Störgeräuschreduktion mit einer Mel-Filterbank in Verbindung mit koinzidenten Mikrofonarrays

(Noise Reduction with a Mel Filter Bank in combination with a coincident microphone array)

B. Runow[1], A. Schilling[2], O. Curdt[3]

[1] *Wilhelm-Schickard Institut, Eberhard Karls Universität, Tübingen, Email: bernfried@runow.info*

[2] *Wilhelm-Schickard Institut, Eberhard Karls Universität, Tübingen, Email: schilling@uni-tuebingen.de*

[3] *Hochschule der Medien Stuttgart, Email: curdt@hdm-stuttgart.de*

Abstract

Bei der Abnahme eines akustischen Nutzsignals, beispielsweise eines Sprechers, soll mit Hilfe eines Mikrofons meist ausschließlich dieses Nutzsignal eingefangen werden. Das ist in einer Studioumgebung mit einem guten Mikrofon, das relativ nahe beim Sprecher platziert werden kann, keine besondere Herausforderung.

Befindet sich diese Nutzschallquelle aber in einem für diese Aufgabe akustisch ungünstigen Raum und befinden sich darüber hinaus noch andere Störschallquellen in ihrer Umgebung, dann wird die Abnahme zu einer durchaus anspruchsvollen Aufgabe. Man wird wahrscheinlich zu gerichteten Mikrofonen, gar einem Richtrohr greifen oder den Sprecher mit einem Lavalier-Mikrofon oder Headset verkabeln. Doch jeder weiß, es gibt Situationen, in denen es nicht möglich ist, ein Mikrofon direkt an der Nutzsignalquelle anzubringen oder in einer akzeptablen Entfernung aufzustellen. In einem solchen Fall benötigt man ein Mikrofon mit einer sehr starken Richtwirkung, also einem hohen Bündelungsmaß. Gleichzeitig soll die Richtcharakteristik möglichst frequenzunabhängig sein.

1. Einführung

Ausgehend von einem virtuellen Mikrofon eines koinzidenten Mikrofonarrays, das in Richtung der Nutzschallquelle ausgerichtet ist, wie in Abb. 1 dargestellt, wird ein zweites virtuelles Mikrofon gebildet, dessen Richtcharakteristik die größtmögliche Dämpfung in Richtung der Nutzschallquelle aufweist. Ziel ist es, mit diesem zweiten virtuellen Mikrofon möglichst genau den Störschall einzufangen, bestehend aus Reflektionen des Schalls der Nutzschallquelle und von anderen Schallquellen herrührendem Störschall. Das Ausgangssignal dieses zweiten virtuellen Mikrofons soll also möglichst alle an der Position des koinzidenten Mikrofonarrays auftretenden Störsignale enthalten und darüber hinaus möglichst keine Signalanteile des Nutzsignals.

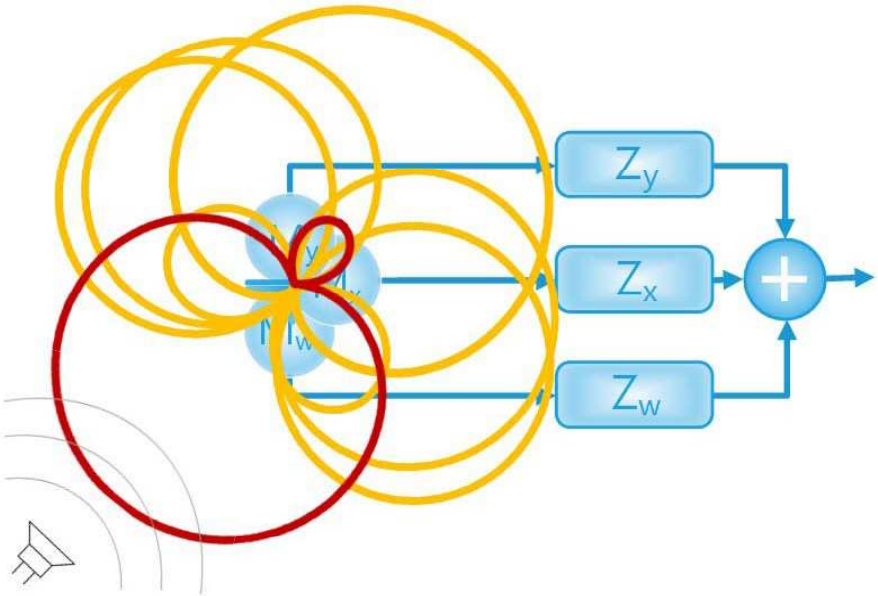

Abb. 1: Koinzidentes Mikrofonarray mit Nutzsignalquelle und den Richtcharakteristiken verschiedener virtueller Mikrofone.
Rot: Superniere in Richtung der Nutzsignalquelle; Gelb: Niere, Acht und Hypernieren mit maximaler Dämpfung in Richtung der Nutzsignalquelle.

ISBN 987-3-9812830-7-5

2. Signalmodell

Liegt die Nutzschallquelle ausgehend vom Mikrofonarray in der Richtung $\theta = 0°$, so wird also neben dem auf das Nutzsignal in Richtung $\theta = 0°$ gerichteten ersten virtuellen Mikrofon ein zweites virtuelles Mikrofon mit möglichst geringem Bündelungsmaß, jedoch einer größtmöglichen Dämpfung in Richtung $\theta = 0°$ für das Störsignal gebildet. Hier bietet sich beispielsweise eine Nierencharakteristik in Richtung $\theta = 180°$ an. Das Ausgangssignal dieses zweiten Mikrofons sei $x_{ambi}(t)$, da es den das Mikrofonarray umgebenden (engl. ambient) (Stör-)Schall beinhalten soll. Ebenso möglich wäre die Verwendung einer virtuellen Acht, Hyperniere, Superniere oder sämtlicher Zwischenformen zur Abnahme des ambienten Störsignals. Diese Richtcharakteristiken weisen jedoch – mit Ausnahme der Acht – ein höheres Bündelungsmaß auf und erscheinen daher weniger geeignet.

Das virtuelle Mikrofon für das Nutzsignal muss hingegen ein Kompromiss aus möglichst großem Bündelungsmaß und hoher Rückwärtsdämpfung sein, um nach Möglichkeit nur das Nutzsignal abzunehmen. Hier bietet sich beispielsweise eine virtuelle Super- oder Hyperniere an. Das Ausgangssignal sei $x_{util}(t)$, dabei ist der Name vom englischen Wort „utile“ abgeleitet.

Auch wenn das virtuelle Mikrofon in Richtung der Nutzschallquelle noch so stark gerichtet ist, das Ausgangssignal $x_{util}(t)$ wird immer einen gewissen Störsignalanteil aufweisen:

$$x_{util}(t) = a_{util}s(t - t_0) + v_{util}(t)\,, \tag{1}$$

wobei $s(t)$ das Signal der Nutzschallquelle ist, das um die Laufzeit t_0, die das Signal von der Schallquelle zum koinzidenten Mikrofonarray benötigt, verzögert und um den Faktor a_{util} gedämpft ist. Das Signal $v_{util}(t)$ beschreibt den Störsignalanteil.

Im Vergleich zu räumlichen Mikrofonarrays erreichen koinzidente Mikrofonarrays unter Anwendung der Gradientensynthese erster Ordnung nur ein verhältnismäßig geringes Bündelungsmaß [5], wodurch gerade bei einer weiter entfernten Nutzsignalquelle der Störsignalanteil steigt. Daher soll eine weitere Erhöhung der Richtwirkung durch die Trennung der Nutzsignalanteile von den Störsignalanteilen erzielt werden.

Von dem Signal $x_{ambi}(t)$ wissen wir, wie der Störschall, der das Mikrofonarray umgibt, beschaffen ist:

$$x_{ambi}(t) = v_{ambi}(t)\,. \tag{2}$$

Zwar wird das Störsignal $x_{ambi}(t)$ nicht aus der Richtung empfangen, in die das virtuelle Mikrofon für die Nutzschallquelle gerichtet ist, es kann jedoch angenommen werden, dass der Störsignalanteil in $v_{util}(t)$ gerade in Hinblick auf die von der Nutzschallquelle verursachten Reflektionen im Raum und weiter entfernte Störsignalquellen ähnlich dem Störsignal $v_{ambi}(t)$ ist:

$$v_{util}(t) \approx v_{ambi}(t)\,. \tag{3}$$

Die Idee, das Störsignal $v_{ambi}(t)$ direkt vom Ausgangssignal $x_{util}(t)$ des virtuellen Mikrofons, das auf die Nutzsignalquelle ausgerichtet ist, abzuziehen, führt nicht zum gewünschten Erfolg. Der Grund dafür ist, dass das Störsignal $v_{ambi}(t)$ nicht exakt dem Störsignal $v_{util}(t)$ entspricht. Dieser Ansatz entspricht einer Matrizierung der Ausgangssignale $x_{util}(t)$ und $x_{ambi}(t)$ im Zeitbereich und verursacht eine ungewollte Änderung der Richtcharakteristik. Das Bündelungsmaß kann sich beispielsweise verschlechtern oder die Ausrichtung verschieben. Die Folge sind weniger Nutzsignalanteile und mehr Störsignalanteile im resultierenden Ausgangssignal, also genau das, was es zu verhindern gilt.

3. Subtraktion im Spektralbereich

3.1. Herleitung

Durch die Subtraktion der Beträge im Spektralbereich ist eine frequenzabhängige Bearbeitung des Signals möglich, womit die frequenzbezogenen Amplituden von $X_{ambi}(e^{j\Omega})$ von den Amplituden von $X_{util}(e^{j\Omega})$ abgezogen werden können: [1]

$$\left|Y(e^{j\Omega})\right| = \left|X_{util}(e^{j\Omega})\right| - \left|X_{ambi}(e^{j\Omega})\right| \cdot w(e^{j\Omega})\,. \tag{4}$$

wobei $\Omega = 2\pi f \,/\, f_s$ die auf die Abtastfrequenz f_s normierte Kreisfrequenz ist und $0 \le w(e^{j\Omega}) \le 1$ ein reeller Faktor, mit dem das Signal $X_{ambi}(e^{j\Omega})$ vor der Subtraktion gewichtet werden kann. Dieser ist abhängig vom ebenso reellwertigen Intensitätsfaktor ι und den beiden Eingangssignalen:

$$w(e^{j\Omega}) = \begin{cases} 1 \cdot \iota & f\text{ür } \left|X_{ambi}(e^{j\Omega})\right| \le \left|X_{util}(e^{j\Omega})\right| \\ \dfrac{\left|X_{util}(e^{j\Omega})\right|}{\left|X_{ambi}(e^{j\Omega})\right|} \cdot \iota & f\text{ür } \left|X_{ambi}(e^{j\Omega})\right| > \left|X_{util}(e^{j\Omega})\right| \end{cases}. \tag{5}$$

Für den einfachen Fall, solange die Amplitude $\left|X_{ambi}(e^{j\Omega})\right|$ kleiner als oder gleich $\left|X_{util}(e^{j\Omega})\right|$ ist, entspricht $w(e^{j\Omega})$ dem Intensitätsfaktor. Sollte $\left|X_{ambi}(e^{j\Omega})\right|$ größer als $\left|X_{util}(e^{j\Omega})\right|$ sein, wird $w(e^{j\Omega})$ mit dem Quotient aus den beiden Amplituden so skaliert, dass die Subtraktion in (4) nicht zu einem negativen Ergebnis für $\left|Y(e^{j\Omega})\right|$ führt. Der Intensitätsfaktor $0 \le \iota \le 1$ ermöglicht es, die Stärke der spektralen Subtraktion zu bestimmen. Mit dem Wert $\iota = 1$ wird der Subtrahend in (4) maximal, mit dem Wert $\iota = 0$ findet keine Subtraktion statt und das Ausgangssignal $Y(e^{j\Omega})$ entspricht dem Eingangssignal $X_{util}(e^{j\Omega})$.

Auch Funktionen, die einen weicheren Übergang von Fall 1 zu Fall 2 schaffen, sind möglich, sollen aber an dieser Stelle nicht diskutiert werden, da damit keine entscheidende, subjektiv wahrzunehmende Verbesserung bei der Umsetzung des Ansatzes erzielt wird.

Das Ausgangssignal $Y(e^{j\Omega})$ erhält den Phasenwinkel von $X_{util}(e^{j\Omega})$:

$$Y(e^{j\Omega}) = \left|Y(e^{j\Omega})\right| \cdot e^{j\angle X_{util}(e^{j\Omega})}\,. \tag{6}$$

3.2. Ergebnis

Im Gegensatz zur Matrizierung im Zeitbereich entsteht bei der spektralen Subtraktion keine Verschlechterung des Bündelungsmaßes und keine Änderung der Ausrichtung. Im Gegenteil, durch die Bearbeitung im Spektralbereich können

die Störsignalanteile gegenüber den Nutzsignalanteilen reduziert werden.

Allerdings entstehen bei der spektralen Subtraktion im resultierenden Ausgangssignal Artefakte, die als „musical tones“ bekannt sind. Diese treten beispielsweise auch bei der verlustbehafteten Audiocodierung auf. Die Stärke des Artefakts ist dabei auch von der relativen Größe des Subtrahenden in Bezug auf den Minuend abhängig. Je mehr sich der Betrag des Subtrahenden dem des Minuenden annähert, desto stärker ist die Störung im Ausgangssignal wahrnehmbar. Das bedeutet, dass durch Reduzierung des Intensitätsfaktors ι die Artefakte soweit minimiert werden können, bis sie nicht mehr wahrgenommen werden. Zwar sinkt dadurch auch die Stärke der Störschallunterdrückung, dennoch liefert dieses Verfahren in einer akustisch schwierigen Umgebung eine deutlich wahrnehmbare Verbesserung gegenüber der Gradientensynthese erster Ordnung.

4. Bearbeitung mit einer Mel-Filterbank

In der Praxis wird die spektrale Subtraktion an den durch die diskrete Fourier-Transformation (DFT) erhaltenen DFT-Koeffizienten vorgenommen, den sogenannten „bins“. Die spektrale Auflösung ist dabei von der Fenstergröße, also von dem der DFT zugeführten Zeitabschnitt abhängig, ebenso die Frequenzwerte, welche die DFT-Koeffizienten repräsentieren. Diese sind bei der Fouriertransformation linear von 0Hz bis zur halben Abtastfrequenz $f_s/2$ verteilt.

Im Gegensatz zur Fourier-Transformation arbeitet das menschliche Gehör hinsichtlich der Frequenz jedoch nicht linear. So löst das Gehör tiefe Frequenzen viel feiner auf als hohe. Der Psychologe Stanley Smith Stevens beschäftigte sich in den dreißiger Jahren des 20. Jahrhunderts mit diesen psychoakustischen Eigenschaften des menschlichen Gehörs.

4.1. Tonheit

Zur Bestimmung der Tonhöhenwahrnehmung schlug Stevens 1937 zusammen mit Volkman und Newmann die Mel-Skala vor.[7] Dabei ist ein Mel, abgeleitet von dem englischen Wort melody, definiert als Einheit für die Tonheit. Diese psychoakustische Größe gibt Auskunft über die menschlich wahrgenommene Tonhöhe und ist insofern interessant, da die Tonheit – wie eingangs erwähnt – nicht proportional zur Frequenz verläuft. Wird bis etwa 500Hz ein Sinuston der doppelt so hohen Frequenz, also einer Oktave, noch etwa als doppelt so hoch wahrgenommen, so verlieren die Frequenz und Tonheit bei höheren Frequenzen ihren näherungsweise linearen Zusammenhang.

Stevens definierte die Mel-Skala ausgehend von dem Ton mit der Frequenz $f = 1000\,Hz$, dem er die Tonheit $z = 1000\,Mel$ zuordnete. Ein doppelt so hoch wahrgenommener Ton besitzt einen doppelt so großen Tonheitswert. Die weiteren Zusammenhänge von Mel-Skala und Frequenz beruhen daher auf psychoakustischen Versuchen.

Mit der folgenden Approximation kann für einen Ton mit bekannter Frequenz die entsprechende Tonheit z ermittelt werden: [2][7]

$$z = 2595Mel \cdot log_{10}\left(1 + \frac{f}{700Hz}\right). \tag{7}$$

Für die inverse Berechnung gilt entsprechend:

$$f = 700Hz \cdot \left(\left(10^{\frac{z}{2595Mel}}\right) - 1\right). \tag{8}$$

Bei der orangefarbenen Kurve in Abb. 2 ist gut zu erkennen, wie der noch näherungsweise lineare Zusammenhang von Frequenz und Tonheit ab etwa 500Hz verloren geht.

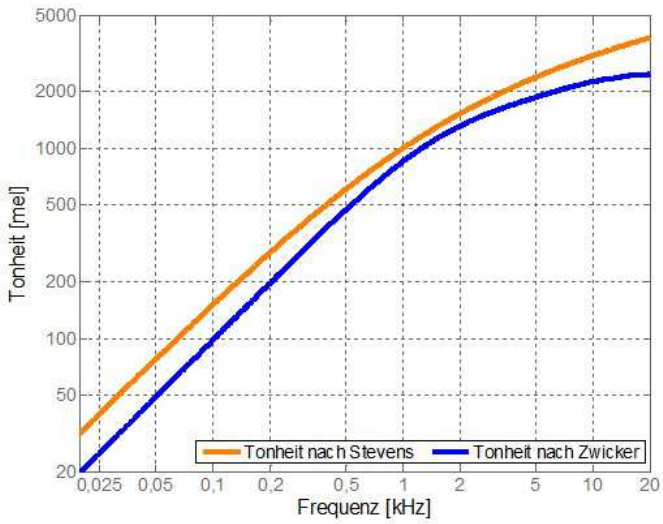

Abb. 2: Verhältnis von Tonheit und Frequenz nach Stevens und nach Zwicker.

Ausgehend von dem Ton c (~131Hz) schlug Eberhard Zwicker später eine Mel-Skala vor, die auf der Bark-Skala aufbaut. Die Bark-Skala ist ebenfalls eine auf psychoakustischen Versuchen beruhende Skala zur Bewertung der Tonheit. Er definierte: [8]

$$131\,Hz = 131\,Mel = 1{,}31\,Bark. \tag{9}$$

Die Bark-Skala und die Mel-Skala nach der Definition Zwickers sind demnach um den Faktor 100 verschoben: 1 Bark = 100 Mel. Im Gegensatz zur Mel-Skala nach Stevens bietet die Skala nach Zwicker durch den tiefer gewählten Bezugspunkt zur Frequenz den Vorteil, dass im linearen Teil, also bis etwa 500Hz, die Tonheit und die Frequenz nahezu identische Werte aufweisen. Die Mel-Skala nach Zwicker kann über die Bark-Skala wie folgt approximiert werden:

$$z = 100Mel \cdot \left(13 \cdot tan^{-1}\left(\frac{0{,}00076 \cdot f}{1Hz}\right) + 3{,}5 \cdot tan^{-1}\left(\left(\frac{f}{7500Hz}\right)^2\right)\right). \tag{10}$$

In Abb. 2 ist sowohl die Mel-Skala nach Stevens als auch die Mel-Skala nach Zwicker im Verhältnis zur Frequenz abgebildet. Neben dem verschobenen linearen Bereich fällt die flachere Kurve bei sehr hohen Frequenzen auf. Die von der Bark-Skala abgeleitete Mel-Skala nach Zwicker sieht also bei hohen Frequenzen eine noch stärkere Frequenzanhebung vor, um die doppelte Tonhöhe zu erzielen.

ISBN 987-3-9812830-7-5

4.2. Mel-Filterbank

Akustische Informationen wie die Tonhöhe, die Lautstärke und die Schalleinfallsrichtung unterteilt das menschliche Gehör zur Auswertung in einzelne Frequenzgruppen [3][4]. Man kann also sagen, die mit dem Ohr empfangenen Informationen werden mit Hilfe einer Filterbank weiterverarbeitet, wobei die Mittenfrequenz und Bandbreite der einzelnen Filter der nichtlinearen Frequenzskala des Gehörs und damit näherungsweise der Mel-Skala folgen. Soll nun ein Audiosignal gemäß dem menschlichen Gehör analysiert werden, so bietet sich der Einsatz einer Mel-Filterbank an. Diese besteht aus M Bandpassfiltern, deren Mittenfrequenzen linear über die Mel-Skala verteilt sind. Möchte man die Mel-Filterbank nicht nur zur Analyse, sondern auch zur Synthese verwenden, so muss die Summe aller Filter die Eigenschaften eines Allpassfilters aufweisen:

$$\sum_{m=1}^{M} H_m(e^{j\Omega}) = 1\,. \tag{11}$$

Auf diese Weise werden durch die Filterbank keine ungewollten Verzerrungen des Audiosignals hervorgerufen. Außerdem müssen die Filter linearphasig sein, so dass die einzelnen Bänder nach der Bearbeitung wieder phasenrichtig aufaddiert werden können. Daher werden meistens sich halb überlappende dreieckförmige Filter verwendet,

$$H_m(e^{j\Omega}) = \begin{cases} 0 & \text{für } f < \zeta(m-1) \\ \dfrac{f-\zeta(m-1)}{\zeta(m)-\zeta(m-1)} & \text{für } \zeta(m-1) \le f \le \zeta(m) \\ \dfrac{\zeta(m+1)-f}{\zeta(m+1)-\zeta(m)} & \text{für } \zeta(m) \le f \le \zeta(m+1) \\ 0 & \text{für } f > \zeta(m+1) \end{cases}, \tag{12}$$

wobei ζ(m) mit m=0, 1, ..., M+1 insgesamt M+2 Frequenzen sind, die linear über die Mel-Skala verteilt sind.
Abb. 3 zeigt eine Mel-Filterbank mit sechs dreieckförmigen Bandpassfiltern sowie dem notwendigen Hoch- und Tiefpassfilter.
Die Mittenfrequenz und Bandbreite der einzelnen Filter wurden mit Hilfe von (7) und (8) berechnet, also der Tonheitsdefinition nach Stevens. Im Gegensatz zur Definition nach Zwicker löst diese hohe Frequenzen etwas genauer auf, was in Hinblick auf die Anwendung der Filterbank zumindest nicht nachteilig ist.

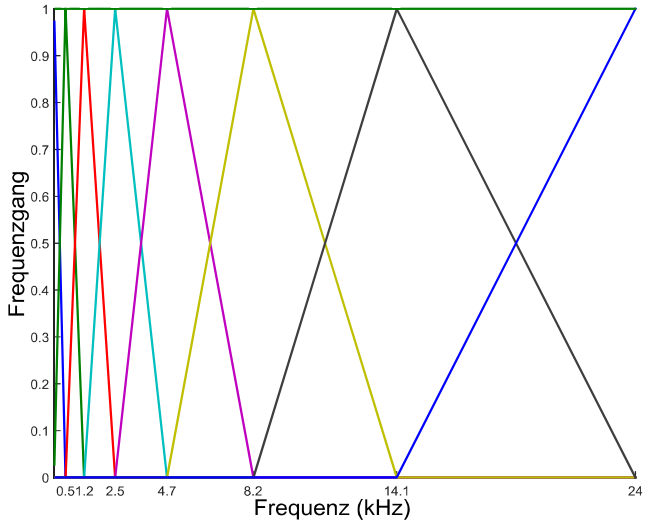

Abb. 3: Mel-Filterbank für M=6, mit Tief- und Hochpass am unteren und oberen Ende. Für die Berechnung der Mittenfrequenzen und der Bandbreiten wurde die Mel-Skala nach Stevens verwendet.

Neben den dreieckförmigen Bandpassfiltern können theoretisch alle Filter für die Mel-Filterbank verwendet werden, deren Frequenzgang bei Überlagerung aller Filter der Filterbank die Forderung nach einem Allpass und einer linearen Phase erfüllt. Bei einer fünfzigprozentigen Überlappung der Filter bietet sich beispielsweise ein Amplitudengang der von-Hann-Fenster-Form an, wie in Abb. 4 zu sehen ist.
Im Gegensatz zu einem dreieckförmigen Filter bietet die von-Hann-Form eine stärkere Betonung des Bereichs um die Mittenfrequenz. Die im Band weiter außen liegenden Frequenzen werden also weniger stark gewichtet.

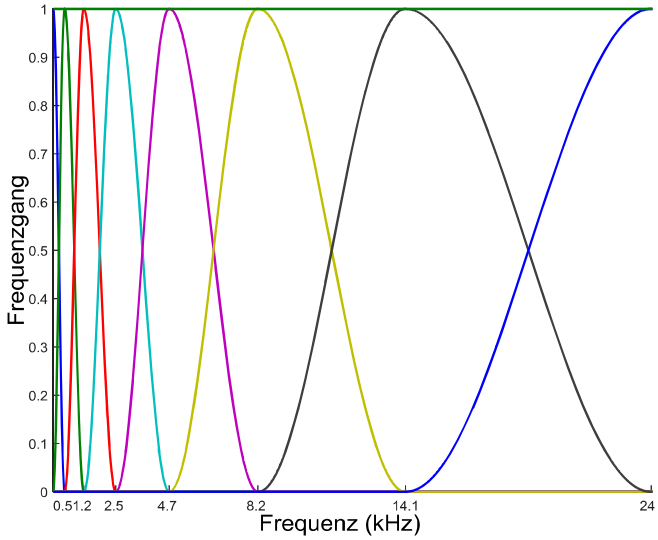

Abb. 4: Mel-Filterbank mit von-Hann-förmigen Bandpassfiltern für M=6, sowie Tief- und Hochpass am unteren und oberen Ende. Für die Berechnung der Mittenfrequenzen und der Bandbreiten wurde die Mel-Skala nach Stevens verwendet.

ISBN 987-3-9812830-7-5

4.3. Analyse

Zur Analyse, wie groß der Störsignalanteil im Ausgangssignal $X_{util}(e^{j\Omega})$ des auf die Nutzschallquelle ausgerichteten virtuellen Mikrofons ist, werden mit Hilfe der Mel-Filterbank M+2 bandbegrenzte Signale sowohl von $X_{util}(e^{j\Omega})$ als auch von $X_{ambi}(e^{j\Omega})$ erstellt:

$$X_m(e^{j\Omega}) = X(e^{j\Omega}) \cdot H_m(e^{j\Omega}) \,. \tag{13}$$

Anschließend können für die einzelnen bandbegrenzten Signale die Leistungswerte ermittelt werden:

$$P_{X,m}(X_m) = \frac{1}{2\pi} \int_0^{2\pi} \left| X_m(e^{j\Omega}) \right|^2 \, d\Omega \,. \tag{14}$$

Ist nun in einem bestimmten Frequenzbereich ein das koinzidente Mikrofonarray umgebender Störschall vorhanden, so ist der Leistungswert des für diesen Frequenzbereich bandbegrenzten Signals $X_{ambi,m}(e^{j\Omega})$ erhöht. Mit dem eingangs vorgestellten Signalmodell (1) und der Folgerung daraus (3) kann davon ausgegangen werden, dass dieser Störsignalanteil zu einem gewissen Grad auch im entsprechend identisch bandbegrenzten Signal $X_{util,m}(e^{j\Omega})$ vorhanden ist.

4.4. Synthese

Um nun diesen detektierten Störsignalanteil aus dem Ausgangssignal $X_{util,m}(e^{j\Omega})$ des auf die Nutzschallquelle ausgerichteten virtuellen Mikrofons zu entfernen, wird ein Korrekturfaktor aus dem Verhältnis von bandbegrenzter Leistung $P_{ambi,m}$ und $P_{util,m}$ berechnet:

$$g_m = 1 - \frac{P_{ambi,m}}{P_{util,m}} \cdot \iota \,, \tag{15}$$

wobei $0 \leq \iota \leq 1$ der bereits von der spektralen Subtraktion bekannte reelle Intensitätsfaktor ist, mit dem die Stärke der Korrektur bestimmt werden kann. Der Korrekturfakor g_m gewichtet beim Wiederzusammensetzen die einzelnen bandbegrenzten Signale $X_{util,m}(e^{j\Omega})$ und realisiert so die frequenzselektive Störsignalunterdrückung:

$$Y(e^{j\Omega}) = \sum_{m=0}^{M-1} X_{util}(e^{j\Omega}) \cdot H_m(e^{j\Omega}) \cdot g_m \,. \tag{16}$$

4.5. Ergebnis

Diese Minderung der Störsignalanteile führt zu einer deutlichen Bündelung der Richtcharakteristik und damit zu einer Erhöhung der Richtwirkung hin zur Nutzschallquelle.

Die Störsignalreduktion mit Hilfe der beschriebenen Mel-Filterbank ist subjektiv wahrnehmbar. Im Gegensatz zur Gradientensynthese erster Ordnung wird selbst in überakustischen Räumen und einer größeren Entfernung der Signalanteil der Nutzsignalquelle deutlich von den Störsignalanteilen freigestellt.

Problematisch ist hingegen das Auftreten von „musical tones"-artigen Artefakten, die auch hier abhängig von der Stärke der Bearbeitung auftreten.

5. Vergleich von spektraler Subtraktion und der Mel-Filterbank

Was die beiden Ansätze zur Störgeräuschreduktion eint, ist das identische Artefakt im Ausgangssignal, das abhängig von der Bearbeitungsstärke auftritt. So kann mit dem Intensitätsfaktor ι die Stärke der Bearbeitung so weit reduziert werden, bis die Störung im Ausgangssignal nicht mehr wahrgenommen werden kann. Allerdings reduziert sich damit auch der positive Effekt zur Erhöhung der Richtwirkung.

Unterschiedlich reagieren die beiden Ansätze auf bestimmte Voraussetzungen bei den Eingangssignalen $x_{util}(t)$ und $x_{ambi}(t)$. Entsprechen die Störsignale in $x_{ambi}(t)$ exakt den Störsignalanteilen in $x_{util}(t)$, also $x_{ambi}(t) = v_{util}(t)$, dann liefert die spektrale Subtraktion das bessere Ergebnis.
Sind sich die Störsignale jedoch nur ähnlich $x_{ambi}(t) \approx v_{util}(t)$, wovon bei der praktischen Anwendung mit einer koinzidenten Mikrofonanordnung auszugehen ist, liefert die Bearbeitung mit der Mel-Filterbank das bessere Ergebnis.
Dieses Verhalten lässt sich relativ einfach theoretisch begründen: Die spektrale Subtraktion arbeitet sehr genau, vergleicht und bearbeitet jeden DFT-Koeffizienten für sich und erzielt daher ein sehr gutes Ergebnis, wenn das Störsignal, das entfernt werden soll, vollständig bekannt ist. Ist das Störsignal jedoch nur näherungsweise bekannt, so führt diese Form der Bearbeitung zwangsläufig zu Fehlern. Der Ansatz mit der Mel-Filterbank bearbeitet hingegen nicht jeden DFT-Koeffizienten für sich, sondern berücksichtigt benachbarte Koeffizienten bei der Korrektur, so dass der Bearbeitungsvorgang eine ‚weichere' Veränderung vornimmt und damit eine größere Toleranz gegenüber dem Störsignal entwickelt. Genau diese Eigenschaft führt dann aber wiederum zu einem größeren Fehler, wenn das Störsignal vollständig bekannt ist.

6. Schlussfolgerung

Mit der Mel-Filterbank steht damit gerade für den praktischen Anwendungsfall ein sehr interessanter Ansatz für die Störgeräuschreduktion zur Verfügung. Die Besonderheit des Verfahrens ist die dem menschlichen Hörverhalten angepasste Signalverarbeitung. Dabei werden tiefe Frequenzen deutlich feiner aufgelöst als hohe.
Durch die vorgestellte Störgeräuschreduktion wird das Nutzsignal von Störsignalen freigestellt und ermöglicht so die Steigerung des Bündelungsmaßes bei koinzidenten Mikrofonarrays. Auch weiter entfernte Nutzsignale können auf diese Weise selbst unter schwierigen Bedingungen eingefangen werden.
Damit die Störsignalreduktion auch mit einem hohen Intensitätsfaktor eingesetzt werden kann, sollte das Auftreten der Artefakte weiter reduziert werden. Dazu könnten beispielsweise psychoakustische Gewichtungsregeln berücksichtigt werden.

ISBN 987-3-9812830-7-5

7. Literatur

[1] Boll, S. F.: Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Tran. on Acoustics, Speech and Signal Processing ASSP-27, 2, 1979, S. 113-120

[2] Fant, G.: Speech Sounds and Features. MIT Press, Cambridge 1973

[3] Hellbrück, J., Ellermeier, W.: Hören - Physiologie, Psychologie und Pathologie. Hogrefe, Göttingen, Bern, Toronto, Seattle, 2004

[4] Moore, B. C.: An Introduction to the Psychology of Hearing. Emerald Group, Bingley, UK, 2012

[5] Runow, B., Curdt, O.: Mikrofonarrays in der professionellen Audioproduktion. Tagungsbericht 28. Tonmeistertagung (2014), S. 263-269

[6] Stadermann, J. R.: Automatische Spracherkennung mit hybriden akustischen Modellen (Dissertation), Technische Universität München, 2005

[7] Stevens, S. S., Volkman, J., Newman, E.: A scale for the measurement of the psychological magnitude of pitch. The Journal of the Acoustical Society of America 8, 1937, S. 185-190

[8] Zwicker, E., Terhardt, E.: Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. Journal of the Acoustical Society of America 68, 1980, S. 1523-1525

ISBN 987-3-9812830-7-5