TIMBRAL QUALITIES OF SEMANTIC STRUCTURES OF MUSIC

Rafael Ferrer and Tuomas Eerola
Finnish Centre of Excellence in Interdisciplinary Music Research
rafael.ferrer-flores@jyu.fi; tuomas.eerola@jyu.fi

ABSTRACT

The rapid expansion of social media in music has provided the field with impressive datasets that offer insights into the semantic structures underlying everyday uses and classification of music. We hypothesize that the organization of these structures are rather directly linked with the "qualia" of the music as sound. To explore the ways in which these structures are connected with the qualities of sounds, a semantic space was extracted from a large collection of musical tags with latent semantic and cluster analysis. The perceptual and musical properties of 19 clusters were investigated by a similarity rating task that used spliced musical excerpts representing each cluster. The resulting perceptual space denoting the clusters correlated high with selected acoustical features extracted from the stimuli. The first dimension related to the high-frequency energy content, the second to the regularity of the spectrum, and the third to the fluctuations within the spectrum. These findings imply that meaningful organization of music may be derived from low-level descriptions of the excerpts. Novel links with the functions of music embedded into the tagging information included within the social media are proposed.

1. INTRODUCTION

Attempts to craft a bridge between acoustic features and the subjective sensation they provoke [3] have usually started with concepts describing instrument sounds, using adjectives or bipolar scales (e.g., bright-dark, static-dynamic) and matching these with acoustic descriptors (such as shape of the envelope and energy distribution) [11, 20].

In this study, we present a purely bottom-up approach to the conceptual mapping between sound qualities and emerging meanings. We utilized social media to obtain a wide sample of music and extract an underlying semantic structure of this sample. Next, we evaluated the validity of the obtained mapping by investigating the acoustic features underlying the semantic structures. This was done by an analyzing of the examples representing the semantic space, and by having participants to rate the similarity of random spliced sound examples representing the semantic space.

Social tagging is an activity, where descriptive verbal characterizations are given to items of interest, such as songs, images, or links as a part of the normal use of the popular online services. Tags can be considered as semantic representations of abstract concepts created essentially for mnemonic purposes and used typically to organize items [14]. Tagging music is not a novel idea, as any labeling scheme such as musical genres may be considered as tags themselves, but in recent years in the context of social networks, tagging has acquired a new relevance and meaning [1].

Despite all the possibilities offered by large databases containing tags, a central problem remains on how to derive an ontology from them [19]. Starting with the assumption of an underlying structure existing in an apparently unstructured set, we consider a sample of tags to extract a semantic structure, explained next.

2. ANALYSIS OF TAGS

2.1 Material

A collection of 6372 songs [7] representing 15 musical genres (Alternative, Folk, Finnish Iskelmä, Pop, World, Blues, Gospel, Jazz, Rock, Classical, Heavy, Soul, Electronic, Hip-Hop, Soundtrack) served as the initial database of music. Musical genres were used in establishing the sample in order to maximize musical variety in the collection and to be compatible with a host of music preference studies (e.g., [6, 22]) that have provided lists of 13 to 15 broad musical genres relevant for most Western adult listeners. The tags related to the songs in this collection were retrieved from an online music service (*last.fm* [1]) with a dedicated API (Application programming interface) named *Pylast* [2] .

2.2 Description of the corpus

The retrieved *corpus* consists of 5,825 lists of tags (mean length of 62.27 tags), each list (*document* in this context) is associated with a piece of music. The number of times each tag had been used in the system until the time of the retrieval was also obtained, representing a measure of "popularity".

[1] http://www.last.fm
[2] http://code.google.com/p/pylast/

In total, the corpus contains 362,732 tags, from which 77,537 are distinct. Each tag is formed by one or more words (M=2.48, SD=1.86), a small proportion of the distinct tags in the corpus contain long expressions (e.g. 6% of the distinct tags are formed by 5 words or more). In this study a tag is considered as a unit representing an element of the *vocabulary*, disregarding the number of words that compose it. Treating tags as *collocations* (i.e. frequent juxtaposition of words) shifts the focus from data processing to concept processing [2], also allowing the tags to function as conceptual expressions [23] instead of words or phrases.

2.3 Lexical layers of the vocabulary

Preprocessing is necessary in any text mining application because retrieved data does not follow any particular set of rules, and there are not standard steps to follow [13].

Three filtering rules where applied to the corpus in the quantitative domain. First, *hapax legomena* (i.e. tags used only once in the corpus), are removed under the rationale of discarding unrelated data. To capture the most prevalent and relevant tags, a second filter uses the associated popularity measure of each tag to eliminate the tags below the mean popularity index of the vocabulary. The third step eliminates tags with three or more words to prune short sentence-like descriptions from the corpus. The subset resulting from such reductions represents 46.6% of the corpus (N=169,052, Vocabulary=2,029 tags).

At this point, data has been de-noised but for the extraction of a meaningful semantic ontology from the tags, a semantic analysis and qualitative filtering is necessary. To categorize the tags at a functional level [24] (e.g. musicological and lexicological), an analysis was performed by using the Brown Corpus [9] as parts-of-speech (POS) tagger, Wordnet database [8] for word sense disambiguation, and Urban Dictionary online [3] and Last.fm database for general reference. Tags are looked-up in these sources and the selection of a category is decided by reviewing each case. The criteria applied in this process favors categories closely related to music, such as genre, artist, instrument, form and company, then adjectives, and finally other types. For instance, "Acid" is a noun but it is also a term extensively used to describe certain musical genres, so it was classified according to its musical function. Proposed categories, percentage of the vocabulary, definition and examples are shown in Table 1. The resulting layers were used to make a finer discrimination of the tags to uncover the semantic structure. Since one of the main motivations of this project was to obtain prototypical timbral descriptions, we focused on tags related to adjectives, nouns, instruments, temporal and verbs.

2.4 Semantic structure

Tag structure (or folksonomy) is obtained by using *latent semantic analysis* (LSA) as a framework [5], a method that has been used before in the domain of musical tags [17, 18]. In this study, detection of semantic structure has three stages: 1) construction of a *Term-Document Matrix*, 2) calculation of similarity coefficients, and 3) cluster analysis. First, a Term-Document Matrix $\mathbf{X} = \{x_{ij}\}$ is constructed. Where each song i, corresponds to a "Document" and each unique tag (or item of the vocabulary) j, to a "Term". The result is a binary matrix $\mathbf{X}(0, 1)$ containing information about the presence or absence of a particular tag to describe a given song. Second, a similarity matrix $n \times n$ $\mathbf{D}$ with elements d_{ij} where $d_{ii} = 0$ for all i, is created by computing similarity indexes between tag vectors x_{i*j} of $\mathbf{X}$ with:

$$d_{ij} = \frac{ad}{\sqrt{(a+b)(a+c)(d+b)(d+c)}} \quad (1)$$

where *a* is the number of (1,1) matches, *b* for (1,0), *c* for (0,1) and *d* for (0,0).

There are several methods to compute similarity coefficients between binary vectors (c.f., [10]). This coefficient was selected because of its *symmetric* quality, which considers the double absence (0,0) as important as (1,1), that presumably has positive impact on ecologic applications [10]. A hierarchical clustering algorithm was used to transform the similarity matrix into a sequence of nested partitions. The method used in the hierarchical clustering was Ward's minimum variance, to find compact, spherical clusters [21] and because it has demonstrated its proficiency in comparison to other methods [12].

After obtaining a hierarchical structure, the clusters are derived from the resulting dendrogram by "pruning" the branches with an algorithm that uses a partitioning around medioids (PAM) clustering method in combination with the height of the branches [15]. Figure 1 shows a two dimensional projection (obtained with multidimensional scaling) of the similarity matrix used in the hierarchical clustering. Each dot represents a tag, and the numbers show the centers of their corresponding clusters. Each number is enclosed in a circle that shows the relative size of the cluster in terms of the number of tags contained in it. A more detailed reference on the content of the clusters can be consulted in Table 2.

2.5 Ranking of musical examples in the clusters

In order to explore any acoustic or musical aspects of the clusters, we need to link the clusters with the specific songs represented by the tags. For this, a $m \times n$ *Term Document Matrix* (TDM) $\mathbf{X} = \{x_{ij}\}$ is constructed, where lists of tags attributed to a particular song are represented as m, and preselected tags as n. A list of tags is a finite set $\{1, ..., k\}$, where $1 \leq k \leq 96$. Each element of the matrix contains a value of the normalized rank of a tag if found on a list, and it is defined by:

$$x_{ij} = \left(\frac{r_k}{k}\right)^{-1} \quad (2)$$

Where r_k is the cardinal rank of the tag j if found in i, and k is the total length of the list. To obtain a cluster profile,

[3] http://www.urbandictionary.com

Categories	%	Definition	Examples
Genre	36.72%	Musical genre or style	Rock, Alternative, Pop
Adjective	12.17%	General category of adjectives	Beautiful, Mellow, Awesome
Noun	9.41%	General category of nouns	Love, Melancholy, Memories
Artist	8.67%	Artists or group names	Coldplay, Radiohead, Queen
Locale	8.03%	Geographic situation or locality	British, American, Finnish
Personal	6.80%	Words used to manage personal collections	Seen Live, Favourites, My Radio
Instrument	4.83%	Sound source	Female vocalists, Piano, Guitar
Unknown	3.79%	Unclassifiable gibberish	aitch, prda, <3
Temporal	2.41%	Temporal circumstance	80's, 2000, Late Romantic
Form	2.22%	Musical form or compositional technique	Ballad, Cover, Fusion
Company	1.72%	Record label, radio station, etc.	Motown, Guitar Hero, Disney
Verb	1.63%	General category of verbs	Chillout, Relax, Wake up
Content	1.03%	Emphasis in the message or literary content	Political, Great lyrics, Love song
Expression	0.54%	Exclamations	Wow, Yeah, lol

Table 1. Main categories of tags, their prevalence, definition and examples.

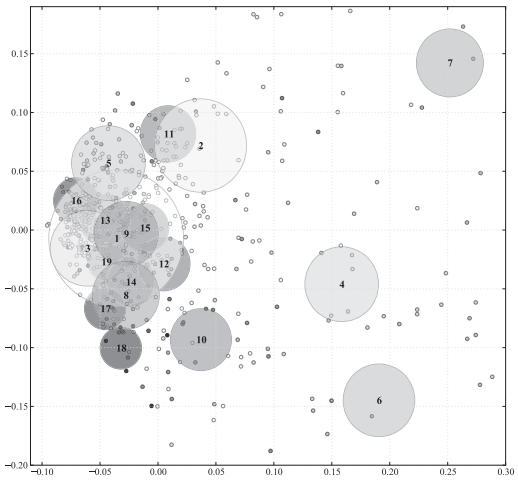

Figure 1. 19 clusters obtained with hierarchical clustering and hybrid pruning.

mean rank of the tag across the TDM is calculated with:

$$\bar{r}_j = \frac{\sum_{i=1}^{m} x_{ij}}{m} \quad (3)$$

Thus the cluster profile or mean ranks vector is defined as:

$$\mathbf{p}_l = \bar{r}_{j \in C_l} \quad (4)$$

C_l denotes a given cluster l for $1 \leq l \leq 19$ (optimal number of clusters for this dataset), and $\mathbf{p}$ is a vector $\{5, ..., k\}$, where $5 \leq k \leq 334$.

Last step aims to obtain ranked lists of songs ordered in terms of its closeness to each cluster profile. This is carried out by calculating the euclidean distance between each song rank vector $x_{i,j \in C_l}$ and the cluster profile $\mathbf{p}_l$:

$$d_i = \sqrt{\sum_{j \in C_l} (x_{ij} - \mathbf{p}_l)^2} \quad (5)$$

The examples of the results can be seen in Table 2, where top artists of each cluster are displayed below central tags of the cluster.

3. EXPERIMENT

In order to explore whether the obtained clusters are perceptually meaningful and to further understand what kinds of acoustic and musical attributes they consist of, empirical data unrelated to the existing structures about the clusters is needed. A similarity rating experiment was designed to assess the timbral qualities of songs pertaining to each of the clusters. We chose to emphasize the low-level, non-structural qualities of music since we wanted to minimize the confounding factors caused by recognition of songs, artists and the subsequent associations with these as well as the lyrical contents of the music. To this end, the stimuli for the experiment consisted of semi-randomly spliced, brief excerpts, explained in detail below.

3.1 Experiment details

3.1.1 Stimuli

Initially, 5-second audio samples were taken from a random middle part (25% after the beginning and 25% before the end) of the 25 top ranked songs (see ranking procedure in section 2.5) from each cluster. For each sample, the temporal position of notes onsets were estimated based on *spectral flux* using MIRToolbox [16]. The highest onset was selected as a reference point from which slices of random length ($150ms \leq t \leq 250ms$) were taken from $10ms$ before the peak onset of each sample, then equalized in loudness, and finally mixed together using a fade in-out of $50ms$ with an overlap window of $100ms$ This resulted in 19 excerpts (each representing a cluster) of variable length, that were finally trimmed to $1750ms$, with a fade in-out of $100ms$ To prepare these 19 excerpts for a similarity rating, the 171 paired combinations were mixed with a silence of $600ms$. between them.

3.1.2 Participants

12 females and 9 males (age M=26.8, SD=4.15) participated to the experiment. 9 of them possessed least one year of musical training. 12 reported listening to music attentively between one and 10 hours per week.

Cluster ID	*Tags proximate to cluster centroids*	Top artists in the cluster
1	*Energetic, Female vocal, Powerful, Hot, Sex*	Amy Adams, Fred Astaire, Kelly Clarkson
2	*Dreamy, Chill out, Haunting, Sleep, Moody*	Nick Drake, Radiohead, Massive Attack
3	*Sardonic, Sarcastic, Cynical, Humorous, Funny*	Alabama 3, Yann Tiersen, Tom Waits
4	*Awesome, Amazing, Male vocalist, Loved, Great*	Guns N' Roses, U2, Metallica
5	*Composer, Cello, Piano, Cello rock, Violin*	Camille Saint-Saëns, Tarja Turunen, Franz Schubert
6	*Female vocalist, Female vocalists, Female, 00s, Sexy*	Fergie, Lily Allen, Amy Winehouse
7	*Mellow, Beautiful, Chillout, Chill, Sad*	Katie Melua, Phil Collins, Coldplay
8	*Hard, Angry, Loud, Aggressive, Rock out*	System of a Down, Black Sabbath, Metallica
9	*60s, 70s, Guitar virtuoso, Sixties, Guitar solo*	Simon & Garfunkel, Janis Joplin, The Four Tops
10	*Feelgood, Summer, Feel good, Cheerful, Gute laune*	Mika, Goo Goo Dolls, Shekinah Glory Ministry
11	*Autumnal, Wistful, Intimate, Sophisticated, Reflective*	Soulsavers, Feist, Leonard Cohen
12	*High school, 90's, 1990s, 1995, 1996*	Fool's Garden, The Cardigans, No Doubt
13	*50s, Saxophone, Trumpet, Tenor sax, Sax*	Miles Davis, Thelonious Monk, Charles Mingus
14	*1980s, 80's, Eighties, 80er, Voci maschili*	Ray Parker Jr., Alphaville, Michael Jackson
15	*Affirming, Lyricism, Life song, Vocalization*	Lisa Stansfield, KT Tunstall, Katie Melua
16	*Choral, A capella, Acapella, Choir, A cappella*	Mediæval Bæbes, Alison Krauss, Blackmore's Night
17	*Voce femminile, Femmina, Voci femminili, Femmine*	Avril Lavigne, The Cranberries, Diana Krall
18	*Tangy, Coy, Sleek, Attitude, Flirty*	Kylie Minogue, Ace of Base, Solange
19	*Rousing, Exuberant, Confident, Playful, Passionate*	James Brown, Does It Offend You, Yeah?, Tchaikovsky

Table 2. Most representative tags and typical artists of each of the 19 clusters.

3.1.3 Procedure

Participants were presented with pairs of sound excerpts in random order using a computer interface and high-quality headphones. Their task was to rate the similarity of sounds on a 9-level Likert scale, whose extremes were labeled as *dissimilar* and *similar*. Before the actual experimental trials, they were given instructions and practice trials to familiarize themselves with the task.

3.1.4 Audio features

To explore the acoustic and musical features underlying the perceptual similarities of the clusters, 41 audio features (listed on Table 3) were extracted from each spliced stimuli using MIR toolbox [16]. The choice of features was restricted to those which would be applicable to spliced examples and would not require high-level feature analysis such as structural repetition or tonality. The extraction was carried out using frame-based approach with $50ms$ analysis frame using 50% overlap.

Category	No.	Feature
Dynamics	1-2	RMS energy
	3-4	Attack time (M, SD)
Rhythm	5-6	Fluctuation peak pos. (M, SD)
	7	Fluctuation centroid (M, SD)
Pitch	8-9	Pitch (M, SD)
	10-11	Chromagram (unwr.) centr. (M, SD)
Harmony	12	Entropy (oct. collap. spectr.) (M)
	13	Roughness (M)
	14	Inharmonicity (M, SD)
Timbre	15-16	Brightness (cut-off 110 Hz) (M, SD)
	17-18	Spectral centroid (M, SD)
	19-20	Zerocross (M, SD)
	20-21	Spread (M)
	22	Spectral entropy (M)
	23	Spectral flux (M)
	24	Flatness (M)
	25	Kurtosis (M)
	26-27	Regularity (M, SD)
	28-29	1st MFCC (M, SD)
	⋮	⋮
	30-41	7th MFCC (M, SD)

Table 3. List of extracted audio features (M= mean, SD= standard deviation)

3.2 Results

Highly consistent pattern of similarities between the 21 participants were obtained (Cronbach $\alpha = 0.94$). For this reason, a mean similarity matrix of the individual ratings was subjected to metric multidimensional scaling (MDS) analysis based on stress minimization by means of majorization (SMACOF) [4]. This yielded adequate low - dimensional projections of the data, from which we focus on 2 - dimensional (stress=0.065) and 3 - dimensional (stress=0.027) solutions.

The organization of the clusters (represented with sliced samples) illustrates a clear organization in terms of the semantic qualities of the clusters (see Figure 2), showing the *Awesome* and *Hard* examples on the left uppermost corner, and the semantically distant, *Autumnal* and *Dreamy* in the lower right-hand corner.

To investigate the perceived organization of the semantic clusters in terms of the acoustic qualities, the 3 dimensions were correlated with the extracted audio features. Highly significant correlations, top five shown in Table 4, were observed for dimensions 1 and 2. We may interpret these correlations in terms of the qualities of the sound spectrum: The first dimension is related to the distribution of energy along the frequency (spectral centroid, flatness, brightness, MFCC1, etc.), where the items in the MDS solution are arranged from the high-frequency energy content in the left to the prevalence of low-frequency energy content in the right. The second dimension may be interpreted as the periodic organization of the spectrum, i.e., whether the spectrum is harmonic (roughness,skewness, spread and fluctuation centroid). The clusters represented by the items in the lower part of the MDS solution possess clearer organization of the spectrum in comparison with the the items high on the MDS solution. The third dimension seem to be related the temporal fluctuation of the spectrum (MFCC6 [SD], Fluctuation position [M], MFCC22 [M]).

Dimension 1			Dimension 2			Dimension 3		
Acoustic feature	***r***		**Acoustic feature**	***r***		**Acoustic feature**	***r***	
MFCC 1 (M)	0.94	***	Fluctuation centroid (M)	-0.72	***	MFCC 6 (SD)	0.51	*
Flatness (M)	-0.86	***	Roughness (M)	0.68	**	Fluctuation position (M)	-0.50	*
Centroid (M)	-0.83	***	Skewness (M)	0.67	**	MFCC 2 (M)	-0.46	*
Brightness (M)	-0.81	***	Spread (M)	-0.65	**	Fluctuation peak (M)	0.45	
Spectral entropy (M)	-0.80	***	Kurtosis (M)	0.57	*	Irregularity (SD)	0.44	

$*** = p < .001$, $** = p < .01$, $* = p < .05$

Table 4. Correlations between the dimensions of the multidimensional scaling solution and acoustic descriptors.

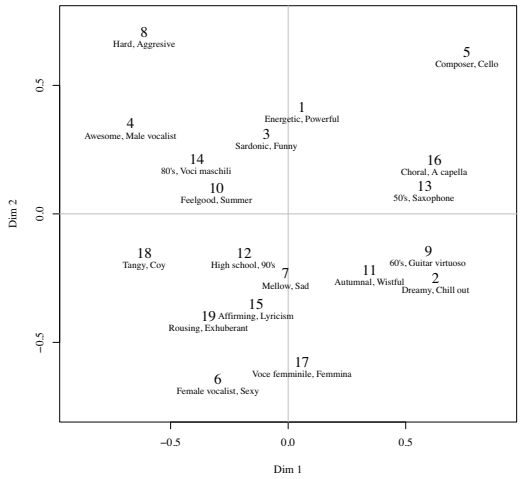

Figure 2. Dimensions 1 and 2 of the MDS with behavioural responses and associated tags

3.3 Discussion

In sum, when brief and spliced excerpts taken from the clusters representing semantic structures of the music descriptions are presented to listeners, they are able to form coherent distances between them. An acoustic analysis of the excerpts was used to label the dimensions embedded in the cluster similarities. This analysis showed clear correlations between the dimensional and timbral qualities of music. However, it should be emphasized that the high relevance of many timbral features is only natural since the timbral characteristics of the excerpts were preserved and structural aspects were masked by the semi-random splicing.

We are careful in not taking these early results to mean literally that the semantic structure of the initial sample would be explainable by means of the same timbral features. This is of course another question which is easily empirically approached using feature extraction of the typical examples representing each cluster and either classify the clusters based on features, or predict the coordinates of the clusters within a low dimensional space by means of regression using a larger set of acoustic features (including those that are relevant for full excerpts such as tonality and structure). However, we are positively surprised at the level of coherence from the part of the listener ratings and their explanations in terms of the acoustic features despite the limitations we imposed on the setting (i.e. discarding tags connected with musical genres), splicing and having a large number of clusters to test. Our intention is to follow this analysis with more rigorous selection of acoustic features (PCA and other data reduction techniques) and use multiple regression to assess whether linear combinations of the features would be necessary for explaining the perceptual dimensions.

4. CONCLUSIONS

The present work provided a bottom-up approach to semantic qualities of music descriptions, which capitalized social media, natural language processing, similarity ratings and acoustic analysis. Semantic structures of music descriptions have been extracted from the social media previously [18] but the main difference here was the careful filtering of such data. We used natural language processing to focus on categories of tags that are meaningful but do not afford immediate categorization of music in a way that, for example, musical genre does.

Although considerable effort was spent on finding the optimal way of teasing out reliable and robust structures of the tag occurrences using cluster analysis, several other techniques and parameters within clustering could also have been employed. We realize that other techniques would probably have led to different structures but it is an open empirical question whether the connections between the similarities of the tested items and their acoustic features would have been entirely different. A natural continuation of the current study would be to predict the typical examples of the clusters with the acoustic features by using either classification algorithms or mapping of the cluster locations within a low dimensional space using correlation and multiple regression. However, the issue at stake here was the connection of timbral qualities with semantic structures.

The implications of the present findings are related to several open issues. The first one is the question whether structural aspects of music are required in explaining the semantic structures or whether the low-level, timbral characteristics are sufficient, as was indicated by the present findings. Secondly, what new semantic layers (as indicated by categories of tags) can be meaningfully connected with the acoustic properties of the music? Finally, if the timbral

characteristics are indeed strongly connected with such semantic layers as *adjectives*, *nouns* and *verbs*, do these arise by means of learning and associations, or are the underlying regularities connected with emotional, functional and gestural cues of the sounds?

5. REFERENCES

[1] J.J. Aucouturier and E. Pampalk. Introduction-from genres to tags: A little epistemology of music information retrieval research. *Journal of New Music Research*, 37(2):87–92, 2008.

[2] J. Brank, M. Grobelnik, and D. Mladenic. Automatic evaluation of ontologies. In Anne Kao and Stephen R.Poteet, editors, *Natural Language Processing and Text Mining*. Springer, USA, 2007.

[3] O. Celma and X. Serra. Foafing the music: Bridging the semantic gap in music recommendation. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):250–256, 2008.

[4] J. de Leeuw and P. Mair. Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, 31(3):1–30, 2009.

[5] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.

[6] M.J. Delsing, T.F. ter Bogt, R.C. Engels, and W.H. Meeus. Adolescents music preferences and personality characteristics. *European Journal of Personality*, 22(2):109–130, 2008.

[7] T. Eerola and R. Ferrer. Setting the standards: Normative data on audio-based musical features for musical genres. In *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music, ESCOM*, 2009.

[8] Christiane Fellbaum, editor. *WordNet: An electronic lexical database.* Language, speech, and communication. MIT Press, Cambridge, Mass, 1998.

[9] W.N. Francis and H. Kucera. *Brown corpus. A Standard Corpus of Present-Day Edited American English, for use with Digital Computers*. Department of Linguistics, Brown University, Providence, Rhode Island, USA, 1979.

[10] J.C. Gower and P. Legendre. Metric and euclidean properties of dissimilarity coefficients. *Journal of classification*, 3(1):5–48, 1986.

[11] J.M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5):1270–1277, 1977.

[12] A.K. Jain and R.C. Dubes. *Algorithms for clustering data*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[13] Anne Kao and Stephen R. Poteet, editors. *Natural Language Processing and Text Mining*. Springer Verlag, 2006.

[14] P. Lamere. Social tagging and music information retrieval. *Journal of New Music Research*, 37(2):101–114, 2008.

[15] P. Langfelder, B. Zhang, and S. Horvath. *dynamicTreeCut: Methods for detection of clusters in hierarchical clustering dendrograms.*, 2009. R package version 1.20.

[16] O. Lartillot, P. Toiviainen, and T. Eerola. A matlab toolbox for music information retrieval. *Data Aalysis, Machine Learning and Applications*, pages 261–8, 2008.

[17] C. Laurier, M. Sordo, J. Serra, and P. Herrera. Music mood representation from social tags. In *Proceedings of the 10th International Society for Music Information Conference, Kobe, Japan*, 2009.

[18] M. Levy and M. Sandler. Learning latent semantic models for music from social tags. *Journal of New Music Research*, 37(2):137–150, 2008.

[19] H. Lin, J. Davis, and Y. Zhou. An integrated approach to extracting ontological structures from folksonomies. In *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*, page 668. Springer, 2009.

[20] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes. *Psychological Research*, 58(3):177–192, 1995.

[21] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.

[22] P.J. Rentfrow and S.D. Gosling. Message in a ballad: the role of music preferences in interpersonal perception. *Psychol Sci*, 17(3):236–242, 2006.

[23] J.M. Siskind. Learning word-to-meaning mappings. *Models of language acquisition: inductive and deductive approaches*, pages 121–153, 2000.

[24] B. Zhang, Q. Xiang, H. Lu, J. Shen, and Y. Wang. Comprehensive query-dependent fusion using regression-on-folksonomies: a case study of multimodal music search. In *Proceedings of the seventeen ACM international conference on Multimedia*, pages 213–222. ACM, 2009.