R&D Spirits

Musical Knowledge Programming for a New Form of Music Distribution

Keiji Hirata
Distinguished Researcher
NTT Communication Science Laboratories

The music paradigm is changing from one-way supply from a handful of people to two-way supply among a large number of people. Computer power is enabling anyone to become a musician by making it easy to create and use melodies. What possibilities does such mass-produced and mass-consumed music offer for digital communications? We put this question to Keiji Hirata, a Distinguished Researcher at NTT Communication Science Laboratories known for his unique research on music information-processing systems.

Toward a new communication medium enabling anyone to use music as simply as words

—Dr. Hirata, what research theme are you currently working on?

In a few words, I would say "music information processing." More specifically, I am working on building "musical knowledge programming" as a new paradigm for music information processing to enable music data to be freely manipulated on a computer. Ever since the invention of the phonograph by Thomas Edison, the established norm has been for music produced by a small minority of artists to be enjoyed by a large number of ordinary people. That, however, is not the only way in which music can be enjoyed. I think that many people would like to move over to the music-making side but their lack of musical knowledge prevents them from doing so. I have been researching this problem with an eye to solving it by adopting a "collage metaphor" and using knowledge representation technology. The most recent version of this scheme is called Music Resonator, which I presented at this spring's open house (**Fig. 1**). The name "Music Resonator" is based on the concept of "resonant communication," proposed by NTT group, and what I want to convey with this name is the idea of resonant interaction between people through music.

—What kind of system is Music Resonator?

Music Resonator lets you search through specialized databases for existing music that you would like to make use of and perform various operations on that music to create original musical arrangements on a client. In short, it allows you to perform a series of semi-automatic processes including retrieval of music, extraction and synthesis of desired fragments, and arrangement of new pieces. Though the system is still in the prototype stage, my aim for the final form is to provide a set of operations that users, even those with no musical knowledge, can react to in an intuitive manner. Frankly speaking, musical arrangement by conventional computer systems is still far from this point in terms of musical quality. With a view toward future business possibilities, I want Music Resonator to be capable of producing high-quality sounds attractive to the young generation.

—What is the purpose of this research? Is it producing something useful?

I would like to make music the third medium of communication after text and images. Music can be used, for example, as ring tones for cell phones, as background music for e-mail, and as a signature for entries in a web log (blog). Of course, even current

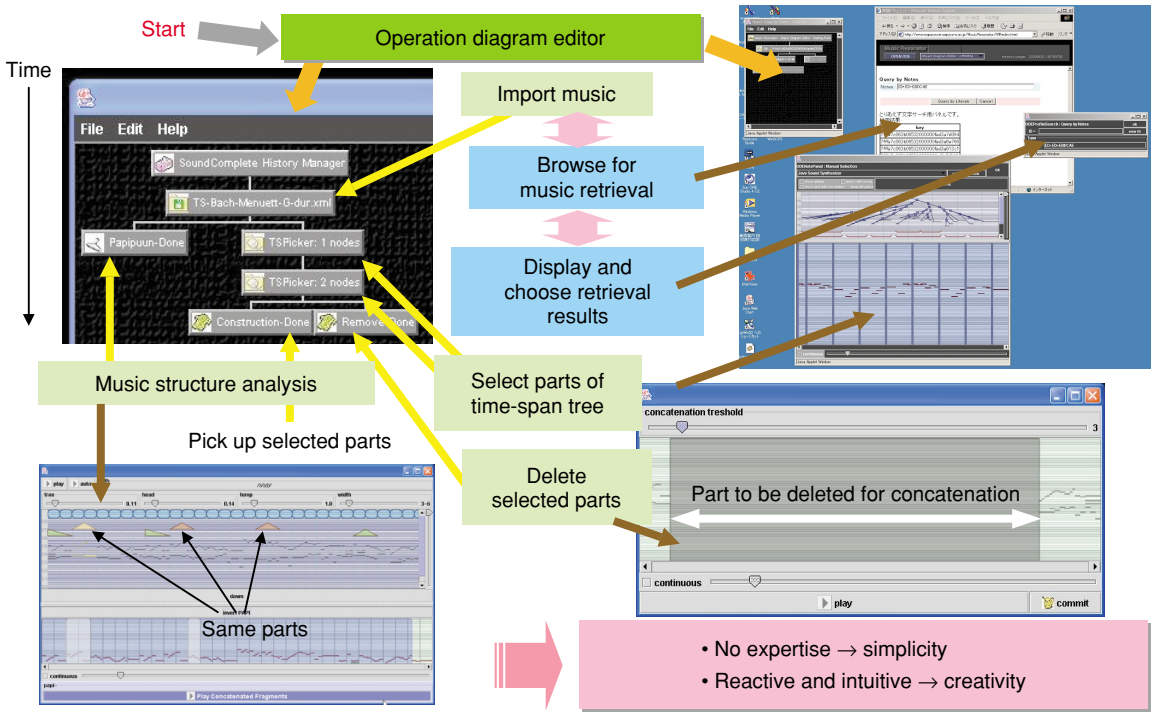

Fig. 1. Music Resonator.

technology could be used to achieve such applications in themselves. But what I would like to aim at is not simply music attached to something but rather music that users themselves mass produce- and mass-consume as easily as they do words and images. My goal here is to make self-expression and communication by means of computers and networks more rich, colorful, and profound.

Annotation technology for converting musical meaning into data

—What are some of the technical features of Music Resonator?

As I mentioned, Music Resonator is based on the metaphor of a collage. This term usually refers to an artistic composition made of various materials (such as paper, cloth, or wood) glued onto a surface, but it also means an assembly of diverse fragments. Music Resonator consists of four main processes: retrieval, selection, transformation, and concatenation (**Fig. 2**). For all of these processes, I use annotation technology, which I think is the main feature of Music Resonator (**Fig. 3**). There are two kinds of musical information: 1) superficial information such as interval, duration, and loudness represented by staff notation and 2) underlying information that indicates musical meaning and music structure. In the same way that a sentence has a structure consisting of a subject, verb, and other elements, rather than being just a sequence of letters, each of the notes making up a musical piece has meaning. This meaning must be understood if a computer is to be used to retrieve and compare musical pieces. I succeeded in doing this by integrating superficial information and deep information using annotations.

At the retrieval stage, for example, when looking for a musical piece having a phrase similar to phrase A, a computer would not be able to determine similarity solely on the basis of information provided by staff notation. This would be possible, however, if the structure and meaning of notes could be attached in the form of annotations that could then be compared. Likewise, in the selection stage, a wrong fragment might be obtained if a phrase is not segmented at an appropriate boundary position. To prevent this, I use annotations to enable phrase boundaries to be identified. In the transformation stage, we can consider two types of operations: “interpolation” for creating a melody between two melodies, say, melody A and melody B, and “imitation” for mimicking an existing

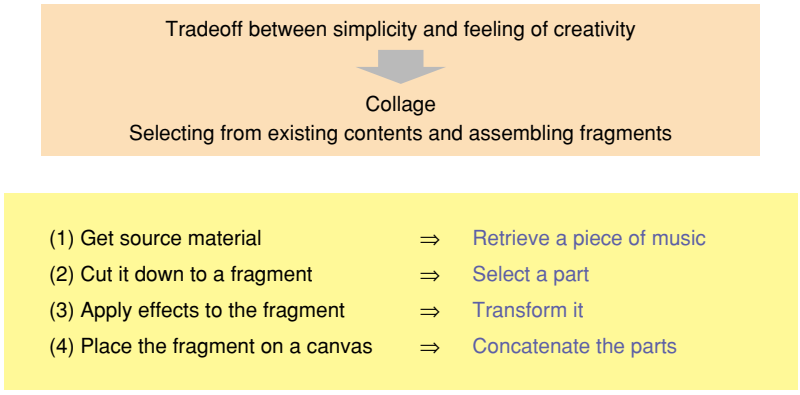

Fig. 2. Using the collage metaphor.

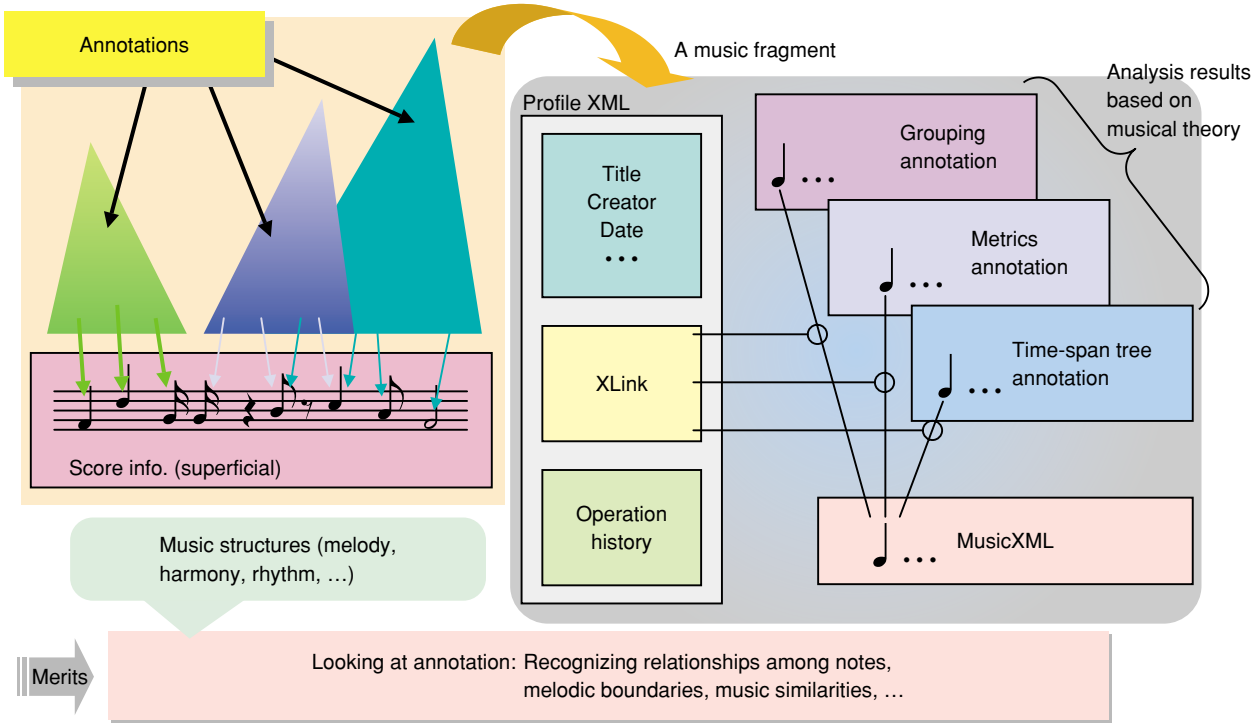

Fig. 3. Musical annotation in Music Resonator.

melody. Music Resonator does not yet include an imitation algorithm, but I plan to incorporate one based on annotations.

Finally, in the concatenation stage, I use annotations in a mechanism for making a natural juxtaposition of phrases. When connecting two phrases, the idea here is to select a part of an existing musical piece that has annotations similar to those of the phrases to be concatenated and to borrow a note sequence from the part. When attempting to make a computer do things that human beings do unconsciously, the logic behind each action must be considered. This was extremely difficult to do for data related to human emotion, as in the case of music. However, I succeeded, and I believe that overcoming this problem through the use of annotation technology is one major achievement of my research.

—What has been your biggest breakthrough up to now?

To handle music on a computer, musical data must be treated in much the same way as numbers. This was quite difficult in the past, but has become possible in this research. This, if anything, is a breakthrough. I won't go into the details here, but let me say that the very useful properties possessed by data expressed by annotations enabled this technology to be developed.

—What form do you think your research will take in three and five years?

Once Music Resonator reaches a practical level, I believe it could be used to provide services in a form that combines various media, not just music. From time immemorial, human beings have been using a variety of communication media including characters, pictures, and sound in a composite form. To appreciate a certain painting, for example, one might read an explanation of that painting's period or listen to music that was popular at that time. Thus, networks of information and contents that cover multiple media have been prepared. Amazon.com is a good example of such a network. This site shows information about other items bought by customers who bought a certain CD or book. It also shows recommendation lists describing items in the same category as an item searched for. I think that this kind of information network might be useful for developing business with the system.

An evolving system to be linked with text and images

—Dr. Hirata, is similar research going on in Japan or overseas?

Technology for processing only superficial musical information is being researched in various forms, but I may be the only one in the world proposing a method for "calculating" music much like numbers. In this regard, I am confident that my research is unique.

—What has been the reaction to your research results?

People that have read my papers have all responded with comments like "difficult" and "hard to understand." Moreover, the system does not use much annotation data yet, which means that I cannot perform the demonstrations that I have in mind. Therefore, I could hardly say that people have a good understanding of my research. To help rectify this situation, I am expending considerable effort on creating and archiving annotation data. Of course, I want the system itself to be appealing, but I also want to accumulate enough annotation data so that many people will be able to enjoy the pleasure of using music as a means of communication. The work is actually slow-going and a bit overwhelming at times—like counting the number of sand grains on a beach. But once this data exceeds a certain critical mass, I expect users to begin exchanging data on their own, resulting in a data pool that grows rapidly.

—Are you collaborating with other research institutions?

In my current research, I am working with two people outside NTT Laboratories. The first is Professor Katashi Nagao of Nagoya University. He has been helping me with the extensible markup language (XML) and annotation technologies, but his main research theme is natural language processing, and he has also been involved with automatic translation systems. My other research partner is Shu Matsuda, who has mainly been writing the program for the system. He is the technical director of a software house called Digital Art Creation while also being an excellent musician, and in these capacities, he has contributed much to the system through programming and musical arrangements.

—What issues do you expect to face from here on?

From a technological perspective, I want to make a system that is as complete and extensive as possible, but I also need to determine how to demonstrate the usefulness of this system, or in other words, how the system can be used to create business opportunities. As I mentioned at the beginning, the current norm for music creation and consumption is for a relatively small number of artists to create works to be enjoyed by a large number of ordinary people. There is no business model in which money would be paid for music easily created by ordinary people. That is a problem. For me to make the fruits of my research truly useful, I need to consider not just music-information processing technology, but also a framework whereby people will pay for and make use of such

music. Of course, I don't expect to be able to handle everything in such a business venture. For the Music Resonator to evolve, I must consider means of linking with text and images that can support future business possibilities.

Researching "systems that make you think"

—How did you begin researching music information processing?

Well, I've liked music for a long time. My university major was information engineering, and in addition to researching parallel-processing computers for artificial intelligence, I formed a band with fellow students and enjoyed giving performances. I also came to think how wonderful it would be if there were a system that enabled people to create and enjoy music even if they could not read sheet music or play musical instruments. I began research on such a system in graduate school at a time when the field of music information processing was not as well recognized as it is now. While I obtained my degree on the basis of my research in computer architecture, I also conducted research on music information processing in parallel.

—What research themes have you been involved with up to now?

After entering NTT, I spent some time studying a field related to the mathematics of computer science and hardly wrote any papers. But my studies from those days became the foundation of my present research. Then, in 1990, I became a member of Japan's 5th Generation Computer Project and was seconded to a research laboratory of the former Ministry of International Trade and Industry (MITI) for three years. I returned to NTT in 1993 and began researching concurrent logic programming languages related to my university research and came to develop a language called "Oda Nobunaga." During this time, I also continued researching music information processing in the background.

My first music system, called "Herbie-kun," was completed in 1996. This system could perform music based on chords. Then, after much trial and error, I researched and developed a system called "Pa-Pi-Pun" in 1999 for performing music based on chord progressions, another called "Ba-Bi-Bun" in 2000 for mimicking other people's arrangements, still another called "Ha-Hi-Hun" in 2001 for adding human-like expression to performances, one called "Papipuun" in 2002 for summarizing a musical piece in a semi-automatic manner, and finally a system called "Sound Complete" in 2003 that was the forerunner to Music Resonator. By the way, my reason for choosing the name "Herbie-kun" for the first system was to pay homage to Herbie Hancock, one of my favorite musicians. You can find details on these systems on my personal home page. I urge anyone interested in this field to take a look. (http://www.brl.ntt.co.jp/people/hirata/)

—Dr. Hirata, what have you tried to achieve through research?

On a somewhat idealistic level, I would like to make society better. I believe it is very important that each and every person give more thought to how he or she can help make the world better. For myself, that means providing a situation that makes people think. Specifically, I would like to stimulate the "creative juices" in people. Whenever people attempt to make something that they like, their thinking becomes more concentrated, and the act of exchanging things made in this way gives birth to communication. This is why I would like to provide and disseminate environments for creating various kinds of media and not just music for personal expression. That is my basic goal.

I have been researching music information processing continuously for the last 20 years, but it's only recently that people began to pay attention to it. With the spread of the Internet and with recent personal computers including speakers, a microphone, and support for MIDI (musical instruments digital interface) and MP3 processing, it looks as if the environment needed for music information processing has finally arrived. On the other hand, it leads to the confusion and pressure that I am feeling in this interview, but I hope to be a pioneer in opening the way to this new form of communication media called music.

One researcher's policy: be an interesting, fun-filled person

—How do you think your research will develop in the years to come?

Over the next two years, I hope to complete a system that will bring Music Resonator to a large number of people. The work that I do is essentially basic research, and turning Music Resonator into a busi-

ness will no doubt require the help of NTT Group companies. But first I must achieve a level of quality that can capture the interest of people at those companies. I believe that the quality of music generated by Music Resonator here has already reached a satisfactory level, so one problem is to construct an interface that is easy to use. At this stage of development, moreover, I assume the hardware for running Music Resonator to be the personal computer, but I can easily envision the cell phone as well. Achieving the same functions as those on a PC on a very compact device like a cell phone is a great challenge in terms of the user interface, and I would like to tackle this problem provided that I have energy to spare.

—What are your thoughts on working in NTT Laboratories?

During my student days, NTT Laboratories was admired as a place filled with researchers that were outstanding leaders in their fields. Of course, it is still a very exciting environment filled with many world-class researchers in a wide variety of fields. And it is no exaggeration to say that NTT Laboratories is blessed with great funding power, now more than ever, and that it is also blessed with the ability to collaborate with NTT Group companies. Thus, there is no lack of opportunity to generate novel ideas. At the same time, I don't like the somewhat conservative nature of the laboratories, which results in a lack of understanding of and support for research that might be risky but could hold great possibilities. I think there is some room for improvement in this respect.

—Dr. Hirata, what is your personal policy as a researcher?

To perform research that other people find interesting, you yourself must first be an interesting, fun-filled person. I believe that an interesting person who works on an interesting idea will more than likely produce interesting results. But if you are dull, I don't think your research will take root in the real world. For this reason, I try to keep aware of what people today find interesting and enjoyable to prevent myself from atrophying. For example, I check advertisements on commuter trains very carefully and ask my colleagues to show me what they have inside their briefcases. I think that using one's own method to observe people's lifestyles and changing attitudes is a must for us researchers.

Interviewee profile

■ Career highlights

Keiji Hirata received the D.Eng. degree in information engineering from the University of Tokyo, Tokyo in 1987. He then joined NTT Basic Research Laboratories. He spent 1990 to 1993 at the Institute for New Generation Computer Technology (ICOT), where he engaged in R&D of parallel inference machines. In 1999, he moved to NTT Communication Science Laboratories. He was promoted to Distinguished Researcher in 2001. His research interests include musical knowledge programming and interaction. Dr. Hirata received the Takahashi Award for a paper presented at the annual convention of the Japan Society for Software Science and Technology in 1987 and the IPSJ Best Paper Award from the Information Processing Society of Japan (IPSJ) in 2001. He co-edited and wrote the book chapter "Musical Knowledge Representation on a Computer" in the book "The World of Computers and Music: Foundations to Frontiers" (in Japanese) in 1998, and co-translated the book "Computer Music Tutorial" by Curtis Roads in 2001. He is a member of IPSJ, the Japanese Society for Artificial Intelligence, and the Japan Society for Software Science and Technology.