Music-understanding technologies to realize the full potential of digital music

Proposing New Ways to Enjoy Music for Supporting the Digital Content Industry

Engendering a "Content-Symbiotic Society" in which content can be created and published without hesitation

Media Interaction Group, Information Technology Research Institute

(**Our life and society will change in this way!**)

⬇

Music has been digitized and it is now possible to access a huge number of musical pieces. Technologies by which computers can automatically analyze and understand this vast body of music will make daily appreciation of music more enjoyable and enable anyone to easily create music. This change could lead to a drastic change of the culture around music. AIST's music-understanding technologies will open up new ways to enjoy music in the future and will provide a boost to content industries.

Making computers "understand" music

Masataka Goto, a Prime Senior Researcher of the Information Technology Research Institute, knows that music-understanding technologies are necessary. He explains: "When I was shopping for CDs to encounter unfamiliar but interesting music in my twenties, I would go to a trial-listening corner and repeatedly press the fast-forward button to listen to chorus sections ("hook") of a song. Since then, digital music has become the norm and, realizing that I would not be able to hear everything even if I spent my whole life listening to music, I decided that it would be helpful if computers could do the listening for me." From this point, he started on a quest to work out how computers could be made to understand music in a way that is useful for people.
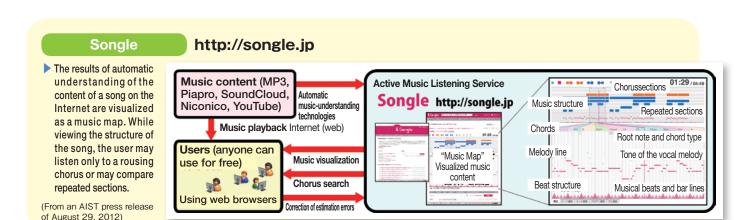
When humans listen to music, they do things like humming along to the melody, clapping in time with the beat, and instantly identifying the chorus. Although these things are easy for humans, it is difficult to implement automatic understanding of music by a computer if its mechanism is not clear. Since music is one of the most complicated signals with multiple sounds being mixed together, existing

technologies were not enough to overcome the difficulties of understanding music. Toward achieving automatic music-understanding technologies, Goto and his colleagues have been researching and developing a wide range of core technologies for over 20 years and have become world leaders in the field.

A web service "Songle" (launched in 2012) enables anyone to easily try some of such technologies. Songle is distinct from ordinary music players in that a user can view the entire structure of a song as a music map that is visualized by using music-understanding technologies. A user can jump and listen to the chorus with just a push of a button, which is useful not only for trial listening but also for comparison between repeated choruses.

Aggregating the power of the crowd to improve usefulness

Songle has applied music-understanding technologies to over 700,000 songs that have already been open to the public on the Internet, and their results can be enjoyed by everyone in the form of music maps. There are some errors in the music maps, but this is to be expected. The reality is that AIST is conducting an important experiment in how to make imperfect technologies useful for society. Songle is provided with a function that enables anyone who can find and fix errors to contribute by making corrections. In this way, Songle can make use of a synergy between the capabilities of computers with their great processing power but imperfect understanding and humans with their ability to spot and fix errors. In other words, we are studying systems of harmonious cooperation between computers and humans that will be critical to society in the future.

Another web service "Songrium" (launched in 2013) uses the automatic music-understanding results from Songle and enables a user to understand relationships between large numbers of songs from a variety of viewpoints. In Songrium, a user can also skip ahead to the chorus and can easily find songs with similar moods to their favorite songs, videos in which people are singing and dancing along with those favorite songs, and the like. In an era in which new original songs are first released on the Internet, such new

Impact in the following fields!

Community life
- Services
- Sports and leisure
- Education

Industry
- IT and Telecommunications
- Electronics
- Services
- Software and content

Songle   http://songle.jp

▶ The results of automatic understanding of the content of a song on the Internet are visualized as a music map. While viewing the structure of the song, the user may listen only to a rousing chorus or may compare repeated sections.

(From an AIST press release of August 29, 2012)

Music content (MP3, Piapro, SoundCloud, Niconico, YouTube)

Automatic music-understanding technologies

Music playback   Internet (web)

Users (anyone can use for free)

Using web browsers

Music visualization

Chorus search

Correction of estimation errors

Active Music Listening Service

Songle   http://songle.jp

"Music Map" Visualized music content

Music structure — Chorus sections
Repeated sections
Chords — Root note and chord type
Melody line — Tone of the vocal melody
Beat structure — Musical beats and bar lines

---

ways to enjoy music are continuing to open up.

Researching singing synthesis technology for creating music

The advance of technology is bringing about a new culture of music. Singing synthesis technology, which has been attracting attention since 2007, has produced, for the first time in history, a culture that can positively enjoy songs with synthesized voices as main vocals. This innovative culture is an expression of Japan's strengths, and AIST is pursuing research to take this culture further forward.

One such singing synthesis technology is called VocaListener ("Bokarisu"). This technology enables the synthesis of singing voices that imitate the way a user sings. The user need only input a sample of their singing into VocaListener, which then controls commercially available singing synthesis software with various voice timbres to naturally synthesize voices that imitate the singing.

In this era of digital music, music-understanding technology is supporting access to the huge body of songs on the Internet, and singing synthesis technology is expanding the scope for enjoyment in creating music. These trends will not be limited to music but will spread to other forms of content, such as videos.

Goto adds that "There is a problem of whether, as the digitalization of content produces a society in which nothing is forgotten, the future might be buried under the huge and continually growing body of content from the past. For listeners, selecting music is getting more difficult, and for creators, their creations can easily just disappear into obscurity. In addition, as more and more similar content appears, the risk that a person's creations will be criticized for being too similar to other pieces increases, and there is a danger that we will end up with a society in which people hesitate to create and publish their own content. By continuing with research and development powered by AIST's core technologies, we are aiming for a "Content-Symbiotic Society" in which everyone can enjoy appreciating and creating music without hesitation while paying due respect to older content."

Songrium   http://songrium.jp

▲ Songrium visualizes various musical relations between music videos on video sharing websites. The user can keep encountering new songs and review a history of old songs.
(From an AIST press release of August 27, 2013)

Toward making singing synthesis more natural

Singing synthesis software, such as Hatsune Miku created by Crypton Future Media, uses VOCALOID, a singing synthesis technology developed by Yamaha. In the past it was difficult to produce natural-sounding singing with this technology but AIST overcame this problem with "VocaListener". Yamaha started to sell a commercial version of VocaListener in 2012.
(From a Yamaha press release of September 18, 2012)

Given the user's singing voice and the lyrics, it can synthesize the singing voice mimicking the user.

Example of the user's singing voice

Lyrics "Tachidomarutoki, mata futo furikaeru"   Analysis

Tone of the voice   Loudness of the voice   VocaListener

Guessing parameters   Time   Time   Natural singing voice like a human

Singing synthesis software A   Synthesis

Singing synthesis software B   Synthesis

Several singing synthesis softwares   Several sound source (singer's voice)   Even by switching either a singing synthesis software, or a sound source (singer's voice), it can always synthesize the singing voice in the same way (tone and loudness of the voice).