Proceedings of Meetings on Acoustics

Volume 21, 2014 http://acousticalsociety.org/

167th Meeting of the Acoustical Society of America
Providence, Rhode Island
5 - 9 May 2014

Session 4pMUa: Musical Acoustics

4pMUa3. Automatic music accompaniment allowing errors and arbitrary repeats and jumps

Shigeki Sagayama, Tomohiko Nakamura, Eita Nakamura*, Yasuyuki Saito, Hirokazu Kameoka and Nobutaka Ono

***Corresponding author's address: National Institute of Informatics, Tokyo, 114-0014, Tokyo, Japan, eita.nakamura@gmail.com**

Automatic music accompaniment is particularly useful for exercises, rehearsals and personal enjoyment of ensemble music and one-hand piano performances. As musicians may make errors and want to correct them, or they may want to skip hard parts in the score, the system should allow errors as well as arbitrary repeats and skips. Detecting such repeats/skips, however, involves a large complexity of search for a player's score position in the entire score for every input event. Several efficient algorithms have been developed to cope with this problem under practical assumptions used in an online automatic accompaniment system named "Eurydice". In Eurydice for MIDI instruments, music performance is modeled by a hidden Markov model and maximum probability estimation is applied to the polyphonic MIDI input to yield an accompanying MIDI output (e.g., orchestra sound). Another version of Eurydice accepts monaural audio signal input and accompanies to it. Other issues such as treating ornaments, tempo estimation, and accompaniment algorithms are also discussed.

Published by the Acoustical Society of America through the American Institute of Physics

©2014 Acoustical Society of America [DOI: 10.1121/1.4904932]
Received 3 Sep 2014; published 11 Dec 2014

1 Introduction

Automatic music accompaniment is a technique which enables machines to accompany humans in live performance based on a musical score. The technique enables performers to play ensemble music without human accompanists and extends the acoustics of musical instruments into the area of live electronics [1, 2]. As well as its importance in artistic contexts, automatic accompaniment created vast demand for use in practice sessions, rehearsals and for the personal enjoyment of ensemble music, and has therefore become a popular subfield of music information processing (see Ref. [3] for a dated review, and more recent works [4–10], to mention a few).

To successfully synchronize accompaniments, real-time estimation of a performer's score position, called score following, is required. This is a nontrivial problem since human performances widely vary even if they are based on the same score. For example, the tempo and its changes, performance mistakes, ornaments, acoustic variations, and noises are typical sources of uncertainties in music performance. A score-following algorithm must contain a set of complex rules to correctly identify score positions, and stochastic models such as the hidden Markov model (HMM) applied to music performance are widely used to construct score-following algorithms in a principled way.

In music performances during practices, players may make repeats to practice some sections again or skips to omit some sections. Allowing arbitrary repeats and skips is particularly important for automatic accompaniment. Repeats and skips in score following were first discussed in Refs. [5, 11], where only jumps to particular score positions were considered. Later methods for handling arbitrary repeats and skips were developed in Ref. [12]. Since then, our group has developed techniques to handle arbitrary repeats and skips as well as performance mistakes and ornaments [12–25]. The main purpose of this paper is to review the highlights of our work and related solutions. A basic model for music performance in symbolic (i.e. MIDI) signals is explained in the next section, and the problem of large computational cost and its solution will be detailed in Sec. 3. An analogous method for audio score following will be explained in Sec. 4. Issues regarding musical ornaments are considered in Sec. 5, and tempo estimation and accompaniment algorithms are discussed in Sec. 6. Conclusions are given in Sec. 7.

2 Stochastic performance model

To construct a score-following algorithm, we first model music performance as a stochastic process. We consider polyphonic MIDI signals in this and the next sections (audio signals are considered in Sec. 4). Given a stochastic model that yields the probability of a sequence of intended score positions and of generated performed notes based on a score, the score-following problem can be restated as one of finding the most probable sequence of intended score positions given a performance signal. HMM is particularly suited for this problem because it can effectively describe sequential, erroneous, and noisy observations of music performance, and several computationally efficient inference algorithms for music processing have been developed [26, 27].

The use of temporal information is important for score following of polyphonic performances since the clustering of performed notes into musical events, e.g., chords or arpeggios, often becomes ambiguous without it. An HMM was proposed to explicitly describe the temporal information. There are two equivalent representations of the model: one describes time as a dimension in a state space, and the other is based on a consideration of the output probability of inter-onset intervals (IOIs). The former representation is explained in the following. A similar HMM was proposed in Ref. [28, 29] in the context of music/rhythm transcription, and a preliminary attempt to consider such an HMM for score following was done in Ref. [12].

Let i label a set of notes whose onsets are simultaneous in the score, which will be called a musical event or a "chord". The state space of the model is the set of all musical events in the score, and a state represents an intended musical event i_m for each performed note indexed by m ($m = 1, \cdots, M$), where M is the total number of performed notes [25]. The pitch and onset time of the m-th performed note are denoted by p_m and t_m. The music performance can be modeled as a two-stage stochastic process of choosing the intended musical events first and then outputting the observed performed notes. The first stage is described as transitions between states, and the temporal information can be described as output of IOI $\delta t_m = t_m - t_{m-1}$ at each transition. Assuming that the probability of choosing the state i_m is only dependent on the previous state as $P(i_m|i_{m-1}) = a_{i_{m-1}i_m}$ and the output probability of pitch and IOI is only dependent on the current and the previous states as $P(p_m, \delta t_m|i_{m-1}, i_m) = b_{i_{m-1}i_m}(p_m, \delta t_m)$, the model can be described by an HMM. The transition probability matrix $(a_{ij})_{i,j}$ describes how players proceed in the score during performance (Fig. 1), and output probabilities describe how they actually produce performed notes.

3 Models of arbitrary repeats and skips, and fast Viterbi algorithm

As shown in Fig. 1, large repeats and skips between score positions can be described by a transition probability a_{ij} with large $|j-i|$. Since it is difficult to anticipate all score positions

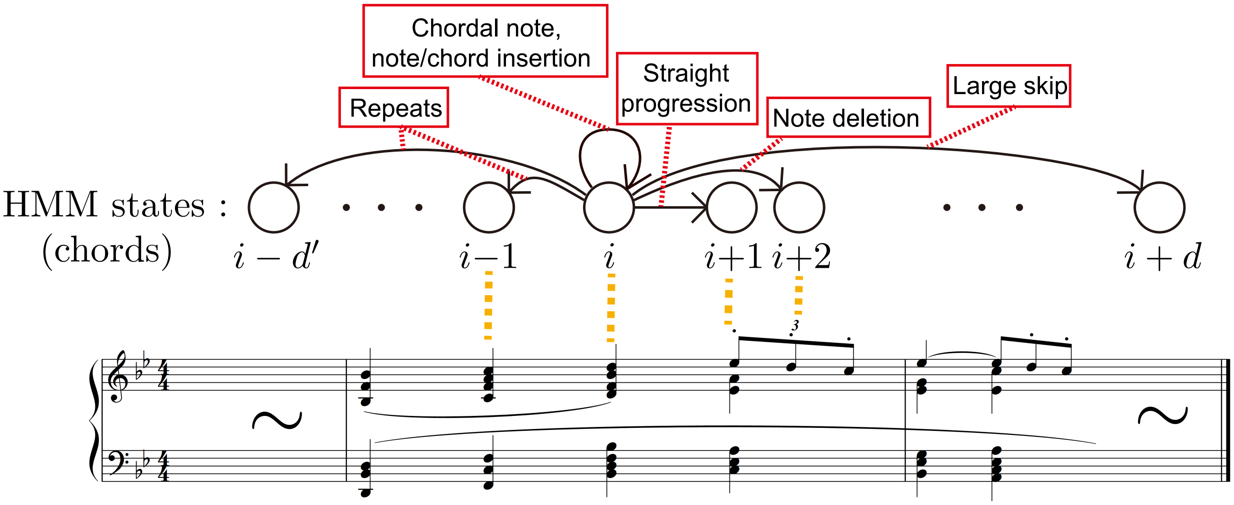

Figure 1: Transitions of the HMM for a simple passage and their interpretations [25].

from and to where players make repeats and skips, it is practical to consider arbitrary repeats and skips, which can be expressed as $a_{ij} \neq 0$ for all i and j. In this case, all score positions and transitions must be taken into account at every time, and the computational cost for the inference is large for long scores. For example, a Viterbi update requires $\mathcal{O}(N^2)$ complexity, where N is the number of states, which is too large for real-time processing for $N \gtrsim 500$.

There are solutions to reduce the computation cost by using simplified models, one of which is a uniform repeat/skip probability model, a_{ij} is constant for large $|j-i|$. It can be shown that the computational complexity can be reduced to $\mathcal{O}(DN)$ when a_{ij} is constant for $j < i - D_1$ or $j > i + D_2$ $(D = D_1 + D_2 + 1)$. In practice a value of D between 3 and 10 is sufficient, significantly reducing the complexity. We can generalize the model to an outer-product HMM, where for large $|j-i|$ a_{ij} is an outer-product of two vectors, while retaining computational efficiency. Details of the models and analyses of tendencies in repeats and skips in actual performance data are given in Ref. [25].

Results of measuring the processing time for a Viterbi update are shown in Fig. 2. We see that the processing time is clearly reduced by the fast Viterbi algorithm for $N \gtrsim 200$, and remains within a few tens of milliseconds for $N \lesssim 10000$. Thus, typical concert-size pieces can be processed without any sensible delay with the proposed score-following algorithm. Evaluations in Ref. [25] show that repeats and skips are followed within about three chords, depending on scores, after resumption with the uniform repeat/skip probability model, and the outer-product HMM improves the result by about one chord.

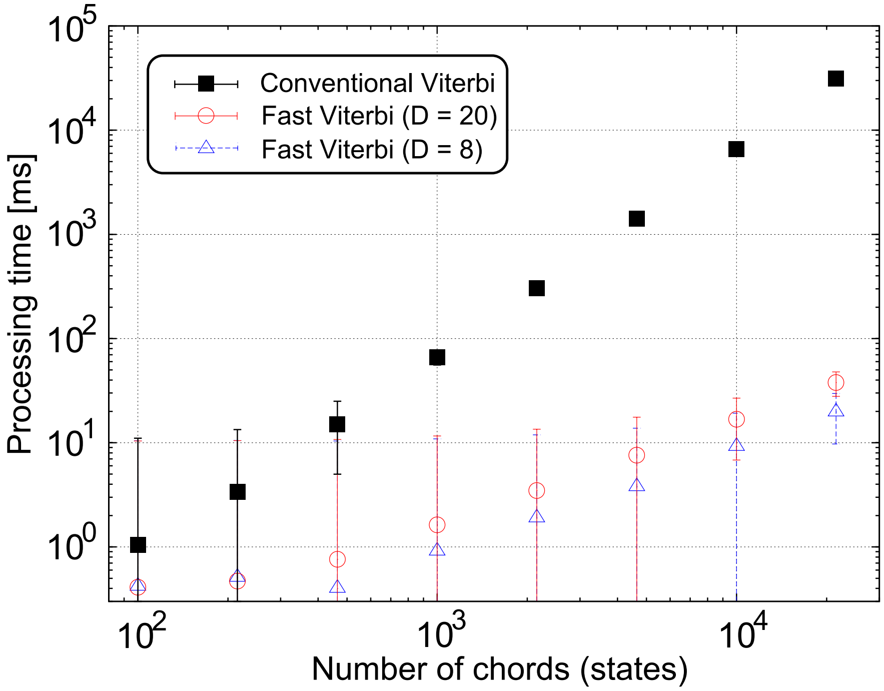

Figure 2: Averaged processing time for a Viterbi update [25]. The squares indicate results using the online version of conventional Viterbi algorithm and the triangles and circles indicate results using the fast Viterbi algorithm. Errors indicate 1σ confidence intervals. The computational environment involved an Intel(R) Core(TM) i5-2540M CPU, with 8 GB of RAM and the Windows 7 Professional 64-bit OS.

The above technique to derive the fast Viterbi algorithm can also be applied to the forward and backward algorithms. Since HMM and similar dynamic-programming matching methods are extensively used for information processing, the fast inference algorithms are useful for other applications. For example, it is applied to music structure analysis in Ref. [30].

4 Audio score following

A score-following algorithm for audio signals handling arbitrary repeats and skips was proposed by our group in Ref. [23] (see Ref. [16] for a related work). A model similar to the one used for the symbolic case can be built based on an HMM for audio signals with some

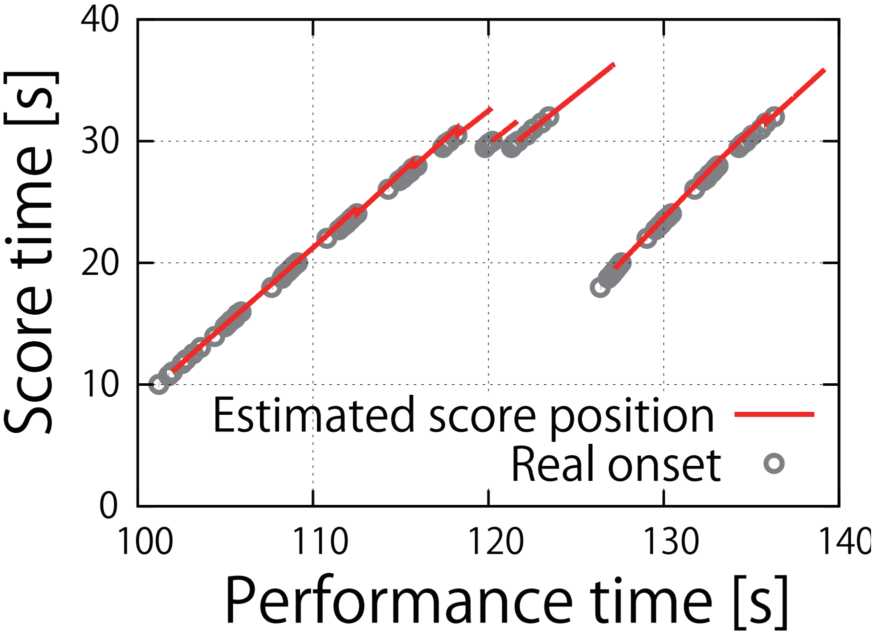

Figure 3: The score-following result for a clarinet performance with repeats. The gray circle indicates performed onsets, and the red line indicates the estimated score positions.

differences. First, the output of the signal model consists of acoustic features instead of symbolic pitches, for which the normalized spectrum of a constant-Q transform [31] can be used. The output probability is described with a mixture of multivariate Gaussian distributions. Second, since the model outputs such acoustic features for each frame, transitions between states occur within a constant time shift. This makes a big difference in modeling note durations compared to the MIDI case. As a simplest choice, we can model durations in terms of the self-transition probability of each state corresponding to each note in the score. A more refined model such as a variable duration model or a semi-Markov model could be used, but typically, the computation cost increases significantly.

Another essential difference is that rests in the score and pauses made by the player must be carefully treated in the case of audio signals, which can be represented by silence states. Since there are frequently pauses before repeats and skips, a model with a state representing such pauses can be constructed. A refined inference algorithm for the model can also be derived. A score-following result for a clarinet performance with repeats is illustrated in Fig. 3. A phase reconstruction algorithm for generating accompaniment sound from a recorded audio file was discussed in Refs. [32, 33]

5 Models of ornaments for MIDI performance

As is generally known in music practice, musical ornaments such as trills and arpeggios are somewhat arbitrary and indeterminate, because of the speed of performed notes in such cases [34, 6]. For an accurate score following of performances with ornaments, indeterminacies must be properly captured in the performance model. Based on a careful observation of the realization of ornaments, an extended model of music performance was constructed which describes the indeterminacies. For example, the variations in speed of ornamental notes are described with distributions of corresponding IOIs, which are obtained by analyzing actual performances.

Especially for a complex polyphonic passage, in which several ornaments can be superposed, proper construction of HMM states is also important. States which represent chord, trill, and short appoggiatura in a general sense were introduced in the model, and a state-construction algorithm was derived which can be applied for a quite general polyphonic passage. Detailed discussions on ornaments are given in a paper to appear elsewhere [35].

6 Tempo estimation and accompaniment algorithm

So far we have not discussed tempo. Given a situation where tempo depends on a particular player and may change during a performance, one must in some way perform tempo estimation in real time. Although a joint model of score position and tempo can in general be used to simultaneously estimate both, such a model typically necessitates large computational cost (see, for example, Ref. [8]). To avoid a significant increase of this cost, we alternately perform tempo estimation and score-position estimation.

Since the tempo is indirectly observed through detected onset times which are subjected to noise, an estimation based on a stochastic model is used, where the tempo is represented as a latent variable and the IOI is the visible quantity. A model based on a linear dynamical system [36, 37] can be used for this. To treat sudden pauses typically made by a player during practices, the model is extended to represent such pauses as noise sources for the visible IOI. An efficient tempo estimation can be done with a switching Kalman filter.

Finally, given the information on the current score position and tempo of the performance, the accompaniment sound is generated. There are two complementary methods for accompaniment generation implemented in our system. One is called the "waiting mode" of accompaniment where the computer waits until the next expected performed note is played and reacts as soon as it is played. In this way, a sudden tempo change or a sudden pause is best treated. The other one is called the "non-waiting mode", where the next performed note is first anticipated with the estimated tempo and the accompaniment is generated according to the prediction. To keep good synchronization, the computer waits when the player gets behind the accompaniment, and catches up when the player gets ahead the accompaniment, as performed notes are observed. With the latter mode of accompaniment, delays in the score following module or the sound control module can be more appropriately treated, and the accompaniment can lead the player, which is preferable in some instances.

Two demonstration videos of our “Eurydice” accompaniment system presented in the Automatic Accompaniment Demonstration Concert at the 167th Meeting of the Acoustic Society of America can now be watched on the Internet [38].

7 Discussions and conclusions

In this paper, we have reviewed the development of automatic music accompaniment systems including recent progress in light of their use for music practice and personal enjoyment. Based on a stochastic method, basic techniques, especially score following, have been developed for music performances having mistakes, arbitrary repeats and skips, with convincing results. The developed technique to reduce the computation cost for score-position estimation of music performance with these properties can be applied for other applications where alignment between observed sequences (e.g., arbitrary performance, speech, etc.) and their corresponding documents (e.g., score, text, etc.) involves both local deformations (e.g., mistakes and added noise) and global deformations (i.e., transitions) in the observed sequence.

For future directions, the generation of expressive accompaniments would be a next challenge. Here, how to reflect player’s, as well as general, musicality in the accompaniment is an interesting problem. Adaptation of the accompaniment generation using the rehearsal data would be important as well [39, 7].

Acknowledgements

This work is supported in part by Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science, No. 23240021 (S.S. and N.O.), No. 26240025 (S.S., Y.S., and N.O.), and No. 25880029 (E.N.).

References

[1] R. Dannenberg, “An on-line algorithm for real-time accompaniment,” *Proc. ICMC*, pp. 193–198, 1984.

[2] B. Vercoe, "The synthetic performer in the context of live performance," *Proc. ICMC*, pp. 199–200, 1984.

[3] N. Orio, S. Lemouton, and D. Schwarz, "Score following: State of the art and new developments," *Proc. NIME*, pp. 36–41, 2003.

[4] D. Schwarz, N. Orio, and N. Schnell, "Robust polyphonic MIDI score following with hidden Markov models," *Proc. ICMC*, pp. 442–445, 2004.

[5] B. Pardo and W. Birmingham, "Modeling form for on-line following of musical performances," *Proc. of the Twentieth National Conference on Artificial Intelligence*, 2005.

[6] A. Cont, "A coupled duration-focused architecture for realtime music to score alignment," *IEEE Trans. PAMI*, **32(6)**, pp. 974–987, 2010.

[7] C. Raphael, "The informatics philharmonic," *Communications of the ACM*, **54(3)**, pp. 87–93, 2011.

[8] T. Otsuka et al., "Real-time audio-to-score alignment using particle filter for coplayer music robots," *EURASIP J. Advances in Signal Processing*, **2011**, 384651, 2011.

[9] C. Joder, S Essid, and G. Richard, "A conditional random field framework for robust and scalable audio-to-score matching," *IEEE TASLP*, **19(8)**, pp. 2385–2397, 2011.

[10] A. Arzt, G. Widmer and S. Dixon, "Adaptive distance normalization for real-time music tracking," *Proc. EUSIPCO*, pp. 2689–2693, 2012.

[11] C. Oshima, K. Nishimoto and M. Suzuki, "A piano duo performance support system to motivate children's practice at home (in Japanese)," *J. of Information Processing Society of Japan (IPSJ)*, **46(1)**, pp. 157–171, 2005.

[12] H. Takeda, T. Nishimoto and S. Sagayama, "Automatic accompaniment system of MIDI performance using HMM-based score following (in Japanese)," *Tech. Rep. IPSJ Special Interest Group on Music and Computer (SIGMUS)*, **MUS-66 (18)**, pp. 109–116, 2006.

[13] H. Takeda, T. Nishimoto and S. Sagayama, "Score following of polyphonic MIDI performance based on dynamic programming (in Japanese)," *Proc. Acoustic Society of Japan (ASJ)*, **3-2-4**, pp. 723–724, 2006.

[14] H. Takeda, T. Nishimoto and S. Sagayama, "Automatic accompaniment using score following of MIDI performance using HMM (in Japanese)," *Proc. ASJ*, **1-7-10**, pp. 571–572, 2006.

[15] Y. Saito, T. Nishimoto and S. Sagayama, "Synchronization of musical performance between human player and automatic-accompaniment system using head motion estimation (in Japanese)," *Proc. ASJ*, **3-1-16**, pp. 1075–1076, 2011.

[16] K. Suzuki, Y. Ueda, Y. Saito, N. Ono and S. Sagayama, "Repetition-adaptive automatic accompaniment based on acoustic score following using HMM (in Japanese)," *Tech. Rep. SIGMUS*, **MUS-89 (29)**, pp. 1–6, 2011.

[17] E. Nakamura, R. Yamamoto, Y. Saito, S. Sako and S. Sagayama, "Modeling Performance Indeterminacies for Polyphonic Midi Score Following and Its Application to Automatic Accompaniment (in Japanese)," *Tech. Rep. SIGMUS*, **MUS-96 (14)**, pp. 1–6, 2012.

[18] E. Nakamura, R. Yamamoto, Y. Saito, S. Sako and S. Sagayama, "Modeling ornaments in polyphonic MIDI score following and its application to automatic accompaniment (in Japanese)," *Proc. ASJ*, **2-10-5**, pp. 929–930, 2012.

[19] T. Nakamura, Y. Mizuno, K. Suzuki, E. Nakamura, Y. Higuchi, S. Fukayama and S. Sagayama, "Automatic accompaniment robust to errors and repetitions in acoustic musical performances (in Japanese)," *Proc. ASJ*, **2-10-6**, pp. 931–934, 2012.

[20] E. Nakamura, Y. Saito and S. Sagayama, "Merged- output hidden Markov model and its applications in score following and hand separation of polyphonic keyboard music (in Japanese)," *Tech. Rep. SIGMUS*, **EC-27 (15)**, pp. 1–6, 2013.

[21] T. Nakamura, E. Nakamura and S. Sagayama, "Fast score following for acoustic signal of musical performance with repeats and skips (in Japanese)," *Proc. IPSJ*, **4R-10**, pp. 283–284, 2013.

[22] T. Nakamura, E. Nakamura and S. Sagayama, "Fast score following for audio signals of musical performance with errors, arbitrary repeats and skips (in Japanese)," *Tech. Rep. SIGMUS*, **MUS-99 (43)**, pp. 1–5, 2013.

[23] T. Nakamura, E. Nakamura and S. Sagayama, "Acoustic Score Following to Musical Performance with Errors and Arbitrary Repeats and Skips for Automatic Accompaniment," *Proc. SMC*, pp. 299–304, 2013.

[24] E. Nakamura, H. Takeda, R. Yamamoto, Y. Saito, S. Sako and S. Sagayama, "Score following handling performances with arbitrary repeats and skips and automatic accompaniment (in Japanese)," *J. IPSJ*, **54(4)**, pp. 1338–1349, 2013.

[25] E. Nakamura, T. Nakamura, Y. Saito, N. Ono and S. Sagayama, "Outer-product hidden Markov model and polyphonic MIDI score following," *JNMR*, **43(2)**, pp. 183-201, 2014.

[26] C. Raphael, "Automatic segmentation of acoustic musical signals using hidden Markov models," *IEEE Trans. PAMI*, **21(4)**, pp. 360–370, 1999.

[27] P. Cano, A. Loscos and J. Bonada, "Score-performance matching using HMMs," *Proc. ICMC*, pp. 441–444, 1999.

[28] N. Saito, M. Nakai, H. Shimodaira and S. Sagayama, "Hidden Markov model for restoration of musical note sequence from the performance (in Japanese)," *Tech. Rep. SIGMUS*, pp. 27–32, 1999.

[29] T. Otsuki et al., "Musical rhythm recognition using hidden Markov model (in Japanese)," *J. IPSJ*, **43(2)**, pp. 245–255, 2002.

[30] Y. Zou, Y. Kamamoto, N. Ono and S. Sagayama, "Structural analysis of musical signal by dynamic programming and its application to audio coding (in Japanese)," *Proc. ASJ*, **3-6-12**, pp. 1049–1050, 2012.

[31] J. Brown, and M. Puckette, "An efficient algorithm for the calculation of a constant Q transform," *J. Acoust. Soc. Am.*, **92**, pp. 2698–2701, 1992.

[32] Y. Mizuno, J. Le Roux, N. Ono and S. Sagayama, "Real-time time-scale/pitch modification of music signal by stretching power spectrogram and consistent phase reconstruction (in Japanese)," *Proc. ASJ*, **2-8-4**, pp. 843–844, 2009.

[33] Y. Mizuno, H. Tachibana and S. Sagayama, "Real-time time-scale modification of a music signal using phase reconstruction for synchronous playback in conducting/accompaniment system (in Japanese)," *Proc. ASJ*, **1-3-12**, pp. 897–898, 2011.

[34] R. Dannenberg and H. Mukaino, "New techniques for enhanced quality of computer accompaniment," *Proc. ICMC*, pp. 243–249, 1988.

[35] E. Nakamura, N. Ono, S. Sagayama and K. Watanabe, "A Stochastic Temporal Model of Polyphonic MIDI Performance with Ornaments," to appear.

[36] C. Raphael, "A probabilistic expert system for automatic musical accompaniment," *J. Computational and Graphical Statistics*, **10(3)**, pp. 487–512, 2001.

[37] A. T. Cemgil, B. Kappen, P. Desain, and H. Honing, "On tempo tracking: Tempogram representation and Kalman filtering," *JNMR*, **29(4)**, pp. 259–273, 2000.

[38] Demonstrating video clips of our system can be downloaded at: https://www.youtube.com/watch?v=KgnR2BzrafU and https://www.youtube.com/watch?v=fW6VKiC4k34

[39] B. Vercoe and M. Puckette, "Synthetic rehearsal: Training the synthetic performer," *Proc. ICMC*, pp. 275–278, 1985.