

## Verteilte Mikrofone im Kraftfahrzeug

Patrick Vicinus<sup>1</sup>, Wolf Baumann<sup>2</sup>

<sup>1</sup> PEIKER acoustic, 61381 Friedrichsdorf, Deutschland, Email: patrick.vicinus@peiker.de

<sup>2</sup> TU-Berlin, Fachgebiet Elektronik und medizinische Signalverarbeitung

### Einleitung

Freisprechtelefonie findet zunehmende Anwendung beim Automobilkunden. Die Positionierung der Mikrofone durch den Fahrzeughersteller wird dabei leider immer weniger durch akustische Gegebenheiten als durch Vorgaben der Designer festgelegt. Liegt eine Anordnung aus verteilten Mikrofonen vor, kann durch Filterung und Summation der Signale (*Filter and Sum*) eine einkanalige Variante des Nutzsignals berechnet werden. Die Filterkoeffizienten sind so zu bestimmen, dass das resultierende Nutzsignal im Gegensatz zu den einzelnen Kanälen an Sprachverständlichkeit gewinnt. Zur Vereinfachung werden lediglich SNR (*Signal to Noise Ratio*) und Lautstärke als Maß der Sprachverständlichkeit bewertet.

### Modellbeschreibung

Im Folgenden sei die hier beschriebene verallgemeinerte Konfiguration aus verteilten Mikrofonen gegeben: Im Fahrzeuginnenraum befinden sich  $I$  Mikrofone,  $K$  Sprecher und  $M$  Störer. Jeder Sprecher und jeder Störer gelangen über ihre zugeordnete Raumübertragungsfunktion (*Room Transfer Function: RTF*)  $h_{ki}$  bzw.  $h_{mi}$  zum Mikrofon  $i$ . Das Signalgemisch  $x_i$  am Mikrofon  $i$  aus Signal  $s_k$  und Störung  $n_m$  hat demnach folgende Form:

$$x_i = \left( \sum_K s_k * h_{ki} \right) + \left( \sum_M n_m * h_{mi} \right) \quad (1)$$

bzw. bei zeitdiskreter Betrachtung und unter Zuhilfenahme der Matrixschreibweise bei einer Länge der RTF von  $L$ :

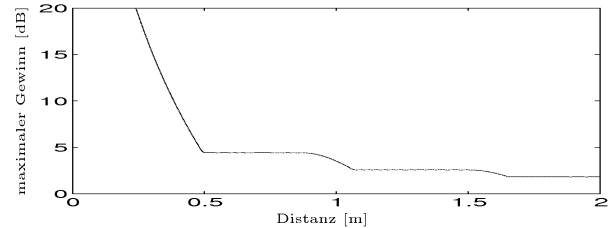
$$X_i = \mathbf{H}_{K_i}^H \mathbf{S}_i + \mathbf{H}_{M_i}^H \mathbf{N}_i, \quad (2)$$

mit

$$\mathbf{S}_i^T = \left\{ s_{t-L+1}^{(1)} s_{t-L+2}^{(1)} \dots s_t^{(1)} s_{t-L+1}^{(2)} s_{t-L+2}^{(2)} \dots s_t^{(2)} \dots s_{t-L+1}^{(K)} s_{t-L+2}^{(K)} \dots s_t^{(K)} \right\} \quad (3)$$

$$\mathbf{H}_{K_i}^H = \left\{ h_{K_0}^{*(0)} h_{K_1}^{*(0)} \dots h_{K_{L-1}}^{*(0)} h_{K_0}^{*(1)} h_{K_1}^{*(1)} \dots h_{K_{L-1}}^{*(1)} \dots h_{K_0}^{*(K)} h_{K_1}^{*(K)} \dots h_{K_{L-1}}^{*(K)} \right\} \quad (4)$$

Die Matrizen  $\mathbf{N}_i$  und  $\mathbf{H}_{M_i}^H$  haben gleiche Form. Bei Übergang in den Frequenzbereich und unter Verzicht der Darstellung der Frequenzabhängigkeit ergibt sich für die gefilterte Addition das einkanalige Zielsignal in jedem Frequenzband



**Abbildung 1:** Maximaler additiver SNR-Gewinn einer Frequenz im Bereich zwischen 300Hz und 3000Hz in Abhängigkeit von der Distanz der Mikrofone, der theoretisch durch Einbeziehung der Minimierung einer diffusen Störung erreicht werden kann.

$$\mathbf{Y} = C_1^* X_1 + \dots + C_I^* X_I = \mathbf{C}^H \mathbf{X}. \quad (5)$$

### Ansätze

Eine Maximierung des *SNR* kann prinzipiell in drei Varianten erfolgen: Maximierung der Ausgangsleistung, Minimierung der Rauschleistung und die Maximierung des Quotienten aus Ausgangs- und Rauschleistung. Im Folgenden wird untersucht, unter welchen Bedingungen diese Varianten zur Verbesserung der Sprachverständlichkeit genutzt werden können.

### Maximierung der Ausgangssignalleistung

Eine Maximierung der Ausgangssignalleistung führt nur dann zu einer Maximierung des *SNR*, wenn der Gewinn an *SNR* durch eine Minimierung der Störung unter gleichen Nebenbedingungen nicht vorhanden bzw. nicht signifikant ist. Sowohl die Maximierung als auch die Minimierung der Störleistung unter der Nebenbedingung einer konstanten Norm der Koeffizienten kann als Eigenwert-Problem formuliert werden:

$$\min: \mathbf{C}^H \mathbf{N} \mathbf{N}^H \mathbf{C} \quad \mathbf{C}^H \mathbf{C} = 1 \quad (6)$$

$$\Rightarrow \mathbf{C}^H \mathbf{N} \mathbf{N}^H + \lambda \mathbf{C}^H = 0 \quad \mathbf{C}^H \mathbf{C} = 1 \quad (7)$$

Die Minimierung wird durch den Eigenvektor des kleinsten Eigenwertes und die Maximierung durch den Eigenvektor des größten Eigenwertes der Kovarianzmatrix der Mikrofonensignale erreicht. Das Verhältnis aus größtem und kleinstem Eigenwert der Kovarianzmatrix der Störung beschreibt das maximale additive relative Maß an Störgeräuschunterdrückung, das durch Einbeziehung der Minimierung der Störung erreicht werden kann.

Abbildung 1 zeigt dieses Maß in Abhängigkeit der Distanz der Mikrofone eines äquidistanten Arrays. Betrachtet wird hier der sprachverständlichkeitsrelevante Frequenzbereich zwischen 300Hz und 3000Hz unter Annahme eines diffusen Störfeldes. Verallgemeinert konvergiert dieser Ansatz zum optimalen SNR-Wert, wenn der Abstand  $d$  der Mikrofone  $d > \frac{\lambda_{max}}{2}$  ist und die Leistung der Störung an jedem Mikrofon gleich ist. Wobei hier  $\lambda_{max}$  die maximale berücksichtigte Wellenlänge ist.

Werden bei der Maximierung die Nebenbedingungen des *MVDR Beamformers* (s.u.) berücksichtigt, berechnen sich die Filterkoeffizienten zu:

$$\begin{aligned} \mathbf{C}_{max} &= EV_{max}(\mathbf{S}_{xx}) \\ \mathbf{C}^H &= \max(\mathbf{C}_{max}) \mathbf{C}_{max}^H \end{aligned} \quad (8)$$

## MVDR Beamformer

Bei steigender Korrelation der Störung an den Mikrofonen nimmt die Ungenauigkeit des obigen Ansatzes zu. Ein bekanntes Verfahren zur Maximierung des SNRs bei korrelierter Störung ist der Ansatz des MVDR (Minimum Variance Distortionless Response) Beamformers. Das Verfahren basiert auf einer Minimierung der Störung unter Annahme eines Signalmodells, das die akustische Übertragungsfunktion (RTF)  $\mathbf{H}_K$  des Nutzsignals zwischen den Mikrofonen als bekannt voraussetzt. Für die Kovarianzmatrix des Signals gilt dann:

$$\mathbf{S}_{ss} = S_f \cdot \mathbf{H}_K \mathbf{H}_K^H \quad (10)$$

Dient das Nutzsignal am Mikrofon mit lautestem Signalpegel  $H_{max}$  als Signalreferenz, dann berechnen sich nach diesem Ansatz die Filterkoeffizienten zu

$$\mathbf{C}^H = \frac{\mathbf{H}_{max} \mathbf{H}_K^H \mathbf{S}_{nn}^{-1}}{\mathbf{H}_K^H \mathbf{S}_{nn}^{-1} \mathbf{H}_K} \quad (11)$$

Die Schätzung der Raum-Übertragungsfunktionen  $\mathbf{H}_K$  erweist sich als schwierig. Liegt weder eine Schätzung der Statistik zweiter Ordnung vom Nutzsignal noch vom Störgeräusch vor, kann durch eine geeignete Analyse der Eigenvektoren (*Principle Component Analysis: PCA*) die RTF geschätzt werden. Allerdings führt im Falle gerichteter Störer erst die Ausnutzung der zeitlichen Eigenschaften zu zuverlässigen Schätzungen [1].

Ist die Störung unkorreliert zum Nutzsignal und sind ihre Statistiken bekannt, kann durch Subtraktion der bekannten Störgeräusch- und Signalgemisch-Kovarianzmatrix eine Nutzsignal-Kovarianzmatrix gebildet werden:

$$\hat{\mathbf{S}}_{ss} = \mathbf{S}_{xx} - \mathbf{S}_{nn} \quad (12)$$

Dabei bezeichnet  $\hat{\mathbf{S}}_{ss}$ ,  $\mathbf{S}_{xx}$  und  $\mathbf{S}_{nn}$  die Kovarianzmatrix des Nutzsignals, der Mikrofon-signale und der Störung. Durch die Wahl des Eigenvektors zum größten Eigenwert als *RTF* ist gewährleistet, dass der Beamformer auf

den größtmöglichen Leistungsanteil des Nutzsignals fokussiert. Da für diesen Ansatz allerdings die Statistiken des Signals und der Störung mit ähnlicher Zeitkonstante geschätzt werden müssen, ist er für den praktischen Einsatz weniger geeignet: Da für die Schätzung der Nutzsinalstatistiken eine relativ kleine Zeitkonstante notwendig ist, muss die Störung zeitgleich geschätzt werden. Ohne Modell-Bildung ist dies nicht möglich.

## Maximierung des SNRs

Im Folgenden wird ein Verfahren hergeleitet, das unabhängig zu den *RTFs* selbst bei gerichteten Störern die optimalen Filterkoeffizienten berechnet. Die direkte Maximierung des SNRs führt zu folgendem Ergebnis [2]:

$$\mathbf{C}_{snr} = EV_{max} \{ \mathbf{S}_{nn}^{-1} \mathbf{S}_{xx} \} \quad (13)$$

Dabei sind  $\mathbf{C}_{snr}$  die Koeffizienten, die das SNR maximieren. Da die Schätzung der Störkovarianzmatrix multiplikativ und nicht additiv berücksichtigt wird, kann sie über lange Zeitkonstanten in Sprachpausen geschätzt werden. Ansatzbedingt ist die Norm der Koeffizienten bis hier noch unbestimmt. Des Weiteren gehen die *RTFs* der akustischen Mikrofonwege nicht direkt aus der Lösung hervor. Unter Annahme gleicher Modelle lassen sich die Lösung dieses Ansatzes und die des Ansatzes des „*MVDR Beamformers*“ gleichsetzen:

$$\mathbf{C}_{snr}^H \stackrel{!}{=} \mathbf{H}_K^H \mathbf{S}_{nn}^{-1} \cdot b \quad (14)$$

Für den Koeffizientenvektor  $\mathbf{C}^H$  und die Schätzung der *RTF*  $\mathbf{H}_K$  ergibt sich schließlich mit  $a = \max(\mathbf{C}_{snr}^H \mathbf{S}_{nn})$ :

$$\mathbf{H}_K = a^{-1} \cdot \mathbf{S}_{nn} \mathbf{C}_{snr} \quad \mathbf{C}^H = \frac{1}{\mathbf{C}_{snr}^H \mathbf{H}_K} \mathbf{C}_{snr}^H \quad (15)$$

## Zusammenfassung

Es wurden drei Verfahren vorgestellt, die das SNR für Sprache im Fahrzeuginnenraum mit Hilfe einer verteilten Mikrofonanordnung optimieren. Durch das Zusammenführen zweier Verfahren können sowohl die im Sinne der Aufgabenstellung optimalen Koeffizienten bestimmt, als auch die *RTFs* des Nutzsignals von Mikrofon zu Mikrofon geschätzt werden. Die Kenntnis der *RTFs* kann in weitergehenden Verarbeitungsstufen ausgenutzt werden. Das Bilden einer Störreferenz und die Lokalisierung des Sprechers seien hier nur zwei Beispiele.

## Literatur

- [1] A New GSC-Based Adaptive Array, Fang-Biau Ueng and Shiann-Jeng Yu, National Chung-Hsing University, Taiwan
- [2] Optimum Array Processing. Van Trees, Wiley, 2002