Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SP Signal Processing

Vadim Titov, Henning Möllers, Leon Mannweiler

Beamforming

Multichannel Speech Enhancement

Universität Hamburg
Signal Processing (SP)

May 4, 2022

Gliederung

1. Motivation
2. Grundlagen
3. Algorithmen
4. Ausblick

Motivation

Motivation

- Sprachverbesserung ermöglicht bessere Kommunikation in geräuschvoller Umgebung

(a) [1]

(b) [2]

Anforderungen an Sprachverbesserung

- Vom Algorithmus wird Gesprächsqualität bzw. Verständlichkeit erwartet

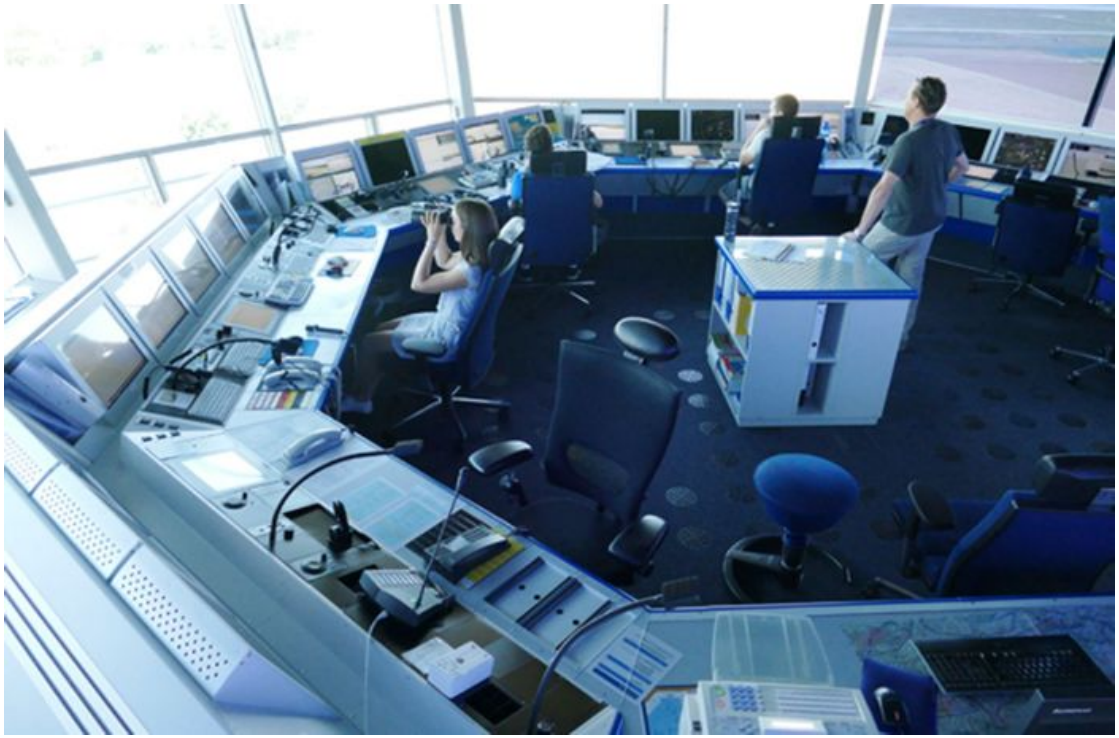

Figure: [3]

Anforderungen an Sprachverbesserung

- Ebenso wichtig ist dem Anwender die Geschwindigkeit der Verarbeitung

Figure: [4]

Mehrere Mikrofone bringen einen Vorteil

- Durch technologischen Fortschritt wird der Einsatz mehrerer immer kleiner werdenden Mikrofonen öfter zum Standard in vielen Geräten
- Mehr als ein Mikrofon liefert zusätzliche räumliche Information

  $\rightarrow$ Einsatz von effizienteren Algorithmen möglich

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SP
Signal Processing

Grundlagen

Beamforming

- Ziel: Beabsichtigte **räumliche Selektivität** erreichen (die sich ggf. im Zeitverlauf unterscheidet)

Figure: [5]

Beamforming

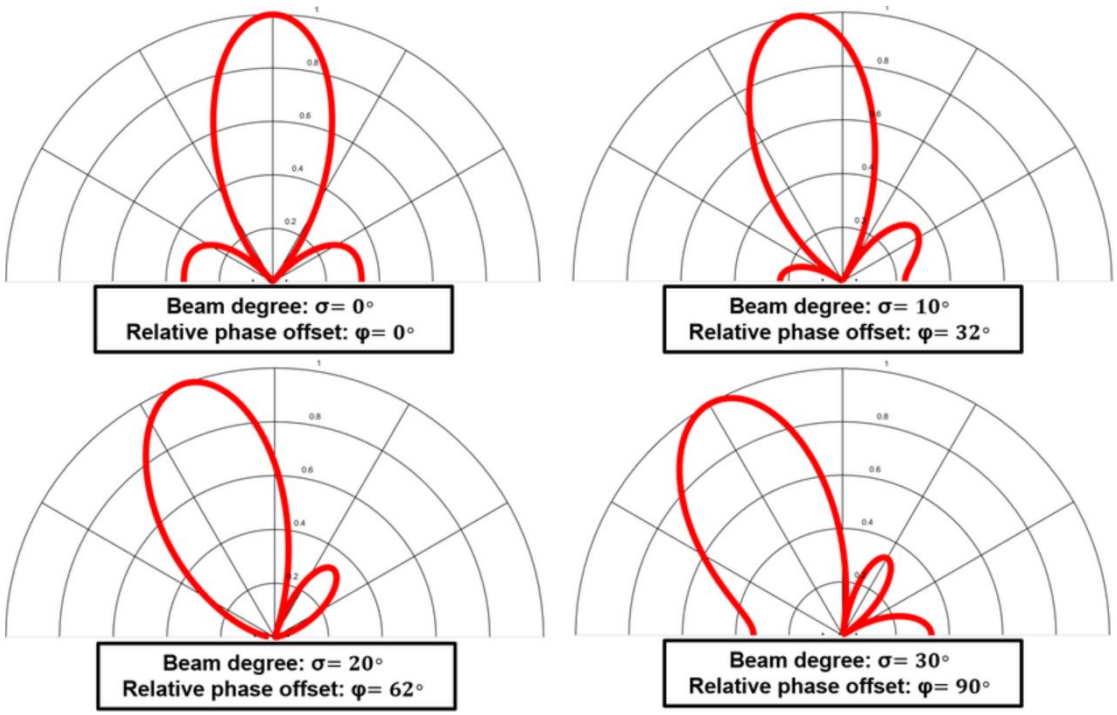

Figure: [6]

- Je nach Ausrichtung ändert sich meine räumliche Selektivität
  $\rightarrow$ **Beam-Pattern**

Wir haben aber doch keine Richtmikrofone!

Wie dann?

- Eingehende Signale in spezifischer Art und Weise kombinieren
  $\rightarrow$ Algorithmen
- Daten**un**abhängige Algorithmen: Position der Signalquellen bekannt
- Datenabhängige Algorithmen: Position der Signalquellen nicht bekannt

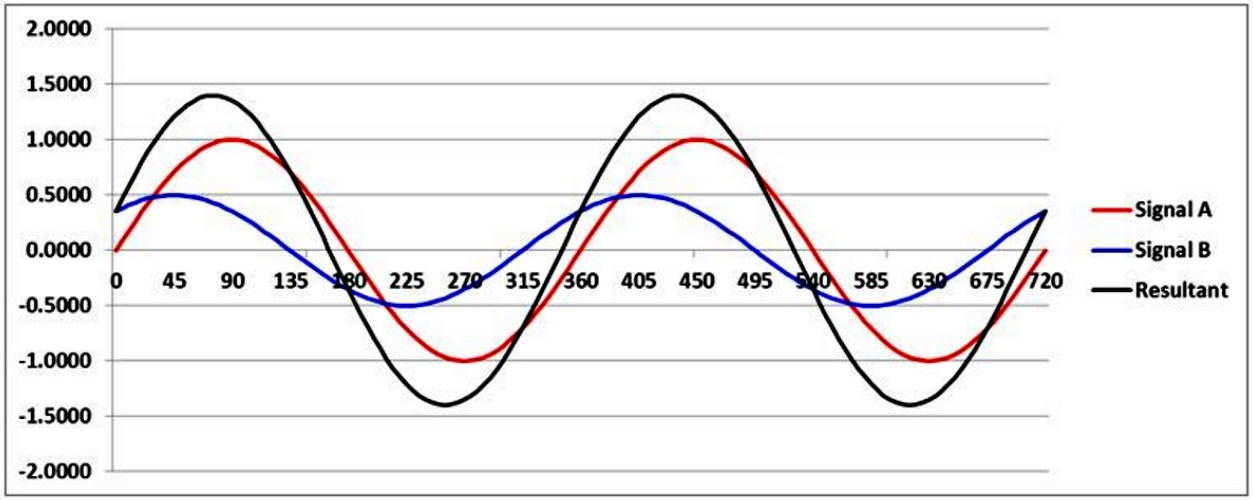

Figure: [7]

Beamformer

- Implementierung eines Beamforming-Algorithmus → **Beamformer**

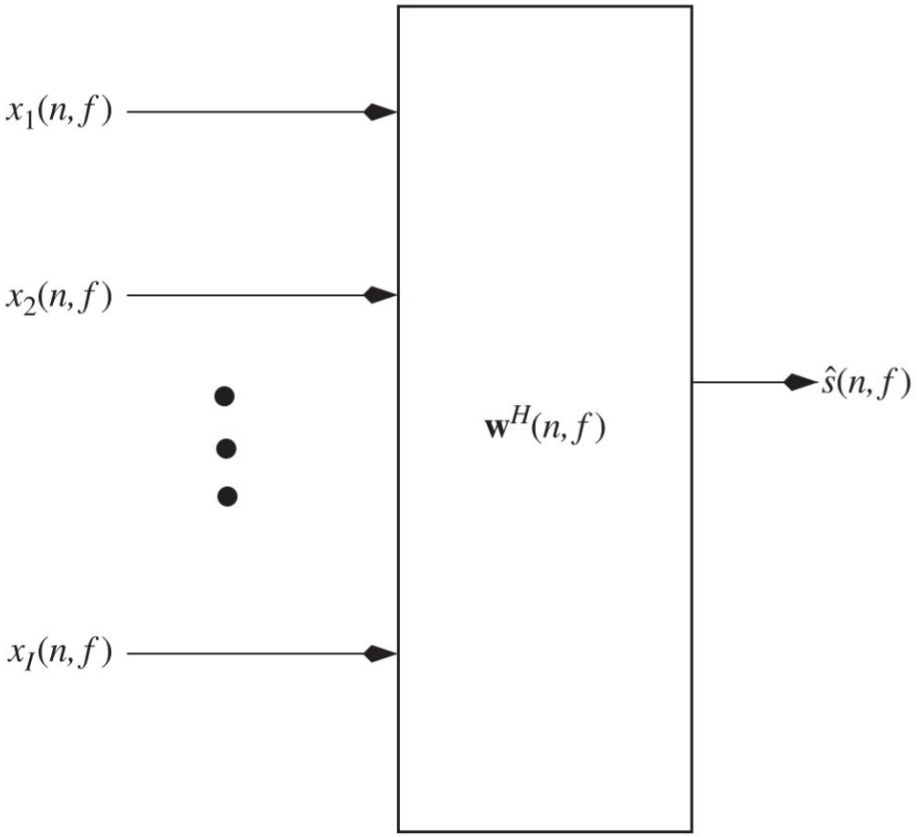

Figure: [8, S. 210]

Schallfelder

Was?

- Entfernung zwischen Signal-Quelle und Mikrofon-Array
- Nahfeld: $r < 2\lambda$[1]
- Fernfeld: $r > 2\lambda$

Warum?

- Fernfeld: Signal trifft mit ebenen Wellen und mit gleicher Signalstärke auf die Mikrofone $\rightarrow$ einfachere Berechnungen

$\Rightarrow$ wichtige Annahme für Algorithmen

- Einordnung: Sprache mit $300 - 3400Hz$
  $\Rightarrow$ Wellenlänge $\lambda = 0,1 - 1,1$ Meter $\Rightarrow$ Fernfeld ab $\sim 2,2$ Meter

[1] es existieren verschiedenste Formeln/Annahmen

Schallfelder

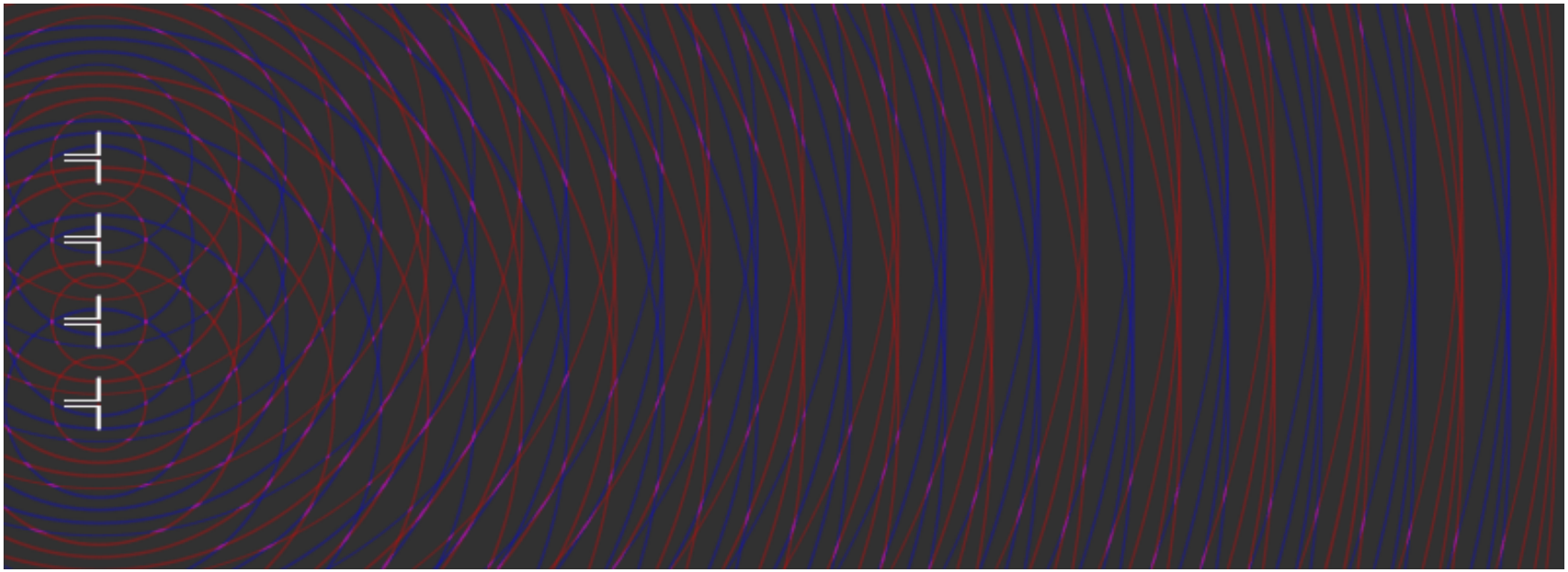

Figure: [9]

Array-Topologien

- Anzahl, Abstand, Anordnung der Mikrofone
- Zentral: **Uniform Linear Array**
- Symmetrie ermöglicht einfachere Berechnungen
  $\rightarrow$ Annahme für Algorithmen

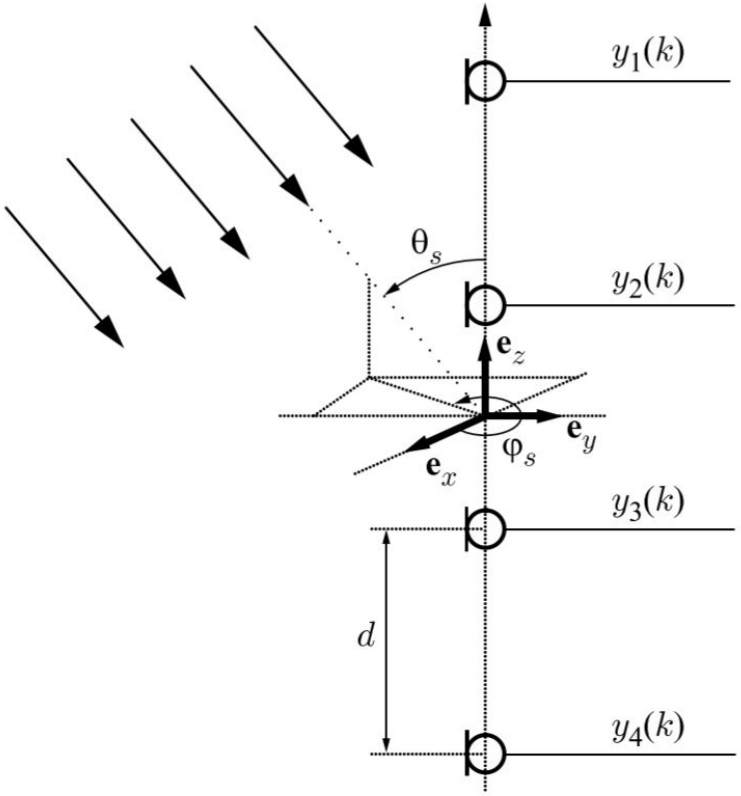

Figure: [10, S. 474]

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SP Signal Processing

Algorithmen: Data-Independent (Fixed) Beamformers

Data-Independent (Fixed) Beamformers

- Signalstatistiken werden nicht berücksichtigt ($\rightarrow$ data-independent).
- Position der Signalquellen muss a-priori bekannt sein (Annahme: Fernfeld).
- Filterkoeffizienten sind statisch ($\rightarrow$ fixed).

Delay-and-sum Beamformer

Algorithmus:

- Eingangssignale werden abhangig vom Signaleinfallwinkel verzögert.
- Ausgangssignal wird bezüglich der Anzahl der Eingangssignale normalisiert (gemittelt), um eine Verzerrung zu vermeiden.

Verzögerte Einzelsignale:

$$\tilde{y}_\ell(t) = \tilde{s}_\ell(t) + \tilde{n}_\ell(t) = s(t + \Delta\tau_\ell - T_\ell) + n_\ell(t - T_\ell), \tag{1}$$

wobei ℓ der Index des Mikrofons, T_ℓ die entsprechende und $\Delta\tau_\ell$ die relative Verzögerung ist.

Ausgangssignal:

$$\hat{s}(kT) = \frac{1}{N_M} \sum_{\ell=1}^{N_M} \tilde{y}_\ell(kT). \tag{2}$$

Delay-and-sum Beamformer

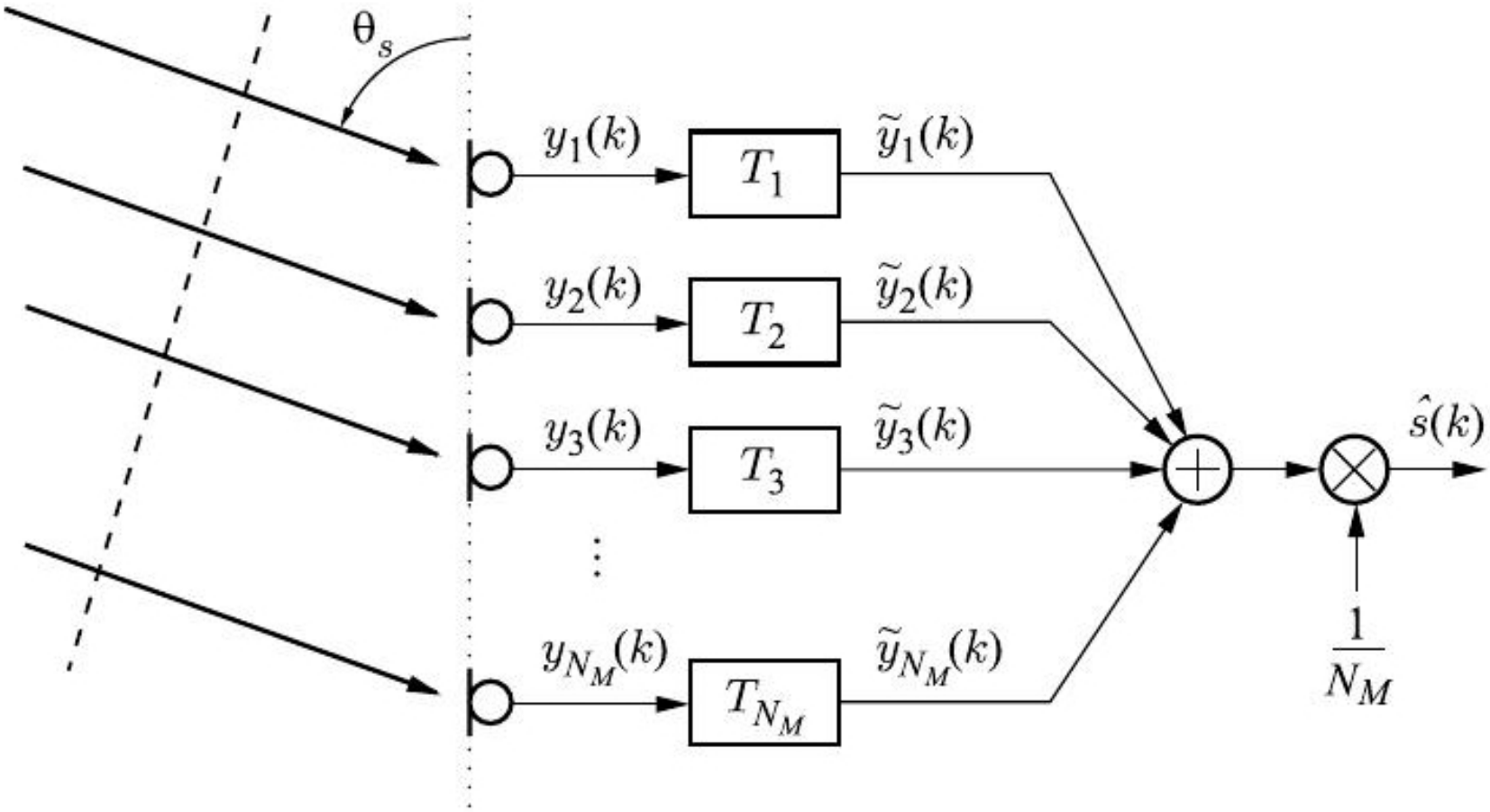

Figure: Delay-and-sum Beamformer, [10], S.477

Filter-and-sum Beamformer

Algorithmus:

- Eingangssignale werden entsprechend der Filterkoeffizienten verstärkt oder abgeschächt
- und anschliessend aufsummiert.

Ausgangssignal:

$$\hat{s}(k) = \sum_{\ell=1}^{N_M} \mathbf{h}_\ell^T \mathbf{y}_\ell(k), \tag{3}$$

wobei $\mathbf{y}_\ell(k)$ das Eingangssignal und $\mathbf{h}_\ell(k)$ die Filterkoeffizienten jeweils in Vektorschreibweise sind.

Filter-and-sum Beamformer

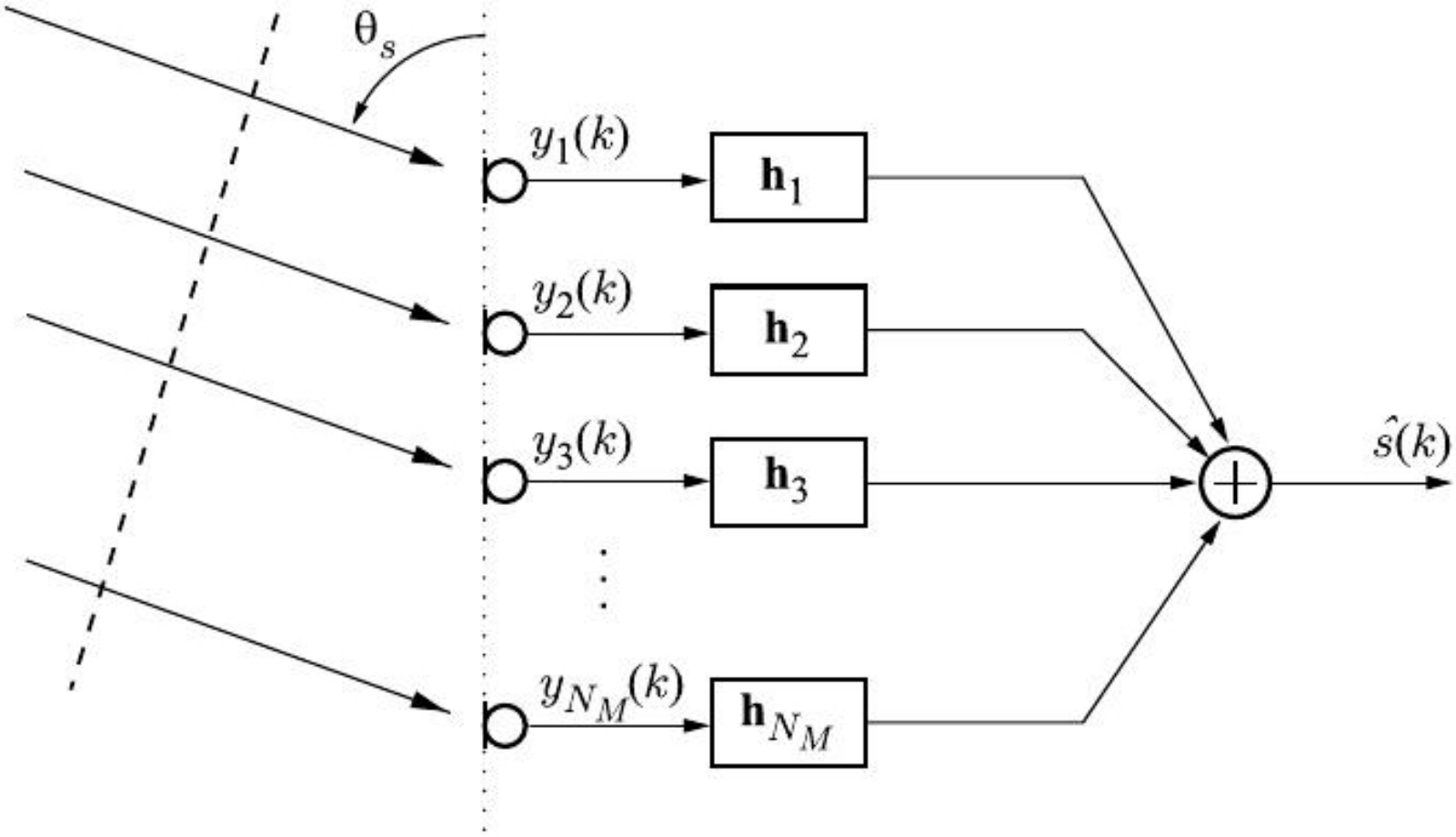

Figure: Filter-and-sum Beamformer,[10], S.479

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SP Signal Processing

Algorithmen: Data-Dependent (Adaptive) Beamformers

Data-Dependent (Adaptive) Beamformers

- Signalstatistiken werden berücksichtigt
  ($\rightarrow$ data-dependent).
- Position der Signalquellen a-priori unbekannt ($\rightarrow$ Source Localization).
- erfordert Kenntnis oder Schätzung der Übertragungsfunktionen (ATF oder RTF).
- Filterkoeffizienten sind anpassbar
  ($\rightarrow$ adaptiv).

Data-Dependent (Adaptive) Beamformers

Eingangssignal (Annahme):

$$
\begin{aligned}
\mathbf{x} &= \mathbf{h}_0 \mathbf{s}_0 + \sum_{p=1}^{P-1} \mathbf{h}_p \mathbf{s}_p + \mathbf{n} \\
&= \mathbf{h}_0 \mathbf{s}_0 + \mathbf{v} = \mathbf{s} + \mathbf{v}, \\
\mathbf{s} &\sim \mathbb{C}\mathcal{N}(\mu_\mathbf{s} = 0, \Sigma_\mathbf{s}), \\
\mathbf{v} &\sim \mathbb{C}\mathcal{N}(\mu_\mathbf{v} = 0, \Sigma_\mathbf{v}), \\
&\text{unkorreliert.}
\end{aligned} \tag{4}
$$

$\mathbf{s}$ ist das gewünschte Signal, $\mathbf{n}$ das Rauschen, $\mathbf{v}$ alle unerwünschten Signale und $\mathbf{h}$ die akkustische Übertragungsfunktion (ATF).

Ausgangssignal:

$$
\begin{aligned}
y &= \mathbf{w}^H \mathbf{x} \\
&= \mathbf{w}^H (\mathbf{s} + \mathbf{v}) \\
&= \mathbf{w}^H \mathbf{s} + \mathbf{w}^H \mathbf{v}.
\end{aligned} \tag{5}
$$

LCMV Beamformer

Der Linearly Constrained Minimum Variance (LCMV) Beamformer, auch Frost-Beamformer genannt, optimiert folgendes Kriterium:

$$\mathbf{w}_{\text{LCMV}} = \underset{\mathbf{w}}{\text{argmin}}\ \mathbf{w}^H \mathbf{\Phi}_{\mathbf{xx}} \mathbf{w} \text{ bzgl.} \mathbf{C}^H \mathbf{w} = \mathbf{g}, \tag{6}$$

wobei $\mathbf{w}^H \mathbf{\Phi}_{\mathbf{xx}} \mathbf{w} = E\{\mathbf{w}^H \mathbf{x}\mathbf{x}^H \mathbf{w}\} = E\{|(\mathbf{w}^H \mathbf{x})(\mathbf{x}^H \mathbf{w})|\} = E\{|y|^2\}$ die Ausgangsleistung,

$\mathbf{\Phi}_{\mathbf{xx}} = E\{\mathbf{x}\mathbf{x}^H\}$ die $M \times M$-dimensionale Kreuzleistungs-Spektraldichte (fouriertransformierte Kreuzkorrelationsmatrix) der empfangenen Signale

und $\mathbf{C}^H \mathbf{w} = \mathbf{g}$ die $M \times P$-dimensionale Constraintsmatrix ist, wobei P die Anzahl der Beschränkungen (constraints) ist, denen entsprochen wird.

LCMV Beamformer

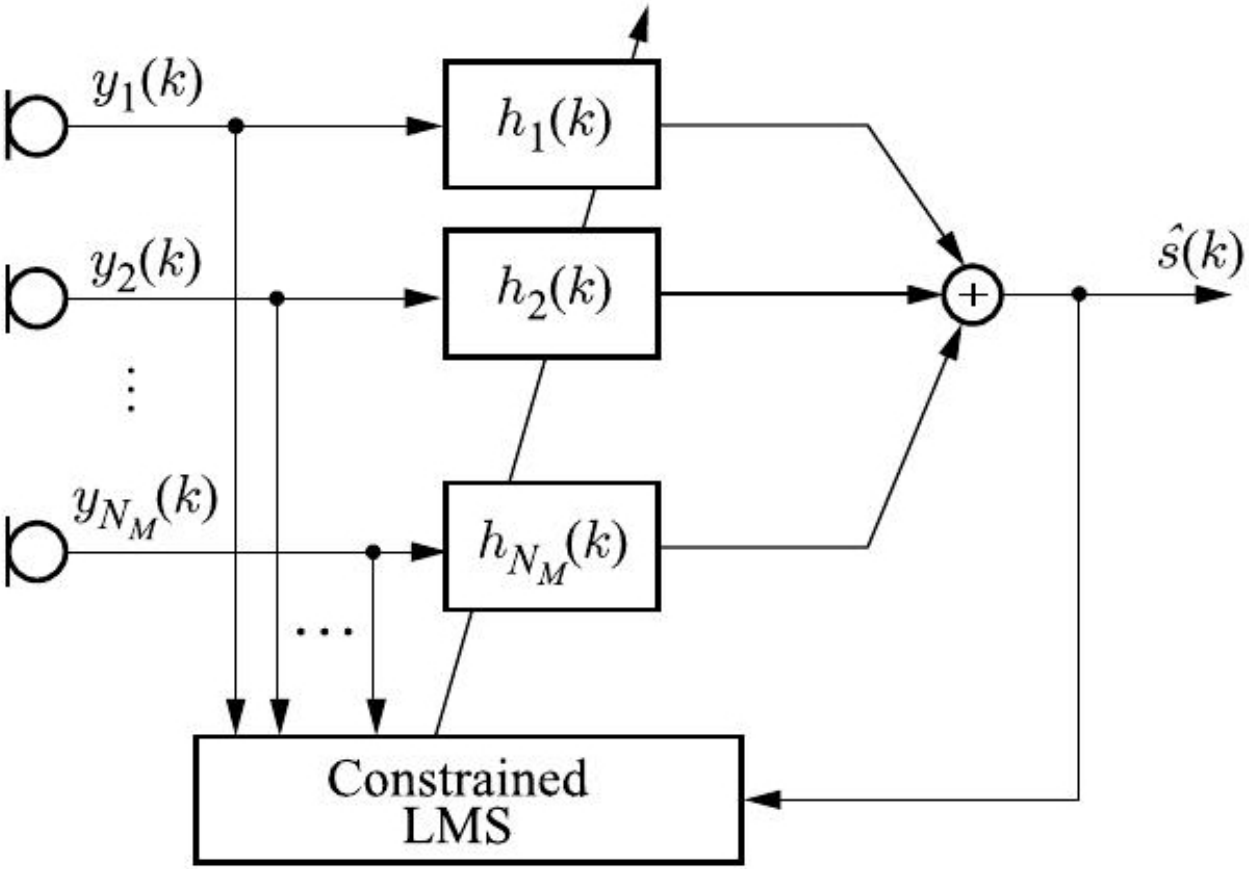

Figure: LCMV (Frost) Beamformer,[10], S.496

MVDR Beamformer

Der Minimum Variance Distortionless Response (MVDR) Beamformer, auch Capon-Beamformer genannt, ist ein Spezialfall des LCMV Beamformers.

Die Leistung des Ausgangssignals wird bezüglich *einer einzigen* linearen Bedingung (constraint) minimiert:

$$\begin{aligned}\mathbf{w}_{\text{MVDR}} &= \underset{\mathbf{w}}{\text{argmin}}\ \mathbf{w}^H \boldsymbol{\Phi}_{\mathbf{xx}} \mathbf{w} \text{ bzgl.} \mathbf{w}^H \mathbf{h}_0 = 1 \\ &= \frac{\boldsymbol{\Phi}_{\mathbf{xx}}^{-1} \mathbf{h}_0}{\mathbf{h}_0^H \boldsymbol{\Phi}_{\mathbf{xx}}^{-1} \mathbf{h}_0}.\end{aligned} \tag{7}$$

Bei unkorrelierten Eingangssignalen mit Mittelwert Null:

$$\mathbf{w}_{\text{MVDR}} = \frac{\boldsymbol{\Phi}_{\mathbf{vv}}^{-1} \mathbf{h}_0}{\mathbf{h}_0^H \boldsymbol{\Phi}_{\mathbf{vv}}^{-1} \mathbf{h}_0}. \tag{8}$$

MVDR Beamformer

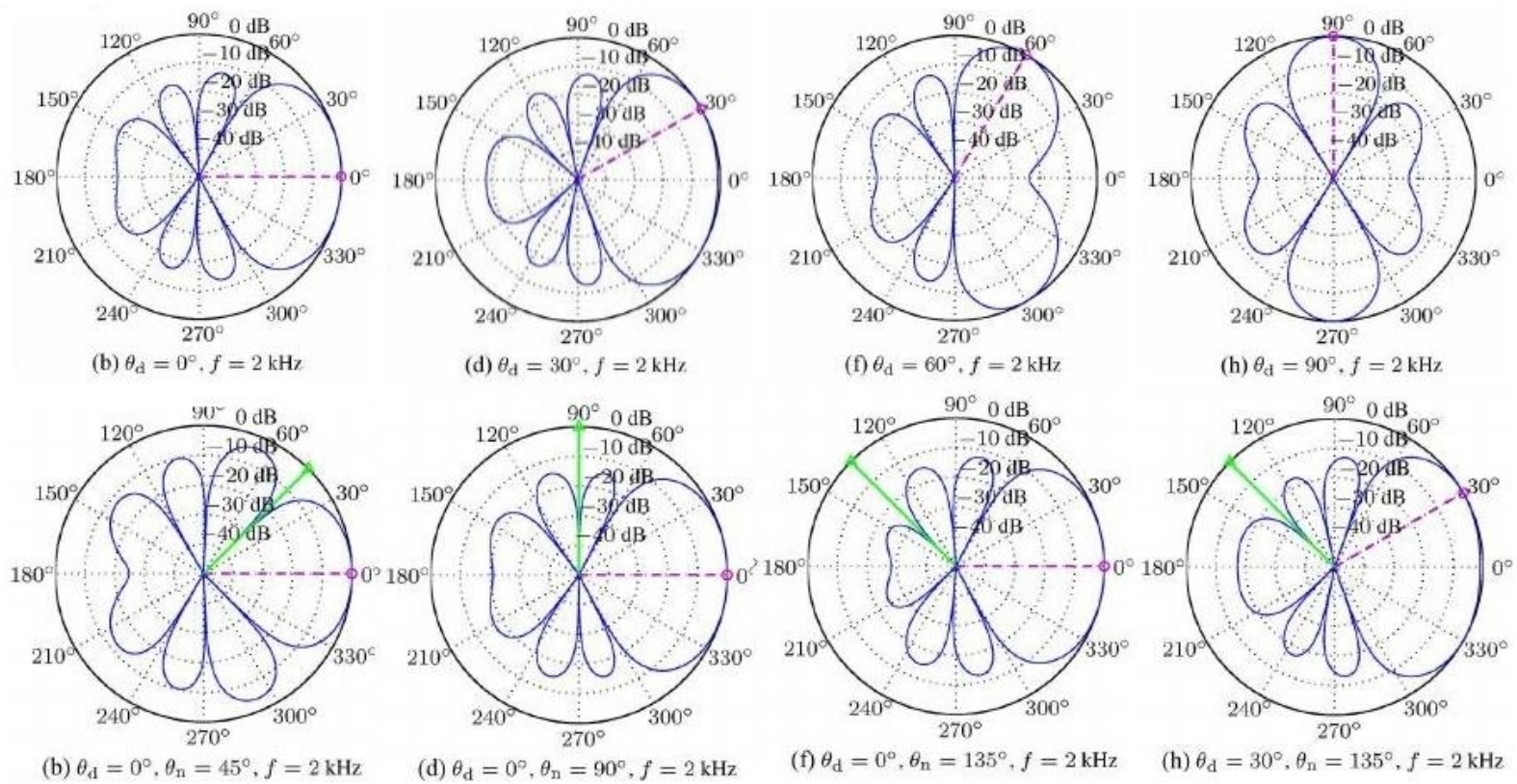

Figure: Beampattern des MVDR Beamformers in spatially white noise (oben) und point source + white noise (unten),[11], S. 71-73 & 75

MVDR Beamformer

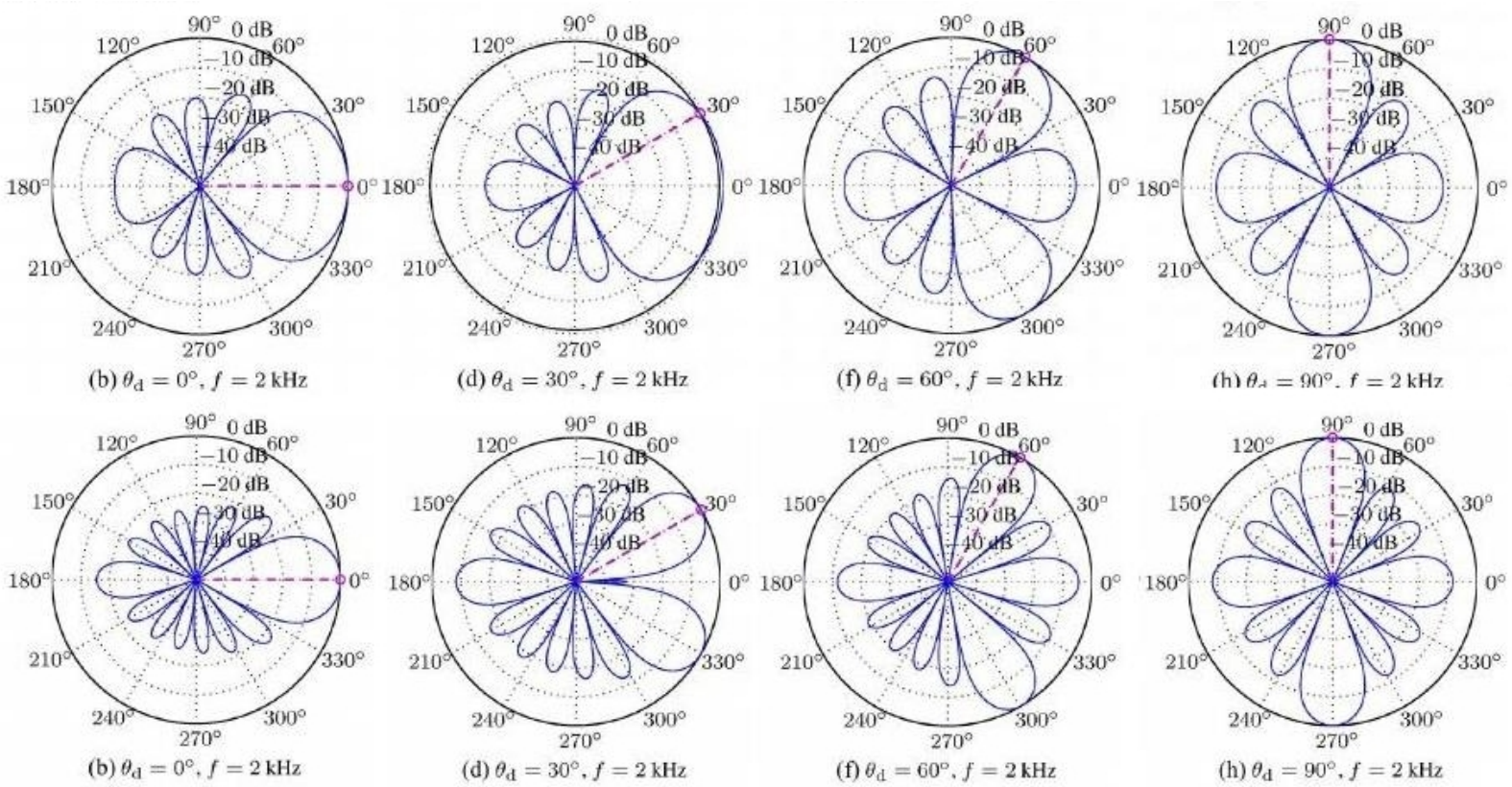

Figure: Beampattern des MVDR Beamformers in diffuse noise (oben) und diffuse noise + white noise (unten),[11], S. 71-73 & 75

Generalized Sidelobe Canceler (GSC)

GSC trennt einen Beamformer in zwei Komponenten auf:

$$\mathbf{w} = \mathbf{w}_0 - \mathbf{w}_n, \tag{9}$$

wobei:

- ein fixed Beamformer $\mathbf{w}_0$ die P auferlegten Beschränkungen einhält,
- und eine adaptive Komponente $\mathbf{w}_n$ bestehend aus einer $M-P$-dimensionalen Blocking Matrix $\mathbf{B}$, die orthogonal zu $\mathbf{w}_0$ ist, gefolgt von einem Rauschunterdrücker $\mathbf{g}$ (Noise Canceller), der das Restrauschen mit $M-P$ Freiheitsgraden adaptiv unterdrückt.

Generalized Sidelobe Canceler (GSC)

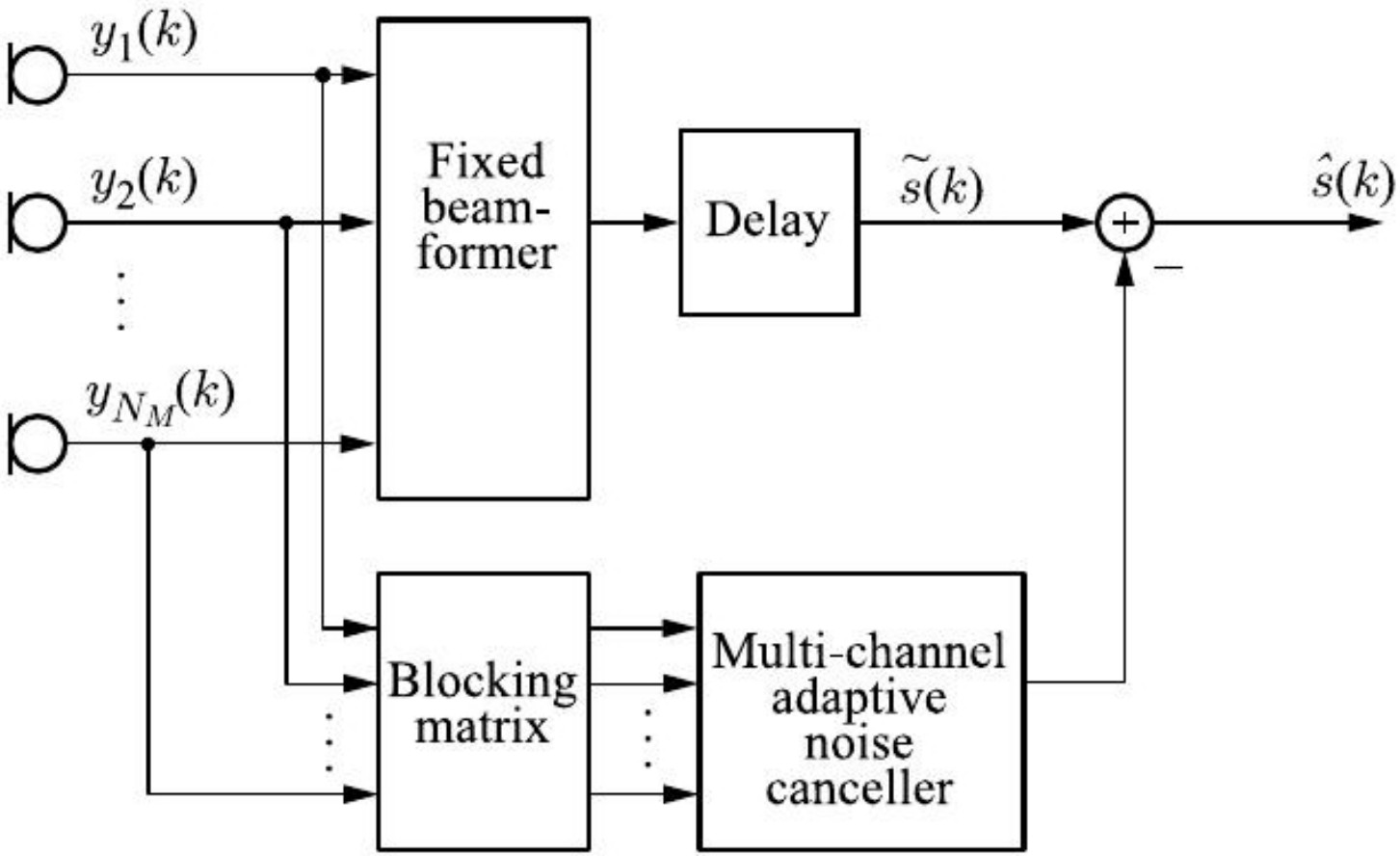

Figure: GSC Beamformer,[10], S.498

MVDR Beamformer Vor- und Nachteile

Vorteile (ggü. fixed/data-independent Beamformern):

- integriert Rauschen und Interferenzen in eine optimale Lösung.
- setzt Nullen in Richtung aller Störquellen.
- höhere räumliche Auflösung.
- Nebenkeulen sind kleiner und glatter.

Nachteile:

- Fehleranfällig.
- anfällig für Selbstnullung.
- genaue Kovarianzmatrix und Steuerungsvektor nicht immer verfügbar.
- Anpassung der Filterkoeffizienten erfordert Matrixinversion.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SP Signal Processing

Ausblick

Hürden

- Sprache ist nicht stationär und hat einen hohen Dynamikbereich
- Reflektionen der Schallwellen an der Umgebung
- Durch das Bewegen des Sprechers verändert sich die Umgebung
- Die Konfiguration des Mikrofon Arrays ist vorher nicht bekannt
- Mikrofon Arrays erfordern eine höhere Rechenleistung

Mögliche Ansätze

- Ad Hoc Mikrofon Array
- Neuronale Netze

Figure: [12]

Danke fürs Zuhören

Quellen I

[1] URL: https://www.computerbild.de/artikel/cb-News-Handy-Apple-Keine-Geraeuschunterdrueckung-mehr-beim-iPhone-13-31646279.html (visited on 05/01/2022).

[2] URL: https://www.aier.org/article/this-pandemic-is-not-homeschoolings-moment/ (visited on 05/01/2022).

[3] URL: https://www.stuttgarter-nachrichten.de/inhalt.flughafen-stuttgart-was-passiert-wenn-fluglotsen-ausfallen.5e25a08f-5b21-4576-8324-e65b2bc37dc3.html (visited on 05/01/2022).

[4] URL: https://www.mobil-bleiben.de/hilfsmittel/hoergeraete/hoergeraete-uebersicht/ (visited on 05/01/2022).

[5] URL: https://www.reddit.com/r/interestingasfuck/comments/kjw80j/acoustic_horns_invented_in_1939_to_identify/ (visited on 04/28/2022).

[6] URL: https://www.researchgate.net/figure/Simulated-beam-patterns-with-predefined-beam-direction-degrees-Three-left-side_fig2_337569387 (visited on 04/28/2022).

[7] URL: https://www.prosoundweb.com/acoustic-feedback-part-1-the-fundamentals/2/ (visited on 04/28/2022).

[8] E. Vincent, T. Virtanen, and S. Gannot. *Audio source separation and speech enhancement*. John Wiley & Sons, 2018.

[9] URL: https://www.radartutorial.eu/06.antennas/an60.en.html (visited on 05/01/2022).

Quellen II

[10] P. Vary and R. Martin. *Digital speech transmission: Enhancement, coding and error concealment*. John Wiley & Sons, 2006.

[11] C. Pan, J. Chen, and J. Benesty. "Performance Study of the MVDR Beamformer as a Function of the Source Incidence Angle". In: *Audio, Speech, and Language Processing, IEEE/ACM Transactions on* 22 (Jan. 2014), pp. 67–79.

[12] URL: https://youtu.be/TQnUF9kxfPE?t=1052 (visited on 05/02/2022).

[13] B. Van Veen and K. Buckley. "Beamforming: a versatile approach to spatial filtering". In: *IEEE ASSP Magazine* 5.2 (1988), pp. 4–24.

[14] S. Doclo et al. "Multichannel Signal Enhancement Algorithms for Assisted Listening Devices: Exploiting spatial diversity using multiple microphones". In: *IEEE Signal Processing Magazine* 32.2 (2015), pp. 18–30.