Exploring Sound Source Separation for Acoustic Condition Monitoring in Industrial Scenarios

Estefanía Cano, Johannes Nowak, and Sascha Grollmisch
Fraunhofer Institute for Digital Media Technology IDMT
Ilmenau, Germany
{cano, noa, goh}@idmt.fraunhofer.de

***Abstract*—This paper evaluates the application of three methods for Sound Source Separation (SSS) in industrial acoustic condition monitoring scenarios. To evaluate the impact of SSS, we use a machine learning approach where a classifier is trained to detect a specific operating machine. The evaluation procedure is based on simulated and measured data, comprising three different machine sounds as targets and 10 interfering signals. Various intermixing levels of target and interfering signal are taken into account, using three different signal-to-interference ratios. Results show that the chosen source separation methods, originally developed for music analysis, work well for industrial signals, significantly improving the classification accuracy.**

I. Introduction

Machine condition monitoring is one of the most efficient strategies for carrying out maintenance in a great variety of industries [1]–[3]. The main idea behind condition monitoring is to assess the condition of a machine during operation by evaluating arbitrary sensor data. The collected sensor data is subsequently fed into a monitoring system to identify and classify machine malfunctions, allowing more cost-effective maintenance and manufacturing.

The focus of this paper is on Acoustic Condition Monitoring (ACM) which refers to the evaluation of the acoustic signature of a machine in an industrial production line. Several ACM scenarios and methods have been proposed [4]–[6]. However, with recent advances in Machine Learning (ML) techniques, and its successful use in a number of applications [7], the combination of ACM with ML methods is becoming more and more powerful [8]–[10].

In practical applications, however, the presence of interfering signals can significantly affect the classification accuracy. To minimize these effects, more focused acoustic analyses can be used, like for example microphone arrays for improved directivity. Additionally, more powerful ML methods such as Deep Neural Networks (DNN) can also minimize the effects of interferences on the performance of the ACM system [**?**]. Finally, as is the case of this study, Sound Source Separation (SSS) methods can be applied to separate the target sound from the interfering sources in the acoustic signal [11]–[14].

This paper evaluates the separation performance of three SSS algorithms, and the impact of the separation stage on a subsequent classification task for ACM applications. In the analysis, various target and interfering signals are evaluated using different interference levels and mixing conditions.

II. Methods for Sound Source Separation

This section briefly describes the three methods for SSS which are applied in the experiment described later in Section III, namely Azimuth Discrimination and Resynthesis (ADRess) [15], Frequency Domain Source Identification and Manipulation in Stereo Mix (FDSI) [16], and Kernel Additive Modeling for Interference Reduction (KAMIR) [17]. Even though these algorithms were developed within the music separation community, the fact that they can handle stereo signals without making strong assumptions about the harmonicity or continuity of the target source, makes them promising candidates for industrial condition monitoring applications. The three algorithms are described in the following.

The first SSS algorithm used is ADRess which works under the assumption that only an interaural intensity difference between the right and the left channel exists for a single source, exploiting this fact for image localization in stereo recordings. This method uses gain scaling and phase cancellation techniques on the frequency-azimuth plane in order to perform the separation [15].

The second algorithm FDSI achieves the stereo separation by defining a similarity measure between the Short-time Fourier Transformation (STFT) of both input channels. In order to determine the lateral direction and the panning index of the sound source, an ambiguity resolving function is applied in combination with the sinusoidal energy preserving panning law [16].

The third algorithm is KAMIR which aims at removing leakage of a given sound source into a microphone destined to capture other sources. This is a common scenario when recording ensembles such as a Jazz quartet where the microphone used to record the saxophone, very often records information from the piano, the drums or the bass. KAMIR is based on the Kernel Additive Model (KAM) proposed in [18] and assumes that each source is predominant in its dedicated channel. By using an approach based on a generalized Wiener filter, KAMIR can estimate the sources in an iterative manner [17].

III. Experiment

The experiment described in this section evaluates the separation performance of the three proposed SSS methods and analyzes the influence of the separation accuracy on the subsequent classification task. The following sections provide

ISBN 978-0-9928626-7-1

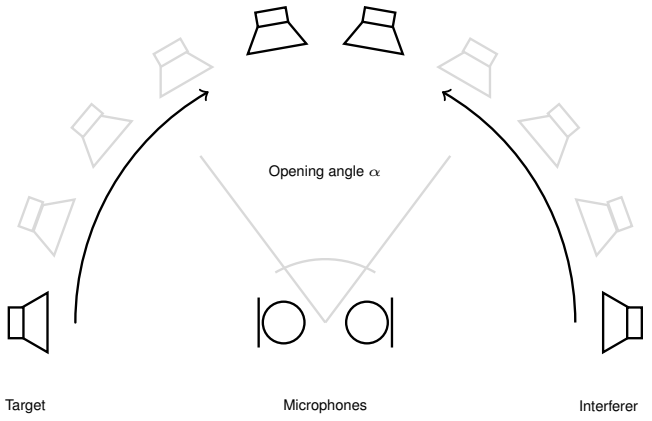

Fig. 1. Measurement setup showing the microphone and loudspeaker positioning as well as the opening angle α.

information on the derivation of the data fundamental for analysis, the evaluation of the source separation performance under simulated and realistic conditions, and their influence on the classification accuracy.

A. Dataset

The experiment is based on simulated and measured data, comprising stereo files which contain target and interference signals with different interchannel mixing levels.

In the simulations, both signals are mixed in each stereo channel using various mixing degrees. In the best-case situation both signals are completely separated to the left and right channel and optimal separation is possible. In the worst-case scenario, where the sources are separated by only 20^o of panning, both separation and classification are much more challenging tasks.

For more realistic measurements, the same conditions as the ones used in the simulations are recreated using two loudspeakers for playback and two cardioid microphones for recording. The interfering signal is played back on the right loudspeaker while the target signal is presented on the left. The opening angle α between the loudspeakers is then iteratively decreased in 10^o steps by moving both loudspeakers, hence, increasing the crosstalk between the microphones. See Fig. 1 for a depiction of the measurement setup. The recordings were conducted in the Class A anechoic chamber at Fraunhofer IDMT, using a microphone stereo setup (A/B technique). The microphones were facing opposite of each other with the left microphone pointing to the left, and the right microphone pointing to the right. The positions of the microphones were not changed during the measurements. The corresponding loudspeaker-microphone distance was 150 cm and both microphones were set up at a distance of 30 cm. TABLE I gives an overview of the 17 test conditions evaluated in the experiment. In addition to the 17 conditions representing various opening angles, all conditions were also evaluated for three Signal-Interference-Ratios (SIR): In the first, the target is 5 dB louder than the interfering signal (+5 dB SIR), in the second both signals have equal levels (0 dB SIR), and in the third, the interference is louder than the target (−5 dB SIR). A total of three target machines (TM) and ten interfering noises served as test material. The target machines included three different types of DC engines which are referred to as TM1, TM2, and TM3 in the following descriptions. The interfering signals included machine noises such as fans, drills, and washing machines as well as environmental sounds, human noises, and white noise.

TABLE I
OVERVIEW OF TEST CONDITIONS REPRESENTING THE INTERCHANNEL MIXING LEVEL AS OPENING ANGLE α BETWEEN TARGET AND INTERFERER SIGNAL.

Conditions	Opening angle α
C1	180^o
C2	170^o
C3	160^o
C4	150^o
C5	140^o
C6	130^o
C7	120^o
C8	110^o
C9	100^o
C10	90^o
C11	80^o
C12	70^o
C13	60^o
C14	50^o
C15	40^o
C16	30^o
C17	20^o

B. Sound Source Separation

The SSS methods described in Section III were applied on both the simulation and measurement data in the attempt to remove the interfering sounds from the target sources. As a mathematical measure of degree of separation, the Signal-to-Distortion Ratio (SDR) was used [19]. It is important to note that the focus of these experiments was on improving classification accuracy of machine sounds. The improvement of the perceptual separation quality is not a goal in itself. Thus, the choice of SDR as a quality measure remains valid.

1) Simulations: The following analysis evaluates the source separation performance for the simulated data over all conditions, i.e., opening angles, for the three SSS algorithms and the three SIRs. The separation results are plotted in Fig. 2, showing the SDRs as boxplots over all panning angles with the boxes comprising all target-noise combinations. The rows represent the different algorithms, with ADRess shown in the top, FDSI in the middle, and KAMIR in the bottom row, whereas the different SIRs are given in the column plots.

As expected, results show that for all algorithms and all SIRs, the separation performance decreases with decreasing opening angle, i.e., with increasing interchannel mixing levels. Overall, it can be seen that ADRess provides the highest separation performance while FDSI and KAMIR yield slightly

ISBN 978-0-9928626-7-1 © EURASIP 2017

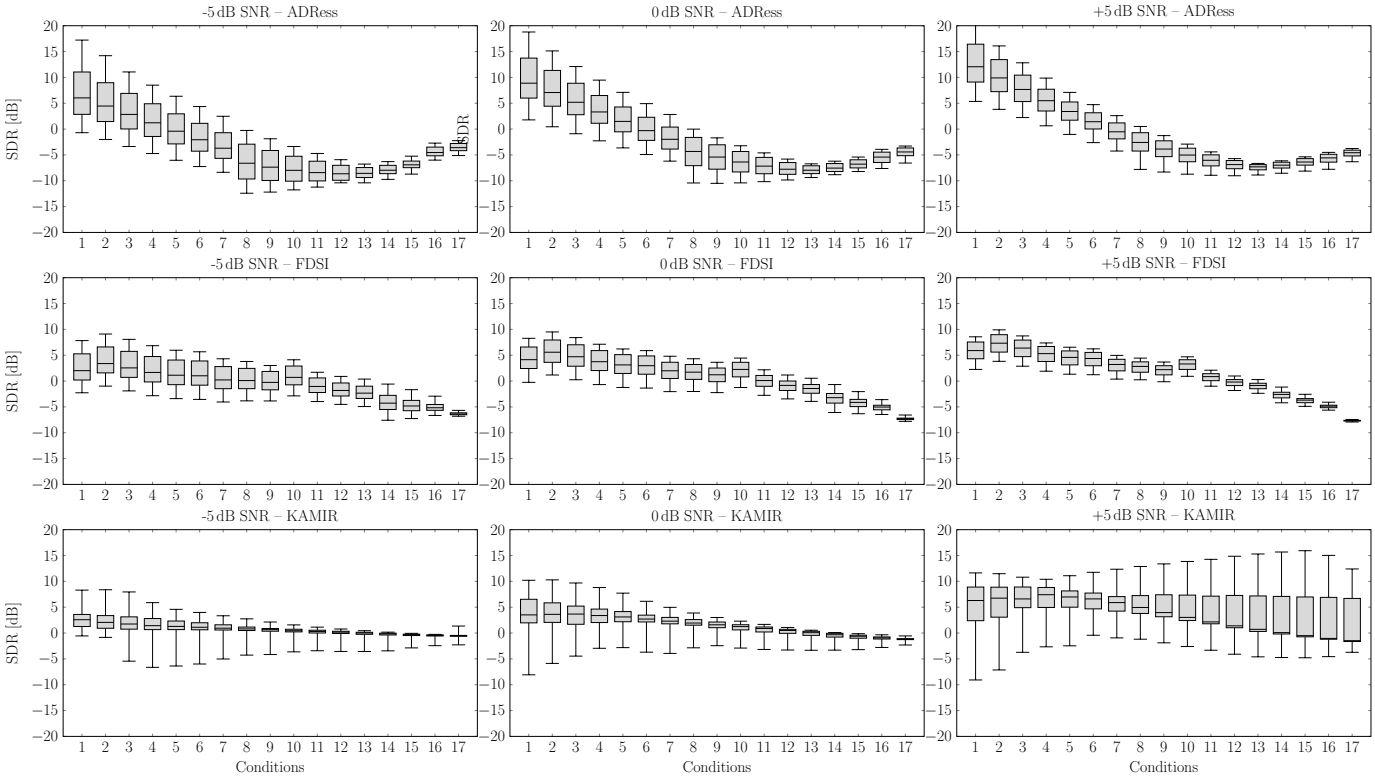

Fig. 2. Separation performance shown as boxplots for all simulated signals over all conditions, with rows representing the three SSS methods and columns the corresponding SIRs. Note that distributional outliers, i.e., data lying outside the 99.3 % distribution, are omitted in the plots for readability.

lower SDR values. As can be seen in the top row of Fig. 2, the separation performance of ADRess shows decreasing SDRs up to Condition C12 or C13, even reaching negative values. After C13, SDR values slightly rise again. Also note that the width of the boxes decreases with decreasing opening angle, indicating less variances across all target and interfering signals. Although the overall behavior is similar for all SIRs, it can be seen that the separation performance is better when the amplitude of the interference is lower than that of the target, it is slightly lower if they have similar levels, and even lower if the interfering signal is louder than the target. The second algorithm FDSI shows a different behavior. Over all SIRs, separation performance drops in a nearly linear fashion, with SDRs below zero for conditions C12 and higher. Like for ADRess, the variances between target and interfering signals decrease for decreasing opening angles. The third algorithm KAMIR shows a similar behavior to FDSI for SIRs $0\,\mathrm{dB}$ and $+5\,\mathrm{dB}$, although SDR values below zero are only observable for conditions C16 and C17. For $-5\,\mathrm{dB}$ SIR, the variances across all target-interferer combinations increase for decreasing opening angles, but the median of the distribution shows the expected tendency of decreasing SDR. It has to be noted that, as the opening angles decrease and the crosstalk between microphones increases, the assumption made in KAMIR only weakly holds, i.e., that the target source is predominant in its dedicated channel. This might explain the increased variance observed for $-5\,\mathrm{dB}$ SIR. The simulation results suggest that the ADRess algorithm is the most suitable method for the separation scenario described, followed by FDSI and KAMIR. However, comparing the SDR scale between the three methods, ADRess also provides a wider separation range, indicating higher sensitivity to opening angles and mixing levels.

In the following, the same experiment is conducted for the measured data to evaluate the separation performance under more realistic conditions.

2) Measurements: The separation results for the measured data are shown in Fig. 3. In general, results show a similar trend as in the simulations but with clearly higher variances, i.e., larger box widths. Also, the separation performance is higher if the target signal is less masked by the interference, whereas for higher SIRs, separation accuracy drops for all methods. Also, as expected for realistic conditions, the overall performance in terms of SDR is lower compared to the simulations. Like in the simulations, the performance for ADRess drops for decreasing opening angles, also providing SDR values below zero. However, in contrast to the simulations, the first six conditions are mostly on the same level and show no significant differences. This might be due to the crosstalk between the channels which cannot be avoided in practical applications. The results for FDSI are in accordance to the simulation results with a more or less linear SDR decrement for decreasing opening angles. The separation performance of KAMIR exhibits the lowest SDR values, but with less

ISBN 978-0-9928626-7-1 © EURASIP 2017

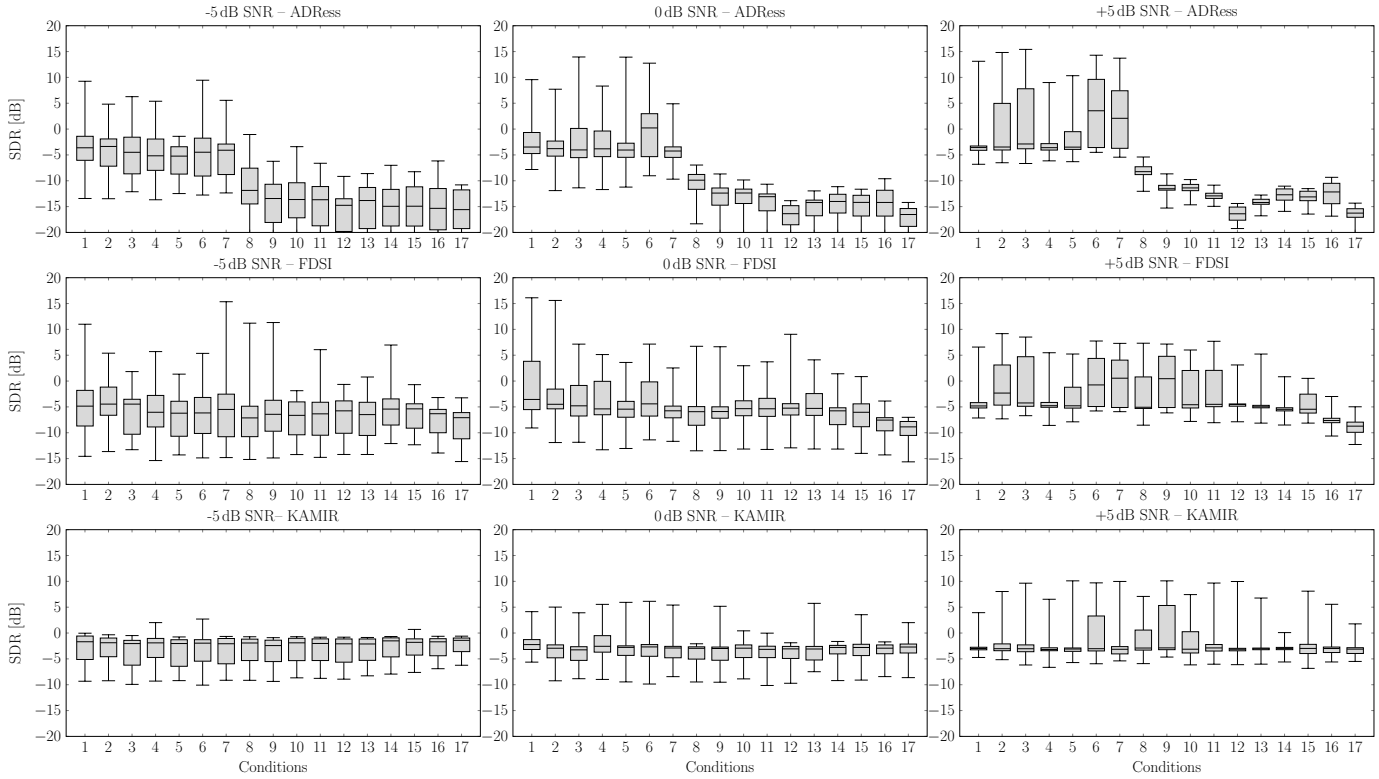

Fig. 3. Separation performance shown as notched boxplots for all recorded signals over all conditions, with rows representing the three SSS methods and columns the corresponding SIRs. Note that distributional outliers, i.e., data lying outside the 99.3 % distribution, are omitted in the plots for readability.

sensitivity to changes in the opening angle.

Overall, results indicate that the ADRess method is best suitable for target-interferer-separation, followed by FDSI and KAMIR. As expected, their separation performances are improved for higher SIRs. Note that, due to the frequency-dependent directivity of the microphones and loudspeakers, the intermixing levels between the target and interfering signals do not exactly replicate the situation of the simulations. However, as shown in the plots, results are still comparable.

C. Classification

This section presents results from a classification task designed to evaluate the impact of SSS in the classification accuracy of a method trained to detect a specific running engine (TM3) within an industrial environment.

a) Model training: A Support Vector Machine (SVM) binary classifier was trained to recognize the running engine of TM3 against other environmental and industrial sounds. A new training dataset was recorded with 20 hours of audio material (10hrs running engine/10hrs other industrial noises) including clean signals of the target machine as well as mixed signals of the target machine with six types of interfering noises such as a fan engine, washing machine, speech, laughter, radio, or people walking around, mixed at different SIRs ($+5$ dB, 0 dB and -5 dB). To improve the classifier robustness against microphone characteristics, 20 different microphones were used in the recordings. The SVM classifier was trained with spectral features including flatness and spectral centroid, using an RBF kernel with grid search.

b) Testing: For testing, another dataset of 2 hours of audio material was recorded (1hr running engine/1hr other industrial noises), including clean signals of the interfering sounds and mixtures of two interfering sounds at different SIRs. Additionally, to discard effects of room acoustics, the test dataset was recorded in a different room. A classification accuracy of 96.7% was obtained with this dataset.

To evaluate the impact of SSS techniques on our condition monitoring scenario, the dataset described in Section III-A was also used for testing, including the separated sounds obtained with the three separation algorithms. The signals from TM3 were used as target while the signals from TM1 mixed with the 10 interfering noises were used as the other condition of the binary classifier.

c) Results: It can be observed in Fig. 4 that there is a considerable drop in classification performance when interfering noises are present (Mix) compared to clean signals. As in realistic industrial scenarios interfering noises are bound to happen, and the ability to minimize the effects of interferences in the classification task is of utmost importance. In general, an improvement in classification accuracy can be observed for all algorithms and SIRs when sound separation is applied, both for the simulations and the measurements. As expected, classification accuracy is in general slightly higher for the simulation data than for the measurements. For the mixed data

ISBN 978-0-9928626-7-1 © EURASIP 2017

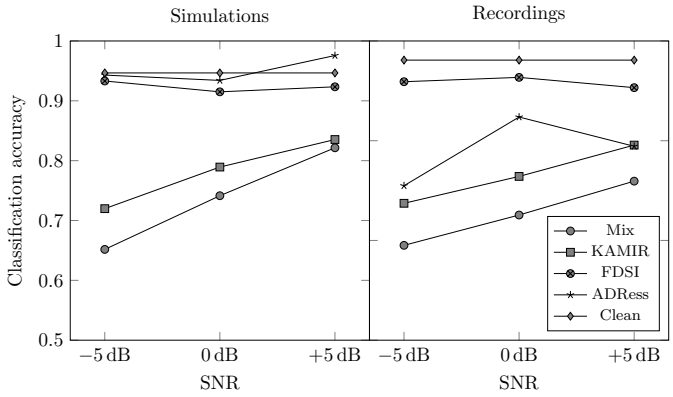

Fig. 4. Measurement setup showing the microphone and loudspeaker positioning as well as the opening angle α.

as well as for the KAMIR algorithm, a clear improvement in classification accuracy can be observed with higher SIRs. In contrast, this trend is not clearly observed for ADRess or FDSI, although they provide higher classification accuracies.

IV. Conclusion

This paper addressed the application of SSS methods for industrial acoustic condition monitoring scenarios. Based on data derived from simulations and real measurements, the separation performance of three SSS methods was evaluated, namely ADRess, FDSI, and KAMIR, by relating the influence of the separation accuracy on a subsequent classification task using an SVM. The data comprised stereo files with different mixes of target and interference signals, including a variety of signal types at different SIRs in both simulations and real recordings. Results showed that the three algorithms can achieve separation between target and interference, with performance decreasing for increasing mixing level. Results from the measurement data were in accordance to the simulations although the overall performance was higher in the ideal scenarios. The subsequent classifier also performed well, providing accuracies between $59\,\%$ and $97\,\%$ depending on the mixing level. Results indicate that the ADRess model performed best under various recording and simulation conditions. As expected, the classification accuracy decreases for decreasing SIR. Overall, results suggest that—as a proof-of-concept—SSS methods from music information retrieval may also be suitable for separating machine sounds. In this case, classification accuracy was used as a performance measure to validate the contribution of SSS in industrial monitoring scenarios. For future work, a more thorough analysis of the influence of realistic scenarios is proposed, taking different reflective environments and the influence of spatially distributed interfering noises into account.

Acknowledgment

The recordings and simulations used for this study were conducted with the support of Abhijatha K Banashankarappa and Brijesh B Parappa in the facilities of Fraunhofer IDMT.

References

[1] R. Randall, *Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications*, ser. EBL-Schweitzer. Wiley, 2011.

[2] F. P. G. Mrquez, A. M. Tobias, J. M. P. Prez, and M. Papaelias, "Condition monitoring of wind turbines: Techniques and methods," *Renewable Energy*, vol. 46, pp. 169 – 178, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148112001899

[3] X. Li, "A brief review: acoustic emission method for tool wear monitoring during turning," *International Journal of Machine Tools and Manufacture*, vol. 42, no. 2, pp. 157 – 165, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0890695501001080

[4] X. Liu, X. Wu, and C. Liu, "A comparsion of acoustic emission and vibration on bearing fault detection," in *Proc. IEEE Int. Conf. Transportation Mech. Elect. Eng. (TMEE2011)*, ChangChun, China, 2011, pp. 922–926.

[5] F. Elasha, M. Greaves, D. Mba, and A. Addali, "Application of acoustic emission in diagnostic of bearing faults within a helicopter gearbox," in *Proc. 4th Int. Conf. Through-life Eng. Services (CIRP)*, vol. 38, ChangChun, China, 2015, pp. 30–36.

[6] A. Purarjomandlangrudi and G. Nourbakhsh, "Acoustic emission condition monitoring: An application for wind turbine fault detection," *Int. J. Res. Eng. Technol. (IJRET)*, vol. 2, no. 5, pp. 907–918, 2013.

[7] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] N. F. Ince, C.-S. Cao, M. Kaveh, A. Tewfik, and J.F. Jabuz, "A machine learning approach for locating acoustic emission," *EURASIP J. Advances Signal Process.*, vol. 2010, no. 15, 2010.

[9] J. H. Zhou, C. K. Pang, Z. W. Zhong, F. L. Lewis, and J.F. Jabuz, "Tool wear monitoring using acoustic emissions by dominant-feature identification," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 547–559, 2011.

[10] W. L. Woon, A. El-Hag, and M. Harbaji, "Machine learning techniques for robust classification of partial discharges in oilpaper insulation systems," *IET Sci. Meas. & Technol.*, vol. 10, no. 3, pp. 221–227, 2016.

[11] A. Liutkus, J.L. Durrieu, L. Daudet, and G. Richard, "An overview of informed audio source separation," in *Proc. 14th Int. Workshop Image Anal. Multimedia Interactive Services (WIAMIS)*, Paris, France, 2013, pp. 1–4.

[12] D. FitzGerald, A. Liutkus, and R. Badeau, "Projection-based demixing of spatial audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1560–1572, Sept 2016.

[13] S. Uhlich, F. Giron, and Y. Mitsufuji, "Deep neural network based instrument extraction from music," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2135–2139.

[14] A. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1652–1664, Sept 2016.

[15] D. Barry, B. Lawlor, and E. Coyle, "Sound source separation: azimuth discrimination and resynthesis," in *Proc. 7th Int. Conf. Digital Audio Effects (DAFx'04)*, Naples, Italy, 2004.

[16] C. Avendano, "Frequency-domain source identification and manipulation in stereo mixes for enhancement, suppression and re-panning applications," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New York, USA, 2003, pp. 55–58.

[17] T. Prätzlich, R. M. Bittner, A. Liutkus, and M. Müller, "Kernel additive modeling for interference reduction in multi-channel music recordings," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Brisbane, Australia, 2015, pp. 584–588.

[18] A. Liutkus, D. Fitzgerald, Z. Rafii, B. Pardo, and L. Daudet, "Kernel additive models for source separation," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4298–4310, 2014.

[19] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Language Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.