Modelling of Sound Events with Hidden Imbalances Based on Clustering and Separate Sub-Dictionary Learning

Chaitanya Narisetty, Tatsuya Komatsu and *Reishi Kondo*
Data Science Research Laboratories
NEC Corporation, Japan
Email: c-narisetty@cp.jp.nec.com, t-komatsu@ew.jp.nec.com, kondoh@ct.jp.nec.com

***Abstract*—This paper proposes an effective modelling of sound event spectra with a hidden data-size-imbalance, for improved Acoustic Event Detection (AED). The proposed method models each event as an aggregated representation of a few latent factors, while conventional approaches try to find acoustic elements directly from the event spectra. In the method, all the latent factors across all events are assigned comparable importance and complexity to overcome the hidden imbalance of data-sizes in event spectra. To extract latent factors in each event, the proposed method employs clustering and performs non-negative matrix factorization to each latent factor, and learns its acoustic elements as a sub-dictionary. Separate sub-dictionary learning effectively models the acoustic elements with limited data-sizes and avoids over-fitting due to hidden imbalances in training data. For the task of polyphonic sound event detection from DCASE 2013 challenge, an AED based on the proposed modelling achieves a detection F-measure of 46.5%, a significant improvement of more than 19% as compared to the existing state-of-the-art methods.**

***Index Terms*—Data-Size-Imbalance, Acoustic Event Detection, Non-Negative Matrix Factorization, Dictionary Learning**

I. Introduction

The ubiquity of recorded audio in the present day paves way for development of smart applications that can be readily integrated with voice activated devices like Amazon Echo or Google Home. Being invariant to occlusions and brightness, audio is ideally suited for security applications likes Acoustic Event Detection (AED) and Scene Identification [1]. This paper focuses on AED which aims to detect the occurrences of sound events in a given audio signal. The event detection performance is significantly dependent on the effectiveness of modelling audio signals gathered for training. Hence the performance of AED degrades with noise, polyphony and hidden data-size-imbalances in audio signals. Polyphony refers to the simultaneous existence of two or more sound events in an audio signal. These degradations demand the need for effective event models with multi-label event classifiers that are robust to the interference from background noises.

The most frequently used approaches for AED are based on Mel-Frequency Cepstral Coefficients (MFCCs) trained using Gaussian Mixture Models [2] and Non-Negative Matrix Factorization (NMF) [3]. The recent advancements in AED demonstrate the effectiveness of Convolutional Recurrent Neural Networks (CRNNs) for modelling large audio databases [4][5]. The capability of CRNNs to model complex and non-linear dependencies in the audio signals is offset by its challenges of being computationally intensive and often prone to severe over-fitting [6]. These challenges motivate the use of matrix factorization methods to extract linear dependencies in audio signals. The additional constraint of having non-negative factors is to model a given audio as an additive representation of different sound sources. The entire set of linear dependencies (dictionary) among all sound events can be extracted at once by performing NMF on the spectrogram of an entire audio database, consisting of several sound events [7]. Hence the performance of NMF based AED methods heavily rely on the ability of their dictionaries to represent sound events. In situations where the data-size of a few events is much higher than the others, dictionary learned by performing NMF of the entire database overlooks the events with limited data-sizes. One of the methods to model such an imbalance is to normalize the spectra of each event with their respective data-sizes [8]. It is also possible to estimate the overall dictionary as a composition of event-wise dictionaries to overcome the imbalance among sound events [9].

A similar but hidden imbalance among the exemplars of an individual sound event was observed in NEC's field trials [10]. For the purpose of illustration, consider a *Piano* sound event consisting of 10 exemplars for 3 notes $C4$-$G4$-$E4$ played simultaneously for 1 second and only 1 exemplar for a $C4$ note played for 0.5 seconds. *Piano* event dictionary learned using NMF fails to model the under-represented $C4$ note. If prior information detailing the imbalance of exemplars is hidden, the dictionary learned for such an event fails to represent the exemplars with limited data-sizes.

We propose an effective dictionary learning of sound events to overcome such data-size-imbalances by assuming an event to be explained from a few latent factors. This assumption motivates an effective strategy for dictionary generation where each event is clustered into latent factors namely sub-events. Each sub-event is an aggregation of multiple acoustic elements (a structural element which can be represented by a single basis). Clustering brings forth the underlying data-size-imbalance hidden in each event. Sub-dictionaries learned from each sub-event improve the modelling of under-represented acoustic elements. The overall set of sub-dictionaries over all sound events effectively models the entire audio database, and when paired with support vector machines, improves on the drawback of existing NMF based AED methods.

ISBN 978-90-827970-1-5

The remainder of this paper is organized as follows. In Section II we briefly discuss a conventional NMF based dictionary learning. Section III describes an AED method based on the proposed dictionary learning and SVM classifier training. Performance evaluations of the proposed method for the above *Piano* illustration and the Polyphonic Sound Event Detection task from Detection and Classification of Acoustic Scenes and Events (DCASE) 2013 are detailed in Section IV. This discussion ends with a few concluding remarks in Section V.

II. Non-Negative Matrix Factorization

Before detailing the proposed dictionary learning method, we briefly explain the concept of NMF as it is predominantly used for modelling of sound events. NMF is a set of matrix decomposition techniques which approximately factorize a given positive matrix V of size $p \times q$ into positive low-rank matrices W and H of sizes $p \times r$ and $r \times q$ respectively, where $r < \min(p, q)$. It is being assumed that V consists of q features each of which is a p-dimensional vector. In the context of an audio signal obtained from an additive mixture of sound sources, the audio spectrogram V can be factorized into a set of few basis vectors W and activation vectors H. NMF is formulated as,

$$V \approx WH \text{ s.t. } V \succeq 0, W \succeq 0, H \succeq 0. \tag{1}$$

This approximate factorization is guided by a cost function minimization. For our present work, Kullback-Leibler (KL) Divergence is taken to be the cost function. Iterative update rules used to estimate W and H for NMF optimized using KL-Divergence are put forth by Lee and Seung [11].

Problem with NMF dictionaries: W represents a dictionary which consists of fundamental vectors necessary to reconstruct V. Columns of H represent a new feature space and indicate the parts of dictionary W that are being used to represent a given spectra [12]. A natural application of such dictionary dependent features H is to identify activations extracted from a given test spectra, that are similar to activations learned from training spectra. Without a proper dictionary, the activations necessary to represent a spectrum cannot be obtained. From the spectrogram of entire *Piano* event described earlier, a dictionary learned using NMF aims to represent most parts of the spectra and ends up over-fitting the $C4$-$G4$-$E4$ sound. Thus there are no basis vectors capable of modelling the $C4$ sound. To overcome this inability of conventional NMF to represent hidden acoustic elements with limited data-sizes, we propose an improved dictionary learning method.

III. AED with Proposed Dictionary Learning

A. Dictionary Learning

Block diagram for the proposed method consists of two important parts: clustering event spectra and separate sub-dictionary learning as shown in Fig. 1. Consider an audio database of M sound events with each event's respective spectrogram denoted by E_m, $1 \leq m \leq M$. MFCC coefficients for each spectrum in E_m are extracted and trained using a

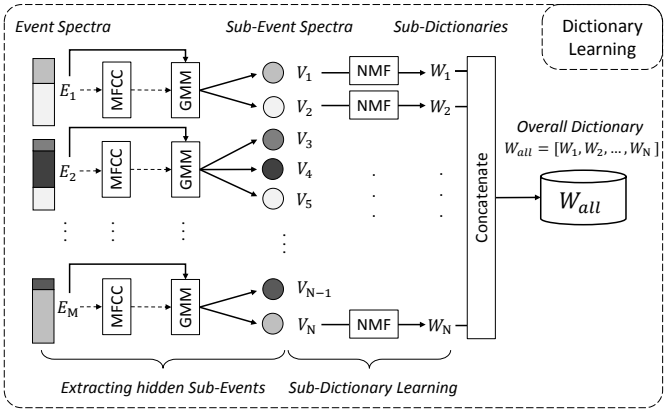

Fig. 1: Block diagram of the proposed clustering and separate sub-dictionary learning for modelling the hidden imbalance of data-sizes in each event spectra.

Gaussian mixture model (GMM). Each of the event spectra E_m is clustered into a set of sub-event spectra using the trained GMM. Let the set $\{V_n\}, 1 \leq n \leq N$ represent the entire set of sub-event spectra, where N is the total number of sub-events clustered across all events. Separately, a sub-dictionary W_n is learned using NMF from their respective sub-event spectra V_n. All sub-dictionaries are concatenated to output an overall dictionary W_{all} for the entire audio database i.e.

$$W_{all} = [W_1, W_2, \ldots W_N]. \tag{2}$$

The motivation behind clustering event spectra is to extract the hidden data-size-imbalances among its acoustic elements. In the *Piano* event illustrated earlier, the unbalanced acoustic elements are the well-represented spectra of $C4$-$G4$-$E4$ and under-represented spectra of $C4$. As it is not always possible to construct event databases which take such imbalance into consideration, mixture models such as GMM are capable of modelling both the under-represented and well-represented parts of an event [13]. The extracted MFCCs reduce the dimensionality of original spectrum and improve the stability of mixture modelling. Separately learned sub-dictionaries make W_{all} capable of representing the acoustic elements with limited data-sizes, which are overlooked by the conventional event-wise NMF dictionaries. Hence an AED based on the proposed dictionary learning effectively models sound events, thereby improving the overall detection performance.

B. Classifier Training and Event Detection

The above dictionary learning is an integral part of the overall acoustic event modelling and detection as shown in Fig. 2. Let Λ_D be the development spectra used for training classifiers. In this work, Λ_D is the spectrogram of polyphonic audio signals obtained from a mixture of different sound events. Supervised NMF is performed over the entire development spectra Λ_D by using a fixed basis matrix W_{all} to estimate the activations matrix H_D. A conventional approach to train event classifiers involves the use of this activations matrix H_D. The columns of H_D, along with their respective event labels are used for classifier training to generate event classifiers.

ISBN 978-90-827970-1-5 © EURASIP 2018

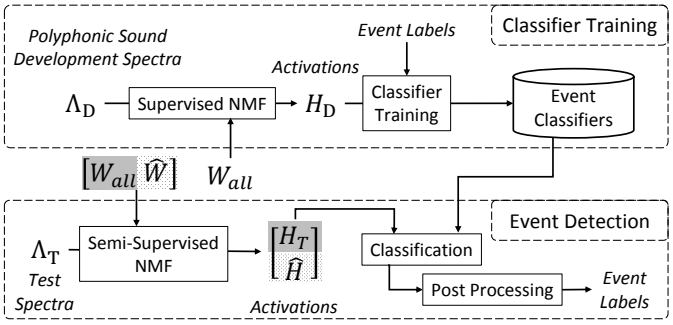

Fig. 2: Block diagram for training event classifiers from the labelled spectra of polyphonic sound events and using these classifiers to detect sound event labels from test spectra.

For detecting sound events in any given test spectra Λ_T, a semi-supervised NMF with an extended basis matrix $[W_{all}, \widehat{W}]$ is employed [14]. As the test spectra often contains an added noise, a small noise dictionary $\widehat{W}$ is appended to the columns of W_{all}. During this semi-supervised NMF, W_{all} remains fixed and only $\widehat{W}$ is estimated. The extended portion of the basis matrix $\widehat{W}$ is estimated using NMF to model the additional noise elements. This increment of the overall basis matrix increases the dimensionality of the estimated activations matrix. This is formulated as,

$$\Lambda_T \approx \begin{bmatrix} W_{all} & \widehat{W} \end{bmatrix} \begin{bmatrix} H_T \\ \widehat{H} \end{bmatrix} \text{ s.t. } \Lambda_T, \widehat{W}, H_T, \widehat{H} \succeq 0. \quad (3)$$

The part of activations matrix corresponding to W_{all} is H_T and that of $\widehat{W}$ is $\widehat{H}$. Only the event activation matrix H_T is used for event detection, while $\widehat{H}$ is ignored. The trained event classifiers classify each column of H_T to output a binary eventroll that indicates which event(s) take place in each column of H_T. The extracted binary eventroll is later post-processed to output the final event labels.

Many types of classifiers are used in literature to train these activations. Gemmeke et al. [9] uses a Hidden Markov Model (HMM) with linear time warping. Thresholding based classifiers to ascertain the existence of a particular event have been used in [15]. However, Support Vector Machines (SVMs) with linear kernels have shown promising results for training event classifiers [16], [17], [18] and will be used in this paper. The activations corresponding to a particular event are used to train a binary linear-SVM classifier against all the activations of the remaining events.

IV. Simulations and Results

A. Dictionary Learning for the Piano Event

As detailed in the previous sections, consider the *Piano* event with a data-size-imbalance among its $C4$-$G4$-$E4$ and $C4$ exemplars as shown in Fig. 3(a). Ideally, the dictionary for this event should contain two basis vectors which are either $\{C4\text{-}G4\text{-}E4, C4\}$ or $\{G4\text{-}E4, C4\}$. The sampling rate of the exemplars is 8kHz. The spectrogram for this event is estimated from 40ms time frames of the signal with a 10ms frame shift.

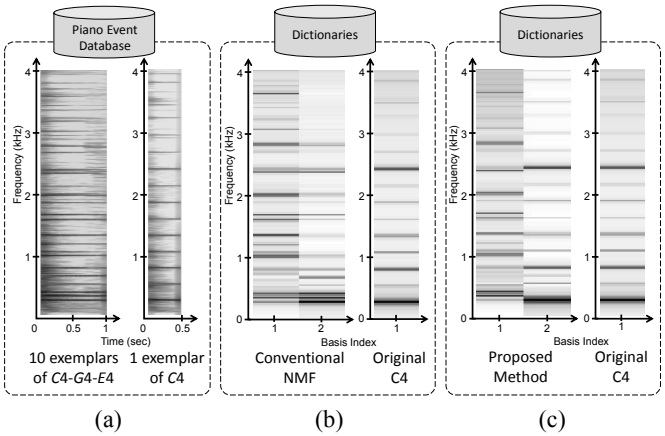

Fig. 3: (a) *Piano* sound event with $20:1$ data-size-imbalance of spectra and its dictionaries extracted using (b) Conventional NMF which can collectively represent only the $C4$-$G4$-$E4$ and (c) Proposed Method which can represent both $G4$-$E4$ and $C4$.

When a dictionary is learned from this spectrogram using conventional NMF, the resulting two basis vectors are low and high frequency representations of the $C4$-$G4$-$E4$ sound and are unable to represent the $C4$ sound as shown in Fig. 3(b). A single basis vector obtained from the $C4$ exemplar using NMF is shown on the side for visual comparison. Alternatively using the proposed method, we first extract 25 MFCCs excluding the 0^{th} coefficient from each spectrum. A GMM with two components models the MFCCs and clusters the entire event into two sub-event spectra. Then sub-dictionaries with two basis vectors are learned from each sub-event. The resulting W_{all} contains a total of four basis vectors. There are only two distinct acoustic elements ($C4$, $C4$-$G4$-$E4$) in the overall event, so the dimensionality of W_{all} can be reduced to 2 by identifying the two basis vectors with minimum correlation between them. These are shown in Fig. 3(c) and it can be seen that the second basis vector closely resembles the single basis vector obtained from $C4$ exemplar. Also, both basis vectors put together closely represent the $C4$-$G4$-$E4$ sound. This demonstrates the ability of the proposed method in learning dictionaries that represent the entire sound event.

For realistic sound events, GMM based clustering of event spectra is dependent on the initialization parameters, and therefore not reliable in modelling unbalanced acoustic elements accurately. If clustering of event spectra can separate the spectra of $C4$-$G4$-$E4$ and $C4$ accurately, normalizing each of the clustered sub-event spectra by their data-size should be sufficient for NMF to extract an effective dictionary from the entire normalized event spectra [8]. Hence we additionally compare the proposed method with clustering and normalized sub-dictionary learning. We evaluate the reconstruction error (log-likelihood) of the $C4$ exemplar at 100 random initializations using the proposed dictionary learning, clustering and normalized dictionary learning and the conventional NMF dictionary learning. A histogram of the 100 log-likelihoods for each method is shown in Fig. 4. Among the 5 histogram

ISBN 978-90-827970-1-5 © EURASIP 2018

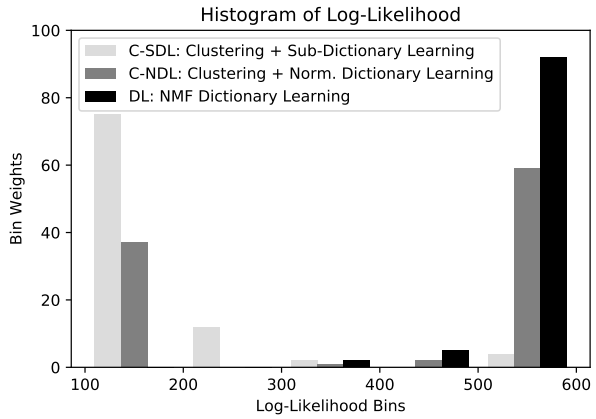

Fig. 4: Histogram of Log-Likelihoods from the reconstruction of the $C4$ exemplar from dictionaries extracting using C-SDL, C-NDL and DL methods. The first and last histogram bins indicate accurate and failed reconstructions respectively.

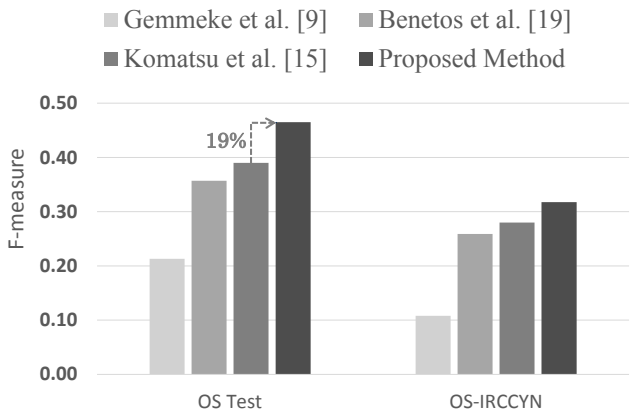

Fig. 5: Comparison of Frame-based $\mathcal{F}$-measures for the polyphonic event detection based on the proposed modelling of sound events for OS Test and OS-IRCCYN Databases with current state-of-the-art methods.

bins, the first bin indicates an accurate reconstruction and the last bin indicates a failed reconstruction. It is evident that the conventional NMF fails to represent the $C4$ exemplar at almost all initializations. The clustering and normalized dictionary learning succeeds to reconstruct at less than 40% of all initializations, while the proposed dictionary learning succeeds at more than 75%. This demonstrates the robustness of the proposed method to uncertainties in clustering effective sub-event spectra.

B. Polyphonic Sound Event Detection

Training and Testing: In this paper, we evaluate the proposed method for polyphonic sound event detection task from the IEEE challenge for DCASE 2013 [19]. The dictionaries for each sound event are learned from the isolated training database consisting of $M = 16$ sound events. Event spectra are estimated from 40ms frame sizes with a 10ms frame shift. Each event spectra is first clustered into 2 sub-event spectra by extracting 23 MFCCs excluding the 0^{th} coefficient and then modelling using a 2 component GMM. Then sub-dictionaries with 3 basis vectors are learned from each of the clustered sub-event spectra. The size of overall dictionary W_{all} is $N = 96$.

Spectra of OS (Office Synthetic) development database consisting of 9 polyphonic synthetic recordings (90 seconds each) is used for classifier training. Note that the dimensionality of MFCCs, GMM and sub-dictionaries are optimized for the OS development database. Due to a large number of sound events and polyphonic nature of the development database, class weight for event activations to be trained using binary linear-SVM classifiers is increased to 3 times. Similar to the training phase, test spectra is obtained from OS test database which contains 12 recordings (120 seconds each) with different levels of polyphony and background noises. A second database OS-IRCCYN [20] is also used for evaluating the proposed method. Both databases share the same ground truths, however the latter is obtained from audio recorded at IRCCYN, France [21]. Size of the noise dictionary $\widehat{W}$ is 1. For post-processing, the generated binary eventroll is filtered using a median filter of length 3. Gaps less than 250ms in the eventroll are filled and the final event labels with duration more than 200ms are considered to be valid.

Evaluations and Comparisons: $\mathcal{F}$-measure based metrics are used evaluation in this paper. We consider three different evaluation metrics: a 10ms frame-based $\mathcal{F}$-measure ($\mathcal{F}_{fb}$) from DCASE 2013, 100ms segment-based ($\mathcal{F}_{sb}$) and class-wise segment-based $\mathcal{F}$-measures ($\mathcal{F}_{cwsb}$) from DCASE 2016. Frame-based metrics are evaluated for each recording and their average is noted, while segment-based metrics are evaluated over the entire database.

To evaluate the proposed method, we compare it with three most relevant NMF based AED methods. First method proposed by Gemmeke et al. uses an event-wise NMF for dictionary learning and extracts event-likelihoods using Hidden Markov Models (HMM) with linear time warping [9]. Second is the probabilistic latent component analysis (PLCA) with integrated linear dynamical systems (LDS) proposed by Benetos et al. [21] for polyphonic sound event tracking. The third is the sparsely activated mixture of local dictionaries (MLD) based dictionary learning proposed by Komatsu et al. and trained using SVM classifiers [16]. For this comparison, the three NMF based AED methods are evaluated for both OS test and OS-IRCCYN databases, and their $\mathcal{F}_{fb}$ are shown in Fig. 5. An AED based on the proposed dictionary learning achieves a frame-based $\mathcal{F}$-measure of 46.5% for the OS test database and shows a significant improvement of 19.2% over the next best state-of-the-art AED method. The proposed dictionary learning has improved event modelling thereby outperforming the existing NMF based state-of-the-art methods. To the best of our knowledge, the reported $\mathcal{F}_{fb}$ of 46.5% for OS test database using the proposed method is highest in existing literature.

ISBN 978-90-827970-1-5 © EURASIP 2018

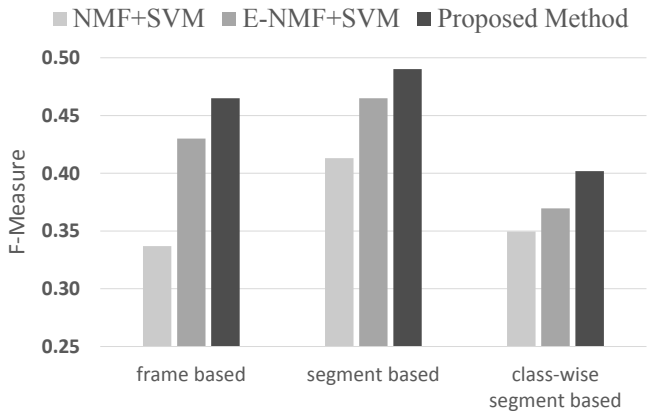

Fig. 6: Significance of the proposed modelling of sound events across various $\mathcal{F}$-measure metrics for the polyphonic DCASE 2013 OS Test Database.

Further evaluating the significance of proposed dictionary learning, we compare it with two conventional NMF-SVM based AED methods. First method learns the overall dictionary with 35 basis vectors by doing NMF to the entire audio spectra. The second method performs an event-wise NMF and groups together all the event dictionaries. Size of these event dictionaries are tested from 1 to 6, with 3 being the best. The two conventional NMF based AED methods are denoted as NMF+SVM and E-NMF+SVM respectively. A binary linear-SVM classifier is trained in all the three methods. $\mathcal{F}$-measures for the three methods are shown in Fig. 6. The proposed method achieves the highest $\mathcal{F}_{fb}, \mathcal{F}_{sb}, \mathcal{F}_{cwsb}$ of $46.5\%, 49\%$ and 40.2% respectively. This comparison shows the significance of learning sub-dictionaries separately.

V. Conclusions

In this work, we propose an effective dictionary learning of sound events with hidden data-size-imbalances, for the task of acoustic event detection. The imbalance among data-sizes of acoustic elements present in the spectra of each sound event are first estimated by clustering and the sub-dictionaries from each cluster spectra are learned separately. An illustrative example shows that the overall set of sub-dictionaries is better able to model the hidden imbalances as compared to conventional NMF based dictionary learning methods. We further demonstrate the superiority of an AED method based on the proposed dictionary learning, over three existing state-of-the-art AED methods. Rigorous mathematical work integrating the clustering and sub-dictionary learning into the NMF formulation is left for future work.

References

[1] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events," in *IEEE Transactions on Multimedia*, vol. 17, no. 10, 2015, pp. 1733–1746.

[2] L. Vuegen, B. Broeck, P. Karsmakers, J. Gemmeke, B. Vanrumste, and H. Hamme, "An mfcc-gmm approach for event detection and classification," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013, pp. 1–3.

[3] Q. Zhou and Z. Feng, "Robust sound event detection through noise estimation and source separation using NMF," DCASE2017 Challenge, Tech. Rep., September 2017.

[4] S. Adavanne and T. Virtanen, "A report on sound event detection with different binaural features," DCASE2017 Challenge, Tech. Rep., September 2017.

[5] H. Lim, J. Park, and Y. Han, "Rare sound event detection using 1D convolutional recurrent neural networks," DCASE2017 Challenge, Tech. Rep., September 2017.

[6] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," in *IEEE Signal Processing Magazine*, vol. 29, no. 6, 2012, pp. 82–97.

[7] C. V. Cotton and D. P. Ellis, "Spectral vs. spectro-temporal features for acoustic event detection," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 69–72.

[8] D. Guillamet, M. Bressan, and J. Vitria, "A weighted non-negative matrix factorization for local representations," in *IEEE Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2001, pp. I–I.

[9] J. F. Gemmeke, L. Vuegen, P. Karsmakers, B. Vanrumste *et al.*, "An exemplar-based nmf approach to audio event detection," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013, pp. 1–4.

[10] T. Komatsu, M. Tani, T. Toizumi, N. Chaitanya, M. Kato, Y. Arai, O. Hoshuyama, Y. Senda, and R. Kondo, "An acoustic monitoring system and its field trials," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2017, pp. 1341–1346.

[11] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems (NIPS)*, 2001, pp. 556–562.

[12] C. Joder and B. Schuller, "Exploring nonnegative matrix factorization for audio classification: Application to speaker recognition," in *VDE Proceedings of Speech Communication; 10. ITG Symposium*, 2012, pp. 1–4.

[13] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," in *IEEE transactions on speech and audio processing*, vol. 3, no. 1, 1995, pp. 72–83.

[14] F. Weninger, J. Feliu, and B. Schuller, "Supervised and semi-supervised suppression of background music in monaural speech recordings," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 61–64.

[15] S. Wang and J. Ortiz, "Non-negative matrix factorization of signals with overlapping events for event detection applications," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 5960–5964.

[16] T. Komatsu, T. Toizumi, R. Kondo, and Y. Senda, "Acoustic event detection method using semi-supervised non-negative matrix factorization with a mixture of local dictionaries," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 Workshop (DCASE)*, 2016, pp. 45–49.

[17] F. Weninger, C. Kirst, B. Schuller, and H.-J. Bungartz, "A discriminative approach to polyphonic piano note transcription using supervised non-negative matrix factorization," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6–10.

[18] Y. Zilu and Z. Guoyi, "Facial expression recognition based on nmf and svm," in *IEEE International Forum on Information Technology and Applications (IFITA)*, vol. 3, 2009, pp. 612–615.

[19] D. Giannoulis, E. Benetos, D. Stowell, M. Rossignol, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events: An ieee aasp challenge," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013, pp. 1–4.

[20] E. Benetos, "OS-IRCCYN datasets for Sound Event Detection," https://archive.org/details/OS-IRCCYN/, [Online; accessed 14-June-2018].

[21] E. Benetos, G. Lafay, M. Lagrange, M. D. Plumbley, E. Benetos, G. Lafay, M. Lagrange, and M. D. Plumbley, "Polyphonic sound event tracking using linear dynamical systems," in *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 25, no. 6, 2017, pp. 1266–1277.

ISBN 978-90-827970-1-5 © EURASIP 2018