Grand Challenges in Music Information Research

Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
m.goto@aist.go.jp

Abstract

This paper discusses some grand challenges in which music information research will impact our daily lives and our society in the future. Here, some fundamental questions are how to provide the best music for each person, how to predict music trends, how to enrich human-music relationships, how to evolve new music, and how to address environmental, energy issues by using music technologies. Our goal is to increase both attractiveness and social impacts of music information research in the future through such discussions and developments.

1998 ACM Subject Classification H.5.5 Sound and Music Computing, J.5 Arts and Humanities–Music, H.5.1 Multimedia Information Systems, I.5 Pattern Recognition

Keywords and phrases Music information research, grand challenges, music processing, music signals

Digital Object Identifier 10.4230/DFU.Vol3.11041.217

1 Introduction

Music information research is gaining a lot of attention [15, 7, 11, 2]. It has a long history as shown by attempts to use a computer to compose music from the time of the invention of the computer, such as the "Illiac Suite for String Quartet" of 1957. Results from music information research have spread widely throughout society, including synthesizers which have become essential for the production of popular music, and music distribution services over mobile phones. The field of music information research covers all aspects of music and all aspects of people's music activities, and is related to a variety of topics such as signal processing, transcription, sound source segregation, identification, analysis, understanding, retrieval, recommendation, classification, distribution, synchronization, conversion, processing, summarization, composition, arrangement, songwriting, performance, accompaniment, score recognition, sound synthesis, singing synthesis, generation, assistance, encoding, visualization, interaction, user interfaces, databases, annotation, and social tags related to music. The aims of music information research as an academic field are to study mechanisms for

- listening to and understanding music,
- creating and performing music,
- distributing, retrieving, and recommending music,
- communication between people through music, and
- qualities intrinsic to music

from the viewpoints of science (revealing the truth) and engineering (making useful systems).

The importance of music information research was not recognized until the 1990s, however. This was transformed dramatically after 2000 when the general public started listening to music on computers in daily life. It is now widely known as an important research field, and new researchers are continually joining the field worldwide. Although music information research sometimes needed an argument to be recognized as serious research instead of

Multimodal Music Processing. *Dagstuhl Follow-Ups, Vol. 3.* ISBN 978-3-939897-37-8.
Editors: Meinard Müller, Masataka Goto, and Markus Schedl; pp. 217–226
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

research for fun more than a decade ago, such misconceptions become a thing of the past. This change has been caused by the fact that the general public is aware that all music will eventually be digitized, created, distributed, used, shared, etc. There will be further demand for new music listening interfaces, retrieval, and recommendations. Academically, one of the reasons many researchers are involved in this field is that the essential unresolved issue is the understanding of complex musical audio signals that convey content by forming a temporal structure while multiple sounds are interrelated [11, 3, 4, 15]. Additionally, there are still appealing unresolved issues that have not been touched yet, and the field is a treasure trove of research themes.

This paper discusses some grand challenges that could further increase both the attraction and social impacts of music information research in the future. Please note that some discussions in this paper are intentionally provocative to trigger controversial discussions and stimulate new ideas.

2 Grand Challenges

How can music information research contribute to building a better world and making people happy? How can it contribute to solving the global problems our worldwide society faces? This paper discusses some grand challenges that could be tackled by music information research and could also convince the general public that this research has social impacts for a better, sustainable world and is really important for enriching their lives.

2.1 Can music information research provide music or music videos optimized to the individual?

The goal here is to provide the best music for each person by generating or finding appropriate context-aware music. Music preferences vary from person to person, and even the same person may want to listen to different music (or watch music videos) depending on their situation or mood. If it is technically possible to automatically generate (compose) optimal songs or select such songs from a huge lineup of existing music according to such preferences, situations, or moods, people could not stop using this technology that always provides super happiness and joy. This would have a big impact on society, though such a technology would be controversial if people are really addicted to it. In order to achieve this, technology that is able to understand music and music videos in the same way people do is important. Current technology is not able to do this in terms of

- the ability of understanding people's preferences and situations,
- the quality of automatically generated music or music videos,
- the accuracy of music selection (retrieval and recommendation), or
- the depth of automatic music understanding.

However, there is room for discussion regarding whether a completely automated system is the best. For example, an approach of making an interactive system that assists people's activities is also appealing [6].

Because it is difficult to automatically generate new music or music videos from scratch, the music or video provided could be *2nd generation (secondary or derivative) content*. In the 2nd generation content, musical elements and ideas of existing music or video called *1st generation (primary or original) content* [16] are reused in the creation of new songs. Even 3rd, 4th, or *N-th generation content* can be considered by reusing the generated content again and again as shown in Figure 1. Satoshi Hamano named this style of content creation *N-th order derivative creation*. The reuse or customization of existing music might be a more

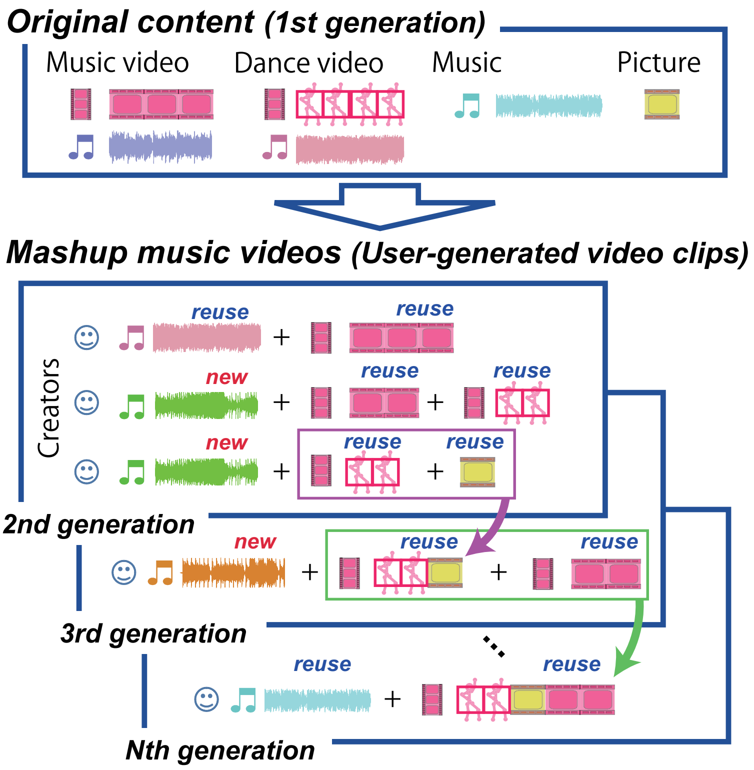

Figure 1 Generation of mashup music videos (user-generated music video clips) by reusing existing original content [16].

natural approach for the future. For example, "mash-ups" that ingeniously combine and mix different songs, and "touch-ups" (customizations) that modify or customize elements of existing songs (changing the timbre, phrase, and volume balance of singing voices and musical instruments) [6] are important in the discussion of music creation. In fact, in recent years, there have been increasing activities to intentionally provide songs and elements so that other people can use them for the N-th order derivative creation [14, 8].

2.2 Can music trends be predicted?

The goal here is to predict music trends by predicting hit songs to cause or prevent a "music pandemic". Is it technically possible to predict hit songs? Alternatively, is it technically possible to provide reasons why a song is not selling? There are actual studies on "hit song science" [17], but technology that is able to predict global or local trends with a high level of precision has not yet been achieved. Prediction of trends is difficult to derive from only the content of music, and it is necessary to globally and exhaustively incorporate information on the Web as social information in order to achieve results that could not be achieved using only technology for analyzing and understanding audio signals [19].

Putting aside the pros and cons of the surveillance society aspect, such trend prediction would become more feasible if it were possible to obtain a worldwide history of what kinds of music everyone is listening to. That is, through the further spread of music distribution technology, it will become possible to record the history of all music playback and sharing this while maintaining anonymity. By making it possible to record the history of what individuals listen to using automatic song identification technology even in live performances [20, 12, 18], it is likely that it will be possible to predict music trends with a high level of precision. However, at the same time, it is interesting to speculate whether, once it becomes possible to provide music optimized for individuals, diversification of music will accelerate and trends will become less likely to occur (thus preventing a "music pandemic"), or people will want to hear what others are listening to, resulting in huge trends (as if a "music pandemic" had occurred).

2.3 Can the relationship between people and music be made richer?

The goal here is to enrich human-music relationships by reconsidering the concept of originality. Digitizing all music from past to present will enable humankind to instantly access all music for the first time ever. Moreover, music will continue to accumulate. The number of accessible songs has monotonically increased, and the number of musical pieces registered on flat-rate music distribution services such as Napster 2.0 has reached 15 million musical pieces. Access will become even easier in the future with progress in music information retrieval and recommendation technology. This itself is historically inevitable, and is desirable as it will make people's music lifestyles more convenient. However, music information research holds the key to whether this will eventually enrich the relationship between people and music.

In the past, new artists needed to try to ensure that their songs were not buried among all the other songs that were on the market, but in the future, it may become even more difficult to get people to listen to music because it is buried among an enormous number of all songs from the past to the present. Moreover, once it becomes possible to automatically compute similarities with all past songs in terms of partial elements such as melody, lyrics, chord progression and arrangement, it will become clear that all songs may have similar aspects to other songs. This is because all creations are affected by other works on a subconscious level. In some cases, it may be technically possible to point out past songs that are partially similar to a song that has just been created. It will be interesting to see how this transforms copyright concepts, and the concept of the originality of music may need to be reconsidered.

So does this mean that human will be unable to overcome the music of the past and lose the will to create new music? Will new music no longer be needed? I don't think so. Essentially, the important things about music are not originality and copyright, but rather how it inspires and makes people happy, and its overall appeal and quality as a work of art. Furthermore, the joy of expression itself is another driving force behind music creation. We may see the arrival of an era in which we go back to the origin of music in a time when it could only be enjoyed in a live concert without the ability to record as more emphasis is placed on using music to bring joy and pleasure to people "here and now." Technological advances could bring about a new music culture that is more centered on emotional, touching experiences.

2.4 Will music information research bring about the evolution of music itself?

The goal here is to push new music evolution forward by enabling new music representations to emerge or enhancing human abilities of enjoying music. The emergence of new technology has already created new musical expressions. This will inevitably continue to create new musical expressions in the future. For example, automatic pitch-correction technology of vocals is already being used on a routine basis in the production of commercial music (popular music, in particular). It has become an absolute necessity for correcting pitch at points in a

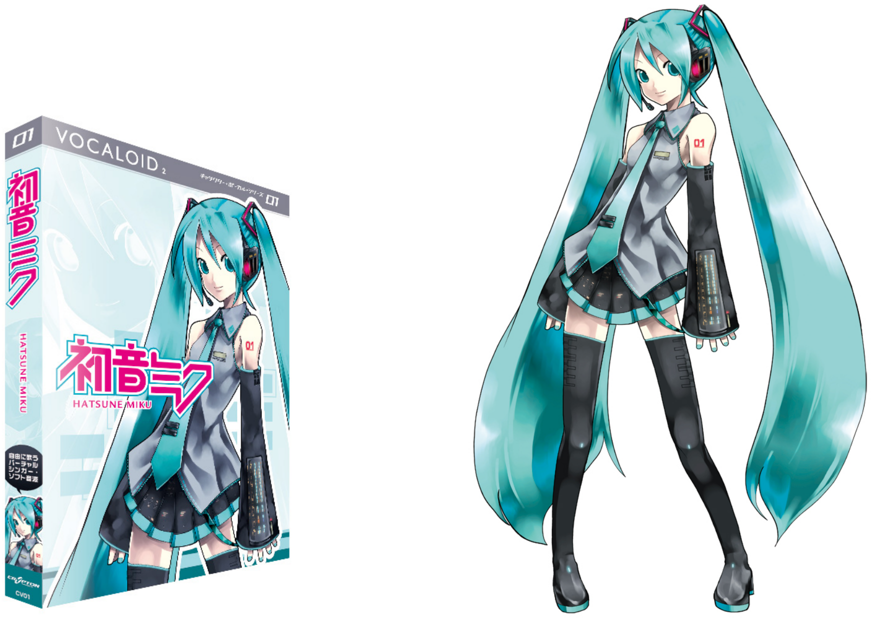

Figure 2 Singing synthesis software *Hatsune Miku* with a cute synthesized voice and an illustration of a cartoon girl. (Courtesy of © CRYPTON FUTURE MEDIA, INC.)

song where the singer is less than skillful and for making corrections to achieve a desired effect. Furthermore, since 2007, singing synthesis technology represented by Yamaha's VOCALOID [10] has gained much attention in Japan. Both amateur and professional musicians have started to use singing synthesizers as their main vocals, and songs sung by computer singers rather than human singers have become popular and are now being posted in large numbers on video sharing services likes *Nico Nico Douga* (`http://www.nicovideo.jp/video_top/`) in Japan and *YouTube* (`http://www.youtube.com`). Even compact discs featuring compilations of songs created using singing-synthesis technology are often sold and appear on popular music charts in Japan [9]. In particular, *Hatsune Miku* [14, 1] is the name of the most popular software package based on VOCALOID and has a cute synthesized voice with an illustration of a cartoon girl as shown in Figure 2. Although Hatsune Miku is a virtual singer, she has already had live concerts with human musicians in Japan, USA, and Singapore (Figure 3). As music synthesizers generating various instrumental sounds are already widely used and have become indispensable to popular music production, it is historically inevitable that singing synthesizers will become more widely used and likewise indispensable to music production. Initially, synthesizers could easily be distinguished from the sound of natural instruments, and this itself led to the creation of unique expressions, but now the quality is high enough that they cannot be differentiated by the general public, and they are used in the majority of popular music. There is no reason that the same will not happen for singing synthesis. The only uncertainty is how soon this will happen.

I hypothesize that the complexity of music created by humankind as audio signals is monotonically increasing. However, there is a limit to the complexity the general public finds enjoyable, and increases in complexity using the approaches of contemporary music have had difficulty in gaining popularity. I believe that the "mash-ups" mentioned earlier hold one of the keys to the next evolution of music from this perspective. Mash-ups are a music production technique in which multiple songs (or their components such as only

Figure 3 Live concert by Hatsune Miku at MIKUNOPOLIS 2011 [13] in Los Angeles, USA on July 2nd, 2011. (Courtesy of © CRYPTON FUTURE MEDIA, INC. and © MIKUNOPOLIS 2011)

the vocals or accompaniment) are used as material to be mixed together and combined as if they were parts of the same song from the beginning. By referring to the musical memory already in the mind of the listener, these mash-ups are able to raise the level of complexity acceptable for enjoyment while retaining popularity. In the days when there were no electronic instruments, it was only possible to create music based on units of single notes (individual instrumental notes) on a musical score, but advances in technology have made it possible to produce music using musical fragments of several bars (one phrase) as units or loop material. Mash-ups are musical productions using whole songs as units or material, making it easier to achieve complex audio signals that would be inconceivable when creating a song from scratch. From the viewpoint of listeners, on the other hand, when the songs used as material to be mixed together are already in the memory, they can enjoy songs that would normally be too complex to enjoy.

Has the tempo of music also monotonically increased throughout the history of humankind? If that is the case, the same song would be shorter in length if the tempo were increased, and the number of songs that can be listened to per unit of time can be expected to increase. This is convenient for the “era of access to an enormous number of songs” mentioned above. If that is the case, how fast can songs be made to still be enjoyed by the human brain? Furthermore, what kinds of technologies can be used to assist and train this? It is intriguing whether the human hearing and capability of the brain are able to keep pace and improve when the tempo is systematically increased by 5 BPM[1] every year by (worldwide) laws. I know this idea is especially provocative, but it is worth thinking about the evolution of music in a think-outside-the-box way.

[1] BPM (Beats Per Minute) is a unit indicating the tempo of a performance based on the number of beats in a minute.

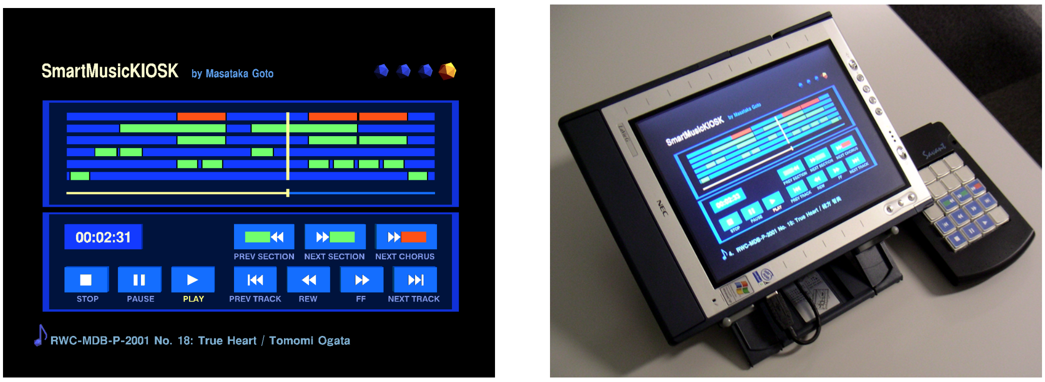

Figure 4 SmartMusicKIOSK screen display. This ia a music listening station with a chorus-search function. The lower window presents the playback operation buttons and the upper window provides a visual representation of a song's contents. A user can actively listen to various parts of a song while moving back and forth as desired on the visualized song structure (the "music map" in the upper window).

2.5 Can music information research contribute to addressing environmental issues and energy issues?

The goal here is to contribute to solving the global problems our worldwide society faces. Environmental issues can be addressed by contributing to a reduction in the use of resources through efforts to increase online music that eliminates the need for physical media (tapes, records, CDs and DVDs). Advances in technology have brought us to an era in which "music" as packaged media can be seen as "information" not affected by physical media, but physical media are still being distributed. Just as overwhelming convenience brought about the transition from record distribution to CD distribution, overwhelming convenience is required for the transition from distribution of physical media using many environmental resources to the distribution of information. Music understanding technology is one means of providing this convenience, and convenience is expected to be improved in various ways such as in "Active Music-Listening Interfaces" [6]. For example, SmartMusicKIOSK [5], an Active Music-Listening Interface with an automatic chorus-section detection technology enables automatic visualization of song structure to listen to parts that are interesting (Figure 4).

With regard to the energy issue, music can be considered a form of high-quality entertainment that does not require much energy. The resources and energy required for music production is less than for production of motion pictures, and will further decrease significantly through the spread of digital music production environments. The *N-th order derivative creation* and mash-ups that reuse (or "recycle") existing songs are also positioned as energy-efficient ways to produce music, and the development of technologies to assist such production is vital. Furthermore, music can be listened to repeatedly, and it is possible to listen to the same song many times. In fact, repeated listening essentially enables the listener to notice a song's appeal. Listening support such as the "Active Music-Listening Interfaces" [6] mentioned above, which enable a deeper understanding of existing music, can contribute to this. Furthermore, advances in music distribution technology will lower distribution costs, and if it is possible to listen to only one's preferred music thanks to advances in music information retrieval and recommendation technology, energy spent on music one is not interested can be reduced. This could be named as "energy-conscious" music production and appreciation. "Music happiness per energy" can thus be increased.

3 Conclusion

In the future, what will be necessary to further increase the appeal of music information research in addition to addressing the above grand challenges?

First, we must develop technology that contributes to building a better world and making people happy, and is essential for society. We hold the key to the creation of a mentally rich future society, and it is vital that academia and industry are seriously engaged in interaction and mutual development. We would need to further discuss what should be done to contribute to the advancement of the music industry and the creation of new industries, and how a contribution can be made to the future of music production and music appreciation.

Second, the importance of our research field must be emphasized as must the further understanding that additional investment in research and development is required. To do this, we must produce researchers who will generate a variety of appealing research results of the highest quality, and also make an effort to talk about our dreams for the future. Such activities will lead to large projects and a diversity of funding, and it would be good to promote great advances in research with a variety of financial backing.

Third, we must promote the field of music information research much more, and make it easy for anyone to feel comfortable participating in it. I would like to expand the research field as a whole, to make possible exciting results from a more diverse range of research.

This paper was written with the aim of contributing the three points above, and I hope that such discussions will continue to be active throughout the field as a whole. However, this must not involve moving in the direction of creating the shell around "music information research" and becoming stuck inside it. What is necessary is a range of activities that span boundaries between fields and reorganize learning from a broader perspective. The field of music information research is also expected to make great strides such as merging with spoken language processing and image processing. I look forward to what the future holds in ten years time.

References

1 Cabinet Office, Government of Japan. Virtual idol. In *Highlighting JAPAN through images*, volume 2, pages 24–25, 2009.
http://www.gov-online.go.jp/pdf/hlj_img/vol_0020et/24-25.pdf

2 Michael Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4):668–696, 2008.

3 Masataka Goto. Music scene description project: Toward audio-based real-time music understanding. In *Proc. of the 3rd International Conference on Music Information Retrieval (ISMIR 2003)*, pages 231–232, 2003.

4 Masataka 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.

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

6 Masataka Goto. Active music listening interfaces based on signal processing. In *Proc. of the 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*, pages 1441–1444, 2007.

7 Masataka Goto and Keiji Hirata. Invited review: Recent studies on music information processing. *Acoustical Science and Technology (edited by the Acoustical Society of Japan)*, 25(6):419–425, 2004.

8 Masahiro Hamasaki, Hideaki Takeda, and Takuichi Nishimura. Network analysis of massively collaborative creation of multimedia contents: Case study of Hatsune Miku videos on Nico Nico Douga. In *Proc. of the 1st international conference on Designing interactive user experiences for TV and video (uxTV' 08)*, pages 165–168, 2008.

9 Hideki Kenmochi. VOCALOID and Hatsune Miku phenomenon in Japan. In *Proc. of the First Interdisciplinary Workshop on Singing Voice (InterSinging 2010)*, pages 1–4, 2010.

10 Hideki Kenmochi and Hayato Ohshita. Vocaloid – commercial singing synthesizer based on sample concatenation. In *Proc. of the 8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*, pages 4011–4010, 2007.

11 Anssi Klapuri and Manuel Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, 2006.

12 Frank Kurth and Meinard Müller. Efficient index-based audio matching. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):382–395, February 2008.

13 Crypton Future Media. Mikunopolis in Los Angeles. `http://mikunopolis.com`.

14 Crypton Future Media. What is the "HATSUNE MIKU movement"? `http://www.crypton.co.jp/download/pdf/info_miku_e.pdf`.

15 Meinard Müller, Daniel P. W. Ellis, Anssi Klapuri, and Gaël Richard. Signal processing for music analysis. *IEEE Journal on Selected Topics in Signal Processing*, 5(6):1088–1110, 2011.

16 Tomoyasu Nakano, Sora Murofushi, Masataka Goto, and Shigeo Morishima. DanceReProducer: An automatic mashup music video generation system by reusing dance video clips on the web. In *Proc. of the 8th Sound and Music Computing Conference (SMC 2011)*, pages 183–189, 2011.

17 Francois Pachet and Pierre Roy. Hit song science is not yet a science. In *Proc. of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 355–360, 2008.

18 Joan Serrà, Emilia Gómez, Perfecto Herrera, and Xavier Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Transactions on Audio, Speech and Language Processing*, 16:1138–1151, October 2008.

19 Malcolm Slaney. Web-scale multimedia analysis: Does content matter? *IEEE MultiMedia*, 18(2):12–15, 2011.

20 Avery Wang. The Shazam music recognition service. *Communications of the ACM*, 49(8):44–48, 2006.