Songs2See and GlobalMusic2One: Two applied research projects in Music Information Retrieval at Fraunhofer IDMT

Christian Dittmar, Holger Großmann, Estefanía Cano, Sascha Grollmisch, Hanna Lukashevich, and Jakob Abeßer

Fraunhofer IDMT
Ehrenbergstr. 31, 98693 Ilmenau, Germany
{dmr,grn,cano,goh,lkh,abr}@idmt.fraunhofer.de
http://www.idmt.fraunhofer.de

Abstract. At the Fraunhofer Institute for Digital Media Technology (IDMT) in Ilmenau, Germany, two current research projects are directed towards core problems of Music Information Retrieval. The Songs2See project is supported by the Thuringian Ministry of Economy, Employment and Technology through granting funds of the European Fund for Regional Development. The target outcome of this project is a web-based application that assists music students with their instrumental exercises. The unique advantage over existing e-learning solutions is the opportunity to create personalized exercise content using the favorite songs of the music student. GlobalMusic2one is a research project supported by the German Ministry of Education and Research. It is set out to develop a new generation of hybrid music search and recommendation engines. The target outcomes are novel adaptive methods of Music Information Retrieval in combination with Web 2.0 technologies for better quality in the automated recommendation and online marketing of world music collections.

Key words: music information retrieval, automatic music transcription, music source separation, automatic music annotation, music similarity search, music education games

1 Introduction

Successful exploitation of results from basic research is the indicator for the practical relevance of a research field. During recent years, the scientific and commercial interest in the comparatively young research discipline called Music Information Retrieval (MIR) has grown considerably. Stimulated by the ever-growing availability and size of digital music catalogs and mobile media players, MIR techniques become increasingly important to aid convenient exploration of large music collections (e.g., through recommendation engines) and to enable entirely new forms of music consumption (e.g., through music games). Evidently, commercial entities like online music shops, record labels and content aggregators have realized that these aspects can make them stand out among their competitors and foster customer loyalty. However, the industry's willingness to fund basic research in MIR is comparatively low. Thus, only well described methods have found successful application in the real world. For music recommendation and retrieval, these are doubtlessly services based on collaborative filtering[1] (CF). For music transcription and interaction, these are successful video game titles using monophonic pitch detection[2]. The two research projects in the scope of this paper provide the opportunity to progress in core areas of MIR, but always with a clear focus on suitability for real-world applications.

This paper is organized as follows. Each of the two projects is described in more detail in Sec. 2 and Sec. 3. Results from the research as well as the development perspective are reported. Finally, conclusions are given and future directions are sketched.

2 Songs2See

Musical education of children and adolescents is an important factor in their personal self-development regardless if it is about learning a musical instrument or in music courses at school. Children, adolescents and adults must be constantly motivated to practice and complete learning units. Traditional forms of teaching and even current e-learning systems are often unable to provide this motivation. On the other hand, music-based games are immensely popular [7], [17], but they fail to develop skills which are transferable to musical instruments [19]. Songs2See is set out to develop educational software for music learning which provides the motivation of game playing and at the same time develops real musical skills. Using music signal analysis as the key technology, we want to enable students to use popular musical instruments as game controllers for games which teach the students to play music of their own choice. This should be possible regardless of the representation of music they possess (audio, score, tab, chords, etc.). As a reward, the users receive immediate feedback from the automated analysis of their rendition. The game application will provide the students with visual and audio feedback regarding fine-grained details of their performance with regard to timing (rhythm), intonation (pitch, vibrato), and articulation (dynamics). Central to the analysis is automatic music transcription, i.e., the extraction of a scalable symbolic representation from real-world music recordings using specialized computer algorithms [24], [11]. Such symbolic representation allows to render simultaneous visual and audible playbacks for the students, i.e., it can be translated to traditional notation, a piano-roll view or a dynamical animation showing the fingering on the actual instrument. The biggest additional advantage is the possibility to let the students have their favorite song transcribed into a symbolic representation by the software. Thus, the students can play along to actual music they like, instead of specifically produced and edited learning pieces. In order to broaden the possibilities when creating

[1] see for example `http://last.fm`
[2] see for example `http://www.singstargame.com/`

exercises, state-of-the-art methods for audio source separation are exploited. Application of source separation techniques allows to attenuate accompanying instruments that obscure the instrument of interest or alternatively to cancel out the original instrument in order to create a play-along backing track [9]. It should be noted that we are not striving for hi-fi audio quality with regards to the separation. It is more important, that the students can use the above described functionality to their advantage when practicing, a scenario in which a certain amount of audible artifacts is acceptable without being disturbing for the user.

2.1 Research results

A glimpse of the main research directions shall be given here as an update to the overview given in [12]. Further details about the different components of the system are to be published in [17]. Past research activities allowed us to use well-proven methods for timbre-based music segmentation [26], key detection [6] and beat grid extraction [40] out of the box. In addition, existing methods for automatic music transcription and audio source separation are advanced in the project.

Automatic music transcription We were able to integrate already available transcription algorithms that allow us to transcribe the drums, bass, main melody and chords in real-world music segments [11], [33]. In addition, we enable a manual error correction by the user that is helpful when dealing with difficult signals (high degree of polyphony, overlapping percussive instruments etc.). With regard to the project goals and requirements of potential users it became clear that is is necessary to also transcribe monotimbral, polyphonic instruments like the piano and the guitar. Therefore, we conducted a review of the most promising methods for multi-pitch transcription. These comprise the iterative pitch-salience estimation in [23], the maximum likelihood approach in [13] and a combination of the Specmurt principle [34] with shift-invariant non-negative factorization [35]. Results show, that it is necessary to combine any of the aforementioned multi-pitch estimators with a chromagram-based [37] pre-processing to spare computation time. That is especially true for real-time polyphonic pitch detection. For the monophonic real-time case, we could successfully exploit the method pointed out in [18].

Audio source separation We focused on assessing different algorithms for audio source separation that have been reported in the literature. A thorough review, implementation and evaluation of the methods described in [14] was conducted. Inspection of the achievable results led to the conclusion, that sound separation based on tensor factorization is powerful but at the moment computationally too demanding to be applied in our project. Instead, we focused on investigating the more straightforward method described in [31]. This approach separates polyphonic music into percussive and harmonic parts via spectrogram diffusion. It has received much interest in the MIR community, presumably for its simplicity. An alternative approach for percussive vs. harmonic separation has been published in [10]. In that paper, further use of phase information in sound separation problems has been proposed. In all cases, phase information is complementary to the use of magnitude information. Phase contours for musical instruments exhibit similar micromodulations in frequency for certain instruments and can be an alternative of spectral instrument templates or instrument models. For the case of overlapped harmonics, phase coupling properties can be exploited. Although a multitude of further source separation algorithms has been proposed in the literature, only few of them make use of user interaction. In [36], a promising concept for using an approximate user input for extracting sound sources is described. Using a combination of the separation methods described in [9] and [20] in conjunction with user approved note transcriptions, we are able to separate melodic instruments from the background accompaniment in good quality. In addition, we exploit the principle described in [41] in order to allow the user to focus on certain instrument tracks that are well localized within the stereo panorama.

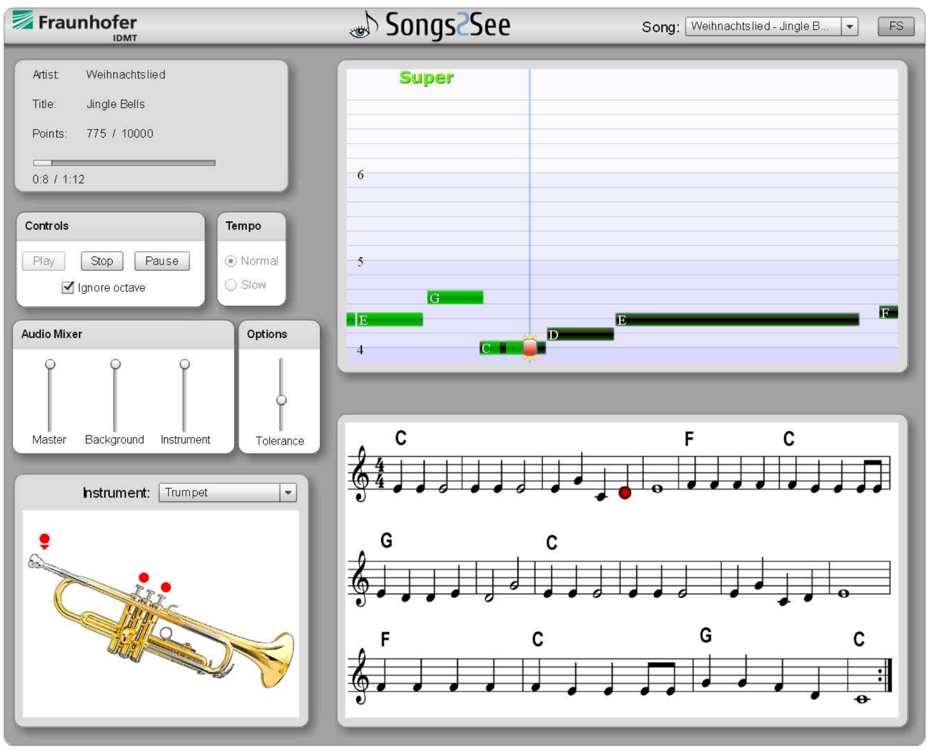

Fig. 1. Screenshot of Songs2See web application

2.2 Development results

In order to have "tangible" results early on, the software development started in parallel with the research tasks. Therefore, interfaces have been kept as generic as possible in order to enable adaption to alternative or extended core algorithms later on.

Songs2See web application In order to achieve easy accessibility to the exercise application, we decided to go for a web-based approach using Flex[3]. Originally, the Flash application built with this tool did not allow direct processing of the microphone input. Thus, we had to implement a streaming server solution. We used the open source Red5[4] in conjunction with the transcoder library Xuggler[5]. This way, we conducted real-time pitch detection [11] in the server application and returned the detected pitches to the web interface. Further details about the implementation are to be published in [18]. Only in their latest release in July 2010, i.e., Adobe Flash Player 10.1, has Flash incorporated the possibility of handling audio streams from a microphone input directly on the client side. A screenshot of the prototype interface is shown in Fig. 1. It can be seen that the user interface shows further assistance than just the plain score sheet. The fingering on the respective instrument of the student is shown as an animation. The relative tones are displayed as well as the relative position of the pitch produced by the players' instrument. This principle is well known from music games and has been adapted here for the educational purposes. Further helpful functions, such as transpose, tempo change and stylistic modification will be implemented in the future.

Songs2See editor For the creation of music exercises, we developed an application with the working title *Songs2See Editor*. It is a stand-alone graphical user interface based on Qt[6] that allows the average or expert user the creation of musical exercises. The editor already allows to go through the prototypical work-flow. During import of a song, timbre segmentation is conducted and the beat grid and key candidates per segment are estimated. The user can choose the segment of interest, start the automatic transcription or use the source separation functionality. For immediate visual and audible feedback of the results, a piano-roll editor is combined with a simple synthesizer as well as sound separation controls. Thus, the user can grab any notes he suspects to be erroneously transcribed, move or delete them. In addition the user is able to seamlessly mix the ratio between the separated melody instrument and the background accompaniment. In Fig. 2, the interface can be seen. We expect that the users of the editor will creatively combine the different processing methods in order to analyze and manipulate the audio tracks to their liking. In the current stage of

[3] see `http://www.adobe.com/products/flex/`
[4] see `http://osflash.org/red5`
[5] see `http://www.xuggle.com/`
[6] see `http://qt.nokia.com/products`

development, export of MIDI and MusicXML is already possible. In a later stage support for other popular formats, such as TuxGuitar will be implemented.

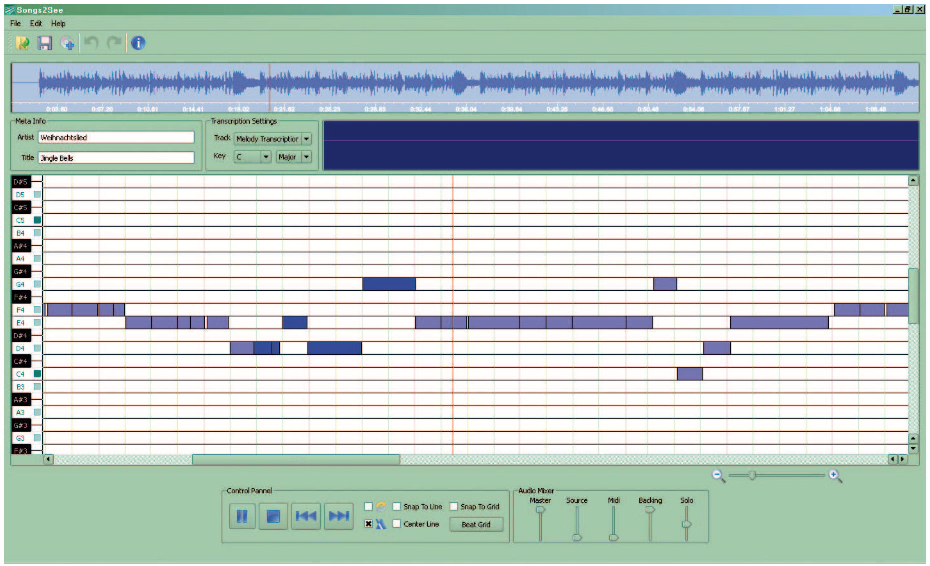

Fig. 2. Screenshot of Songs2See editor prototype

3 GlobalMusic2One

GlobalMusic2one is developing a new generation of adaptive music search engines combining state-of-the-art methods of MIR with Web 2.0 technologies. It aims at reaching better quality in automated music recommendation and browsing inside global music collections. Recently, there has been a growing research interest in music outside the mainstream popular music from the so-called western culture group [39],[16]. For well-known mainstream music, large amounts of user generated browsing traces, reviews, play-lists and recommendations available in different online communities can be analyzed through CF methods in order to reveal similarities between artists, songs and albums. For novel or niche content one obvious solution to derive such data is content-based similarity search. Since the early days of MIR, the search for music items related to a specific query song or a set of those (Query by Example) has been a consistent focus of scientific interest. Thus, a multitude of different approaches with varying degree of complexity has been proposed [32]. Another challenge is the automatic annotation (a.k.a. "auto-tagging" [8]) of world music content. It is obvious that the broad term "World Music" is one of the most ill-defined tags when being used to lump all “exotic genres” together. It lacks justification because this category comprises such a huge variety of different regional styles, influences, and a mutual mix up thereof. On the one hand, retaining the strict classification paradigm for such a high variety of musical styles inevitably limits the precision and expressiveness of a classification system that shall be applied to a world-wide genre taxonomy. With GlobalMusic2One, the user may create new categories allowing the system to flexibly adapt to new musical forms of expression and regional contexts. These categories can, for example, be regional sub-genres which are defined through exemplary songs or song snippets. This self-learning MIR framework will be continuously expanded with precise content-based descriptors.

3.1 Research results

With automatic annotation of world music content, songs often cannot be assigned to one single genre label. Instead, various rhythmic, melodic and harmonic influences conflate into multi-layered mixtures. Common classifier approaches fail due to their immanent assumption that for all song segments, one dominant genre exists and thus is retrievable.

Multi-domain labeling To overcome these problems, we introduced the “multi-domain labeling” approach [28] that breaks down multi-label annotations towards single-label annotations within different musical domains, namely *timbre*, *rhythm*, and *tonality*. In addition, a separate annotation of each temporal segment of the overall song is enabled. This leads to a more meaningful and realistic two-dimensional description of multi-layered musical content. Related to that topic, classification of singing vs. rapping in urban music has been described in [15]. In another paper [27] we applied the recently proposed Multiple Kernel Learning (MKL) technique that has been successfully used for real-world applications in the fields of computational biology, image information retrieval etc. In contrast to classic Support Vector Machines (SVM), MKL provides a possibility of weighting over different kernels depending on a feature set.

Clustering with constraints Inspired by the work in [38], we investigated clustering with constraints with application to active exploration of music collections. Constrained clustering has been developed to improve clustering methods through pairwise constraints. Although these constraints are received as queries from a noiseless oracle, most of the methods involve a random procedure stage to decide which elements are presented to the oracle. In [29] we applied spectral clustering with constraints to a music dataset, where the queries for constraints were selected in a deterministic way through outlier identification perspective. We simulated the constraints through the ground-truth music genre labels. The results showed that constrained clustering with the deterministic outlier identification method achieved reasonable and stable results through the increment of the number of constraint queries. Although the constraints were enhancing the similarity relations between the items, the clustering was conducted in the static feature space. In [30] we embedded the information about the constraints to a feature selection procedure, that adapted the feature space regarding the constraints. We proposed two methods for the constrained feature selection: *similarity-based* and *constrained-based*. We applied the constrained clustering with embedded feature selection for the active exploration of music collections. Our experiments showed that the proposed feature selection methods improved the results of the constrained clustering.

Rule-based classification The second important research direction was rule-based classification with high-level features. In general, high-level features can again be categorized according to different musical domains like rhythm, harmony, melody or instrumentation. In contrast to low-level and mid-level audio features, they are designed with respect to music theory and are thus interpretable by human observers. Often, high-level features are derived from automatic music transcription or classification into semantic categories. Different approaches for the extraction of rhythm-related high-level features have been reported in [21], [25] and [2]. Although automatic extraction of high-level features is still quite error-prone, we proved in [4] that they can be used in a rule-based classification scheme with a quality comparable to state-of-the-art pattern recognition using SVM. The concept of rule-based classification was inspected in detail in [3] using a fine granular manual annotation of high-level features referring to rhythm, instrumentation etc. In this paper, we tested rule-based classification on a restricted dataset of 24 manually annotated audio tracks and achieved an accuracy rate of over 80%.

Novel audio features Adding the fourth domain *instrumentation* to the multi-domain approach described in Sec. 3.1 required the design and implementation of novel audio features tailored towards instrument recognition in polyphonic recordings. Promising results even with instruments from non-European cultural areas are reported in [22]. In addition, we investigated the automatic classification of rhythmic patterns in global music styles in [42]. In this work, special measures have been taken to make the features and distance measures tempo independent. This is done implicitly, without the need for a preceding beat grid extraction that is commonly recommended in the literature to derive beat synchronous feature vectors. In conjunction with the approach to rule-based classification described in Sec. 3.1, novel features for the classification of bass-playing styles have been published in [5] and [1]. In this paper, we compared an approach based on high-level () features and another one based on similarity measures between bass patterns. For both approaches, we assessed two different strategies: classification of patterns as a whole and classification of all measures of a pattern with a subsequent accumulation of the classification results. Furthermore, we investigated the influence of potential transcription errors on the classification accuracy. Given a taxonomy consisting of 8 different bass playing styles, best classification accuracy values of 60.8% were achieved for the feature-based classification and 68.5% for the pattern similarity approach.

3.2 Development results

As with the Songs2See project, the development phase in GlobalMusic2One started in parallel with the research activities and is now near its completion. It was mandatory to have early prototypes of the required software for certain tasks in the project.

Annotation Tool We developed an Qt-based *Annotation Tool* that facilitates the gathering of conceptualized annotations for any kind of audio content. The program was designed for expert users to enable them to manually describe audio files efficiently on a very detailed level. The Annotation Tool can be configured to different and extensible annotation schemes making it flexible for multiple application fields. The tool supports single labeling, and multi-labeling, as well as the new approach of multi-domain labeling. However, strong labeling is not enforced by the tool, but remains under the control of the user. As a unique feature the audio annotation tool comes with a automated, timbre-based audio segmentation algorithm integrated that helps the user to intuitively navigate through the audio file during the annotation process and select the right segments. Of course, the granularity of the segmentation can be adjusted and every single segment border can be manually corrected if necessary. There are now approx. 100 observables that can be chosen from while annotating. This list is under steady development. In Fig. 3, a screenshot of the Annotation Tool configured to the Globalmusic2one description scheme can be seen.

PAMIR framework The Personalized Adaptive MIR framework (PAMIR) is a server system written in Python that loosely strings together various functional blocks. These blocks are e.g., a relational database, a server for computing content-based similarities between music items and various machine learning servers. PAMIR is also the instance that enables the adaptivity with respect to the user preferences. Basically, it allows to conduct feature-selection and automated picking of the most suitable classification strategy for a given classification problem. Additionally, it enables content-based similarity search both on song and segment level. The latter method can be used to retrieve parts of songs, that are similar to the query, whereas the complete song may exhibit different properties. Visualizations of the classification and similarity-search results are delivered to the users via a web interface. In Fig. 4, the prototype web interface to the GlobalMusic2One portal is shown. It can be seen that we feature an intuitive similarity map for explorative browsing inside the catalog as well as target-oriented search masks for specific music properties. The same interface allows the instantiation of new concepts and manual annotation of reference songs. In the upper left corner, a segment player is shown that allows to jump directly to different parts in the songs.

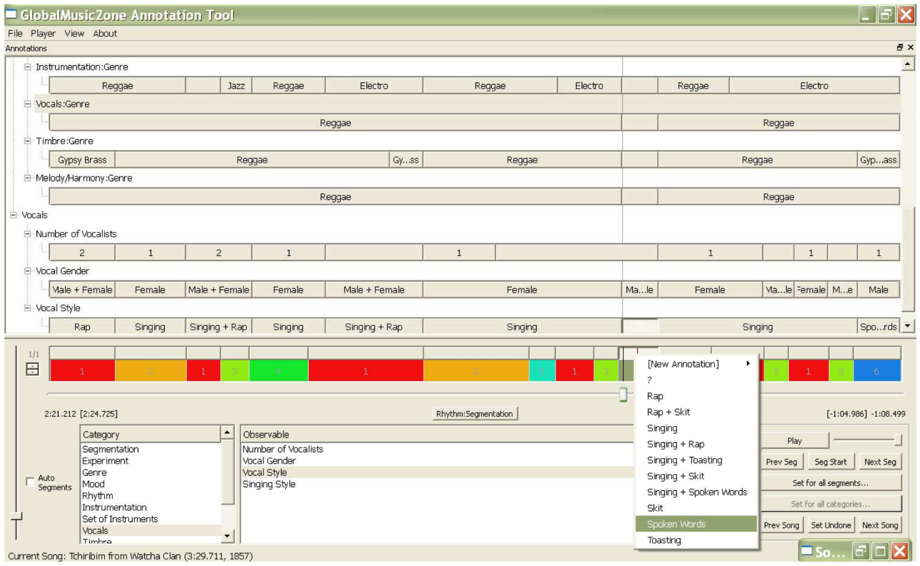

Fig. 3. Screenshot of the Annotation Tool configured to the Globalmusic2one description scheme, showing the annotation of a stylistically and structurally complex song.

4 Conclusions & Outlook

In this paper we presented an overview of the applied MIR projects Songs2See and GlobalMusic2One. Both address core challenges of MIR with strong focus to real-world applications. With Songs2See, development activities will be strongly directed towards implementation of more advanced features in the editor as well as the web application. The main efforts inside GlobalMusic2One will be concentrated on consolidating the framework and the web client.

5 Acknowledgments

The Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project *Songs2See*[7], enabling transnational cooperation between Thuringian companies and their partners from other European regions. Additionally, this work has been partly supported by the German research project *GlobalMusic2One*[8] funded by the Federal Ministry of Education and Research (BMBF-FKZ: 01/S08039B).

[7] see `http://www.songs2see.eu`

[8] see `http://www.globalmusic2one.net`

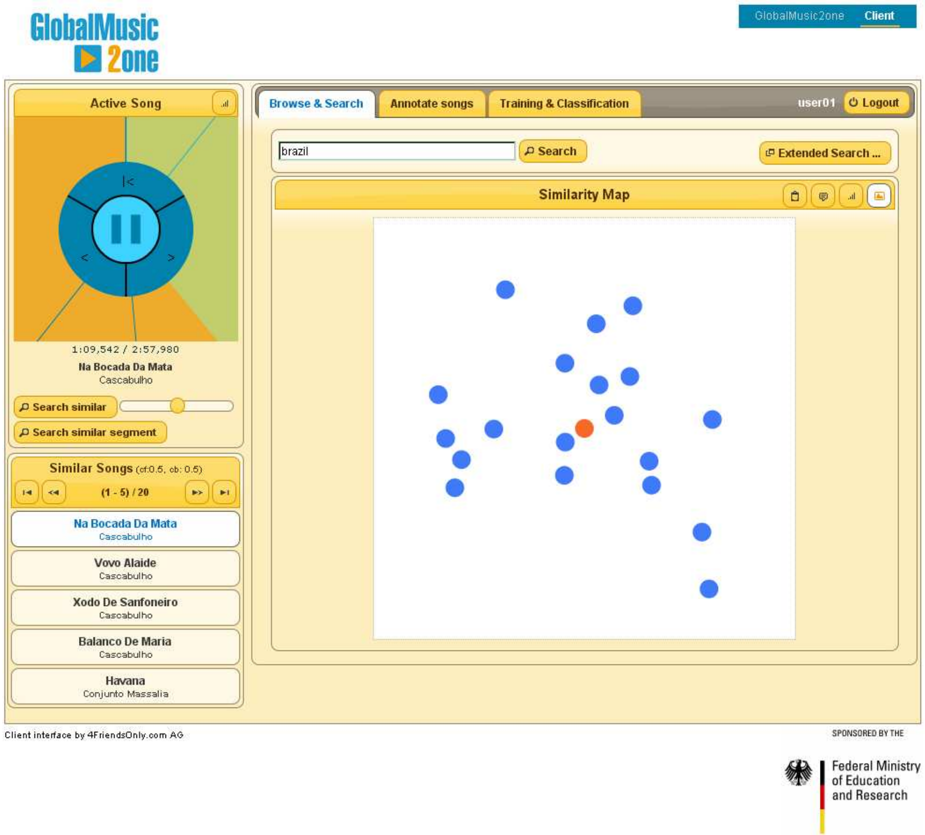

Fig. 4. Screenshot of the GlobalMusic2One prototype web client

References

1. J. Abeßer, P. Bräuer, H. Lukashevich, and G. Schuller. Bass playing style detection based on high-level features and pattern similarity. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, Netherlands, 2010.
2. J. Abeßer, C. Dittmar, and H. Großmann. Automatic genre and artist classification by analyzing improvised solo parts from musical recordings. In *Proceedings of the Audio Mostly Conference (AMC)*, Piteå, Sweden, 2008.
3. J. Abeßer, H. Lukashevich, C. Dittmar, P. Bräuer, and F. Krause. Rule-based classification of musical genres from a global cultural background. In *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, Malaga, Spain, 2010.
4. J. Abeßer, H. Lukashevich, C. Dittmar, and G. Schuller. Genre classification using bass-related high-level features and playing styles. In *Proceedings of the 10th*

International Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan, 2009.
5. J. Abeßer, H. Lukashevich, and G. Schuller. Feature-based extraction of plucking and expression styles of the electric bass guitar. In *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, USA, 2010.
6. D. Arndt, G. Gatzsche, and M. Mehnert. Symmetry model based key finding. In *Proceedings of the 126th AES Convention*, Munich, Germany, 2009.
7. A. Barbancho, I. Barbancho, L. Tardon, and C. Urdiales. Automatic edition of songs for guitar hero/frets on fire. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, New York, USA, 2009.
8. T. Bertin-Mahieux, D. Eck, F. Maillet, and P. Lamere. Autotagger: a model for predicting social tags from acoustic features on large music databases. *Journal of New Music Research*, 37(2):115–135, 2008.
9. E. Cano and C. Cheng. Melody line detection and source separation in classical saxophone recordings. In *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx)*, Como, Italy, 2009.
10. E. Cano, G. Schuller, and C. Dittmar. Exploring phase information in sound source separation applications. In *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, 2010.
11. C. Dittmar, K. Dressler, and K. Rosenbauer. A toolbox for automatic transcription of polyphonic music. In *Proceedings of the Audio Mostly Conference (AMC)*, Ilmenau, Germany, 2007.
12. C. Dittmar, H. Großmann, E. Cano, S. Grollmisch, H. Lukashevich, and J. Abeßer. Songs2See and GlobalMusic2One - Two ongoing projects in Music Information Retrieval at Fraunhofer IDMT. In *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, Malaga, Spain, 2010.
13. Z. Duan, B. Pardo, and C. Zhang. Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions. *IEEE Transactions on Audio, Speech, and Language Processing*, PP(99):1 –1, 2010.
14. D. Fitzgerald, M. Cranitch, and E. Coyle. Extended nonnegative tensor factorization models for musical sound source separation. *Computational Intelligence and Neuroscience*, 2008.
15. D. Gärtner. Singing / rap classification of isolated vocal tracks. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, Netherlands, 2010.
16. E. Gómez, M. Haro, and P. Herrera. Music and geography: Content description of musical audio from different parts of the world. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, 2009.
17. S. Grollmisch, C. Dittmar, and E. Cano. Songs2see: Learn to play by playing. In *Proceedings of the 41st AES International Conference on Audio in Games*, London, UK, 2011.
18. S. Grollmisch, C. Dittmar, E. Cano, and K. Dressler. Server based pitch detection for web applications. In *Proceedings of the 41st AES International Conference on Audio in Games*, London, UK, 2011.
19. S. Grollmisch, C. Dittmar, and G. Gatzsche. Implementation and evaluation of an improvisation based music video game. In *Proceedings of the IEEE Consumer Electronics Society's Games Innovation Conference (IEEE GIC)*, London, UK, 2009.

20. M. Gruhne, K. Schmidt, and C. Dittmar. Phoneme recognition on popular music. In *8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.
21. P. Herrera, V. Sandvold, and F. Gouyon. Percussion-related semantic descriptors of music audio files. In *Proceedings of the 25th International AES Conference*, London, UK, 2004.
22. M. Kahl, J. Abeßer, C. Dittmar, and H. Großmann. Automatic recognition of tonal instruments in polyphonic music from different cultural backgrounds. In *Proceedings of the 36th Jahrestagung fr Akustik (DAGA)*, Berlin, Germany, 2010.
23. A. Klapuri. A method for visualizing the pitch content of polyphonic music signals. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, 2009.
24. A. Klapuri and M. Davy, editors. *Signal Processing Methods for Music Transcription.* Springer Science + Business Media LLC, New York, 2006.
25. T. Lidy, A. Rauber, A. Pertusa, and J. M. Iesta. Improving genre classification by combination of audio and symbolic descriptors using a transcription system. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.
26. H. Lukashevich. Towards quantitative measures of evaluating song segmentation. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, Pennsylvania, USA, 2008.
27. H. Lukashevich. Applying multiple kernel learning to automatic genre classification. In *Proceedings of the 34th Annual Conference of the German Classification Society (GfKl)*, Karlsruhe, Germany, 2010.
28. H. Lukashevich, J. Abeßer, C. Dittmar, and H. Großmann. From multi-labeling to multi-domain-labeling: A novel two-dimensional approach to music genre classification. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, 2009.
29. P. Mercado and H. Lukashevich. Applying constrained clustering for active exploration of music collections. In *Proceedings of the 1st Workshop on Music Recommendation and Discovery (WOMRAD)*, Barcelona, Spain, 2010.
30. P. Mercado and H. Lukashevich. Feature selection in clustering with constraints: Application to active exploration of music collections. In *Proceedings of the 9th Int. Conference on Machine Learning and Applications (ICMLA)*, Washington DC, USA, 2010.
31. N. Ono, K. Miyamoto, J. L. Roux, H. Kameoka, and S. Sagayama. Separation of a monaural audio signal into harmonic/percussive components by complememntary diffusion on spectrogram. In *Proceedings of the 16th European Signal Processing Conferenc (EUSIPCO)*, Lausanne, Switzerland, 2008.
32. T. Pohle, D. Schnitzer, M. Schedl, P. Knees, and G. Widmer. On rhythm and general music similarity. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, Kobe, Japan, 2009.
33. M. Ryynänen and A. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32:72–86, 2008.
34. S. Sagayama, K. Takahashi, H. Kameoka, and T. Nishimoto. Specmurt anasylis: A piano-roll-visualization of polyphonic music signal by deconvolution of log-frequency spectrum. In *Proceedings of the ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA)*, Jeju, Korea, 2004.
35. M. Shashanka, B. Raj, and P. Smaragdis. Probabilistic latent variable models as nonnegative factorizations. *Computational Intelligence and Neuroscience*, 2008.

36. P. Smaragdis and G. J. Mysore. Separation by "humming": User-guided sound extraction from monophonic mixtures. In *Proceedings of IEEE Workshop on Applications Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 2009.
37. M. Stein, B. M. Schubert, M. Gruhne, G. Gatzsche, and M. Mehnert. Evaluation and comparison of audio chroma feature extraction methods. In *Proceedings of the 126th AES Convention*, Munich, Germany, 2009.
38. S. Stober and A. Nürnberger. Towards user-adaptive structuring and organization of music collections. In M. Detyniecki, U. Leiner, and A. Nürnberger, editors, *Adaptive Multimedia Retrieval. Identifying, Summarizing, and Recommending Image and Music. 6th International Workshop, AMR 2008, Berlin, Germany, June 26-27, 2008. Revised Selected Papers*, volume 5811 of *LNCS*, pages 53–65, Heidelberg / Berlin, 2010. Springer Verlag.
39. G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2):1–24, 2007.
40. C. Uhle. *Automatisierte Extraktion rhythmischer Merkmale zur Anwendung in Music Information Retrieval-Systemen*. PhD thesis, Ilmenau University, Ilmenau, Germany, 2008.
41. M. Vinyes, J. Bonada, and A. Loscos. Demixing commercial music productions via human-assisted time-frequency masking. In *Proceedings of the 120th AES convenction*, Paris, France, 2006.
42. T. Völkel, J. Abeßer, C. Dittmar, and H. Großmann. Automatic genre classification of latin american music using characteristic rhythmic patterns. In *Proceedings of the Audio Mostly Conference (AMC)*, Piteå, Sweden, 2010.