Vol.2009-MUS-82 No.2
2009/11/2

デモンストレーション：音楽情報処理の研究紹介 VIII

藤島 琢哉	ヤマハ株式会社
帆足 啓一郎	KDDI 研究所
後藤 真孝	産業技術総合研究所
	＜発表者＞
Fabien Gouyon	INESC Porto
Toshio Modegi	Media Technology Research Center, Dai Nippon Printing Co., Ltd.
Eric Nichols	Indiana University
Meinard Mueller	Saarland University and MPI Informatik
Jianfeng Xu	KDDI R & D Laboratories, Inc.
Olivier Lartillot	Finnish Centre of Excellence of Interdisciplinary Music Research
Chris Cannam	Queen Mary University of London
Yushen Han	School of Informatics, Indiana University
Shimpei Tatsumi	Kobe Univ. Grad. School of Sci.
Amélie Anglade	Queen Mary University of London, Centre for Digital Music
Kazuyoshi Yoshii	National Institute of Advanced Industrial Science and Technology (AIST)
Hiromasa Fujihara	National Institute of Advanced Industrial Science and Technology (AIST)
Kazuhiro Nakadai	Honda Research Institute Japan
Emiru Tsunoo	University of Tokyo
Tae Hun Kim	University of Tokyo
Jean-Julien Aucouturier	Temple University Japan Campus
Yuu Mizuno	University of Tokyo
Naoki Yasuraoka	Kyoto University
Takeshi Mizumoto	Kyoto University

上記発表者のデモンストレーションを通じ音楽情報処理の研究事例を紹介する．

Demonstrations: Introduction of Research on Music Informatics VIII

Takuya Fujishima	Yamaha Corporation
Keiichiro Hoashi	KDDI R & D Laboratories, Inc.
Masataka Goto	National Institute of Advanced Industrial Science and Technology (AIST)
	<presenters>
Fabien Gouyon	INESC Porto
Toshio Modegi	Media Technology Research Center, Dai Nippon Printing Co., Ltd.
Eric Nichols	Indiana University
Meinard Mueller	Saarland University and MPI Informatik
Jianfeng Xu	KDDI R & D Laboratories, Inc.
Olivier Lartillot	Finnish Centre of Excellence of Interdisciplinary Music Research
Chris Cannam	Queen Mary University of London
Yushen Han	School of Informatics, Indiana University
Shimpei Tatsumi	Kobe Univ. Grad. School of Sci.
Amélie Anglade	Queen Mary University of London, Centre for Digital Music
Kazuyoshi Yoshii	National Institute of Advanced Industrial Science and Technology (AIST)
Hiromasa Fujihara	National Institute of Advanced Industrial Science and Technology (AIST)
Kazuhiro Nakadai	Honda Research Institute Japan
Emiru Tsunoo	University of Tokyo
Tae Hun Kim	University of Tokyo
Jean-Julien Aucouturier	Temple University Japan Campus
Yuu Mizuno	University of Tokyo
Naoki Yasuraoka	Kyoto University
Takeshi Mizumoto	Kyoto University

Through the demonstrations by the presenters above, researches in the field of music information processing are exhibited.

Introduction

Takuya Fujishima (Yamaha Corporation)

Since its first appearance in 2004, the demo session in SIGMUS has been an opportunity for researchers to present not only their latest achievement, but also their ongoing development which would require more sophistication for a formal presentation, as well as systems which were already presented and which therefore wouldn't possibly be revisited otherwise. It generously accommodates "last-minute" submissions and even welcomes on-site proposals for demonstrations to encourage researchers to present their latest works in a casual manner.

This time, in conjunction with 10th International Society for Music Information Retrieval Conference in Japan, the demo session aims at facilitation of communication and mutual understanding between two major research communities in this domain: ISMIR and SIGMUS.

19 demonstrations have been accepted and will be presented today. The topics widely range from a novel algorithm design to an application prototype development, reflecting the fertility of this interdisciplinary research domain. Another interesting aspect of this exhibition will be the versatility in the context of the peculiar cultural background each presentation stands on.

The author expects that all the attendees will enjoy intensive discussions through this precious opportunity, to develop further their research interest.

1. RAMA: An Interactive Artist Network Visualization Tool

Fabien Gouyon, Diogo Costa (INESC Porto)
Luis Sarmento(LIACC/FEUP)

We propose a web-based tool for visualizing and interact-ing with networks of music artists, named RAMA (stand-ing for Relational Artist Maps, available at http://rama.inescporto.pt). The data covers hundreds of thousands of artists and three million user-generated tags, collected from Last.fm ' s API. Artist names are pre-sented in connected graphs and main visualization fea-tures concern (i) artist similarities, (ii) artist popularity and (iii) associated tags. Differing from existing artist network visualization tools, RAMA emphasizes com-monalities as well as main differences between artist categorizations (user-defined tags). A number of interac-tive features are also provided to enhance user browsing experiences, and to foster music discovery.

2. High-precision Audio-MIDI Encoder (Auto-F second version) Supporting Vocal Synthesis by MIDI

Toshio Modegi (Dai Nippon Printing Co., Ltd.)

We have been developed a general acoustic signals to MIDI code converter tool " Auto-F ", which enables to playback voice-like signals with a standard MIDI synthesizer by encoding given vocal acoustic signals to MIDI data. Since then, we have been improving temporal analysis precisions, by stretching temporally given waveform signals, in order to produce more realistic vocal signals with a MIDI sound module. In this demo, we can present this improved version of our audio to MIDI encoder tool by presenting either Japanese or English speech MIDI-format samples with a GM standard MIDI software synthesizer.

Moreover, applying this tool, we are developing a novel voice synthesizing system based on harmonically synthesizing musical sounds, which can generate MIDI data and playback voice-like signals with a MIDI synthesizer by giving Japanese plain (kana) texts, referring to our developing phoneme MIDI code database. In this presentation, we propose an algorithm separating a set of 20 consonant and vowel phoneme MIDI codes from Japanese male or female 71-syllable recorded waveforms by logically multiplying a set of selected MIDI converted syllable data. In this demo, we can also present a prototype version of our developing Japanese speech synthesis tool based on a MIDI software sequencer.

3. Cantonese Melody-Composition Assistant

Eric Nichols, Matthew Hurley (Indiana University)

Recently we developed a system to generate musical rhythms to fit a given English text, for use as a composer's or songwriter's assistant. The system works by generating rhythms which match the natural syllabic stresses of English with the metrical structure of the music. In a tonal language such as Cantonese, an analogous idea is to generate melodies — as opposed to rhythms — to fit a given text. The Cantonese dialect of Chinese is especially interesting in this problem because Cantonese relies on at least six different tones (pitch heights and contours). Mandarin Chinese, in contrast, only uses four different tones. It has been shown that Cantonese popular melodies exhibit a strong correspondence between the spoken tonal contour and the contour of the matching melody.

We are currently developing a system to generate melodies matching a given Cantonese text. The user of the system specifies the text as well as additional constraints on the desired melody such as the harmony and phrase structure. The system generates a collection of melodies to fit the text, taking into account the natural tonal contour of the text (as specified in a database of Cantonese pronunciation), in addition to the other constraints.

4. Multimodal Presentation and Browsing of Music

Meinard Mueller (Saarland University and MPI Informatik)
Christian Fremerey, David Damm, Michael Clausen (Bonn University)
Frank Kurth (Frauenhofer-FKIE)

Recent digitization efforts have led to large music collections, which contain music documents of various modes comprising textual, visual and acoustic data. In this demo, we present a multimodal music player for presenting and browsing digitized music collections consisting of heterogeneous document types. In particular, we concentrate on music documents of two widely used types for representing a musical work, namely visual music representation (scanned images of sheet music) and associated interpretations (audio recordings). We introduce a novel user interface for multimodal (audio-visual) music presentation as well as intuitive navigation and browsing. Our system offers high quality audio playback with time-synchronous display of the digitized sheet music associated to a musical work. Furthermore, our system enables a user to seamlessly crossfade between various interpretations belonging to the currently selected musical work. Finally, based on a visual query by marking certain measures in the sheet music, our system is able to identify corresponding passages in audio recordings.

5. Music-driven Dancer using Example Motions

Jianfeng Xu, Koichi Takagi, Akio Yoneyama (KDDI R & D Laboratories, Inc)

Music and dance are two kinds of main entertainment in our life. Moreover, the ability of dancing to the music might be uniquely human being. It is very natural for people to induct the beat instants in both music and dance, resulting in the handclaps in many cases. In this demo, we present an automatic system to generate a dancing motion that is synchronized with a piece of input music. The features of our system include: 1) the generated motion is synchronized with the music; 2) a completely automatic system is provided to ease the usage; 3) user interaction is supported such as dance genre selection and motion intensity control; 4) 3D content is created that can provide free viewpoint. Basically, our system re-orders the selected motion clips in a motion capture (MoCap) database according to the beat and intensity of the input music. In more detail, our system is composed of two parts. In the first part, which is off-line, we re-organize the MoCap database into motion graphs by the motion beat and intensity. In the second part, which is on-line, a dancing motion is generated from the motion graphs to synchronize with the music, where we search an optimized path in the motion graphs by dynamic programming.

6. MIRtoolbox: an innovative environment, on top of Matlab, for music and audio analysis

Olivier Lartillot, U. Jyvaskyla, Petri Toiviainen (Finnish Centre of Excellence of Interdisciplinary Music Research)

MIRtoolbox is a Matlab toolbox dedicated to the extraction of musical features from audio files, including routines for statistical analysis, segmentation and clustering. MIRtoolbox integrates a user-friendly syntax that enables to easily combine low and high-level operators into complex flowcharts. The modular design of MIRtoolbox is guided by a philosophy of expertise capitalization: techniques developed for certain domains of music analysis are turned into general operators that could be used for different analytical purposes.

Each feature extraction method can accept as argument an audio file, or any preliminary result from intermediary stages of the chain of operations. Also the same syntax can be used for analyses of single audio files, batches of files, series of audio segments, multi-channel signals, etc. For that purpose, the data and methods of the toolbox are organised in an object-oriented architecture.

Memory management mechanisms allow the analysis of large-scale corpus, through the integration of automated chunk decomposition mechanisms and of distinctive processes for flowchart design and evaluation. A set of meta-functions have been designed that enable the integration of user-defined algorithms with the help of simple templates.

7. A Demonstration of Sonic Visualiser

Chris Cannam, Gyorgy Fazekas, Katy Noland (Queen Mary University of London)

Sonic Visualiser is an application designed to assist the study and comprehension of the contents of audio recordings, particularly of music – providing facilities for automatic extraction, browsing, and editing of time-synchronous audio descriptors. It is a powerful and useful tool for researchers investigating the properties of musical recordings.

We will demonstrate the most important features, which include visualisation of audio as waveforms or spectrograms at multiple resolutions; annotation by importing files, text entry, tapping the computer keyboard or using a MIDI device; and looped and time-stretched playback. Sonic Visualiser can run feature extraction plugins to automatically calculate annotations, and we will demonstrate some of the numerous Vamp feature extraction plugins that are available for download, such as beat trackers, pitch detectors, and visualisation tools.

We will also discuss the Vamp plugin software development kit, the new Vampy plugin for Python developers, and Sonic Visualiser's sister application for batch feature extraction, Sonic Annotator.

8. Phase Estimation By Kalman Smoothing to Show Harmonicity in Woodwind with an Application to Enhance Woodwind Music at a Low Sampling Rate

Yushen Han, Christopher Raphael (Indiana University)

This demo is about a novel technique based on state-space model to decouple the phase and the amplitude of partials. This technique results in a representation of separated phase and amplitude for each harmonic of a musical sound. The phase representation obtained by this technique explores the assumption that the harmonics are largely in phase in woodwind instruments. With this harmonicity assumption, the sample-by-sample phase representation makes it possible to project unwrapped phase from one harmonic to another. Several notes recorded in separation and extracted in real music will be presented. Also the state-space model for Kalman smoothing to estimate the phase will be illustrated. An application is developed to enhance a degraded woodwind music recording at a low sampling rate by estimating and reconstructing missing high harmonics from observed low harmonics, guided by score-following. The demo ends with an audible reconstruction of this excerpt.

9. Musiking Self-Motivated System: A model of one's musical playing

Shimpei Tatsumi (Kobe University)

In contrast to a conventinal musicological view, in that music is divided into "sequence" and "tone", recent musicians and researchers attempt to consider "Musiking Body Movement". But they don't have presented any executive formal theory on "body". In this work a unique aproach is proposed, a system motivated to move by itself as an interface between haptic and auditory information, which are formalized to "concept lattice" on the Lattice theory.

The system, Musiking Self-Motivated System (MSMS), determines its next movement by comparing model of neighboring and remote conditions respectively as haptical and auditory. Instead of allocating "tone" to coressponding "sequence", MSMS makes sounds from its movement then sounds make MSMS's movement, which is able to be regarded as sequential sounds. MSMS's trajectories are spatial and temporary biased as if one is playing instrumentals in improvising.

10. Computing genre statistics of chord sequences with a flexible query system

Amélie Anglade (Queen Mary University of London)
Rafael Ramirez (Universitat Pompeu Fabra)
Simon Dixon (Queen Mary University of London)

A query system allowing the user to look for occurrences and statistics of chord sequences is introduced. Through a friendly interface the user can query chord sequences of any length, specifying for each chord either the root note, the category, the degree or a combination of these. One extreme but admissible example of chord sequence to query is: [A7 IV B Vmin]. Sequences containing gaps such as [Amin G7 ... C9 ... Emaj7] are also allowed. The database is composed of 856 music pieces covering Classical (235 pieces), Jazz (338) and Popular music (283). Each genre contains in turn three subgenres: Baroque, Classical and Romantic periods for Classical music; pre-bop, bop

Vol.2009-MUS-82 No.2
2009/11/2

and bossanova for Jazz; pop, blues and celtic music for Popular music. Depending on the task the coverage of a chord sequence on each genre/subgenre or on only one given genre can be looked at. The name of the music pieces and the precise location of the chord sequence in the piece can be displayed if required. The user can also specify if he wants to consider only pieces in a given tonality. Finally the user can work on chord changes only or can take into account the harmonic rhythm.

11. MusicCommentator: A System of Generating Music-Synchronized Comments

Kazuyoshi Yoshii, Masataka Goto (AIST)

This paper presents a system called MusicCommentator that suggests possible comments on appropriate temporal positions in a musical audio clip. In an online video sharing service, many users can provide free-form text comments for temporal events occurring in clips not for entire clips. To emulate the commenting behavior of users, we propose a joint probabilistic model of audio signals and comments. The system trains the model by using existing clips and users' comments given to those clips. Given a new clip and some of its comments, the model is used to estimate what temporal positions could be commented on and what comments could be added to those positions. It then concatenates possible words by taking language constraints into account. Our experimental results showed that using existing comments in a new clip resulted in improved accuracy for generating suitable comments to it.

12. A Novel Method for Estimating F0 and Phonemes of Singing Voice in Polyponic Music

Hiromasa Fujihara, Masataka Goto (AIST)

A novel method will be demonstrated that can be used to concurrently recognize the F0 and phoneme of a singing voice (vocal) in polyphonic music. Our method stochastically models a mixture of a singing voice and other instrumental sounds without segregating the singing voice. It can also estimate a reliable spectral envelope by estimating it from many harmonic structures with various fundamental frequencies (F0s).

13. A Robot Singer with Music Recognition Based on Real-Time Beat Tracking

Kazuhiro Nakadai (Honda Research Institute Japan)
Kazumasa Murata (Tokyo Tech)
Ryu Takeda, Hiroshi G. Okuno (Kyoto University)
Yuji Hasegawa, Hiroshi Tsujino (HRI-JP)

A robot that can provide an enjoyable user interface is a challenging application for music information processing, because the robot should cope with high-power noises including self voices and motor noises.

This paper proposes noise-robust musical beat tracking by using a robot-embedded microphone by introducing two proposed methods, i.e., spectro-temporal pattern matching and echo cancellation, and describes its application to a robot singer which steps, sings, and scats according to musical beats predicted by using a robot-embedded microphone, as a first step to realize a robot which makes a music session with people. We, then, evaluated it in detail at the following three points: adaptation to tempo changes, robustness of environmental noises including periodic noises generated by stepping, singing and scatting, and human-robot interaction by using a clapping sound. The results showed that our beat-tracking robot improved noise-robustness and adaptation to tempo changes drastically so that it can make a simple sound session with people.

14. Rhythm Map: Automatic Music Analysis and Application to Genre Classification

Emiru Tsunoo, Nobutaka Ono, Shigeki Sagayama (University of Tokyo)

This demonstration analyzes rhythmic patterns which are bar-long and repeated frequently. First the input music instruments are processed harmonic/percussion sound source separation and the percussive sounds are emphasized. Then using the algorithm we have proposed which is a combination of dynamic programming and k-means clustering algorithm, bar-long rhythmic patterns are extracted and bar lines are estimated automatically. By this analysis this demonstration shows the structures of the input songs and it makes easy to replace for instance rhythmic patterns according to the structures. It also shows the application to audio genre classification. Genres are characterized not only by timbral information but also rhythm patterns and bass-line patterns information. Those kind of temporal features can be extracted only by bar line segmentation. By k-means-like clustering method that we have proposed, bar-long rhythm patterns and bass-line patterns represent particular genre are extracted and these patterns are used for genre classification.

15. HMM-based automatic accompaniment for piano with jumping capability

Tae Hun Kim (University of Tokyo)
Gustav Larsson (Royal Institute of Technology(KTH))
Haruto Takeda, Stanislaw Raczynski, Takuya Nishimoto, Shigeki Sagayama (University of Tokyo)

This research aims at automatic accompaniment that synchronizes the accompanying parts with the music being performed by human. To achieve this goal, the method of score following that estimates performer's beat position in music score is developed, and we have built an automatic accompaniment system which plays accompaniment parts in the tempo determined by the results of score following.

Since real human performance may include performance error or repetition of the same phrases, score cannot be followed by simple matching of performed notes with the notes in score in time order. To estimate the most probable score position for a given real performance, we formulate the score following as a probabilistic inverse problem using Hidden Markov Models.

In this demonstration, we present automatic accompaniment system for piano with jumping capability. If a score is given, the system is able to estimate the position of the performed note and accompany the performance. Even if user performs the score with some repetition and jumps, the system follows the score correctly, thus proper accompaniment is promised.

16. The "hearing object" class in Temple University Japan

Jean-Julien Aucouturier (Temple University Japan Campus)
Toshimasa Yamanaka (University of Tsukuba)

This poster describes an ongoing experiment in teaching audio/music pattern recognition to non technical students in Temple University Japan, in a class running from Sept - Dec. 2009. Temple students are audio/video media majors, with no computer programming experience and little mathematical background if any. The class runs in parallel with a design class in the Faculty of Design of the University of Tsukuba: students from Temple and Tsukuba must collaborate to create "hearing objects" which incorporate machine listening of sound and music. While the practical goal of the class is to create the content for a design exhibition to be held in Tokyo later this year, the students learn many concepts of audio technology and pattern recognition on the way, in a non-technical and practical manner. We will describe some of the students projects, some of the pedagogical methodology as well as some of the practical sound/music recognition problems encountered, many of them new and interesting even for the expert researcher.

17. Real-Time Control System of Time-Scale and Pitch of Polyphonic Signals via Phase Reconstruction from Expanded/Contracted Power Spectrogram

Yuu Mizuno, Nobutaka Ono, Shigeki Sagayama (University of Tokyo)

Our system enables to control the playback-speed and pitch of audio signals in real time. Modification of time-scale and pitch with user's preference will much enrich music-listening environment and several remixing, and the modification also has various music applications with high-quality recorded sound, such as karaoke and automatic accompaniment systems, which currently use MIDI data.

There have been several reports about this problem, e.g., synchronous overlap and add (SOLA), pitch synchronous overlap and add (PSOLA), and phase vocoder algorithm. Each of these methods, however, has its own disadvantages. Generally the time-domain methods, such as SOLA and PSOLA degrade the quality severely with polyphonic signals. In contrast, although frequency-domain methods can apply to polyphonic signals, the phase vocoder is not necessarily sufficient in terms of sound quality.

Aiming to natural and good-quality sound for polyphonic music signals, our algorithm is based on expansion/contraction of power spectrogram. First, the input power spectrogram is expanded (or shrunken) on time and frequency axis for time-scale and pitch modification by modifying the frame shift and the frame length of short-time Fourier analysis, respectively. Then, a signal waveform is synthesized from the modified power spectrogram with iteratively estimating phase spectrogram consistent to the power spectrogram. The sliding block analysis makes it possible to be implemented in real time. By using linear prediction, our method separates the pitch components and the tone and phonetical features and control the pitch preserving the tone features.

We have implemented a real-time control system that runs under the Linux environment. Input audio signals are played back in real time with arbitrary speed and pitch.

18. Changing Timbre and Phrase in Existing Musical Performances as You Like

Naoki Yasuraoka, Takehiro Abe, Katsutoshi Itoyama, Toru Takahashi, Tetsuya Ogata, Hiroshi G. Okuno (Kyoto University)

We present a new music manipulation method that can change the timbre and phrases of an existing instrumental performance in a polyphonic sound mixture. This method consists of three primitive functions: 1) extracting and analyzing of a single instrumental part from polyphonic music signals, 2) mixing the instrument timbre with another, and 3) rendering a new phrase expression for another given score.

A single instrumental part is extracted by using an integrated tone model that consists of harmonic and inharmonic tone models with the aid of the score of the single instrumental part. The extracted model parameters are decomposed into their averages and deviations. The former is treated as instrument timbre and is customized by mixing, while the latter is treated as phrase expression and is customized by rendering. The resulting customized part is re-mixed with the remaining parts of the original sound mixture to generate new polyphonic music signals.

19. Thereminist Robot: Development of a Robot Theremin Player based on a Theremin's Pitch Model

Takeshi Mizumoto (Kyoto University)
Hiroshi Tsujino (Honda Research Institute Japan)
Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno (Kyoto University)

We present a Thereminist robot system that plays the Theremin based on a Theremin ' s pitch model. The Theremin, which is a 1920s electronic musical instrument, is played by moving a player's hand position in the air without touching it. It is difficult to play the Theremin because the relationship between the hand position and Theremin's pitch (pitch characteristics) is non-linear and varies according to the electromagnetic field (hereafter called environment). These characteristics cause two problems: (1) Adapting to the environment change is required and (2) a naive design tends to depend on robot's particular hardware. We implement the coarse-to-fine control system on the Thereminist robot using newly proposed two pitch models: parametric and nonparametric ones. The Thereminist robot works as below: first, the robot calibrates the pitch model by parameter fitting with the Levenberg-Marquardt method. Second, the robot moves its hand in a coarse manner by feedforward control based on the pitch model. Finally, the robot adjusts its position by feedback control (Proportional-Integral control). In these steps, the robot can play a required pitch quickly, because the robot moves its hand using the pitch model without listening to the Theremin's sound Thus, the time to play the exact pitch is shorter than when only feedback control is used. Three experiments were conducted to evaluate the robustness against the number of samples, environment change, and types of robots. The results revealed that our pitch model describes using only 12 samples of pitches for estimation of the parameters, and adapts if the environment changes. In addition, our system works on two different robots: HRP-2 and ASIMO.