes of high degree order that are linked to many more sparsely connected nodes (Fig.3). In the graph containing the affective adjectives a dense cluster defines an emotional arch going from 'happy' and 'romantic' over 'melancholy' to 'sad', extending into yet another subgraph that connects the textures of 'soft, dark' and 'cool'. Separated from the more aroused aspects reflected in the nodes 'sexy, funny' and 'angry' at the periphery. Altogether the graph of affective adjectives is strongly integrated into the graph containing the 3'rd LSA component top words. Here the degree centrality of the node 'love' is more than doubled from 5 to 11 due to new connections to affective adjectives, including both the sensory percept 'soft' and the 'happy-sad' contrasts of positive and negative valence. Which are in turn strongly interconnected as the terms 'soft' and 'happy-sad' are linked to 9 and 8 other words respectively, resulting in an overall network transitivity of 0.50. That is, the most central nodes in the network function not only as hubs, but also couple the subgraphs of 3'rd LSA component and the affective adjectives directly. And thereby makes it possible to connect remote nodes within each subgraph by only a few associative steps.

Projecting the lyrics into the LSA semantic space we define the cosine similarity between the individual lines making up each of the lyrics against twelve affective adjectives frequently applied at *last.fm* : 'happy, funny, sexy, romantic, soft, mellow, cool, angry, aggressive, dark, melancholy, sad'. To provide a measure of ground truth we explore the similarity between tag-clouds describing songs and our analysis, by comparing the most frequently applied *last.fm* emotional tags against the LSA values of the corresponding lyrics. First, taking the affective adjectives from 'happy' to 'sad' as a twelve-dimensional vector, this is compared against each of 16505 tags constituting the *last.fm* tag-clouds of the corresponding songs. Calculating the cosine similarity between the twelve affective adjectives and all of the *last.fm* tags describing the songs, makes it possible to extend the comparison to idiosyncratic expressions like 'a bit sad' that can thus similarly be represented as a twelve dimensional vector: [0.39, 0.11, -0.08, -0.15, 0.02, 0.03, -0.06, 0.23, 0.01, -0.07, 0.19, 0.98]. Subsequently the *last.fm* tag vectors corresponding to a tag cloud for a particular song are selected; e.g. the four vectors representing 'fun, sexy, mellow, cool' are compared against the average LSA vector derived from the lyrics and ranged according to their cosine similarity. Our analysis is based on 24798 lyrics selected from LyricWiki by using artist entries retrieved from the Wikipedia "List of Musicians", associated with the genres: alternative rock, blues, brit pop, dream pop, gothic rock, indie rock, indie pop, pop punk, R&B, soul, hard rock, reggae and heavy metal. Additionally applying the restriction that the corresponding *last.fm* tag clouds should contain one or more of the twelve affective adjectives applied in the LSA analysis. The cosine similarity values have been calculated using a LSA software package developed by DTU Informatics which is available for download including the above described HAWIK literature corpus [35].

4. RESULTS

Similar to an emotional space, the columns of the LSA derived lyrics matrices reflect a vertical span from positive to negative valence. The upper rows in the columns correspond to active positive emotions like 'happy' followed by more passive aspects like 'mellow' and 'cool' towards the center of the columns. Further down the values in the columns correspond to aroused negative feelings like 'angry' while the bottom rows in the matrix reflect aspects of low arousal and negative valence such as 'melancholic' and 'sad'.

Taking the song "Rehab" (Fig.4) as an example, the most frequently applied *last.fm* tags are 'mellow, sexy, cool, happy' whereas the top LSA values from the lyrics are 'funny, angry, cool, aggressive'. Comparing *last.fm* tags against LSA emotions based on their cosine similarity they agree most significantly on the adjective 'fun(ny)' 0.87 followed by the terms 'cool' 0.29 'sexy' 0.19 and 'mellow' 0.04. In the latent semantics derived from the lyrics, the upper half of the lyrics matrix is characterized by a horizontal band in row 2 corresponding to almost sustained triggering of 'funny' reflecting positive valence, with less pronounced activations of 'cool' and 'angry' components in row 7 and 8. Whereas the rows of 'happy' at the top and 'sad' emotions at the very bottom remain mostly negatively correlated until activated twice towards the end.

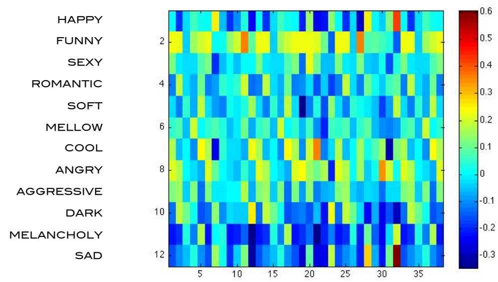

Figure 4: "Rehab" (Amy Winehouse) LSA derived emotions over time reflected in the lyrics: 'funny, angry, cool, aggressive', last.fm tags for the song: 'mellow, sexy, cool, happy', max last.fm/LSA similarity: 'fun(ny)' 0.87 'cool' 0.29 'sexy' 0.19 'mellow' 0.04

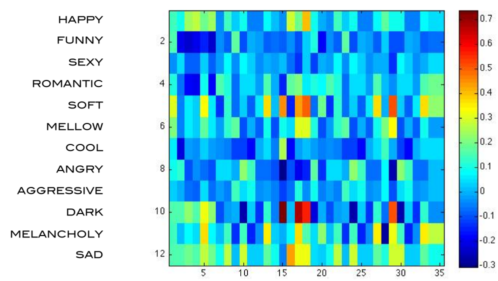

Figure 5: ""My immortal" (Evanescence) LSA derived emotions over time reflected in the lyrics: 'soft, sad, melancholy, happy', last.fm tags for the song: 'mellow, sad, melancholy, romantic', max last.fm/LSA similarity: 'sad' 0.90 'melancholy' 0.78 'soft' 0.75 'romantic' 0.40

Almost the reverse distribution of emotions is triggered by the lyrics of "My immortal" (Fig.5), where the three bottom rows of the matrix are saturated reflecting mostly 'dark' as well as 'melancholy' and 'sad' components. These aspects are coupled with 'soft' components in row 5, while the upper rows of the matrix now remain largely negatively correlated.

Similar to how ‘funny’ and ‘angry’ was triggered simultaneously in the previous example, here in these lyrics ‘soft’ and ‘mellow’ stand out together centered around two peaks. The top LSA values from the lyrics appear as ‘soft, sad, melancholy’ and ‘happy’. Whereas the *last.fm* tags from the song come out as ‘sad’ combined with elements of ‘melancholy, soft’ and ‘romantic’. That is, the distribution of feelings triggered by “My immortal” is shifted towards negative aspects of valence coupled with ‘soft’ elements. This also comes out when comparing *last.fm* tags against LSA emotions, where they agree on the terms ‘sad’ 0.90 ‘melancholy’ 0.78 ‘soft’ 0.75 and ‘romantic’ 0.40.

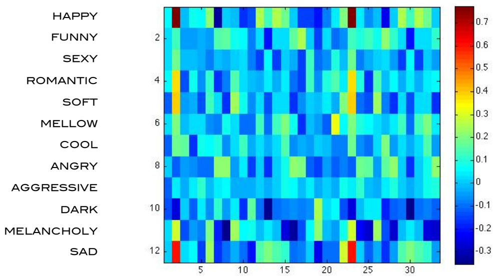

Figure 6: “The Scientist” (Coldplay) LSA derived emotions over time reflected in the lyrics: ‘happy, sad, angry, mellow’, last.fm tags for the song: ‘mellow, cool, sad, melancholy’, last.fm/LSA similarity: ‘sad’ 0.42 ‘romantic’ 0.42 ‘mellow’ 0.38 ‘melancholy’ 0.19

While the previous two songs represent predominantly positive or negative aspects of valence, the lyrics of “The Scientist” trigger feelings from both ends of the scale simultaneously characterized by two ‘happy-sad’ peaks mixed with ‘romantic-soft’ feelings (Fig.6). Positive and negative extremes of valence are activated in synchrony, reflected in the four emotions ‘happy, sad, angry’ and ‘mellow’ being triggered in the LSA analysis. Overall the matrix is sparse and as a consequence the loadings of the two peaks are preserved in the average LSA values for the lyrics. While the LSA analysis comes out as more ‘happy’ than ‘sad’, the *last.fm* tags tip the balance emphasizing more negative aspects in ‘mellow, cool, sad’ and ‘melancholy’. That is, the LSA and *last.fm* vectors agree primarily on the affective adjectives ‘sad, romantic, mellow’ and ‘melancholic’.

In order to determine to what degree these correlations between LSA and *last.fm* entries are statistically significant, a randomized permutation test was run to refute the null hypothesis of the rows and columns being independent. Initially plotting the frequency of the entries on a gradient greyscale (Fig. 7), the left matrix indicates whether the top two entries in the LSA and *last.fm* matrix occur less or more frequently relative to null. Whereas the black and white matrix to the right indicates which of the overly or less frequent combinations of entries in the co-occurrence matrix are beyond random and thus remain significant at $p < 0.05$ for $n = 200$.

Comparing the feelings retrieved from 24798 lyrics based on LSA against the *last.fm* tag-clouds describing the corresponding songs, the correlations in the upper left corner of the matrices that remain statistically significant are the off-diagonal LSA derived emotions of ‘happy’ and ‘funny’ that are confused with the *last.fm* tag ‘sexy’. While in the lower right corner of the matrices the correlation which appears beyond random is the off-diagonal LSA activation of ‘sad’ that is confused with the *last.fm* term ‘melancholy’. Likewise in the right part of the matrix the LSA derived feeling of ‘soft’ when reflected in the lyrics is correlated beyond random with the word ‘dark’ in the *last.fm* tagclouds.

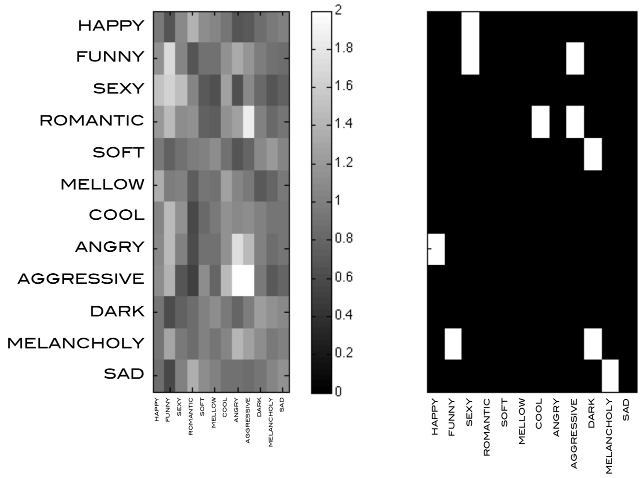

Figure 7: Observed frequencies of rows and columns corresponding to LSA derived emotions and *last.fm* affective tags relative to null (left), whereas the black and white plot (right) shows matrix entries to be considered beyond random and statistically significant for $p < 0.05$ and $n = 200$.

5. DISCUSSION

Comparing the feelings retrieved from twentyfive thousand lyrics based on LSA against the *last.fm* tag-clouds describing the corresponding songs, there is primarily agreement on the contrasts between positive and negative valence captured by the affective adjectives in the upper left and lower right corner of the matrix. Whereas the more subdued emotions along the diagonal show little agreement, due to the disproportionate number of *last.fm* songs which are tagged as ‘mellow’ or ‘cool’ and thereby being confused with the LSA row entries in the matrix. Further analyzing to what degree these correlations are statistically significant (Fig.7), what remains beyond random is the juxtaposition of ‘happy’ and ‘funny’ versus ‘sad’, although skewed off-diagonally due to differences in how these terms are interpreted in the LSA corpus and the *last.fm* social network. And additionally the permutation test indicates a correlation between the LSA feeling ‘soft’ retrieved from the lyrics and the term ‘dark’ when applied as a *last.fm* tag to describe a song. We find this particularly salient as we in a parallel study [36], have applied a 3-way Tucker tensor decomposition to the LSA patterns of emotions triggered by the lyrics over time, to find higher order factors that capture similarities across an even larger data set of fifty thousand songs. Here the results similarly indicate that the most highly correlated components among emotions, time and songs is concentrated in five emotional topics (Fig.8)

These sparse mixtures of affective adjectives can largely be

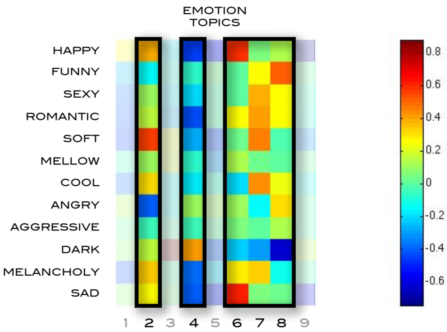

Figure 8: Emotional topics retrieved from a Tucker tensor decomposition of 50.274 lyrics, characterized by mixtures of 2.'soft', 4.'dark' 6.'happy-sad', 7.'soft-cool' and 8.'funny-angry' feelings [36]

thought of as contrasts along dimensions that span the psychological axes of valence and arousal and in the widest sense define a low dimensional structure of the input. Interpreted as principal components the 'happy-sad' mixture could be thought of as representing the maximal contrast potentially biased towards positive or negative valence, which as complementary aspects define the emotional range. Which suggests that the LSA derived matrices of lyrics reflect the overall bias in the songs along the dimensions of valence and arousal. However local maxima in the lyrics could potentially shift the balance from 'happy' to 'sad' based on only a few large peaks. And thus cognitively define the overall lasting impression of how we perceive a song rather than being based on a summation of average intensity. Meaning a mapping of highly granular sequential LSA values for each line in the lyrics against *last.fm* tags that overall describe the corresponding songs, remains far from trivial due to the lack of temporal information within the tag-clouds.

Most of the LSA peaks in the time-series samples analyzed seem to be built on top of simultaneously triggered pairs of 'happy' and 'sad' contrasts, in line with earlier psychological studies finding that these basic level affective terms capture half of the variance in emotional words [17]. Also the 'soft', 'cool' and 'dark' textures identified in the lyrics appear salient as they might not only be understood as abstract concepts, but in a larger context reflect sensorimotor percepts of touch or timbre metaphorically mapped onto feelings [37], similar to the examples described above of semantic neighborhoods formed around words such as 'hand' or 'dark' (Fig.3). Constrained to the two psychological dimensions of valence and arousal, the combinations of emotional topics could provide the affective building blocks that define a low dimensional representational structure of the input. Which would allow us to reconstruct the original signal from a incomplete set of linear affective mixtures forming sequential patterns that temporally reflect the emotional load. Embedded as cognitive components that we are able to retrieve as latent semantics when the bottom-up generated input raises above the background noise of core affect and top-down trigger distinct feelings in response to what we perceive.

Recent studies of haptic sensations indicate that simply holding a heavy object or encountering a rough surface unconsciously influence social judgments and decision tasks, as our conceptual knowledge seems to be grounded in aspects of touch that mentally trigger a multitude of associations [38]. In terms of language this provides a mental scaffolding for expanding the sensation of holding something heavy into metaphorically relating whether something is perceived as a weighty matter or a lighthearted insubstantial idea that cannot be taken seriously. Or our earliest experiences gained from touching objects become an embodied cognition of surfaces that abstractly translate into having a rough day or on the contrary feeling that everything runs smoothly. Textures like 'soft' therefore not only semantically lend themselves to interpretations along multiple dimensions altering their meaning to being lenient, a drink or jazz for easy listening but are also related to the two higher order factors: stability and plasticity, which in behavioral psychology capture the correlations among the Big Five personality traits describing how agreeable, conscientious and neurotic we are, complemented by how inclined we are to be open and extravert [39]. These two personality factors which determine how we maintain emotional stability while leaving sufficient plasticity to adapt to change, are in terms of neuropsychology linked to the neurotransmitters serotonin related to mood regulation and dopamine involved in reward mechanisms respectively. That is, textures such as 'soft' or 'dark' metaphorically extend into the very foundation for our behavioral responses. Which might again be projected onto the contrasts of 'happy' and 'sad' spanning the axes of valence and arousal within a psychological plane.

We could in that sense think of the simplified cognitive LSA model outlined above as both deriving the underlying meaning in texts by capturing the correlations of words in a vector space of reduced dimensionality. As well as emulating behavioral responses to emotions that are transformed into subjective experiences, once they are expressed in words and metaphorically mapped onto spatial structures and aspects of motion within time [40]. Meaning, that the contrasts of 'happy-sad' might be interpreted as not only principal emotional components, but also reflect memories capturing pleasure and pain of past experiences, that as feelings are conceptualized as bodily states integral to establishing our sense of self [41]. And in that respect we might view the bottom-up retrieved percepts of tension and relaxation and the top-down triggered mixtures of affective components, as meaningful chunks of semantics which wrapped as recursively nested strings of text exploit an underlying compact structure of how we perceive the world [42]. And thus provide a basis for designing affective interfaces for recommender systems, as it seems that even if we turn down the volume both the emotional context as well as overall structural elements can be extracted from songs by coupling latent semantics with cognitive components.

6. REFERENCES

[1] A.D. Patel: *Music, language and the brain*, Oxford University Press, 2008

[2] D.J. Levitin, M. Vinod: *Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence*, NeuroImage, 20:4, pp. 2142-2152, 2003

[3] I. Peretz, R.J. Zatorre: *Brain organization for music processing*, Annual review of psychology, 56, pp. 89-105, 2005

[4] P. Janata, S.T. Grafton: *Swinging in the brain: shared neural substrates for behaviors related to sequencing and music*, Nature Neuroscience, 6:7, pp. 682-687, 2003

[5] D.E Callan, V. Tsytsarev, T. Hanakawa, A.M. Callan, M. Katsuhara, H. Fukuyama, R. Turner: *Song and speech: Brain regions involved with perception and covert production*, NeuroImaga, 31:3, pp. 1327-1342, 2006

[6] S. Koelsch: *Neural substrates of processing syntax and semantics in music*, Current Opinion in Neurobiology, 15, pp. 207-212, 2005

[7] A.D. Patel: *Music, biological evolution and the brain*, Emerging disciplines, pp. 91-144, Rice University Press, 2010

[8] F. Lerdahl, C.L. Krumhansl: *Modeling tonal tension*, Music Perception, 24:4, pp. 329-366, 2010

[9] G. Rizzolatti, M.A. Arbib *Language within our grasp*, Trends in Neurosciences, 21:5, pp.188-194, 1998

[10] V. Gallese *Embodied simulation: From neurons to phenomenal experience*, Phenomenology and the Cognitive Sciences, 4, pp.23.48, 2005

[11] F. Feldman, S. Narayanan *Language that puts you in touch with your bodily feelings*, Psychological Science, 20:8, pp.974-980, 2009

[12] J. Foroni, G. Semin *Embodied meaning in a neural theory of language*, Brain and Language, 89, pp.385-392, 2004

[13] S. Duncan, L.F. Barret: *Affect is a form of cognition: a neurobiological analysis*, Cognition & Emotion, 21:6, pp.1184-1211, 2007

[14] S. Ortigue et al.: *Electrical neuroimaging reveals early generator modulation to emotional valence*, NeuroImage, 21, pp.1242-1251, 2004

[15] T.K. Landauer, S.T. Dumais: *A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge*, Psychological Review, 104:2, pp.211-240, 1997

[16] X. Hu, M. Bay, S. Downie: *Creating a simplied music mood classification ground-truth set*, Proceedings of the 8th ISMIR conference, pp.309-310, 2007

[17] J.A. Russell.: *A circumplex model of affect*, Journal of Personality and Social Psychology, 39:6, 1980

[18] J. Possner et al.: *The neurophysiological bases of emotion: an fMRI study of the affective circumplex using emotion-denoting words*, Human Brain Mapping, 30:3, pp.883-895, 2008

[19] M.M. Bradley, 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

[20] Y. Hu, X. Chen, D. Yang: *Lyric-based song emotion detection with affective lexicon and fuzzy clustering method*, Proceedings of the 10th ISMIR conference, 2009

[21] M.D. Korhonen et al.: *Modeling emotional content of music using system identification*, IEEE Transactions on systems, man and cybernetics, 36:3, pp. 588-599, 2006.

[22] Y. Yang, et al.: *Music emotion classification: a fuzzy approach*, IEEE Transactions on Audio, Speech and Language Processing, 16:.2, pp. 448-457, 2008.

[23] Y. Yang, et al.: *A regression approach to music emotion recognition*, Proceedings of 14th annual ACM international conference on multimedia, pp. 81-84, 2006.

[24] D. Turnbull et al.: *Semantic annotation and retrieval of music and sound effects*, IEEE Transactions on Audio, Speech and Language Processing, 16:2, pp. 467-476, 2008.

[25] D. Eck et al.: *Automatic generation of social tags for music recommendation*, Proceedings of NIPS, 2007.

[26] J. Aucouturier, B. Defreville, F. Pachet: *The bag-of-frames approach to audio pattern recognition: a sufficient model for urban soundscapes but not for polyphonic music*, Acoustical society of America, 122:2, pp. 881-891, 2007.

[27] M. Levy, M. Sandler: *Music information retrieval using social tags and audio*, IEEE Transactions on Multimedia, 11:3, pp. 383-395, 2009.

[28] M.K. Petersen, L.K. Hansen, A. Butkus: *Semantic contours in tracks based on emotional tags*, Lecture Notes in Computer Science, 5493, pp. 45-66, Springer, 2009.

[29] G. Furnas et al.: *Information retrieval using a singular value decomposition model of latent semantic structure*, Proceedings of 11th annual international ACM SIGIR conference, 1988.

[30] W. Kintsch: *Comprehension - a paradigm for cognition*, Cambridge University Press, 1998.

[31] W. Kintsch, V.L. Patel, A.K. Ericsson: *The role of long-term working memory in text comprehension*, Psychologia, 42, pp. 186-198, 1999.

[32] T.L. Griffiths, M. Steyvers, J.B. Tenenbaum: *Topics in semantic representation*, Psychological review, 114:2, pp. 211-244, 2007

[33] T.M. Fruchterman, M.J. Thomas, E.M. Reingold: *Graph drawing by force-directed placement*, Software - Practice And Experience, 21:11, pp. 1129-1164, 1991

[34] M. Girvan, M.E.J. Newman: *Community structure in social and biological networks*, PNAS, 99:12, pp. 7821-7826, 2002

[35] DTU Informatics: *LSA software for latent semantic analysis*, (Mac OSX, Linux) including HAWIK and LYWIK text corpora, http://dl.dropbox.com/u/5442905/LSA.zip., 2009.

[36] M.K. Petersen, M. Morup, L.K.Hansen: *Cognitive components as latent semantics*, IEEE Xplore, CIP 2nd International Workshop on Cognitive Information Processing, 2010

[37] A. Damasio: *Feelings of emotion and the self* , Annals of the New York Academy of Sciences, 1001, pp. 253-261, 2003

[38] J.M. Ackermann, C.C. Nocera, J.A. Bargh: *Incidental haptic sensations influence social judgments and decisions*, Science, 328, pp. 1712-1715, 2010

[39] C.G. DeYoung, J.B. Peterson, D.M. Higgins: *Higher-order factors of the Big Five predict conformity: are there neuroses of health*, Personal and Individual Differences, 33, pp. 533-552, 2002

[40] G. Lakoff, M. Johnson: *Philosophy in the flesh: the embodied mind and its challenge to western thought*, Basic Books, 1999

[41] A. Damasio: *Looking for Spinoza*, Vintage, 2003

[42] E.B. Baum: *What is thought?*, MIT Press, 2004