AUTOMATIC TRANSCRIPTION OF GUITAR TABLATURE FROM AUDIO SIGNALS IN ACCORDANCE WITH PLAYER'S PROFICIENCY

Kazuki Yazawa, Katsutoshi Itoyama, Hiroshi G. Okuno

Graduate School of Informatics, Kyoto University, Japan

ABSTRACT

We describe a method for automatically transcribing guitar tablatures from audio signals in accordance with the player's proficiency for use as support for a guitar player's practice. The system estimates the multiple pitches in each time frame and the optimal fingering considering playability and player's proficiency. It combines a conventional multipitch estimation method with a basic dynamic programming method. The difficulty of the fingerings can be changed by tuning the parameter representing the relative weights of the acoustical reproducibility and the fingering easiness. Experiments conducted using synthesized guitar audio signals to evaluate the transcribed tablatures in terms of the multipitch estimation accuracy and fingering easiness demonstrated that the system can simplify the fingering with higher precision of multipitch estimation results than the conventional method.

Index Terms— Dynamic programming (DP), guitar tablature transcription, multipitch estimation, music signal processing, performance proficiency.

1. INTRODUCTION

Tablature (figure 1) is a format of musical scores for string instruments such as guitar and bass, and many guitar players are familiar with it. Tablature indicates the strings and fret positions by numbers, which enables players to play a guitar intuitively without much musical knowledge. Since a guitar can produce the same pitch with various strings, multiple fingerings are possible for a sequence of pitches. Tablature eliminates such ambiguity of fingerings, thus players can easily play a guitar. Therefore, tablature is helpful to many guitar players, particularly beginners, in practicing the guitar.

While tablature can motivate most guitar players to practice, players cannot always obtain tablatures for the pieces they want to play. It has become more difficult to find objective tablatures due to the popularization of consumer generated media (CGM). Moreover, the difficulty of a tablature may not suit the player's proficiency. For example, many beginners often want to play musical pieces that have a level of difficulty greater than their performance proficiency, causing them to lose their motivation to practice. In contrast, expert players care more about accuracy than they do about ease of playing since they are more proficient at their instruments. For all of these reasons, a system is needed for automatically transcribing tablatures from audio signals in accordance with the player's proficiency.

In general, guitar players consider both acoustical reproducibility and fingering easiness when determining the fingering from an audio signal. Acoustical reproducibility means how similar the pitches and timbres played with the fingering are to the actual ones, and fingering easiness means how easily players can play the piece. The trade-off between them varies with the player's proficiency. On the one hand, most experts select the fingering with which they can produce sound similar to that of the actual performance, even if it is difficult to play. On the other hand, beginners tend to select the easier fingering and tolerate the degraded performance. This means that the transcription system should enable users to manually tune the relative weights of these features.

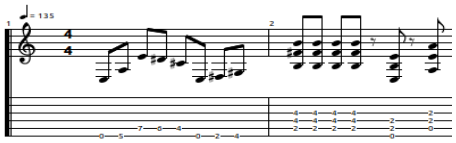


Fig. 1. Example of tablature.

Here we present a method for transcribing tablatures from audio signals in accordance with the player's proficiency. The proposed method estimates the fingering satisfying the constraints on guitar performance by modeling the fingering estimation as a longest path search problem on a weighted directed acyclic graph and solving it by a dynamic programming (DP) [1]. The difficulty of the estimated fingerings can be changed by tuning the weights for acoustical reproducibility and fingering easiness.

2. OVERVIEW OF PROPOSED METHOD

The procedure used in our proposed method is mostly the same as that in our prior work [2]. First, the multiple pitches in each time frame are estimated by using an existing method and then the optimal fingering is estimated on the basis of the results. Next, the unplayable combinations of pitches in the original result of multipitch estimation are suppressed by using the estimated optimal fingering.

Latent harmonic allocation (LHA) [3], which is a conventional multipitch estimation method using machine learning, is used to estimate the appearance degree of multiple pitches in each time frame. LHA approximates harmonic structures of instrumental sounds using a Gaussian mixture model (GMM) and estimates the model parameters using Bayesian estimation. With LHA, we can estimate N_{tk}, the appearance degree of the k-th pitch in the t-th time frame, by inputting the frequency spectrum of the audio signal.

The optimal fingering is estimated on the basis of the results of LHA and predetermined fingering costs. A fingering is regarded as the changes with time of the fingering configurations used, and it is modeled by using a weighted directed acyclic graph (figure 2). The graph is described in detail in the next section. The optimal sequence of fingering configurations $C^* = \{c^*_{p_1}, \ldots, c^*_{p_T}\}$ is estimated by searching the longest path of the graph using DP [1].

Any pitches that cannot be played with the optimal fingering C^* from the original LHA results are suppressed. The modified appearance degree of the k-th pitch in the t-th time frame is defined as

$$\tilde{N}_{tk} = \begin{cases} N_{tk} & (k \in K_{p_t}) \\ 0 & (otherwise) \end{cases},$$

where $K_{p_t} = \{k_{p_t1}, \ldots, k_{p_t6}\}$ represents a combination of the pitches that can be played with all six strings and the optimal finger-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

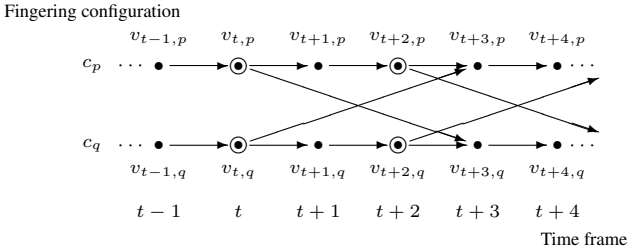


Fig. 2. Weighted directed acyclic graph for fingering estimation. Vertical axis represents fingering configuration, and horizontal axis represents time frame. Directed edges represent transitions of configurations. Onset occurs at t-th and $(t+2)$-th time frame, when configurations may be changed. Here $D = 3$ and there are two example configurations.

ing configuration $c^*_{p_t}$. The presence or absence of pitches is finally determined by making a threshold decision on $\tilde{N}_{tk}$. That is, threshold parameter α is set, and all pitches that satisfy $\tilde{N}_{tk} > \alpha \max \tilde{N}_{tk}$ are regarded as played in the t-th time frame.

The strings and fret positions to be played in each time frame are estimated by using the result of fingering estimation C^* and that of the threshold decision on $\tilde{N}_{tk}$. The actual tablature can be transcribed by combining these results with other information such as the results of an existing beat tracking method [4].

3. WEIGHTED DIRECTED ACYCLIC GRAPH

Here we formalize the weighted directed acyclic graph used to estimate the optimal fingering and describe a way to define the locations of the vertexes and edges and the weight of each edge. In our prior work [2], the weights of the edges are defined only by acoustical reproducibility. By contrast, the proposed method defines the weights by the weighted sum of acoustical reproducibility and fingering easiness. This enables fingering to be estimated on the basis of a player's proficiency. Furthermore, we slightly modified the three constraints regarding guitar performance that determine the locations of vertexes and edges. Here we represent use of the p-th fingering configuration in the t-th time frame as v_{tp} and the edge from v_{tp} to v_{uq} as e_{tupq}.

3.1. Three constraints

The locations of vertexes and edges are determined by three constraints on guitar performance. The first constraint is "playable configuration constraint," which ensures the playability of the fingering configuration used in each time frame. In our prior work [2], playable configurations were theoretically enumerated based on only reach of fingers and number of fingers, which included some configurations that is actually unplayable. In the proposed method, playable configurations are enumerated more strictly by using a standard guitar chordbook [5]. First, we enumerate the templates of common fingering configurations, which contain information about only the relative finger positions, by referring the chordbook. We then enumerate all playable fingering configurations by arranging these templates on any of the fret positions on a guitar fingerboard. Here we assume a standard guitar with 20 frets and 6 strings tuned to normal tuning (EADGBE). In this case, the total number of enumerated fingering configurations is $P = 1401$. By making these fingering configurations correspond to vertexes of the graph, $v_{t1}, \ldots, v_{tP}$, we can ensure the playability of a fingering configuration in each time frame.

The second constraint is "configuration change timing constraint," which allows a fingering configuration to change only at onset time frames. Onset time frames are detected on the basis of N_{tk} flux, $NF_t = \sum_k \max(0, N_{tk} - N_{(t-1)k})$, which is an application of spectral flux [6]. When using N_{tk} flux instead of spectral flux, all of the non-harmonic energy that happens at the attack are disregarded. However, we use N_{tk} flux because it was experimentally confirmed that onsets can be detected more accurately with it. We regard any time frame for which the value of NF_t is higher than a certain threshold $\beta \max_t NF_t$ as an onset time frame. Here β is a threshold parameter.

The third constraint is "configuration continuity constraint," which forces fingering configurations to be used at least for D continuous time frames after the configuration changes from one to another. This constraint reflects the fact that guitar players cannot move their fingers too quickly. Here we assume that the minimum duration D is a fixed value for the entire piece and independent of the fingering configurations and the timing of configuration changes.

For the second and third constraints, an edge of the graph e_{tupq} satisfies either of the following two conditions:

- $p = q$, $u = t + 1$.
- $p \neq q$, $u = t + D$, and t is an onset time.

An edge of the graph, $e_{\hat{t}(\hat{t}+D)pq}$ $(p \neq q)$, represents a change in the used fingering configuration from the p-th one to the q-th one in the $\hat{t}$-th time frame. In this case, the q-th configuration is used for the following D time frames.

3.2. Weight of each edge

The weight of each edge e_{tupq} is defined on the basis of both acoustical reproducibility (AR) and fingering easiness (FE). Since there is generally a trade-off relationship between them that depends on the proficiency of the player, we define the weights of the edges as the weighted sum of AR and FE:

$$W_{tupq} = \sum_{t'=t}^{u-1} \{wAR(X_{t'+1}, c_{p_{t'+1}}) + (1-w)FE(c_{p_{t'}}, c_{p_{t'+1}})\},$$

where w is the parameter used to change the relative weights of AR and FE and thereby reflect the player's proficiency in the fingering estimation. The lower the w, the higher the priority of FE against AR, resulting in easier fingering for beginners. Conversely, the higher value the w, the higher the priority of AR against FE. When $w = 1.0$, no consideration is given to FE, and the fingering that maximizes AR is estimated. By tuning w in accordance with the proficiency of users, who will play the transcription, the tablature of appropriate difficulty is transcribed.

AR and FE are defined as a hierarchical structure, as shown in figure 3. The details are described below.

3.2.1. Acoustical reproducibility

AR represents how accurately the actual sound of the performance is reproduced with a fingering. The main factors that affect on sound are pitch and timbre. While a player can produce the same pitch with different combinations of strings, the timbre for each string is slightly different, thus players can play sound closer to the actual sound by considering differences in timbre. However, compared with other string instruments such as the violin [7], such differences are often disregarded in guitar performance, and many players tend to consider the fingering easiness to be more important than exactly reproducing the timbre when determining the fingering. Therefore, we approximate AR by using only pitch reproducibility (PR):

$$AR(X_t, c_p) \approx PR(X_t, c_p) = \sum_{k \in K_p} N_{tk}.$$

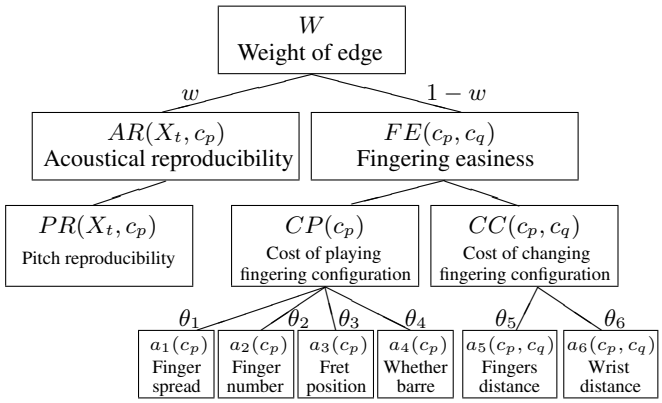


Fig. 3. Hierarchical structure of weights of edges in figure 2. Weights are defined as weighted sum of AR and FE, and each feature is defined as combination of further lower features.

This formula means that the pitch reproducibility of spectrum X_t for fingering configuration c_p is defined as the summation of the appearance degrees of the six pitches $K_p = \{k_{p1}, \ldots, k_{p6}\}$ that can be played with the configuration. For this calculation, duplicate pitches in K_p are eliminated in order to prevent the configurations that have pitch duplications from unfairly getting a higher score.

3.2.2. Fingering easiness

FE is determined by both the easiness of playing a fingering configuration and that of changing the fingering configuration from one to another [8, 9]. Therefore, here, we calculate the cost of playing fingering configurations (CP) and that of changing the configuration (CC) for all playable ones in advance and define FE as

$$FE(c_p, c_q) = \frac{1}{1 + (CP(c_q) + CC(c_p, c_q))}.$$

CP is defined by four features, "a_1: width of finger spread," "a_2: number of fingers used," "a_3: fret position of forefinger," and "a_4: whether it is a barre chord," by referring to previous work on guitar fingering [8, 10, 11]. In prior work on CC [12, 13], the cost was calculated by summing up the Manhattan distances of all fingers between two configurations. However, strictly considering the actual guitar performance, this way of determining the cost is insufficient because fingers do not move independently. Therefore, we calculate CC in the following steps. First, all fingers used in one fingering configuration are moved horizontally along the fretboard so that the fret position of the forefinger on one configuration is equal to that on the other configuration. This movement distance is regarded as that of player's wrist (a_5). After that, the Manhattan distances of all fingers between the two configurations are calculated and summed up (a_6). CC is defined the weighted sum of a_5 and a_6.

Thus CP and CC are defined as:

$$CP(c_p) = \sum_{i=1}^{4} \theta_i a_i(c_p),\ CC(c_p, c_q) = \sum_{i=5}^{6} \theta_i a_i(c_p, c_q),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_6)$ is a parameter used to determine the relative weight of each feature. We can reflect each player's tendency and his or her strong and weak points for the fingering estimation by changing the value of $\boldsymbol{\theta}$. For example, for those who have small hands and cannot extend their fingers very widely, we can estimate the fingering so that it has less load for him or her by enlarging the value of θ_1. We can also estimate the fingering with less load for those who are not good at playing barre chord configurations by enlarging the value of θ_4.

Table 1. Results of multipitch estimation. Each value is mean for all guitar parts. *LHA* represents conventional method without post-processing.

	LHA	Proposed									
Value of w		1.0	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
Precision	0.607	0.649	0.650	0.651	0.655	0.656	0.662	0.669	0.662	0.641	0.617
Recall	0.612	0.637	0.632	0.629	0.623	0.614	0.590	0.539	0.456	0.368	0.313
F-measure	0.604	0.634	0.632	0.632	0.631	0.626	0.614	0.587	0.530	0.459	0.404

4. EXPERIMENTAL EVALUATION

To evaluate the performance of the proposed method, we conducted experiments for examining the multipitch estimation accuracy and the fingering easiness for various values of parameter w.

4.1. Experimental conditions

We used 93 guitar parts, extracted from 11 jazz pieces and 56 popular ones in the RWC music database [14], as experimental data. Only the first 60 seconds of each part was used in order to reduce the computation time. There were some silent sections, and the average sounding time for all parts was 34.7 seconds. A MIDI version of each piece was used to enable quantitative evaluation for multipitch estimation. The audio signals were recorded with a MIDI synthesizer (Yamaha MOTIF-XS) and transformed into wavelet spectrograms using Gabor wavelets with a time resolution of 20 ms. The ground truths of multipitch estimation were constructed from corresponding MIDI data.

To evaluate the potential performance of the system, we optimized the threshold parameter of multipitch estimation α so as to maximize the F-measure for each part and condition. The threshold parameter for onset detection β was set to 0.10 experimentally. The fingering configuration duration D was set to 200 (ms), considering that ordinary guitar players cannot change the fingering configuration more than five times a second. All parameters of the $\boldsymbol{\theta}$ were set to be equal this time.

As evaluation criteria for multipitch estimation, we used the precision, recall and F-measure for the time frames. For comparison, we evaluated the multipitch estimation accuracy of conventional LHA with the same dataset. To evaluate the fingering easiness, we used the number of fingering configurations used and the entire fingering cost (that is CP + CC) divided by the number of sounding time frames for each piece. We investigated how these values changed while varying the value of w.

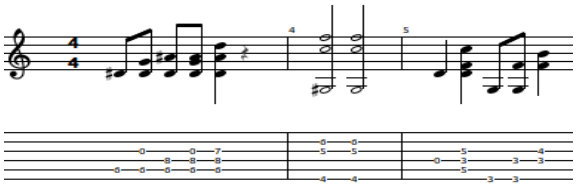
4.2. Experimental results

The results of multipitch estimation are shown in Table 1. When the value of w was high, the recall and F-measure with the proposed method exceeded those of the conventional one. This means that the constraints added to LHA worked well and eliminated some undesirable pitches. The precision of the proposed method was higher than that of the conventional one for all w. This means that our method can simplify the estimated fingering with high precision. Moreover, it seems from table 1 that the relation of w to precision is quadratic, with a peak at $w = 0.40$. This result may suggest that the experimental data were played in such ratio of the acoustical reproducibility and fingering easiness.

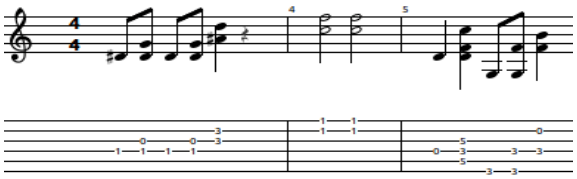
The number of fingering configurations used and the entire fingering cost are shown in table 2. Two examples of produced tablatures are shown in figure 4 for reference. They show that the lower the value of w, the easier the estimated fingering.

Table 2. Number of used fingering configurations and entire fingering cost. Each value is mean for all parts.

Value of w	1.0	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
No. of configurations	19.0	19.3	19.5	18.7	18.0	16.6	14.2	10.4	6.1	4.2
Fingering cost	54.7	36.3	29.6	22.1	16.2	12.4	10.1	9.0	8.6	8.4



(a) $w = 1.00$



(b) $w = 0.50$

Fig. 4. Examples of transcribed tablatures (RM-J007).

5. DISCUSSION

5.1. Validation of fingering cost

Since the fingering cost for the fingering estimation is manually defined by only a few features of the fingering configurations, there still remains doubt about the validity of the cost. Therefore, we validated the cost using tablatures rated by their difficulties [15]. The total number of the used tablatures is 24, and each of them is ranked on a scale of one to six based on the difficulties by the site manager.

The relationship between the fingering cost and the difficulty of the tablatures is shown in figure 5. The correlation coefficient between them was 0.735, thus it can be said that the way of defining the fingering cost is appropriate to some extent. Moreover, the correlation coefficients between each feature of the cost and the difficulty are shown in table 3. It is clear from the table that the fret position and the moving distances of a wrist and fingers are highly related to the difficulty of fingering.

5.2. Relation to prior work

Several methods for estimating fingering from a sequence of pitches have been proposed. Tuohy and Potter [9] proposed a tablature generation method using a genetic algorithm (GA), and others [8, 13] modeled guitar fingerings as a graph search problem and solved it by using a DP technique. Since these methods assume that the input contains only notes that can be played with a guitar, they cannot be used directly for noisy multipitch estimation results including unplayable combinations of pitches. We solved this problem by estimating the optimal fingering on the basis of noisy results obtained using a conventional multipitch estimation method in combination with a basic dynamic programming method and then performing post-processing on it. The fingering decision method proposed by Hori and others [10] is similar to that of our system, and their method supports inputs of unplayable pitch combinations. Therefore, we can make a similar system by directly combining LHA and their method, although it does not consider the difficuly of transcribed tablatures.

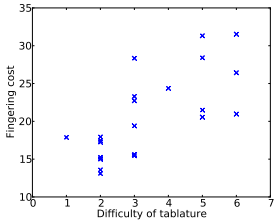


Fig. 5. Relationship between fingering cost and difficulty of tablatures. Correlation coefficient is 0.735.

Table 3. Correlation coefficient between each feature of fingering cost and difficulty of tablatures.

Feature	a_1	a_2	a_3	a_4	a_5	a_6
Correlation coefficient	0.189	0.219	0.798	0.178	0.824	0.756

There are also methods for transcribing a tablature directly from an audio signal. Grady and Rickard [16] proposed a transcription method using a guitar with special equipment to record the sounds played by each string separately. Hrybyk and Kim [17] and Paleari *et al.* [18] combined audio and visual data to identify a player's fingering. Fiss and Kwasinski [11] proposed a real-time transcription system considering the constraints on guitar performance. Barbancho *et al.* [19] used the difference in inharmonicity due to the strings played as a clue for fingering estimation. Other researchers [12, 20] used HMM to model the fingering configuration transitions. Although these are effective methods, there are problems with each of them such as limitation on use due to the requirement for a specially equipped guitar [16] or for visual data corresponding to the audio data [17, 18] and insufficiency of enumerated fingering configurations [12]. Our method requires only audio data and supports over 1000 configurations. Moreover, it considers the proficiency of the player, which is a big difference from previous methods.

5.3. Future work

Currently, a player's proficiency and tendencies are reflected by manually tuning the parameter values. To reduce the burden on the user, we plan to automatically estimate the optimal values on the basis of the player's actual performance or tablatures that the player can already play. We plan to apply an existing method for automatically tuning the optimal values of the parameters by using a steepest descent method [8] and one for evaluating player proficiency by using a fuzzy analytic hierarchy process [21].

Experiments using several values of $\boldsymbol{\theta}$ should be conducted to confirm that our system can reflect the tendency of a player's fingering. Since MIDI data may have less fluctuation, we plan to conduct experiments using real audio data in order to confirm the robustness of our system against noise. We also plan to compare our method with other transcription methods. Other applications of our method, such as music arrangement [10, 22] and music information retrieval (MIR) [23, 24], will also be considered.

6. CONCLUSION

We have developed a tablature transcription method using a conventional multipitch estimation method and a basic dynamic programming method for fingering estimation. The results of experiments showed that the system can transcribe tablatures of various difficulties with highly precise multipitch estimation. Future work includes automatically tuning the optimal parameter values, conducting additional experiments, and investigating applications to other fields of music information processing. This research was partially supported by KAKENHI (S) No. 24700168.

7. REFERENCES

[1] C. E. Leiserson, R. L. Rivest, C. Stein, and T. H. Cormen, *Introduction to algorithms*, The MIT Press, 2001.

[2] K. Yazawa, D. Sakaue, K. Nagira, K. Itoyama, and H. G. Okuno, "Audio-based guitar tablature transcription using multipitch analysis and playability constraints," in *Proc. ICASSP*, 2013, pp. 196–200.

[3] K. Yoshii and M. Goto, "A nonparametric Bayesian multipitch analyzer based on infinite latent harmonic allocation," *IEEE Trans. on ASLP*, vol. 20, no. 3, pp. 717–730, 2012.

[4] D. P. Ellis, "Beat tracking by dynamic programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.

[5] K. Natsubayashi, "k. natsu. brand. 81 - the acoustic guitar site for all beginners -," http://www9.ocn.ne.jp/˜knatsu/, (date last viewed 11/04/13).

[6] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Trans. on SAP*, vol. 13, no. 5, pp. 1035–1047, 2005.

[7] A. Maezawa, K. Itoyama, K. Komatani, T. Ogata, and H. G. Okuno, "Automated violin fingering transcription through analysis of an audio recording," *Computer Music Journal*, vol. 36, no. 3, pp. 57–72, 2012.

[8] A. Radisavljevic and P. Driessen, "Path difference learning for guitar fingering problem," in *Proc. ICMC*, 2004, vol. 28.

[9] D. R. Tuohy and W. D. Potter, "A genetic algorithm for the automatic generation of playable guitar tablature," in *Proc. ICMC*, 2005, pp. 499–502.

[10] G. Hori, H. Kameoka, and S. Sagayama, "Input-output HMM applied to automatic arrangement for guitars," *Journal of Information Processing*, vol. 21, no. 2, pp. 264–271, 2013.

[11] X. Fiss and A. Kwasinski, "Automatic real-time electric guitar audio transcription," in *Proc. ICASSP*, 2011, pp. 373–376.

[12] A. M. Barbancho, A. Klapuri, L. J. Tardon, and I. Barbancho, "Automatic transcription of guitar chords and fingering from audio," *IEEE Trans. on ASLP*, vol. 20, no. 3, pp. 915–921, 2012.

[13] D. Radicioni and V. Lombardo, "Guitar fingering for music performance," in *Proc. ICMC*, 2005, pp. 527–530.

[14] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Popular, classical and jazz music databases," in *Proc. ISMIR*, 2002, vol. 2, pp. 287–288.

[15] K. Yamame, "Guitar TAB. PTB," http://yamame.chu.jp/guitar/, (date last viewed 11/04/13).

[16] P. D. O'Grady and S. T. Rickard, "Automatic hexaphonic guitar transcription using non-negative constraints," in *Proc. ISSC*, 2009, pp. 1–6.

[17] A. Hrybyk and Y. Kim, "Combined audio and video analysis for guitar chord identification," in *Proc. ISMIR*, pp. 159–164.

[18] M. Paleari, B. Huet, A. Schutz, and D. Slock, "A multimodal approach to music transcription," in *Proc. ICIP*, 2008, pp. 93–96.

[19] I. Barbancho, L. J. Tardon, S. Sammartino, and A. M. Barbancho, "Inharmonicity-based method for the automatic generation of guitar tablature," *IEEE Trans. on ASLP*, vol. 20, no. 6, pp. 1857–1868, 2012.

[20] Y. Ueda, Y. Uchiyama, T. Nishimoto, N. Ono, and S. Sagayama, "HMM-based approach for automatic chord detection using refined acoustic features," in *Proc. ICASSP*, 2010, pp. 5518–5521.

[21] K. Yasuhiro, E. Norio, and M. Masanobu, "Evaluating performance proficiency for a chord performance on guitar using fuzzy AHP," 2009, vol. 1, p. 12.

[22] D. R. Tuohy and W. D. Potter, "GA-based music arranging for guitar," in *Proc. Congress on Evolutionary Computation*, 2006, pp. 1065–1070.

[23] J. S. Downie, "Music information retrieval," vol. 37, no. 1, pp. 295–340, 2003.

[24] R. Macrae and S. Dixon, "Guitar tab mining, analysis and ranking," in *Proc. ISMIR*, 2011, pp. 453–458.