FRONTIERS OF MUSIC INFORMATION RESEARCH BASED ON SIGNAL PROCESSING

Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST), Japan
m.goto [at] aist.go.jp

ABSTRACT

How will music information research open up the future to new ways of enjoying music? Today it is the fate of all music in the world to be digitized, and the era in which we can listen to a large selection of our favorite music anytime and anywhere has arrived. In the past we generally listened to music passively, but in the future we will be able to enjoy music in a more active manner. In this keynote paper, I will discuss how music technologies will open up new possibilities in both music appreciation and music creation.

Index Terms— Music information research, music technologies

1. INTRODUCTION

Music information research has gained in importance academically, industrially, socially, and culturally, thereby attracting an increasing amount of interest. The history of music information research is a long one, as attempts to apply computers to the automatic composition of music began not long after the invention of the computer, the best example being the Illiac Suite in 1957. Research achievements in this field have spread widely through society, such as music streaming services and the sound synthesizers now essential to music production.

In addition, many people have, especially since the start of this century, come to realize that all music will eventually be digitized and will be created, delivered, retrieved, shared, and disseminated in that form. As a result, the importance of music information research has become widely recognized and new researchers throughout the world have begun to enter this field. Indeed, with the growing awareness that the already huge and ever increasing amount of music content must be dealt with in a better way, research on music information retrieval, music recommendation, and new music interfaces has become active.

Another reason for this growing interest in music information research from an academic viewpoint is the difficulty of understanding audio signals that convey content by forming a temporal structure consisting of multiple and interrelated sounds. Automatic understanding of such signals is a fundamental unresolved problem. Other unresolved but intriguing problems that have yet to be touched provide a treasure-trove of future research themes. The number of international conferences and workshops presenting achievements in music information research continues to increase, and given the importance and future possibilities of this field, research projects with sizable budgets have been launched one after another in Europe, the United States, and Japan.

Along with the spread of digital music content in society, industrial applications of music technologies have already expanded in various forms and made contributions to society. Singing information processing [1,2] typified by the synthesis of singing voices has been attracting attention since 2007, and the most surprising change in this regard is the birth in Japan of the world's first *culture that actively enjoys songs with synthesized singing voices as main vocals*, thereby breaking down the long-cherished view that *listening to a non-human singing voice is worthless*. This is a feat that could not have been imagined prior to 2007, when the singing-voice synthesis software *Hatsune Miku* [3, 4] based on *VOCALOID* [5] singing synthesis technology first went on sale. Further advances in music information research may break down other entrenched attitudes about music listening and give birth to another new culture of music appreciation.

In the light of this increasing importance of music information research, this paper presents my ideas on hypotheses and grand challenges in the hope that music information research will make more contributions to academia and society while opening up new possibilities in the future. I have already discussed five challenges in my article "Grand Challenges in Music Information Research" [6]. Here, in Section 4 I purposely take up more controversial topics that have much room for debate. Although some of them are not very technically feasible at this point in time, I hope that our research community will foster new ideas by freeing ourselves from current attitudes about music.

2. FUTURE POSSIBILITIES WITH MUSIC INFORMATION RESEARCH

Music information research holds a variety of possibilities for the future. Each and every researcher carves out a future that he or she believes in, which results in various research approaches that could be classified, for example, from the following viewpoints:

- "Analysis and understanding" versus "generation and creation"
- "Signal processing" versus "symbol processing"
- "Real-time and interactive" versus "non-real-time and non-interactive"
- "Creating a system" versus "learning about people"
- "Dealing with one song" versus "dealing with a set of songs"
- "Dealing with the song itself" versus "dealing with metadata about a song"
- "Communication by performing music" versus "meta-level communication while listening to music"
- "Professional-oriented" versus "amateur-oriented"
- "Automatic" versus "manual or semiautomatic"

In thinking about the future direction of this research field, in this paper I focus on the following two important approaches related to the last of the above viewpoints.

978-1-4799-2186-7/14/$31.00 ©2014 IEEE

(1) *Research on augmenting human abilities* (achieving "musical ability support")

In addition to supporting training processes enabling people to develop their musical abilities, this approach will include support for temporarily augmenting (enhancing) human abilities. For example, a person who does not have the ability to compose could produce a musical piece by using a composition support system, or a person who is not good at singing could produce a main vocal by using a singing synthesis system. Even electronic musical instruments (sound synthesizers) could also be considered as support for augmenting the musical ability of a person who cannot play a certain musical instrument. I will discuss "musical ability support" in more detail in Section 3.

(2) *Research on automating musicians' abilities* (achieving a "virtual musician")

This approach attempts to realize on a computer various abilities possessed by human musicians. Research themes targeted by this approach include automatic composition, automatic arrangement, automatic generation of lyrics, automatic performance (automatic playback of music files, automatic improvisation, and performance rendering), automatic accompaniment, automatic recognition of sheet music (optical music recognition), and automatic transcription. This approach may achieve human abilities by imitating them or may achieve functions that simply exceed human abilities. In the automatic control of musical instruments, for example, renderings that cannot be performed by humans are frequently used regardless of actual human ability. On the other hand, automatic performance can also refer to efforts at automatically generating an improvised performance in the manner of a jam session or at performing a certain musical score in a human-like manner. In Section 4 I will continue this discussion and elaborate on the future direction of the virtual musician concept.

Approaches (1) and (2) above are not independent and just differ in their focus. If an ability achieved by a virtual musician is used with the aim of supporting a person without that ability, it then becomes a technology in the category of "musical ability support" in approach (1). For example, the automatic composition system *Orpheus* [7] that enables a user to automatically compose a song by entering lyrics and selecting provided options would obviously be classified as approach (2) but if viewed as a technology for augmenting the ability of a human to compose music would be classified as approach (1).

It is interesting to investigate how much contribution a person has to make in order to feel that the process is not totally automatic and that he or she is simply being supported. What is necessary for a person using music technologies to be able to think "I am engaging in self-expression by using my own abilities that have simply been augmented"? If it were simply a matter of pushing a button, it would be difficult to think of the results obtained as self-expression no matter how much the technology might be referred to as content-creation support.

3. AUGMENTING HUMAN ABILITIES

As described in "Research on augmenting human abilities' in Section 2, a variety of approaches have come to be taken with the aim of augmenting abilities and overcoming limitations through the power of technology. I think that the enhancement of human abilities can take several directions, which I classify as follows:

(a) *Training*: Improve human ability itself

(b) *Support with an immediate effect*: Temporarily improve human ability through the power of technology

(b-1) *Make anyone a musician*: Empower a person lacking musical abilities

(b-2) *Give a musician outstanding ability*: Make a musician with ability even better

"Training" here corresponds to providing people with the means of continually or permanently improving their own abilities through technology. Examples of this type of support include vocal training [8–10] and performance training [11–17]. Although human ability can be improved through steady effort and practice even without technology-based support, this research aims to improve human ability in a more efficient and appropriate manner.

"Support with an immediate effect," in contrast, corresponds to the enhancing of some kind of musical ability in a person for the time that technology-based support is being provided even if that person's ability itself will not be fundamentally changed. In other words, this type of support aims to empower people only for the time that the technology in question is being used — there is no need to wait for an improvement in human ability. Depending on who is being empowered and to what extent, I divide this type of support into "Make anyone a musician" and "Give a musician outstanding ability."

In "Make anyone a musician," it is assumed that the user is an ordinary person without musical training. Research in this area aims to use the power of technology to make easy what is usually difficult for that person. Efforts in this regard include composition support in a way that reflects the intent of the user in some form [7, 18–21], performance support that enables the user to perform as desired through operations on a computer even if that person cannot play a musical instrument [12,22–25], and vocal support that synthesizes or processes singing voices [4, 5, 26, 27]. This approach targets even trained musicians who lack certain abilities and would like to receive support for them.

In "Give a musician outstanding ability," however, the user is assumed to be a trained musician and is enabled to give musical expressions that are difficult for an ordinary musician. A musician whose abilities have been enhanced through computer support could be expected to seek out new forms of expression. If we were to simply think of this as "support for outstanding performances," then even a high-speed mechanical performance by synthesizers would be this kind of support, as would physically augmenting the body control of a musician by using some devices.

In reality, there could be a "Training" effect when one is receiving "Support with an immediate effect," and there could be an intermediate support between making anyone a musician and giving a musician outstanding ability. So these kinds of support can also take on various overlapping forms.

The enhancement technologies discussed above were focused on augmenting human abilities in relation to "music generation" as in music production and music performance. In the following, however, I introduce some examples of technologies enhancing human abilities in relation to "analysis and understanding" as in music appreciation: "Augmented Music-Understanding Interfaces" (Section 3.1) and the "OngaCREST Project: Similarity-Aware Information Environment" (Section 3.2), both of which my colleagues and I have been working on. Various researchers are also conducting studies on other means of augmenting human abilities.

3.1. Augmented Music-Understanding Interfaces

I proposed a research approach called *Augmented Music-Understanding Interfaces* [28, 29] that aims to improve one's ability to listen to music. These interfaces are examples of the "Make anyone a musician" approach described in this section.

The goal of this approach is to enrich music listening experiences by deepening each person's understanding of music. Music listening experiences depend on music-understanding abilities. Although the music-composing/performing abilities of musicians are often discussed, the music-understanding abilities of casual listeners have not been discussed or studied much. It is difficult, for example, to define music-understanding abilities and express the results of understanding. In addition, it is difficult to know how others understand music. Music listening is usually an individual experience, and it is impossible to directly observe how others understand elements in music. Similarly, one often does not notice what one does not understand when listening to music. Methods that listeners can use to better understand music or improve their ability to understand music have not yet been established and will have to be discovered.

3.1.1. Examples of Augmented Music-Understanding Interfaces

My colleagues and I have therefore pursued a research approach of building *Augmented Music-Understanding Interfaces* [28, 29] that facilitate deeper understanding of music by using automatic music-understanding technologies based on signal processing. The following are examples based on this approach.

- *SmartMusicKIOSK*: Music listening station with a chorus-search function [30,31]

  SmartMusicKIOSK is a music playback interface that visualizes the entire song structure consisting of chorus sections and repeated sections to enable within-song browsing of popular music and trial listening of songs. It augments the ability of the user to understand music structure and musician intent.

  It can, for example, provide a person who usually listens to music without any concern for music structure an opportunity to listen to music while being aware of that structure. In addition, by enabling the user to listen to the chorus sections of a song in succession, it lets the user obtain a more accurate understanding of how lyrics and arrangement change for each repetition of the chorus (as a reflection of musician intent).

- *LyricSynchronizer*: Automatic synchronization between music and lyrics [32]

  LyricSynchronizer is a user interface that displays scrolling lyrics with the phrase currently being sung highlighted in a manner similar to that used in *karaoke* machines. It augments the ability of the user to understand the lyrics of a song.

  For example, LyricSynchronizer enables a person who usually listens to songs without concern for the lyrics to view a song's lyrics synchronized with playback or to click on a word that appears interesting and listen from that word. In this way, the user pays attention to a song's lyrics while listening to it, and this may provide the user with a better understanding of the message conveyed by the lyrics.

- *INTER*: Musical instrument equalizer for music audio signals [33]; *Drumix*: Music playback interface with a real-time drum-part editing function [34]

  INTER is a music playback interface that allows the user to adjust the volume of each track in a song corresponding to a different musical instrument. This enables remixing of the song by equalizing in units of musical instruments. *Drumix* is a music playback interface that allows the user to edit the drum part of a song in real time during playback just as if another drummer were performing a different drum pattern. These interfaces augment the ability of the user to listen for and understand different parts corresponding to individual musical instruments.

  For example, a person who is usually unaware of the drum pattern or timber of drum sounds in a song can use Drumix to edit those drum characteristics, which should help that person distinguish between bass-drum sounds and snare-drum sounds and learn how drum patterns can have a great effect on the feeling conveyed by a song. We believe that interfaces for such casual experience of derivative creation can be associated with content-creation support technologies and contribute to an "age of billions of creators."

These interfaces were originally developed as *Active Music Listening Interfaces* [35] enabling people to enjoy music in active ways but were later found to be good examples of Augmented Music-Understanding Interfaces. The SmartMusicKIOSK interface can be experienced as part of the active music listening service *Songle* (http://songle.jp) [36,37].

3.1.2. Discussion

The following functions provided by the interfaces described above are important in augmenting music understanding:

- Visualization of a song's content (musical elements)
- Synchronization of information display with music playback
- Provision of interactive interfaces for controlling music playback and musical elements

Visualization of music content plays an important role in augmenting (temporarily supporting) people's understanding of music because *understanding is deepened through seeing*. And *music touch-up* (personalization or customization by making small changes to elements in existing music) [35], like those made using Drumix, also helps music understanding because *understanding is deepened through editing*. It lets users naturally observe why music is composed and arranged in a certain way while casually enjoying the content modification.

We also found that interfaces able to make users aware of what they tend to not notice are useful. User interaction enhances immersive music listening experiences and promotes a deep appreciation of music.

3.1.3. Toward Music-Understanding Ability Training Interfaces

The above Augmented Music-Understanding Interfaces can, from the viewpoint of music appreciation, be classified as a "Make anyone a musician" approach to providing "support for *temporarily* augmenting the ability to listen to music" in Section 3. The next step will be to develop interfaces that *permanently* improve one's ability to understand music, as it is improved in the "Training" described in Section 3. Although there are many ways to improve one's ability to understand a foreign language, such as through language schools and training materials, there are virtually no systematic means of improving music-understanding abilities. Most music schools and training materials including music-dictation training are intended for musicians and creators rather than casual listeners.

I therefore proposed the research approach *"Music-Understanding Ability Training Interfaces"* [29] as an important

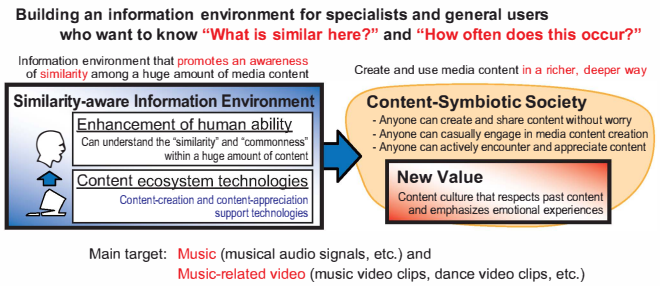

Fig. 1. Overview of OngaCREST Project.

future direction towards enabling a greater number of people to enjoy music in more depth from more diverse views. The more people there are who understand how delicious good food can be, the more likely it is that delicious food will be prepared. In the same way, the more people there are who understand the attraction and magic of music, the more likely it is that attractive music will be created. Augmenting and improving music-understanding ability is important to that end.

3.2. OngaCREST Project: Similarity-Aware Information Environment

Since judging music similarity is an essential ability if one is to understand music, I introduce (as an example of the "Support with an immediate effect" approach described in Section 3), a five-year research project, named *OngaCREST*, intended to provide various types of support for music creation and appreciation based on the implementation and enhancement of this similarity judging ability. The OngaCREST Project is officially entitled *"Building a Similarity-aware Information Environment for a Content-Symbiotic Society,"* and is now in progress as a fiscal-year-2011 selected research project (Research Director: Masataka Goto) in the research area of *"Creation of Human-Harmonized Information Technology for Convivial Society"* of the CREST (Core Research for Evolutional Science and Technology) funding program provided by the JST (Japan Science and Technology Agency).

As shown by the project overview in Figure 1, this project aims to build a similarity-aware information environment in which people are able to know similarities among vast amounts of media content so as to foster a content culture that respects past content and gives greater importance to emotionally touching experiences. Furthermore, by developing technologies that can open up future possibilities in content creation and appreciation, we aim to promote a society in which anyone can engage in content creation and appreciation. The main types of media content targeted in this project are music and music-related videos.

3.2.1. Limits of human ability in judging similarity

The amount of digital content that can be accessed by people has been increasing and will continue to do so in the future. This is desirable, but unfortunately makes it easier for the work of content creators to become buried within a huge amount of content and makes it harder for viewers and listeners to select content. Furthermore, since the amount of similar content is also monotonically increasing, creators will be more concerned that their content might invite unwarranted suspicion of plagiarism. All kinds of works are influenced by existing content and it is difficult to avoid the unconscious creation of content partly similar in some way to prior content.

However, human ability with regard to similarity is limited. Judging similarity between two things in front of one's eyes is usually a simple task, but the speed of judging similarity is limited — searching for similar content among a million items of content cannot be done for all practical purposes. Moreover, while humans are able to make accurate judgments based on past experience, their ability is limited in judging "commonness" as in determining the probability that some event out of all others will occur. When an uncommon event happens to be frequently observed recently, for example, people tend to wrongly assume that it is likely to occur. And when a frequent event happens to not be encountered, people tend to wrongly assume that it is rare. Consequently, with the coming of an "age of billions of creators" in which anyone can enjoy creating and sharing works, the monotonic increase in content means that there is a growing risk that one's work will be denounced as being similar to someone else's. This situation could lead to a society that makes it difficult for people to create and share content without worry.

3.2.2. Information environment enabling people to use similarity and commonness

In light of the above, we started the OngaCREST Project to build an environment in which specialists and general users alike can know the answers to the questions "What is similar here?" and "How often does this occur?" Here we aim to make it possible for people to continue creating and sharing content without worry. Furthermore, we want to make it easy for anyone to enjoy the content creation process by developing *content-creation support technology* enabling "high commonness" elements (such as chord progressions and conventional genre-dependent practices) to be used as knowledge common to mankind. We also want to promote a proactive approach to encountering and appreciating content by developing *content-appreciation support technology* that enables people to encounter new content in ways based on its similarity to other content.

We hope to contribute to the creation of a culture that can mutually coexist with past content while paying appropriate respect to it. This will become possible by supporting a new music culture that enables creators to take delight in finding their content being reused in much the same way that researchers take delight in finding their articles being cited. We feel that the value of content cannot be measured by the extent to which it is not similar to other content — pursuing originality at all costs in content does not necessarily bring joy to people. Fundamentally, content has value by inducing an emotional and joyous response in people. We would like to make it a matter of common sense that content with emotional appeal and high-quality form has value. In fact, we would like to see conditions in which it is exactly the referring to many works that gives content its value, similar to the situation with academic papers. Through this approach, we aim to create a content culture that emphasizes emotionally touching experiences.

In the digital content society, future content is apt to be overwhelmed by the huge amount of ever-increasing past content but without being forgotten. The OngaCREST Project takes up the challenge of creating a content-symbiotic society capable of rich, sustainable development in a "cannot-be-forgotten society" brought about through digitization. Our aim here is to enable people to feel the symbiosis between past and future content and to create a society that can enjoy a huge amount of content through a symbiosis between people and content.

As described above, the OngaCREST Project can be positioned as an endeavor to enhance and support human ability in judging similarity and creating and appreciating content, all through the power of technology. Our achievements to date in

relation to the content-appreciation support technology described above include the web-based active music listening service *Songle* (http://songle.jp) [36, 37] and music browsing assistance service *Songrium* (http://songrium.jp) [38,39].

4. WILL COMPUTERS CREATE AND APPRECIATE MUSIC IN THE FUTURE?

The virtual musician concept presented in Section 2 under the "Research on automating musicians' abilities" heading will cover both "automatic creation" and "automatic appreciation." In the following, I discuss the possibility of a future in which computers can automatically create and appreciate music. At this point in time, this is considered a grand challenge that may be somewhat divorced from reality, if not science fiction.

4.1. Automatic creation

First consider "automatic creation." One of many future possibilities is that it might be possible to totally automate the generation of musical pieces and even lyrics through the power of technology and that this automation will give birth to a culture in which people actively enjoy such music. Although many researchers have already researched automatic composition, I myself could not stop feeling that songs I usually listen to should be composed by human beings. However, since the traditional attitude that "listening to a non-human singing voice is worthless" was broken down, it is just a matter of time before technologies break down the prejudice "listening to a non-human composition is worthless."

When we enjoy a song whose main vocal is generated by a singing synthesis technology on existing video-sharing services like *Niconico* (`http://www.nicovideo.jp/video_top/`) and *YouTube* (`http://www.youtube.com`), it is already impossible to tell from the sound whether the creator of the song is a human being or a computer. In the past, listeners could easily detect human intervention from human vocals even if the sounds of all the musical instruments could be synthesized with such high quality that non-specialists could not tell the difference from human performances. However, now that synthesized singing voices have come to be accepted for main vocals, we often enjoy music in an environment in which such human intervention is no longer obvious and we cannot notice even if all the videos posted on a video-sharing service are created by a completely automatic system.[1]

Such an environment holds great potentials. For example, it is conceivable that people could listen to a posted work with music (lyrics, singing, performance) automatically generated by computer and make time-synchronized user comments without noticing the automated aspect of that work. Then, on the computer side, algorithms could automatically observe the number of plays, analyze the content of those comments, and use that information as a basis for making improvements. This capability could therefore make it possible to repost a work any number of times while making improvements resulting in the generation of even better music.

How would we as human beings respond if computers began singing in the true sense of the term? That is, how would we respond if a computer were in some sort of state that could be associated with emotion and began to express itself through singing as a form of self-assertion? Additionally, if we were to be uncomfortable with the words "automatic generation of music," what would we think if the mechanism of such automatic generation was actually a "virtual collaboration" between people and computers based on N-th generation of derivative creation from a huge number of Consumer Generated Media (CGM) works posted by people? In other words, how would we react if the number of songs created by a huge number of users within a CGM culture were to reach a massive number and new attractive songs came to be automatically generated by computer on the basis of the results of automatic music understanding targeting those songs?

[1] In actuality, though, there are still no complete technologies for totally automating even non-singing tasks like composition/performance and mastering in a human manner. So human intervention could also be detected from that evidence.

4.2. Automatic appreciation

The above discussion assumes the audience, at least, to be human beings. However, could computers be an audience as well? Let's consider a future in which "automatic appreciation" is a possibility.

At present, the research on automatic music understanding by computer is often concerned with describing musical elements and phenomenon (music scene descriptions [40, 41]) expressed by musical audio signals, classifying music by genre, and tagging pre-trained labels associated with emotions and other characteristics. Such research has produced, for example, the *MusicCommentator* system [42] that automatically generates comments on a musical piece by using machine learning of the relationship between music and time-synchronized user comments made on the *Niconico* video-sharing service. Nevertheless, technology that enables a computer to "appreciate" music in a form that includes preferences and value judgments like a human being has not yet been realized. In short, the feasibility of doing so with current technology is still low, but what would we as human beings think if this became possible?

After creating a musical piece, a human being usually expects other people to listen to it. The creator of a work would of course place value on the act of creating a musical piece regardless of whether someone else listens to it but in general would hope that another human being catches the messages or feelings expressed in the piece. The possibility is therefore high that the creator of a musical piece would not place much value on it if his or her musical piece were listened to by a computer and not people. On the other hand, we cannot rule out the possibility that this traditional attitude of "there is no value in having one's music listened to by a non-human audience" will someday be made obsolete.

Furthermore, in a future in which both automatic creation and automatic appreciation have become technically possible, it is conceivable that computers will function as both creators and appreciators of music. If it cannot be assumed that only people appreciate music, human hearing organs and human power of understanding will no longer limit music expression and the possibility arises that musical expression will be transformed, perhaps to forms with super-fast tempos or extremely complex acoustical expressions. This is interesting from a science-fiction perspective, but in reality a world in which people create and appreciate music should continue even in a future with computers creating and appreciating music. As has happened in the past, new technologies will open up new musical expressions, but those expressions will fall into a range that can be accepted and enjoyed by human beings.

5. ENLARGING THE CONTRIBUTION OF MUSIC INFORMATION RESEARCH TO SOCIETY

The use of music technologies in industry and society will continue to expand as various component technologies advance, and music information research will contribute to society as an essential tech-

nology. In the following, I would like to discuss the questions "What is music-induced deep emotion?" and "How can music information research contribute to human happiness and value enhancement?" as two fundamental human problems. I point out that this discussion will be centered on hypotheticals, but my aim here is to disseminate ideas as a foundation for enabling music information research to make even bigger contributions to society.

5.1. What is music-induced deep emotion?

What kind of deep emotion could be induced by music created by the automatic-creation technologies described in Section 4.1? How would music created through the content-creation support technologies as described in Section 3 be accepted? If such music were simply wonderful, could automatically created music touch people's deep emotion? Or even if the quality of such music were high, would listeners be incapable of having emotionally touching experiences once they found out that the music was a non-human creation?

In this regard, I hypothesize that at least three kinds of deep emotion relevant here.

(i) Content-induced deep emotion
This is deep emotion that is elicited by the content itself: that is, emotion induced without relation to the process used to create that content or to any social or cultural significance that the content may have.

(ii) "Olympics-type" deep emotion
This is deep emotion that is elicited because a fellow human being is the creator of the content in question. For example, given a piece of music that includes incredibly agile playing of some instrument, no emotion at all may be felt if it were known that the piece is automatically performed by a computer, but emotion could arise if a human being were performing the same musical piece in front of the audience. This is analogous to feeling no emotion if a human being and an automobile were to compete in an Olympic event and the automobile reached the finish line first.

(iii) Context-driven deep emotion
This is deep emotion that is elicited when one knows the context in which the content came about. Deep emotion might be induced when one knows the social, cultural, or personal background in which the content was created. When the content was created in an unusual situation, for an unusual reason, or by a person overcoming a difficulty, it could elicit deep emotion. Even the "Olympics-type" deep emotion described above could be interpreted as context-driven deep emotion.

Although here I take up only the above three factors, there are other factors that could contribute to eliciting deep emotion. For example, deep emotion may be elicited by an experience shared with another person when enjoying the content together. While this is similar to context-driven deep emotion, the context in this case is on the side of the audience rather than that of the content.

In the past, it was difficult to divide up deep emotion into the above factors (i)–(iii). Whenever content was presented, it was easy to see who was presenting that content and in what form. Nowadays, however, given the common practice of posting self-created content on a video-sharing service, an anonymous work has a chance to be highly evaluated if it induces deep emotion that depends solely on the content itself. On the other hand, if the content creation process is disclosed, the possibility of Olympics-type deep emotion could increase. If the anonymous work is presented along with some context, context-driven deep emotion might be generated.

In the future, then, how will people's deep emotion be elicited when further advances in technology make automatic creation possible and make it easy for anyone to create and perform music? If people's emotion is elicited only by the content itself, nothing will be changed. On the other hand, if Olympics-type deep emotion is indispensable, people cannot give a high evaluation to a work known to have been created so easily. If, however, people could accept a context in which a certain virtual character (like a virtual idol or computer-generated pop star) creates a musical piece, both content-induced deep emotion and context-driven deep emotion may be triggered.

A similar discussion has arisen with our singing synthesis technology called *VocaListener* [43], which can easily synthesize a natural singing voice by having the user simply sing (singing-to-singing synthesis). Given a song produced by such natural singing synthesis, it is interesting to note that there are people who would give the song a high evaluation solely on the basis of the content itself as well as people who would give the song a poor evaluation simply because they knew the song was created by the power of technology[2]. There may also be cases in which such a song using VocaListener could be highly evaluated by people who feel context-driven deep emotion after learning of our many years of research activities trying to contribute to music culture and industry.

5.2. How can music information research contribute to human happiness and value enhancement?

One of the goals that the field of music information research should be aiming for over the long-term is to contribute to human happiness through the power of music technologies. In this regard, I hypothesize that "happiness" is determined by the derivative (slope) of a person's psychological perception of reality and not by that person's absolute amount of resources. This hypothesis originates from observations and introspection and has its limits, but even someone having only a small amount of resources can feel happy if the derivative of his or her psychological perception of reality is positive. Conversely, even someone having a large amount of resources cannot feel happy if the derivative of his or her psychological perception of reality is negative. "Sustainable development" is important to a society since it is then easy to make the value of this derivative positive; if the society is only "sustainable," the derivative is zero.

5.2.1. Enhancing value per unit resource

I believe that our new concept of "happiness per unit energy" (happiness ÷ energy) will become important in the future. Media content like music is indispensible to leading an emotionally enriching life; having a means of obtaining happiness and a sense of fulfillment is directly connected to the quality of life. In particular, music content is high-quality entertainment requiring a minimal amount of energy while being resistant to repeated listening — the amount of resources and energy used in the production of music have dropped with the proliferation of digital content production environments. In this regard, N-th generation of derivative creation can be viewed as an effective means of condensing the good points of multiple items of existing content and increasing happiness. There is also an aspect of recycling (reusing) content here, which means that 2nd generation (secondary or derivative) content can be interpreted as a good

[2] When VocaListener is used, however, the creator has much leeway in adjusting and manipulating musical expression. The more people know about VocaListener, the more probable it is that Olympics-type deep emotion will be triggered.

and energy-efficient means of producing content as an unintended consequence. Research-and-development efforts devoted to increasing "happiness per unit energy" (i.e., reducing energy consumption while increasing happiness) should become increasingly important in the years to come.

This "happiness per unit energy" can be extended to "value per unit resource" as a more general concept. In the past, only time and cost (including those of human resources) were usually considered as unit resources and value enhancement was focused on improving productivity. In today's world, however, it is important that we consider energy in addition to time and cost while trying to increase our happiness.

5.2.2. Facilitating contributions by music information research

In the information society, one's psychological perception of reality can be greatly affected not just by physical space but also by information space. There is therefore much room for music technology (or information technology on the whole) to make a contribution. It is easier in information space than in physical space to create conditions that can avoid resource competition or conflict, and as a result, technology that aims to enhance value here holds great possibilities. Optimizing or maximizing value, however, is not necessarily our true goal here. If we were to actually succeed in maximizing value, value could then only decrease and the derivative of one's psychological perception of reality would be negative. Furthermore, attempting to maximize value often requires a great amount of resources.

Consequently, if the above ideas are correct, answering the question "How can people be made to continually feel over the long term that the derivative of their psychological perception of reality is positive?" becomes a challenge that must be faced. Although a psychological perception of reality corresponds to no physical quantity, an illusion of such may be created and effective. If we take this perspective when supporting people by using music technology, an approach different from that in the past may appear. For example, in the "Training" described in Section 3, a positive derivative can be felt over the long term because ability increases gradually. Even if there were a technology that could instantly enable a person the best in the world at some ability, it would not make that person happy because after that the derivative would no longer be positive. In the "Support with an immediate effect" approach described in Section 3, the technology will enhance ability only temporarily, but its temporariness would be beneficial since a positive derivative in feeling an improvement in ability can be obtained again and again as many times as the user quits using and then reuses that technology (however, we can see the importance of making it difficult to sense that the derivative has become negative at the instant that the user quits).

Additionally, given the ongoing, monotonic increase in the volume of content, we have to consider methods for achieving a state in which the derivative of the psychological perception of reality is positive for new content creators. For example, venues for presenting content currently take the form of nationwide or worldwide conventions, but it could be beneficial here to adopt the approach taken in the sports world, where both national and local contests are held. This is because a new creator at a national convention may have to compete with many other people of different levels of ability in terms of content quality, which raises the possibility that the creator's content will be buried (resource competition easily occurs). A local convention, in contrast, is apt to feature many creators of similar ability, which can foster an environment of friendly competition in which a creator's content can be appreciated without being buried (resource competition hardly occurs). Popularization of such local conventions should therefore have a positive effect for new creators. Such a step-by-step process of improvement has the possibility of inducing in creators a long-term feeling that the derivative of their psychological perception of reality is positive.

I do not claim that all of the ideas I have presented above are new, as I can imagine that there are others who have made similar hypotheses. That being said, the importance of music information research to the happiness of mankind is unchanged.

6. CONCLUSION

With the hope that music information research will advance to even higher levels and open up new possibilities in the future, I have ventured to present some stimulating discussions in this paper. Although I have focused on singing synthesis, Augmented Music-Understanding Interfaces, and the OngaCREST Project as examples to advance my arguments, there are many other research topics that can be discussed in the same context. I expect music information research based on signal processing to be expanded dramatically by the diverse contributions of many researchers. In particular, cooperation and in-depth discussions among researchers of various communities are essential for human happiness and for value enhancement. I hope to see further developments in music information research through a melding of diverse fields such as signal processing, machine learning, web service, human-computer interaction, speech processing, hearing, psychoacoustics, and image processing.

Acknowledgments: I thank (in alphabetical order by surname) Hiromasa Fujihara, Katsutoshi Itoyama, Tomoyasu Nakano, Hiroshi G. Okuno, and Kazuyoshi Yoshii, who have worked with me to build the systems presented in this paper. This work was supported in part by CREST, JST.

7. REFERENCES

[1] M. Goto, T. Saitou, T. Nakano and H. Fujihara, "Singing information processing based on singing voice modeling," in *Proc. of ICASSP 2010*, pp. 5506–5509, 2010.

[2] M. Goto, "Singing information processing," in *Proc. of the 12th IEEE International Conference on Signal Processing (IEEE ICSP 2014)*, 2014.

[3] Crypton Future Media, "What is the HATSUNE MIKU movement?." `http://www.crypton.co.jp/download/pdf/info_miku_e.pdf`, 2008.

[4] H. Kenmochi, "VOCALOID and Hatsune Miku phenomenon in japan," in *Proc. of the First Interdisciplinary Workshop on Singing Voice (InterSinging 2010)*, pp. 1–4, 2010.

[5] H. Kenmochi and H. Ohshita, "Vocaloid — commercial singing synthesizer based on sample concatenation," in *Proc. of Interspeech 2007*, pp. 4009–4010, 2007.

[6] M. Goto, "Grand challenges in music information research," in *Dagstuhl Follow-Ups: Multimodal Music Processing* (M. Muller, M. Goto and M. Schedl, eds.), pp. 217–225, Dagstuhl Publishing, 2012.

[7] S. Fukayama, D. Saito and S. Sagayama, "Assistance for novice users on creating songs for Japanese lyrics," in *Proc. of ICMC 2012*, 2012.

[8] D. M. Howard and G. F. Welch, "Microcomputer-based singing ability assessment and development," *Applied Acoustics*, vol. 27, pp. 89–102, 1989.

[9] T. Nakano, M. Goto and Y. Hiraga, "MiruSinger: A singing skill visualization interface using real-time feedback and music CD recordings as referential data," in *Proc. of ISM 2007 Workshops (Demonstrations)*, pp. 75–76, 2007.

[10] D. Hoppe, M. Sadakata and P. Desain, "Development of real-time visual feedback assistance in singing training: a review," *Journal of Computer Assisted Learning*, vol. 22, pp. 308–316, 2006.

[11] Y. Takegawa, T. Terada and T. Tsukamoto, "Design and implementation of a piano practice support system using a real-time fingering recognition technique," in *Proc. of ICMC 2011*, 2011.

[12] C. Oshima, K. Nishimoto and N. Hagita, "A piano duo support system for parents to lead children to practice musical performances," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 3, no. 2, 2007.

[13] S. Ferguson, A. Vande Moere and D. Cabrera, "Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians," in *Proc. of the 9th International Conference on Information Visualisation*, pp. 97–102, 2005.

[14] J. Yin, Y. Wang and D. Hsu, "Digital violin tutor: An integrated system for beginning violin learners," in *Proc. of ACM Multimedia 2005*, pp. 976–985, 2005.

[15] D. Fober, S. Letz and Y. Orlarey, "VEMUS - feedback and groupware technologies for music instrument learning," in *Proc. of SMC 2007*, 2007.

[16] M. Sadakata, D. Hoppe, A. Brandmeyer, R. Timmers and P. Desain, "Real-time visual feedback for learning to perform short rhythms with expressive variations in timing and loudness," *Journal of New Music Research*, vol. 37, no. 3, pp. 207–220, 2008.

[17] T. Knight, N. Bouillot and J. R. Cooperstock, "Visualization feedback for musical ensemble practice: A case study on phrase articulation and dynamics," in *Proc. of Visualization and Data Analysis 2012*, 2012.

[18] F. Pachet, "The Continuator: Musical interaction with style," *Journal of New Music Research*, vol. 32, no. 3, pp. 333–341, 2003.

[19] F. Pachet and P. Roy, "Markov constraints: steerable generation of Markov sequences," *Constraints*, vol. 16, no. 2, pp. 148–172, 2011.

[20] M. McVicar, S. Fukayama and M. Goto, "AutoRhythmGuitar: Computer-aided composition for rhythm guitar in the tab space," in *Proc. of ICMC SMC 2014*, pp. 293–300, 2014.

[21] M. McVicar, S. Fukayama and M. Goto, "AutoLeadGuitar: Automatic generation of guitar solo phrases in the tablature space," in *Proc. of the 12th IEEE International Conference on Signal Processing (IEEE ICSP 2014)*, pp. 599–604, 2014.

[22] T. M. Nakra, "Synthesizing expressive music through the language of conducting," *Journal of New Music Research*, vol. 31, no. 1, pp. 11–26, 2002.

[23] J. Patten, B. Recht and H. Ishii, "Interaction techniques for musical performance with tabletop tangible interfaces," in *Proc. of ACE 2006*, 2006.

[24] S. Jordà, G. Geiger, M. Alonso and M. Kaltenbrunner, "The reactable: Exploring the synergy between live music performance and tabletop tangible interfaces," in *Proc. of the 1st International Conference on Tangible and Embedded Interaction*, pp. 139–146, 2007.

[25] T. Baba, M. Hashida and H. Katayose, "VirtualPhilharmony: A conducting system with heuristics of conducting an orchestra," in *Proc. of NIME 2010*, pp. 263–270, 2010.

[26] K. Saino, H. Zen, Y. Nankaku, A. Lee and K. Tokuda, "An HMM-based singing voice synthesis system," in *Proc. of Interspeech 2006*, pp. 1141–1144, 2006.

[27] T. Nakano and M. Goto, "VocaRefiner: An interactive singing recording system with integration of multiple singing recordings," in *Proc. of SMC 2013*, pp. 115–122, 2013.

[28] M. Goto, "Augmented music-understanding interfaces," in *Proc. of SMC 2009 (Inspirational Session)*, 2009.

[29] M. Goto, "Music listening in the future: Augmented Music-Understanding Interfaces and Crowd Music Listening," in *Proc. of the AES 42nd International Conf. on Semantic Audio*, pp. 21–30, 2011.

[30] M. Goto, "SmartMusicKIOSK: Music listening station with chorus-search function," in *Proc. of UIST 2003*, pp. 31–40, 2003.

[31] M. Goto, "A chorus-section detection method for musical audio signals and its application to a music listening station," *IEEE Trans. on ASLP*, vol. 14, no. 5, pp. 1783–1794, 2006.

[32] H. Fujihara, M. Goto, J. Ogata and H. G. Okuno, "LyricSynchronizer: Automatic synchronization system between musical audio signals and lyrics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1252–1261, 2011.

[33] K. Itoyama, M. Goto, K. Komatani, T. Ogata and H. G. Okuno, "Instrument equalizer for query-by-example retrieval: Improving sound source separation based on integrated harmonic and inharmonic models," in *Proc. of ISMIR 2008*, pp. 133–138, 2008.

[34] K. Yoshii, M. Goto, K. Komatani, T. Ogata and H. G. Okuno, "Drumix: An audio player with functions of realtime drum-part rearrangement for active music listening," *IPSJ Journal*, vol. 48, no. 3, pp. 1229–1239, 2007.

[35] M. Goto, "Active music listening interfaces based on signal processing," in *Proc. of ICASSP 2007*, 2007.

[36] M. Goto, K. Yoshii, H. Fujihara, M. Mauch and T. Nakano, "Songle: A web service for active music listening improved by user contributions," in *Proc. of ISMIR 2011*, pp. 311–316, 2011.

[37] M. Goto, J. Ogata, K. Yoshii, H. Fujihara, M. Mauch and T. Nakano, "PodCastle and Songle: Crowdsourcing-based web services for retrieval and browsing of speech and music content," in *Proc. of the First International Workshop on Crowdsourcing Web Search (CrowdSearch 2012)*, pp. 36–41, 2012.

[38] 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 the 9th International Symposium on Open Collaboration (WikiSym + OpenSym 2013)*, pp. 1–10, 2013.

[39] M. Hamasaki, M. Goto and T. Nakano, "Songrium: A music browsing assistance service with interactive visualization and exploration of a web of music," in *Proc. of the 23rd International World Wide Web Conference (WWW 2014)*, pp. 523–528, 2014.

[40] M. Goto, "Music scene description project: Toward audio-based real-time music understanding," in *Proc. of ISMIR 2003*, pp. 231–232, 2003.

[41] 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*, vol. 43, no. 4, pp. 311–329, 2004.

[42] K. Yoshii and M. Goto, "MusicCommentator: Generating comments synchronized with musical audio signals by a joint probabilistic model of acoustic and textual features," in *Proc. of ICEC 2009*, pp. 85–97, 2009.

[43] T. Nakano and M. Goto, "VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation," in *Proc. of SMC 2009*, pp. 343–348, 2009.