

Sprachsignalverzerrung für einen Kunstkopf mit zeitveränderlichem Resonanzvolumen

Tobias Hübschen¹, Alexander Wilbrandt¹, Michael Brodersen^{1,2}, Gerhard Schmidt¹

¹ *Digitale Signalverarbeitung und Systemtheorie, 24143 Kiel, Deutschland, Email: thu@tf.uni-kiel.de*

² *Draeger Safety AG & Co. KGaA, 23560 Lübeck, Deutschland, Email: michael.brodersen@draeger.com*

Kurzfassung

Bei der anwendungsnahen Evaluierung von Sprachkommunikationssystemen ist es üblich einen Kunstkopf zur Wiedergabe des Quellsignals zu verwenden. Solch ein Kunstkopf ist klassischerweise statisch, das heißt der Mund bewegt sich nicht und auch eine Atmung wird vom Kunstkopf nicht simuliert. Für Anwendungen wie die Evaluierung der Kommunikationseinheiten von Atemschutzmasken ist es jedoch unabdinglich den Kunstkopf zusätzlich atmen zu lassen, da ein signifikanter Teil der in Atemschutzmasken implementierten Signalverarbeitung speziell darauf ausgerichtet ist, natürliche und durch den Luftstrom in der Maske hervorgerufene Atemgeräusche zu unterdrücken. Wird die Atmung durch den Anschluss eines Blasebalges an einen modifizierten Kunstkopf simuliert, so ist allerdings zu beobachten, dass klassische Ansätze zur Verzerrung des Sprachsignals nur suboptimale Ergebnisse liefern. Dies ist mitunter durch die zeitveränderliche Anpassung des Resonanzraumes zu begründen. Für diesen konkreten Fall wird in diesem Beitrag ein neues zeitvariantes Entzerrverfahren als Alternative gegenüber konventionellen zeitinvarianten Verfahren vorgestellt.

Einleitung

Zuletzt konnte der Deutsche Feuerwehrverband jährlich bis zu 180.000 Brandeinsätze der deutschen Feuerwehren verzeichnen [1]. Solche Brandeinsätze sind für das jeweilige Einsatzteam besonders kritisch, da Feuerwehrfrauen und -männer sich hier oft selbst den zu bekämpfenden Gefahren aussetzen müssen. Ein erfolgreicher Einsatz ist so in der Regel nur durch entsprechende Schutzausrüstung möglich. Fester Bestandteil dieser Ausrüstung sind Atemschutzmasken, welche zwar vor giftigen Dämpfen schützen und ein nahezu normales Atmen ermöglichen, gleichzeitig aber die Kommunikationsfähigkeit mit der Umwelt signifikant einschränken. Aus diesem Grund haben sich inzwischen Atemschutzmasken auf dem Ausrüstungsmarkt etabliert, welche mit einem passiven oder aktiven Kommunikationsmodul ausgestattet sind.

Folglich lässt sich solch eine Maske mit Kommunikationsmodul als generisches Kommunikationssystem verstehen, welches bestimmte qualitative Anforderungen erfüllen muss. Wie auch bei anderen Kommunikationssystemen besteht hier somit die Notwendigkeit die Qualität, und insbesondere die Verständlichkeit, der übertragenen Sprache zu evaluieren. Als Aufbau zur reproduzierbaren Gewinnung der übertragenen, bzw. im Allgemeinen degradierten, Signale bietet es sich an, die Maske einem Kunstkopf aufzusetzen und den Mundlautsprecher des Kopfes zur Einspeisung realistischer Sprachsignale zu verwenden. Während die Verwen-

dung eines klassischen Kunstkopfes für viele Kommunikationssysteme genügt, bleibt so im beschriebenen Anwendungsfall jedoch ein besonderer Effekt unberücksichtigt: die Atmung. Denn durch den Aufbau und das Prinzip von Vollschutzmasken ist es unausweichlich, dass Atemgeräusche in Mikrofone einkoppeln oder der Luftstrom selbst an Mikrofonen Störungen erzeugt. Folglich muss für den beschriebenen Anwendungsfall ein „atmender“ Kunstkopf entwickelt werden.

In diesem Beitrag wird der Prototyp eines atmenden Kunstkopfes vorgestellt und insbesondere die Notwendigkeit und Umsetzung einer zeitvarianten Sprachsignalverzerrung thematisiert. Hierfür wird zunächst beispielhaft eine Atemschutzmaske mit Kommunikationsmodul vorgestellt, bevor der grundsätzliche Aufbau und die Funktionen des aktuellen Kunstkopfprototypen beschrieben werden. Anhand von Messungen mit diesem Prototypen wird anschließend die Notwendigkeit einer zeitvarianten Verzerrung aufgezeigt, sodass dieser Beitrag abschließend eine Methode zur Umsetzung solch einer zeitvarianten Verzerrung vorstellt.

Feuerwehrmasken

Feuerwehrmasken (Abb. 1 links) werden frontal auf das Gesicht gesetzt und bilden so eine Barriere für Luft und Schall. Um den Träger mit Frischluft zu versorgen, wird an die Maske ein Kompressor mit Sauerstoffflasche angeschlossen, der im Einsatz für die unverkennbaren Atemgeräusche verantwortlich ist. Die Dämpfung der Sprache durch die Maske in Kombination mit Kompressorgeräuschen sorgt folglich für die schwierige Ausgangslage bezüglich der Kommunikation des Maskenträgers mit Außenstehenden. Aus diesem Grund haben einige Maskenhersteller Kommunikationseinheiten (Abb. 1 rechts) in ihre Masken integriert. Diese Einheiten bestehen beispielsweise aus einem Mikrofon im Inneren, einem Lautsprecher an der Außenseite und einer Signalverarbeitungseinheit. Aufgabe der Signalverarbeitung ist es hauptsächlich Sprache zu verstärken, Atemgeräusche zu unterdrücken und allgemein die Verständlichkeit des Gesagten zu erhöhen. Genau diese Funktionalitäten müssen auf ihre Robustheit geprüft werden, weshalb es unabdingbar ist für



Abbildung 1: Beispiel für eine Atemschutzmaske (links) mit zugehöriger Kommunikationseinheit (rechts).



Abbildung 2: Blasebalg mit Kunstkopf (links), Detailansicht Blasebalg (mitte) und entkoppelter Kunstkopf mit Zuführungsschlauch (rechts).

Funktionstests den Einsatz der Maske möglichst realitätsnah auf einem „atmenden“ Kunstkopf zu simulieren.

Atmender Kunstkopf

Ein atmungsfähiger Kunstkopf sollte grundsätzlich aus zwei Komponenten bestehen: zum einen aus einem Kunstkopf, der mit einer Luftzuführung (Luftröhre) versehen ist, zum anderen aus einer Luftstrom erzeugenden Einheit (Pumpe). Abb. 2 zeigt solch einen Aufbau mit einem 3D-gedruckten Kunstkopf, der auf einem Blasebalg montiert ist. Da dieser Blasebalg im Betrieb eine akustische Störquelle ist, die jede Messung mit dem Kunstkopf verfälschen würde, wurden Kunstkopf und Blasebalg allerdings für den Betrieb räumlich getrennt und mit einem Zuleitungsschlauch verbunden. Der resultierende Druckabfall an der Mundöffnung wurde mittels eines Anemometers als nicht signifikant bewertet.

Die Ausprägung der Atmung wird durch eine entsprechende Ansteuerung des Blasebalges über eine serielle Schnittstelle festgelegt. Hier sind beispielsweise Parameter wie Atemfrequenz und Atemvolumen wichtig. Da ein Kunstkopf im Allgemeinen einen menschlichen Sprecher simulieren soll, ist es unabdingbar das Atemverhalten an die ausgegebene Sprache anzupassen bzw. mit dieser zu koppeln. So lässt sich für den Kontext der Atemschutzmasken motivieren, dass der Kunstkopf während des Sprechens leicht ausatmen und in kurzen Sprachpausen einatmen soll. In längeren Sprachpausen kann die Atmung unabhängig von der Sprache betrachtet werden, sodass hier Ruhe- bzw. durch das Aktivitätslevels der simulierten Person bedingte Stressatmung mit einer festen Atemfrequenz zutrage kommen kann. Eine grafische Darstellung solch einer Kopplung ist in Abb. 3 gegeben. Der Atmungsverlauf, ausgedrückt durch die Pumpenposition des Blasebalges, wurde offline mittels einer Sprachaktivitätsdetektion und angegebener Atmungsparameter berechnet und kann zur Laufzeit synchron zur Sprachausgabe nachgefahren werden.

Damit der atmende Kunstkopf zuverlässig in der Messtechnik verwendet werden kann, muss das Übertragungsverhalten des Kunstkopfsystems analysiert, und anschließend auch kompensiert, werden. Da das System durch die Bewegung der Pumpe nicht, wie üblicherweise, statisch sondern dynamisch ist, muss zunächst die Auswirkung dieser Systemdynamik bestimmt werden. Hierzu wurden Leistungsdichtespektren am Mundreferenzpunkt in Abhängigkeit der statischen Pumpenposition gemessen und verglichen. Abb. 4 zeigt die Abweichung der Leistungsdich-

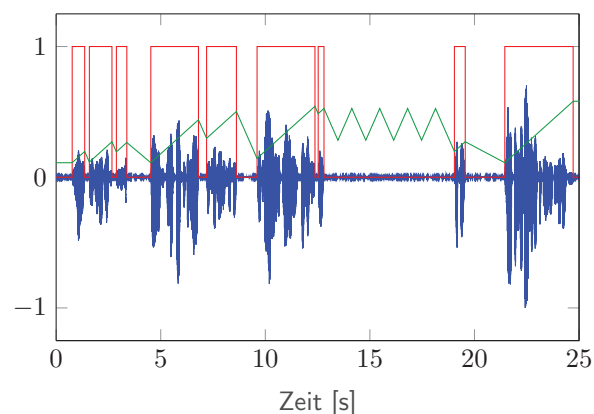


Abbildung 3: Beispiel für Kopplung von Sprache und Atmung. Signalverläufe von Sprache (—), Sprachdetektion (—) und Pumpenposition (—).

tespektren bei voller und leerer Pumpe. Es wird deutlich, dass bei Frequenzen bis 100 Hz das Übertragungsverhalten des Systems signifikant durch die Pumpenposition beeinflusst wird. So sind bei messtechnikrelevanten Frequenzen um 50 Hz Abweichungen bis zu 3 dB zu beobachten, die durch die unterschiedlichen Resonanzvolumina zu begründen sind. Hier lässt sich zusätzlich vermuten, dass sich dieser Effekt bei einem größeren Blasebalg weiter verstärken würde. Aus dieser Beobachtung folgt, dass bei einer Kompensation des Übertragungsverhaltens (Entzerrung) die Systemdynamik, bzw. die Zeitvarianz des Systems, berücksichtigt werden muss.

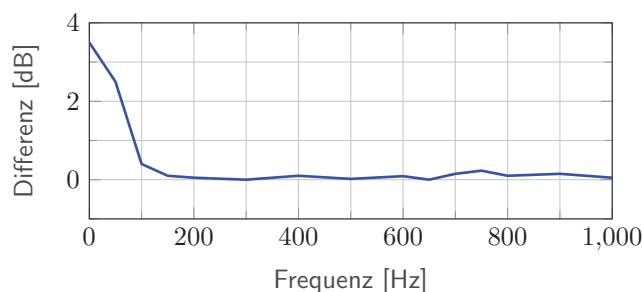


Abbildung 4: Abweichung zwischen den Leistungsdichtespektren (gleiche Anregung ohne Entzerrung) am Mundreferenzpunkt bei vollem und leerem Blasebalg im Bereich 0 – 1000 Hz.

Entzerrung

Bei einer Signal- bzw. Systementzerrung wird ein auszugebendes Signal $x(n)$ vorgefiltert. Diese Filterung kann als

$$x_{\text{eq}}(n) = \mathbf{p}^T \cdot \mathbf{x}(n) \quad (1)$$

geschrieben werden, wobei die Definition $\mathbf{x}(n) = [x(n), \dots, x(n - P + 1)]^T$ gilt. Der Vektor \mathbf{p} enthält somit die Koeffizienten eines FIR-Filters der Länge P . Dieses Filter wird Entzerrfilter genannt und für jede Anwendung so entworfen, dass das an einem Referenzpunkt im Raum aufgenommenen Signal $y(n)$ festgelegte spektrale Eigenschaften aufweist. Im vorigen Abschnitt wurde gezeigt, dass für die beschriebene Anwendung die spektralen Eigenschaften des Signals $y(n)$ von der Pumpenposition $s(n) \in [0, 1]$ des

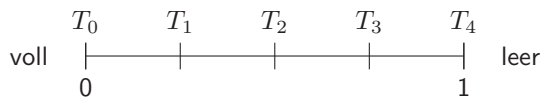


Abbildung 5: Aufteilung des Pumpvolumens in $N = 4$ Intervalle. T_0 entspricht einem vollen und T_4 einem entleerten Blasebalg.

Blasebalges abhängig sind. Somit sollte auch das angewendete Entzerrfilter von der Pumpposition, und folglich von der Zeit, abhängig sein.

Um solch eine Abhängigkeit herzustellen kann das Pumpvolumen zunächst in N Pumpintervalle aufgeteilt und deren Grenzen mit $N + 1$ äquidistanten Pumppositionen $T_\lambda \in [0, 1]$ beschrieben werden. Der ganzzahlige Index λ zählt dabei die Intervallgrenzen mit aufsteigender Position von 0 bis N durch. Solch eine Aufteilung ist in Abb. 5 beispielhaft für $N = 4$ aufgetragen.

Für jede der äquidistanten Pumppositionen kann dann ein eigenes Entzerrfilter \mathbf{p}_λ entworfen werden. Hierfür muss zunächst

$$S_{\Delta,\lambda}(\mu) = \frac{S_{yy}(\mu|T_\lambda)}{S_{xx}(\mu)} \quad (2)$$

als Quotient der Leistungsdichtespektren [2] von $x(n)$ sowie von $y(n)$ bedingt durch Pumpposition T_λ und in Abhängigkeit der Frequenzstützstelle μ berechnet werden. Dieser Quotient wird anschließend nach

$$\bar{S}_{\Delta,\lambda}(\mu) = (1 - \beta(\mu))\bar{S}_{\Delta,\lambda}(\mu - 1) + \beta(\mu)\bar{S}_{\Delta,\lambda}(\mu) \quad (3)$$

über die Frequenz rekursiv geglättet, wobei (3) eine Glättung in positive Frequenzrichtung beschreibt. Zur Vermeidung von Glättungsartefakten wird anschließend eine analoge Glättung in negativer Frequenzrichtung durchgeführt. Startwerte für die Glättung sind jeweils die entsprechenden Funktionswerte selbst. Die Glättungsvariable $\beta(\mu)$ ist frequenzabhängig und ermöglicht mit

$$\beta(\mu) = \frac{z(f(\mu + 1)) - z(f(\mu))}{f(\mu + 1) - f(\mu)} \quad (4)$$

eine psychoakustisch motivierte Glättung mit zunehmender Stärke bei hohen Frequenzen. Die Funktion $f(\mu)$ bildet die Frequenzindizes auf die entsprechenden Frequenzen in Hz und $z(f)$ die Frequenzen dann auf die Bark-Skala [3] ab. Anhand der so geglätteten Quotienten lassen sich dann unter anderem mittels der Levinson-Durbin-Rekursion [4] linearphasige Filter \mathbf{p}_λ entwerfen.

Die Filter \mathbf{p}_λ gelten für die jeweiligen Pumppositionen T_λ , allerdings nicht für die Positionen dazwischen. Um eine Entzerrung für jede beliebige Pumpposition zu ermöglichen, kann der Interpolationsansatz

$$\mathbf{x}_{\text{eq}}(n) = g_1(n) \cdot \mathbf{p}_i^T \mathbf{x}(n) + g_2(n) \cdot \mathbf{p}_j^T \mathbf{x}(n) \quad (5)$$

gewählt werden. Hierfür werden die Indices zeitvariant zu

$$i = \lfloor N \cdot s(n) \rfloor \quad \text{und} \quad j = \lceil N \cdot s(n) \rceil \quad (6)$$

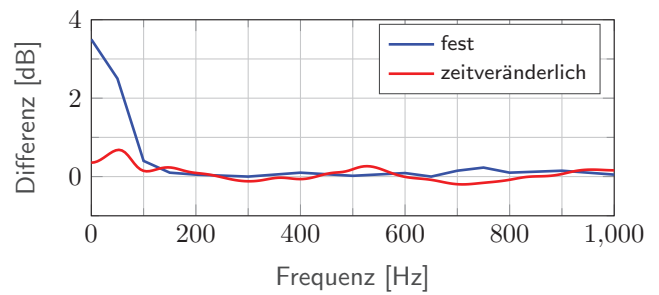


Abbildung 6: Abweichung zwischen den Leistungsdichtespektren des Ausgangs bei vollem und leerem Blasebalg im Bereich 0 – 1000 Hz. Vergleich von festem und pumppositionsabhängigem Entzerrer.

gewählt, sodass sich die Interpolationskoeffizienten zu

$$g_1(n) = N \cdot (T_j - s(n)) \quad (7)$$

und

$$g_2(n) = N \cdot (s(n) - T_i) \quad (8)$$

ergeben. Abb. 6 zeigt in Anlehnung an Abb. 4 dass mit dem vorgestellten Ansatz spektrale Abweichungen bedingt durch die Pumpposition ausgeglichen werden können.

Zusammenfassung

In den vorigen Abschnitten wurde zuerst die Notwendigkeit für einen „atmenden“ Kunstkopf erläutert und im nächsten Schritt ein entsprechender Prototyp vorgestellt. Anhand dieses Prototyps wurde gezeigt, dass die Entzerrung des Systems an die Pumpposition des angeschlossenen Blasebalges gekoppelt sein muss. Eine mögliche Umsetzung solch einer Entzerrung wurde anschließend dargelegt und validiert. Mit dieser besonderen Entzerrung ist es nun möglich das neue Kunstkopfsystem verlässlich zur Evaluierung von Feuerwehmasken einzusetzen. Dennoch sollten weitere Effekte des dynamischen Systemverhaltens, bspw. hervorgerufen durch den Luftstrom, untersucht werden.

Literatur

- [1] Deutscher Feuerwehrverband: Feuerwehr-Statistik, URL: www.feuerwehrverband.de/statistik.html, Letzter Zugriff: 29.01.2019
- [2] Welch, P.: The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics, vol. 15, no. 2, 1967 Acustica 90 (2004), 2203-2503
- [3] Zwicker, E und Fastl, H.: Psychoacoustics, Facts and Models, Springer, 1990
- [4] Vary, P. und Martin, R: Digital Speech Transmission: Enhancement, Coding and Error Concealment, Wiley, 2006