

Unterdrückung der Umgebungsgeräusche bei der akustischen Qualitätskontrolle

D. Filbert, M. Knaak
Institut für Mess- u. Automatisierungstechnik
TU Berlin, e-mail: dieter.filbert@tu-berlin

1 Einleitung

In der akustischen Qualitätskontrolle werden hauptsächlich zwei Schallereignisse, der Körperschall und der Luftschall zur Prüfung verwendet. Der Körperschall wird benutzt, weil er weitgehend Umgebungsgeräusche ausschließt, da der Sensor direkt am Meßobjekt befestigt ist. Gemessen wird im Allgemeinen die Beschleunigung oder der Schwingweg. Der Nachteil besteht aber darin, dass bestimmte Fehlerarten schlecht diagnostiziert werden können.

Der Luftschall wird mit Mikrofonen gemessen. Dabei wird im Allgemeinen der Schalldruck als Signalgröße verwendet. Um Umgebungsgeräusche weitgehend auszuschließen, wird häufig eine schalldämmende Kabine benutzt. Dies ist eine aufwendige und teure Vorgehensweise. Im Folgenden werden deshalb Verfahren angegeben, die den Störschall aus der Umgebung mit Hilfe von Signalverarbeitungsmethoden dämpfen.

2 Methoden der Störgeräuschunterdrückung

Die Methoden der Störgeräuschunterdrückung entstammen hauptsächlich dem Freisprechen beim Telefonieren. Daneben gibt es aber auch die Methoden der Blinden Quellentrennung und der Mustererkennung. Auch Hidden-Markov-Modelle werden zur Störunterdrückung verwendet [1].

2.1 Raumakustische Eigenschaften

Beim Entwurf eines Geräuschreduktionssystems liegt die Problematik darin, daß eine standardisierbare Geräuschsituation kaum vorliegt. Daher ist es nicht verwunderlich, daß es eine Vielzahl von Geräuschreduktionsverfahren gibt. Es ist also nötig, ein Verfahren zu entwickeln, das der Geräuschsituation in der Qualitätskontrolle angepaßt ist. Dabei ist zu klären, wie das zu prüfende Objekt als Schallquelle modellierbar ist. Ebenso sind die Störquellen und die Schallwege sinnvoll zu modellieren. Im Allgemeinen kann davon ausgegangen werden, dass das Testobjekt innerhalb des Nachhallradiuses liegt, d.h., dass die Mikrofone nahe am Testobjekt angeordnet sind. Weiterhin kann davon ausgegangen werden, dass die Störquellen weiter entfernt sind. Im Idealfall stellt der Schall des Testobjekts ein freies Schallfeld dar, und der Schall der Störquellen ist ein diffuses Schallfeld. Aufgrund der unterschiedlichen Eigenschaften dieser Schallfelder läßt sich der Störschall vom Nutzschaall trennen. Das real vorliegende Störschallfeld läßt sich in der Praxis exakt nicht angeben. Jedoch für den Grenzfall des diffusen Schallfeldes lassen sich einige quantitative Aussagen machen [2].

Ein diffuses Schallfeld liegt vor,

- wenn sich eine Punktquelle in einem Raum mit der Nachhallzeit $T_H \rightarrow \infty$ befindet. Durch die große Anzahl von wiederholten Schallreflektionen wird ein diffuses Schallfeld erzeugt.
- Wenn die Anzahl der Punktquellen gegen unendlich geht. Die Nachhallzeit ist beliebig. Die Punktquellen seien unabhängig voneinander und gleichmäßig auf den Raumbegrenzungsflächen verteilt. Durch die Vielzahl der Schallquellen wird auch bei geringer Nachhallzeit ein diffuses Schallfeld erzeugt.

Im Folgenden soll ausschließlich von Methoden mit Mehrkanalverarbeitung ausgegangen werden. Deshalb ist die Kreuzkorrelation der Störsignalkomponenten in den Mikrofonkanälen von besonderer Bedeutung. Für ein diffuses Schallfeld ist die normierte Kreuzkorrelation zwischen zwei Mikrofonsignalen bei Anregung mit der Frequenz ω

$$R_{ss}(d, \omega) = \frac{\sin\left(\omega \frac{d}{c}\right)}{\omega \frac{d}{c}} \cos(\omega\tau) \quad (1)$$

Darin ist d der Abstand zwischen den Mikrofonen, c ist die Schallgeschwindigkeit und τ die Zeitdifferenz des Eintretens des Schallereignisses an den beiden Mikrofonen. Eine Auswertung zeigt, daß die Kreuzkorrelation nur für höhere Frequenzen gering wird oder anders ausgedrückt, dass die Mikrofone einen gewissen Mindestabstand aufweisen müssen.

2.2 Beamforming und adaptive Filterung

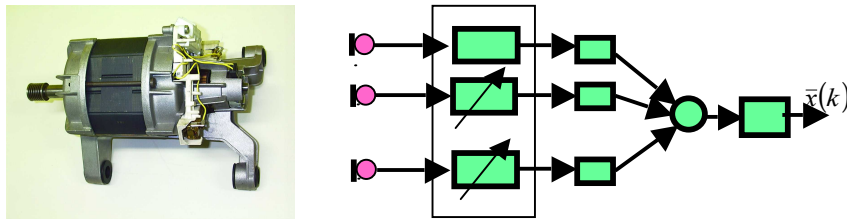


Bild 1: Aufbau des Prüfstands (Prinzip)

Der Aufbau des Prüfstandes richtet sich im Prinzip nach Bild 1. Ein Mikrofonarray nimmt die Schallsignale des Prüfobjekts und der Umgebungsgeräuschquellen auf. In einer Vorverarbeitungsstufe werden die Signale so verzögert, dass das Nutzsignal verstärkt und die Störsignale geschwächt werden. Ein nachgeschaltetes Filter sorgt für eine weitere Verbesserung der Störschallunterdrückung. Vereinfacht sollen die Verhältnisse, wie in Bild 2 dargestellt, mit zwei Mikrofonen angegeben

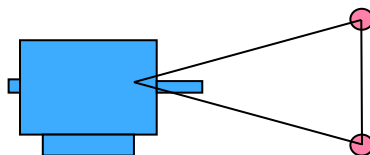


Bild 2: Laufzeiten

werden. Die Laufzeit ergibt sich zu

$$\tau_i = \frac{f_T}{c} \cdot r_{si} = \frac{f_T}{c} \cdot \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2 + (z_s - z_i)^2} \quad (2)$$

Im Mikrofon i entsteht eine Verzögerung $\tau_{1,i}$ gegenüber dem Mikrofon 1

$$\tau_{1,i} = \tau_1 - \tau_i = \frac{f_T}{c} \cdot d_{1,i}$$

Damit ergibt sich das Signalmodell mit den Schallsignalen s des Testobjekts

$$x_i(k) = \alpha_{1,i} \cdot s(k + \tau_{1,i}) + n_i(k) \quad (3)$$

$\alpha_{1,i}$ ist die relative Amplitudendämpfung der Schallsignale. Überlagert sind die Signale von dem Störschall $n_i(k)$. Die Signale der Mikrofone werden so verzögert, daß die Signale, die vom Testobjekt kommen, gleichphasig werden. Das Ausgangssignal des Beamformers ergibt sich damit zu

$$\bar{x}(k) = \frac{1}{M} \sum_{i=1}^M \alpha_i \alpha_{1,i} \cdot s(k + \tau_{1,i} - \hat{\tau}_{1,i}) + \frac{1}{M} \sum_{i=1}^M \alpha_i \cdot n_i(k - \hat{\tau}_{1,i}) \quad (4)$$

mit $\hat{\tau}_{1,i}$ = geschätzte Zeitverzögerung; M = Anzahl der Mikrofone.

Ausgehend von der Tatsache, daß der Schall des Testobjekts ein räumlich kohärentes Feld darstellt und dass die Störschalle sowohl räumlich kohärent als auch inkohärent sein können, folgt für das Ausgangssignal des Beamformers

$$\begin{aligned} \bar{x}(k) = \frac{1}{M} \sum_{i=1}^M \alpha_i \cdot s(k + \Delta\tau_i) + \frac{1}{M} \sum_{i=1}^M \alpha_i \cdot n_D(k + \tau_{n,1i} + \Delta\tau_i - \tau_{1i}) + \\ + \frac{1}{M} \sum_{i=1}^M \alpha_i \cdot n_{R,i}(k + \Delta\tau_i - \tau_{1i}) \end{aligned} \quad (5)$$

In dieser Gleichung werden in dem Signal $n_{R,i}(k)$ die Störsignale zusammengefasst, die auf den räumlich inkohärenten Schalleinfall zurückgehen und vom Mikrofon i empfangen werden. Dagegen beschreibt $n_D(k)$ das Signal einer beliebig im Raum angeordneten Störsignalquelle, von der Schallanteile räumlich kohärent auf das Mikrofonarray einfallen. Dies können auch Signale sein, die durch Reflexion des Nutzschalls entstehen. τ_{1i} bezeichnet die Laufzeitdifferenzen der empfangenen Störsignale und $\Delta\tau_i$ den Laufzeitschätzfehler. $\tau_{n,1i}$ ist die Verzögerungszeit von Schallen aus entfernten Quellen.

Die Störunterdrückung von kohärentem Schall ergibt sich über das Verhältnis der Signalleistung am Ausgang zur Signalleistung am Eingang

$$NR = -10 \log \left(\frac{\sigma_{\bar{n}_D}^2}{\sigma_{n_D}^2} \right) \quad NR = -10 \log \left(\frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \frac{\alpha_i \alpha_j R_{n_D n_D}(\tau_{n,ij} + \Delta\tau_{ij} - \tau_{ij})}{R_{n_D n_D}(0)} \right) \quad (6)$$

$R_{n_D n_D}$ sind die Kreuzkorrelationsfunktionen zwischen den einzelnen Mikrofonen. Werden die Störsignale aus räumlich kohärentem Schalleinfall durch ein Sinussignal beliebiger Frequenz charakterisiert, so folgt für die Störsignaldämpfung am Ausgang des Beamformers

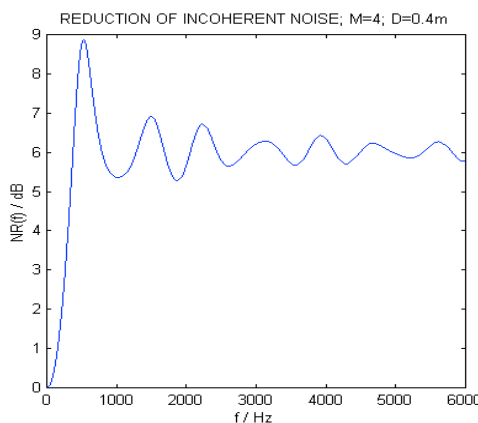
$$NR = -10 \log \left(\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j \cos \left(\frac{2\pi f}{f_T} (\tau_{n,ij} + \Delta\tau_{ij} - \tau_{ij}) \right) \right) \quad (7)$$

Aus dieser Herleitung ergeben sich die bekannten Richtcharakteristiken, die frequenzabhängig und über die Gewichtungsfaktoren α zu beeinflussen sind [3]. Eine interessante Möglichkeit besteht darin, das Richtdiagramm so zu entwerfen, dass eine räumlich feste Störquelle stark gedämpft wird. Für einen solchen Fall kann die Störunterdrückung theoretisch den Wert unendlich annehmen.

Die Störunterdrückung von inkohärentem Schall ergibt sich ohne Herleitung entsprechend

$$NR = -10 \log \left(\frac{1}{M^2} \sum_{i=1}^M \alpha_i^2 + \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1, j \neq i}^M \alpha_i \alpha_j \cdot \cos \left(\frac{2\pi f}{f_T} (\Delta\tau_{ij} - \tau_{ij}) \right) \text{sinc} \left(2\pi f \frac{d_{ij}}{c} \right) \right) \quad (8)$$

In Bild 3 ist die Störunterdrückung für ein quadratisches Mikrofonarray mit vier Mikrofonen im Abstand von 40 cm über der Frequenz aufgetragen. Zu erkennen ist, daß es eine untere Grenzfrequenz gibt. Für hohe Frequenzen nähert sich die Dämpfung immer mehr dem Wert $10 \log(M)$. Dies gilt unter der Voraussetzung, daß die Gewichtungsfaktoren $\alpha = 1$ sind.



Als Fazit aus diesen Untersuchungen ergibt sich, dass die Störunterdrückung mit Hilfe eines Beamformers allein nicht sehr groß ist. Deshalb wird ein zusätzliches Nachfilter eingesetzt. Hierzu hat sich im wesentlichen das Wiener Filter bewährt [4]. Das Wiener Filter minimiert die mittleren quadratischen Fehler des geschätzten Nutzsignals. Für diese Filterklasse ergibt sich aus der Fouriertransformierten der Wiener-Haupt-Gleichung

$$R_{ss}(k) = \sum_{m \in \mathbb{Z}} h_{opt}(m) \cdot R_{xx}(k-m) \quad (9)$$

direkt die Übertragungsfunktion des Wiener Filters

zu

$$H_{opt}(f) = \frac{P_{ss}(f)}{P_{xx}(f)} \quad (10)$$

Darin bezeichnet $P_{ss}(f)$ das Autoleistungsdichtespektrum ungestörten Prozesses und $P_{xx}(f)$ das Autoleistungsdichtespektrum des gestörten Prozesses. Die Spektren müssen aus den Signalen $x(k)$ geschätzt werden. Die Schätzung des Leistungsdichtespektrums des gestörten Signals kann unproblematisch mit Standardverfahren erfolgen. Das Leistungsdichtespektrum des ungestörten Signals kann, wenn es stationär ist, aus Messungen bei völliger Ruhe der Umgebung ermittelt werden. Bei mehrkanaliger Messung ergibt sich eine weitere Möglichkeit. Dazu werden zunächst die Kreuzleistungsdichtespektren $P_{xi,j}$ jeweils zweier Kanäle berechnet.

$$P_{xi,j} = P_{ss} + P_{snj} + P_{sni} + P_{ni,nj} \quad (11)$$

Aus dem Mittelwert aller Kreuzleistungsdichtespektren $\bar{P}_{xi,j}$ ergibt sich das Autoleistungsdichtespektrum des ungestörten Signals $P_{ss}(f)$. Wenn die Störungen n nicht mit dem Nutzsignal s korreliert sind, verschwinden die beiden mittleren Terme und das Leistungsdichtespektrum der Störung $P_{ni,nj}$ verrin-

gert sich mit $1/\sqrt{M}$. Außerdem kann der Imaginärteil zu Null gesetzt werden, da $P_{ss}(f)$ reell sein muss und ein imaginärer Anteil nur aus dem Spektrum des Störgeräusches stammen kann. Somit wird $P_{ss}(f) \cong \bar{P}_{xi,j}$.

In praktischen Untersuchungen zeigte sich, daß die theoretische Obergrenze der Störsignaldämpfung nicht erreicht wird. Mit einer weissen Rauschquelle als Störsignal wurden unter Verwendung von 16 Mikrofonen nur 7,3 dB und mit 4 Mikrofonen nur 3,7 dB erreicht.

2.3 Blinde Quellentrennung

Bei der blinden Quellentrennung wird davon ausgegangen, daß durch n Mikrofone die Signale $x(k)$ aufgenommen wurden, die durch eine Mischung von n Quellen $s(k)$ entstanden sind. Das Ziel besteht in der Rekonstruktion der Quellensignale. Diese Methode kann für die Unterdrückung der Nebengeräusche verwendet werden, wenn eine Quelle als das zu prüfende Objekt die anderen als Störquellen angesehen werden.

Wie der Name blinde Quellentrennung sagt, sind dazu keine modellhaften Information weder über die Quellensignale noch über die Übertragungsstrecke bzw. die Mischung erforderlich. An die Signale wird lediglich die Anforderung gestellt, daß sie statistisch unabhängig sind. Diese Anforderung kann als gewährleistet angenommen werden, wenn die Signale einen unterschiedlichen physikalischen Ursprung haben.

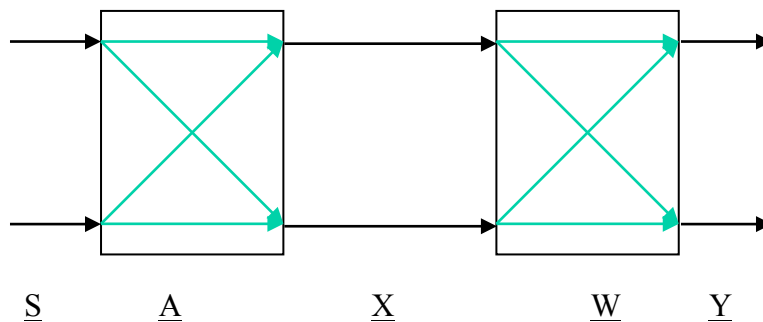


Bild 4: Mischung und Entmischung von Signalen

Als Modell der Mischung wird die Übertragungsstrecke durch einen Operator \mathbf{A} beschrieben, wie es Bild 4 verdeutlicht. Darin wird der Vektor \mathbf{s} durch den Operator \mathbf{A} auf den Vektor \mathbf{x} abgebildet (Mischung)

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (12)$$

Im einfachsten Fall enthält die Matrix skalare Elemente, d.h. die Mischung entsteht durch die Überlagerung der Quellensignale mit unterschiedlicher Gewichtung. In der Akustik sind die Elemente von \mathbf{A} Filterfunktionen, die auch die Überlagerungen mit den reflektierten Quellensignalen und somit die Raumakustik berücksichtigen. Da sich eine Filterung im Zeitbereich durch eine Multiplikation im Frequenzbereich ausdrücken läßt, werden im Folgenden nur noch skalare Mischmatrizen beschrieben [5]. Die blinde Quellentrennung wendet die Entmischungsmatrix \mathbf{W} als Operator auf den Vektor der Meßsignale an, um damit die Signale $\mathbf{y}(k)$ zu erhalten.

$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (13)$$

Gesucht wird nun ein Operator \mathbf{W} , der garantiert, dass die Signale $\mathbf{y}(t)$ ebenfalls statistisch unabhängig sind. Dann kann eine Identität zwischen \mathbf{y} und \mathbf{s} (ohne Berücksichtigung einer Skalierung und einer Permutation) gezeigt werden [6]. Also wurde der Signalvektor \mathbf{s} rekonstruiert. Es gilt damit $\mathbf{I} = \mathbf{A} \cdot \mathbf{W}$, wobei Permutationen der Einheitsmatrix zugelassen werden.

Die statistische Unabhängigkeit kann über statistische Verbundmomente (Kumulanten) gemessen werden. Kumulanten sind statistische Kennwerte, die Aussagen über die Form der Wahrscheinlichkeitsdichtefunktion machen. Nach [7] verschwinden alle Kumulanten jeder Ordnung für statistisch unabhängige Signale.

Das Ziel besteht also darin, durch eine geeignete Wahl von \mathbf{W} die Kumulanten zu minimieren.

In einem ersten Schritt werden die Kumulanten zweiter Ordnung (Kovarianzen) minimiert und die Signale damit dekorreliert. Dazu muß der Vektor \mathbf{x} in Richtung der größten Energieanteile (Hauptachsen) gedreht werden. Dieser Schritt wird auch als Hauptachsentransformation oder Karhunen-Loeve-Transformation bezeichnet. Das Problem kann als Eigenwertproblem aufgefaßt werden. Es muß also ein orthonormaler Operator \mathbf{U} gefunden werden, der folgendes Eigenwertproblem löst

$$\mathbf{R}_{xx} \cdot \mathbf{u}_i = \lambda_i \cdot \mathbf{u}_i \quad (14)$$

Darin ist $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^T\}$ die Kovarianzmatrix zwischen den Komponenten von $\mathbf{x}(k)$. Wie in [8] gezeigt ist, ist \mathbf{R} für verbunden stationäre Prozesse symmetrisch. Damit existieren orthogonale Eigenräume, so dass sich \mathbf{R}_{xx} folgendermaßen zerlegen läßt

$$\mathbf{R}_{xx} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^T \quad (15)$$

Darin hat \mathbf{U} die Eigenvektoren \mathbf{u}_i als Spalten, $\mathbf{\Lambda}$ ist eine Diagonalmatrix mit λ_i als Elemente. Letztlich werden die Varianzen auf $\sigma^2 = 1$ normiert (Weißes). Die Ausgangssignale ergeben sich zu

$$\mathbf{y}_z = \mathbf{\Lambda}^{-\frac{1}{2}} \cdot \mathbf{U}^T \cdot \mathbf{x} \quad (16)$$

und sind statistisch in zweiter Ordnung unabhängig (dekorreliert).

$$\mathbf{R}_{yy} = E\{\mathbf{y}_z \cdot \mathbf{y}_z^T\} = \mathbf{\Lambda}^{-\frac{1}{2}} \cdot \mathbf{U}^T \cdot \mathbf{R}_{xx} \cdot \mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I} \quad (17)$$

Die Dekorrelation erreicht jedoch noch keine Trennung von nicht-gaußschen Signalen. Die statistische Unabhängigkeit ist gewährleistet, wenn sich die Verbundwahrscheinlichkeitsdichte in das Produkt der partiellen Wahrscheinlichkeitsdichten aufspalten läßt. Das wird mit Dekorrelation nur bei Signalen erreicht, die sich vollständig durch Statistik zweiter Ordnung beschreiben lassen. Anderenfalls müssen Korrelationen höherer Ordnung ebenfalls entfernt werden. Korrelation höherer Ordnung werden häufig durch Kumulanten vierter Ordnung (Kurtosis) detektiert. Die Kumulanten dritter Ordnung (Schiefe) sind bei symmetrischen Signalen Null und somit für die Bestimmung der statistischen Abhängigkeit nicht geeignet

Da die Kumulanten höherer Ordnung nur sehr fehlerhaft geschätzt werden können, kann auch auf andere Verfahren zurückgegriffen werden, die den unterschiedlichen spektralen Gehalt der Signale ausnutzen. In diesem Fall wird ausgenutzt, daß Quellsignale für beliebige Zeitverzögerungen unabhängig sind:

$$E\{s_i(k)s_j(k+\tau)\} = 0 \quad (18)$$

Weiterhin gilt unter der Voraussetzung, daß die Signale nicht weiß sind:

$$E\{s_i(k)s_i(k+\tau)\} \neq 0 \quad (19)$$

Das heißt, dass durch eine weitere Transformation \mathbf{O}^T erreicht werden muss, dass die Kovarianzmatrizen $\mathbf{R}_{yy}(\tau) = E\{y(k) \cdot y^T(k+\tau)\}$ für alle Zeitverzögerungen τ die Form einer Diagonalmatrix erhalten. Daraus ergibt sich ein erweitertes Eigenwertproblem

$$\mathbf{R}_{yy}(\tau) = \mathbf{O}\mathbf{\Lambda}\mathbf{O}^T \quad (20)$$

Zur Lösung dieses Problems bietet sich die Verbunddiagonalisierung mit erweiterten Jacobi-Rotationen an [9].

3 Quellenverzeichnis

- [1] Ephraim, Y., et.al.: On the application of hidden Markov models for enhancing noisy speech. Int. Conf. Acoust., Speech, Signal Proc., Proceed., New York, 1988, S. 533-536
- [2] M. Drews: Mikrofonarrays und mehrkanalige Signalverarbeitung zur Verbesserung gestörter Sprache. Dissertation, TU Berlin, 1999.
- [3] R. Zielinski: Mikrofon-Arrays mit superdirektiven Eigenschaften zur Sprachsignalübertragung. Frequenz 50 (1996), S.198-204.
- [4] R. Le Bouquin-Jeannès, A. Akbari Azirani, G. Faucon: Enhancement of Speech Degraded by Coherent and Incoherent Noise Using a Cross-Spectral Estimator, IEEE Trans. Speech and Audio Processing 5 (1997), S. 484-487.
- [5] R. Lambert: Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures, Ph.D. thesis, Univ. of Southern Calif., San Diego, 1996
- [6] P. Comon: Independent component analysis, a new concept? Signal Processing, 36: S. 287-314, 1994
- [7] C. Nikias, A. Petropulu: High-Order Spectra Analysis, Prentice Hall, Englewood Cliff, 1993
- [8] S. Haykin: Adaptive Filter Theory, MacMillan Publishing, Englewood Cliffs, 1996
- [9] K. Abed-Meraim, A. Belouchrani, J.F. Cardoso, E. Moulines: Asymptotic performance of second order blind source separation, Proc. of ICASSP, (4), S. 277-284, 1994

