Songrium: A Music Browsing Assistance Service with Interactive Visualization and Exploration of a Web of Music

Masahiro Hamasaki Masataka Goto Tomoyasu Nakano

National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
songrium-ml [at] aist.go.jp

ABSTRACT

This paper describes a music browsing assistance service, *Songrium* (`http://songrium.jp`), which increases user enjoyment when listening to songs and allows visualization and exploration of a *"Web of Music"*. We define a Web of Music in this paper to be a network of *"web-native music"*, which we define in turn to be music that is published, shared, and remixed (has derivative works created) entirely on the web. Songrium was developed as an attempt to realize a Web of Music, by showing relations between both original songs and derivative works and offering an enriched listening experience. Songrium has analyzed over 600,000 music video clips on the most popular Japanese video-sharing service, *Niconico*, which contains original songs of web-native music and their derivative works such as covers and dance arrangements. Analysis of over 100,000 original songs reveals that over 500,000 derivative works were generated and have contributed to enrich the Web of Music.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services; H.5.5 [**Sound and Music Computing**]: Signal analysis, synthesis, and processing

General Terms

Design, Human Factors

Keywords

Web of Music, Web-native music, Music interface, Visualization, Web service, User-generated content

1. INTRODUCTION

Despite being a creative process, composing a musical piece or music video clip is often inspired by existing work, and as such, various types of relations between musical entities can be found. Similarity relations, for example, include those based on lyrics, melody line, chord progression, genre, mood, musical instrument, etc. Derivative relations, such as new arrangements, remixes, and covers, are also popular. These relations among songs form a large and complex network which, in contrast to a web of documents, is not immediately accessible. For example, in the web of documents [6], hyperlinks allow people to quickly and automatically see which documents are related to which. In a web of musical works, on the other hand, the lack of a tool for viewing relations among songs or music video clips precludes users from exploring this rich and complex space.

In this paper we propose that with recent advances in web-crawling, metadata analysis, music signal processing, machine learning, and data visualization techniques, we are close to realizing a *"Web of Music"*. We demonstrate this by describing a music browsing assistance service, *Songrium* (http://songrium.jp), which we have developed in order to allow users to explore music while seeing and utilizing various relations among music video clips on video sharing websites.

Although YouTube[1] is the most popular video sharing service in terms of views and uploads of music video clips, there are too many missing nodes and links to realize a web of musical works. The majority of the popular music content on YouTube is commercial music, and most commercial music video clips are uploaded for promotional purposes. These songs/videos are usually created independently of the web and subsequently uploaded as finished works. It is therefore common, for example, for some works (nodes) and relations (links between works) by an artist to be missing.

In contrast, music content related to the singing synthesis technology *VOCALOID* [14] on the most popular Japanese video sharing service, *Niconico*[2], largely features content native to the web. The music and derivative content contained within were born, listened to, and distributed on the web. For this reason, we describe this type of music as *web-native music*, which we believe to be the key factor in the formation of the Web of Music. The current main target of Songrium is therefore original songs using VOCALOID and their derivative works on Niconico.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2576944.

[1] http://www.youtube.com
[2] http://nicovideo.jp

In the present paper, we describe the Web of Music as seen through the lens of Songrium. After gathering and analyzing over 600,000 video clips of web-native music from Niconico, we found that web-native music has a strong power of generating derivative works, which are five times more common than original songs. Interestingly, we found that some derivative works received more views than their original songs.

2. WEB-NATIVE MUSIC ON NICONICO

We define music satisfying all the following three conditions to be *web-native music*:

(1) It is generally assumed that new original songs are first released on the web (without CD release or radio play, for example), with a unique URL identifying the source and release date.

(2) Creators do not hesitate to create and release derivative works of original songs on the web.

(3) After releasing original or derivative works, their creators can publicly receive feedback on the web and be encouraged to create more content.

The consequence of condition (1) is that everybody can instantly listen to new content and be assured of its source. The lack of barriers to viewing content is clearly attractive to people. Moreover, the identifying URL means that the popularity and other statistics of web-native music can be automatically tracked by systems such as Songrium. Condition (2) means that there is a huge potential to generate derivative works from original content, which is crucial to the formation of a mature Web of Music. Marshall *et al.* [16] reported that remixing and republishing can pose issues in some web music, highlighting the need for web-native music to have well-defined and accepted rules for the creation of derivative works. The final condition in our definition motivates creators to continue making new content and derivative works, forming a positive feedback loop for both creators and listeners. Satisfying all three conditions, VOCALOID songs on Niconico are an ideal example of web-native music, as outlined below.

VOCALOID is a singing synthesis technology [14] which forms a subset of the music content creation community on Niconico. This technology has begun to be used to synthesize the main vocal melody of songs, and many examples have been published as original works on the website. Despite the impressive technology, the vocals produced by VOCALOID are easily identifiable as not of human origin, meaning that both creators and listeners naturally accept that these songs are first published on the web. Niconico therefore serves as a forum for VOCALOID creators and listeners to coalesce. Furthermore, because page views and publishing date are managed by the website, metadata can easily be used to analyze the music content and provide ranking of video clips based on page views, etc.

There are many different products based on VOCALOID and each product has a different vocal timbre. Most products have an associated character image, with Hatsune Miku [3] being the most well-known. Right after releasing Hatsune Miku, Crypton Future Media (developer of Hatsune Miku) officially started allowing users to reuse its character for derivative works with their original license: Piapro Character License [4]. Following this, users started to create music videos, such as promotion videos for musicians, with such original songs and drawings. Some users even went as far as to create 3D models of Hatsune Miku and create 3D animation videos [9, 8]. Subsequently, many songwriters publish karaoke (full song without vocals) version of their own original songs, prompting some users to sing these songs and also publish derivative works recorded in video clips.

Niconico is one of the most popular video communication services in Japan today. As with similar services (YouTube, etc.), users are able to upload and view videos, but Niconico has additional feature of a text communication which overlays recent comments by anonymous users on the video, synchronized to the video playback. These overlaid comments create a sense of a shared watching/listening experience, labelled by Satoshi Hamano as *Pseudo Synchronized Communication*. Essentially, users feel part of a community by viewing the impressions, interpretations and feelings expressed in the comments. These time-synchronized comments also give feedback to creators from listeners and often motivative the creation of new content.

Figure 1 shows the number of monthly published original songs and their derivative works from September 2007 to August 2012. In this paper, we define the term 'derivative work' to be a video clip reusing a part of or whole of

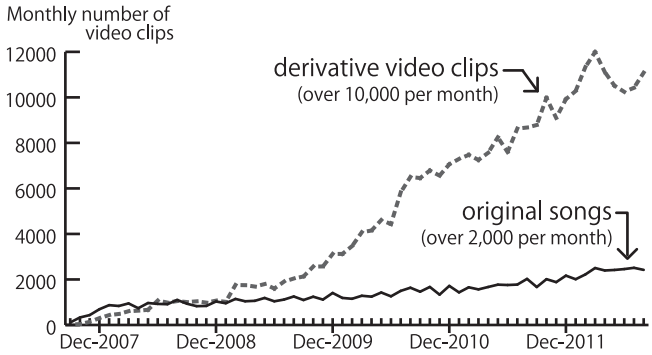

Figure 1: The monthly number of published original songs and their derivative works in the period September 2007 – August 2012.

Table 1: Classification results of 517,024 derivative works. Since some derivative works have multiple categories, the total number of classifications is greater than the number of derivative works.

category	# of works
(a) Singing	346,192
(b) Dancing	22,808
(c) Arranging and Performing	30,766
(d) Featuring 3D characters	28,311
(e) Creating Music video	8,174
(f) Others	80,880

[3] `http://www.crypton.co.jp/mp/pages/prod/vocaloid/cv01_us.jsp`

[4] From December 2012, they use Creative Commons license for foreign users.

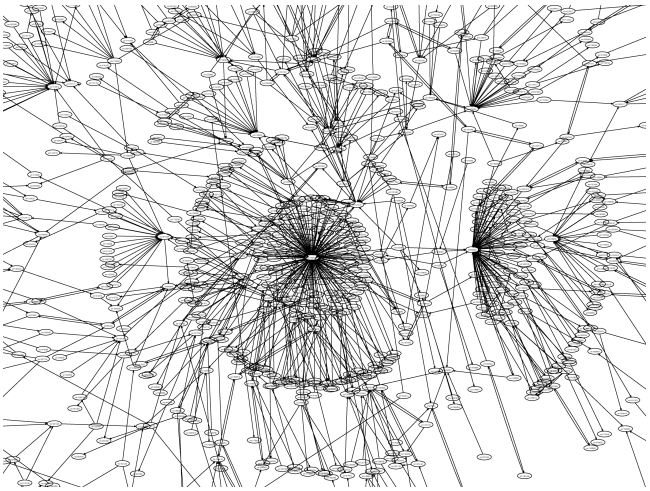

Figure 2: Re-using network of original songs and their derivative works in VOCALOID community on Niconico.

video clip of VOCALOID original song. According to this figure, the number of uploaded original songs and derivative works has been increasing rapidly. 14% of original songs have a derivative work whose page view is higher than original work, indicating that derivative works are also attractive content.

80% of derivative works are published after one month of the original publication date, while 40% are published after one year, showing that derivative works extend the longevity of original works. Table 1 shows classification results of 517,024 derivative works of 108,215 VOCALOID original songs that we gathered from Niconico. We defined six categories of derivative works: (a) Singing a song, (b) Dancing to a song, (c) Performing a song on musical instruments, (d) Featuring 3D characters in music video, (e) Creating a music video for a song, and (f) Others. The first three categories derived from official categories from Niconico; the other two categories are derived from our previous work [9, 8]. "Others" includes, for example, videos which review or rank existing videos, or use VOCALOID songs as the background music to other video content.

A reuse network of this large number of derivative works in the VOCALOID community on Niconico forms a huge directed graph where each node denotes a different video clip and each edge has a direction from derivative work to an original work. Figure 2 shows a small excerpt of the huge reuse network in the VOCALOID community. This network forms part of a Web of Music, but is not in plain sight for users.

3. SONGRIUM FOR VISUALIZATION OF WEB OF MUSIC

By visualizing the Web of Music, Songrium enables people to have a better understanding of various relations in the web-native music and an enriched interactive experience with the Web of Music. It was difficult for people listening to original songs to notice that there exist various derivative works of them, such as cover versions, singing or dancing video clips, and music video clips with 3D animations. By providing people with easy, intuitive access to those derivative works, Songrium allows them to not only find interesting music video clips but also understand and respect creators of music and video clips.

Songrium uses web mining and music understanding technologies together with advanced data visualization techniques to achieve unique functions, such as Music Star Map, Planet View, Bubble Player, and Singing Voice Analysis.

3.1 Music Star Map and Planet View

Music Star Map is a function that visualizes original songs. Original songs are embedded in a two-dimensional space, mapped automatically based on audio feature similarity. The position of a song on the map is such that songs in close proximity have similar moods (estimated by audio feature analysis, see 4.2). Figure 3-(A) portrays a screenshot of this function. Furthermore, when a user clicks an original song on Music Star Map, its derivative works appear as colorful icons and orbit the selected song. We call this view "Planet View". Figure 3-(B) shows a screenshot of Planet View.

In Figure 3-(B), each circle icon denotes a derivative work with attributes represented by the icon orbit, size, color, and velocity. The distance from the center is indicative of the publishing date, with the most recent work in the outermost orbit. The size of each icon reflects the number of page views; the color indicates one of the following derivative categories: Blue (Singing), Red (Dancing), Green (Arranging and Performing), Purple (3D characters in music video), Yellow (Creating music video), and White (Others). Finally, the velocity (orbit speed) of an icon represents how many times the content has been favourited by users of the system.

The official embedded video player of the Niconico service is shown at the upper-right corner to play back a video clip of the selected original song (Figure 3-(C)). Our music-listening interface has a chorus-search function for trial listening, *SmartMusicKIOSK* [4], which is shown below the embedded player (Figure 3-(D)). Songrium has an original social tagging framework called 'Arrow Tag' that allows to annotate a relation between music content [7]. Figure 3-(E) shows a list of Arrow Tags.

3.2 Bubble Player

Bubble Player visualizes the history of VOCALOID songs. It plays groups of associated songs published within a user-specified time frame on continuous playback. The interface shows the growth in popularity of songs, arranged by published date, in an animated display and only plays songs that satisfy certain user-specified conditions (play count, etc.). Thus, this feature allows users to experience a group of songs in one continuous movie, giving a clear, intuitive picture of how trends on video-sharing services behave.

Figure 4 shows a screen shot of Bubble Player and examples of animation. When a user specifies a period, Songrium displays groups of songs published during the specified period in chronological order, giving the user a full perspective on the trends and transitions in published song groups over time. Each song is represented by a "bubble" (a colored circle). New song bubbles appear in accordance with their respective published dates and congregate in an animation. The colors of the bubbles correspond to the voice synthesis

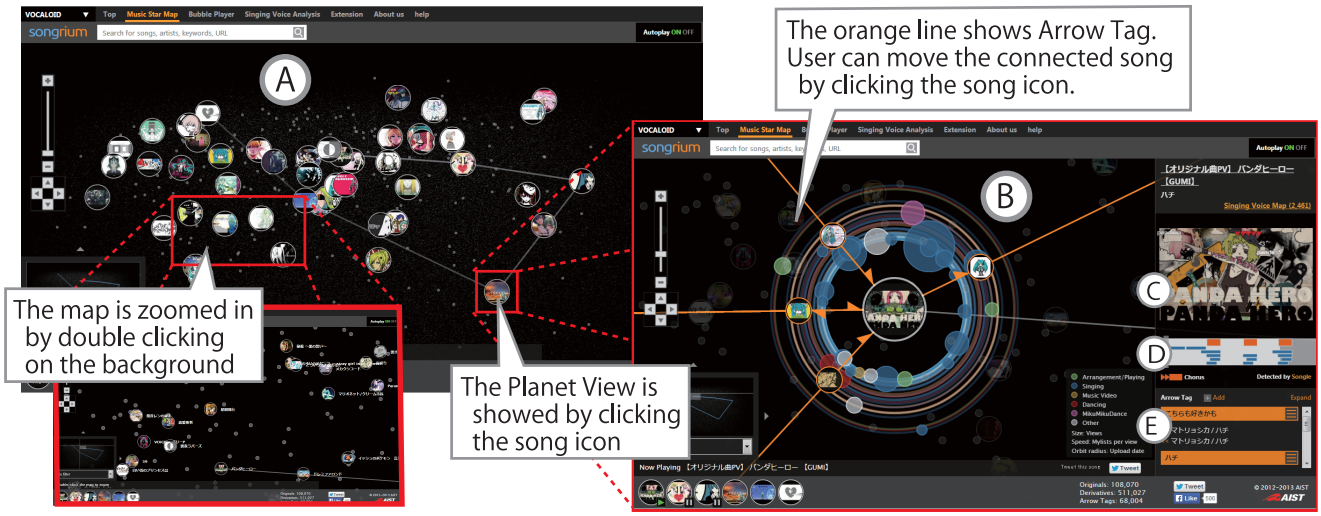

Figure 3: Screenshot of the (A) "Music Star Map" and (B) "Planet View" interface in Songrium. The former visualizes original songs; the latter their derivative works. Both are connected seamlessly. (A) All original songs are arranged in a two-dimensional space with similar songs positioned in close proximity. Any particular area can be expanded for viewing by double clicking, and then scrolled to by dragging. (B) After selecting an original song on Music Star Map, users can view its derivative works rotating around the selected song in the center. (C) Embedded video player of Niconico for video playback. (D) Playback interface for trial listening (SmartMusicKIOSK). (E) Social annotated relations called Arrow tags [7] to and from this song instance.

library used in the VOCALOID software, while the sizes of the bubbles indicate play counts. Songrium also plays truncated chorus sections of certain songs that satisfy conditions specified by the user, making it easy to see what kinds of songs were being published as more and more bubbles appear. In the background, the display shows pieces of text describing related events.

3.3 Singing Voice Analysis

"Singing a VOCALOID song" (in which a human singer covers a VOCALOID song) is currently the most popular category of derivative work. These derivative singing videos represent one of the most-watched and most popular sets of content on video-sharing services, with some original songs having hundreds or even thousands of derivative singing videos. The derivative singing videos for a given song are all based on the same melody and song structure, meaning their defining characteristics lie solely in the properties of voice and vocal interpretations of the singers. Singing Voice Analysis visualizes these derivative works based on measurements of male and female singing voice characteristics.

Figure 5 shows a screenshot of the Singing Voice Analysis. Red and blue circles in the middle of the screen represent singing videos. The higher the circle (larger y-coordinate), the more feminine the singer's voice; the lower the circle (smaller y-coordinate), the more masculine the voice. Through this visualization of voice properties, users can browse a wide array of singing videos through the lens of vocal attributes.

4. IMPLEMENTATION

Songrium consists of three components: a web crawler, a web server, and a user interface. The web crawler automatically checks updated music video clips related to VOCALOID on Niconico on a daily basis. These are then classified using the technique described in Section 4.1, and analyzed in the audio domain in order to understand the musical content using the method described in Section 4.2. The user interface is implemented using HTML5, SVG, JavaScript, the JavaScript library D3.js, and the embedded video player of the Niconico service.

Songrium service was released to the public at `http://songrium.jp` on August 7, 2012. As of December 2013, 108,215 original songs and 517,024 derivative works have been registered in Songrium.

4.1 Classification of Video Clips

Every music video clip on Songrium is classified automatically as an original song or a derivative work. Because Niconico supports social tags for each clip and tags of some kinds such as "Original Song" and "be enshrined in the Hall of Fame song" are usually put on original songs on Niconico, we can rely on these tags. However, even if some original songs have no such tag, Songrium automatically classifies them correctly by crawling a set of related web sites to generate the "white list" of VOCALOID original songs. In the case of derivative works, these can be readily identified when the description text of the video clip includes a hyperlink to the original video clip from which it was derived. These hyperlinks almost always exist on Niconico because users like to acknowledge the original video clip.

When a derivative work is incorporated, its relation to the original song is estimated automatically. The derivative works are classified into one of the six predefined categories described in Section 3.1. With the exception of category *Others* all the remaining categories are extremely popular and have their own unique social tags on Niconico. Using

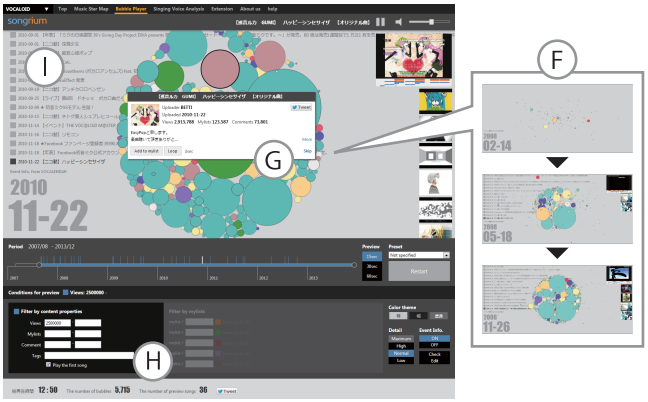

Figure 4: Screenshot of the "Bubble Player" of Songrium showing various types of information. (F) Bubble animation intuitively shows how many songs were published during a user-specified period. (G) Chorus playback and metadata of songs that satisfy user-specified conditions. (H) User input form for playback conditions. (I) Related events are shown in the background (sourced from open calendar data on the web).

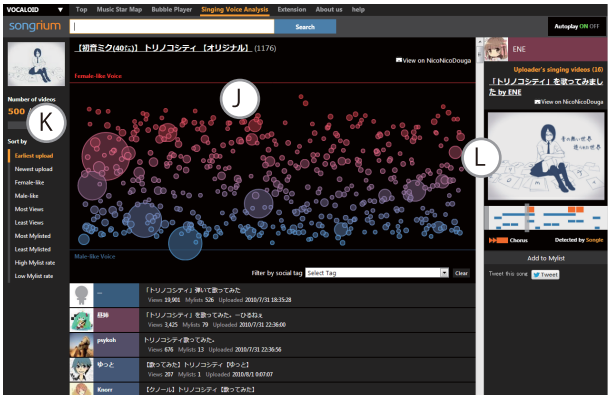

Figure 5: Screenshot of the "Singing Voice Analysis" of Songrium, showing visualization of the proportions of male-type and female-type vocals of a song's derivative works. (J) Circles indicate derivative work. $y-$axis shows the vocal type; x-axis can be changed from the left panel. (K) Users can choose the order of the circle in the x-axis, e.g., number of page views, publishing date, and so on. (L) Embedded video player of Niconico for derivative work playback, featuring SmartMusicKIOSK.

these tags, Songrium can produce a reliable classification of derivative works.

Moreover, Songrium enables users to easily report an error in any of the above classification of video clips, extraction of links, or estimation of relations to further improve the user experience.

4.2 Understanding Music in Video Clips

User interfaces of Songrium utilize not only web-mined text metadata but also results of understanding audio signal on music content. This combination expands variety of music browsing assistance.

Chorus sections and repeated sections are estimated by using the audio-based chorus-section detection method *RefraiD* [4], which was also used as the back-end for the original SmartMusicKIOSK interface. By analyzing relations between various repeated sections, the RefraiD method can detect all the chorus sections in a song and estimate both the start and end points of each section. It can also detect modulated (transposed) chorus sections, which are especially common in pop music.

The position of each original song is mapped to a two-dimensional space of the Music Star Map based on analysis of audio features. Although any feature vector designed for computing music similarity could be used, in our current implementation, we use a 200-dimensional audio feature vector obtained using learned latent representations [10]. To determine the (x-y) position of a song on the Music Star Map, its 200-dimensional feature vector is projected onto two dimensions using principal component analysis (PCA), where only the first two components are retained.

To estimate male and female singing voice characteristics from singing video clips, we use a probability estimation model with extracted *reliable frames* [2] including the vocal part. In order to extract such frames from video clips with polyphonic music, we first estimate the melody using the audio-based F_0 estimation method *PreFEst* [3]. Following this, LPMCC (mel-cepstral coefficients of LPC spectrum) of vocal and ΔF_0 are estimated and then combined into a single feature vector at each frame. Then, reliable frames are classified as vocal by using a vocal/non-vocal GMM [2].

Finally, the male/female probabilities of each reliable frame are estimated using a probabilistic Support Vector Machine (SVM) [1], which are reduced to a single value by taking the median across the SVM output.

5. RELATED WORK

Songrium, at its core, is a music browsing assistance service. Most previous research into interactive music browsing focused on visualizations to explore musical collections [17, 18, 5, 21, 12]. Given the multiple dimensions associated with music data, one particular visualization technique often attempted is visualization of a music collection in a two-dimensional plane [17, 21, 12]. Our Music Star Map (see Section 3.1) is one particular example of this. Interactive interfaces are also important for user experiences; [5] assists a user in discovering songs and [18] assists a user in finding artists. In contrast to the advances in interactive music browsing mentioned above, Songrium visualizes not only original songs, but also their derivate works and respective histories, facilitating the effortless browsing of web-native musical content.

Music recommendation [22][20] is an automated method to give users the opportunity to encounter unfamiliar but potentially interesting songs. Songrium also assists such user activities by visualizing the Web of Music. Kamalzadeh reports 50% of active listeners would like to choose songs one after another [13]. Furthermore, just 9% use online recommendation and 10% use shuffle when listening to a collection. This result indicates that active listeners enjoy not only listening songs but also choosing songs. On this point, visualiz-

ing the Web of Music can provide an excellent experience for active listeners, having a complementary relationship with music recommendation.

Some researchers have also attempted to investigate linked music data. Jacobson proposed music similarity ontology [11], while Raimond [19] and Kolozali [15] proposed an ontolgy to describe a process of creating music. The purpose of these studies was to construct a Web of Data for music. On the other hand, our present research aims to investigate and assist the existing Web of Music.

6. CONCLUSIONS

In this paper, we proposed a new music browsing assistance service called *Songrium* that visualizes *VOCALOID* music including original songs and their derivative works on the video sharing site *Niconico*. Our target music content was *web-native music*: music content that was born, listened to, and distributed on the web. Songrium provides various visualization tools to assist users in grasping the relations among web-native music and also helps them discover new content. Specifically, *Music Star Map* shows audio similarity relations and *Planet View* shows derivative relations. *Bubble Player* shows a history of web-native music content and *Singing Voice Analysis* helps users to find derivative works based on the characteristics of the sung voice.

For future work, we will continue to run the Songrium service and improve it based on user feedback. In this paper, we mentioned only VOCALOID music, but web-native music is available from many other sources, which we hope to exploit in the near future.

7. ACKNOWLEDGMENTS

We thank Keisuke Ishida for the web service implementation of Songrium and Matt McVicar for his helpful comments on earlier drafts of this paper. We also thank anonymous users of Songrium for editing social annotations. This work was supported in part by CREST, JST.

8. REFERENCES

[1] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, 2011.

[2] H. Fujihara, M. Goto, T. Kitahara, and H. G. Okuno. A modeling of singing voice robust to accompaniment sounds and its application to singer identification and vocal-timbre-similarity based music information retrieval. *IEEE Trans. on ASLP*, 18(3):638–648, 2010.

[3] M. Goto. A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, 43(4):311–329, 2004.

[4] M. Goto. A chorus-section detection method for musical audio signals and its application to a music listening station. *IEEE Transaction on ASLP*, 14(5):1783–1794, 2006.

[5] M. Goto and T. Goto. Musicream: Integrated music-listening interface for active, flexible, and unexpected encounters with musical pieces. *IPSJ Journal*, 50(12):2923–2936, 2009.

[6] W. Hall. The ever evolving web: The power of networks. *International Journal of Communication*, 5:651–664, 2011.

[7] M. Hamasaki and M. Goto. Songrium: A music browsing assistance service based on visualization of massive open collaboration within music content creation community. In *Proc. of WikiSym/OpenSym 2013*, pages 4:1–4:10, 2013.

[8] M. Hamasaki, H. Takeda, T. Hope, and T. Nishimura. Network analysis of an emergent massively collaborative creation community: How can people create videos collaboratively without collaboration? In *Proc. of ICWSM 2009*, 2009.

[9] M. Hamasaki, H. Takeda, and T. Nishimura. Network analysis of massively collaborative creation of multimedia contents - case study of Hatsune Miku videos on Nico Nico Douga -. In *Proc. of uxTV 2008*, pages 165–168, 2008.

[10] P. Hamel, M. E. P. Davies, K. Yoshii, and M. Goto. Transfer learning in MIR: Sharing learned latent representations for music audio classification and similarity. In *Proc. of ISMIR 2013*, pages 9–14, 2013.

[11] K. Jacobson, Y. Raimond, and M. Sandler. An ecosystem for transparent music similarity in an open world. In *Proc. of ISMIR 2009*, pages 33–38, 2009.

[12] Z. Juhasz. Low dimensional visualization of folk music systems using the self organizing cloud. In *Proc. of ISMIR 2011*, pages 299–304, 2011.

[13] M. Kamalzadeh, D. Baur, and T. Moller. A survey on music listening and management behaviors. In *Proc. of ISMIR 2012*, pages 373–378, 2012.

[14] H. Kenmochi and H. Ohshita. Vocaloid – commercial singing synthesizer based on sample concatenation. In *Proc. of Interspeech 2007*, pages 4011–4010, 2007.

[15] S. Kolozali, M. Barthet, G. Fazekas, and M. Sandler. Knowledge representation issues in musical instrument ontology design. In *Proc. of ISMIR 2011*, pages 465–470, 2011.

[16] C. C. Marshall and F. M. Shipman. Saving, reusing, and remixing web video: Using attitudes and practices to reveal social norms. In *Proc. of WWW 2013*, pages 885–895, 2013.

[17] E. Pampalk and S. Dixon. Exploring music collections by browsing different views. *Computer Music Journal*, 28(2):49–62, 2004.

[18] E. Pampalk and M. Goto. Musicsun: A new approach to artist recommendation. In *Proc. of ISMIR 2007*, pages 101–104, 2007.

[19] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The music ontology. In *Proc. of ISMIR 2007*, pages 417–422, 2007.

[20] F. Ricci. Context-aware music recommender systems: Workshop keynote abstract. In *Proc. of WWW 2012 Companion*, pages 865–866, 2012.

[21] M. Schedl, C. Hoglinger, and P. Knees. Large-scale music exploration in hierarchically organized landscapes using prototypicality information. In *Proc. of ICMR 2011*, pages 17–20, 2011.

[22] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE Transactions on Audio, Speech & Language Processing*, 16(2):435–447, 2008.