

INCREASING CONSUMER VALUE IN VIRTUAL MUSIC ENVIRONMENTS

[HTTP://VIRTUALGOODS.TU-ILMENAU.DE/2003/CONSUMERVALUE.PDF](http://virtualgoods.tu-ilmenu.de/2003/consumervalue.pdf)

STEPHAN BAUMANN & OLIVER HUMMEL

German Research Center for Artificial Intelligence

Erwin Schrödinger Str.

67663 Kaiserslautern, Germany

E-mail: stephan.baumann@dfki.de

Consumer value is an essential aspect for the success of commercial platforms dealing with virtual music. It leads to successful adoption by end users. New value propositions being enabled by technological innovations are possible. In this paper we present recent work in the field of music information retrieval (MIR) satisfying the requirements, which have been proposed in value-based modeling of MIR systems in e-commerce. Two different scenarios are described: a central server approach integrating phonetic matching, and a P2P-based recommendation system supporting fan and genre communities. Furthermore we describe a novel artist recommendation approach using cultural features, which also could be used in both scenarios.

1. Introduction

The digital content industry has to think about new value propositions for customers. The distribution of digital music is one of the most attracting and challenging topics for end users, content owners and musicians these days. In spite of the ongoing legal debates about consumer behavior and illegal file sharing services we find a lot of potential for convenient man-machine-interfaces to music on the technical side. The focus of this paper is the motivation how technological innovations enable new value propositions. After the presentation of a recent approach [1] to value-based modeling in e-commerce scenarios we will describe two such scenarios and their underlying technical solutions. First a central server approach offering music information to end-customers is described which is the largest German music portal satisfying 5 million page impressions per month [2]. Second a new P2P research framework handling essential features for future commercial scenarios in the area of P2P based music and community sharing is presented [3]. A novel approach about cultural recommendations which could be integrated in both scenarios is outlined at last. For all of these applications our approach offers a maximum of usability and a minimum amount of manual indexing of the underlying large-scale musical data. We subsumed

these objectives under the term *super-convenience* in our former work [4]. The paper is outlined as follows: we describe the general value-based modeling in chapter 2, we link the presented dimensions of the value propositions to the technical solutions in chapter 3, details about our experiences are concluded in chapter 4.

2. Value-based modeling in a world of virtual goods

The work described in [1] shows perspectives from a business point of view why our technological solutions make sense in a world of virtual goods, resp. virtual music. Value-based approaches to e-commerce offer a quantitative analysis how to cope with the problem of illegal file sharing. It seems to be possible in a reasonable way to convert special consumer segments of illegal P2P file sharing platforms to legal offers. Required ingredients are reduction of search time, superior quality of service and innovative interactive features. This can be reached if one enhances existing P2P networks with intelligent processing components such as automated similarity recommendations, semi-automated meta-tagging and user-friendly query processing. This holds also true for the upgrading of central server platforms with such technical features. Gordijn showed in his Ph.D thesis [1] that a quantification of e-commerce scenarios for virtual music is possible according to the following value equation:

$$\text{Consumer Value} = \frac{\sum \text{receipts}}{\sum \text{sacrifices}}$$

The sum of the receipts represents the benefits a user experiences while the sacrifices represent the total amount of costs to consume a product (e.g. to download a music track). The consumer only buys if the consumer value ratio is greater than 1. Obviously the data connection fee to be paid to the ISP and the fee to download a track are part of the sacrifices for a consumer. Further receipts and sacrifices are based on valuing the experiences during music download. As a starting point Gordijn introduces Holbrook's [5] framework for exploitation of the dimensions of different value types. If this framework is applied to music download in a listen-once scenario the following parameters result:

Value dimension	<i>Extrinsic</i>	<i>Intrinsic</i>
<i>Active</i>	EFFICIENCY Time: select, upload, download	PLAY Interactive track play
<i>Reactive</i>	EXCELLENCE Presentation quality	ESTHETICS Track beauty

Table 1: Value types according to Holbrook's framework & Gordijn's listen-once parameters

With these parameters in place quantification of consumer value for different end-user segments and download scenarios are possible. Gordijn did this kind of research for students and yuppies in legal and illegal variants of music download and came out with the following results:

- Selection time is the key for creating value gaps between illegal and legal offers
- Exploiting a short search time for legal content leads to low inconvenience fees and has long-term effects
- Track beauty has been hard to evaluate but we will show some technical options to increase this experience by providing similar artist recommendations based on audio and cultural features.

3. Examples for value-based music scenarios

3.1. Correction of phonetic misspellings in a music information portal

In this case study we implemented a phonetic matching for a music information portal together with a spin-off company. The approach was connected to the search field for artists operating on a database of approx. 70.000 artists. After implementation end users entries may contain typing errors and phonetic misspellings. The system is able to connect artist names which sound similar to each other, i.e. it is still able to produce results when there is phonetic similarity (such as e.g. „fil collins“ vs. „phil collins“). In contrast to other methods of non-exact search (such as Levenstein [6] or Soundex) our method is optimized to the musical domain.

The reduction of search time could be proven by repetitive evaluation of the web logs of the site. The convenience factor increased dramatically since manual reformulations of ill-formed queries are not necessary anymore. Furthermore traffic has increased by promotional activities emphasizing this outstanding feature. In figure 1 a typical example is presented.



Figure 1. Artist search with phonetic matching in www.musicline.de

3.2. Community features and audio analysis for a prototypical P2P platform

Peer-to-Peer Community Features

We will first make some remarks why a mainstream music service should follow a P2P paradigm in order to be successful. It is a very natural paradigm to process the content where it is, in this case at the private computers of the consumers. Distributed computing and storage power, bandwidth, fault-tolerance and reliability are the characteristics of such a network. In addition to this a P2P protocol offers the possibility to implement peer clustering by content to realize such issues as collaborative retrieval on top. This step offers the potential to have automatic and dynamic community building (e.g. special genre interest groups) over the time the network evolves. These features fall into both categories of the value type framework: efficiency = reduction of search time and esthetics = exploring the beauty of recommended, previously unknown but relevant new tracks.

JXTA was chosen as the platform for developing our integrated approach. A filesharing application with some basic features based on the JXTA protocol has been developed open source in the MyJXTA2 project. We used straightforward the following functionalities of the client to realise our own version MPeer:

- (1) Peer Groups: can be used to model different interest groups according to specific styles, genres or artist fan communities.
- (2) Group Chat, 1:1 Chat: supporting the verbal communication about preferred music has been one of the cornerstones in our design of an intelligent P2P music platform since we aim at extracting cultural metadata from chats in the future (see Fig. 3).
- (3) File Sharing: there is nothing special about this feature, but we realized that we would have a great benefit by adding our special additional tags (e.g. pre-computed audio profiles, extended metatags such as “similar artists”, etc.) into the existing MP3 file format which is supported at hand by its mime type.
- (4) Metatag representation: by using ID3v2 as part of the MP3 file format we were able to supply the interesting musical features using a well-established format, having a critical mass of users, resp. standard applications such as MP3 players, playlist and cataloguing tools to read and store this format. In this way simple filesharing enters a totally new dimension of quality since each MP3 file can be interpreted as a self-contained music knowledge container.
- (5) File Search: at this point we added again our component for a fuzzy match of phonetic misspelled queries to correct entities in the artist field. (see Fig.4).
- (6) Presentation of metadata attributes and values: after a computation of similar songs by using the music similarity metrics we added the according artists as a new meta tag. Furthermore standard ID3 tags such as artist and genre are presented (see Fig.2).

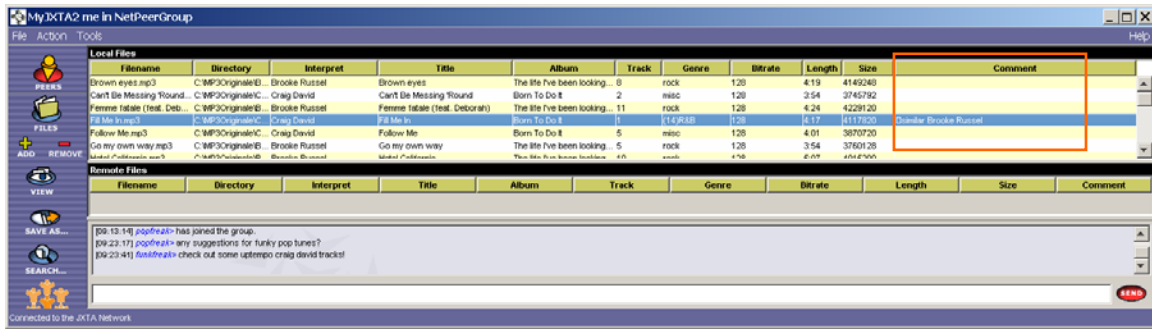


Figure 2. MPeer Client: Embedded recommendations (comment field)

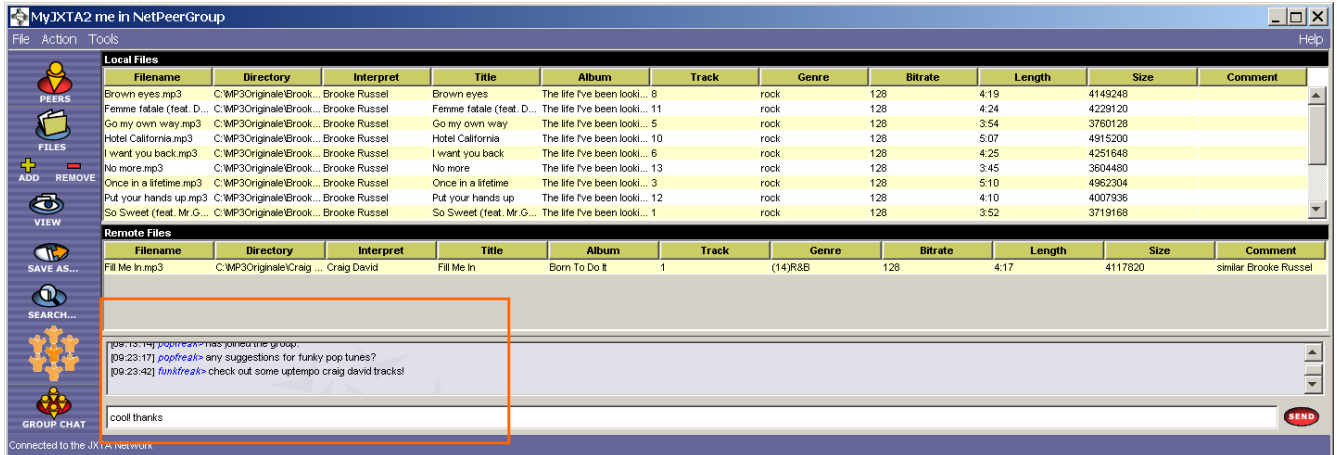


Figure 3. MPeer Client: Embedded chat

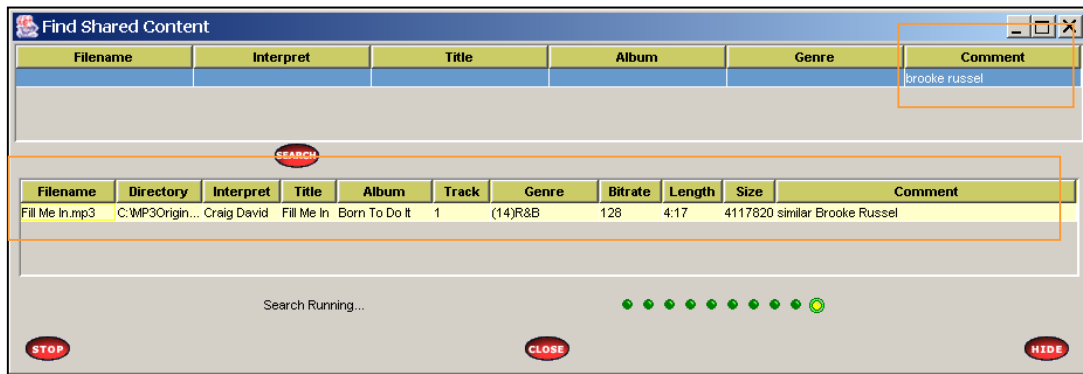


Figure 4. MPeer Client: Metatag Search

Audio Analysis

The aforementioned automatic audio analysis recognizes properties about loudness, tempo and timbral features. These features can be used for the determination of music similarity. Lots of authors have presented different approaches in the recent past to tackle this problem which is a research issue in music information retrieval (MIR). An excellent introduction to the entire MIR field and related communities can be found in [7].

While we are interested in providing recommendations for similar music other authors have coped with the problem of identifying audio tracks in large music databases even for distorted audio queries [8]. Such methods are essential to realize stable digital right management systems (DRM) either for central server or P2P approaches. In our perspective on consumer value we focus on the value for end-users, which is not directly increased by using DRM technology.

Grimm and Nützel [9] presented an alternative model without copy protection for a friendly P2P approach enabling users to re-distribute musical content. Users do not have to care for DRM and have the freedom to receive, consume and re-distribute multimedia content. Re-distribution allows for earning credits and represents therefore increased consumer value. In this way their approach can be seen as complementary to our work.

Wang, Li, Shi [10] evaluated four different P2P models with integrated content-based music retrieval. They showed how acceleration of retrieval could be achieved in large-scale distributed music networks. Therefore they reach an increase of consumer value by improving the extrinsic value of efficiency for music selection. The paper outlines a very generic content-based retrieval method missing details about the audio features and similarity measure. It seems to be optimised for exact retrieval using audio extracts or sung queries. In contrast to our approach recommendations for similar music based on sound or cultural context are neglected.

For the extraction of basic features such as loudness and psychoacoustic features, we used the approaches of Pfeiffer which have been compiled in a toolset under GPL license, available at CSIRO [11]. In parallel we use mel frequency cepstral coefficients and a clustering approach to generate audio profiles. The extracted features are stored as a feature vector in a special user-defined ID3 tag. In this way every peer makes an initial indexing of its files during setup or by entering entities missing this tag. In order to perform a similarity computation based on the audio profiles we use the Earth Moving Distance and Kullback Leibler Distance, as suggested in [12]. Even working with a standard Nearest Neighbor (NN) classifier delivered interesting results for cross-genre recommendations of music “sounding” similar. Professional musicians and non-musicians in a very limited scope have performed subjective evaluation of these results since this process is very time-consuming. In the average 3 out of 5 top hits which we computed have been rated as similar in the subjective evaluations.

3.3. Generation of cultural profiles for artist recommendation

“How much music is in some words?”

Our approach to generate cultural recommendations for similar artists follows a recent tradition of authors [13,14] tackling the problem not with audio-based content analysis. Instead we rely on the acquisition, filtering and condensing of text-based information that can be found in the web. The beauty of this approach lies in the possibility to access so-called cultural metadata which is indeed the agglomeration of several independent - originally subjective - perspectives about music, i.e. artists.

The advantages of such a technique can be summarized as follows:

- Incorporation of semantics: in contrast to content-based audio processing the usage of web reviews offers the possibility to access descriptive semantics about the musical work of an artist.
- Instant availability: in contrast to collaborative filtering techniques such as used by commercial portals (e.g. Amazon) we have no bootstrapping phase in web-based approaches. After the release of new songs or albums immediate access to according reviews is possible.
- Time-awareness: in contrast to the static representation of an audio-based approach the dynamics of changing cultural context, resp. artist relations are included in a web-based representation. Table 1 shows these effects by snapshots of different points in time. The information sources for this example have been *Yahoo Launch Top Ten Music Service* and the most often searched artists in an *OpenNap* network.

Yahoo (12/02)	Yahoo (02/03)	OpenNap (08/01)
Depeche Mode	Depeche Mode	Culture Beat
Madonna	Madonna	Thompson Twins
Sting	Eurythmics	New Order
*NSYNC	The Cure	Blondie
Duran Duran	Enya	Erasure
Dido	Sade	Duran Duran
Alanis Morissette	Sting	Roxette
Nelly Furtado	Pink	Eurythmics
Tears For Fears	U2	Ace of Base
INXS	Phil Collins	Wham
Foreigner	The Cranberries	Depeche Mode
Sade	Duran Duran	A-Ha

Table 1. Dynamics of cultural similarity, e.g. the *Pet Shop Boys*

Technical Approach

Our approach relies solely on the usage of textual features that are contained in the HTML documents. Other parts such as link structure, image or audio content are not considered. As a starting point we crawl about 50 pages per artist that contain reviews of the musical work and apply some filtering techniques to extract the meaningful parts out of these pages (see Figure 3). In subsequent steps part-of-speech tagging, term weighting and the computation of artist similarities are performed. The part-of-speech (POS) tags offer the ability to get rid of some *language noise* that is indeed necessary for the authors to formulate their reviews. We followed the work of Whitman [13] and used the same categories, namely nouns and adjectives. Different combinations of these categories are tested in order to achieve a maximum of performance in predicting similar artists. We chose similar to Whitman:

- single occurrences of terms (n1) e.g. *self*
- pairs of terms (n2), e.g. *daft punk*
- single occurrences of nouns e.g. *rock*
- single occurrences of adjectives e.g. *tricky*



Figure 5. Focus of the term extraction on an artist website

Since we had no natural language parser for noun phrases at hand we implemented the following simple approximation:

- adjective-noun pairs or adverb-adjective-noun triples, e.g. *mindless self indulgence*

We rely on the well-known TFIDF weightings that can be found in standard IR (Information Retrieval) to put bias on important, resp. most-discriminating terms of a document. The computation of TFIDF is based on the number of occurrences of a term in the 50 documents that have been crawled for an artist, denoted as term frequency TF. The document frequency DF of a term counts the number of documents with occurrences of this term in the entire artist collection of size n. The product $TFIDF = TF * \log(n / DF)$ can be used as a weighting score for each term in the collection. The representation of an artist consists of an n-dimensional vector with n = number of unique terms in the set of 50 documents and the TFIDF weightings as according values.

Furthermore different vector types can be stored by using the aforementioned POS combinations. On top of this vector representation the cosine measure has been used to compute similarity between two artists, reps. their TFIDF vectors. The most similar artists to a given artist share common terms and therefore yield larger values. Table 2 shows some examples of TFIDF-weighted vector representations for the *Bloodhound Gang*. We highlighted the ten adjective terms and ten phrases with the highest normalized TFIDF values at each case.

adj Terms	TFIDF	Phrases	TFIDF
funny	0.20463	bad touch	0.86982
juvenile	0.14242	safe version	0.80009
queer	0.12907	fierce beer coaster	0.40004
anomalous	0.09314	inevitable return	0.40004
cultural	0.08607	great white dope	0.40004
annoying	0.07558	still funny	0.40004
limp	0.07339	mindless self indulgence	0.40004
funniest	0.06887	bad language	0.36956
fierce	0.06497	lower volume	0.36956
tricky	0.06497	good beat	0.34793

Table 2. Most relevant phrases for the cultural profile of *Bloodhound Gang*

The cultural profiles and according artist recommendations create for the end user interesting artist's spaces for interactive exploration. This could be used as a driver to enhancement the value types *play (=fun)* and *aesthetics* in the consumer value framework of Holbrook. From a technical point of view this approach can be integrated into the central server approach as being described in 3.1 as well as the P2P setting of chapter 3.2. Furthermore we plan to set up an integrated feature vector consisting of merged audio and textual features by combining the audio analysis of chapter 3.2 with the cultural approach.

4. Conclusion

We have shown that technological solutions may act as value enabling tools in e-commerce scenarios for virtual music. The theoretical framework of Gordijn as well as the quantitative evaluation has been applied in a first real-world application. By adding phonetic matching to a central server portal consumer value has been increased. Advanced features for artist recommendations based on audio and cultural features may act in the same way. We showed the technical feasibility of these ideas in a prototypical P2P environment. We foresee an essential need to integrate the advanced approaches in the forthcoming commercial download platforms of major players to succeed in the virtual music marketplace.

References

1. J. Gordijn, Value based requirements engineering: Exploring innovative e-commerce ideas, *Ph.D thesis*, Vrije Universiteit Amsterdam, (2002).
2. www.musicline.de
3. S. Baumann, Music Similarity Analysis in a P2P environment, *Proceedings of the WIAMIS 2003*, London, UK, April 9-11, (2003).
4. S. Baumann, A. Klüter, Super Convenience for Non-Musicians: Querying MP3 and the Semantic Web, *Proceedings of the ISMIR 2002*, Paris, France, Oct 13-17, (2002).
5. M. B. Holbrook, Consumer Value: A Framework for Analysis and Research, *Routledge, New York, NY*, (1999).
6. A. Weigel, S. Baumann, A modified levensthein-distance for handwriting recognition, *Proceedings of the 7th International Conference on Image Analysis and Processing*, Bari, September, (1993).
7. J. Futrelle, J. S. Downie, Interdisciplinary Communities and Research Issues in Music Information Retrieval, *Proceedings of the ISMIR 2002*, Paris, France, Oct 13-17, (2002).
8. E. Allamanche, J. Herre, O. Hellmuth, B. Froeba, T. Kastner, M. Cremer, Content-based Identification of Audio Material Using MPEG-7 Low Level Description, *Proceedings of the ISMIR 2001*, Indiana University Bloomington, Indiana, USA, October 15-17, (2001).
9. R. Grimm, J. Nützel, Peer-to-Peer Music-Sharing with Profit but Without Copy Protection, *Proceedings of the Wedelmusic 2002*, Darmstadt, Germany, December 9-11,(2002).
10. C. Wang, J. Li, S. Shi, A Kind of Content-Based Music Information Retrieval Method in a Peer-to-Peer Environment, *Proceedings of the ISMIR 2002*, Paris, France, Oct 13-17, (2002).
11. S. Pfeiffer, T. Vincent, Formalization of MPEG-1 compressed domain audio features, *Report No 01/96 of CSIRO Mathematical and Information Sciences*, Australia, Dec, (2001).
12. B. Logan, Mel Frequency Cepstral Coefficients for Music Modelling, *Proceedings of the Int. Symposium on Music Information Retrieval (ISMIR) 2000*, Plymouth, USA, October (2000).
13. B. Whitman. P. Smaragdis, Combining Musical and Cultural Features for Intelligent Style Detection, *Proceedings of the ISMIR 2002*, Paris, France, October 13-17, (2002).
14. Pachet F., Aucouturier J., Representing Musical Genre: A State of the Art, *Journal of New Music Research* (2002).