

Intelligentes akustisches Monitoring durch ausgewählte Mikrofonierungskonzepte

Tobias Fritsch¹, Joachim Bös¹, Sascha Grollmisch¹, Saichand Gourishetti¹,
Peter Hofmann¹ und Judith Liebetrau¹

¹ Fraunhofer Institute for Digital Media Technology IDMT, 98693 Ilmenau, Germany, E-Mail: joachim.boes@idmt.fraunhofer.de

Einleitung

Ein erfahrener Tonmeister weiß vor der Aufnahme, wo er welches Mikrofon aufstellen muss, damit er den gewünschten Klang erfasst. Für weniger Erfahrene bietet die Literatur ausführliche Informationen, Hinweise und Tabellen z.B. mit Abstrahlcharakteristiken von Musikinstrumenten [1].

Für das Mikrofonieren von Maschinen, insbesondere zur automatisierten Fehlererkennung und Zustandsüberwachung, stellt sich das Problem grundlegend anders dar. Erfahrungsschätze sind oft technologiespezifisch und ein umfassendes Nachschlagewerk ist, allein wegen des Umfangs, nicht vorhanden. Zudem entzieht sich die Qualität der Mikrofonierung einer direkten Kontrolle durch den Menschen. Es steht hier nicht der gute Klang oder eine saubere räumliche Staffelung im Vordergrund. Stattdessen muss der Fehlerindikator von einer künstlichen Intelligenz(KI) störungsfrei gefunden und ausgewertet werden können. Das Wort Fehlerindikator bezeichnet hierbei den Teil des Gesamtschalls, der geeignete und robuste Informationen über den Zustand der zu überwachenden Anlage enthält.

Bei einem rutschenden Keilriemen beispielsweise wäre der Fehlerindikator nach menschlicher Wahrnehmung intuitiv ein quietschendes Geräusch. Auch in diesem scheinbar einfachen Fall müssen jedoch unterschiedene Aspekte bedacht werden: Ist „Quietschen“ überhaupt ein geeigneter Fehlerindikator, oder gibt es andere Geräusche die schon viel früher auf einen rutschenden Keilriemen hinweisen können? Wo entsteht das Geräusch, durch welche Spektralanteile oder deren komplexes Verhalten über die Zeit kann es charakterisiert werden und in welche Richtung wird es vornehmlich abgestrahlt? Was ist folglich die optimale Mikrofonauswahl und Position? Welchen Einfluss haben Störungen, wie Wind- und Motorgeräusche und wie lassen sie sich abbildern?

Um diese Fragen zufriedenstellend zu Beantworten bedarf es eingehender Untersuchungen. Dabei ist das zuverlässige Auftreten oder nicht Auftreten des Fehlerindikators erforderlich. Diese Bedingung ist in der Praxis nicht immer erfüllt. Um dennoch möglichst allgemeingültige Erkenntnisse über das effektive Zusammenwirken von Mikrofonierung und der Auswertbarkeit durch KI zu erhalten wird hier ein Datensatz vorgestellt, der systematische Untersuchungen und Vergleiche von verschiedenen Mikrofonierungskonzepten und ihrem Zusammenspiel mit KI-Methodiken erlaubt.

Stand der Technik

In den letzten Jahren wurde mehrere Datensätzen für die akustische Qualitätskontrolle [2, 3, 4] und Anomaliedetektion [5, 6] veröffentlicht. Dies demonstriert das

wachsende wissenschaftliche Interesse an diesem Forschungsbereich. Die Datensätze umfassen vielfältige Anwendungsfälle wie die Zustandsüberwachung von Elektromotoren, Pumpen und Ventilatoren aber auch die Material- und Oberflächenerkennung von Werkstücken anhand des Klangs. Jedoch sind diese Aufnahmen entweder mit nur einem Mikrofon oder einem festen Setup aufgenommen, welches keine detaillierte Auswertung hinsichtlich unterschiedlicher Mikrofonplatzierung oder den Einfluss der Abstrahlcharakteristik auf die Erkennungsgenauigkeit ermöglicht. Auch die notwendigen Schritte, welche zur Auswahl und Platzierung der Mikrofone führten, sind wenig erörtert. Lediglich der Einfluss des Abstandes zwischen Mikrofon und Druckluftleckage wurde bisher näher untersucht [4]. Entsprechend sind den Autoren keine Quellen oder veröffentlichten Datensätze für die akustische Zustandsüberwachung bekannt, die eine systematische Untersuchung dieser Einflussfaktoren ermöglichen.

Erstellung eines Datensatzes für das Evaluieren von Mikrofonierungskonzepten

Innerhalb in einer Industriehalle wird es kaum möglich sein, kontrollierbare Bedingungen zu schaffen, um die optimale Mikrofonposition herauszuarbeiten. Wir versprechen uns jedoch einen Qualitäts- und Kostenvorteil von Heuristiken zu Mikrofonierungskonzepten. Um diese erfolgreich ableiten zu können muss der Versuchsaufbau einerseits realitätsnahe Bedingungen abbilden, andererseits gut kontrollierbar sein, um die Mikrofonierung ohne äußere Einflüsse zu betrachten. Die Anwendung der allgemein gehaltenen Heuristiken muss dann selbstverständlich an die Gegebenheiten vor Ort angepasst werden.

Versuchsaufbau

Statt eine Maschine technisch so zu Manipulieren, dass sie verschiedene reproduzierbare Fehlerindikatoren von sich gibt, deren Klangcharakteristik, Abstrahlort und Richtung bekannt sind, wird ein achtkanaliges Beamforming-Array verwendet. Es gibt den Fehlerindikator in neun Kombinationen aus Abstrahl- und Öffnungswinkel wieder. In dem Beispiel des quietschenden Keilriemens ist dieses Signal das isolierte Quietschen. Im Datensatz trägt es den Namen „Maschinenfehler“. Alle Geräusche der normal funktionierenden Maschine werden über einen Lautsprecher kurz hinter dem Beamforming-Array abgespielt. Im Beispiel sind dies periodische Rollgeräusche. Dieses Signal heißt im Datensatz „Maschinenlärm“. Störgeräusche aus der Umgebung werden über einen dritten, schräg gegenüberliegenden

Lautsprecher wiedergeben. Dies könnten beispielsweise Motoren- und Windgeräusche sein. Im Datensatz trägt dieses Signal den Namen „Umgebungsgeräusch“. Das vierte Signal „Umgebungsfehler“ berücksichtigt die Möglichkeit, dass bei Maschinen in der Umgebung ähnliche Fehler auftreten können. Dieses Signal wird gemeinsam mit dem „Umgebungsgeräusch“ über den dritten Lautsprecher wiedergegeben.

Inmitten der Lautsprecher-Anordnung befindet sich ein „Mikrofonierungs-Bereich“. Hier werden im Verlauf des Versuches 34 Mikrofone positioniert. Es handelt sich dabei um MT Gefell M360 1/4" Messmikrofone mit Kugelcharakteristik und Sennheiser MKH 800 P48 Studiomikrofone mit umschaltbarer Nieren- und Supernierencharakteristik. Der beschriebene Aufbau steht in einem reflexionsarmen Vollraum der Klasse 1 mit einer unteren Grenzfrequenz von 63 Hz. Der Versuchsaufbau mit einem beispielhaften Mikrofonaufbau ist als Draufsicht in Abbildung 1 skizziert.

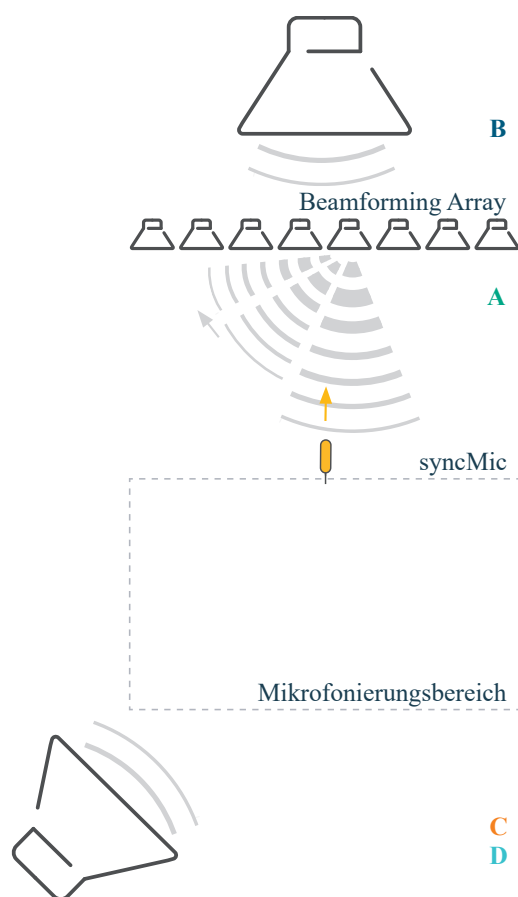


Abbildung 1: Skizze des Versuchsaufbaus im reflexionsarmen Raum als Draufsicht. Gezeigt sind das Beamforming-Array, aus welchem das Signal „Maschinenfehler“ (A) wiedergegeben wird, sowie die zwei Lautsprecher aus denen die Signale „Maschinenlärm“ (B), „Umgebungsgeräusch“ (C) und „Umgebungsfehler“ (D) wiedergegeben werden. Außerdem die Position des Synchronisationsmikrofones (syncMic) und der Mikrofonierungsbereich. Die Skizze ist Unmaßstäblich.

Datensätze

Die Ziele bei der Auswahl der Daten sind Realismus und Kontrollierbarkeit. Um beides zu erreichen wird ein synthetischer und ein in realen Situationen aufgenommener

Datensatz vorgestellt.

Synthetischer Datensatz

Der synthetische Datensatz basiert auf gefiltertem Rauschen, welches durch Gates, Kompressoren und andere Effekte verändert wird. Der „Maschinenfehler“ ist eine Reihe zufälliger Klickgeräusche und ähnelt einer Teilchenentladung oder entstehenden Rissen. Der „Maschinenlärm“ ist ein aperiodisches Wummern und ähnelt einer Rotationsmaschine unter Lastwechseln. Als „Umgebungsgeräusch“ dient gefiltertes Rauschen. Der „Umgebungsfehler“ enthält ebenfalls eine Reihe zufälliger Klicks des „Maschinenfehlers“.

Druckluftdatensatz

Der zweite Datensatz basiert auf dem bestehenden IDMT-ISA-COMPRESSED-AIR Datensatz¹ [4]. Der „Maschinenfehler“ besteht aus aufgenommenen Strömungsgeräuschen. Diese stammen aus Aufnahmen von drei verschiedenen Arten absichtlich erzeugter Undichtigkeiten in verschiedener Stärke bei Druckluftanlagen. Im Datensatz wird dies in Ventileinstellungen $\in [5, 5; 9]$ ausgedrückt. Der „Maschinenlärm“ kommt von der Aufnahme einer hydraulischen Anlage und ist im Wesentlichen Quietschen und Rumpeln. Der „Umgebungsgeräusch“ ist eine Aufnahme aus einer Maschinenhalle und enthält neben sprechenden Menschen diverse andere Hintergrundereignisse wie Maschinenlärm und Hammerschläge. Der „Umgebungsfehler“ ist wie schon im synthetischen Datensatz von der gleichen Beschaffenheit wie der „Maschinenfehler“, also eine Leckage, nur dass ein anderer Zeitausschnitt wiedergegeben wird.

Permutation der Signale

Um zu untersuchen, wie stark der Einfluss des Störschalls ist und ab welchem Pegel der Fehlerindikator (Maschinenfehler) erkannt wird, werden die Wiedergebepegel der vier Signale in verschiedenen Permutationen variiert.

Für den Druckluftdatensatz entstehen 48 Lautstärke-Permutationen, welche in Abbildung 2 gezeigt werden. Bei dem synthetischen Datensatz werden die Signale „Umgebungsgeräusch“ und „Umgebungsfehler“ jeweils zusätzlich mit 50 % Pegel variiert. Es ergeben sich folglich 108 Permutationen. Jede dieser Permutationen wird mit jeder der neun Beamformingeinstellungen wiedergegeben. Das führt bei dem Druckluftdatensatz mit drei Leckagefehlern und den jeweils acht Luftdrücken zu 9288 einzelnen Aufnahmen. Bei dem synthetischen Datensatz gibt es keine Unterscheidung in Fehlerkategorien. Dafür wird der Datensatz vier mal mit randomisierten Zeitabschnitten wiederholt. Generell werden die jeweils drei Sekunden langen Ausschnitte aus den vier Signalen für jede der Permutationen zufällig aus den mehrminütigen synthetischen oder aufgenommenen Audiodateien herausgeschnitten. Dabei entstehen keine Pausen oder sprunghafte Übergänge.

Synchronisation des Datensatzes

Vor jeder Aufzeichnung wird ein 1 kHz Ton mit der Dauer von 100 ms aus dem Lautsprecher für den „Maschi-

¹Verfügbar auf <https://www.idmt.fraunhofer.de/en/publications/datasets/isa-compressed-air.html>

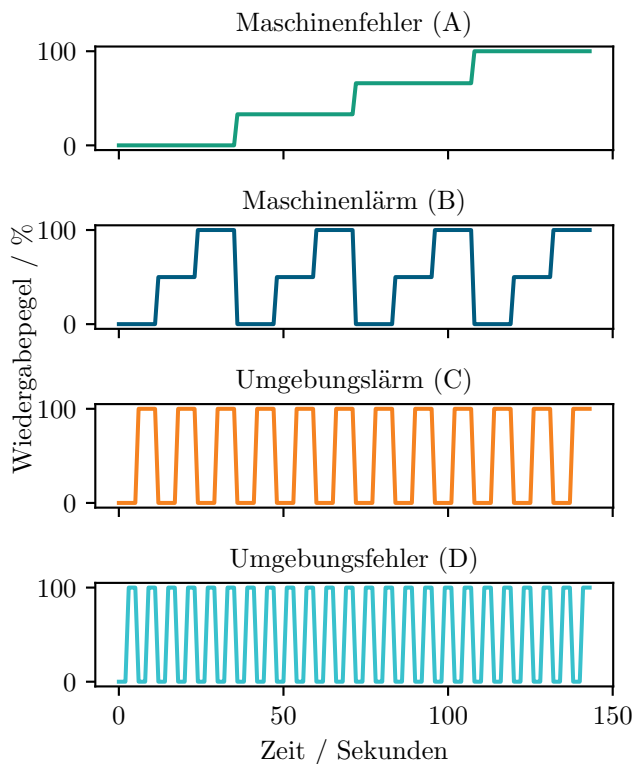


Abbildung 2: Permutation verschiedener Wiedergabepegel der vier Signale in der zeitlichen Abfolge

nenlärm“ wiedergegeben. Hierdurch sind die Laufzeitunterschiede zwischen den Mikrofonen und dem in jedem Aufbau befindlichen „Synchronisierungsmikrofon“ (syncMic) bekannt und könnten bei Bedarf kompensiert werden. Der Synchronisationston wurde nicht über das Beamforming-Array wiedergegeben. Dies soll den unbekanntem Entstehungsort des Fehlerindikators repräsentieren.

Die Daten im Datensatz sind synchron aber nicht laufzeitkompensiert. Dies entspricht am ehesten den realen Bedingungen. Da die dreisekündigen Permutationen jeweils einzeln aufgenommen werden ergibt sich beim Zerschneiden in einem festen Dreisekundenraster zu Beginn oder am Ende der Abschnitte eine laufzeitabhängige Stille. Dies entspricht nicht dem realistischen Verhalten. Die Abschnitte sind daher so auf 2,9 Sekunden gekürzt, dass der ursprüngliche Zeitbezug erhalten bleibt.

Limitierungen des Aufbaus

Die Notwendigkeit zur Aufteilung der beiden Maschinensignale auf unterschiedliche Lautsprecher ergibt sich aus der limitierten Lautstärke des Beamforming-Arrays, gerade für tiefe Frequenzen. Dies kann aufgrund von Reflexionen und Auslöschungen als Problem gesehen werden, entspricht aber dem in der Praxis ebenfalls unbekanntem Entstehungsort des Fehlerindikators.

Das Beamforming-Array bündelt ab etwa 500 Hz stark genug. Folglich wurde das Signal „Maschinenfehler“ entsprechend hochpassgefiltert. Um das Signal „Umgebungsfehler“ nicht schon durch das bloße Vorhandensein eines Tieftonanteils vom „Maschinenfehler“ unterscheiden zu können, wurde auch dieses Signal hochpassgefil-

tert. Dennoch unterscheiden sich beide Signale durch die Übertragungseigenschaften der unterschiedlichen Lautsprecher. In realistischen Fällen ist kaum zu erwarten, dass zeitgleich zwei nahezu identische Fehlerindikatoren aus unterschiedlichen Richtungen abgegeben werden, daher wird dies akzeptiert. Welche Auswirkung die unterschiedlichen Frequenzgänge der beiden Fehlersignale am Ende auf die Auswertung hat, hängt von den angewendeten KI-Methodik ab und lässt sich zum jetzigen Zeitpunkt kaum einschätzen.

Die Datensätze werden in einer reflexionsarmen Umgebung aufgenommen, was dem Anspruch an Realitätstreue entgegensteht. Da die Reflexions-Bedingungen aber zwischen verschiedenen Anwendungsszenarien, wie unter der Motorhaube, in Industriehallen oder Offshore-Windparks stark variieren, wurde ihr Einfluss für unsere verallgemeinernden Betrachtungen möglichst ausgeschlossen. Bei einem konkreten Anwendungsfall müssen die vorherrschenden Reflexionsbedingungen betrachtet gegebenenfalls optimiert werden.

Nachteilig an dem vorgestellten Aufbau ist die Reduktion des Umgebungs-lärms auf lediglich eine Schallquelle. Diese Einschränkung wird bewusst in Kauf genommen um den Aufbau- sowie Auswertungsaufwand gering und den Aufbau übersichtlich zu halten. Die Erwartung ist, dass eine „rotationssymmetrische“ Betrachtung des Aufbaus möglich ist. So lässt sich untersuchen, welche KI-Methodiken mit Hilfe von Mikrofonen in der Umgebung der Störquelle eine saubere Unterscheidung zwischen Fehlerindikator und Störschall ermöglicht. Die gewonnenen Erkenntnisse sollten sich auf eine Vielzahl an Störschallquellen und Mikrofonen übertragen lassen.

Betrachtet man die gemessenen Isobarendarstellungen der Abstrahlcharakteristik des Beamforming-Arrays, von denen eine beispielhaft in Abbildungen 3 gezeigt wird, fallen schnell die Nebenkeulen auf, die sich ab 5 kHz bilden. Auch realistische Systeme haben derartige Nebenkeulen, das Artefakt im Bereich von 15 kHz könnte bei der Auswertung dennoch Probleme verursachen. Gegebenenfalls müssen die gemessenen Daten entsprechend tiefpassgefiltert werden, um die Nebenkeulen und das Artefakt ganz oder teilweise auszublenden. Die hier gezeigte Isobarendarstellung ist beispielhaft. Gemeinsam mit dem Datensatz werden alle neun Isobardarstellungen veröffentlicht.

Auswertung

Die folgende Untersuchung stellt einen ersten beispielhaften Schritt zur detaillierten Auswertung des Datensatzes dar. Ein einfach trennbarer Merkmalsraum ermöglicht die Nutzung von simpleren Ansätzen des maschinellen Lernens, welche auch mit weniger Trainingsdaten robuste Ergebnisse erzielen können. Hierbei kann die Platzierung und Auswahl der Mikrofone eine entscheidende Rolle für das Gesamtsystem spielen. Abbildung 4 zeigt das Ergebnis einer Hauptkomponentenanalyse (Principal Component Analysis PCA) mit 2 Hauptkomponenten für unterschiedliche Umgebungs- und Maschinenfehlerpegelstufen. Als Eingangsdaten wurden über Fenster von einer Sekunde Länge mit einer Überlappung von einer halben Sekunde die Mittelwerte und Standardabweichun-

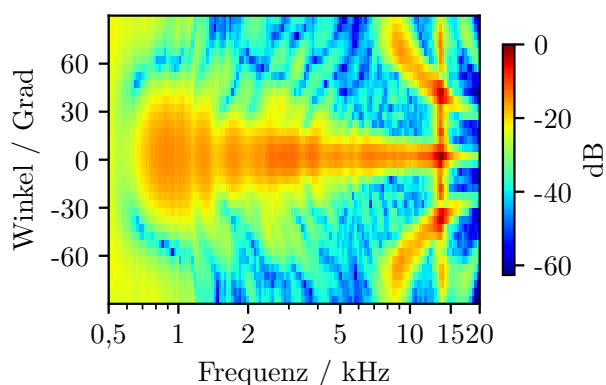


Abbildung 3: Isobarendarstellung des Beamforming-Arrays für die Einstellung: Abstrahlwinkel = 0° bezogen auf die 0° -Achse des Arrays und halber Öffnungswinkel = $2,5^\circ$ bezogen auf den Abstrahlwinkel

gen von etablierten Akustikmerkmalen genommen. Hierzu zählen Mel Frequency Cepstral Coefficients (13 Koeffizienten), RMSE, Centroid, Bandbreite, Flatness und Rolloff². Wie in Abbildung 4 ist ohne Umgebungslärm selbst bei der Nutzung einfacher Merkmale eine deutliche Trennung zwischen IO (A 0%) und nIO (A 33% und mehr) erkennbar. Mit Umgebungslärm verschlechtert sich die Trennbarkeit deutlich und es besteht eine große Überlappung zwischen den Klassen.

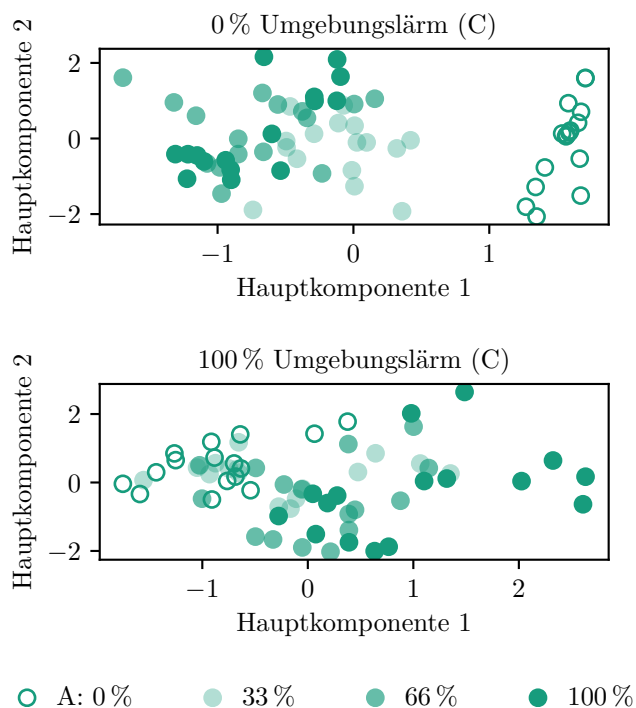


Abbildung 4: Auswirkung des Umgebungslärmes (C) mit 0% und 100% Pegel auf die Unterscheidbarkeit des „Maschinenfehlers“ in 4 Pegelstufen (A 0% bis 100%) bei Verwendung einer auf das Array gerichteten Superniere am synthetischen Datensatz.

²Weitere Informationen zu den Merkmalen finden sich unter <https://librosa.org/doc/main/feature.html#spectral-features>.

Ausblick

Die gezeigten Auswertungen sind erst der Anfang der Forschung, die durch den vorgestellten Datensatz ermöglicht wird. Fortgeschrittenere Methoden wie Quellentrennung, an den Anwendungsfall angepasste Merkmale oder der Einsatz von KI können auch unter Störschall gute Trennbarkeit erreichen [7, 2, 8] und stellen entsprechend wichtige zukünftige Forschungsarbeiten für diesen Datensatz dar. Außerdem ist es denkbar, Methoden zu testen, bei denen frequenzselektiv bestimmte Richtcharakteristika aus den Mikrofonen erlernt werden und eventuell Mikrofone eingespart werden können. Weitere potentiell beantwortbare Fragen sind: Kann ich mit Aufnahmen von der Hauptabstrahlachse trainieren und das Mikrofon für den Dauerbetrieb später in einer Nebenecke positionieren? Lassen sich die Vorteile von Mikrofonen mit Richtcharakteristik auch mit einem geometrisch nicht einwandfreien Aufbau aus Kugelmikrofonen erreichen?

Durch die geplante Veröffentlichung des Datensatzes könnten bisher ungeahnte Untersuchungen auf den Datensatz fußen, welche die angehende Beantwortung der Fragen ermöglichen.

Literatur

- [1] M. Dickreiter, *Mikrofonaufnahme: Aufnahmeräume, Instrumente, Mikrofone, Stereo- & Surroundaufnahme ; mit 44 Tabellen*, 4th ed. Stuttgart: Hirzel, 2011.
- [2] S. Grollmisch, J. Abeßer, J. Liebetrau, and H. Lukashevich, “Sounding Industry: Challenges and Datasets for Industrial Sound Analysis,” in *EUSIPCO*, A Coruña, Spain, 2019.
- [3] S. Grollmisch, D. Johnson, T. Krüger, and J. Liebetrau, “Plastic Material Classification using Neural Network based Audio Signal Analysis,” in *SMSI*, Nürnberg, Germany, 2020.
- [4] D. Johnson, J. Kirner, S. Grollmisch, and J. Liebetrau, “Compressed Air Leakage Detection Using Acoustic Emissions with Neural Networks,” in *InterNoise*, Seoul, South Korea, 2020.
- [5] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, “ToyADMOS: A Dataset of Miniature-Machine Operating Sounds for Anomalous Sound Detection,” in *WASPAA*, New Paltz, NY, USA, 2019.
- [6] H. Purohit, R. Tanabe, T. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, “Mimii dataset: Sound dataset for malfunctioning industrial machine investigation and inspection,” in *DCASE*, New York, USA, 2019.
- [7] E. Cano, J. Nowak, and S. Grollmisch, “Exploring sound source separation for acoustic condition monitoring in industrial scenarios,” in *EUSIPCO*, Kos island, Greece, 2017.
- [8] D. Johnson and S. Grollmisch, “Techniques Improving the Robustness of Deep Learning Models for Industrial Sound Analysis,” in *EUSIPCO*, Amsterdam, The Netherlands, 2020.