Freisprechen mit adaptiver geräuschabhängiger Bandbreite

Martin Schießl[1], Klaus Linhard[2], Harald Schnepp[2]

[1] *Monkey Engineering, 89134 Blaustein, Deutschland, Email: schiessl@monkey-engineering.de*

[2] *Daimler AG, 89013 Ulm, Deutschland, Email: klaus.linhard@daimler.com, harald.schnepp@daimler.com*

Einleitung

Für Freisprechen in geräuscherfüllter Umgebung (Beispiel Fahrzeug) wird seit vielen Jahren als Geräuschreduktion die sogenannte „Spektrale Subtraktion“ in der Praxis angewandt. Es wurden viele Verbesserungen des Verfahrens publiziert, um z.B. die Adaption an das Geräusch oder die Vermeidung von Artefakten zu erhöhen, z.B. [1].

Hier wird ein Verfahren vorgestellt, das eine Art Vorstufe für eine Geräuschreduktion darstellt. In dieser Vorstufe wird die Bandbreite adaptiv geregelt, und anschließend eine Geräuschreduktion durchgeführt. Die nachfolgende Geräuschreduktion kann eine Spektrale Subtraktion sein, oder aber eine andere Verarbeitung. Anstelle des üblichen festen Kompromiss-Frequenzgangs hat das Gesamtsystem jetzt einen optimierten, geräuschabhängigen Frequenzgang. Bei wenig Geräusch wird das System breitbandig, bei starkem tieffrequenten Geräusch wird der Frequenzgang im Bassbereich eingeschränkt.

Der Kompromiss-Frequenzgang

Bei dem Anwendungsbeispiel Freisprechen im Kraftfahrzeug werden üblicherweise Mikrofone mit einem Freisprech-Frequenzgang verwendet, wie er z.B. in der ITU-T P.1110 empfohlen wird [2]. Zunächst ist der Frequenzgang im akustischen Freifeld vorgegeben. Durch den Mikrofon-Einbau im Fahrzeug ergibt sich dann der tatsächliche Frequenzgang, auch als „send frequency response mask“ bezeichnet. Dieser beinhaltet den Freifeldfrequenzgang und die Fahrzeugakustik. Dafür soll hier die Bezeichnung Referenzfrequenzgang H_{ref} verwendet werden. In Abbildung 1 ist ein Beispiel für H_{ref} eingezeichnet.

Die Fahrzeuggeräusche sind oft tieffrequent, so dass durch den Hochpass-Charakter des Freisprech-Referenzfrequenzgangs (z.B. ab 300 Hz) bereits viel Geräusch entfernt werden kann ohne die Sprache zu stark zu beeinträchtigen. Freisprechsysteme haben in aller Regel eine Eingangssignal-Filterung, die dazu benutzt wird H_{ref} final festzulegen, d.h. es werden feste Kompromiss-Werte eingestellt. Somit wird in der Regel bei geringem Geräusch zu viel gefiltert und bei sehr starkem Geräusch zu wenig. Ein Beispiel für ein sehr starkes Geräusch sind das Geräusch und die Windschläge, die vom Gebläse eines Fahrzeugs bei Maximal-Einstellung erzeugt werden.

Umschaltbare Frequenzgänge

Ein erster Optimierungsschritt ist die Umschaltung von verschiedenen Eingangs-Filtern, abhängig von der erkannten Geräuschsituation. In einem einfachen Fall kann

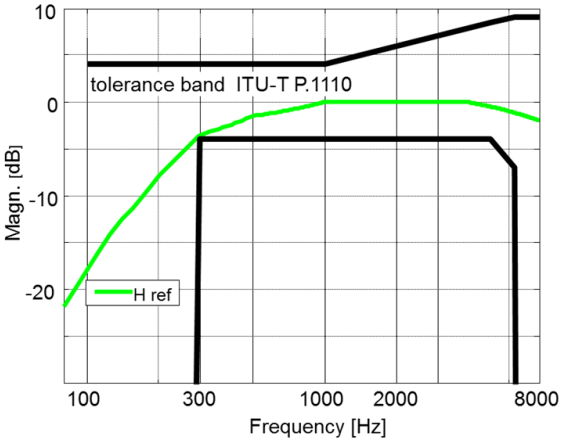

Abbildung 1: Mikrofon-Frequenzgang im Fahrzeug: Toleranzband und Referenzfrequenzgang H_{ref} [2].

z.B. bei starkem Gebläse-Geräusch ein zusätzlicher Hochpass eingeschaltet werden. Dadurch würde das eine Kompromissfilter gewissermaßen in 2 oder noch weitere Kompromissfilter weiterentwickelt. Zusätzlich müsste ein Auswahlmechanismus entworfen werden.

Adaptiver Frequenzgang mit Wiener Filter

Der elegantere Weg ist ein adaptives Optimalfilter zu verwenden. Der klassische Ansatz dafür ist das Wiener Filter, das durch mitlaufende Schätzungen des Nutz- und Störsignals adaptiv erweitert wird. Das Fehlerkriterium ist die Minimierung des mittleren quadratischen Fehlers $e(k)$ zwischen den gestörten Nutzsignal $x(k)$ und dem ungestörten Nutzsignal $s(k)$, $e(k) = x(k) - s(k)$. Das gestörte Nutzsignal ist die Addition: $x(k) = s(k) + n(k)$. Dabei bezeichnet $n(k)$ das Störsignal, k die diskrete Zeit.

Mit $y(k)$ als das Ausgangssignal, $h(k)$ als Impulsantwort des Wiener Filters und „$*$“ als Faltungsoperation gilt:

$$y(k) = h(k) * x(k). \tag{1}$$

Im Frequenzbereich mit der diskreten Frequenz i und der Wiener Lösung

$$H(i) = \frac{\left|S(i)\right|^2}{\left|X(i)\right|^2}, \tag{2}$$

ergibt sich unter Annahme unkorrelierter Signale [1]:

$$H(i) = \frac{|X(i)|^2 - |N(i)|^2}{|X(i)|^2}, \quad (3)$$

$$H(i) = 1 - \frac{|N(i)|^2}{|X(i)|^2} \quad (4)$$

und somit für das Ausgangssignal $Y(i)$:

$$Y(i) = H(i)X(i). \quad (5)$$

In Gleichung 2 und einigen folgenden wäre korrekterweise die Leistungsdichte für die Spektren von N und S zu verwenden. Hier soll zu Gunsten der besseren Lesbarkeit darauf verzichtet werden.

$S(i)$ und $N(i)$ sind zunächst unbekannt und müssen geschätzt werden. Da die Schätzung über der Zeit ausgeführt wird, ist das Wiener Filter zeitvariant.

$$\hat{H}(k,i) = 1 - \frac{|\hat{N}(k,i)|^2}{|X(k,i)|^2}, \quad (6)$$

$$\hat{H}(k,i) = 1 - \frac{|\hat{N}(k,i)|^2}{|\hat{N}(k,i)|^2 + |\hat{S}(k,i)|^2}. \quad (7)$$

Obige Gleichung 6 ist im Wesentlichen die Gleichung, wie sie auch bei der als „Spektrale Subtraktion“ [1] bezeichneten Geräuschreduktion verwendet wird, wobei im Nenner das bekannte Eingangssignal $X(k,i)$ verwendet wird und für das Geräusch eine Schätzung benötigt wird.

Adaption der Störsignal-Schätzung

Im weiteren Verlauf möchten wir Gleichung 7 verwenden. Die Schätzung von $N(k,i)$ gibt:

$$|\hat{N}(k+1,i)|^2 = (1-\alpha)\,|\hat{N}(k,i)|^2 + \alpha|X(k,i)|^2, \quad (8)$$

wobei der Glättungsparameter α abhängig davon geregelt wird, ob $|X(k,i)|^2$ größer oder kleiner als der Schätzwert $|\hat{N}(k,i)|^2$ ist:

$$\alpha(k,i) = \alpha_0 \cdot \frac{|\hat{N}(k,i)|^2}{|X(k,i)|^2}. \quad (9)$$

Zur Anwendung von Gleichung 7 verbleibt die Schätzung von $|\hat{S}|^2$. Wir treffen folgende Vereinbarungen:

$$H(k,i) = 1; \qquad i > i_0, \quad (10)$$

$$|\hat{S}(k,i)|^2 := |S_{\text{ref}}(i)|^2. \quad (11)$$

Das Filter H wird nur bis zur diskreten Frequenz i_0 (z.B. 400 Hz) eingesetzt und das Sprachsignal durch ein Referenz-Sprachspektrum $|S_{\text{ref}}|^2$ ersetzt.

Referenz-Sprachspektrum

Dadurch, dass das Sprachspektrum $|S_{\text{ref}}|^2$ als konstant angenommen wird ergeben sich keine tonalen Störungen (im Englischen „musical tones“), wie sie bekanntermaßen bei der der Spektralen Subtraktion auftreten. Das Filter H ändert sich nur langsam, es passt sich dem geschätzten Geräusch an. Diese Schätzung erfolgt nach üblichen Verfahren mit einer großen Zeitkonstante (Größenordnung 1 Sekunde, durch entsprechende Wahl von α_0).

Ein geeigneter Mittelwert für $|S_{\text{ref}}|^2$ ist z.B. das genormte mittlere Sprach-Spektrum, oder auch das sogenannte „Artifical Voice“. Artifical Voice Spektren sind für männliche und weibliche Sprecher bekannt [3]. Es kann z.B. der Mittelwert dieser beiden Spektren verwendet werden. Alternativ kann auch ein bekanntes Langzeitspektrum von Sprache verwendet werden [4]. Die Normlautstärke des Sprechers ist durch die bekannte Anwendung im Fahrzeug in Form des Mittelwerts der Lautstärke bekannt. Wird lauter gesprochen ist das Filter etwas zu stark eingreifend, wird leiser gesprochen ist das Filter etwas zu wenig eingreifend. Das Filter bleibt allerdings in seiner grundlegenden Wirkung erhalten, es öffnet sich bei wenig Geräusch und schließt sich bei starkem Geräusch. Bei Bedarf kann $|S_{\text{ref}}|^2$ dem tatsächlichen Sprachpegel oder auch dem Sprechspektrum langsam nachgeführt werden.

Abbildung 2 zeigt die hier gewählte Vorgehensweise zur Ermittlung von $|S_{\text{ref}}|^2$. Artificial Voice nach [3] wurde im Fahrzeug über den Kunstkopf abgespielt, und die geglättete Version des Spektrums („art. voice smooth“) eingesetzt. Zur Verifikation wird das Spektrum des verwendeten realen Sprechers mit angegeben.

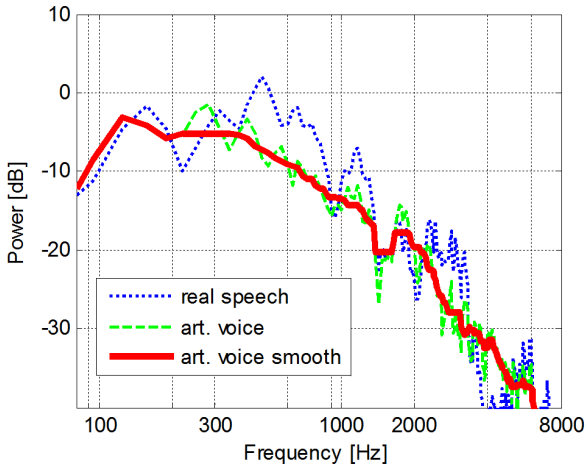

Abbildung 2: Ermittlung des Sprachspektrums $|S_{\text{ref}}|^2$ und Vergleich mit realem Sprecher.

Ergebnisse

Es soll die Wirkungsweise der geräuschabhängigen Bandbreite an 3 Beispielen gezeigt werden, bei Fahrgeschwindigkeit 100 km/h und 140 km/h und bei einer maximalen Gebläsestufe (hier Stufe 7 von 7). Dabei erfolgt ein Vergleich mit dem konstanten Kompromissfrequenzgang H_{ref}. Es wird ein Freisprechmikrofon mit einem linearen Freifeldfrequenzgang von ca. 80 Hz bis über 10 kHz verwendet. Die aufgezeichneten Daten werden anschließend

mit einem Filter H_{ref} oder alternativ mit dem adaptiven Filter H verarbeitet. Der Vorteil durch das adaptive Filter im Vergleich zum Kompromissfilter H_{ref} liegt z.B. bei dem Geräuschszenario 100 km/h im Bereich 100 bis 200 Hz bei 5 bis 10 dB mehr Sprachpegel (Abbildung 3).

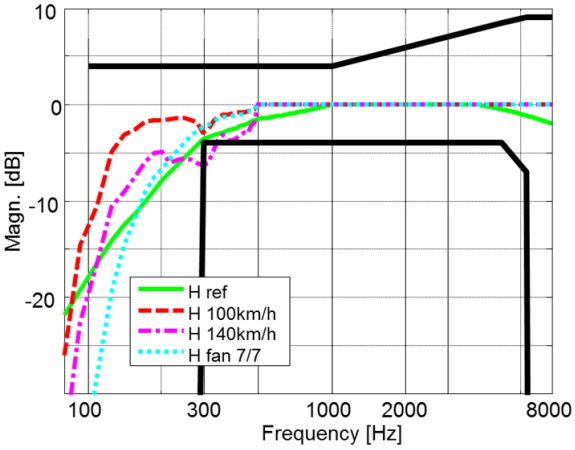

Abbildung 3: Kompromissfilter H_{ref} im Vergleich zum adaptiven Filter H.

Nachfolgend soll gezeigt werden, dass das adaptive Filter tatsächlich zu einem besseren Signal-zu-Rauschabstand SNR führt. Dabei wird der Fall mit dem Kompromissfilter H_{ref} mit den Fall ohne Filter und 2 Varianten mit adaptivem Filter verglichen. $H_{\mathrm{N(est)}}$ bezeichnet den Fall wobei das Geräusch adaptiv geschätzt wird und $H_{\mathrm{N(true)}}$ den Fall mit dem tatsächlichen Geräuschspektrum. Das tatsächliche Geräuschspektrum ist nur im Fall der Simulation bekannt wobei Sprache und Geräusch nachträglich addiert werden. Für die Praxis gilt der Fall $H_{\mathrm{N(est)}}$ wobei die Geräuschschätzung dem Spektrum in den Sprachpausen näherungsweise folgt (Gleichung 8).

Eine SNR-Auswertung wurde im Bereich 80 bis 400 Hz durchgeführt (siehe Abbildung 4). Die übliche SNR-Berechnung

$$\mathrm{SNR} := \frac{\sum |S(i)|^2}{\sum |N(i)|^2} \tag{12}$$

wird durch die nachfolgende Form ersetzt, bei der neben dem Geräusch auch die lineare Verzerrung der Sprache durch die Filterung als Fehler mit erfasst wird

$$\mathrm{SNR} := \frac{\sum |S(i)|^2}{\sum |H(i)N(i)|^2 + \sum |(1-H(i))S(i)|^2}. \tag{13}$$

$H_{\mathrm{N(true)}}$ zeigt immer den höchsten SNR-Wert und $H_{\mathrm{N(est)}}$ zeigt immer einem besseren Wert als der Kompromissfilter H_{ref} und das Original ohne Filter.

Diskussion und Anmerkungen

Gerade bei der Wideband-Telefonie ab ca. 80 Hz würde durch einen Kompromiss-Frequenzgang bei geringen Geräuschen zu viel Sprachqualität eingebüßt. Bei der älteren Narrowband-Telefonie mit einem Frequenzbereich

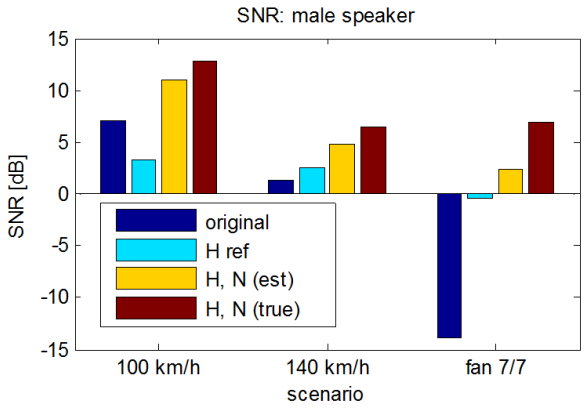

Abbildung 4: Signal-zu-Rauschabstände für verschiedene Filter und Fahrsituationen.

von ca. 300 Hz bis 3,6 kHz war diese Problematik permanent vorhanden und konnte nicht verbessert werden.

Systeme mit dem hier vorgeschlagenem adaptivem Frequenzgang sind nicht mehr eindeutig bestimmt, d.h. bei Frequenzgangmessungen ist ein Rauschspektrum zu definieren. Die Anlehnung dieses Entwurfs an das Wiener Filter erfolgte auch mit Hinblick darauf, dass dieses Vor-Filter mit geringem Aufwand in die Spektrale Substraktion (Wiener Filter) integriert werden kann.

Das Fehlerkriterium des mittleren quadratischen Fehlers ist nicht psycho-akustisch optimiert, aber durchaus ein übliches Verfahren zur Geräuschreduktion. Modifikationen des Fehlerkriteriums sind bekannt, wurden hier aber nicht verwendet.

Zusammenfassung

Bei einer Freisprecheinrichtung wurde das Kompromissfilter H_{ref} durch ein adaptives Filter H ersetzt, das als Wiener Filter entworfen wird. Dieses Filter passt sich langsam an das sich verändernde Geräusch an und erzeugt damit keine Artefakte (musical tones). Dieses Filter wird so eingesetzt, dass sich die Bandbreite in den tiefen Frequenzen verändert, d.h. bei starken tieffrequenten Geräuschen die Bass-Bandbreite verringert. Es kann gezeigt werden, dass das adaptive Filter den SNR-Wert bei verschiedenen Szenarien deutlich verbessert und somit ein hochwertiges artefaktfreies Signal für eine nachfolgende Geräuschreduktion darstellt.

Literatur

[1] Linhard, K.; Haulick, T.: Spectral noise subtraction with recursive gain curves. ICSLP-1998, pp. 1479 - 1482

[2] ITU-T, P.1110: Wideband hands-free communication in motor vehicles, 12/2009

[3] ITU-T, P.50: Artificial Voices, 09/1999

[4] Byrne, D., et al.: An international comparison of long-term average speech spectra, J. Acoust. Soc. Am 96 (4), (10/1994)