Predicting Musical Meaning in Audio Branding Scenarios

Martin Herzog[*1], Steffen Lepa[*2], Jochen Steffens[*3], Andreas Schoenrock[*4], Hauke Egermann[#5]

[*] *Audio Communication Group, Technische Universität Berlin, Germany*

[#] *York Music Psychology Group, University of York, UK*

[1]herzog@tu-berlin.de, [2]steffen.lepa@tu-berlin.de, [3]jochen.steffens@tu-berlin.de, [4]andreas.schoenrock@tu-berlin.de, [5]hauke.egermann@york.ac.uk

ABSTRACT

This paper describes the concept of applying automatic music recommendation to the audio branding domain. We describe our approach of developing a prediction model for the perceived expressive content of music which is based on a large-scale listening experiment. We present an orthogonal 4-factor model for measuring musical expression as outcome variable, whereas audio- and music features as well as lyric-based features are introduced as prediction variables in the model. Furthermore, we describe Random Forest Regression as a concept for feature selection required to develop a Multi-Level Regression Model, which is taking individual listener parameters into account. Finally, we present first results from a preliminary stepwise regression model for perceived musical expression.

Keywords: Music Branding, Audio Branding, Music Recommendation, Musical Semantics, Prediction Model, Random Forest Regression, GMBI, Music Information Retrieval

I. INTRODUCTION

In the field of Audio Branding, companies become more and more interested in systems for automated music recommendation. In this type of application, suitable music pieces are automatically selected from a large music archive and subsequently presented to consumers in order to communicate specific expressions. These expressions shall then contribute to a strategically-planned brand image perceived by the recipients (Müllensiefen & Baker, 2015). Operational scenarios include marketing activities like point of sale background music, music on websites or music used in audiovisual advertisements.

A significant amount of research has already been carried out to investigate the correlations between musical attributes on one side and *emotional qualities* on the other (Schmidt et al., 2012; Song et al., 2012; Yang & Chen, 2012). Algorithmic tools employing this knowledge already help private music enthusiasts to navigate through nowadays' endless digital music archives and to let them discover new titles and artists. Thus, through algorithmic emotion-based recommendation, music can unfold its functionality of mood-management, social-bonding and distinction, identity formation or any other kind of ritual affect-laden everyday use (Schäfer et al., 2013).

Our project, however, investigates the associative *semantic meaning* of music for listeners. The aim of the presented study is therefore to test the feasibility of predicting the music-induced activation of branding relevant semantic associations. In order to achieve this, we aim to find statistical prediction models for brand attributes (such as 'young', 'urban', 'trustworthy' or 'playful') based on a variety of low- and high-level audio and music features and based on the moderating influence of inter-individual differences of groups of music listeners. Based on this, we aim on developing a prototype system for automatic music recommendation within the audio branding domain. To initially conceptualize this scenario, music branding can be interpreted as a special case of sign-based communication. An adapted version of Egon Brunswik's 'lense model' (Brunswik, 1955) exemplifies this approach (see figure 1).

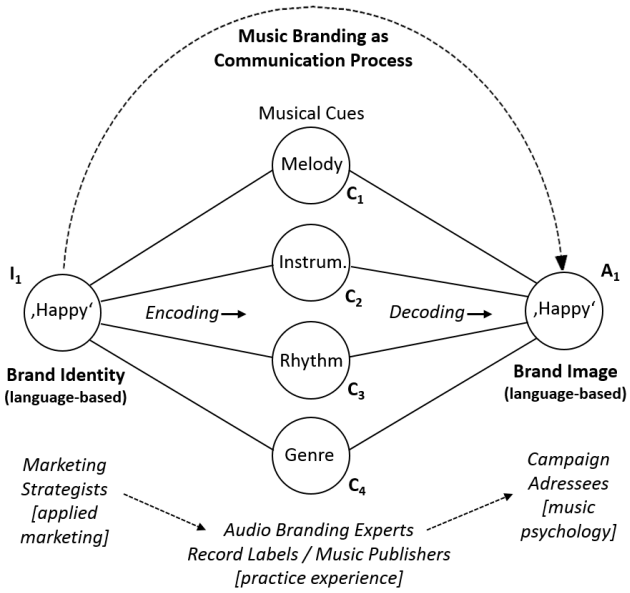

Figure 1. Music Branding as Communication Process

II. LISTENING EXPERIMENT

To find statistical determinants for perceived semantic expression of music, we conducted a large-scale online listening experiment to systematically gather ratings on the musical expression perceived from a larger number of different music titles. Therefore n = 3.485 participants were recruited from three different countries (UK, Spain, Germany), three different age cohorts (18-34; 35-51; 52-68), three different educational backgrounds (ISCED 0-2; 3-4; 5-8), and both genders using countrywise crossed-quotas.

The music corpus for this experiment consisted of 183 music excerpts, representing 61 different music styles grouped into 10 different genres. After reporting socio-demographics and performing a listening test to calibrate their audio output volume, subjects were asked to listen to a set of four randomly assigned excerpts with a duration of

approximately 30 seconds each (typically comprising a part of a verse and chorus). A randomized title selection for single respondents was carried out in a systematic way, ensuring that each title would receive the same amount of ratings from each consumer cluster. After each stimulus, participants had to rate the fit between the excerpt and each item of the preliminary General Music Branding Inventory (GMBI). The GMBI is a new psychometric instrument for assessing the music-induced association of attributes, which are frequently and reliably used in the field of music branding (Steffens et al., 2017). It consists of 51 attributes that were rated using a Likert scale from 1 ("very bad fit") to 6 ("very good fit"). For each stimulus, respondents should also indicate how well they knew the excerpt and how much they liked it. Finally, the participants also reported their degree of focus throughout the experiment as well as their genre preferences, degree of musicality, and the audio setup used for the experiment.

A second iteration of this multinational listening experiment will be carried out in 2017 with 6.000 participants. It is aiming on cross-validating the present results and enlarging the training data set.

III. PREDICTING MUSICAL MEANING

This section describes our approach of predicting the perceived expressive content of popular music based on the comprehensive empirical ground truth resulting from our listening experiment. We discuss the components required to build a statistical model for the prediction of music-induced semantic associations, highlight the major challenges for this task and present first results.

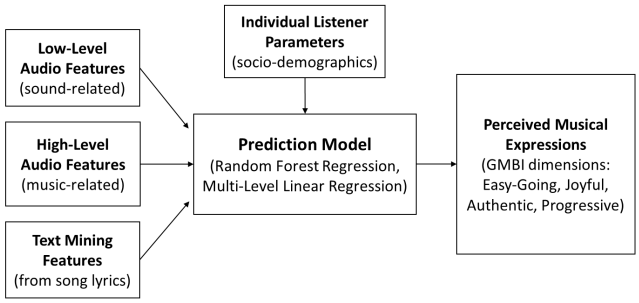

Figure 2. Prediction model and components overview

A. Perceived Musical Expressions

The underlying theoretical idea of our prediction approach is a parametric orthogonal feature space of semantic musical expression. Every conceivable musical piece should have its own location in this multi-dimensional space, predictable based on the original ground truth data from our listening experiment. Taking the GMBI fit ratings from the experiment and performing exploratory factor analysis, the orthogonal dimensions ("Easy-Going", "Joyful", "Authentic", and "Progressive") of this space were developed and further refined by employing Exploratory Structural Equation Modeling (Asparouhov & Muthén, 2009), drawing on so-called *orthogonal target rotations* in order to arrive at PCA-like orthogonal so-called *ESEM factors* (see Bonneville-Roussy et al., 2013 and Lepa & Seifert, 2015 for applied examples of ESEM).

Table 1. Orthogonal ESEM factor loading matrix for GMBI (loadings > 0.5 set in bold)

Item/Factor	Easy-Going	Joyful	Authentic	Progressive
confident	0.141	0.481	0.486	0.202
loving	**0.647**	0.312	0.346	0.057
friendly	0.483	**0.608**	0.248	0.063
honest	0.412	0.370	**0.549**	0.060
trustworthy	0.475	0.361	**0.517**	0.109
happy	0.197	**0.750**	0.161	0.137
beautiful	**0.570**	0.363	0.454	0.123
soft	**0.798**	0.100	0.173	0.053
warm	**0.632**	0.407	0.323	0.008
bright	0.323	**0.530**	0.330	0.203
stimulating	0.212	**0.551**	0.449	0.270
relaxing	**0.783**	0.126	0.258	0.074
chilled	**0.657**	0.122	0.186	0.174
detailed	0.293	0.281	**0.582**	0.238
simple	0.386	0.197	0.098	0.072
pure	0.497	0.282	**0.511**	0.108
unique	0.202	0.273	**0.561**	0.280
reflective	**0.506**	0.116	**0.516**	0.213
intellectual	0.373	0.099	**0.596**	0.239
modern	0.149	0.242	0.049	**0.770**
classic	0.359	0.080	**0.547**	-0.177
young	0.126	0.377	0.017	**0.664**
innovative	0.200	0.280	0.431	**0.544**
solid	0.298	0.327	**0.548**	0.150
fresh	0.273	**0.543**	0.279	0.397
inviting	0.435	**0.555**	0.397	0.176
integrating	0.352	0.406	0.473	0.225
adventurous	0.038	0.424	0.485	0.370
familiar	0.397	0.351	0.416	0.042
serious	0.261	-0.071	**0.564**	0.174
playful	0.152	**0.601**	0.202	0.213
funny	0.099	**0.511**	0.218	0.258
male	-0.085	0.140	0.333	0.109
female	0.382	0.186	0.064	0.150
passionate	0.297	0.420	**0.534**	0.101
sexy	0.322	0.404	0.307	0.306
epic	0.245	0.163	**0.597**	0.258
personal	0.412	0.266	**0.520**	0.161
inspiring	0.398	0.398	**0.545**	0.241
creative	0.214	0.410	**0.506**	0.340
magical	0.426	0.264	0.496	0.267
exciting	0.113	**0.511**	**0.502**	0.312
futuristic	0.076	0.048	0.176	**0.705**
retro	0.164	0.173	0.375	-0.108
timeless	0.400	0.287	**0.541**	-0.008
contemporary	0.268	0.243	0.210	**0.542**
urban	0.067	0.214	0.183	**0.517**
natural	**0.521**	0.353	0.435	0.007
authentic	0.288	0.406	**0.571**	0.074
glamorous	0.381	0.265	0.421	0.253
cool	0.222	0.462	0.355	0.420

When applied to the listening experiment data our approach led to a well-fitting orthogonal ESEM solution (X^2=22510.842; df=1077; p<0.01; RMSEA=.039; CFI=.925; SRMR=.026), which draws on all original 51 GMBI items.

B. Predictors for Perceived Musical Expressions

For predicting perceived musical expression in terms of the 4 developed dimensions, our project will draw on three different variable groups. These groups are: low-level audio features, high-level audio features and text mining features (figure 2). Although this work is still in progress, we deem it worthy to discuss our ideas about suitable predictors.

1) Low-Level Audio Features. This category of features comprises on one hand recording-related features such as stereo spread and beats per minute of a song. On the other hand it contains sound-related audio features describing loudness, roughness and sharpness of a musical stimulus. Although not yet researched in depth or analyzed by music branding practitioners, these sound-related audio features may play a significant role in predicting the musical expression perceived by consumers. Our analysis draws on such features extracted from the IRCAM Timbre Toolbox (Peeters et al., 2011) and similar software packages. Within our research group, additional new features will be developed by refining and combining low-level features to new complex ones.

2) High-Level Audio Features. This set of features is based on audio signal analysis as well, but in contrast to low-level audio features, it is drawing on various concepts from music theory. These features are typically deemed highly relevant by music branding experts for selecting suitable pieces of music. Within our project we (inter alia) employ the software packages IRCAMBEAT and IRCAMSUMMARY (Kaiser & Peeters, 2013; Peeters & Papadopoulos, 2011) to either extract these features directly or their fundamental data structures. Based on that, musical features requiring a higher level of abstraction are developed by our team, e.g. specific types of melody successions and chord progressions.

Finally, Machine Learning is employed (by our project partners at IRCAM) to automatically classify songs in terms of genre, style, instrumentation, intensity, and further high-level features. The ground truth for these predictors comes from 9428 titles from the HEARDIS music library which were tagged by music branding experts from the company.

3) Text Mining Features. Most pieces in our music sample contain song lyrics which are providing an additional source of the perceived semantic expressions of music titles as reflected in our ground truth data. Therefore, we plan to also take lyric-based predictors into account in our modeling approach. However, this idea entails a new challenge, since feature extraction schemes for e.g. emotional labeling of lyrics are non-trivial and a research subject in itself (Kim et al., 2010). In a first step, we will carry out a benchmark of existing text mining tools such as Synesketch (Krcadinac et al. 2013), Word2Vec-networks (Wolf et al., 2014) and the IBM Tone Analyzer ("IBM Corp. Tone Analyzer.") regarding their potential for explaining another portion of the perceived expression of (text-based) music. The Tone Analyzer for example uses linguistic analyses to detect basic emotions (*happiness, sadness, anger, fear, disgust,* and *surprise*) which might constitute strong predictors of perceived semantic expression of music. To exploit this information, we extracted the lyrics contained in our sample from the Music Lyrics Database (MLDb).

C. Individual Listener Parameters

Members of different social milieus and generations tend to attribute different semantic meanings to the very same musical pieces (Bonneville-Roussy et al., 2013). In our approach of predicting the perceived expressive content of music, we therefore also address the challenge of inter-individual differences in the association to music. Our gathered ground truth data contains information about listeners' countries, different age cohorts, different educational backgrounds, and both genders. In the second iteration of our listening experiment we will also draw on the so-called SINUS-Milieus which are deemed relevant by marketing practitioners to identify and address relevant target groups. We will test if membership in these consumer clusters would produce significantly different perceptions of musical expressions. The moderating influence of these individual listener parameters will then be tested in a Hierarchical Linear Regression Model.

D. Modelling Approach

Our aim is to combine all gathered ground truth data in a regression model predicting each music title's position in the described feature space. However, this leads to the challenge of feature selection, which needs to be addressed in order to handle the complexity, redundancy and huge amount of possible predictors. Therefore, we will conduct *Random Forest Regression* for each orthogonal dimension of our musical expression feature space.

In random forest regression a large number of decision trees is used, which are grown independently in order to predict the outcome variable. For each tree, the number of predictor variables is limited to a small subset of the available explanatory variables. Furthermore, only a random subset of the ground truth data is used for each individual tree (Pawley & Müllensiefen, 2012). Calculating the relative rank of each predictor compared across all trees will lead to a Monte-Carlo-like approach to identify the best set of predictors. Thus, we avoid facing typical regression problems like multi-collinearity and interaction complexity. From the many decision trees grown within a random forest, the average level of hierarchy is determined for all available explanatory variables. This will provide the best subset of predictor variables which can then be turned into a classical regression model accordingly.

The second challenge for our modeling approach is the different consumer groups as described in the listening experiment (see II). We expect the need of differential regression parameters for these groups as found e.g. by Chamorro-Premuzic et al. (2010). Therefore, we will extend the regression model to a multivariate multi-level regression model (Hox, 2010) with random effect parameters for social milieus. In this way, different regression parameters can be used for different subject clusters. Additionally, a 'fixed' mean effect is estimated which can be used in music branding scenarios where no specific target group parameters are available. Exploiting existing data from our first online listening experiment, we will use socio-demographics (gender, birth-cohorts and

education) as cluster variables. Ground truth from our second study in 2017 will allow us to also draw on the so-called SINUS-Milieus allowing for grouping people according to their lifestyle and values.

E. First results

Since random forest regression is still in progress, we developed a preliminary general prediction model for each of the four factors (*Easy-Going, Joyful, Authentic,* and *Progressive*) based on our current set of low-level and high-level audio features and for all target groups. To address correlation between predictors we used stepwise regression (PIN=.05, POUT=.10). Table 2 depicts the key characteristics of each individual model.

Table 2. Preliminary stepwise regression models for orthogonal musical expressions

Model	Pred. included	R^2 (adjusted)	p	df
Easy-Going	31	.25	< .001	12980
Joyful	24	.13	< .001	12987
Authentic	32	.15	< .001	12979
Progressive	25	.22	< .001	12988

The four models are based on a first preliminary set of 118 predictor variables in total, not yet containing high-level music descriptors such as melody and harmony features. The column "Pred. included" gives the number of different variables employed in each model, whereas Table 3 depicts the 10 most influential predictors per model.

Table 3. Overview of most influential predictors per model

	Easy-Going		**Joyful**	
No	Feature	R^2	Feature	R^2
1	intensity	.08	pop appeal	.02
2	female vocals	.09	genre Classical	.04
3	speed	.10	Intensity	.05
4	style Traditional-Folk	.11	style Rock & Roll	.06
5	style Hip-Hop	.13	style Folkloric	.08
6	style Punk	.14	genre Soul/Funk	.08
7	style Balearic	.14	country Canada	.09
8	style Funk	.15	genre Hip-Hop	.09
9	style Downbeat	.16	speed	.10
10	style Rock & Roll	.17	style Oriental	.10
	Authentic		**Progressive**	
No	Feature	R^2	Feature	R^2
1	complexity	.05	genre Dance	.08
2	style Reggaeton	.06	publishing year	.12
3	style Hist. Classical	.07	complexity	.14
4	style EDM	.09	speed	.15
5	female vocals	.10	style Dubstep	.15
6	publishing year	.10	style Balearic	.16
7	style Dancehall	.11	style Hip-Hop	.17
8	style Flamenco	.11	style Indie-Dance	.17
9	style Reggae	.11	style Dancehall	.17
10	style Rare-Groove	.12	style Cont. Classical	.18

F. Discussion

Our preliminary results indicate that especially high-level features such as *intensity*, *complexity* and *pop appeal* of a song as well as information about *style* and *genre* contribute most to the perception of all the four expression dimensions. These are still based on annotations from audio branding experts and will later on be substituted by machine-learning data. On the other hand, only *speed* of a track could be identified as a relevant low-level feature for the perceived expressions. A reason for the high influence of high-level features might lie in their complexity: They usually express a variety of lower-order audio features which are in this way aggregated to one semantical concept such as *intensity*.

A similar 'wholistic' function applies to genres and styles which are in addition associated with cultural influences. However, we expect additional predictive power from music features such as rhythmic styles, melody and harmony progressions to be incorporated in the next stage of our modelling approach.

IV. CONCLUSION AND OUTLOOK

Our contribution illustrates the concept of applying automatic music recommendation to the audio branding domain. It describes results derived from a large-scale online listening experiment and introduces our approach for predicting perceived musical expressions. It includes the application of Music Information Retrieval, Machine Learning, Structural Equation Modeling and Random Forest Regression techniques to adequately model brand-music-consumer relationships. Moreover, our work presents first results from four preliminary regression models.

A second large-scale listening experiment will be carried out in 2017. Eventually, we will integrate all developed high-level and low-level audio features as well as lyric-based features in one comprehensive multi-level regression model. Comparing their predictive power will also shed a light on the question which realm is more dominant in conveying musical expression perceived by listeners.

Concluding, we think our work offers a unique and innovative approach towards semantic music analysis which is applicable far beyond the field of audio branding.

ACKNOWLEDGEMENT

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 688122.

REFERENCES

Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology*, *105*(4), 703–717. https://doi.org/10.1037/a0033770

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/37/h0047470

Chamorro-Premuzic, T., Fagan, P., & Furnham, A. (2010). Personality and uses of music as predictors of preferences for music consensually classified as happy, sad, complex, and social. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4), 205–213.

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications*. New York: Routledge.

IBM Corp. (2017). Tone Analyzer. (n.d.). Retrieved April 24, 2017, from https://www.ibm.com/watson/developercloud/doc/tone-analyzer/index.html

Kaiser, F., & Peeters, G. (2013). A simple fusion method of state and sequence segmentation for music structure discovery. In *ISMIR (International Society for Music Information Retrieval)* (p. -). NA, France. Retrieved from https://hal.archives-ouvertes.fr/hal-01106873

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, O. G., Richardson, P., Scott, J., … Turnbull, D. (2010). Emotion Recognition: a State of the Art Review. In *11th International Society for Music Information and Retrieval Conference*.

Krcadinac, U., Pasquier, P., Jovanovic, J., & Devedzic, V. (2013). Synesketch: An Open Source Library for Sentence-Based Emotion Recognition. *IEEE Transactions on Affective Computing*, *4*(3), 312–325. https://doi.org/10.1109/T-AFFC.2013.18

Lepa, S., & Seifert, M. (2015). Musikalische Vorlieben oder Alltagsästhetische Schemata? Zur relativen Bedeutung von Demographie-, Sozialisations- und Persönlichkeitsvariablen für die Optimierung digitaler Musikempfehlungssysteme. *Jahrbuch Der Deutschen Gesellschaft Für Musikpsychologie*, *25*, 116–141.

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *The American Psychologist*, *52*(5), 509–516.

Müllensiefen, D., & Baker, D. J. (2015). Music, Brands, & Advertising: Testing What Works. In K. Bronner, C. Ringe, & R. Hirt (Eds.), *Audio Branding Acaddemy Yearbook 2014/2015* (pp. 31–51). Baden-Baden: Nomos.

Pawley, A., & Müllensiefen, D. (2012). The science of singing along: A quantitative field study on sing-along behavior in the north of England. *Music Perception: An Interdisciplinary Journal*, *30*(2), 129–146.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, *130*(5), 2902–2916. https://doi.org/10.1121/1.3642604

Peeters, G., & Papadopoulos, H. (2011). Simultaneous Beat and Downbeat-Tracking Using a Probabilistic Framework: Theory and Large-Scale Evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(6), 1754–1769. https://doi.org/10.1109/TASL.2010.2098869

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00511

Schmidt, E. M., Scott, J. J., & Kim, Y. E. (2012). Feature Learning in Dynamic Environments: Modeling the Acoustic Structure of Musical Emotion. In *ISMIR* (pp. 325–330). Citeseer.

Song, Y., Dixon, S., & Pearce, M. (2012). Evaluation of Musical Features for Emotion Classification. In *ISMIR* (pp. 523–528). Citeseer.

Steffens, J., Lepa, S., Egermann, H., Schönrock, A., & Herzog, M. (2017). Entwicklung eines Systems zur automatischen Musikempfehlung im Kontext des Audio Brandings. Paper presented at DAGA 2017 - Annual Conference of the German Acoustic Society, Kiel (Germany).

Wolf, L., Hanani, Y., Bar, K., & Dershowitz, N. (2014). Joint word2vec Networks for Bilingual Semantic Representations. *International Journal of Computational Linguistics and Applications*, *5*(1), 27–42.

Yang, Y.-H., & Chen, H. H. (2012). Machine recognition of music emotion: A review. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *3*(3), 40.