

SCHRITTWEITENSTEUERUNG ZUR QUALITÄTSERHALTENDEN STEREO-ECHOUNTERDRÜCKUNG

Maximilian Gauger

Fachgebiet Theorie der Signale, Technische Universität Darmstadt
Merckstr. 25, D-64283 Darmstadt
gauger@t.tu-darmstadt.de

ÜBERBLICK

Die Unterdrückung akustischer Echos für Mono-Anwendungen kann als ein weitgehend gelöstes Problem angesehen werden. Werden jedoch an die Realitätsnähe einer Anwendung erhöhte Anforderungen gestellt, bietet sich der Einsatz von Stereo-Übertragungen an.

Im Stereofall ist jedoch durch das sog. Nichteindeutigkeitsproblem die korrekte Einstellung der verwendeten adaptiven Filter sehr viel schwieriger. Gängige Ansätze verzerren daher das Eingangssignal mehr oder weniger stark. Die vorliegende Arbeit versucht, diese Verzerrung zu minimieren oder ganz zu vermeiden, indem die Schrittweite der Adaption in Abhängigkeit der Eigenschaften des Eingangssignals vermindert wird.

1. EINLEITUNG

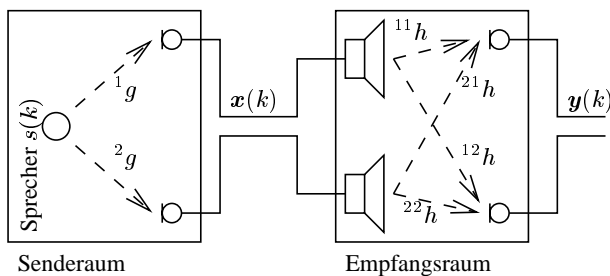


Fig. 1. Beispiel für ein Stereo-Konferenzsystem

Für die Verwendung einer Stereo-Echounterdrückung gibt es hauptsächlich zwei Einsatzfelder: Zum einen (Video-)Konferenzsysteme, bei denen zum Erreichen eines besseren räumlichen Eindrucks eine Stereoübertragung des Audiosignals wünschenswert ist, zum anderen sog. Command-and-Control-Systeme, bei denen der Benutzer Geräte mittels Sprachsteuerung anspricht. In beiden Fällen ist die Echounterdrückung unabdingbar: Bei Konferenzsystemen (Abb. 1) würde die Kommunikation durch das Sprecherecho ungemein behindert; bei Command-and-Control-Systemen würde die Erkennungsrate der Spracherkennung ohne Echounterdrückung durch das ins Mikrofon eingekoppelte Lautsprechersignal enorm verringert.

Gegenüber der Echounterdrückung für Monosignale ergeben sich bei Verwendung von Stereosignalen zwei Probleme: Zum einen genügt es nun nicht mehr, nur eine einzelne Impulsantwort zu schätzen, sondern je nach Anwendung müssen zwei oder vier

Impulsantworten gleichzeitig geschätzt werden, was zusätzliche Rechenleistung erfordert. Zum zweiten konvergiert der adaptive Algorithmus sehr langsam oder gar nicht, da die beiden Kanäle, die, wie in Abb. 1 gezeigt, aus der gleichen Quelle erzeugt wurden und daher stark korreliert, d.h. linear abhängig sind.

Um die Konvergenz des adaptiven Algorithmus zu verbessern, muss daher die lineare Abhängigkeit zwischen den beiden Kanälen verringert werden. Dazu wird in gängigen Ansätzen eine nichtlineare Verzerrung verwendet, wobei entweder die Phase oder die Amplitude verzerrt werden kann. Zur Phasenverzerrung kommt ein zeitvarianter Allpass zum Einsatz [1], wodurch allerdings der Stereoeffekt beeinträchtigt werden kann. Für die Amplitudenverzerrung hat sich als effektivstes und überraschender Weise am wenigsten hörbares Verfahren eine teilweise Gleichrichtung des Signals bewährt [2]. Dieses Verfahren beeinträchtigt zwar den Stereoeffekt nicht, hat aber negative Auswirkungen auf die Signalqualität.

Gerade für hochwertige Audiosignale, bei denen Verzerrungen erheblichen Einfluss auf die Signalqualität haben, sind also weiterhin Verbesserungen der Adaption erwünscht, die ohne oder mit möglichst wenig Modifikationen des Signals auskommen.

2. STEREO-ECHOUNTERDRÜCKUNG

2.1. Das Nichteindeutigkeitsproblem

Nimmt man die Anordnung aus Abb. 1 als Grundlage, lässt sich im ungestörten Fall (kein zusätzliches lokales Signal im Empfangsraum) das an Mikrofon 1 im Empfangsraum gemessene Signal wie folgt im z -Bereich beschreiben:

$${}^1Y(z) = S(z) ({}^1G(z) {}^{11}H(z) + {}^2G(z) {}^{21}H(z)) \quad (1)$$

Dabei bezeichnet $S(z)$ die z -Bereichs-Darstellung des Sprachsignals und ${}^cG(z)$ die Übertragungsfunktion des Pfades zwischen Sprecher und Mikrofon c . $x(k)$, ${}^{ab}H(z)$ ist die Übertragungsfunktion zwischen Lautsprecher a und Mikrofon b im Empfangsraum. Dabei werden alle vorhandenen Komponenten als linear angenommen.

Um das Echo vollständig kompensieren zu können, muss aus den zugänglichen Lautsprechersignalen ${}^1X(z)$ und ${}^2X(z)$ das Mikrofonsignal ${}^cY(z)$ geschätzt werden, wobei die Schätzung mit ${}^c\hat{Y}(z)$ bezeichnet wird. Für die Kompensation muss also gelten:

$${}^c\hat{Y}(z) = {}^cY(z)$$

Das wiederum lässt sich erreichen durch ein Wertepaar

$$\{ {}^{11}\hat{H}(z), {}^{21}\hat{H}(z) \}$$

bei dem

$$\begin{aligned} {}^1G(z) {}^1H(z) + {}^2G(z) {}^2H(z) = \\ {}^1G(z) {}^1\hat{H}(z) + {}^2G(z) {}^2\hat{H}(z) \end{aligned} \quad (2)$$

Die Existenz dieses Wertepaars bedeutet aber nicht notwendiger Weise, dass auch ${}^1\hat{H}(z) = {}^1H(z)$ und ${}^2\hat{H}(z) = {}^2H(z)$ gelten. Es existiert also eine unendlich große Zahl an Lösungen. Dieser Effekt heißt *Nichteindeutigkeitsproblem* oder englisch *non-uniqueness problem*[3].

Durch Raumänderungen im Senderraum, Bewegungen des Sprechers usw. ist allerdings die lineare Abhängigkeit der beiden Lautsprecher signale nicht konstant, sondern zeitvariant. Der vorgestellte Algorithmus versucht diese Eigenschaft auszunutzen.

2.2. Echounterdrückung: Der NLMS-Algorithmus

Der *Normalized Least Mean Squares*-Algorithmus[4, 5] ist ein gängiges adaptives Verfahren zur Schätzung der Impulsantwort. Im Stereofall erfolgt die Berechnung des Fehlersignals und Anpassung der Schätzung nach folgender Vorschrift:

$${}^c\hat{y}(k) = \sum_{l=0}^N {}^1x(k-l) \cdot {}^1c\hat{h}(l) + {}^2x(k-l) \cdot {}^2c\hat{h}(l) \quad (3)$$

$${}^ce(k) = {}^cy(k) - {}^c\hat{y}(k) \quad (4)$$

$${}^1c\hat{h}_{k+1}(l) = {}^c\hat{h}_k(l) + \mu \frac{1}{\sigma_x^2} {}^ce(k) \cdot {}^1x(k-l) \quad (5)$$

$${}^2c\hat{h}_{k+1}(l) = {}^c\hat{h}_k(l) + \mu \frac{1}{\sigma_x^2} {}^ce(k) \cdot {}^2x(k-l) \quad (6)$$

Die Adaptionsgüte des adaptiven Filters lässt sich durch den sog. *Systemabstand* ausdrücken:

$$d = \sum_{k=0}^{\infty} \left(\|{}^1c\hat{h}(k) - {}^1c\hat{h}(k)\|^2 + \|{}^2c\hat{h}(k) - {}^2c\hat{h}(k)\|^2 \right) \quad (7)$$

3. ADAPTIVE SCHRITTWEITENSTEUERUNG

Der oben kurz vorgestellte Algorithmus funktioniert gut für unkorrelierte oder schwach korrelierte Eingangssignale. Bei korrelierten Eingangssignalen dagegen kommt, wie bereits dargestellt, das Nichteindeutigkeitsproblem zum Tragen.

Ausgehend vom Wissen, dass die Korrelation zwischen den Kanälen zeitvariant ist, ist es nahe liegend, nicht wie in [6] vorgeschlagen die Verzerrung, sondern die Adaptionsgeschwindigkeit in Abhängigkeit von der Korrelation zu steuern. Dazu wird in Gl. 5 und 6 die Schrittweite μ bei starker Korrelation zwischen den Kanälen verringert; bei schwacher Korrelation wird sie wieder freigegeben.

Die Schrittweite μ muss sich, um ein stabiles Verhalten des Algorithmus zu garantieren, im Intervall $[0, 2)$ bewegen, verwendet wird üblicher Weise das Intervall $[0, 1]$.

Der Korrelationskoeffizient ist definiert als

$$\gamma_x = \frac{{}^1\mathbf{x}_k^T {}^2\mathbf{x}_k}{\sqrt{({}^1\mathbf{x}_k^T {}^1\mathbf{x}_k) ({}^2\mathbf{x}_k^T {}^2\mathbf{x}_k)}} \quad (8)$$

Dieser liegt im Intervall $[-1, 1]$. Damit bietet es sich an, die Schrittweite wie folgt anzupassen:

$$\mu_k = 1 - \gamma_x^2 \quad (9)$$

Bei dieser Steuerung wird die Adaption in den Bereichen mit starker Korrelation gebremst, was verhindert, dass die Schätzung durch die Nichteindeutigkeit divergiert.

4. EXPERIMENTELLE ERGEBNISSE

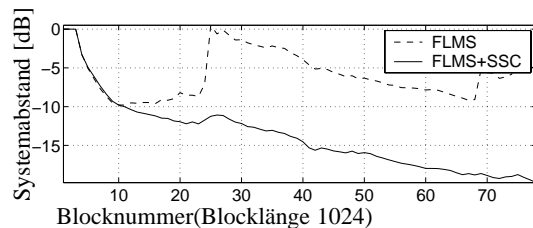


Fig. 2. Systemabstand (Eingang: Stereosignal, einzelner Sänger, keine Raumänderungen, $f_A = 16000$ Hz)

Abb. 2 vergleicht die konventionelle Methode mit der Verwendung der adaptiven Schrittweitensteuerung. Die obere Kurve (gestrichelt) zeigt den Systemabstand ohne adaptive Steuerung. Es ist deutlich zu erkennen, dass das adaptive Filter ungefähr ab Block 10 schlechter konvergiert als die Variante mit Schrittweitensteuerung und ab Block 25 beginnt, zu divergieren.

Bei Verwendung der adaptiven Schrittweite (untere Kurve) dagegen wird die Adaption an den kritischen Stellen abgebremst bzw. angehalten. Das divergierende Verhalten des ursprünglichen Verfahrens wird unterbunden, das Filter konvergiert zuverlässig zur korrekten Schätzung hin.

5. ZUSAMMENFASSUNG

Das Problem der Stereo-Echounterdrückung wurde hier vorgestellt und das Nichteindeutigkeitsproblem besprochen. Mit einem neuartigen Ansatz zur Modifikation der Schrittweite lässt sich die Verzerrung der Eingangssignale entweder verringern oder ganz vermeiden, was eine bessere Signalqualität zulässt.

6. LITERATUR

- [1] M. Ali, "Stereoacoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *ICASSP-98*, 1998, vol. 6, pp. 3689–3692.
- [2] J. Benesty, D. Morgan, and M. Sondhi, "A better understanding and an improved solution to the problems of stereoacoustic echo cancellation," in *ICASSP-97*, 1997, vol. 1, pp. 299–302.
- [3] M. Sondhi, D. Morgan, and J. Hall, "Stereoacoustic echo cancellation - an overview of the fundamental problem," *IEEE Signal Processing Letters*, vol. 2, no. 8, August 1995.
- [4] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River NJ, 1996.
- [5] George Moschytz and Markus Hofbauer, *Adaptive Filter*, Springer, Berlin, 2000.
- [6] T. Gänslar and J. Benesty, "New insights to the stereoacoustic echo cancellation problem and an adaptive non-linearity solution," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 686–696, 2001.