

# Verbesserung der Sprachverständlichkeit für Atemschutzmasken mittels Signalbearbeitung mit nichtlinearen Kennlinien

Bastian Gröger<sup>1</sup>, Michael Brodersen<sup>2,3</sup>, Gerhard Schmidt<sup>2</sup>

<sup>1</sup> E-Mail: groeger.bastian@gmx.de

<sup>2</sup> Christian-Albrechts-Universität zu Kiel, 24143 Kiel, E-Mail: mibr/bagr/gus@tf.uni-kiel.de

<sup>3</sup> Dräger Safety AG & Co. KGaA, 23560 Lübeck, E-Mail: michael.brodersen@draeger.com

## Einleitung

Bei Feuerwehreinsätzen ist die Verwendung einer Atemschutzmaske aus Sicherheitsgründen unabdingbar. Durch die geschlossene Atemschutzmaske wird die von der Einsatzkraft ausgehende Sprache aber auch stark gedämpft, wodurch die Verständigung unter den Feuerwehrleuten erschwert wird. Eine integrierbare Kommunikationseinheit soll einerseits die Sprachverständlichkeit verbessern und andererseits die Kommunikation per Funk sowie die Verstärkung des Gesprochenen über die Lautsprecher der Einheit ermöglichen. Das in der Kommunikationseinheit integrierte Mikrofon befindet sich dabei außerhalb des abgeschlossenen Systems der Atemschutzmaske hinter einer Sprechmembran mit einer Resonanzfrequenz von 800 Hz, was eine Dämpfung von essenziellen Sprachanteilen über 1400 Hz zur Folge hat und zudem das Signal nichtlinear beeinflusst. Folglich ist die Sprachverständlichkeit des Mikrofonsignals beeinträchtigt. Der in diesem Beitrag vorgestellte Exciter-Algorithmus soll dem Signal die gedämpften Harmonischen (Vielfache der Grundfrequenz) wieder aufprägen und so eine verbesserte Sprachverständlichkeit erwirken. Die durch den Exciter erzielte Verbesserung der Sprachverständlichkeit wird mittels eines CMOS-ähnlichen Tests (*Comparison Mean Opinion Score*) evaluiert.

## Eigenschaften von Atemschutzmasken

Atemschutzmasken schützen das Gesicht und bieten Schutz vor toxischen Gasen und Rauch (siehe [1]). Die Maske umschließt das Gesicht luftdicht, Nase und Mund werden dabei von einer Innenmaske bedeckt, um das Be-



Abbildung 1: Atemschutzmaske auf einem Kunstkopf.

schlagen des Visiers beim Ausatmen zu verhindern (siehe Abb.1). Der Raum vor Mund und Nase wird mit frischer Luft aus einer auf dem Rücken getragenen Flasche versorgt. Durch die Abdichtung der Atemschutzmaske am Gesicht, wird die Sprache stark gedämpft. Um diese Dämpfung zu minimieren, ist vor dem Mund eine Sprechmembran angeordnet. Die Membran kann aus Metall oder Kunststoff bestehen und weist, aufgrund ihrer Eigenschaften, eine Resonanz bei 800 Hz auf. Die Auswirkungen der Sprechmembran auf die Sprachcharakteristik sind in Abb.3 anhand von zwei Spektrogrammen dargestellt. Das Signal von Spektrogramm (a) ist dabei mit einem Mikrofon ohne Atemschutzmaske in einer störungsfreien Umgebung aufgenommen worden. Das Signal von Spektrogramm (b) ist über das Mikrofon der Kommunikationseinheit aufgenommen worden. Beide Signale stammen vom selben Sprecher, der jeweils den gleichen Satz gesprochen hat. Im Vergleich der Spektrogramme ist erkennbar, dass die Maskenumgebung die Frequenzbereiche unter 500 Hz und über 1400 Hz dämpft. Der Hauptsprachanteil, der für die Verständlichkeit der Sprache maßgebend ist, liegt im Frequenzbereich von 1 kHz bis 4 kHz. Eine Beeinträchtigung dieses Bereichs geht mit einer deutlichen Beeinträchtigung der Sprachqualität einher [2]. Um das Fehlen der Spektrallinien oberhalb von 1400 Hz zu kompensieren, wird der Exciter angewandt, wodurch Harmonische im hochfrequenten Frequenzbereich entstehen sollen.

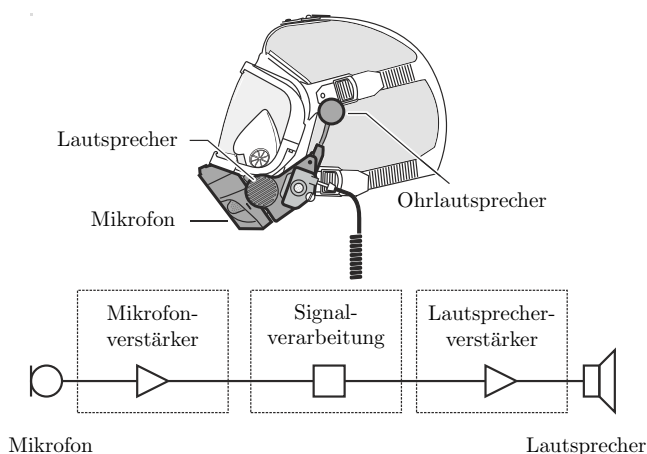


Abbildung 2: Struktureller Überblick der Kommunikationseinheit (nach [1]).

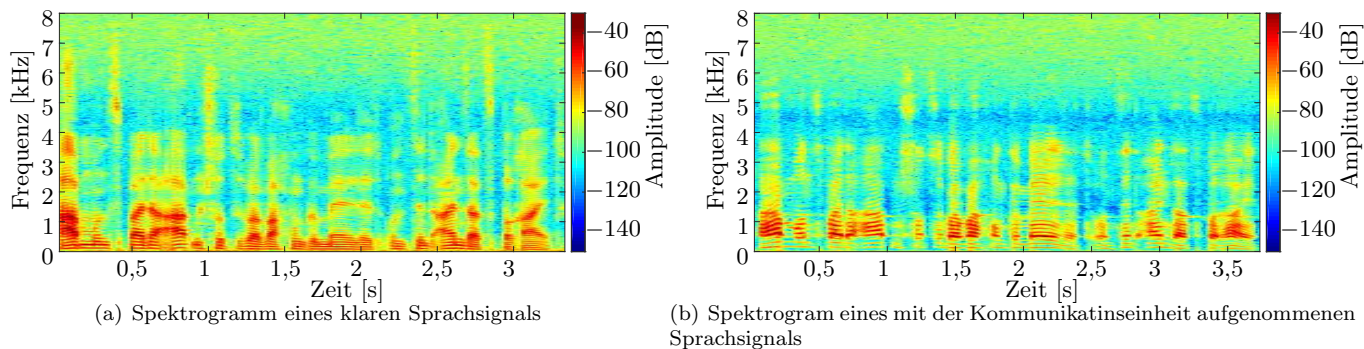


Abbildung 3: Vergleich von zwei Spektrogrammen; (a) klares Sprachsignal; (b) mit dem Mikrofon der Kommunikationseinheit aufgenommenes Sprachsignal.

### Kommunikationseinheiten für Atemschutzmasken

Um die Kommunikation zu verbessern, werden Kommunikationseinheiten verwendet, welche die Sprache mit einem Mikrofon vor der Sprechmembran aufzeichnen, digital verarbeiten und anschließend verstärkt über die Außenlautsprecher der Maske ausgeben, wodurch die Kommunikation im direkten Umfeld erleichtert wird. Das Mikrofonsignal kann auch über lange Distanzen mit einem externen Funkgerät (Taktischer Funk) oder über kürzere Distanzen mit dem Funksystem der Kommunikationseinheit (Team Funk) übertragen werden. Eingehende Funksignale können über die Ohrlautsprecher der Kommunikationseinheit abgespielt werden. Ein struktureller Überblick der Kommunikationseinheit wird in Abb. 2 gezeigt.

### Der Exciter

Die Erzeugung von Harmonischen ist in der Signalverarbeitung auch als Verzerrung bekannt und wird eher als ungewollter Nebeneffekt angesehen. Dabei gibt es auch Gegenbeispiele, bei denen eine Verzerrung des Signals als durchaus positiv wahrgenommen wird. Ein Röhrenverstärker kann zum Beispiel geradzahlige Harmonische erzeugen, die oft als angenehm und warm empfunden werden [3]. Ein ähnliches Ziel verfolgen Exciter [3]. Sie erzeugen Harmonische, die oftmals im Originalsignal kaum oder gar nicht vorhanden waren. Die Erzeugung der Harmonischen basiert auf der Verwendung von nichtlinearen Kennlinien. In Abbildung 4 ist die Übersicht des Exciter-Signalfussgraphen zu sehen. In dem Verzerrerpfad werden die Harmonischen des Eingangssignals  $x(n)$  erzeugt und mit dem linearen Faktor  $g_h$  gewichtet, sodass  $x_h(n)$  entsteht. Im Verzögerungspfad

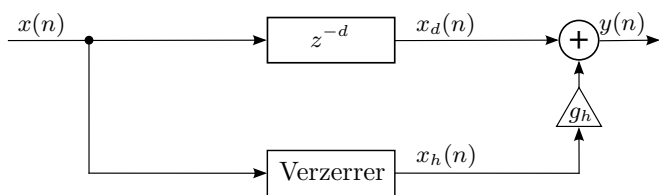


Abbildung 4: Übersicht des Exciter-Signalfussgraphen (nach [3]).

wird  $x(n)$  um  $d$  Abtastwerte verzögert, sodass die Laufzeit des Verzerrerpfad über das verzögerte Signal  $x_d(n)$  ausgeglichen wird. Zuletzt werden beide Pfade addiert, sodass das Ausgangssignal

$$y(n) = x_d(n) + g_h \cdot x_h(n) \tag{1}$$

entsteht. Der Verzerrerpfad besteht, wie in Abbildung 5 zu sehen ist, im Wesentlichen aus der Verzerrer-Kennlinie  $K_g$  für die Erzeugung der gerade Harmonischen und der Verzerrer-Kennlinie  $K_u$  für die Erzeugung der ungeraden Harmonischen. Den beiden Kennlinien ist ein Filter  $H_{ex}(z)$  vorgelagert, das bestimmt, welche Frequenzanteile von  $x(n)$  zur Erzeugung der Harmonischen genutzt werden,  $x_{ex}(n)$  ist dann der Eingang der nichtlinearen Kennlinien. Die Verzerrerausgänge  $x_g(n)$  und  $x_u(n)$  werden durch die Faktoren  $\tau$  bzw.  $1 - \tau$  gewichtet, wodurch die Klangfarbe (*timbre*) bestimmt wird. Der Faktor  $\tau$  kann im reellen Wertebereich zwischen 0 und 1 liegen und bestimmt welchen prozentualen Anteil die geraden bzw. ungeraden Harmonischen bei der Erzeugung von  $x_h(n)$  haben sollen. Für den Ausgang ergibt sich also

$$y(n) = x_d(n) + g_h \cdot (\tau \cdot x_g(n) + (1 - \tau) \cdot x_u(n)) \tag{2}$$

Im Folgenden werden zwei Exciter-Konfigurationen präsentiert, die unterschiedliche Frequenzbereiche ansprechen.

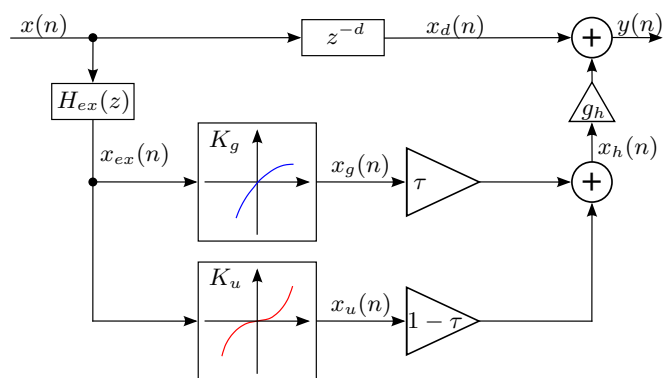
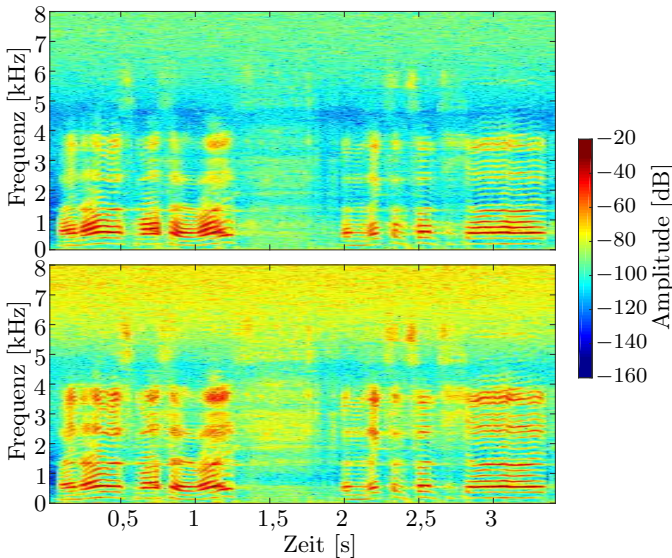


Abbildung 5: Vollständiger Exciter-Signalfussgraph (nach [3]).

### Der Höhen-Exciter

Der Höhen-Exciter wurde mit dem Ziel implementiert, Harmonische in den Höhen ( $f \geq 2$  kHz) des Eingangssignals zu erzeugen, die den subjektiven Höreindruck



**Abbildung 6:** Signal nach Bearbeitung mit dem Höhen-Exciter mit  $\tau = 0,8$  und  $g_h = 5$ ; Eingangssignal  $x$  (oberer Plot); Signal nach Bearbeitung mit dem Höhen-Exciter  $y$  (unterer Plot).

verbessern. Wie einleitend bereits erwähnt wurde, wird ein Großteil der Höhen durch die Nutzung der Maske gedämpft - der Höhen-Exciter soll diesem Verhalten entgegenwirken.

### Vorfilter

Damit die Harmonischen aus den Höhen des Eingangssignals entstehen, wurde das Filter  $H_{ex}(z)$  als digitaler Butterworth-Hochpass zweiter Ordnung entworfen. Die 3dB-Grenzfrequenz dieses stabilen IIR-Filters ist mit  $f_{g,HE} = 1,6\text{kHz}$  gegeben. Das Filter hat eine maximale, aufgerundete Gruppenlaufzeit von 3 Abtastwerten ( $d = 3$ ), was bei einer Abtastrate von 16 kHz zu einer sehr geringen Signalverzögerung von  $\frac{3}{16\text{kHz}} = 187,5\ \mu\text{s}$  führt.

### Nichtlineare Kennlinien

Für die Erzeugung der geraden Harmonischen wurde die Kennlinie

$$K_g : x_g(n) = -0,5 \cdot x_{ex}(n)^2 + x_{ex}(n) \quad (3)$$

verwendet, für die Erzeugung der ungeraden Harmonischen die Kennlinie

$$K_u : x_u(n) = |x_{ex}(n)| \cdot x_{ex}(n) . \quad (4)$$

Der Parameter  $\tau$  wurde, nach eigenem subjektiven Hörempfinden, auf  $\tau = 0,8$  festgelegt. Die optimale Dimensionierung des Parameters  $g_h$  scheint, nach eigenem subjektivem Empfinden, den Höreffekt sehr stark zu beeinflussen.

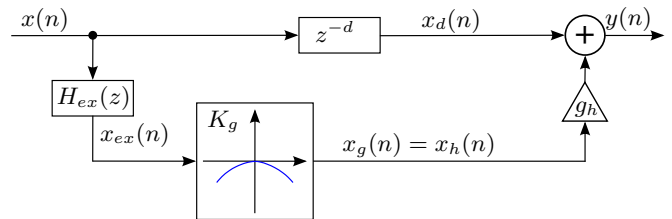
### Aufgeprägte Frequenzanteile

In Abbildung 6 ist ersichtlich, welche Auswirkungen der Höhen-Exciter auf den Spektrogramm eines Sprachsignals hat. Das obere Spektrogramm zeigt das Eingangssignal  $x(n)$ , das untere zeigt das Ausgangssignal  $y(n)$  des Höhen-Exciters, wobei  $\tau = 0,8$  und  $g_h = 5$  gewählt wurden. In den Spektrogrammen sind die Harmonischen an den roten Ausprägungen zu erkennen. Es ist deutlich zu erkennen, dass der Höhen-Exciter die Harmonischen ab einer Frequenz von ca. 1,4 kHz verstärkt. Im Bereich

um 1 kHz sinkt die Amplitude teilweise, durch die Verwendung des Höhen-Exciters. Die Geräuschanteile über 5 kHz sollten mittels Geräuschreduktion entfernt werden [4].

### Der Mitten-Exciter

Der Mitten-Exciter wurde mit dem Ziel entworfen, Harmonische im Bereich von 1,4 kHz bis 2 kHz zu erzeugen, die in den Aufnahmen mit der Kommunikationseinheit kaum vorhanden sind. Der Signalfussgraph des Mitten-Exciters aus Abb. 7 ähnelt dem des Höhen-Exciters mit dem Unterschied, dass nur eine nichtlineare Kennlinie verwendet wird. Der Timbre-Faktor  $\tau$  wird hierbei nicht mehr benötigt, da die Harmonischen nur durch eine nichtlineare Kennlinie entstehen. Damit gilt auch  $x_g(n) = x_h(n)$ .



**Abbildung 7:** Signalflussgraph des Mitten-Exciters.

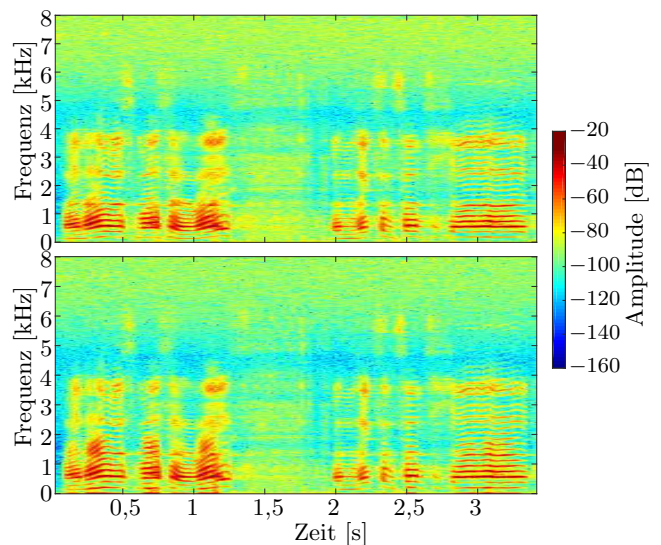
### Vorfilter

Damit entsprechende Harmonische im Bereich von 1,4 kHz bis 2 kHz entstehen, ist das Filter  $H_{ex}(z)$  als digitaler Butterworth-Bandpass vierter Ordnung entworfen worden. Die 3dB-Grenzfrequenzen dieses stabilen IIR-Filters sind mit  $f_{g1,ME} = 650\text{Hz}$  und  $f_{g2,ME} = 1050\text{Hz}$  gegeben. Das Filter hat eine maximale, aufgerundete Gruppenlaufzeit von 25 Abtastwerten ( $d = 25$ ), was bei einer Abtastrate von 16 kHz zu einer geringen Signalverzögerung von  $\frac{25}{16\text{kHz}} = 1,6\text{ms}$  führt.

### Nichtlineare Kennlinie

Für die Erzeugung der Harmonischen wurde folgende Kennlinie verwendet:

$$K_a : x_a(n) = -0,5 \cdot x_{ex}(n)^2 . \quad (5)$$



**Abbildung 8:** Signal nach Bearbeitung mit dem Mitten-Exciter mit  $g_h = 7,5$ ; Originalsignal  $x$  (oberer Plot); Signal nach Bearbeitung mit dem Mitten-Exciter  $y$  (unterer Plot).

### Aufgeprägte Frequenzanteile

Abbildung 8 zeigt die Auswirkungen des Mitten-Exciters auf das Spektrogramm eines Sprachsignals. Das obere Spektrogramm zeigt das Eingangssignal  $x$ , das untere Spektrogramm zeigt das Ausgangssignal  $y$  des Mitten-Exciters. Es wurde das gleiche Eingangssignal wie beim Höhen-Exciter verwendet. Der skalare Verstärkungsfaktor der Harmonischen ist hierbei  $g_h = 7,5$ . Beim Vergleich der Spektrogramme ist ersichtlich, dass vor allem im Bereich von ca. 1,5 kHz bis 2 kHz Harmonische erzeugt werden. Auch im Bereich unter 500 Hz werden Harmonische erzeugt, diese sind in Kombination mit der Kommunikationseinheit allerdings unbedeutend, da die Lautsprecher in diesem Frequenzbereich keinerlei Töne wiedergeben.

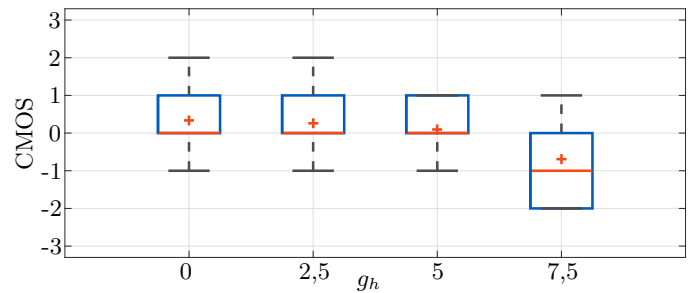
### Evaluierung mittels CMOS-Tests

Um die erläuterten Exciter-Algorithmen zu parametrieren und zu evaluieren, werden an CMOS-Tests (*Comparison Mean Opinion Score*) angelehnte Tests durchgeführt [5]. Hier bewertet eine Testperson zwei Sprachsequenzen S1 und S2 gleichzeitig, indem sie die subjektiv empfundene Sprachverständlichkeit bei unterschiedlichen linearen Verstärkungsfaktoren  $g_h$  vergleicht. Die Bewertungspunkte aus Tabelle 1 werden dabei der Sequenz S1 zugeteilt, Sequenz S2 erhält jeweils die gegenteilige Bewertung. Abb. 9 zeigt die Ergebnisse der CMOS-ähnlichen Tests, die aus der Gegenüberstellung der linearen Verstärkungsfaktoren  $g_h$  entstanden sind, im Zusammenhang mit dem Höhen-Exciter und dem Mitten-Exciter bei einem Bewertungsumfang von 10 Testpersonen. Die Ergebnisse sind als sog. Boxplots dargestellt. Die schwarzen Begrenzungen stellen den oberen und unteren Whisker dar, 90% bzw. 10% der Ergebnisse sind kleiner oder gleich dieser Grenze. Die blauen Begrenzungen zeigen das obere und untere Quartil, 75% bzw. 25% der Ergebnisse sind kleiner oder gleich dieser Grenze. Die rote Begrenzung zeigt den Median und das rote Kreuz den Mittelwert. Aus Abb. 9 (a) ist zu erkennen, dass die Testpersonen im Mittel den Höhen-Exciter mit dem Verstärkungspegel  $g_h = 5$  bevorzugen, das schlechteste Ergebnis erzielt das Originalsignal ( $g_h = 0$ ). Aus Abb. 9 (b) ist zu erkennen, dass die Testpersonen im Mittel den Mitten-Exciter mit dem Verstärkungspegel

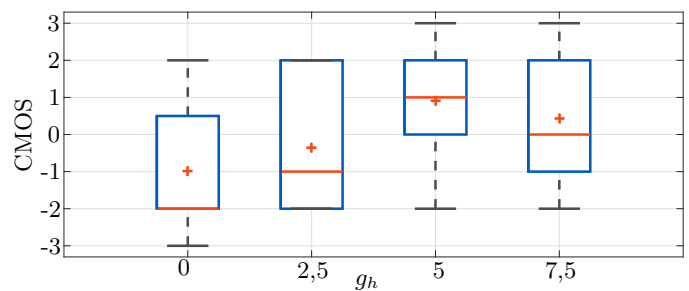
Im Vergleich der Sprachverständlichkeit erscheint/erscheinen	CMOS
S1 viel besser als S2	3
S1 besser als S2	2
S1 etwas besser als S2	1
beide Sequenzen gleich	0
S2 etwas besser als S1	-1
S2 besser als S1	-2
S2 etwas besser als S1	-3

**Tabelle 1:** CMOS-Skala bei der Bewertung der Sprachverständlichkeit (nach [5]).

$g_h = 0$  und  $g_h = 2,5$  bevorzugen. Demnach folgt, dass der Mitten-Exciter keine eindeutige Verbesserung der Sprachverständlichkeit erzeugen kann.



(a) CMOS-Testergebnisse des Mitten-Exciters



(b) CMOS-Testergebnisse des Höhen-Exciters

**Abbildung 9:** Darstellung der CMOS-Testergebnisse im Zusammenhang mit dem Höhen-Exciter in (a) und dem Mitten-Exciter in (b) auf der Bewertungsgrundlage von 10 Testpersonen bei der Bewertung von 4 Sprechern; bewertet wurde die Sprachverständlichkeit.

### Zusammenfassung

Durch zwei Exciter sollten die hohen Frequenzanteile des Mikrofonsignals der Kommunikationseinheit wiederhergestellt werden, die durch das abgeschlossene System der Atemschutzmaske und die Sprechmembran gedämpft werden. Dabei werden nichtlineare Kennlinien auf das gefilterte Signal angewendet, die Harmonische im Signal erzeugen. Der Höhen-Exciter erzeugt Harmonische über 2 kHz und verbessert, nach den subjektiven CMOS-Testergebnissen, die Sprachverständlichkeit. Der Mitten-Exciter erzeugt Harmonische von 1,4 kHz bis 2 kHz, bewirkt, nach den subjektiven Testergebnissen, allerdings keine Verbesserung der Sprachverständlichkeit.

### Literatur

- [1] Dr.-Ing. A. Volmer, Dr.-Ing. M. Romba, C. Schmidt und M. Housseem Harbi: Optimization of Speech Intelligibility for Fire Fighters' Full Face Masks, DAGA, 2013
- [2] A. Friesecke: Die Audio-Enzyklopädie. Ein Nachschlagewerk für Tontechniker, K. G. Saur Verlag, 2007
- [3] P. Bulling, K. Linhard, A. Wolf, G. Schmidt, A. Theiss, M. Grimm: Nichtlineare Kennlinien zur Verbesserung der Sprachverständlichkeit in geräuschbehafteter Umgebung, DAGA, 2016
- [4] V. K. Rajan, C. Baasch, M. Krini, G. Schmidt: Improvement in Listener Comfort Through Noise Shaping Using a Modified Wiener Filter Approach, Proc. ITG, 2014
- [5] International Telecommunication Union (Hrsg.): SERIES P: TELEPHONE TRANSMISSION QUALITY. Methods for objective and subjective assessment of quality, G.800, International Telecommunication Union, 1996