ELECTRIC GUITAR PLAYING TECHNIQUE DETECTION IN REAL-WORLD RECORDINGS BASED ON F0 SEQUENCE PATTERN RECOGNITION

Yuan-Ping Chen, Li Su, Yi-Hsuan Yang
Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
qoo0972@hotmail.com, lisu@citi.sinica.edu.tw, yang@citi.sinica.edu.tw

ABSTRACT

For a complete transcription of a guitar performance, the detection of playing techniques such as bend and vibrato is important, because playing techniques suggest how the melody is interpreted through the manipulation of the guitar strings. While existing work mostly focused on playing technique detection for individual single notes, this paper attempts to expand this endeavor to recordings of guitar solo tracks. Specifically, we treat the task as a time sequence pattern recognition problem, and develop a two-stage framework for detecting five fundamental playing techniques used by the electric guitar. Given an audio track, the first stage identifies prominent candidates by analyzing the extracted melody contour, and the second stage applies a pre-trained classifier to the candidates for playing technique detection using a set of timbre and pitch features. The effectiveness of the proposed framework is validated on a new dataset comprising of 42 electric guitar solo tracks without accompaniment, each of which covers 10 to 25 notes. The best average F-score achieves 74% in two-fold cross validation. Furthermore, we also evaluate the performance of the proposed framework for bend detection in five studio mixtures, to discuss how it can be applied in transcribing real-world electric guitar solos with accompaniment.

1. INTRODUCTION

Over the recent years there has been a flourishing number of online services such as Chordify [1] and Riffstation [2] that are dedicated to transcribing the chord progression of real-world guitar recordings [10]. As manual transcription demands on musical training and time, such services, despite not being perfect, make it much easier for music lovers and novice guitar learners to comprehend and appreciate music, thereby creating valuable educational, recreational and even cultural values.

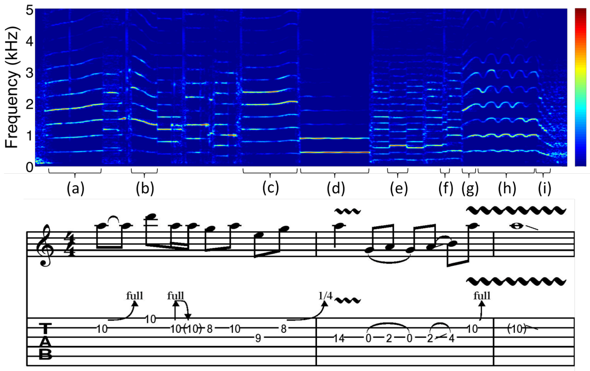

Figure 1. The spectrogram and tablature of a guitar phrase that contains the following techniques: bend (a, b, c, g), vibrato (d, h), hammer-on & pull-off (e) and slide (f, i).

For solo guitar recordings, a note-by-note transcription of the pitches and the playing techniques associated with each note is needed. While the sequence of notes constitutes a melody, playing techniques such as bend and vibrato determine how the notes are played and accordingly influence the expression of the guitar performance. As shown by the guitar tablature in Figure 1, a complete transcription of a guitar performance should contain the notations of the playing techniques. [3]

Unlike pitch estimation or chord recognition, research on playing technique detection is still in its early stages. Most of the existing work, if not all, is only concerned with audio recordings of pre-segmented individual single notes. For example, Abeßer *et al.* [1] collected around 4,300 bass guitar single notes to investigate audio based methods to distinguish between 10 bass guitar playing techniques. Reboursière *et al.* [20] evaluated a number of audio features to detect 6 playing techniques over 1,416 samples of hexaphonic guitar single notes. More recently, Su *et al.* [18] recorded 11,928 electric guitar single notes and investigated features extracted from the cepstrum and phase derivatives to detect 7 playing techniques. It is,

[1] http://chordify.net/ (accessed: 2015-7-15)

[2] http://play.riffstation.com/ (accessed: 2015-7-15)

[3] Fretted stringed instruments such as the guitar usually employ the tablature as the form of musical notation. Various arrows and symbols are used in a guitar tablature to denote the playing techniques. To “generate” the tablature from an audio recording, one would also need to predict the finger positions on the guitar fret, which is beyond the scope of this paper.

however, not clear how these methods can be applied to detect playing techniques in a real-world guitar solo track, such as the one shown in Figure 1.

The only exception, to our best understanding, is the work presented by Kehling *et al.* [16], which extended the work presented in [1] and considered playing technique detection in 12 phrases of guitar solo. They proposed to use onset and off detection first to identify each note event in a guitar solo track, after which the statistics (*e.g.* minimum, maximum, mean, or median) of frame-level spectral features over the duration of each note event are extracted and fed to a pre-trained classifier for playing technique detection. Using the multi-class support vector machine (SVM) with radial basis function (RBF) kernel, they obtained 83% average accuracy in distinguishing between the following 6 cases: *normal*, *bend*, *slide*, *vibrato*, *harmonics*, and *dead notes*. It appears that lower recall rates are found for slide, vibrato, and bend: the recall rates are 50.9%, 66.7%, and 71.3%, respectively.

Although Kehling *et al.*'s work represented an important step forward in playing technique classification, their approach has a few limitations. First, using the whole note event as a fundamental unit in classification cannot deal with techniques that are concerned with the transition between successive notes, such as pull-off and hammer-on, which are also widely used in guitar. Second, extracting features from the whole note may include information not relevant to techniques that are related to only the beginning phase of note events, such as bend and slide. Third, existing techniques for onset and offset detection may not be robust to timbre variations commonly seen in guitar performance [2, 14], originating from the predominant use of sound effects such as distortion or delay [9]. Onset and offset detection would be even more challenging in the presence of accompaniments such as bass and drums.

In light of the above challenges, we propose in this work a new approach to playing technique detection in guitar, by exploiting the time sequence patterns over the melody contour. Given a guitar recording, our approach firstly employs a melody extraction algorithm to estimate the melody contour, *i.e.* sequence of fundamental frequency (F0) estimates. Then, we apply a number of signal processing methods to analyze the estimated melody contour, from which candidate regions of target playing techniques are identified. Because the candidates are identified from the melody contour, the proposed approach can deal with techniques employed during the transition or the beginning phase of notes. The candidate selection algorithms are designed in such a way that emphasizes more on recall rates. Finally, we further improve the precision rates by extracting spectral and pitch features from the candidate regions and using SVM for classification.

The effectiveness of the proposed approach is validated on a new dataset comprising of 42 electric guitar solos taken from the teaching material of the textbook, *Rock Lead Basics: Techniques, Scales and Fundamentals for Guitar*, by Danny Gill and Nick Nolan [13]. While the guitar phrases employed in Kehling *et al.*'s work are not associated with any sound effect [16], the phrases we take from this book are recorded with distortion sound effect and are perceptually more melodic and realistic. Moreover, according to the data from the book, we consider the following five playing techniques in this work: *slide*, *vibrato*, *bend*, *hammer-on*, and *pull-off*, which are viewed as the most frequently used and fundamental techniques in rock lead guitar by the textbook authors.

The guitar solos collected from the book are not accompanied by any other instruments. To examine how the proposed approach can be applied to real-world recordings with accompaniment, we also conduct a preliminary evaluation using 5 well-known guitar solo tracks with different tones and accompaniments. The use of a source separation algorithm as a pre-processing step to suppress the accompaniments is also investigated.

2. DATASETS AND PLAYING TECHNIQUES

Two datasets are employed in this work. The first one is composed of 42 tracks of unaccompanied electric guitar solo obtained from the CD of the textbook by Danny Gill and Nick Nolan. The duration of the tracks is about 15–20 seconds, summing up to about 10 minutes. The tracks are recorded by a standard tuned electric guitar with clean tone and distortion sound effect, covering 10–25 notes per track. For evaluation purposes, we have the timestamps of the playing techniques employed in each track carefully annotated by an experienced electric guitar player, with the help of the corresponding guitar tablature. In total, we have 143 pull-offs, 70 hammer-ons, 143 bends, 74 slides, and 61 vibratos. While the audio tracks are copyright protected, we have made the annotations publicly available with the research community through a project webpage. [4]

The first dataset contains a variety of different possible realizations of the techniques in real-world performances. To illustrate this, we combine a few passages of different phrases and show in Figure 1 the spectrogram and guitar tab. The five playing techniques and their possible variations are described below.

- **Bend** refers to stretching the string with left hand to increase the pitch of the bended note either gradually or instantly. The region (a) in Figure 1 shows a note *full-bended* from `A4` to `B4` gradually. In (b), the note is *pre-bended* to `B4`, *i.e.* bend the note without sounding it, and then *released* to `A4` with the hitting of string. Region (c) shows a *half-step* bend commonly seen in Blues. A *grace note* bend is when you strike the string and at the same time bend the note to the target, as shown in (g).
- **Vibrato** represents minute and rapid variations in pitch. Regions (d) and (h) of Figure 1 show a very subtle vibrato with smaller extent and a wide vibrato with larger extent, respectively.

[4] `http://mac.citi.sinica.edu.tw/GuitarTranscription`. Note that we label the instant of transition between two notes for pull-off and hammer-on, the middle of the employment of bend and slide, and the whole duration (including the beginning and end timestamps) for vibrato.

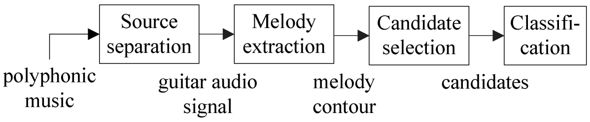

Figure 2. Flowchart of the proposed approach to guitar playing technique detection.

- **Hammer-on** is when a note is sounded, a left hand finger is used to quickly press down a fret that is on the same string while the first note is still ringing.
- **Pull-off** is when you have strummed one note and literally pull off of the string to a lower note. Rapid and successive use of pull-of and hammer-on is often referred to as *trill*, which is illustrated in (e).
- **Slide** refers to the action of slide left hand finger across one or more frets to reach another note. A slide between `B3` and `D4` is shown in (f). There are *shift* slides and *legato* slides. A guitar solo usually begins or ends with another variant known as *slide from/into nowhere*," which is illustrated in (i).

The second dataset, on the other hand, consists of 5 excerpts of real-world guitar solo (with accompaniment) clipped from the following famous recordings: segments 1'48"–2'39" and 2'51"–3'23" from *Bold as Love* by Jimi Hendrix, segments 0'17"–1'26" and 3'50"–4'33" from *Coming Back to Life* by Pink Floyd, and segment 4'22"–5'04" from *Wet Sand* by Red Hot Chili Peppers. The first two are both played in fuzzy tone (akin to overdrive), the third one with reverb effect in clean tone, the fourth one in overdrive, and the fifth one is played with the distortion effect. The excerpts last 3 minutes 57 seconds in total. We also manually label the playing techniques for evaluation.

3. PROPOSED APPROACH

3.1 Overview

Kehling *et al.* [16] employs a two-stage structure in detecting playing techniques in audio streams. The first stage uses onset and offset detection to identify each note event from the given audio track, and the second stage applies a pre-trained classifier to the note events for multiclass classification. A similar two-stage structure is also adopted in the proposed approach, but in our first stage we make use of the melody contour extracted from the given audio track, and employ a number of algorithms to identify candidates of playing techniques from the melody contour. Different candidate selection algorithms are specifically designed for the 5 playing techniques. Depending on the target playing technique, the input to the second-stage classifier can be temporal segments falling between note events or fragments of whole note events. In this way, the proposed approach can deal with techniques such as hammer-on and pull-off, while Kehling *et al.*'s approach cannot.

Figure 2 shows the flowchart of the proposed approach, which includes source separation as an optional pre-processing step to cope with instrumental accompaniments. We provide the details of each component below.

3.2 Source Separation

In real-world guitar performance, the guitar solo is usually mixed with strong bass line, percussion sounds, or others. Due to the accompaniments, the performance of estimating the melody contour of the lead guitar may degrade.

We experiment with the robust principal component analysis (RPCA) algorithm [6, 7, 15] to separate the sound of the lead guitar from the accompaniments, before extracting the melody. Given the magnitude spectrogram $\mathbf{D} \in \mathbb{R}^{t \times m}$ of the mixture, where t denotes temporal length and m the number of frequency bins, RPCA seeks to decompose $\mathbf{D}$ into two matrices of the same size, a low-rank matrix $\mathbf{A}$ and a sparse matrix $\mathbf{E}$, by solving the following convex optimization problem:

$$\min_{\mathbf{A},\mathbf{E}:\, \mathbf{D}=\mathbf{A}+\mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \,, \tag{1}$$

where the trace norm $\|\cdot\|_*$ and l_1 norm $\|\cdot\|_1$ are convex surrogate of the rank and the number of nonzero entries of a matrix, respectively [6], and λ is a positive weighting parameter. As the *background* component of a signal is usually composed of repetitive elements in time or frequency, its spectrogram is likely to have a lower rank comparing to that of the *foreground*. RPCA has been applied to isolating the singing voice (foreground) from the accompaniment (background) [15]. We use the same idea, assuming that the guitar solo is the foreground (i.e. in $\mathbf{E}$).

3.3 Melody Extraction

Melody extraction has been an active field of research in the music information retrieval society for years [5, 8, 19]. It is concerned with the F0 sequence of only the main melody line in a polyphonic music recording. Therefore, it consists of a series of operations for creating candidate pitch contours from the F0 estimates and for selecting one of the pitch contours as the main melody. We employ the state-of-the-art melody extraction algorithm proposed by Salamon and Gòmez [21], for its efficiency and well-demonstrated effectiveness. Specifically, we employ the implementation of the MELODIA algorithm developed by the authors for an open-source library called Essentia [3]. It is easy to use and the estimate is in general accurate.

3.4 Candidate Selection (CS)

We propose to mine the melody contour for the following time sequence patterns specific to each playing technique. Following this process of pattern finding, we can find candidates of the playing techniques scattered in the time flow of a music signal. We refer to this process as candidate selection, or CS for short.

- Bend: arc-like or twisted trajectories.
- Vibrato: sinusoidal patterns.
- Slide: upward or downward stair-like patterns.
- Hammer-on, pull-off: two adjacent parallel horizontal lines resulting from two notes of different F0s.

Clearly, such patterns may not necessarily correspond to true instances (or, true positives) of the playing techniques. For example, sounding two notes with pick picking

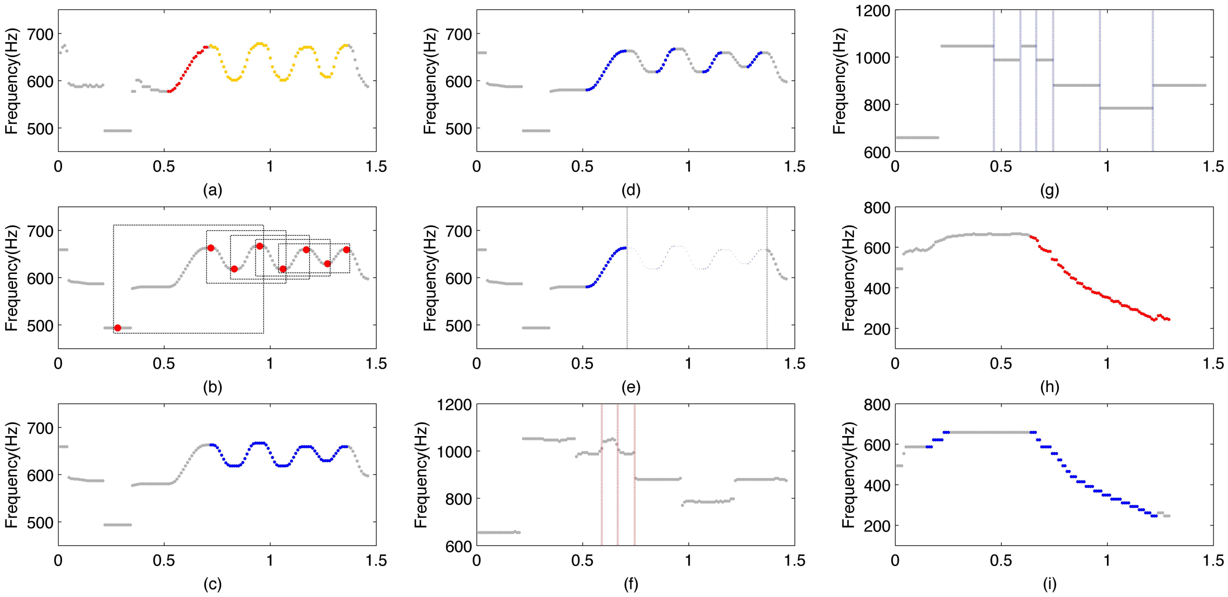

Figure 3. The procedure of candidate selection (best seen in color). (a) The raw melody contour of a bend (red segment) and a vibrato (yellow segment). (b) The processed melody contour by median filter, note tracking and mean filter. Four local extrema of pitch value create a window to determine vibrato. (c) The candidate segment for vibrato (blue). (d) The candidate segments for bend (blue). (e) The candidate segments for bend, after excluding candidates of vibrato (blue). (f) The raw melody contour of a pull-off and a hammer-on. The red vertical lines show the groundtruth instants of the playing techniques. (g) The processed melody contour by note tracking and quantization, and the blue vertical lines denote the candidates instants. (h) The raw melody contour of a "slide into nowhere" (red segment). (i) The processed melody contour by quantization, and the selected candidates for slide (blue segments).

also results in a pitch trajectory of two parallel horizontal lines akin to the case of hammer-on or pull-off. There might also be errors in the estimate of the melody contour (*e.g.* when the lead instrument is silent, the estimated melody contour may correspond to the sounds of other instruments). Therefore, the purpose of the CS process is actually to identify the candidates with high recall rates (*i.e.* not missing the true positives) and moderate precision rates (*i.e.* it is fine to have false positives). In the next stage, we will use SVM classifiers that are discriminatively trained to distinguish between true positives and false positives by exploiting both timbre and pitch features computed from these candidates. Because the CS process only considers pitch information, the additional use of timbre information in the classification stage has the potential to boost the precision rates.

As described below, the CS process is accomplished with a few simple signal processing methods for simplicity and efficiency. The methods are illustrated in Figure 3.

3.4.1 Vibrato and Bend

We use similar procedures to select the candidates of vibrato and bend, because the two techniques share the same arc-like trajectories. Indeed, a vibrato can be viewed as succession of bend up and then releasing down. The two techniques mainly differ in the number of the cycles. The following operations are firstly employed to process the (raw) melody contour estimated by MELODIA [3].

- First, we flatten the rugged raw contour and remove the outliers produced by the melody extraction algorithm by a 10 points (100ms in 44.1 kHz sampling rate) *median filter*, whose length is approximately shorter than a period of vibration. The median filter also slightly corrects octave errors made by melody tracking.
- Second, we perform a simple *note tracking* step by grouping adjacent F0s into the same note if the pitch difference between them is smaller than 80 cents, according to the auditory streaming cues [4]. The step leads to a number of segments corresponding to different note events, from which segments shorter than 80ms are discarded, assuming that the a single note should last at least 80ms, approximately the length of a semiquaver in 180 BPM.
- Finally, we convolve each segment with a 5 points (50ms) *mean filter* with hop of 10ms for smoothing.

The segments are then considered as possible note events. We then use different ways to detect vibrato and bend. For vibrato, we search for all the local maxima and minima in each note [12]. A temporal fragment of four consecutive extrema within a note is considered as a vibrato candidate if the following conditions meet: 1) the temporal distance between two neighboring extrema should fall within 30ms and 400ms for valid vibrato rate, i.e., the modulation frequency from 1.25Hz to 16.67Hz; 2) the pitch difference between neighboring extrema should be smaller than 225 cents, which is slightly larger than a whole note; 3) dividing the fragment into three shorter fragments of pitch sequence by the four extrema, the variance in the logarithmic pitch of each short fragment should be larger than an empirical threshold. Please see Figure 3(c) for an example.

On the other hand, we consider a temporal fragment as a bend candidate if the following conditions meet: 1) it is not a vibrato candidate; 2) the pitch sequence continuously ascends or descends for more than 80ms; 3) the pitch difference between two neighboring frames is smaller than 50 cents. An example can be found in Figure 3(e).

3.4.2 Pull-off and Hammer-on

While bend and vibrato can last a few frames, pull-off and hammer-on are considered as the temporal instance (*i.e.* a frame) during the transition of notes. Therefore, without using either a median or mean filter, we perform the note tracking procedure described in Section 3.4.1, and then *quantize* each F0 to its closest semitone in terms of cent. After this, we consider all the temporal instances in the middle of two notes as a candidate for both pull-off and hammer-on, as long as the following conditions meet: 1) the gap between the note transition is shorter than 20ms; 2) the pitch cannot be away from its closest semitone by 35 cents. The former condition is set, because it is known that the contact of pick (or right hand finger) and the string would temporarily stop the vibration of the string when a note is sounded by plucking the string, thereby creating the gap in the note transition [20]. The latter condition is set because there might be such gaps within the employment of a vibrato or a bend due to the F0 quantization.

Because each candidate for pull-off or hammer-on only lasts one frame, to characterize the temporal moment, we use a 100ms fragment centering at the candidate frame for the feature extraction step described in Section 3.5.

3.4.3 Slide

To recognizing the ladder-like pitch sequence pattern, we simply quantize all the F0s into its closest semitone without any pre-processing, in order not to falsely remove the transition notes of a bend (which is usually around tens of milliseconds). After quantization, we search for the ladders in the melody contour with the following rules: 1) the number of steps should be at least three (*i.e.* slide across at least three frets); 2) the length of transitional steps (notes) should fall within 10 to 70ms, according to our empirical observation from the data; 3) the pitch difference between neighboring steps should be exactly one semitone (*i.e.* a fret). Please refer to Figure 3(i) for an example.

3.5 Feature Extraction and Classification

After applying CS, we would have candidates of the 5 playing techniques spreading over the input guitar track. As we have mentioned, our design of the signal processing methods and the setting of some parameter values have been informed by the need of reaching high recall rate. It is then the job of the classifier to identify false positives of the techniques and improve precision rates.The candidates are represented by the following three sets of audio features.

- **TIMBRE** (T) includes the statistics of the following features: spectral centroid, brightness, spread, skewness, kurtosis, flux, roll-off, entropy, irregularity, roughness, inharmonicity, zero-crossing rate, low-energy ratio, and their 1st-order time difference. We use the mean, standard deviation (STD), maximum, minimum, skewness, kurtosis as the statistics measure, so there are $13\times8\times2=208$ features in total.
- **MFCC** (M) contains mean and STD of the 40-D Mel-frequency cepstral coefficients and its 1st-order time difference, totalling 160 features. Both the TIMBRE and MFCC sets are computed by the open-source library MIRtoolbox [17].
- **Pitch** (P) is computed from the log scale F0 sequence on the processed (instead of the raw) melody contour. Except for vibrato, we use the following 6 features for all the playing techniques: skewness, kurtosis, variance, the difference between the maximum and minimum, and the mean and STD of the 1st-order time difference. For vibrato, as there are 3 short temporal fragments for each candidate (see Section 3.4.1), we calculate the 6 features for each of the fragment, and additionally use the variance of difference between the four pitch extrema in log scale and the variance of the temporal distance between the four pitch extrema, totalling 20 features.

4. EXPERIMENT

4.1 Experimental Setup

For short-time Fourier transform, we use the Hamming window of 46ms and 10ms overlap under the sampling rate of 44.1 kHz. For MELODIA, we set the lowest and highest possible F0 to 77Hz (`E2b`) and 1400 (`F6`) respectively, considering the frequency range of a standard-tuned guitar plus additionally half step tolerance of inaccurate tuning. We train 5 binary linear kernel SVMs [11], one for each technique, [5] and employ z-score normalization for the features. The parameter C of SVM is optimized by an inside cross validation on the training data. We conduct training and testing 10 times under a two-fold cross validation scheme and report the average result, in terms of precision, recall and F-score. An estimate of bend or slide is deemed correct as long as the ground truth timestamp falls between the detected bend or slide segment. An estimate of pull-off or hammer-on is deemed correct if the detected instant of employment falls between the interval of ground truth instant with a tolerance time-window of 50ms. Vibrato is evaluated in the frame level, *e.g.* the recall of vibrato is the proportion of frames labeled as vibrato which are detected as vibrato. For evaluation on the studio mixtures, the SVM is trained over the 42 unaccompanied phrases. Source separation is only performed for the 5 studio mixtures.

[5] It would have been better if a multi-class classification scheme is adopted to avoid possible overlaps of the estimates of different techniques. We leave the issue as a future work.

	Bend	Vibrato	Pull-off	Hammer-on	Slide
Recall	94.4	94.2	94.4	94.3	85.1
Precision	53.1	63.0	30.1	24.7	15.0
F-score	68.0	75.5	45.7	39.2	25.5

(a)

	Bend	Vibrato	Pull-off	Hammer-on	Slide
Recall	86.2	79.5	73.6	65.7	58.6
Precision	89.3	89.1	75.3	66.7	56.8
F-score	87.7	84.0	74.4	66.3	57.7

(b)

Table 1. Recall, precision, and F-scores (in %) of playing technique detection in the unaccompanied set using (a) CS only and (b) CS+SVM{MFCC,TIMBRE,Pitch}.

4.2 Expriment Result

4.2.1 Evaluation on Unaccompanied Guitar Solos

Table 1 shows the per-class result of playing technique detection over the 42 unaccompanied guitar solos, using either (a) only candidate selection (CS) or (b) both CS and SVM. The following observations can be made.

- Except for slide, the proposed CS methods can lead to recall rates higher than 94% for the considered playing techniques. Slide appears to be the most challenging one, as its detection can be affected by octave errors from the melody extraction step.
- By comparing Tables 1(a) and (b), we see that the second-stage SVM can remarkably improve the precision rates, and accordingly the F-scores, for all the playing techniques. This validates the effectiveness of the proposed approach.
- Bend and vibrato appear to be easier to detect, with F-scores 87.7% and 84.0%, respectively. Although it is not fair to compare the numbers with the ones reported in [16] due to differences in settings and datasets, the performance of the proposed approach seems to be promising. Interestingly, slide appears to be the most challenging case in our study and the one presented by [16], with comparable F-scores (57.7% versus 50.9%).

Figure 4 provides the F-scores of using different features in the SVM. Although not shown in the figure, MFCC appears to be the best performing individual feature set among the three. Better result is seen by concatenating the features (*i.e.* early fusion). Pitch features contribute more to the detection of hammer-on but less for others, possibly because pitch information has been exploited in CS.

4.2.2 Evaluation on Real-World Studio Mixtures

As bend detection is found to be promising, we focus on bend detection for our evaluation over the 5 studio mixtures, which include in total 85 bends. Figure 5 compares the F-score of bend detection of various methods, including the case when RPCA is employed before melody extraction. It is not surprising that the F-scores are lower than that obtained for the unaccompanied tracks. However, it is interesting to note that the best result can be obtained by CS only, regardless of whether RPCA or SVM is used. Actually, the result of using CS+SVM degrades a lot comparing to the case of CS only, except for the case that pitch features are considered in SVM. The performance of CS+SVM can be improved by using RPCA, but the result is still inferior to the result of CS only. We conjecture that the inferior result of CS+SVM can be attributed to the difference between the data used for training the SVM (*i.e.* the unaccompanied tracks) and the data for testing (*i.e.* the mixtures). The result might be better if we have a few training data that are with accompaniment. However, if such data are not available, it seems to be advisable to use the CS process only for the bend detection in mixtures.

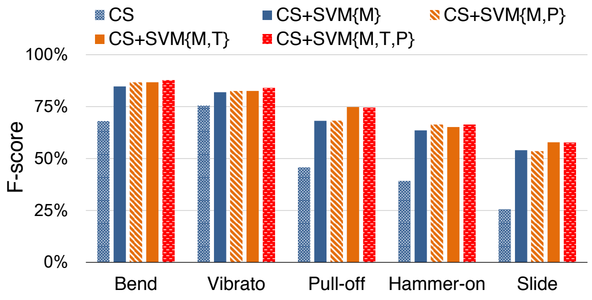

Figure 4. F-scores of playing technique detection in 42 unaccompanied guitar solos using various methods.

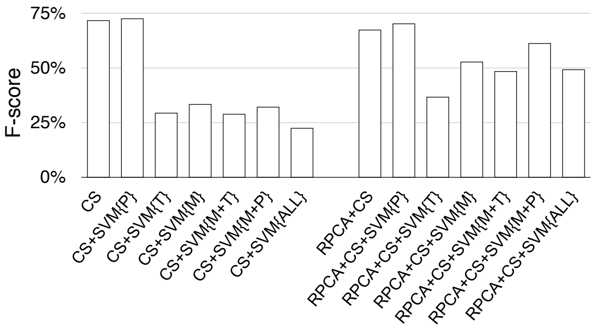

Figure 5. F-scores of bend detection of 5 accompanied guitar solos, without (left) or with (right) RPCA.

5. CONCLUSION

In this paper, we have presented a two-stage approach for detecting 5 guitar playing techniques in guitar solos. The proposed approach is characteristic of its use of time sequence patterns mined from the melody contour of the lead guitar for candidate selection in the first stage, and then using classifiers to refine the result in the second stage. The F-scores for the unaccompanied set range from 57.7% to 87.7% depending on the playing techniques. The average F-score across the techniques reaches 74%. We have also evaluated the case of bend detection for a few guitar solos with accompaniment, and shown that the best F-score 67.3% is obtained by candidate selection alone.

6. ACKNOWLEDGMENT

This work was supported by the Academia Sinica Career Development Program.

7. REFERENCES

[1] J. Abeßer, H. Lukashevich, and G. Schuller. Feature-based extraction of plucking and expression styles of the electric bass guitar. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pages 2290–2293, 2010.

[2] S. Bock and G. Widmer. maximum filter vibrato suppression for onset detection. In *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx)*, 2013.

[3] D. Bogdanov, N. Wack, E. Gòmez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra. Essentia: an audio analysis library for music information retrieval. In *Proc. Int. Soc. Music Information Retrieval Conf.*, pages 493–498, 2013. [Online] `http://essentia.upf.edu`.

[4] A. S. Bregman, editor. *Auditory scene analysis*. MIT Press, 1990.

[5] P. M. Brossier. *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis, Queen Mary, University of London, 2006.

[6] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011.

[7] T.-S. Chan, T.-C. Yeh, Z.-C. Fan, H.-W. Chen, L. Su, Y.-H. Yang, and R. Jang. Vocal activity informed singing voice separation with the iKala dataset. In *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, pages 718–722, 2015.

[8] A. De Cheveigné and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

[9] J. Dattorro. Effect design, part 2: Delay line modulation and chorus. *J. Audio engineering Society*, 45(10):764–788, 1997.

[10] W. B. de Haas, J. P. Magalhães, and F. Wiering. Improving audio chord transcription by exploiting harmonic and metric knowledge. In *Proc. Int. Soc. Music Information Retrieval Conf.*, pages 295–300, 2012.

[11] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *J. Machine Learning Research*, 2008. `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`.

[12] A. Friberg and E. Schoonderwaldt. Cuex: An algorithm for automatic extraction of expressive tone parameters in music performance from acoustic signals. *Acta Acustica united with Acustica*, 93(3):411–420, 2007.

[13] D. Gill and N. Nolan. *Rock Lead Basics: Techniques, Scales and Fundamentals for Guitar*. Musicians Institute Press, Los Angeles, California, 1997.

[14] A. Holzapfel, Y. Stylianou, A. C. Gedik, and B. Bozkurt. Three dimensions of pitched instrument onset detection. *IEEE Trans. Audio, Speech, and Language Processing*, pages 1517–1527, 2010.

[15] P.-S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson. Singing-voice separation from monaural recordings using robust principal component analysis. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pages 57–60, 2012.

[16] C. Kehling, J. Abeßer, C. Dittmar, and G. Schuller. Automatic tablature transcription of eletric guitar recordings by estimation of score-and instrument-related parameters. In *Proc. Int. Conf. Digital Audio Effects*, 2014.

[17] O. Lartillot and P. Toiviainen. MIR in Matlab (II): A toolbox for musical feature extraction from audio. In *Proc. Int. Soc. Music Information Retrieval Conf.*, pages 127–130, 2007. [Online] `http://users.jyu.fi/~lartillo/mirtoolbox/`.

[18] L.Su, L.-F. Yu, and Y.-H. Yang. Sparse cepstral and phase codes for guitar playing technique classification. In *Proc. Int. Soc. Music Information Retrieval Conf.*, pages 9–14, 2014.

[19] M. Müller, D. P. W. Ellis, A. Klapuri, and G. Richard. Signal processing for music analysis. *IEEE J. Sel. Topics Signal Processing*, 5(6):1088–1110, 2011.

[20] L. Reboursière, O. Lähdeoja, T. Drugman, S. Dupont, C. Picard, and N. Riche. Left and right-hand guitar playing techniques detection. In *Proc. Int. Conf. New Interfaces for Musical Expression*, 2012.

[21] J. Salamon and E. Gòmez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Trans. Audio, Speech, and Language Processing*, 20(6):1759–1770, 2012.