Technical University of Denmark

DTU

Extracting moods from songs and BBC programs based on emotional context

Petersen, Michael Kai; Butkus, Andrius

Link back to DTU Orbit

Citation (APA):
Petersen, M. K., & Butkus, A. (2008). Extracting moods from songs and BBC programs based on emotional context. International Journal of Digital Multimedia Broadcasting, 11. DOI: 10.1155/2008/289837

DTU Library
Technical Information Center of Denmark

Research Article

Extracting Moods from Songs and BBC Programs Based on Emotional Context

Michael Kai Petersen and Andrius Butkus

Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, 2800 Kongens Lyngby, Denmark

Correspondence should be addressed to Michael Kai Petersen, mkp@imm.dtu.dk

The increasing amounts of media becoming available in converged digital broadcast and mobile broadband networks will require intelligent interfaces capable of personalizing the selection of content. Aiming to capture the mood in the content, we construct a semantic space based on tags, frequently used to describe emotions associated with music in the *last.fm* social network. Implementing latent semantic analysis (LSA), we model the affective context of songs based on their lyrics, and apply a similar approach to extract moods from BBC synopsis descriptions of TV episodes using TV-Anytime atmosphere terms. Based on our early results, we propose that LSA could be implemented as machinelearning method to extract emotional context and model affective user preferences.

1. INTRODUCTION

When both digital broadcast streams and the content itself are adapted to the small screen size of handheld devices, it will literally translate into hundreds of channels featuring rapidly changing mobisodes and location-aware media, where it might no longer be feasible to select programs by scrolling through an electronic program guide. In order to automatically filter media according to personalized preferences, this will require metadata which not only defines traditional genre categories but also incorporates parameters capturing the changing mobile usage contexts. Since 2005, the broadcaster BBC has made their program listings available as XML formatted TVA TV-Anytime [1] metadata, which allows for describing media using complementary aspects, such as content genre, format, intended audience, intention, or atmosphere. We have previously in a related paper [2] analyzed how especially atmosphere metadata describing emotions may facilitate identifying programs that might be perceived as similar even though they belong to different genre categories. Also in music it appears that despite the often idiosyncratic character of tags, defined by hundred thousands of users in social networks like *last.fm*,

people tend to agree on the affective terms they attach to describe music [3, 4]. A mounting question might therefore be: could we possibly apply machine learning techniques to extract emotional aspects associated with media in order to model our perception, and thus facilitate an affective categorization which goes beyond traditional divides of genres?

2. RELATED WORKS

In usage scenarios involving DVB-H mobile TV, where shifting between a few channels might be even more time-consuming than watching the actual mobisode, new text mining approaches to content-based filtering have been suggested as a solution. Reflecting preferences for categories like "fun," "action," "thrill," or "erotic," topics and emotions are extracted from texts describing the programs and incorporated into the EPG electronic program guide data as a basis for generating user preferences [5]. In broadcast context, a similar approach has been implemented to extract both textual and visual concepts for automatic categorization of TV ad videos based on probabilistic latent semantic analysis (pLSA) [6]. As a machine learning method similar

to latent semantic analysis (LSA) [7], it captures statistical dependencies among distributions of visual objects or brand names, and thus enables unsupervised categorization of semantic concepts within the content. Recent neuroimaging experiments, focused on visualizing human brain activity reflecting the meaning of nouns, have demonstrated a direct relationship between the observed patterns in brain scans of regions being activated, and the statistics of word cooccurrence in large collections of documents. The distinct patterns of functional magnetic resonance images (fMRI) triggered by specific terms seem not only to cause similar brain activities across different individuals [8], but also makes it possible to predict which voxels in the brain will be activated according to semantic categories based on word cooccurrence in a large text corpus [9]. Or in other words, the way LSA simulates text comprehension by modelling the meaning of words as the sum of contexts in which they occur appears to have neural correlates.

Over 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 [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 mirror neurons, which transfer sensory information of what we perceive by reenacting them on a motor level. The mirror neuron populations mediate the inputs across audiovisual modalities and 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 reenactment of the perceived motion [17].

Aspects of musical affect have been the focus of a wide field of research, ranging from how emotions arise based on the underlying harmonic and rhythmical hierarchical structures forming our expectations [18–20], to how we consciously experience these patterns empathetically as contours of tensions and release [21], in turn triggering physiological changes in heart rate or blood pressure as has been documented in numerous cognitive studies of the links between music and emotions [22]. But when listening to songs our emotions are not only evoked by low-level cognitive representations but also exposed to higher level features reflecting the words which make up the lyrics. Studies on retrieving songs from memory indicate that lyrics and melody appear to be recalled from two separate versions: one storing the melody and another containing only the text [23], while further priming experiments indicate that song memory is not organized in strict temporal order, but rather that text and tune intertwine based on reciprocal connections of higher-order structures [24].

Taking the above findings into consideration, could we possibly extract affective components from textual representations of media like song lyrics, and model them as patterns reflecting how we emotionally perceive media? Applying LSA as a machine learning method to extract moods in both song lyrics and synopsis descriptions of BBC programs, we describe in the following sections, the methodology used for extracting high level representations of media using emotional tags, the early results retrieved when mapping emotional components of song lyrics and synopsis descriptions, and conclude with a discussion of the potential for automatically generating affective user preferences as a basis for mood-based recommendation.

3. EMOTIONAL TAG SPACE

When investigating how unstructured metadata can be used to describe media, the social music network *last.fm* provides an interesting case. The affective terms which are frequently chosen as tags by *last.fm* users to describe the emotional context of songs 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 [3, 4]. We have selected twelve of these frequently used tags for creating an emotional semantic space. Drawing on standard psychological parameters for emotional assessment, we map these affective terms along the two primary dimensions of *valence* and *arousal* [25], and use these two axes to outline an emotional plane for dividing them within an affective semantic space containing four groups of frequently used *last.fm* tags:

(i) *happy, funny, sexy;*

(ii) *romantic, soft, mellow, cool;*

(iii) *angry, aggressive;*

(iv) *dark, melancholy, sad.*

Within this 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 tags like angry or happy. Applying the selected *last.fm* tags as emotional buoys to define a semantic plane of psychological valence and arousal dimensions, we apply latent semantic analysis (LSA) to assess the correlation between the lyrics and each of the selected affective terms. Applying these affective terms as markers also enables us to compare the LSA-retrieved values against the actual tags users have applied in the *last.fm* tag clouds associated with the songs in our analysis. Additionally, when analyzing the synopsis descriptions of BBC programs we have complemented the *last.fm* tags with a large number of TV-Anytime atmosphere terms similarly used as emotional

buoys. Though the two sets of markers are clearly affected differently by the synopsis, a comparison shows that despite the higher degree of detail in the TV-Anytime vocabulary, the overall emotional context is reflected similarly by the *last.fm* tags and the atmosphere terms. Or in other words, the *last.fm* and TV-Anytime markers provide different granularities for capturing emotions but the larger tendencies in the resulting patterns remain the same.

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

4. RESULTS: SONG LYRICS

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 [26]. 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. Viewed from a neural network perspective, projecting the lyrics into a semantic LSA space line by line, could also in a cognitive sense be interpreted as similar to how mental concepts are constrained by the amount of activation among

the neural nodes representing events and associations in our working memory [27]. In that respect, the cooccurrence matrix formed by the word frequencies of *last.fm* tags and song lyrics might be understood as corresponding to the strengths of links connecting nodes in a mental model of semantic and episodic memory.

4.1. Accumulated emotional components

Projecting the lyrics of thirty songs selected from the weekly top track charts at *last.fm*, we compute the correlation between lyrics and tags against each of the twelve affective terms used as markers in the LSA space, while discarding cosine values below a threshold of 0.09. And in order to compare the retrieved LSA correlation values of lyrics and affective terms against the user-defined tags attached to the song at *last.fm*, we sum up the accumulated LSA values retrieved from each line of the lyrics.

Taking the song "Nothing else matters" as an example, the user defined tags attached to the song as at *last.fm*, include less frequently used tags like *love, love songs, chill, chillout, relaxing, relax, memories,* and *melancholic* which are not among the markers we used for our LSA analysis. We therefore subsequently combine these tags into larger segments of tags in order to facilitate a direct comparison with the LSA-retrieved values (Figure 1). Comparing the accumulated LSA values of emotional components against the user-defined tags at *last.fm*, the terms *melancholy,* and *melancholic*, which describe the most dominant emotions in the tag cloud, could be understood as captured by the affective term *sad* in the LSA analysis. Similarly, if interpreting *love* from the *last.fm* tag cloud as associated with the term *happy* (based on a cosine correlation of 0.56 between the words love and happy), the LSA analysis could be understood to retrieve also aspects of this emotion. Likewise, if *chill* in the *last.fm* tag cloud is understood as associated with *soft* and *mellow* (based on cosine correlations of 0.36 and 0.35, resp.), the LSA analysis also here appears to capture that mood.

Applying a similar approach to a set of thirty songs, we grouped semantically close *last.fm* tags into larger segments consisting of *sad, happy, love,* and *chill* aspects to facilitate a comparison with the LSA-derived correlations between song lyrics and the selected affective terms. Though there is an overlap between the retrieved LSA values and user-defined *last.fm* tags in most of the songs, there is no overall significant correlation between LSA-retrieved values and the exact distribution of tags in the user-defined *last.fm* tag clouds. Essentially, the individual tags in a cloud are "one size fits all" and apply to the song as a whole, whereas the LSA correlation between lyrics and semantic markers reflect the changing degrees of affinity between the song lines and affective components over time. But for a third of the set of songs, as exemplified by "Now at last" Figure 2, the distribution of *last.fm* tags resembled the LSA values if grouped into larger segments. While in the remaining two thirds of the set of songs, as exemplified by the song "Mad World" Figure 3, the overall distribution in *last.fm* tags while
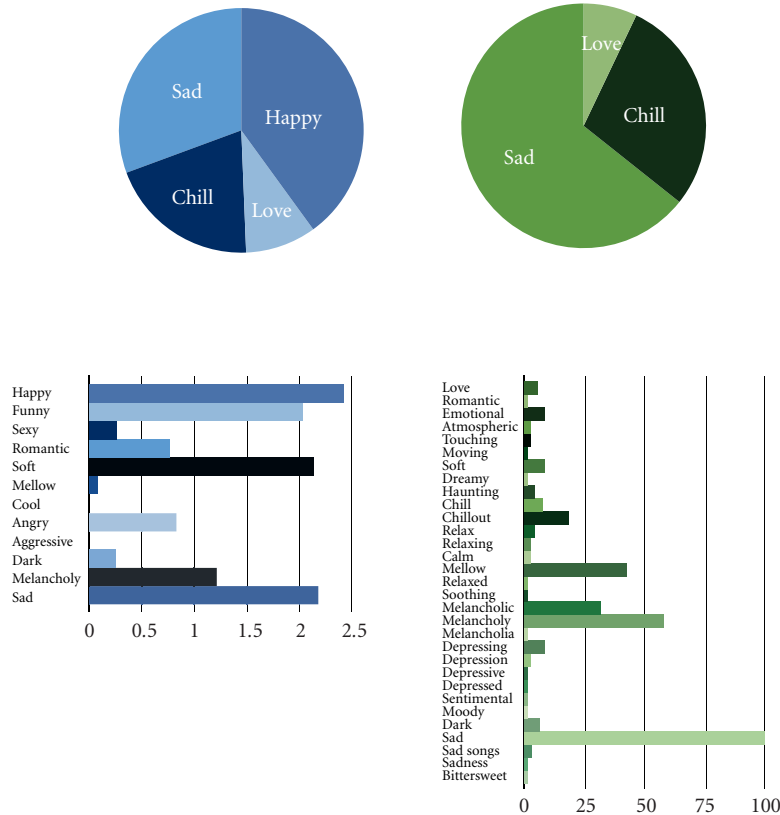
FIGURE 1: Accumulated LSA correlation between the lyrics of the song "Nothing else matters" and 12 affective terms (left), compared to the actual user-defined emotional tags at last.fm (right).

clearly overlapping remain overly biased toward *sad* type of components.

4.2. Distribution of emotional components

Instead of grouping the emotional components into larger segments, we subsequently maintained the LSA values retrieved from each of the individual lines in the lyrics, and proceeded by plotting the values over time to provide a view of the distribution of emotional components. The plots can be interpreted as mirroring the structure of patterns of changing emotions in the songs along the horizontal axis. Vertically, the color groupings indicate which of the aspects of valence and arousal are triggered by the lyrics as well as their general distribution in relation to each other. Any color will signify an activation beyond the cosine similarity threshold level of 0.09, and the amount of saturation from light to dark signifies the degree of correlation between the song lyrics and each of the affective terms. The contribution of each emotional component apparent in the overall LSA values of the lyrics can be made out when considering their distribution as single pixels over time triggered by the individual lines in each of the songs. When analyzing which emotional components appear predominant and overall contribute the most, the LSA plots can roughly be grouped into three categories which can be characterized as

unbalanced distributions, *centered distributions,* and *uniform distributions.*

Going back to the song "Nothing else matters," Figure 4, the plot exemplifies the first *unbalanced* category by in this case having a bottom-heavy distribution of emotional components biased toward *melancholy*. The below curve of accumulated LSA values indicates the contribution of each component over the entire song, where the significant aspects of *melancholy* are clearly separated from the other components.

The *centered distribution* distribution as found in "Now at last" Figure 5, shows a lack of the more explicit emotions like "happy" or "sad" apart from the very beginning, while instead the main contribution throughout the song comes from more passive "mellow" and "soft" aspects. In contrast to the former example, the below curves of accumulated emotional contributions reflect a pattern combining the activation of "happy" or "sad" elements which remain at the initial level, whereas the more passive aspects "mellow" and "soft" are continuously accumulating throughout the song.

A *uniform distribution* of a wide range of simultaneous emotional components is exemplified by "mad world," Figure 6, simultaneously juxtaposing emotional areas around "happy" against "sad" components. This pattern can also be made out in the below curves, where additionally the sudden steep increase in accumulated values starting roughly
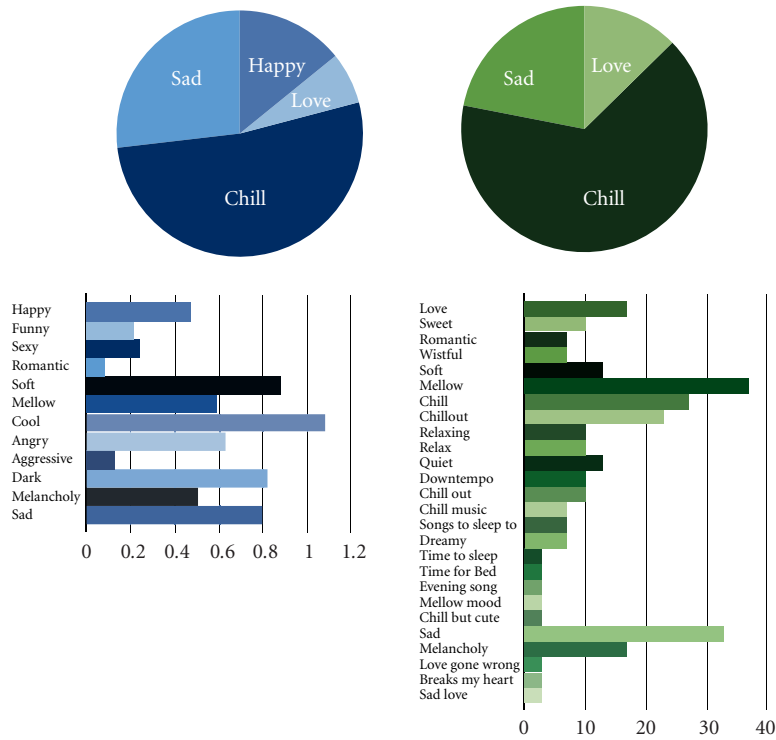
Figure 2: Accumulated LSA correlation between the lyrics of the song "Now at last" and 12 affective terms (left), compared to the actual user-defined emotional tags at last.fm (right).
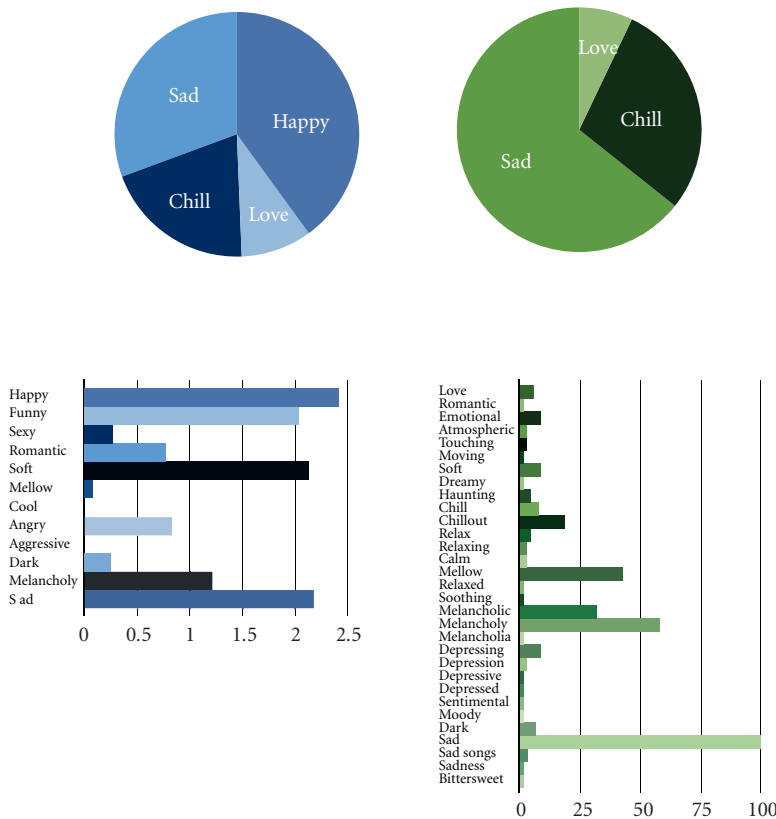

Figure 3: Accumulated LSA correlation between the lyrics of the song "Mad world" and 12 affective terms (left), compared to the actual user-defined emotional tags at last.fm (right).
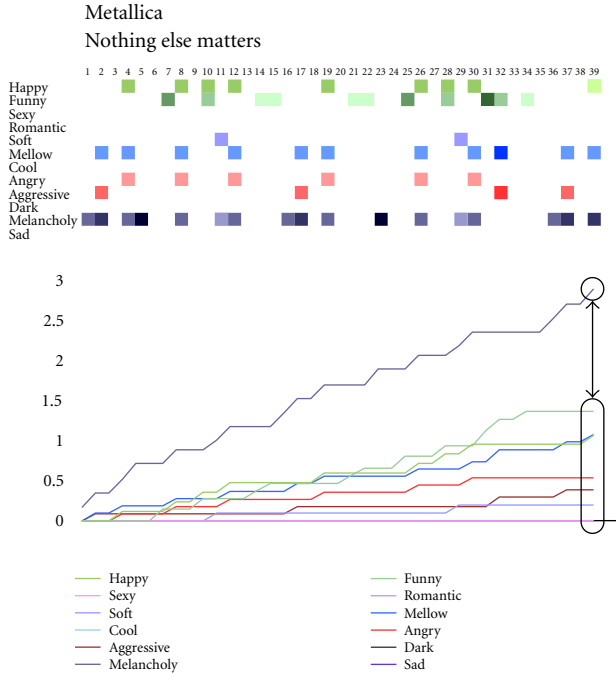
FIGURE 4: LSA correlation between the lyrics of the song "Nothing else matters" and 12 affective terms (top), with accumulated values plotted over the entire length of the song (bottom).

FIGURE 5: Summed up values of LSA correlation between the lyrics of the song "Now at last" and 12 affective terms (top), with accumulated values plotted over the entire length of the song (bottom).
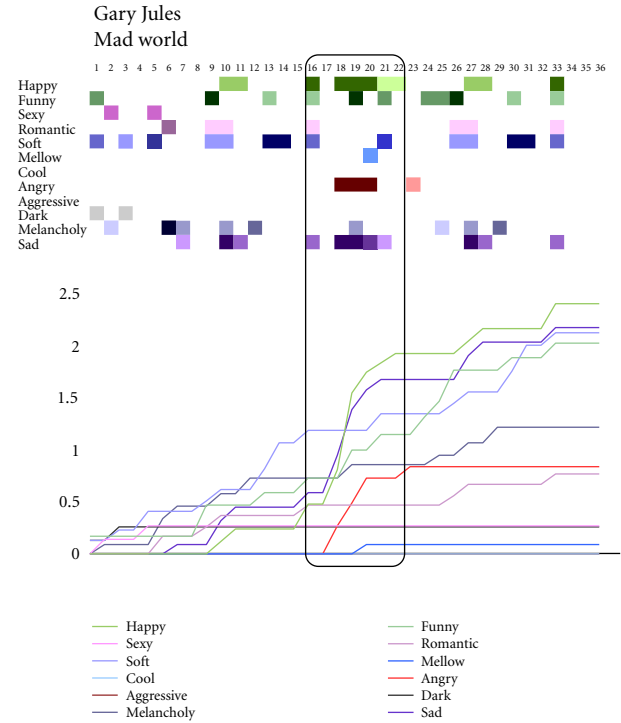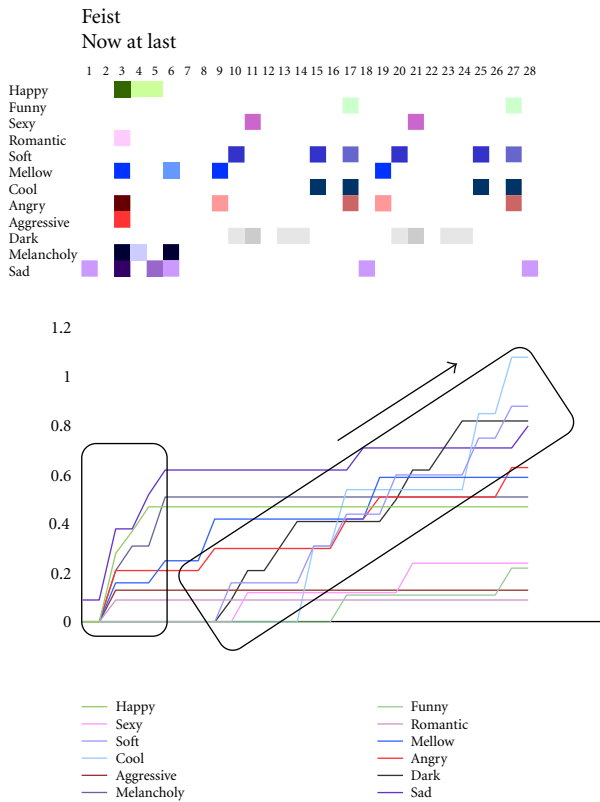

FIGURE 6: Summed up values of LSA correlation between the lyrics of the song "Mad world" and 12 affective terms (top), with accumulated values plotted over the entire length of the song (bottom).

a third into the song also illustrates how the emotional components reflect the overall structure in the song.

The overall saturation defining the amount of correlation between lyrics and emotional markers, as well as the distributional patterns of emotional components throughout the songs seem consistent. Lyrics that appear more or less saturated in relation to the emotional markers used for the LSA analysis remain so over the entire song. The distributional patterns of emotional elements seem throughout the songs to form consistent schemas of contrasting elements, which appear to form sustained lines or clusters that are preserved as pattern once initiated. We suggest that these elements form bags of features, which could be used to categorize and infer patterns as a basis for building emotional playlists. From these features, general patterns emerge, as in the distributions of emotional components in the songs "Wonderwall" and "My Immortal," Figure 7, which appear similar due to a sparsity of central aspects like "soft," while instead emphasizing the outer edges by juxtaposing elements around "happy" against "sad." The opposite character can be seen in the distributions of central elements stressed in the songs "Falling slowly" and "Stairway to heaven," Figure 7, which underline the aspects of "soft" and "mellow" at the expense of "happy" and "sad." Whereas these elements in the songs "Everybody hurts" and "Smells like teen spirit," Figure 8, appear as structural components grouped into clusters, either providing a strong continuous activation

of complementary feelings or juxtaposing these emotional components against each other.

5. RESULTS: BBC SYNOPSIS

Repeating the approach, but this time to extract emotions from texts describing TV programs, we take a selection of short BBC synopses as input, and compute the cosine similarities between a synopsis text vector and each of the selected *last.fm* emotional words. While the previously analyzed lyrics could be seen as integral parts of the original media, a synopsis description is clearly not. It only provides a brief summary of the program, but it nevertheless offers an actual description complementary to the associated *TV-Anytime* metadata genres. We initially analyzed a number of standalone synopsis descriptions to see if would be possible to capture emotional aspects of the BBC programs.

An analysis of the program "News night," based on the short description: *News in depth investigation and analysis of the stories behind the day('s) headline*, triggers the tags "funny" and "sexy" which might not immediately seem a fitting description, probably caused by these emotional terms being directly correlated with the occurrence of the words stories and news within the synopsis. The atmosphere of the lifestyle program "Ready Steady Cook!" might be somewhat better reflected in the synopsis: *Peter Davidson and Bill Ward challenge celebrity chefs to create mouth watering meals in minutes*, which triggers the tag "romantic" as associated with meals. Another singular emotion can be retrieved from the documentary "I am a boy anorexic," which based on the synopsis: *Documentary following three youngsters struggling to overcome their obsessive relationship with food as they recover inside a London clinic and then return to the outside world*, triggers the affective term "dark." We find a broader emotional spectrum reflected in the lifestyle program "The flying gardener" described by the text: *The flying gardener Chris travels around by helicopter on a mission to find Britain('s) most inspirational gardens. He helps a Devon couple create a beautiful spring woodland garden. Chris visits impressive local gardens for ideas and reveals breathtaking views of Cornwall from the air*. The synopsis triggers a concentration of passive pleasant *valence* elements related to the words "soft, mellow" combined with "happy." In this context also the tag "cool" comes out as it has a strong association to the word air contained in the synopsis, while the activation of the tag "aggressive" appears less explainable. This cluster of pleasant elements is lacking in the LSA analysis of the program "Super Vets" which instead evokes a strong emotional contrast based on the text: *At the Royal Vet College Louis the dog needs emergency surgery after a life threatening bleed in his chest and the vets need to find out what is causing the cat* (· · ·) *fits*, where both pleasant and unpleasant active terms like "happy" and "sad" stand out in combination with strong emotions reflected by the tag "romantic." And as can be seen from programs like "The flying gardener" and "Super Vets" (**Figure** 10), the correlation between the synopsis and the chosen tags might often trigger both complementary elements as well as contrasting emotional components.

We proceeded to explore whether we could sum up a distinct pattern reflecting an emotional profile pertaining to a TV series, by accumulating the LSA values of correlation between synopsis texts and emotional tags over several episodes. Similar to our previous approach when analyzing lyrics, where we held the LSA results against the user defined *last.fm* tag clouds, we here compare the LSA values of the synopsis against the *TV-Anytime* atmosphere genres used in the BBC metadata. This classification scheme offers 53 different terms which might be included in the genre metadata to express the atmosphere or perceived emotional response when watching a program. Projecting the synopsis descriptions against 53 *TV-Anytime* terms, used as emotional markers in the LSA analysis, allows for defining more differentiated patterns. At the same time also projecting the BBC synopsis against the previously used *last.fm* tags in the LSA analysis, makes it possible to compare to what extent the choice of using either *TV-Anytime* atmosphere terms or *last.fm* tags as emotional markers in the semantic space is influencing the results.

For analyzing the emotional context in a sequence of synopsis descriptions of the same program, we chose the soap "East Enders," the comedy "Two pints of lager," and sci-fi series "Doctor Who." Initially, plotting the LSA analysis of the soap "East Enders" and comedy "Two pints of lager" against 12 *last.fm* tags (Figures 1 and 2, increased color saturation corresponds to degree of correlation), the distributions of emotional components appear unbalanced in both cases. But whereas the soap has a bottom-heavy bias toward "sad" and "angry" outweighing "happy," the balance is reversed in the comedy which shifts towards predominantly "happy" and "funny" complemented by "soft" and "mellow" aspects. Overall, the distribution in "East Enders" is much more dense and emotionally saturated as exemplified in elements like "angry" reflecting high arousal. In contrast, the lighter character of "Two pints of lager" comes out in the clustering of positive valence elements such as "happy" and "funny," coupled with a general sparsity of excitation within the matrix.

As a second step, projecting the synopsis descriptions against the 53 *TV-Anytime* atmosphere terms of course results in more differentiated patterns. Users at *last.fm* frequently describe tracks as "angry" but as music is rarely described as scary, feelings of fear are lacking. Otherwise, so with the *TV-Anytime* metadata which also captures these aspects in a synopsis with atmosphere terms like "terrifying." Some of these elements are essential for describing the content as is evident in the sci-fi series "Doctor Who," **Figure** 13. Lacking words for these feelings, the *last.fm* tags "Melancholy" and "dark" are triggered, whereas it takes the increased resolution of the *TV-Anytime* atmosphere terms to capture the equally "spooky" and "silly" aspects.

Altogether *TV-Anytime* adds a large number of terms, which rather than describing emotions capture attitudes or perceived responses like "stylish" or "compelling," and as such trigger vast amounts of elements contributing to the atmosphere. In "East Enders" adding elements like "frantic" and "exciting" to the pattern. Similarly, the larger number of comical elements exemplified by words like "crazy, silly," or
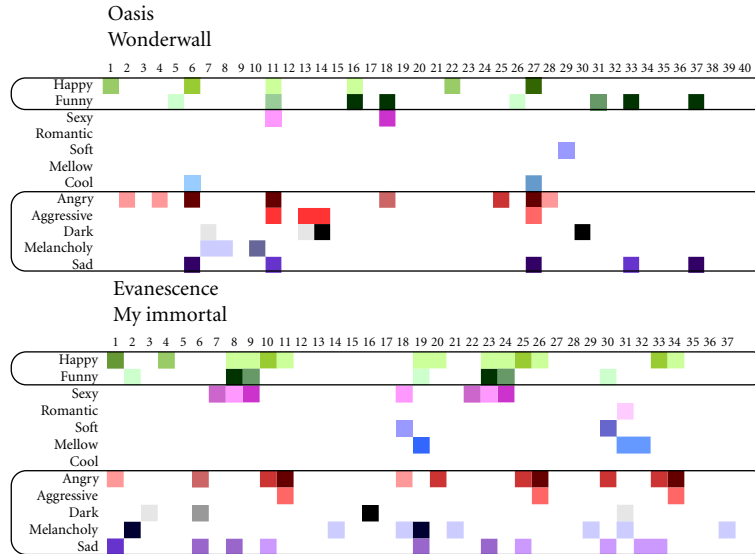
FIGURE 7: Pairwise comparison of patterns reflecting LSA correlation values in the lyrics of the songs "Wonderwall" (top) and "My immortal" (bottom) against 12 affective terms.
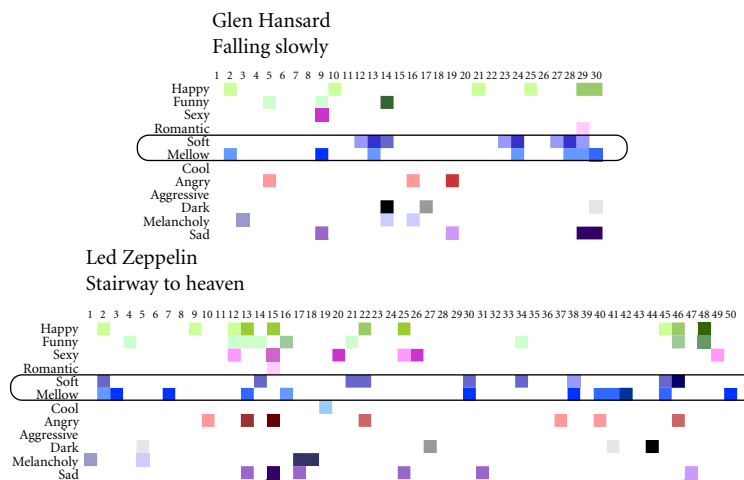
FIGURE 8: Pairwise comparison of patterns reflecting LSA correlation values in the lyrics of the songs "Falling slowly" (top) and "Stairway to heaven" (bottom) against 12 affective terms.

"wacky" provides a much higher emotional granularity in the description of "Two pints of lager". However, the overall bias toward positive or negative valence and arousal within the distributions seem largely preserved, independent of whether *last.fm* or *TV-Anytime* terms are used as emotional markers in the LSA analysis.

Comparing the emotional components retrieved from the LSA analysis of the synopsis texts against the actual *TV-Anytime* atmosphere terms in the BBC metadata, they seem to be largely in agreement. The comedy has been indexed as "humorous, silly, irreverent, fun, wacky, crazy," while based on the synopsis texts alone, most of these components also come out in the LSA analysis. In the case of the soap "East Enders," the episodes are annotated as "gripping, gritty, gutsy." Although these terms are also triggered from the synopsis texts, these aspects might be even more reflected

in the stark accumulated contrasts of "happy" and "sad" components retrieved by the LSA analysis. Similarly, in "Doctor Who" the actual *TV-Anytime* atmosphere terms applied in the BBC metadata *spooky, exciting* are also captured, while the grey patterns of perceived responses seem to add a lot more nuances to this description.

6. CONCLUSIONS

Projecting BBC synopsis descriptions into an LSA space using both *last.fm* tags and *TV-Anytime* atmosphere terms as emotional buoys, we have demonstrated an ability to extract patterns reflecting combinations of emotional components. While each synopsis triggers an individual emotional response related to a specific episode, general patterns still emerge when accumulating the LSA correlation
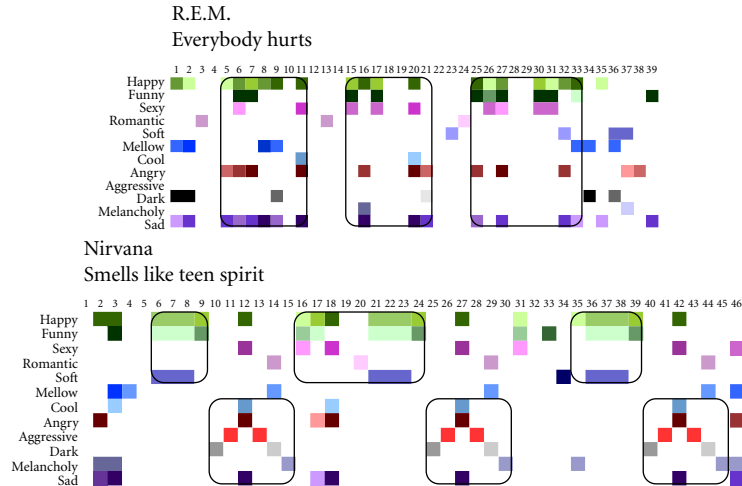
FIGURE 9: Pairwise comparison of patterns reflecting LSA correlation values in the lyrics of the songs "Everybody hurts" (top) and "Smells like teen spirit" (bottom) against 12 affective terms.
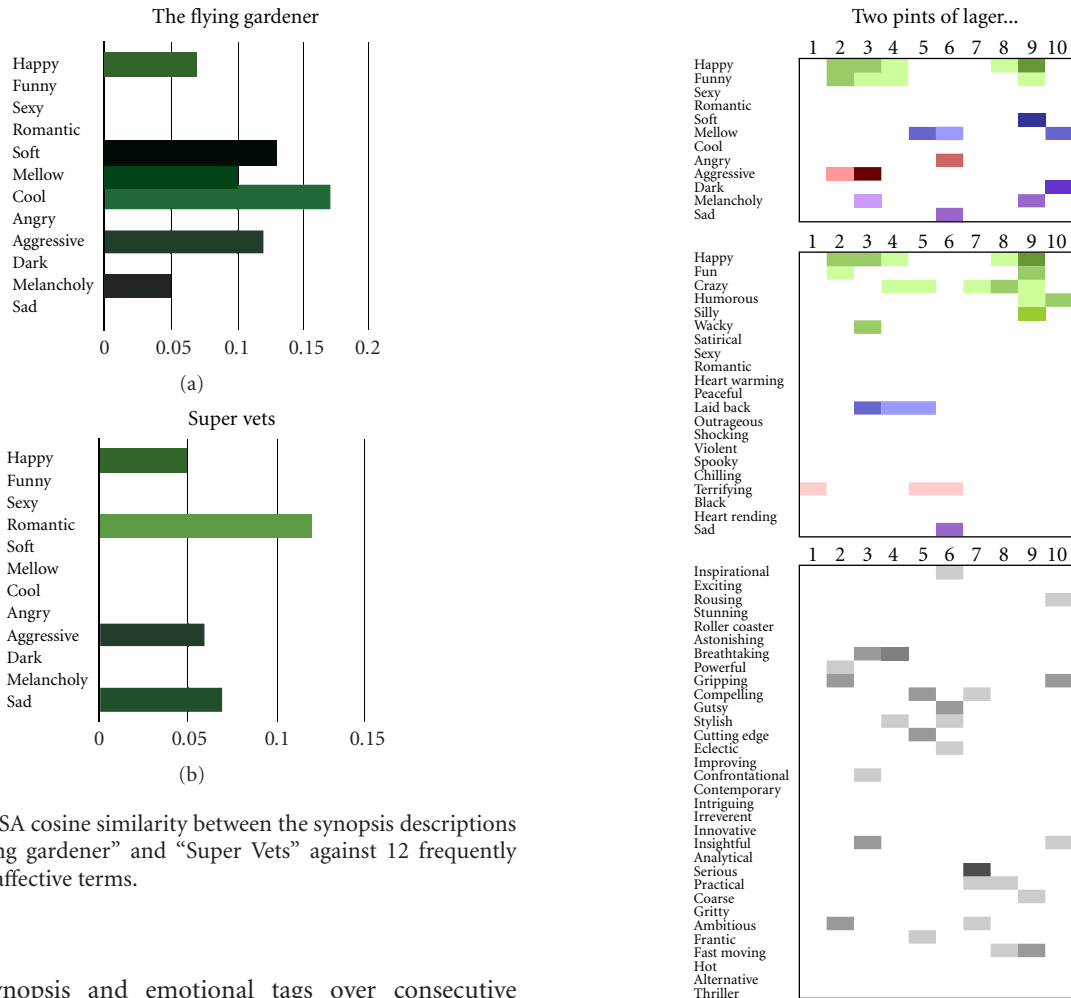

FIGURE 10: LSA cosine similarity between the synopsis descriptions of "The flying gardener" and "Super Vets" against 12 frequently used *last.fm* affective terms.

FIGURE 11: LSA correlation values of 10 episodes of "Two Pints of lager" against 12 last.fm tags (top), 53 tva atmosphere terms (bottom).

between synopsis and emotional tags over consecutive episodes, which enables us to differentiate between a comedy and a soap based on textual descriptions alone. Applying more semantic markers in the analysis allows for capturing additional elements of atmosphere in terms of perceived attitudes or responses to the media being consumed. How-

Figure 12: LSA correlation values of 18 episodes of "East Enders" against 12 last.fm tags (top), and 53 tva atmosphere terms (bottom).
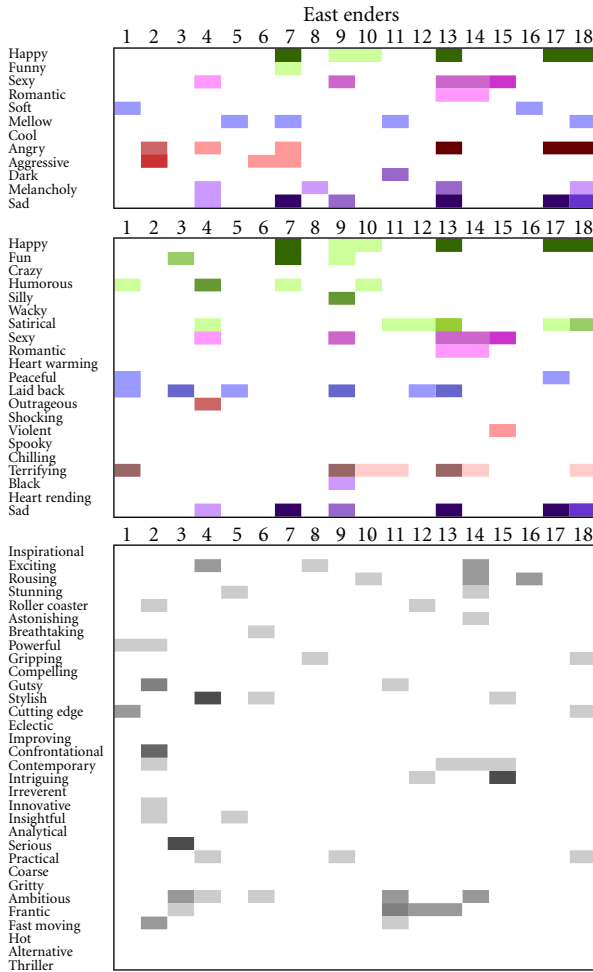

Figure 13: LSA correlation values of 12 episodes of "Doctor Who" against last.fm tags (top), and 53 tva atmosphere terms (bottom).

ever, the overall balance of affective components reflecting the media content seems largely preserved, independent of whether *last.fm* or *TV-Anytime* terms are used as emotional markers in the LSA analysis.

Moving beyond the static LSA analysis of consecutive synopsis descriptions, plotting the components over time might provide a basis for modelling the patterns of emotions evolving when we perceive media. We hypothesize that these emotional components reflect compositional structures perceived as patterns of tension and release, which form the dramatic undercurrents of an unfolding story line. 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 [26]. 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 by the media. Or from a cognitive perspective,

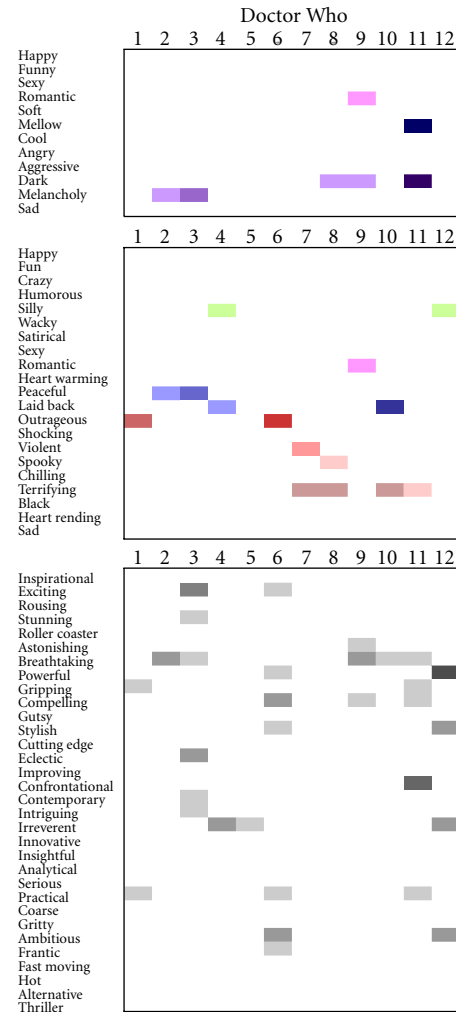the dimensionality reduction enforced by LSA might be interpreted as a simplified model of how mental concepts are constrained by the strengths of links connecting nodes in our working memory [27].

Finding that the emotional context of media can be retrieved by using affective terms as markers, we propose that LSA might be applied as a basis for automatically generating mood-based recommendations. It seems that even if we turn off both the sound and the visuals, emotional context as well as overall formal structural elements can still be extracted from media based on latent semantics.

REFERENCES

[1] ETSI, TV-Anytime. Part 3. Metadata 1. Sub-part 1. Part 1—Metadata schemas, TS 102822-3-1, 2006.

[2] A. Butkus and M. K. Petersen, "Semantic modelling using TV-anytime genre metadata," in *Proceedings of the 5th European Conference on Interactive TV: A Shared Experience (EuroITV '07)*, P. Cesar, K. Chorianopoulos, and J. F. Jensen, Eds., vol. 4471 of *Lecture Notes in Computer Science*, pp. 226–234, Springer, Amsterdam, The Netherlands, May 2007.

[3] M. Levy and M. Sandler, "A semantic space for music derived

from social tags," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR '07)*, pp. 411–416, Vienna, Austria, September 2007.

[4] X. Hu, M. Bay, and S. J. Downie, "Creating a simplified music mood classification ground-truth set," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR '07)*, pp. 309–310, Vienna, Austria, September 2007.

[5] A. Bär, A. Berger, S. Egger, and R. Schatz, "A lightweight mobile TV recommender," in *Proceedings of the 6th European Conference on Interactive TV: A Shared Experience (EuroITV '08)*, M. Tscheligi, M. Obrist, and A. Lugmayr, Eds., vol. 5066 of *Lecture Notes in Computer Science*, pp. 143–147, Springer, Salzburg, Austria, July 2008.

[6] J. Wang, L. Duan, L. Xu, H. Lu, and J. S. Jin, "TV ad video categorization with probabilistic latent concept learning," in *Proceedings of the 9th ACM SIG Multimedia International Workshop on Multimedia Information Retrieval (MIR '07)*, pp. 217–226, Bavaria, Germany, September 2007.

[7] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211–240, 1997.

[8] K. Skreiner, "In the news: machine learning takes on the brain," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 7–8, 2008.

[9] T. M. Mitchell, S. V. Shinkareva, A. Carlson, et al., "Predicting human brain activity associated with the meanings of nouns," *Science*, vol. 320, no. 5880, pp. 1191–1195, 2008.

[10] D. J. Levitin and V. Menon, "Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence," *NeuroImage*, vol. 20, no. 4, pp. 2142–2152, 2003.

[11] S. Koelsch and W. A. Siebel, "Towards a neural basis of music perception," *Trends in Cognitive Sciences*, vol. 9, no. 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, no. 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," in *Proceedings of the 10th International Conference on Music Perception & Cognition (ICMPC '08)*, Sapporo, Japan, August 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, no. 1, 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, no. 3-4, 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, no. 33, pp. 235–241, 2006.

[18] L. B. Meyer, "Meaning in music and information theory," *Journal of Aesthetics and Art Criticism*, vol. 15, no. 7, pp. 412–424, 1957.

[19] D. Temperley, *Music and Probability*, MIT Press, Cambridge, Mass, USA, 2007.

[20] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*, MIT Press, Cambridge, Mass, USA, 2006.

[21] R. Jackendoff and F. Lerdahl, "The capacity for music: what is it, and what's special about it?" *Cognition*, vol. 100, no. 1, pp. 33–72, 2006.

[22] C. L. Krumhansl, "Music: a link between cognition and emotion," *Current Directions in Psychological Science*, vol. 11, no. 2, pp. 45–50, 2002.

[23] I. Peretz, R. Gagnon, and S. Hebert, "Singing in the brain: insights from cognitive neuropsychology," *Music Perception*, vol. 21, no. 3, pp. 71–81, 2004.

[24] I. Peretz, M. Radeau, and M. Arguin, "Two-way interactions between music and language: evidence from priming recognition of tune and lyrics in familiar songs," *Memory and Cognition*, vol. 32, no. 1, pp. 142–152, 2004.

[25] M. M. Bradley and P. J. Lang, "Afective norms for English words (ANEW): stimuli, instruction manual and affective ratings," Tech. Rep. C-1, The Center for Research in Psychophysiology, University of Florida, Gainesville, Fla, USA, 1999.

[26] E. Pöppel, "A hierarchical model of temporal perception," *Trends in Cognitive Sciences*, vol. 1, no. 2, pp. 56–61, 1997.

[27] W. Kintsch, *Comprehension—A Paradigm for Cognition*, Cambridge University Press, Cambridge, UK, 1998.

Composition Comments

1. We reduced the font size of some labels in all Figures. Please check.

2. We changed "(..)" to the highlighted "$(\cdots)$." Please check.

3. Please cite Figures 9, 11, and 12 in the paper.