Verfahren zur Multikanal-Echokompensation in immersiv verknüpften Räumen

Marcel Nophut[1], Robert Hupke[1], Stephan Preihs[1], Jürgen Peissig[1]
[1] *Leibniz Universität Hannover, Institut für Kommunikationstechnik*
Appelstr. 9A, 30167 Hannover, Email: {marcel.nophut}@ikt.uni-hannover.de

Abstract

Im Rahmen des vom Bundesministerium für Wirtschaft und Energie geförderten Projekts „LIPS – Live Interactive PMSE Services“ arbeiten die Projektpartner aus Industrie und Forschung an einer immersiven audiovisuellen Verbindung zwischen entfernten Räumlichkeiten, die es Menschen erlaubt in einer möglichst natürlichen Art und Weise miteinander kommunizieren und sogar musizieren zu können. Dazu werden die Schallquellen im einen Raum aufgenommen, die Signale mit geringer Latenz übertragen und im jeweils anderen Raum über ein Multikanal-Lautsprechersetup wiedergegeben. Eine bidirektionale akustische Verbindung von Räumen erzeugt jedoch eine Feedback-Schleife, die akustische Echos oder Rückkopplungen hervorruft. Um diesem Problem zu begegnen nutzen bestehende mono- und stereophonische Systeme häufig eine adaptive Echokompensation, die die akustischen Übertragungsfunktionen im Raum schätzt, um so die Echos der Lautsprechersignale aus den aufgenommenen Mikrofonsignalen herauszufiltern. Bei Mehrkanalsystemen ist dieses Problem aufgrund der Korrelation der Lautsprechersignale in der Regel nicht eindeutig lösbar. Mit wachsender Anzahl der Kanäle tritt dieses Phänomen, das sogenannte „Non-Uniqueness-Problem“, immer stärker zutage, was zu einer höheren Fehlanpassung der Schätzung führt. Eine vorgeschaltete Dekorrelation der Lautsprechersignale wirkt diesem Problem entgegen und führt zu einer Verbesserung der Schätzergebnisse. Dieser Beitrag stellt einige gängige Algorithmen und Methoden der Multikanal-Echokompensation vor und vergleicht deren Leistungsfähigkeit anhand von Simulationen mit Aufnahmen aus einem realitätsnahen Modellaufbau.

Einleitung

Das Prinzip der akustischen Echokompensation, engl. Acoustic Echo Cancellation (AEC), ist vom einkanaligen Fall seit Jahren bestens bekannt. Mithilfe eines adaptiven Filters wird das im *Near-End Room* wiedergegebene Lautsprechersignal aus dem Mikrofonsignal herausgefiltert, sodass es nicht wieder zurück in den *Far-End Room* übertragen wird und so der Echopfad unterbunden wird.

Gegenüber dem einkanaligen Fall bringt der Multikanal-Fall (siehe Abbildung 1) mit mehreren Lautsprechersignalen eine zusätzliche fundamentale Schwierigkeit mit sich: Da die Lautsprechersignale in der Regel stark miteinander korreliert sind, ist die Lösung des Problems für das adaptive Filter nicht eindeutig zu bestimmen. Im Falle der Konvergenz auf eine von der gewünschten Lösung unterschiedliche Filterantwort kann die Energie

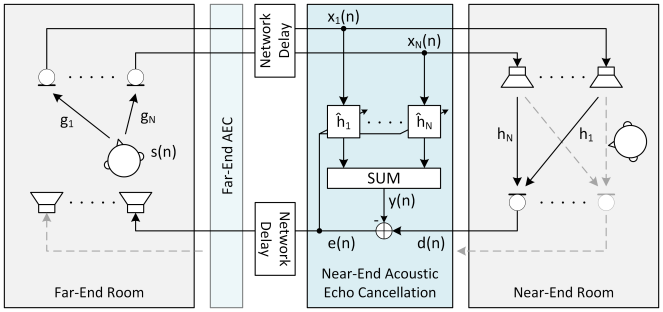

Abbildung 1: Prinzipskizze MCAEC.

des Fehlersignals gering sein, jedoch ist die Fehlanpassung an die Echopfade im Near-End Room gleichzeitig meist sehr groß. Dieses Phänomen wird als *Non-Uniqueness-Problem* bezeichnet.

Dabei sind aus der Literatur zwei grundsätzliche Ansätze bekannt diesem Problem zu begegnen [1]:

- Berücksichtigung der Auto- und Kreuzkorrelationen der Lautsprechersignale innerhalb der adaptiven Filter-Algorithmen mithilfe von Kovarianzmatrizen der Eingangssignale,
- Partielle Dekorrelation der Lautsprechersignale durch einen vorgeschalteten Verarbeitungsschritt.

Für die Multichannel Acoustic Echo Cancellation (MCAEC) wurden in der Vergangenheit schon zahlreiche Verfahren vorgeschlagen und untersucht.

Algorithmen für MCAEC

Die Familie der *Sparse Adaptive Filter* nähert die zu schätzenden Raumimpulsantworten als dünnbesetzte Filterantwort an und nutzt diese Eigenschaft für die Adaption aus. Bekannte Vertreter aus dieser Gruppe sind der Proportionate-Normalized-Least-Mean-Squares-Algorithmus (PNLMS) [2] und der Improved-PNLMS (IPNLMS) [3]. Diese Algorithmen betrachten jedoch nicht die Auto- und Kreuzkorrelationen der Lautsprechersignale. Der *Recursive-Least-Squares-Algorithmus* (RLS) benutzt eine Kovarianzmatrix der Lautsprechersignale im Zeitbereich um die Auto- und Kreuzkorrelationen zu berücksichtigen. Jedoch ist der Rechenaufwand insbesondere bei vielen Kanälen hoch, da die Matrix explizit oder implizit invertiert werden muss. Beim *Frequency-Domain-Adaptive-Filtering-Algorithmus* (FDAF) wird die Kovarianzmatrix der Lautsprechersignale durch die DFT annähernd diagonalisiert. Dadurch kann die Invertierung sehr viel effizienter durchgeführt

werden [4]. Auch die Filterung im Frequenzbereich wirkt sich positiv auf die Recheneffizienz aus. Diese Eigenschaften machen den FDAF sehr attraktiv für die Anwendung der MCAEC.

Ein weiterer vielversprechender Ansatz ist das *Wave-Domain Adaptive Filtering* (WDAF). Hierbei werden Lösungen der Wellengleichung für die Signaldarstellung genutzt, wobei die Filterantwort in der Wave-Domain den Zusammenhang von idealem zum tatsächlichen Schallfeld beschreibt [5, 6]. So kann eine deutliche Dimensionsreduktion erreicht werden. Dies macht das WDAF attraktiv in Kombination mit der Wellenfeldsynthese oder Ambisonics. Jedoch ist der Ansatz nicht auf beliebige Lautsprechersetups und Schallfelder übertragbar und wurde deshalb in der vorliegenden Untersuchung nicht betrachtet.

Für die partielle Dekorrelation wurden ebenfalls zahlreiche Verfahren vorgestellt. Dazu zählt das Hinzufügen von unkorreliertem Rauschen [7] oder von Nichtlinearitäten [8] zu den Lautsprechersignalen oder auch perzeptiv motivierte Ansätze [9]. Durch diese Verfahren wird die Robustheit gegenüber dem Non-Uniqueness-Problem erhöht, jedoch kann auch die Audioqualität der Signale beeinträchtigt werden.

Um die Leistungsfähigkeit von adaptiven Filtern zur Echokompensation zu bewerten, haben sich zwei Metriken in der Literatur durchgesetzt. Das *Echo Return Loss Enhancement* (ERLE) gibt das Verhältnis der Signalleistung des Mikrofonsignals d zur Signalleistung des Fehlersignals e an und dient damit als Maß der Echokompensation. Es berechnet sich über:

$$ERLE = 10\,log_{10}\left(\frac{\sigma_d^2}{\sigma_e^2}\right). \tag{1}$$

Die *System-Distance* (SD) dient als Maß, das die Fehlanpassung der Filterantwort des adaptiven Filters $\hat{\mathbf{h}}_p$ an die originale Impulsantwort $\mathbf{h}_p$ beschreibt und ist definiert durch:

$$SD = 10\,log_{10}\left(\frac{\sum_p \|\mathbf{h}_p - \hat{\mathbf{h}}_p\|^2}{\sum_p \|\mathbf{h}_p\|^2}\right). \tag{2}$$

Hierbei ist p der Index der Lautsprechersignale.

Obwohl das Themenfeld der MCAEC seit vielen Jahren beforscht wird und es in der Literatur schon viele Untersuchungen dazu gibt, unterscheiden sich diese Untersuchungen aber in den Rahmenbedingungen von der von uns angestrebten Anwendung. Für eine immersive Verknüpfung zweier Musiker an entfernten Orten sind sowohl deutlich mehr als zwei Lautsprecherkanäle als auch eine Abtastrate von mehr als 16 kHz nötig. Da die meisten Veröffentlichungen auf dem Gebiet der MCAEC aber auf Sprachkommunikation ausgerichtet sind, werden diese Anforderungen oft nicht beachtet.

Ergebnisse

Aus den oben genannten Gründen wurden eigene Untersuchungen angestrebt und dazu in zwei Laborräumen des Instituts für Kommunikationstechnik (IKT) ein anwendungsnahes Modell-Setup aufgebaut. Der Far-End Room war dabei mit einem Lautsprecher (Neumann KH120) als Schallquelle und vier Mikrofonen (Beyerdynamic MM1) in beliebiger Anordnung und einem Abstand von 1,5 bis 2 m zum Lautsprecher ausgestattet. Im Near-End Room befanden sich vier Lautsprecher des gleichen Typs, ebenfalls in beliebiger Anordnung, und ein Mikrofon. Der Abstand betrug ebenfalls 1,5 bis 2 m. Dabei war in diesem Aufbau jedes Mikrofon im Far-End Room auf genau einen Lautsprecher in Near-End Room geroutet. Es ergab sich also ein 4x1 MISO-System mit 4 Lautsprecherkanälen ($P = 4$). Die Nachhallzeit im Near-End Room beträgt etwa $T_{60} = 250$ ms.

In diesem Aufbau wurden Impulsantworten von den Lautsprechern zu den Mikrofonen gemessen, die dann für die Signalsynthese verwendet wurden. Als Quellensignal wurde ein Musiksignal mit einer Dauer von 10 s verwendet. Als zu untersuchende adaptive Filter-Algorithmen wurden drei klassische Ansätze der MCAEC gewählt: IP-NLMS, RLS und FDAF. Die Filterlänge betrug bei allen $L = 1024$. Die gemessenen Impulsantworten wurden vor der Signalsynthese auf definierte Längen zugeschnitten. Für den Far-End Room wurde die Länge auf $L_{IR,FER} = 8192$ festgelegt. Die Länge im Near-End Room wurde so gewählt ($L_{IR,NER} = 1024$), dass das Filter theoretisch eine perfekte Anpassung erreichen kann. Bei der Synthese der Mikrofonsignale wurde durch Hinzufügen von weißem gaußschen Rauschen ein Signal-Rauschabstand $SNR = 40$ dB eingestellt. Sowohl Double-Talk als auch die partielle Dekorrelation durch Vorverarbeitung wurde in den Untersuchungen nicht berücksichtigt.

Die Simulationsergebnisse sind in vier Plots dargestellt: ganz oben die aneinandergereihten Impulsantworten der vier Kanäle, wobei sowohl die originalen Impulsantworten als auch die letzte Schätzung am Ende der Simulation dargestellt ist, darunter die Zeitsignale von Mikrofon- und Fehlersignal (d bzw. e), darunter der Verlauf der SD über der Zeit und ganz unten der Verlauf des ERLE.

Für die Simulation mit dem IPNLMS-Algorithmus zeigt sich in Abbildung 2 folgendes Verhalten: Das ERLE liegt bei ca. 30 dB und zeugt somit von einer guten Echokompensation. Die Kurve der System-Distance offenbart jedoch eine recht hohe Fehlanpassung. Nach 10 s ist der Wert erst auf -7 dB gesunken. Das ist auf das Non-Uniqueness-Problem zurückzuführen. Dadurch können auch die erkennbaren Artefakte (vor dem Eintreffen des Direktschalls) in der geschätzten Impulsantworten erklärt werden.

Abbildung 3 zeigt die Simulationsergebnisse für den RLS-Algorithmus. Die Kurve der System-Distance zeugt von einem recht langsamen, aber guten Konvergenzverhalten. Die Echokompensation ist mit ca. 35 dB zufriedenstellend. In den geschätzten Impulsantworten zeigen sich jedoch auch hier noch kleine Artefakte. Hierbei ist zu beachten, dass das Non-Uniqueness-Problem auch hier Auswirkungen zeigt, obwohl der RLS Auto- und Kreuzkorrelationen der Lautsprechersignale berücksichtigt.

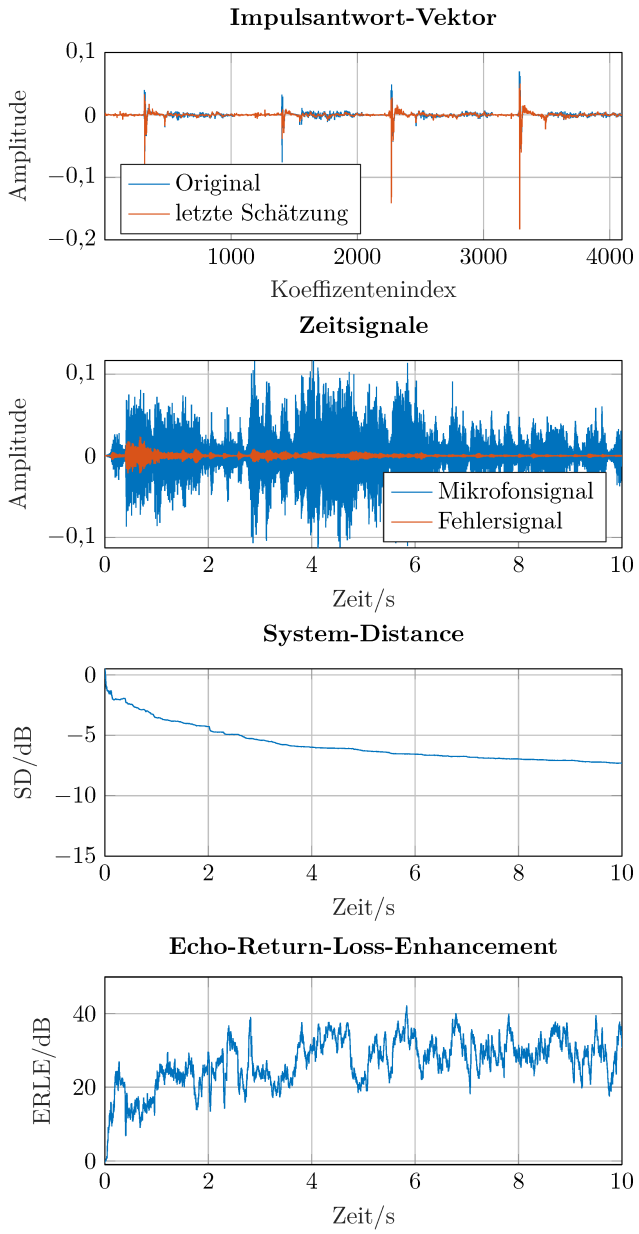

Abbildung 2: Simulationsergebnisse IPNLMS.

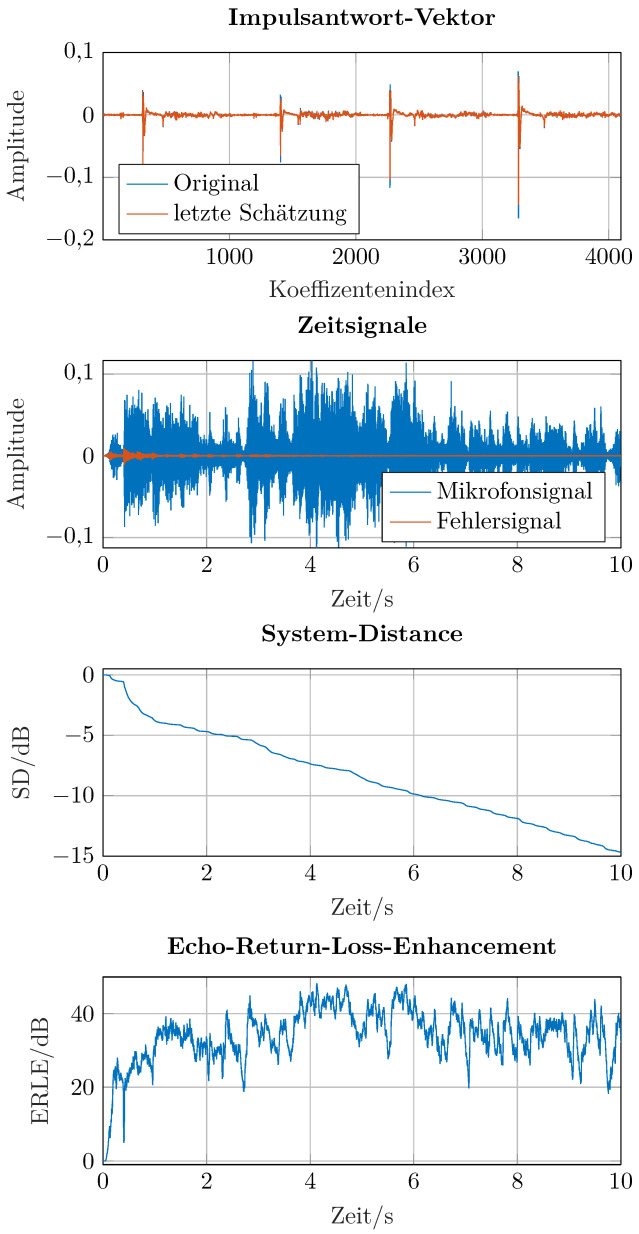

Abbildung 3: Simulationsergebnisse RLS.

In Abbildung 4 sind die Simulationsergebnisse für den FDAF zu sehen. Die Echokompensation ist ähnlich gut wie beim RLS-Algorithmus und auch das Konvergenzverhalten ist zunächst vielversprechend. Nach einigen Sekunden kommt der Algorithmus jedoch vom ursprünglichen Kurs ab und die SD beträgt nach 10 s nur etwa −9 dB. Der Grund für dieses Verhalten ist noch unbekannt.

Zusammenfassung und Ausblick

Es konnte gezeigt werden, dass auch Verfahren, die die Auto- und Kreuzkorrelationen der Lautsprechersignale berücksichtigen, das Non-Uniqueness-Problem nicht ohne weitere Maßnahmen vollständig lösen können. Das deckt sich mit Aussagen aus der Literatur [10]. Der Nachhall im Far-End Room unterstützt die Dekorrelation der Lautsprechersignale und verbessert damit das Konvergenzverhalten der Algorithmen. Neben dem Non-Uniqueness-Problem ist die enorme Anzahl von zu schätzenden Koeffizienten und der daraus resultierende erheblichen Rechenaufwand eine große Herausforderung bei der MCAEC.

In zukünftigen Untersuchungen soll die Kanalanzahl weiter erhöht werden und Verfahren zur partiellen Dekorrelation einbezogen werden. Des weiteren soll die Erweiterung des FDAF zum State-Space-FDAF [11] untersucht werden und durchgängiger Double-Talk betrachtet werden. Wichtig ist dabei auch die Einbeziehung von mehr Vorwissen über die Lautsprechersignale und das zu schätzende akustische System.

Förderung

Das Projekt LIPS ist gefördert vom Bundesministerium für Wirtschaft und Energie unter dem Förderkennzeichen 01MD18010G.

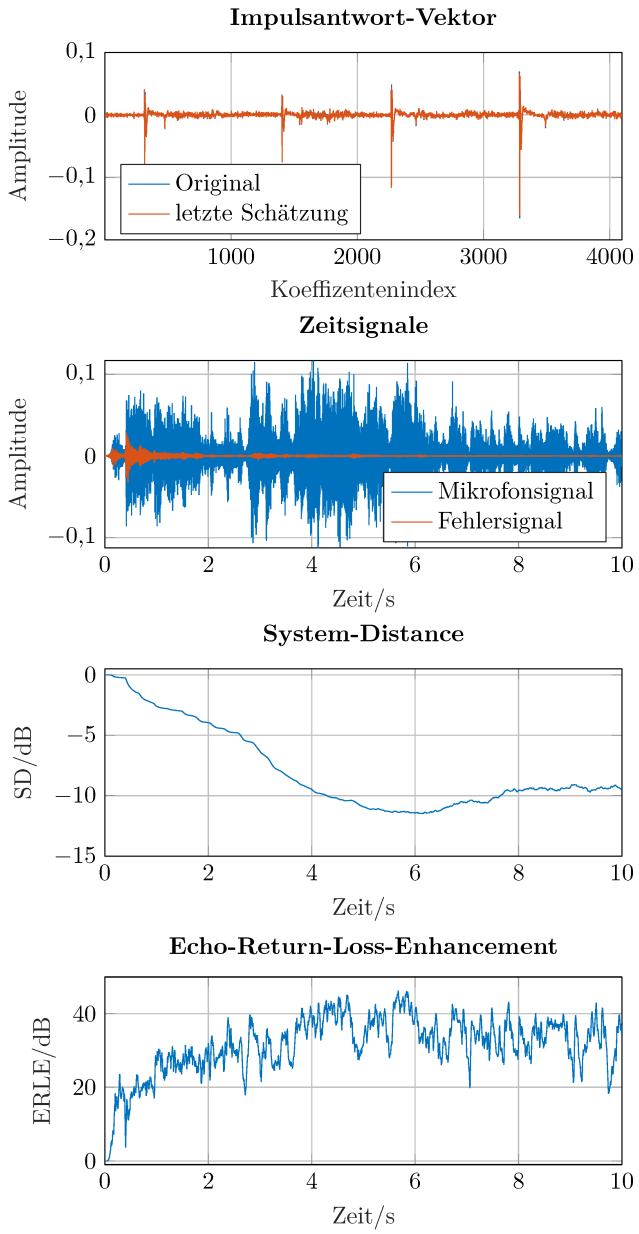

Abbildung 4: Simulationsergebnisse FDAF.

Literatur

[1] Gerald Enzner, Herbert Buchner, Alexis Favrot, and Fabian Kuech, "Acoustic echo control," in *Image, video processing and analysis, hardware, audio, acoustic and speech processing*, vol. 4 of *Academic Press Library in Signal Processing*, pp. 807–877. Acad. Press/Elsevier, Amsterdam, 2014.

[2] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, 2000.

[3] Jacob Benesty and Steven L. Gay, "An improved PNLMS algorithm," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Piscataway, NJ, 2002, pp. II–1881–II–1884, IEEE.

[4] Herbert Buchner, Jacob Benesty, and Walter Kellermann, "Multichannel frequency-domain adaptive filtering with application to acoustic echo cancellation," in *Adaptive Signal Processing*, Signals and Communication Technology, pp. 95–129. Springer, Berlin and Heidelberg, 2003.

[5] H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: acoustic echo cancellation for full-duplex systems based on wave-field synthesis," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Piscataway, N.J, 2004, pp. iv–117–iv–120, IEEE.

[6] Martin Schneider and Walter Kellermann, "Large-scale multiple input/multiple output system identification in room acoustics," in *Proceedings of Meetings on Acoustics, Vol. 19*, 2013, vol. 19, p. 015022.

[7] T. Gansler and P. Eneroth, "Influence of audio coding on stereophonic acoustic echo cancellation," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Piscataway, NJ, 1998, pp. 3649–3652, IEEE Service Center.

[8] D. R. Morgan, J. L. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 686–696, 2001.

[9] Jurgen Herre, Herbert Buchner, and Walter Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007*, Piscataway, NJ, 2007, pp. I–17–I–20, IEEE Operations Center.

[10] Philipp Thune and Gerald Enzner, "Trends in adaptive miso system identification for multichannel audio reproduction and speech communication," in *8th International Symposium on Image and Signal Processing and Analysis (ISPA), 2013*, Piscataway, NJ, 2013, pp. 767–772, IEEE.

[11] Sarmad Malik and Gerald Enzner, "Recursive bayesian control of multichannel acoustic echo cancellation," *IEEE Signal Processing Letters*, vol. 18, no. 11, pp. 619–622, 2011.