

## Entwicklung akustischer Messungen für industrielles maschinelles Lernen

Christian Fuchs<sup>1</sup>, Steffen Klein<sup>1</sup>, Stefan Saller<sup>2</sup>, Daniel Spies<sup>3</sup>, Andreas Schütze<sup>1</sup>, Tizian Schneider<sup>1</sup>

<sup>1</sup> Universität des Saarlandes, Lehrstuhl für Messtechnik, 66123 Saarbrücken, Mail: c.fuchs@lmt.uni-saarland.de

<sup>2</sup> Festo SE & Co. KG, Abteilung TN-MS, 73734 Esslingen, Mail: stefan.saller@festo.com

<sup>3</sup> Festo SE & Co. KG, Abteilung TN-AI, 73734 Esslingen, Mail: daniel.spies@festo.com

### Abstract

Einige Defekte an industriellen Maschinen können von Experten aufgrund von akustischen Ereignissen erkannt werden. Durch die bisherigen Erfolge in der Spracherkennung [1], [2], motiviert dies zur Nutzung maschinellen Lernens für industrielle akustische Messungen wie zum Beispiel bei der akustischen Leckageüberwachung pneumatischer Zylinder.

Da maschinelles Lernen die Messkette um statistische Datenauswertung erweitert [3], stellt es besondere Anforderungen an das Design of Experiment (DoE) und die Messkette. Bei akustischen Messungen spielen Noise und Clipping eine große Rolle. Es gilt also, einen geschickten Dynamikbereich zu finden und einen sinnvollen Nutzsinalbereich mit ausreichendem Signalrauschabstand (SNR) in den relevanten Messsignalbereichen zu suchen.

Als Beispiel werden Messungen an Antrieben von Festo angeführt, bei denen durch Manipulation der Dichtungen Leckagen erzeugt wurden. Antriebsleckagen können im Feld durch eine erhöhte und außerplanmäßige Beanspruchung im Rahmen des Betriebs auftreten.

Um physikalisches Clipping zu vermeiden, wurden robuste MEMS Mikrofone gewählt, sowie Gehäusedesigns mit konischer Einfräsung und der Option, akustische Dämmmaterialien zu integrieren, aufgebaut. Außerdem wurde das Gain auf einen Nutzsinalbereich optimiert und die Einzelmessungen wurden entsprechend segmentiert, um Datenhomogenität zu erreichen.

Zur Erfüllung der statistischen Anforderungen wurde das DoE per Latin Hypercube Sampling (LHS) [4] durchgeführt. Dazu wurden Daten mit verschiedenen Leckagestärken, beaufschlagten Drücken und Einstellungen der Drossel aufgezeichnet. Somit wurde maschinellen Lernen Robustheit zu unterschiedlichen Druck- und Drosselinstellungen sowie die Lernbarkeit der Zielgröße (Leckage) ermöglicht.

Rauschsignale wurden bewusst außer Acht gelassen, um das SNR maximal zu halten. Verschiedene Rauschsignale können als zusätzliche Einflussgrößen für Robustheitstests und weiterführende Vergleiche im Nachhinein eingefügt werden.

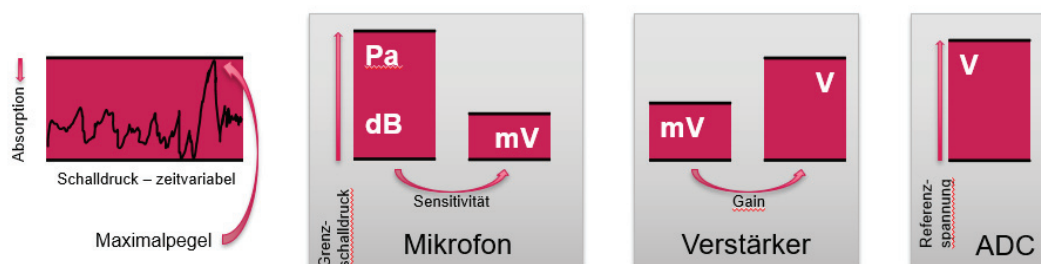
### Datenerfassung

Der Fokus dieses Beitrags liegt auf der Erfassung von Daten und Datensätzen zur Auswertung mit maschinellen Lernen. Häufig besteht die Annahme, dass Daten, die bisher klassisch ausgewertet wurden, gleichermaßen auch mit maschinellen Lernen ausgewertet werden können. Allerdings bestehen besondere Anforderungen an Datensätze für maschinelles Lernen.

Es ist weitläufig bekannt, dass die Menge der Daten, speziell auch die Menge von Fehlerdaten und deren Vollständigkeit im Sinne des DoE maßgeblich sind, um die zugehörigen Fehlerbilder im Vergleich zu „gesunden“ Daten einordnen zu können und die nötige Robustheit gegenüber Störeinflüssen wie Umgebungs- und Prozessparametern zu erreichen [5], [6].

Synchronität und konstante Datenlängen sind für maschinelles Lernen relevant [7]. Im Idealfall wird auf Synchronität und konstante Datenlänge bereits bei der Durchführung von Messungen geachtet. Alternativ kann auch die Möglichkeit sinnvoller Segmentierungen (z.B. durch Trigger-Signale) sichergestellt werden. Eine Kombination beider Ansätze ist ebenfalls umsetzbar und aufgrund verschiedener Messsituationen häufig notwendig.

Die Abschätzung des Erfolgs von maschinellen Lernen in Bezug auf Datenqualität ist nicht trivial. Im akustischen Bereich kann man im ersten Schritt die Beobachtung treffen, dass Experten häufig eine Zuordnung von akustischen Ereignissen zu Fehlerfällen vornehmen können. Im zweiten Schritt besteht die Frage, ob diese Ereignisse in den Messdaten ebenfalls erkennbar sind.



**Abbildung 1:** Analoge Messkette – Vereinfachtes Modell mit Fokus auf *Dynamik* und die Vermeidung von *Clipping*. Luftschall > Mikrofon > Verstärker > Analog-Digital-Wandler. Clipping kann am Mikrofon („Grenzschalldruck“) oder in den meisten Fällen am ADC („Referenzspannung“) auftreten und durch Absorption oder v.a. *Gain*-Settings verhindert werden.

## Datenqualität als Grundlage für ML

Die Zuordnung von akustischen Ereignissen zu Fehlerfällen kann in einigen Fällen von Experten vorgenommen werden. Voraussetzung für das maschinelle Lernen ist aber, dass diese Ereignisse in den Audiodaten abgebildet sind. Damit dies der Fall ist, ist eine hohe Datenqualität erforderlich [8].

Besonders wenn man maschinelles Lernen als Erweiterung der Messkette um statistische Mustererkennung betrachtet, wird nochmals deutlich, dass Verschleißsymptome klar gemessen werden müssen und sich Fehler, wie bei allen klassischen Messsystemen auch, im Verlauf der Messkette aufsummieren. Frühe Fehler in der Messkette (vgl. Abb. 1) können auch von intelligenten Algorithmen nicht korrigiert werden. Ein sorgfältiges DoE und die Datenerfassung unter den folgenden Gesichtspunkten sorgen für eine hohe Datenqualität.

### Mikrofonposition und Messumgebung

Um den Direktschall zu maximieren, sollten sich Mikrofone an der Quelle orientieren, also wird die Membran auf die Quelle gerichtet. Um Umgebungs- und Hallaufnahmen zu vermindern kann eine nach vorne gerichtete Charakteristik (z.B. Niere oder Superniere) verwendet werden. Nebengeräusche werden in der Messumgebung z.B. durch ein abgeschlossenes Labor und die Einhausung des Experiments minimiert.

Das Mikrofon befindet sich nah an der Quelle. Dennoch sollte es weit genug von der Quelle entfernt sein, sodass weder Überschreiten des Grenzschalldrucks noch der Nahbesprechungseffekt eintreten [9], [10]. Um den Direktschall aufzunehmen, sollten sich keine Medien außer Luft zwischen Mikrofon und Quelle befinden, die z.B. Absorption oder Reflexion verursachen würden.

Um Überschreiten des Grenzschalldrucks zu verhindern, können Absorptionsmaßnahmen in der direkten Umgebung des Mikrofons vorgenommen werden. All diese Maßnahmen wirken sich auf die Aussteuerung des Mikrofons aus.

### Design of Experiment

Um die bereits angesprochene Robustheit zu ermöglichen, sind Variationen aller relevanten Einflussgrößen des Messsignals zu erfassen. Zu allen Parametern und Daten wird eine ausführliche Dokumentation benötigt, um den Datensatz vollständig annotieren zu können. Einer der Einflussparameter wird dann als Zielgröße, auf welche trainiert werden kann, definiert.

Die Methode Latin Hypercube Sampling setzt sich zum Ziel eine möglichst homogene Verteilung und statistische Unabhängigkeit der variierten Prozessparameter bei wenigen Experimenten zu erreichen (vgl. Abb. 2). Systematische Fehler und zufällige Korrelationen bei der Datenaufnahme müssen vermieden werden, da sonst Ziel- und Störgrößen von dem von Natur aus statistischen maschinellen Lernen nicht getrennt werden können (Robustheit).

Dazu werden alle Einflussparameter in Abschnitte unterteilt und untereinander zufällig kombiniert. Jeder Parameterabschnitt wird nur einmal getestet.

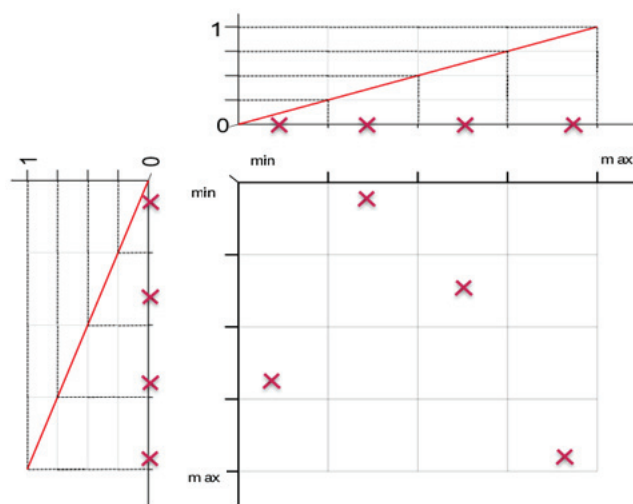


Abbildung 2: Design of Experiment: Latin Hypercube Sampling (LHS) mit zwei Parametern zur Erreichung von Robustheit bei minimaler Anzahl an Messungen.

### Dynamikbereich (Mikrofon, Verstärker und ADC)

Bei akustischen Messungen, speziell auch bei Luftschall, sind häufig Problemstellungen dynamischer Natur vorzufinden. Luftschalldruck ist ein zeitvariables Signal, das hohe dynamische Differenzen bzw. Peaks aufweisen kann. Zur Veranschaulichung der Messkette mit Fokus auf den dynamischen Signalverlauf stellt Abbildung 1 ein vereinfachtes Modell einer analogen Messkette dar, beginnend mit dem Luftschall selbst. Die Signalblöcke aus Abbildung 1 können um die tatsächlichen oberen und unteren Grenzen erweitert werden (vgl. Abb. 3):

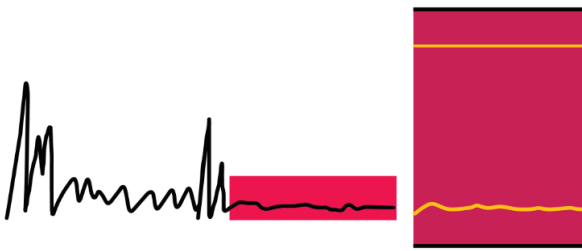


Abbildung 3: Signalblock mit dynamischer Obergrenze (Headroom) und Untergrenze (Grundrauschen).

Um das SNR des Signals über die gesamte Messkette maximal zu halten, ist das Ziel bei jedem Verarbeitungsschritt die obere Grenze auszunutzen, aber nie zu überschreiten (Clipping vermeiden) und alle relevanten Signalanteile ausreichend über die untere Grenze zu heben. Als Einstellmöglichkeit dient dabei der Verstärker.

Um die Verstärkung richtig einzustellen, wird der maximale Pegel des Messsignals inklusive sämtlicher potentieller Nebengeräusche erzeugt und eine Messung durchgeführt. Anhand dieser Messung wird der Dynamikbereich geprüft und die Verstärkung angepasst. Tritt Clipping auf, wird die Verstärkung verringert. Liegen relevante Signale nahe zum Grundrauschen, wird die Verstärkung erhöht. Clipping darf dabei für relevante Signale keinesfalls auftreten. Der Prozess wird so lange wiederholt, bis der Dynamikbereich maximal ausgenutzt, aber nicht überschritten ist.

Treten beide Fälle gemeinsam auf (vgl. Abb. 4), spricht potentiell Clipping relevanter Signale und das Versinken relevanter Signale im Grundrauschen, dann reicht der Dynamikbereich nicht aus und muss (z.B. durch Verstärkung) angepasst werden.

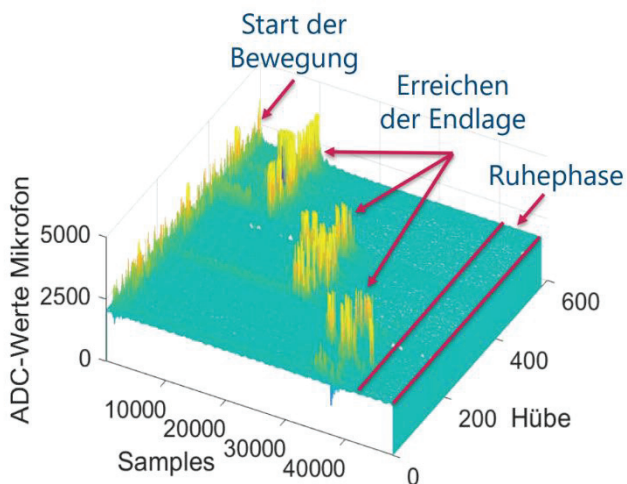


**Abbildung 4:** Skizze eines dynamischen Signals, dessen relevante Informationen nahe des Grundrauschens (z.B. eines ADCs) liegen.

Es stellt sich also zunächst die Frage, welche Signalanteile als relevant eingestuft werden und welche Signalanteile verworfen werden können. Im Beispiel aus Abbildung 4 wird angenommen, dass der Bereich mit der niedrigsten Aussteuerung die Information enthält.

Hier kommen die Leckagemessungen der Festo SE & Co. KG ins Spiel. Hierbei handelt es sich um einen Antrieb mit Kolbenstange, die einfährt und wieder ausfährt. An beiden Seiten des Antriebs befindet sich eine Leckage in der Dichtung. Diese soll jeweils auf der Seite erkannt werden, auf der sich die Kolbenstange nicht befindet, da in diesem Fall ein mit Druck beaufschlagtes Volumen besteht, aus dem Luft entweichen kann.

In Abbildung 5 sieht man einen Waterfall-Plot. Er zeigt mehrere Ausfahrten von mit unterschiedlichen Leckagen präparierten Zylindern. Die Daten sind über den Start der Bewegung synchronisiert und nach Drosselstellung sortiert. Dadurch sind Start und Ende der Bewegung als Peaks nebeneinander erkennbar. Danach folgt die Ruhephase ohne Bewegung.



**Abbildung 5:** Ausfahrende Kolbenstangenbewegungen mit drei unterschiedlichen Drosselstellungen und individuellen Leckagen im Zeitbereich. Eine größere Drosselöffnung führt zu kürzerer Fahrt.

Das Erreichen der Endlage erfolgt bei drei unterschiedlichen Drosselstellungen zu drei unterschiedlichen Zeitpunkten. Dadurch sind nur Daten mit gleicher Drosselstellung von gleicher Fahrtlänge. In der folgenden Ruhephase kann allerdings eine gleichmäßige Datenlänge gewählt werden.

Im Hinblick auf das Beispiel Pneumatikzylinder enthält die Bewegungsphase starke Peaks (Anfahren, Anschlag in die Endposition), wogegen die Ruhephase keine sichtlichen Peaks aufweist. Diese Bereiche können also als zwei unterschiedliche Dynamikbereiche betrachtet werden. Die quasistationäre Ruhephase hebt sich hier graphisch nicht vom Grundrauschen ab, dem Plot kann dort keine Information entnommen werden.

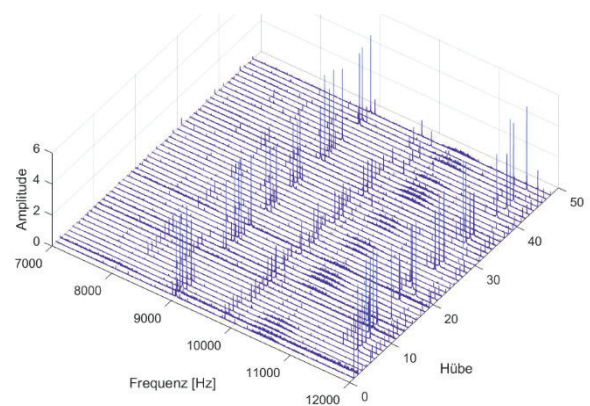
Nun stellt sich die Frage, in welcher Phase die wesentliche Information entnommen werden kann. Eine Leckage ist je nach Ausmaß und Form als leises Zischen hörbar während die Kolbenstange ruht. Während der Fahrt wird das Zischen dagegen akustisch von den Fahrtgeräuschen verdeckt. Dementsprechend liegt die Annahme nah, dass die eindeutigsten Information zur Leckage der Ruhephase zu entnehmen sind.

Die gesuchte Leckageinformation kann im Ruhebereich gehört werden. Die Verstärkung muss also so weit erhöht werden, dass das Signal in der Ruhephase den Dynamikbereich voll ausnutzt und die restlichen Signale im ADC ins Clipping laufen (vgl. Abb. 4). Dieses Clipping kann wiederum als Trigger zur Segmentierung der relevanten Signale verwendet werden.

Eine weitere Fragestellung ist die Bedeutung des abzudeckenden Messbereichs der Zielgröße. In Abbildung 5 wurden Leckagen von 0 l/h bis 25 l/h aufgenommen. Sollen auch größere Leckagen erkannt werden, würden sie ein stärkeres Signal in der Ruhephase liefern.

## Ergebnisse

Zu den Daten in Abbildung 5 wird im Folgenden der Frequenzbereich (7 kHz – 12 kHz) aufgetragen, in dem das Zischen der Leckagen, vergleichbar zu den menschlichen Lauten ‚S‘ oder ‚Sch‘ [11], erwartet wird.

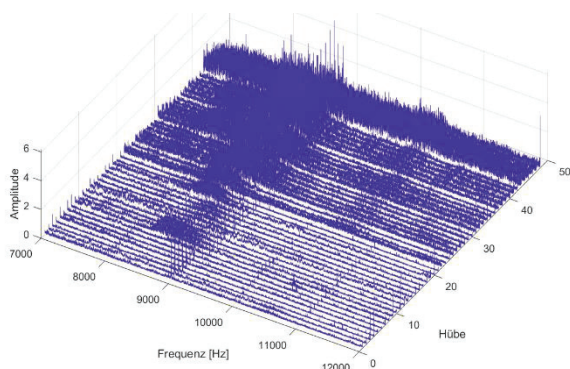


**Abbildung 6:** Ruhephase mit geringer Verstärkung (vgl. Abb. 5). Frequenzbereich (7 kHz – 12 kHz) nach Leckage sortiert. Zunehmende Leckagen (0 l/h bis 25 l/h) können in den Daten nicht erkannt werden.

Um die Signalqualität zu erhöhen wurden die Messungen aus Abbildung 6 mit folgenden Änderungen wiederholt:

- Ausmaß der Leckagen erhöht
- Verstärkung auf Ruhephase optimiert
- Reduzieren der Störpeaks im Frequenzbereich

Die Ergebnisse dieser wiederholten Messung sind in Abbildung 7 dargestellt.



**Abbildung 7:** Ruhephase mit angepasster Verstärkung. Frequenzbereich (7 kHz – 12 kHz) nach Leckage sortiert. Die Zunahme der Leckagen (0 l/h bis 45 l/h) kann in den Daten beobachtet werden.

In Abbildung 7 kann die Zunahme des Zischens anhand eines zunehmenden Rauschens im Frequenzbereich 7 kHz bis 12 kHz erkannt werden. Besonders auffällig ist eine Betonung des Rauschens zwischen 8 kHz und 9 kHz. Anhand dieser Daten besteht die Möglichkeit, das Leckageausmaß bereits mit einfachen Merkmalen, wie z.B. dem RMS in den genannten Frequenzbereichen, quantifizieren zu können.

## Zusammenfassung und Ausblick

Die verbesserte Qualität der Daten in Abbildung 7 im Vergleich zu Abbildung 6 lässt sich hauptsächlich auf die angepassten Dynamikbereiche zurückführen. Diese wurden durch Erweitern des Messsignalbereichs bzw. des Leckageausmaßes als auch durch die Optimierung des Verstärkungsfaktors für den relevanten Dynamikbereich erreicht.

Insgesamt tragen Synchronität, Datenlänge und saubere Annotation bzw. Dokumentation zu folgenden Kernpunkten bei der Analyse bei:

- Rohdaten effizient und übersichtlich einlesbar
- Hohe Verständlichkeit bei der Rohdatenbetrachtung
- Lauffähigkeit von bestimmten Algorithmen (Anforderungen an Datenstruktur)
- Zuverlässige Lernbasis für statistische Auswertung (KI-Methoden)
- Qualität und Darstellbarkeit der Auswertungen

Zusammenfassend lässt sich sagen, dass Mikrofone sich von anderen Sensoren unterscheiden. Der Dynamikbereich kann hier eine wesentliche Rolle spielen, da Luftschall hohe Dynamikbereiche abdeckt, manche Informationen aber in

kleinen Dynamikbereichen (z.B. quasistationäre Ruhephase) zu finden sind. Deshalb kann es helfen „die Lupe richtig einzustellen“.

Eine homogene Verteilung und statistische Unabhängigkeit mit möglichst geringer Anzahl an Messungen im Versuchsraum können per Latin Hypercube Sampling erreicht werden. Im Beispiel wurden die Einflussgrößen Leckage, Druck und Drosseleinstellung im Design of Experiment berücksichtigt.

Da der Aufwand für die Datensatzgenerierung sehr hoch ist, empfiehlt es sich, zunächst den Prozess tiefergehend zu verstehen und die Datenqualität zu maximieren. Danach kann ein vollständiges Design of Experiment mit den relevanten Einflussparametern erstellt und angewandt werden. Die Prozesskenntnis und der Austausch von Anwendungs- und Analysepartnern in Projekten sind wesentlich.

Die beschriebenen Vorgehensweisen sind bewusst generell gehalten, um sie auf sonstige Anwendung von maschinellem Lernen für akustische Messungen übertragbar zu machen. Die Erhöhung der Datenqualität, die dadurch erreicht werden kann, wurde am Beispiel der Leckageerkennung an Pneumatikzylindern gezeigt. Die Information der zu bestimmenden Leckage wurde im Frequenzbereich sichtbar.

## Literatur

- [1] Kinga, S. und Frankel, J.: Speech production knowledge in automatic speech recognition. *The Journal of the Acoustical Society of America* 121 (2007), 723
- [2] Benzeghiba, M.: Automatic speech recognition and speech variability: A review. *Speech Communication*, Volume 49 (2007), Issues 10–11, Pages 763-786, ISSN 0167-6393
- [3] Schneider, T. und Klein, S.: Machine learning in industrial measurement technology for detection of known and unknown faults of equipment and sensors. *tm - Technisches Messen*, vol. 86 (2019), no. 11, pp. 706–718
- [4] Loh, W.-L.: On Latin hypercube sampling. *The Annals of Statistics*, 24 (5) (1996), 2058–2080
- [5] Schneider, T.: Industrial condition monitoring with smart sensors using automated feature extraction and selection. *Measurement Science and Technology*, vol. 29 (2018), no. 9
- [6] Müller, R. und Blum, A.: Schlanke Daten für schlanke Prozesse/Lean Data for Lean Production. *wt Werkstattstechnik online*. 110 (2020), 639-643. 10.37544/1436-4980-2020-09-69
- [7] Dorst, T. und Robin, Y.: Influence of synchronization within a sensor network on machine learning results. *Journal of Sensors and Sensor Systems* (2021)
- [8] Schuh, G. und Nyhuis, P. und Hansen, J.: Produktionsdaten als Enabler für Industrie 4.0 [...]. Springer-VDI-Verl., Düsseldorf, (2015)
- [9] Schneider, M.: Reale Ambisonics-Mikrophone. *pub.dega-akustik.de*, Deutsche Jahrestagung für Akustik, München (2018)
- [10] Mariano, D.: Überblick über verschiedene Erklärungsansätze für das Zustandekommen des Nahbesprechungseffekts [...]. URL: <http://www.sengpielaudio.com/Nahbesprechungseffekt-Mariano.pdf> (2006)
- [11] Linhard, K.: Frequency Domain De-Essing for Hands-free Applications. *Deutsche Jahrestagung für Akustik*, Kiel (2017)