IMPROVING INSTRUMENT RECOGNITION IN POLYPHONIC MUSIC THROUGH SYSTEM INTEGRATION

Dimitrios Giannoulis[†*], *Emmanouil Benetos*[‡*], *Anssi Klapuri*[§], *and Mark D. Plumbley*[†]

† Centre for Digital Music, EECS, Queen Mary University of London, London, UK
‡ Department of Computer Science, City University London, London, UK.
§ Ovelin, Helsinki, Finland & Tampere University of Technology, Finland.

ABSTRACT

A method is proposed for instrument recognition in polyphonic music which combines two independent detector systems. A polyphonic musical instrument recognition system using a missing feature approach and an automatic music transcription system based on shift invariant probabilistic latent component analysis that includes instrument assignment. We propose a method to integrate the two systems by fusing the instrument contributions estimated by the first system onto the transcription system in the form of Dirichlet priors. Both systems, as well as the integrated system are evaluated using a dataset of continuous polyphonic music recordings. Detailed results that highlight a clear improvement in the performance of the integrated system are reported for different training conditions.

Index Terms— Musical instrument recognition, automatic music transcription, music signal analysis

1. INTRODUCTION

Automatic Music Transcription (AMT) systems attempt to convert an acoustic music recording into some form of musical notation. It has many applications related to music information retrieval, musicological analysis, and interactive computer music systems [1]. AMT typically entails the detection of note events within the music piece. Since music is mostly polyphonic, assigning detected notes to instruments is also amongst the central tasks of such a system. The problem of polyphonic musical instrument identification has also been studied on its own [2, Section IV], it is however clearly associated with AMT and often considered as a subtask of the latter.

Instrument identification for polyphonic music is a closely-related task to blind source separation (sources being the musical instruments), where the goal is given a number of mixture signals (in most cases just one) to separate the source signals from the mixture. Although in instrument identification the separation is not a requirement the task could benefit from a pre-processing source separation step that simplifies the problem to that of monophonic recognition which is considerably easier. The opposite approach, to attempt and identify the instruments straight from the mixture and potentially use this information to improve a latter source separation process could be another option.

In the literature, there are instrument identification approaches that first attempt to separate the signals of the various musical instruments at a pre-processing step and then perform instrument identification to the separate signals like for example in [3], or approaches that try to identify the musical instruments directly from the mixture and avoid the complex source separation process as in [4].

Despite the popularity of the instrument recognition task and the significant progress that has been made in AMT research in general, systems are still not able to support end-user applications that can transcribe accurately, reliably and with no constraints any recorded music. Current challenges and problems associated with this static performance of transcription systems have been analysed in [5], where the authors have also highlighted future directions for AMT research. Among the various future directions, one of high interest, mainly because it requires minimal effort and added complexity, is that of information integration. The main idea is to fuse information across different aspects of music or combine methods targeting the same feature. The first for example, would have a set of independent systems that estimate various music content descriptors such as: tempo estimation, key detection, instrument recognition and so on, inform the main AMT system, but also each other where possible, in an attempt to raise the overall system performance. In the second case, the system's performance is attempted to be increased by combining multiple estimators or detectors for a single music aspect, like for example two multi-pitch detectors or two instrument detection systems. That way, and especially if the two systems follow different methodologies, certain difficulties may be overcome. For example in [6], the authors combined successfully a series of pitched instrument onset detectors, which individually have high precision and low recall and managed to obtain an improved detection accuracy for the overall system.

In this paper, we propose a fusion of an independent instrument recognition system [7] with an AMT system [8] in an attempt to improve overall instrument recognition performance. Instrument information extracted from the instrument recognition system is fused into the transcription system using Dirichlet priors [9]. To the authors' knowledge, this is the first attempt in the literature to fuse systems for instrument recognition. Furthermore, this work focuses on performing instrument recognition on complete music recordings and not isolated notes or chords, as was done in [7]. Instrument assignment experiments are performed using the Bach10 polyphonic music dataset [10]. Results show that the fusion of the two systems leads to a significant improvement in terms of instrument assignment performance.

2. AUTOMATIC MUSIC TRANSCRIPTION SYSTEM

In this work, we utilise the transcription system proposed in [8], which is based on shift-invariant probabilistic latent component

* Equally contributing authors. D.G. is funded by a Queen Mary University of London CDTA Research Studentship. E.B. is supported by a City University London Research Fellowship. M.D.P. is supported by EPSRC Leadership Fellowship EP/G007144/1. This work is partly funded by EPSRC Grant EP/H043101/1.

Published in: Proc IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), Florence, Italy, 5-9 May 2014. pp. 5259-5263.

analysis (SI-PLCA) [11]. In SI-PLCA, the input spectrogram $V_{\omega,t}$, which must be scaled to have integer entries, is modeled as the histogram of the draw of N independent random variables (ω_n, t_n), which are distributed according to $P(\omega, t)$ (ω denotes frequency, and t time). The model is shift-invariant due to the fact that inter-harmonic spacings are the same for all pitches in the log-frequency domain, which is utilised in the present model for supporting tuning deviations and frequency modulations.

The model decomposes $P(\omega, t)$ as:

$$P(\omega, t) = P(t) \sum_{f,h,s} P(\omega|s, f, h)P(h|f, t)P(s|f, t)P(f|t) \quad (1)$$

where f denotes pitch in semitone resolution, s instrument source, and h the log-frequency shifting factor. $P(\omega|s, f, h)$ is the pre-extracted and pre-shifted spectral template for pitch f and instrument s, which is shifted across log-frequency according to h. $P(h|f, t)$ is the time-varying shifting parameter, $P(t)$ is the log-spectrogram energy (known quantity), $P(f|t)$ are the pitch activations (used for multi-pitch detection), and finally $P(s|f, t)$ are the time-varying instrument contributions. h is constrained to a semitone range. In the present system, we use as a time-frequency representation the constant-Q transform (CQT), with a log-frequency resolution of 60 bins per octave and a 40ms step [12]. Thus, $h \in [1, \ldots, 5]$.

The unknown model parameters can be iteratively estimated using the Expectation-Maximisation (EM) algorithm [13]. For the expectation step, the following posterior is computed:

$$P(f, h, s|\omega, t) = \frac{P(\omega|s, f, h)P(h|f, t)P(s|f, t)P(f|t)}{\sum_{f,h,s} P(\omega|s, f, h)P(h|f, t)P(s|f, t)P(f|t)}. \quad (2)$$

For the maximisation step, unknown parameters $P(h|f, t)$, $P(s|f, t)$, and $P(f|t)$ are updated using the posterior computed from the expectation step. For brevity we only include the update equation for the instrument contribution $P(s|f, t)$, which is relevant for this work (all maximisation equations can be found in [8]):

$$P(s|f, t) = \frac{\sum_{\omega,h} P(f, h, s|\omega, t)V_{\omega,t}}{\sum_{s,\omega,h} P(f, h, s|\omega, t)V_{\omega,t}}. \quad (3)$$

The update equations for the expectation and maximisation steps are iterated until convergence, with 15-20 updates being sufficient. Sparsity constraints are also applied to the update equations for $P(f|t)$ and $P(s|f, t)$ in order to control the level of polyphony as well as the number of active instruments for producing a note.

The matrix used for multi-pitch detection evaluation is given by $P(t, f) = P(t)P(f|t)$ and the matrix used for instrument assignment evaluation is $P(s, t, f) = P(t)P(f|t)P(s|f, t)$. Since the resulting activations are non-binary, the pitch and instrument activation matrices have to be converted into binary representations (this procedure is also called *note tracking*). Both matrices are thresholded followed by minimum duration pruning set to $\tau = 80ms$, in order to remove detected notes with small durations.

3. MUSICAL INSTRUMENT RECOGNITION SYSTEM

In this section we introduce the musical instrument recognition system we plan to fuse with the system introduced in Section 2. The system utilized is the one proposed in [7]. Polyphonic musical instrument recognition is performed using a *missing feature* approach that deals with occlusions and partial overlaps in the Time-Frequency domain. *Missing feature* (or *missing data*) techniques attempt to perform recognition based on incomplete spectrograms [14]. In this work, among other things, we evaluate the system performance using continuous music recordings rather than artificially created mixtures from isolated notes.

Missing feature techniques try to separate the corrupted or occluded regions of the spectrogram from the ones for which clean information about each source can be extracted and this is achieved by estimating binary masks that separate out the clean source spectrograms from the mixture. The missing regions removed by the mask can either be marginalized out of the classification (excluded) or the observations from these regions can be used as an upper bound for the missing data and bounded marginalization can be applied [15]. Since extracting the missing data mask is, probably, the most difficult part of such approaches, an assumption often made is that prior knowledge, informing us with certainty about which spectro-temporal regions are missing, is available [14].

Missing feature was first introduced to musical instrument recognition by Eggink and Brown in [16]. The method assumed the binary missing data masks were known *a priori* but also included a simple pitch-based mask estimation alternative with significantly worse performance. The missing data were entirely disregarded from the classification step of the algorithm. Two of the paper authors in [7] proposed a missing feature approach for polyphonic musical instrument recognition in which the missing data were treated with bounded marginalization. Features proposed are spectral subband energy level differences calculated from harmonic partial amplitudes. We also proposed ways to estimate binary masks from the data or use a mask summation process to marginalize the binary missing data mask.

In this work the system from [7] is employed in order to compute the instrument assignment probabilities $P(s|f, t)$ and subsequently "feed" these into the AMT system we introduced in Section 2. In order to measure the maximum benefit we can achieve out of the two system integration we decided to use "oracle" masks for the missing data estimation in [7] and disregard, at present, the hard mask estimation process.

The instrument recognition system also estimates internally and independently of the AMT system the pitch probabilities $P(f|t)$ and thus the system could on its own produce a transcription output after some post-processing. However, the system employed in this work is used only to produce a set of conditional probabilities $P(s|f, t)$ for each candidate instrument s, in other words, the instrument contributions as in (3). That is, the probability that the true source that produced the sound corresponding to pitch f at time frame t is instrument s.

Instrument recognition is performed within individual time frames. So let us define the system input with $\mathbf{o}$ to denote the observed time-domain signal of the music mixture. The system models a single frame of the mixture signal o_t at time t as a mixture of harmonic sounds and a residual and calculates the probabilities $P(s|o_t, f), \quad \forall f \in \mathcal{F}$, where $\mathcal{F}$ denotes the set of candidate active pitches detected in frame t. Given an analysis time frame t we can rewrite $P(s|\mathbf{o} = o_t, f)$ as $P(s|t, f)$ since these two are equivalent and thus we obtain the instrument contributions.

The system performs instrument recognition independently in each analysis frame t using only local spectral features extracted from the mixture. It does not include any temporal features and information from the estimated frame-wise class-conditional probabilities is not integrated across time. As a result, the system is perhaps not achieving its full potential as the AMT system of Section 2, however it is computationally very light and easy to integrate as we will show in the following section. Finally, we are interested to see whether by integrating information from this system to a complete

AMT system the overall performance can still be boosted.

4. SYSTEM INTEGRATION

The SI-PLCA framework on which the AMT system of Section 2 is based upon allows the introduction of additional probabilistic factors in the decomposition of the representation matrix with relative ease. In the context of this work, we are interested in incorporating the instrument contribution estimates that are extracted from the system in Section 3 into the model of the AMT system to act as prior information of the instrument identities of the signal in each analysis frame.

A mechanism for imposing priors for estimated parameters in a PLCA model is introduced in [9]. The class conditional densities of PLCA models like $P(s|f,t)$ follow multinomial distributions, as explained in [9]. Therefore priors can be easily introduced in the model as *Dirichlet* distributions which constitute a conjugate prior distribution to a mutlinomial. Dirichlet distributions, denoted as $Dir(\alpha)$, are parameterized by a set of positive and real *hyperparameters* α. In order to satisfy the unit measure assumption for the priors and without loss of generality we impose that $\sum_i \alpha_i = 1$.

We subsequently define the instrument contribution priors Λ_s over all possible pitches f and time frame indices t as:

$$P(\Lambda_s) \propto \prod_t \prod_f \prod_s P(s|t,f)^{\lambda_s \alpha(s|t,f)} \quad (4)$$

where λ_s is a weight parameter utilised in order to allow us to scale the hyperparameters α arbitrarily based on how much we wish to impose the priors in the model for each instrument s. Based on this, we can rewrite the update equation for the instrument contribution parameters in (3) as:

$$P(s|f,t) = \frac{\sum_{\omega,h} P(f,h,s|\omega,t) V_{\omega,t} + \lambda_s \alpha(s|t,f)}{\sum_{s,\omega,h} P(f,h,s|\omega,t) V_{\omega,t} + \lambda_s \alpha(s|t,f)} \quad (5)$$

where $\alpha(s|t,f)$ are the instrument contribution estimates of the system in Section 3. For the proposed system, the value of λ_s was set to 0.2 after experimentation.

5. EVALUATION

5.1. Dataset

For testing the proposed system, we employ the Bach10 dataset [10], which is currently the largest freely available dataset of recorded music for both instrument assignment and multi-pitch detection evaluation. It consists of 10 polyphonic music recordings of four-part J.S. Bach chorales, performed by violin, clarinet, saxophone, and bassoon. Recordings are included as final mixes containing all instruments (which are used for testing), as well as individual tracks for each instrument (which are used for comparative experiments). The dataset also contains pitch ground truth for each instrument.

For training the SI-PLCA-based transcription system of Section 2, we use isolated note samples for violin, clarinet, saxophone, and bassoon from the RWC database [17], using the complete note range of each instrument. In order to extract log-frequency spectral templates for each note of each instrument, we perform unsupervised SI-PLCA on each note sample. For comparative purposes, we also extract note templates from the RWC database for the following instruments: cello, flute, oboe, and piano. Finally, also for comparative purposes, we extract note templates directly from the individual

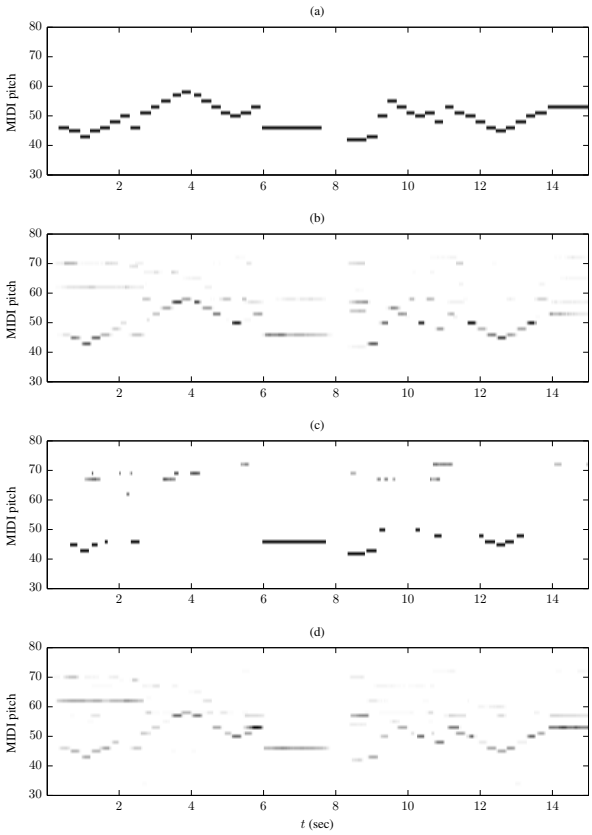

Fig. 1. Piano-roll representations for the bassoon track of recording "Herr Gott" from the Bach10 dataset. (a) Ground-truth. (b) Output of the AMT system. (c) Output of the instrument recognition system. (d) Output of the integrated system.

Bach10 tracks. In order to achieve this, we use the non-negative matrix factorisation (NMF) algorithm with β-divergence [18].

For training the missing feature-based instrument recognition system we utilized a similar procedure using isolated note samples from the RWC database [17]. We trained statistical models representing the various instruments s as described in [7, Section II-D]. On a different experiment we also trained on isolated notes from the Bach10 dataset but perhaps because the lack of data diversity in the tracks did not enable the system to learn meaningful pitch and instrument specific statistical models we performed the training on a mixture of RWC database and Bach10 training samples instead.

5.2. Evaluation Metrics

For assessing the performance of the proposed system, we employ instrument assignment and multi-pitch detection metrics. In all cases, we use the precision, recall, and F-measure metrics, which are commonly used in transcription evaluations [3, 8]:

$$Pre = \frac{N_{tp}}{N_{sys}}, \quad Rec = \frac{N_{tp}}{N_{ref}}, \quad F = \frac{2 \cdot Rec \cdot Pre}{Rec + Pre} \quad (6)$$

where N_{tp} is the number of correctly detected pitches, N_{sys} is the number of pitches detected by the system, and N_{ref} is the number of ground-truth pitches.

As in the MIREX evaluations [19], a detected note is considered correct if its pitch is the same as the ground truth pitch and its onset

System	F_{mp}
AMT system with RWC templates	61.96%
AMT system with Bach10 templates	67.38%

Table 1. Multi-pitch detection results using the system of Section 2.

System	F_v	F_c	F_s	F_b	F_{ins}
[7]	20.10%	12.62%	19.65%	32.37%	**21.18%**
[8]	21.52%	36.01%	21.45%	35.49%	**28.62%**
Integrated system	22.32%	34.03%	29.33%	37.36%	**30.76%**

Table 2. Instrument assignment results for the transcription system, the instrument recognition system, and the proposed integrated system (using training data for 4 instruments from the RWC database).

is within a 50ms tolerance interval of the ground-truth onset. For multi-pitch evaluation, we use the pitch ground-truth each recording and the resulting F-measure is denoted as F_{mp}. For the instrument assignment evaluations we use the pitch ground-truth of each instrument separately, and denote the following metrics (in terms of F-measure): F_v, F_c, F_s, F_b, denoting the F-measure metrics for violin, clarinet, saxophone, and bassoon, respectively. We also define an average instrument assignment metric:

$$F_{ins} = \frac{1}{4}\big(F_v + F_c + F_s + F_b\big) \tag{7}$$

5.3. Results

Instrument assignment and multi-pitch detection experiments are performed using training data from the RWC database for the 4 instruments present in the recordings. Comparative experiments are also performed using training data from the Bach10 dataset and also using training data from the RWC database for a more broad 8-instrument set (also including cello, flute, oboe, and piano). Fig. 1 shows the raw piano-rolls extracted from the two detectors as well as the integrated system, for a bassoon track of the Bach10 dataset.

Multi-pitch detection results for the transcription system of Section 2 can be seen in Table 1. The AMT system reaches a note-based F-measure of 61.96%. It can be seen that the achieved F-measure increases by about 6%-units when training samples from the Bach10 set are used, giving an indication of the upper limit of the algorithm.

Instrument assignment results using RWC data trained for 4 instruments are shown in Table 2. It can be seen that the average instrument assignment performance for the AMT system in terms of F-measure is 28.62%, with the best results reported for clarinet (which has a distinct spectral shape). The instrument recognition system reaches $F_{ins} = 21.18\%$, recognising best the bassoon, having tones in a different pitch range compared to the other instruments. The performance of the integrated system is improved over 2%-units in terms of F_{ins}, showing that fusing detectors can lead to a performance improvement in instrument assignment. The improvement is particularly prevalent for the saxophone, where both detectors exhibit similar performance. In cases where there is a significant gap in performance between the two detectors, the resulting performance improvement might be smaller, or in certain cases there might be a decrease (as shown for the clarinet). The proposed method is also robust in terms of λ_s: by varying its values from 0.1 to 0.5, the F_{ins} improvement is always above 1.6%-units.

In Table 3, instrument assignment results using training data from the Bach10 dataset (for the instrument recognition system also from the RWC database) are shown. For the AMT system, the increase over the RWC-trained system is over 20%, while for the instrument recognition system the increase is over 8%. The integrated system improves upon the AMT system by about 2%. Here, the best performance for the AMT system is reported for the saxophone; the fact that many different saxophone variants exist might indicate that the instrument model used for training from the RWC might not have been the same as in the test recordings. For the instrument recognition system, the best performance is still reported for the bassoon.

System	F_v	F_c	F_s	F_b	F_{ins}
[7]	31.86%	28.84%	20.03%	38.80%	**29.88%**
[8]	39.08%	46.01%	64.98%	50.93%	**50.25%**
Integrated system	43.26%	46.93%	66.09%	53.63%	**52.48%**

Table 3. Instrument assignment results for the transcription system, the instrument recognition system, and the proposed integrated system (using training data for 4 instruments from the Bach10 and RWC databases).

System	F_v	F_c	F_s	F_b	F_{ins}
[7]	7.67%	11.00%	14.43%	24.12%	**14.30%**
[8]	17.87%	30.67%	22.52%	27.71%	**24.69%**
Integrated system	17.39%	30.29%	25.17%	29.22%	**25.52%**

Table 4. Instrument assignment results for the transcription system, the instrument recognition system, and the proposed integrated system (using training data for 8 instruments from the RWC databases).

Finally, results for systems trained on RWC data for 8 instruments are displayed in Table 4. In all cases, the performance drops compared to systems trained only using the 4 instruments present in the recordings. However, an improvement of +0.9% is still reported for the integrated system over the AMT system.

6. CONCLUSIONS

In this work, we proposed a system for instrument recognition in polyphonic music which combines two detectors, namely an automatic music transcription system which supports instrument assignment and an instrument recognition system based on missing feature theory. The instrument recognition system was fused with the AMT system using Dirichlet priors.

Experiments performed on the Bach10 dataset consisting of 4-instrument recordings showed that the integrated system has a clear instrument assignment performance improvement. The improvement was more significant in cases where the performance of the two individual systems before integration was comparable. However, it is worth mentioning that even in the most challenging of the evaluation scenarios, when 8 instrument classes were utilised and the performance of the AMT system was clearly superior to that of the instrument recognition system (potentially because the latter performs the recognition only within single analysis frames), there was still reported improvement in the performance of the integrated system, that can be shown to be statistically significant [20, Ch. 3].

The reported results also demonstrate the level of difficulty in creating a system for identifying instruments in polyphonic music, especially in cases with many harmonic overlaps or when the active instruments belong in the same instrument taxonomy (as is the case with the Bach10 dataset). A significant improvement can be achieved if system parameters can be suited to the instrument sources present in the test signals, as demonstrated by results using the Bach10 dataset for training. To that end, in the future we will work on source-adaptive systems for both instrument assignment and multi-pitch detection.

7. REFERENCES

[1] A. Klapuri and M. Davy, *Signal Processing Methods for Music Transcription*, Springer, 2006.

[2] M. Muller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1088–1110, 2011.

[3] G. Grindlay and D. Ellis, "Transcribing multi-instrument polyphonic music with hierarchical eigeninstruments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1159–1169, Oct. 2011.

[4] J. G. A. Barbedo and G. Tzanetakis, "Musical instrument classification using individual partials," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 111–122, 2011.

[5] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: challenges and future directions," *Journal of Intelligent Information Systems*, 2013, accepted.

[6] A. Holzapfel and Y. Stylianou, "Three dimensions of pitched instrument onset detection," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1517–1527, Aug. 2010.

[7] D. Giannoulis and A. Klapuri, "Musical instrument recognition in polyphonic audio using missing feature approach," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1805–1817, 2013.

[8] E. Benetos, S. Cherla, and T. Weyde, "An effcient shift-invariant model for polyphonic music transcription," in *6th International Workshop on Machine Learning and Music*, Sept. 2013.

[9] P. Smaragdis and G. Mysore, "Separation by "humming": user-guided sound extraction from monophonic mixtures," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2009, pp. 69–72.

[10] Z. Duan, B. Pardo, and C. Zhang, "Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2121–2133, Nov. 2010.

[11] P. Smaragdis, "Relative-pitch tracking of multiple arbitary sounds," *Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3406–3413, May 2009.

[12] C. Schörkhuber and A. Klapuri, "Constant-Q transform toolbox for music processing," in *7th Sound and Music Computing Conf.*, Barcelona, Spain, July 2010.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[14] J. Barker, "Missing data techniques: Recognition with incomplete spectrograms," in *Techniques for Noise Robustness in Automatic Speech Recognition*, T. Virtanen, R. Singh, and Bhiksha Raj, Eds. Wiley, 2012.

[15] J. Barker, M. P. Cooke, and D. P. W. Ellis, "Decoding speech in the presence of other sources," *Speech Communication*, vol. 45, no. 1, pp. 5–25, 2005.

[16] J. Eggink and G. J. Brown, "A missing feature approach to instrument identification in polyphonic music," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003, vol. 5, pp. 553–556.

[17] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: music genre database and musical instrument sound database," in *International Conference on Music Information Retrieval*, Baltimore, USA, Oct. 2003.

[18] R. Kompass, "A generalized divergence measure for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 3, pp. 780–791, 2007.

[19] "Music Information Retrieval Evaluation eXchange (MIREX)," `http://music-ir.org/mirexwiki/`.

[20] E. Benetos, *Automatic transcription of polyphonic music exploiting temporal evolution*, Ph.D. thesis, Queen Mary University of London, Dec. 2012.