THE WORK OF MASATAKA GOTO: HOW TO ENRICH MUSIC UNDERSTANDING AND LISTENING

Gabriel Vigliensoni
Music Technology Area, Schulich School of Music, McGill University
gabriel@music.mcgill.ca

ABSTRACT

The work of Masataka Goto are nowadays focused on how to enrich each person's understanding and enjoying of music using technologies based on signal processing. His methods allows non-musician people to make changes to elements in existing music, converting a passive listening into an active interaction.

1. BIOGRAPHY

Masataka Goto works at the Information Technology Research Institute (ITRI) of the National Institute of Advanced Industrial Science and Technology (AIST), Japan. In this public institution devoted to cutting-edge research in industrial technology toward a sustainable society, he is the leader of the Media Interaction Group. Over the past 17 years, he has received 24 awards, including the Commendation for Science and Technology by the Minister of MEXT "Young Scientist' Prize", DoCoMo Mobile Science Awards "Excellence Award in Fundamental Science", IPSJ Nagao Special Researcher Award, and IPSJ Best Paper Award. Goto has been an active member in the music information retrieval field since 1991. He was the general chair of the ISMIR 2009 together with Ichiro Fujinaga.

2. RESEARCH FIELD

Mr. Goto has worked in 6 interrelated research areas: *Music Audio Signals Understanding*, *Active Music Listening Interfaces* , *Speech*, *Interaction*, *Music Databases*, and *Music Information Retrieval* . Many of them have been developed through several years. The following overview considers the description in the original paper and some add-ons in more recent papers.

2.1 Music Audio Signal Understanding

The work of Goto on *Music Audio Signals Understanding* has been one of his starting points for future development in other areas. Its goal has been to enrich music listening experience by deepening people understanding of music. To achieve this, he has developed novel techniques for automatic music-understanding based on signal processing. His main research projects in the field are *Real-time Music Scene Description System* and *Sound separation system for percussion instruments* . The latter was the first system that performed sound source separation system for polyphonic drum sets, extracting onset time, loudness, and classifying then as different instruments (Goto et al. 1993).

On the other hand, *Real-time Music Scene Description System* has three different signal processing based projects into it. First, in *A real-time F0 estimation of melody and bass lines in musical audio signals* (Goto 1999) he describes a real-time system for estimating the fundamental frequency (F0) of melody and bass lines in complex audio signals containing various other instruments. The method he proposes is called *PreFEst* , predominant-F0 estimation method, and basically estimates the F0 of the most predominant harmonic structure within a limited frequency range of a sound mixture. Thus, the system estimates the relative dominance of every possible F0 by using a probability density function and considers the F0's temporal continuity by using a multiple-agent tracking architecture. The results gave an detection rate of 88.4% for melody and 79.9% for bass lines of 20 seconds long mono audio signals sampled from compact discs of 10 pieces in popular, jazz, and orchestral genres (Goto 2004).

Second, in *A chorus-section detecting method for musical audio signals* (Goto 2002), Mr. Goto developed a method to automatically detects the beginning and end points of chorus and repeated sections in compact-disc recordings of popular music. The system can also detect transposed chorus sections by introducing a similarity measure that enables transposed repetition. To achieve this, a method called *RefraiD* was implemented extracting the 12 dimensional *chroma vectors*, calculating its similarity, organizing and integrating it into groups, and choosing the most probable chorus section. 80% of correct song choruses were achieved with the test set in the experimental implementation.

Finally, the last project in music understanding was first developed by Goto as part of his M.S. thesis. It is called *BTS* or *A real-time beat tracking system for musical acoustic signals*, which is a beat-tracking system that processes acoustic signals of popular music and recognizes temporal positions of beats performed by drums in real-time. It is assumed that the time signature is 4/4 and the tempo remains stable. The system implementation uses four parallel processing techniques to execute heterogeneous processes simultaneously. Thus, a frequency analysis finds notes' onset times for bass and snare drums. A *beat prediction* process using multiple agents is performed to predict the inter-beat-interval, the next beat, and infers the *beat information* (*beat type*, *beat time* and *tempo*). 30 songs were tested and

BTS correctly tracked in real time 27 of them (Goto 1994).

2.2 Active Music Listening Interfaces

All previous work mentioned above using signal processing has been used by Goto to develop interfaces that help *ordinary* people (not professionals) to interactively understand music. *Active music listening* is understand as a way of listening to music through active and enjoyable interactions in that the user listening experience is enriched through visual representations and editing capabilities of music. Active music listening interfaces projects developed by Goto includes: *SmartMusicKIOSK*, *MusicRainbow*, *Cindy*, *LiricSynchronizer*, *Inter*, *Drumix*, *Musicream*, and *MusicSun*.

His most well-known project in the field is SmartMusicKIOSK. It is a playback interface for trial listening in music stores that allows users to skip verses and go direct to choruses. It is based mainly on the *RefraiD system* mentioned above, but additionally shows different parts of a song. All of the system is contained in a tactile display with easy-to-understand graphics. The implementing results in real life situations have demonstrated their usefulness for active listening experience. Furthermore, some applications for *computer-based media players*, *music thumbnail*, and *digital listening station* are proposed by the author as possible key application examples in the future (Goto 2003).

MusicRainbow is a simple interface for discovering similar artists. The graphical user interface place similar artists close to each other in rings with different colors for different types of music. The user can navigate and play the songs with a simple knob and button. In addition, *RefraiD chorus-section detection method* is used to provide a faster navigation through the interface. While the process to compute the similarities between different artist and map them to the circular rainbow is audio-based, the words to label it comes from mining Google via its SOAP interface and extracting filtered words. Finally, based on these words the rainbow is colored (Pampalk and Goto 2006). *MusicSun* is an evolution of *MusicRainbow* that allows the user to change the impact of three different aspects of similarity (*Audio-based similarity*, *Web-based similarity*, and *Word-based similarity*) changing their weights using sliders. The interface was tested with 33 people and although the results in terms of *fun factor*, *interest of future usage*, *quality of the recommendations*, and *learning curve* were very good (over 90% for all of them), improvement were suggested to find better ways to link unknown artists with known artists (Pampalk and Goto 2007).

The rest of the additional active music listening interfaces mentioned above use signal processing techniques to give the user the possibility to interact (visualizing or editing) with: instrument equalization (*INTER*), drum-part sounds editing (*Drumix*), musical pieces unexpected encountered (*Musicream*), virtual dancers driven by beat tracking (*Cindy*), and lyrics automatically sinchronized to CD recording (*LyricSynchronizer*).

2.3 Speech

In terms of speech signal processing, Goto presented two novel techniques—*SpeechSpotter* and *SpeechCompletion*—to help user to control or enter voice commands into a speech recognizer system. While the former enables a user speak to a machine in the midst of a natural human-human conversation, the latter can helps a hesitating user to enter a word or phrase by completing the fragment. Both applications are based on a natural phenomenon that human hesitates by lengthening a vowel (a filled pause), which can be sensed using speech recognition techniques. *Speechspotter* has been tested in an *on-demand information system for assisting human-human conversation* and a *music-playback system for enriching telephone conversation* applications with convenient and robust results without any training (Goto et al. 2001; Goto et al. 2004). In addition, in 2007 Goto presented —as an instance of his research project *Speech Recognition Research 2.0—Podcastle*, a public web service that provides full-text searching of Japanese podcasts on the basis of automatic speech recognition. The application provides the users the possibility to find podcasts that include a search term and read full texts of their recognition results. The system provides a way to gradually increase its performance, usefulness and applicability through correct recognition errors (Ogata et al. 2007).

2.4 Interaction

The first paper published by Goto was on *A distributed cooperative system to play MIDI instruments*, integrating MIDI and LAN protocols. The *Remote Music Control Protocol* was developed as an extension of MIDI, allowing local and remote communication and visualization to play in ensemble (Goto 1992).

Cindy, virtual dancer, was a project that allowed two players (musicians) to interact with each other through music and 3-D computer animation. *Cindy* changed her movements in real-time according to the performance of the musicians. Thus, the musicians not only improvised together hearing their music, but observing the dancer. *Cindy, the virtual dancer* provided them with auditory and visual information to improvise and perform (Goto and Muraoka 1995).

The next step in controlling CG characters was *Virja Session*, a jazz trio session system in which performers could control over distributed workstations the virtual players. As in *Cindy*, the system provided both auditory and visual feedback to enhance and enrich the performance (Hidaka et al. 1996).

2.5 Music Databases

Goto is the chair of the RWC Music Database International Steering Committee. The *RWC Music Database* (Real World Computing database) is the world's first large-scale copyright-cleared music database compiled entirely for research purposes. It contains six original collections in *popular music*, *royalty-free music*, *classical music*, *jazz music*,

music by genre, and *musical instruments sounds*. For all of the pieces, a set of audio signals, standard MIDI files, and text files of lyrics (for songs) is provided. For the musical instruments sounds, they have been recorded at half-tone intervals with several variations of playing styles, dynamics, brands, and musicians. All these material provides a great ground truth to test and research methods and systems with the same common standard (Goto et al. 2002). In addition, an AIST Annotation of the musical pieces was developed in terms of beat structure, melody line, and chorus sections. Cue points have been done to synchronize the standard MIDI files (Goto 2006)

2.6 Japanese research on Music Information Retrieval

In addition to all previous work, Goto has an online repository of Japanese works written in English related to music information retrieval.

3. CONCLUSIONS

The works of Goto on Music Information Retrieval have had an special evolution. He began researching on signal processing of musical signals to automatically extract information of the music scene. Nowadays, he applies this knowledge in the development of interfaces that help human to understand and enrich the music listening experience, understanding that the 21th century poses new ways to listen, perform, record, share and enjoy music.

4. REFERENCES

Goto, M. 1992. A distributed cooperative system for the MIDI control. *Proceedings of the 10th Anniversary International UNIX Symposium*. 161–71.

Goto, M., M. Tabuchi, and Y. Muraoka. 1993. An automatic transcription system for percussion instruments. *Proceedings of the 46th Annual Convention IPS*.

Goto, M., and Y. Muraoka. 1994. A beat tracking system for acoustic signals of music. *Proceedings of the 48th Annual Convention IPS*.

Goto, M. and Y. Muraoka. 1995. Interactive performance of a music-danced CG dancer. *Proceedings of the Workshop on Interactive Systems and Softwares*. 9–18.

Hidaka, I., M. Goto, and Y. Muraoka. 1996. A jazz session system for interplay among all players II. Implementation of a bassist and a drummer. *Information Processing Society of Japan*.

Goto, M., I. Hidaka, and H. Matsumoto. 1999. A virtual jazz session system: VirJa session. *Transactions of Information Processing Society of Japan*. 1910–21.

Goto, M. 1999. F0 estimation of melody and bass lines in real-world musical audio signals. *Information Processing Society of Japan*.

Goto, M., R. Neyama, and Y. Muraoka. 1999. Musical information processing based on remote music control protocol. *Transactions of Information Processing Society of Japan*. 1335–45.

Goto, M. 2001. An audio-based real-time beat tracking system for music with or without drum-sounds. *Journal of New Music Research*. 159–71.

Goto, M., K. Itou, T. Akiba, and S. Hayamizu. 2001. Speech Completion: new speech interface with on-demand completion assistance. *Proceedings of HCI International*. 198–202.

Goto, M. 2002. A real-time music scene description system: a chorus-section detecting method. *Information Processing Society of Japan*. 27–34.

Goto, M., T. Nishimura, H. Hashiguchi, and R. Oka. 2002. RWC music database: popular, classical, and jazz music databases. *Proceedings of 3rd International Conference on Music Information Retrieval*. 287–8.

Goto, M. 2003. SmartMusicKIOSK: Music listening station with chorus-search function. *Proceedings of the 16th annual ACM Symposium on User Interface Software and Technology*. 31–40.

— — —. 2003. A chorus-section detecting method for musical audio signals. *International Conference on Acoustics, Speech, and Signal Processing*. 437–40.

Goto, M. 2004. A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*. 311–29.

Goto, M., K. Kitayama, K. Itou, and T. Kobayashi. 2004. Speech Spotter: On-demand speech recognition in human-human conversation on the telephone or in face-to-face situations. *Proceedings of the 8th International Conference on Spoken Language Processing*.

Goto, M. 2006. AIST Annotation for the RWC Music Database. *Proceedings of the 7th International Conference on Music Information Retrieval*. 359–60.

Ogata, O., M. Goto, and K. Eto. 2007. Automatic transcription for a Web 2.0 service to search podcasts. *Proceedings of the 8th Annual Conference of the International Speech Communication Association*. 2617–20.

Pampalk, E., and M. Goto. 2006. MusicRainbow: a new user interface to discover artists using audio-based similarity and web-based labeling. *Proceedings of the ISMIR International Conference on Music Information Retrieval.*

— — —. 2007. MusicSun: a new approach to artist recommendation. *Proceedings of the ISMIR International Conference on Music Information Retrieval.* 101–104.

Masataka Goto's Home Page. http://staff.aist.go.jp/m.goto/ (accessed November 4, 2009)