ANALYSIS OF EXPRESSIVE MUSICAL TERMS IN VIOLIN USING SCORE-INFORMED AND EXPRESSION-BASED AUDIO FEATURES

Pei-Ching Li\textsuperscript{1} Li Su\textsuperscript{2} Yi-Hsuan Yang\textsuperscript{2} Alvin W. Y. Su\textsuperscript{1}
\textsuperscript{1} SCREAM Lab., Department of CSIE, National Cheng-Kung University, Taiwan
\textsuperscript{2} MAC Lab., CITI, Academia Sinica, Taiwan
p78021015@mail.ncku.edu.tw, lisu@citi.sinica.edu.tw
yang@citi.sinica.edu.tw, alvinsu@mail.ncku.edu.tw

ABSTRACT

The manipulation of different interpretational factors, including dynamics, duration, and vibrato, constitutes the realization of different expressions in music. Therefore, a deeper understanding of the workings of these factors is critical for advanced expressive synthesis and computer-aided music education. In this paper, we propose the novel task of automatic expressive musical term classification as a direct means to study the interpretational factors. Specifically, we consider up to 10 expressive musical terms, such as \textit{Scherzando} and \textit{Tranquillo}, and compile a new dataset of solo violin excerpts featuring the realization of different expressive terms by different musicians for the same set of classical music pieces. Under a score-informed scheme, we design and evaluate a number of note-level features characterizing the interpretational aspects of music for the classification task. Our evaluation shows that the proposed features lead to significantly higher classification accuracy than a baseline feature set commonly used in music information retrieval tasks. Moreover, taking the contrast of feature values between an expressive and its corresponding non-expressive version (if given) of a music piece greatly improves the accuracy in classifying the presented expressive one. We also draw insights from analyzing the feature relevance and the class-wise accuracy of the prediction.

1. INTRODUCTION

The expressive meaning of music is generally related to two inter-dependent factors: the \textit{structure} established by the composer (e.g., mode, pitch, or dissonance) and the \textit{interpretation} of the performer (e.g., expression) [21]. Glenn Gould could phrase the \textit{trills} in a way different from other pianists. Mozart’s \textit{Grazioso} should be interpreted unalike to Brahms’. Although the interplay between the structural and interpretational factors makes it difficult to characterize musical expressiveness from audio signals, it has been pointed out that such analysis is valuable in emerging applications such as automatic music transcription, computer-aided music education, or expressive music synthesis [2,4,7,19]. Accordingly, computational analysis of the interpretational aspects in music expression has been studied for a while. For example, Bresin \textit{et al.} analyzed the statistical behaviors of \textit{legato} and \textit{staccato} played with 9 expressive adjectives (not expressive musical terms) [3]. Grachten \textit{et al.} made both predictive and explanatory modeling on the dynamic markings (e.g., \textit{f}, \textit{p}, \textit{fz}, and \textit{crescendo}) [10]. Ramirez \textit{et al.} considered an approach of evolutionary computing for general timing and energy expressiveness [18]. Marchini \textit{et al.} analyzed the performance of string quartets by the following three terms: \textit{mechanical}, \textit{normal} and \textit{exaggerated} [14]. Recently, Roda \textit{et al.} further considered expressive constants as affective dimensions of music [20]. Related works also include the identification of performers, singers and instrument playing techniques in the context of musical expression [1,6,12,15].

To model specific aspects of the complicated music expression quantitatively, a machine learning based approach is usually taken. Given an audio input, features are extracted to characterize the interpretational aspects of music, such as the dynamics, tempo and vibrato [3,9,12,14].\textsuperscript{1} If the symbolic or score data such as the MIDI or MusicXML are available, one can further introduce more structural aspects including tonality, pitch, note duration and measure, amongst others [10,15,16]. In [14], the synchronized audio, score and even motion data are utilized to generate 4 sets of features, including sound level, note lengthening, vibrato extent and bow velocity, in an attempt to reveal human behaviors while playing the instrument or indicate the structural information of music. This way, the features investigated have music meanings, and can be adopted for specific applications such as the prediction and the generation of expressive performances [10,18].

Among all the objects of music expression, we notice that the \textit{expressive musical terms} (EMT)\textsuperscript{2} have garnered less attention in the literature, although they have been

\textsuperscript{1} Here we assume that any real-world interpretation of an expressive musical term performed by a musician can be “atomized” into several (independent) factors such as dynamics, tempo, and vibrato.

\textsuperscript{2} In this paper, the expressive musical term is defined as the Italian musical term which describes an emotion, feeling, image or metaphor, rather than merely an indication of tempo or dynamics. It includes, but not limited to the emotional terms (see Table 1).

widely used in specifying expressions of classical music for hundreds of years. How the interpretational factors (dynamics, duration or vibrato) are taken for a musician to interpret the terms is still not well understood. This might be due to the lack of a dataset containing various interpretations for a fixed set of classical music pieces.

In this paper we address these issues, and particularly, focus on the classification of expressive musical terms in violin solo music. We compile a new dataset of solo violin excerpts featuring the realization of 10 expressive terms and 1 non-expressive term (e.g., no expression) by 11 different musicians for 10 classical music pieces (Section 2). After collecting the MIDI and MusicXML data for the music pieces, we design a number of dynamic-, duration- and vibrato-based features under a score-informed scheme (Section 3.2). Moreover, we also consider a baseline feature set comprising of standard audio features that can be computed without score information, such as the Mel-frequency cepstral coefficients (MFCCs), spectral flux, spectral centroid, and the zero-crossing rate (Section 3.1). As such features have been widely used in music information retrieval tasks like the classification of mood, genre or instruments [25], we want to know whether they are also useful for classifying the expressive musical terms. However, we should note that many of the baseline features do not bear clear music meanings as the proposed features do. In our experiments, we will evaluate the performance of these features for expressive musical term classification, and analyze the importance of such features (Section 4).

The dataset is referred to as the SCREAM-MAC-EMT dataset. For reproducibility and for calling more attention to this research problem, we have made the audio files of the recordings publicly available online.³

<table>
<thead>
<tr>
<th>Violin pieces</th>
<th>Measure</th>
<th>Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>W. A. Mozart - Variationen</td>
<td>1-24</td>
<td>None, Scherzando, Tranquillo, Con Brio, Maestoso, Risoluto</td>
</tr>
<tr>
<td>T. A. Vitali - Chaconne</td>
<td>1-9</td>
<td>None, Scherzando, Affettuoso, Con Brio, Agitato, Cantabile</td>
</tr>
<tr>
<td>G. Faure - Elegie</td>
<td>2-9</td>
<td>None, Scherzando, Affettuoso, Con Brio, Agitato, Cantabile</td>
</tr>
<tr>
<td>P. I. Tchaikovsky - String Quartet, No. 1, Mov. II</td>
<td>1-16</td>
<td>None, Affettuoso, Tranquillo, Con Brio, Agitato, Cantabile</td>
</tr>
<tr>
<td>M. Bruch - Violin Concerto, No. 1, Mov. I</td>
<td>6, 10 (solo, ad lib.)</td>
<td>None, Scherzando, Affettuoso, Con Brio, Agitato, Cantabile</td>
</tr>
<tr>
<td>A. Vivaldi - La primavera, Mov. I</td>
<td>1-13</td>
<td>None, Scherzando, Affettuoso, Con Brio, Agitato, Cantabile</td>
</tr>
<tr>
<td>A. Vivaldi - La primavera, Mov. II</td>
<td>2-11</td>
<td>None, Affettuoso, Tranquillo, Agitato, Maestoso, Cantabile</td>
</tr>
<tr>
<td>E. Elgar - Salut d’Amour</td>
<td>3-17</td>
<td>None, Affettuoso, Grazier, Agitato, Expressivo, Maestoso</td>
</tr>
<tr>
<td>A. Vivaldi - L’autunno, Mov. I</td>
<td>1-13</td>
<td>None, Affettuoso, Grazier, Agitato, Expressivo, Maestoso</td>
</tr>
<tr>
<td>A. Vivaldi - L’autunno, Mov. II</td>
<td>1-29</td>
<td>None, Tranquillo, Grazier, Con Brio, Expressivo, Risoluto</td>
</tr>
</tbody>
</table>

³https://sites.google.com/site/pclipatty/scream-mac-emt-dataset

The following 10 expressive terms are considered: Tranquillo (calm), Grazier (graceful), Scherzando (playful), Risoluto (rigid), Maestoso (majestic), Affettuoso (affectionate), Expressivo (expressive), Agitato (agitated), Con Brio (bright), and Cantabile (like singing).⁴ In order to have a balanced dataset, we require that each expressive musical term is associated with 5 pieces. This is not easy, because not all of the 20 pieces can be interpreted with diverse expressions. Eventually, some compromises have to be made. For example, we chose Maestoso instead of Cantabile for Elgar’s Salut d’Amour, although the former is somewhat awkward for this music piece. The resulting selection of the music pieces and the assigned expressions is shown in Table 1.

After selecting the music pieces, we recruited 11 professional violinists to perform them one by one in a real-world environment. In addition to the 5 assigned terms, every musician performed a non-expressive (denoted as None) version for each piece. Here, None means mechanical interpretation [14] by which the music is of constant dynamics, constant tempo and no vibrato. The dataset therefore contains 660 excerpts as there are 10 classical music pieces and each piece is interpreted by 6 different versions by all the 11 violinists. We have 110 excerpts of None, and 55 excerpts for each of the 10 expressions.

3. METHOD

Figure 1 shows the proposed system diagram. At the first stage of the system, the input audio signal is aligned with professional violinists, who are active in classical music performance, to select 10 pieces from the list and assign 5 suitable expressive musical terms for each of them. The major criterion of selecting the music pieces, as it turns out, requires that an excerpt has a simple melody that can be effectively manipulated to exhibit different characteristics when being interpreted with different expressions.

In order to have a balanced dataset, we require that each expressive musical term is associated with 5 pieces. This is not easy, because not all of the 20 pieces can be interpreted with diverse expressions. Eventually, some compromises have to be made. For example, we chose Maestoso instead of Cantabile for Elgar’s Salut d’Amour, although the former is somewhat awkward for this music piece. The resulting selection of the music pieces and the assigned expressions is shown in Table 1.

After selecting the music pieces, we recruited 11 professional violinists to perform them one by one in a real-world environment. In addition to the 5 assigned terms, every musician performed a non-expressive (denoted as None) version for each piece. Here, None means mechanical interpretation [14] by which the music is of constant dynamics, constant tempo and no vibrato. The dataset therefore contains 660 excerpts as there are 10 classical music pieces and each piece is interpreted by 6 different versions by all the 11 violinists. We have 110 excerpts of None, and 55 excerpts for each of the 10 expressions.

⁴For more information, see http://www.musictheory.org.uk/res-musical-terms/italian-musical-terms.php
Table 2: Proposed features, the note-level and song-level aggregation methods.

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbreviation</th>
<th>Note-level aggregation</th>
<th>Song-level aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamics</td>
<td>D</td>
<td>M, Max, maxPos</td>
<td>M, S, C₀Μ</td>
</tr>
<tr>
<td>Duration</td>
<td>ND, 1MD, 2MD, 4MD</td>
<td>—</td>
<td>M, S, C₀Μ</td>
</tr>
<tr>
<td></td>
<td>PPD</td>
<td>—</td>
<td>—, C₀Μ</td>
</tr>
<tr>
<td>Vibrolato rate</td>
<td>VR</td>
<td>M, S, MΔ, SΔ, Max, Min, Diff</td>
<td>M, S, C₀Μ</td>
</tr>
<tr>
<td>Vibrolato extent</td>
<td>VE</td>
<td>M, S, MΔ, SΔ, Max, Min, Diff</td>
<td>M, S, C₀Μ</td>
</tr>
<tr>
<td>Global vibrolato extent</td>
<td>GVE</td>
<td>—</td>
<td>M, S, C₀Μ</td>
</tr>
<tr>
<td>Vibrolato ratio</td>
<td>vibRatio</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

3.1 Baseline Features

The baseline features are a rich set of audio features covering dynamics, rhythm, tonal, and timbre. In particular, the baseline features are a rich set of temporal, spectral, cepstral and harmonic descriptors. It contains the mean and standard deviation of spectral centroid, brightness, spread, skewness, kurtosis, roll-off, entropy, irregularity, flatness, roughness, inharmonicity, flux, zero-crossing rate, low energy ratio, attack time, attack slope, dynamics and the mean and standard deviation of first-order temporal difference for all the above features, totaling 4 × 17 = 68 features. Besides, it involves the mean of fluctuation peak and centroid, tempo, pulse clarity and event density, generating 5 features; the mean and standard deviation of mode and key clarity, resulting 4 features. Furthermore, it includes the mean and standard deviation of the 40-D MFCCs, ΔMFCCs (first-order temporal difference) and ΔΔMFCCs (second-order temporal difference), totaling 2 × 120 = 240 features. In sum, we have 317 features extracted by the MIRtoolbox (version 1.3.4) [13].

3.2 Proposed Features

3.2.1 Dynamic Features

The dynamics of each note is estimated from the short-time Fourier transform (STFT). Given a segmented note \(x(n)\) and the Hanning window function \(w(n)\), the STFT is represented as \(X^w(n, k) = M^w(n, k)e^{j\Phi^w(n, k)}\), where \(M^w(n, k)\) is the magnitude part, \(\Phi^w(n, k)\) is the phase part, \(n\) is the time index, and \(k\) is the frequency index. The dynamic level function \(D(n)\) is computed by the summation of the magnitude spectrogram over the frequency bins and is expressed in dB scale:

\[
D(n) = 20 \log_{10} \left( \sum_k M(n, k) \right).
\]  

Three note-level dynamic features are computed from \(D(n)\). Each of them are the mean value of \(D(n)\) (D-M), the maximal value of \(D(n)\) (D-Max) and the proportion of the maximum position to the note length (D-maxPos):

\[
\text{maxPos} = \arg \max_n \frac{D(n)}{\text{length}(D(k))} \times 100\%.
\]  

D-maxPos therefore measures the time a note reaches its maximal energy from its beginning, normalized to the length of the note. All of these three note-level features are then aggregated to song-level by M, S, and C₀Μ, totaling 9 features (see the second row of Table 2). For the \(D(n)\) calculation, frames of 23ms (1014 samples) with an 82% overlap (832 samples), as used in [14], are adopted.

3.2.2 Duration Features

After score alignment and note segmentation, we take the following values as the features: the duration of every single note (ND), measure (1MD), two-measure segment (2MD), four-measure segment (4MD), and the full piece (FPD) (see the third row of Table 2). We expect that these features can capture the interpretation of local tempo variations measured by single notes, downbeats, and phrases. We take M and S on ND, 1MD, 2MD and 4MD to obtain song-level features. FPD itself is already a song-level feature so no aggregation is needed. Moreover, all of these

\[5\] For more details about MusicXML, please refer to http://www.musicxml.com/
five features are processed by $C_M$. Figure 2 shows examples where the same note (a crotchet C5) is interpreted in different ND for distinct expressions.

There are some more implementation details about the duration features. If a music piece has an incomplete measure in the beginning (e.g., Vivaldi’s La primavera Mov. I) then the incomplete measure is merged into the next one and features are computed starting from the first complete measure. If the length of a phrase is not the multiple of the 2 or 4 measures then the remainders are combined as a group. Bruch’s Violin Concerto No. 1 Mov. I (the 5th piece) is an unusual instance that has two ad libitum measures. In this case, 4MD is set at zero. In the parser process, a special part is to eliminate rests and ties because they do not have a unique sound. The former means an interval of silence and the latter has a curved line connecting to its previous note of the same pitch, indicating that they should be played as a single note.

3.2.3 Vibrato Features

Vibrato is an expressional manipulation of pitch corresponding to a frequency modulation of $F_0$ (fundamental frequency) [17]. Because the vibrato is characterized by the rate and extent of the frequency modulation of $F_0$, a precise estimation of the instantand pitch contour is needed. Since the frequency resolution in the STFT representation may not be high enough to represent the instantaneous frequency, we compute the instantaneous frequency deviation (IFD) [11] to estimate the instantaneous frequency:

$$\text{IFD}^w(n,k) = \frac{\partial \Phi^w}{\partial t} = \text{Im} \left( \frac{X^D^w(n,k)}{X^w(n,k)} \right),$$  

where $D^w(n) = w'(n)$. Given the pitch of each note from the score, instantaneous frequency is computed by summing the IFD and the bin frequency of the bin which is nearest to the pitch frequency. Figure 2 also sketches examples of the vibrato contours. We can see large differences in both duration and vibrato among them. For the

![Figure 2: Pitch contours of the first crotchet (C5) of Mozart’s Variationen with 6 expressions: None, Scherzando, Tranquillo, Con Brio, Maestoso and Risoluto.](image)

IFD calculation, a window of 1025 samples at 44.1 kHz sampling rate and a hop size of 64 samples are applied.

After obtaining the vibrato contour of each note, we adopt a moving-average filter with length of one-hundredth of the note length to reduce the spurious variation of the pitch contour. The filter length is empirically set so as not to avoid much distortion and to remove high-frequent noise. Based on the smoothed pitch contour, we consider the vibrato rate (VR) and the vibrato extent (VE). The former means the reciprocal of the time duration of two consecutive peaks, while the latter means the frequency deviation between a peak and its nearby valley. Following [8], we require that a vibrato chain contains more than 3 points and VR is between 3 and 12 Hz; otherwise, the vibrato chain is excluded. For each note, we compute the mean, standard variation, mean of difference (MΔ), standard variation of difference (SΔ), maximum (Max), minimum (Min) and difference (Diff) between the maximal and minimal values of both VR and VE over all frames within a note [24]. These note-level features are also aggregated to song-level features by means of M, S, and $C_M$.

In addition, we consider a note-level feature called global vibrato extent (GVE), meaning the difference of the maximal peak value and the minimal valley value within a vibrato note as shown in Figure 3. GVE is also aggregated to song-level features through M, S, and $C_M$. Finally, we consider a song-level feature called vibrato ratio (vibRatio), defined as:

$$\text{vibRatio} = \frac{\# \text{ vibrato notes}}{\# \text{ notes in a violin piece}} \times 100\%. \quad (4)$$

When no vibrato note is detected or the ND is shorter than 125ms [14], the vibrato features are set at zero.

3.3 Feature Selection and Classification

To evaluate the importance of the adopted features in our task, we perform feature selection on both the baseline and the proposed feature sets. Here, the ReliefF routine of the MATLAB statistics toolbox 6 is employed in the feature selection process [22]. In the training process, ReliefF sorts the features in descending order of relevance (importance). Then, the top-$n'$ most relevant features are taken for SVM modeling. The optimal feature number $n_{opt}$ which results in the best accuracy is obtained by brute-force searching.

6 http://www.mathworks.com/products/
Table 3: Performance of the baseline and the proposed feature sets. ‘All’ indicates the combination of dynamics, duration and vibrato; ‘fusion’ represents the combination of baseline and ‘all.’ \( n \) and \( n_{opt} \) are the original and the optimized number of features respectively; \( c \) and \( \gamma \) are SVM parameters; \( ACC \) indicates the average accuracy.

The RBF-kernel SVM is adopted for classification. Since the dataset is recorded by 11 violinists, we simply take 11-fold cross validation, by using the data of 10 violinists as the training set and the other as the testing set. Then the feature selection is performed in each fold of the cross validation individually. After sending the top-\( n_{opt} \) most relevant features into classification, the resulting performance is obtained from optimizing the parameters \( c \) and \( \gamma \) of the SVM. In this work, the SVM parameters are set according to the highest average classification accuracy across the 11 folds. In the future, we will consider other data splitting settings, for example using an independent held-out set for parameter tuning.

In our classification experiment, we exclude the case of None and consider a 10-class multi-class classification problem, because the calculation of \( C_M \) aggregation method requires that the non-expressive version is known \textit{a priori}. The classification accuracy of random guess is 0.152 on average.

As we want to find out the relevant interpretational factors, we only report below the results obtained by the top-\( n_{opt} \) relevant features selected by ReliefF.

4. EXPERIMENT RESULTS

4.1 Overview

Table 3 lists the original feature number \( n \), the optimal feature number \( n_{opt} \), the average accuracy (the ratio of true positives and the number of data) computed over the 11 folds, and the corresponding optimal \( c \) and \( \gamma \) for each experimental setting. The upper part of the table shows the result of the baseline feature set, where ReliefF selects \( n_{opt} = 107 \) out of 317 features and achieves an accuracy of 0.473. From the lower part of the table, when \( C_M \) aggregation method is considered, the proposed feature set achieves an accuracy of 0.531 when choosing \( n_{opt} = 36 \) out of 69 features, showing a significant improvement from the baseline feature set as validated by a one-tailed t-test (\( p<0.05 \), d.f. = 20). Finally, after fusing the baseline features and all the proposed features, the average accuracy comes to 0.589, using \( n_{opt} = 68 \) out of 386 features.

Table 4: The first 20 ranked features of the feature sets.

4.2 The contrast value \( C_M \)

Table 3 also shows how important using the contrast between the expressive and non-expressive version improves the performance. Comparing the left-hand side (without \( C_M \)) and the right-hand side (with \( C_M \)) of the table, using \( C_M \) constantly improves the average accuracy. Salient improvement can be observed for dynamic features \((p<0.05)\) and duration features \((p<0.05)\), implying that the change of dynamics, note duration, downbeat or phrase might be important interpretation factors when comparing the expressive and non-expressive performance. The improvement is not significant for vibrato \((p>0.5)\), possibly because the ratio of strong vibrato (expressive) and “almost no vibrato” (non-expressive) is not a stable feature. Table 3 also shows that using \( C_M \) on the proposed (‘all’) and the fusion feature sets leads to significant improvement for both cases \((p<0.005)\). Taking the contrast of feature values between expressive and non-expressive performance seems to be critical in modeling musical expression.

4.3 Feature importance analysis

Table 4 lists the top-20 relevant features for the baseline, proposed (‘all’) and fusion feature sets. The list is generated by summing the rank of each feature over the results of 11 folds, and by sorting the summarized rank again.

From the leftmost column, we can see that most of the relevant features in the baseline set are MFCCs. Despite its accuracy is inferior to the proposed features, this result shows the generality of MFCCs in audio classification.

From the middle column, we see that the top-20 proposed features include 11 duration features, 6 dynamic ones, and 3 vibrato ones. Over half of them are duration features. However, we note that the second feature is about vibrato (vibRatio) and the next two are both dynamic features \((D-Max-C_M \text{ and } D-M-C_M)\). It is not trivial to conclude that which factor is the most relevant. Dynamics, duration and vibrato all have contribution on music inter-
Affettuoso has lighter dynamics than other expressions and the last one retains relatively high F-scores because the first two have a similar meaning.

Table 5 clearly reveals their semantic similarity. The most serious confusion occurs between Con Brio and Cantabile. The proposed features, motivated from the basic understanding of dynamics, duration, vibrato, and the information of score, give better performance than the standard feature set in classifying expressive musical terms. Particularly, the contrast of feature values between expressive and non-expressive performance is found critical in modeling musical expression. The importance of the features is also reported. This provides insights into the design of new expression-based features, which may include features for the possible glissando between two adjacent notes, or the variation of the note/measure duration proportion with respect to its measure/excerpt. For future work, we will consider to expand the dataset, to experiment with other features and machine learning techniques, and to devise a mechanism that does not require a non-expressive reference to compute the contrast values.

5. CONCLUSION AND FUTURE WORK

In this study, we have presented a method for analyzing the interpretational factors of expressive musical terms implemented on a new dataset comprising of rich expressive interpretations of violin solos. The proposed features, motivated from the basic understanding of dynamics, duration, vibrato, and the information of score, give better performance than the standard feature set in classifying expressive musical terms. Particularly, the contrast of feature values between expressive and non-expressive performance is found critical in modeling musical expression. The importance of the features is also reported. This provides insights into the design of new expression-based features, which may include features for the possible glissando between two adjacent notes, or the variation of the note/measure duration proportion with respect to its measure/excerpt. For future work, we will consider to expand the dataset, to experiment with other features and machine learning techniques, and to devise a mechanism that does not require a non-expressive reference to compute the contrast values.

6. ACKNOWLEDGMENTS

The authors would like to thank the following three professional violinists for consulting: Chia-Ling Lin (Doctor of Musical Arts, City University of New York; concertmaster of Counterpoint Ensemble), Liang-Chun Chou (Master of Music, Manhattan School of Music; concertmaster of Tainan Symphony Orchestra), and Hsin-Yi Su (Master of Music, New England Conservatory of Music; major violin performance of Tainan Symphony Orchestra). We are also grateful to the 11 professional violinists for their contribution to the development of the new dataset. The paper is partially funded by the Ministry of Science and Technology of Taiwan, under contracts MOST 103-2221-E-006-140-MY3 and 102-2221-E-001-004-MY3, and the Academia Sinica Career Development Award.
7. REFERENCES


