VOLLDUPLEX–KOMMUNIKATION IN KIOSK- UND KONFERENZSYSTEMEN

Christian Gruber

voice INTER connect GmbH, Ammonstraße 35, 01067 Dresden
gruber@voiceinterconnect.de

Kurzfassung: Eine echte Vollduplex–Kommunikation in elektronischen Kioskanwendungen und Konferenzsystemen stellt hohe Anforderungen an die Leistungsfähigkeit der Audiosignalverarbeitung. Bisherige Lösungen mit einer Pegelwaage ermöglichen eine robuste Kommunikation, jedoch ist stets nur eine wechselseitige Übertragung der Audiosignale möglich (Halbduplex), was von den Sprechern eine gewisse Disziplin fordert. Hintergrundgeräusche werden durch das Umschalten mit dem Sprachsignal teilweise übertragen und teilweise unterbrochen, was die Natürlichkeit des Dialoges beeinträchtigt.

Eine echte Vollduplex–Kommunikation hingegen entspricht dem natürlichen Kommunikationsverhalten des Nutzers und erlaubt sowohl gleichzeitiges Sprechen der Beteiligten als auch eine Übertragung der akustischen Kulisse, wodurch die Natürlichkeit verstärkt wird. Vorraussetzung für eine Vollduplex–Kommunikation ist eine sehr gute akustische Echokompensation. Die akustischen Rahmenbedingungen, sowie die verwendeten Gerätekomponenten beschränken jedoch mit den derzeit verfügbaren Echokompensationsalgorithmen die Vollduplex–Fähigkeit einer Anwendung. Diese beschränkenden Faktoren und die damit verbundenen Herausforderungen werden näher betrachtet.

In aktuellen praktischen Systemen gibt es daher graduelle Abstufungen (Teilduplex) durch eine Kombination von einer Pegelwaage mit einer akustischen Echokompensation und Störgeräuschunterdrückung. Der Eindruck der wechselseitigen Übertragung wird dadurch bereits deutlich gemindert. Die Anforderungen an die akustische Echokompensation sind geringer als in einem echten Vollduplex–System.

1 Einleitung

Ein Kommunikationssystem, welches eine bidirektionale Audiosignalübertragung ermöglichen soll, hat den in Abbildung 1 gezeigten prinzipiellen Aufbau. Das Sprachsignal vom fernen Ende $x(k)$ wird über einen Lautsprecher ausgegeben und das Sprachsignal vom nahen Ende $s(k)$ wird vom Mikrofon aufgenommen und an das ferne Ende übertragen. Zusätzlich zum Sprachsignal nimmt das Mikrofon aber auch das vom Lautsprecher in das Mikrofon zurückgekoppelte Echosignal $y(k)$ auf. Dieses Echosignal würde am fernen Ende ohne eine entsprechende Signalverarbeitung als extrem störend wahrgenommen.

2 Anforderungen

Die Dämpfung zwischen dem Sprachsignal vom fernen Ende $x(k)$ und dem an das ferne Ende zurück übertragenen Signal (TCL[1]) muss im Single Talk, d.h. wenn nur der ferne Sprecher

[1] TCL – terminal coupling loss

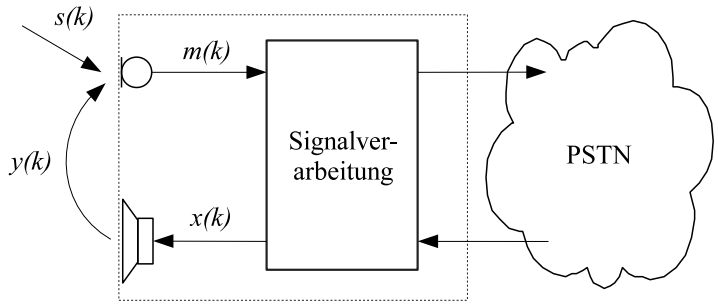

Abbildung 1 - Prinzipaufbau eines bidirektionalen Kommunikationssystems

spricht, mindestens 40 dB betragen [3]. Der Abstand zwischen dem Lautsprecher und dem Mikrofon ist bei Kiosk- und Konferenzsystemen relativ gering (ca. 5 – 20 cm). Je nach Lautsprecher–Mikrofon–Abstand und eingestellter Wiedergabelautstärke beträgt die akustische Rückkopplung, also die Verstärkung zwischen dem Sprachsignal $x(k)$ und dessen Echosignal $y(k)$, 10 – 20 dB. Das Echosignal muss demnach um 50 – 60 dB gedämpft werden (ERLE[2], [1]), um die geforderte Rückkopplungsdämpfung von 40 dB zu erfüllen.

3 Pegelwaagensysteme

Einfache Systeme integrieren eine Pegelwaage, welche wechselseitig den Sende- und den Empfangskanal bedämpft (siehe Abbildung 2). Dadurch wird dauerhaft eine konstant hohe Dämpfung erzielt und die geforderte Echodämpfung wird sichergestellt. Weitere Komponenten sind nicht notwendig. Pegelwaagen funktionieren relativ robust auch bei Störgeräuschen und haben im Vergleich zu Echokompensations–Algorithmen eine geringe Komplexität. Leider ist mit einer Pegelwaage aber auch nur eine wechselseitige Kommunikation möglich (Halbduplex).

4 Systeme mit akustischer Echokompensation

4.1 Funktionsprinzip einer akustischen Echokompensation

Eine akustische Echokompensation (AEC) ist ein adpatives Filter, welches die Aufgabe hat, das akustische Übertragungsverhalten zwischen dem Lautsprechersignal und dem Echosignal, dem

[2]ERLE – Echo return loss enhancement

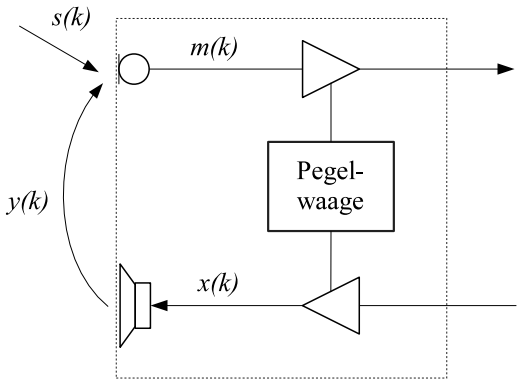

Abbildung 2 - Aufbau eines Kommunkationssystems mit einer Pegelwaage

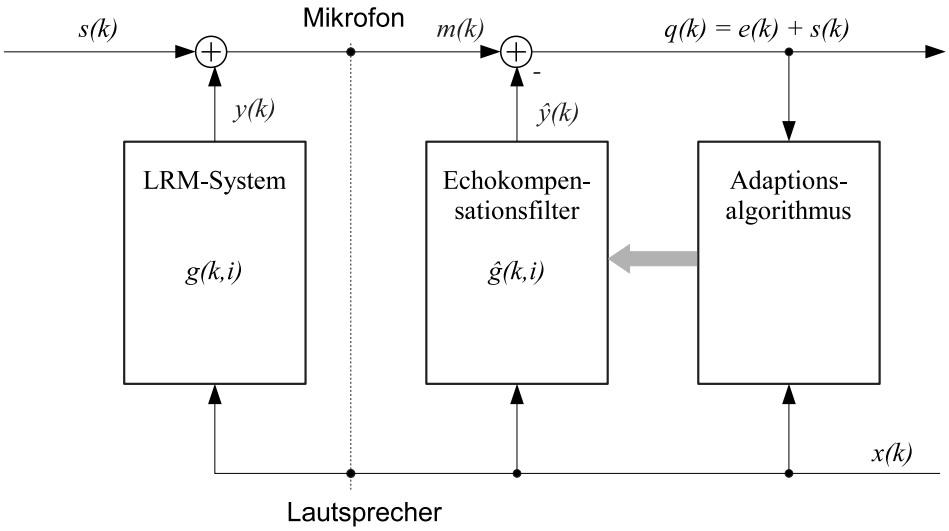

Abbildung 3 - Funktionsprinzip einer akustischen Echokompensation

so genannten Lautsprecher–Raum–Mikrofon–System (LRM–System), nachzubilden (siehe Abbildung 3). Damit wird ein geschätztes Echosignal $\hat{y}(k)$ berechnet und vom Mikrofonsignal $m(k)$ subtrahiert. Das Sprachsignal $s(k)$ wird dadurch nicht beeinträchtigt. Das Echokompensationsfilter wird üblicherweise als nichtrekursives Filter (FIR–Filter) realisiert. Da das LRM–System in erster Näherung durch die Raumimpulsantwort beschrieben werden kann, entspricht die Impulsantwort des Echokompensationsfilters nach erfolgter Adaption der Raumimpulsantwort und die Filterkoeffizienten sind die Abtastwerte der Raumimpulsantwort.

Die Herausforderung bei der Umsetzung einer AEC besteht darin, einen geeigneten Adaptionsalgorithmus zu entwickeln welcher

- eine schnelle Adaption aufweist
- eine große Anzahl von Filterkoeffizienten adaptieren kann
- eine hohe Echodämpfung aufweist
- robust ist gegenüber Störgeräuschen und Gegensprechsignalen
- eine geringe Komplexität aufweist, so dass eine Realisierung z.B. auf einem DSP möglich ist

4.2 Kombinationssysteme mit Pegelwaage und akustischer Echokompensation

Kombiniert man eine Pegelwaage mit einer akustischen Echokompensation (siehe Abbildung 4), wird ein Teil der geforderten Echodämpfung durch die AEC erreicht. Die restliche Echodämpfung wird weiterhin durch die Pegelwaage erzielt. Es ist mit heutigen Echokompensationsalgorithmen möglich, ca. 20 – 30 dB Echodämpfung zu erzielen, ohne dass dabei das Sprachsignal des nahen Sprechers beeinträchtigt wird. Eine weitere Echodämpfung um ca. 10 – 20 dB lässt sich durch Nachverarbeitung mit einer Restechoreduktion erzielen, wodurch das Sprachsignal jedoch bereits teilweise bearbeitet wird und etwas an seiner Natürlichkeit verliert. Man erreicht mit diesen beiden Maßnahmen zusammen ca. 40 – 50 dB Echodämpfung. Die Pegelwaage muss daher nur noch etwa 10 dB dämpfen.

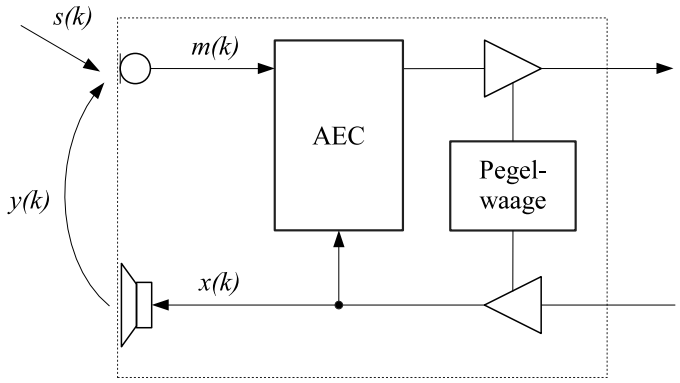

Abbildung 4 - Aufbau eines kombinierten Systems aus Pegelwaage und akustischer Echokompensation

Durch eine solche Kombinationslösung wird bereits eine Teilduplex–Kommunikation möglich. Es entsteht nicht mehr der Eindruck, dass zwischen zwei Übertragungsrichtungen umgeschaltet wird. Stattdessen wird jeweils ein Kanal immer nur leicht gedämpft. Jedoch ist auch dies noch keine echte Vollduplex–Kommunikation. Es wird immer noch hörbar zwischen einer ungedämpften und einer gedämpften Übertragung des Sprachsignals hin- und hergeschaltet. In vielen Fällen, in denen nur einer der beiden Gesprächsteilnehmer spricht, wird dies weniger auffallen. Lediglich die Geräuschkulisse wird lauter und leiser. In einer Gegensprechsituation ist das Verhalten der Pegelwaage aber auf jeden Fall noch hörbar. Außerdem beeinträchtigt die Restechoreduktion teilweise die Sprachsignalqualität des nahen Sprechers. Auch die Wiedergabelautstärke ist stark begrenzt.

4.3 Vollduplex–System

Für eine echte Vollduplex–Kommunikation muss die akustische Echokompensation ohne Unterstützung durch eine Pegelwaage auskommen. Auch die Restechoreduktion darf nur geringfügig eingreifen, da sonst das Nutzsignal zu stark beeinträchtigt wird und damit der natürliche Charakter der Sprache verloren geht. Um dieses Ziel zu erreichen, muss die Echodämpfung noch weiter erhöht werden.

Es gibt zwei wesentliche Ursachen, welche die Dämpfungsleistung aktueller Echokompensationsalgorithmen begrenzen.

- eine begrenzte Echofilterlänge
- nichtlineare Verzerrungen im LRM–System

Die Länge des Echokompensationsfilters ist durch die Anzahl der verwendeten Filterkoeffizienten begrenzt. Ist die Echofilterlänge kleiner als die Länge der Raumimpulsantwort, kann sich das Echokompensationfilter nicht vollständig an die Raumimpulsantwort adaptieren. Es wird nur ein Teil der Raumimpulsantwort nachgebildet und damit auch nur ein Teil des Echosignals kompensiert. Das verbleibende Restechosignal führt zu einer begrenzten Echodämpfung. Die Echofilterlänge lässt sich jedoch nicht beliebig erhöhen, da mit zunehmender Echofilterlänge die Adaptionsgeschwindigkeit der Algorithmen abnimmt. Es muss also ein Kompromiss gefunden werden. Dieser liegt mit heutigen Algorithmen bei einer Echofilterlänge von ca. 50 ms.

Abbildung 5 zeigt ein Beispiel einer Raumimpulsantwort aus einer Büroumgebung. Die Raumimpulsantwort setzt sich zusammen aus einem Teil, welcher durch den Direktschall hervorgerufen wird, und einem Teil, welcher durch den Diffusschall hervorgerufen wird. Der Direkt-

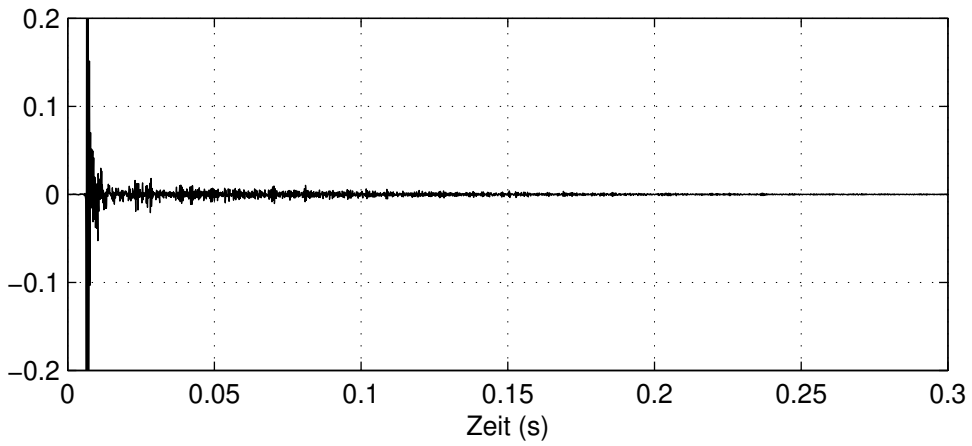

Abbildung 5 - Raumimpulsantwort in einem kleineren Büro (50 m^3)

schallanteil ist bereits nach ca. 20 ms abgeklungen. Dieser Anteil lässt sich bereits mit heutigen Algorithmen gut kompensieren. Wird nur der Direktschallanteil kompensiert, wird das Restechosignal durch den Diffusschallanteil der Raumimpulsantwort erzeugt. Dies ist dann als Nachhall wahrnehmbar.

Die Länge der Raumimpulsantwort hängt von der Nachhallzeit ab. Man kann anhand der Nachhallzeit abschätzen, wie groß die Echodämpfung in Abhängigkeit von der Echofilterlänge bei vollständiger Adaption aller Filterkoeffizienten sein wird. Je größer die Nachhallzeit ist, desto länger muss das Echokompensationsfilter sein. In Tabelle 1 ist für drei verschiedene Umgebungen die Nachhallzeit sowie die Echofilterlänge angegeben, welche benötigt wird, um eine Echodämpfung von 30 dB zu erreichen. Die Echodämpfung, welche durch die Kompensation des Direktschallanteils erzielt wird, ist dabei nicht berücksichtigt.

Tabelle 1 - Erforderliche Echofilterlängen in verschiedenen Umgebungen

Umgebung	Nachhallzeit	Echofilter-länge	Filterkoeff. (8 kHz)	Filterkoeff. (16 kHz)
kleines Büro	300 ms	150 ms	1200	2400
großes Büro	400 ms	200 ms	1600	3200
Foyer	600 ms	300 ms	2400	4800

Die Tabelle zeigt anschaulich, dass eine Echofilterlänge von 50 ms selbst in einem kleineren Büro nicht ausreicht. In größeren Räumen wird mit dieser Echofilterlänge kaum noch eine Echodämpfung erreicht. Es bleibt ein deutlich wahrnehmbarer Nachhall erhalten. Eine große Herausforderung besteht deshalb darin, Echokompensationsalgorithmen zu entwickeln, die auch bei sehr großen Echofilterlängen eine schnelle Adaption ermöglichen. Dabei darf der Ressourcenbedarf nicht zu stark ansteigen, so dass eine praktische Umsetzung mit aktuellen Prozessoren noch möglich ist.

Interessant ist auch die Tatsache, dass sich durch eine Erhöhung des Lautsprecher–Mikrofon–Abstandes die Echorückkopplung (TCL) nicht weiter verringern lässt. Bei höherem Lautsprecher–Mikrofon–Abstand verringert sich zwar die Stärke der akustischen Rückkopplung, jedoch verringert sich in gleichem Maße auch die Echodämpfung (ERLE). Das liegt daran, dass bei einer Verringerung des Lautsprecher–Mikrofon–Abstandes nur die Intensität des Direktschallanteils abnimmt, jedoch nicht die Intensität des Diffusschallanteils. Die Echorückkopplung kann daher nur durch eine größere Echofilterlänge oder durch eine Verringerung der Wiedergabelautstärke reduziert werden.

Die zweite Ursache für eine begrenzte Echodämpfung sind nichtlineare Verzerrungen im LRM–System, d.h. Klirrverhalten. Diese führen zu nichtlinearen Echosignalanteilen und werden durch übliche Echokompensationsalgorithmen nicht kompensiert, da diese von einem linearen Übertragungsverhalten ausgehen. Häufigste Ursachen von nichtlinearen Verzerrungen sind kleine Lautsprecher, geringe Bauvolumina und Begrenzungseffekte in elektronischen Bauteilen, wie z.B. Verstärkern. Häufig kommen auch vibrierende Bauteile hinzu, welche zu „Klapper"–Geräuschen führen.

Um auch nichtlineare Echoanteile kompensieren zu können, gibt es zwei Möglichkeiten:

- Einsatz der Restechoreduktion (NLP[3])
- Erweiterung des Echokompensationsalgorithmus zur Modellierung nichtlinearen Verhaltens

Der erste Ansatz hat den Nachteil, dass durch die Restechoreduktion das Sprachsignal weiter beeinträchtigt wird. Beim zweiten Ansatz geschieht dies nicht, da hier das (nichtlineare) Echosignal geschätzt und vom Mikrofonsignal subtrahiert wird. Zu diesem Thema gibt es bereits mehrere Veröffentlichungen [4]. Ein großes Problem ist jedoch, dass die Komplexität des Algorithmus extrem schnell zunehmen kann, wenn nicht geeignete Einschränkungen bei der Modellierung des nichtlinearen LRM–Systems vorgenommen werden. Außerdem lassen sich damit auch nicht beliebige Nichtlinearitäten modellieren. Die genannten „Klapper"–Geräusche zum Beispiel lassen sich damit nicht kompensieren.

5 Zusammenfassung

Dieser Beitrag zeigt drei Entwicklungsschritte von einer Halpduplex–Kommunikationslösung, bestehend aus einer Pegelwaage, zu einer Vollduplex–Kommunikationslösung mit einer akustischen Echokompensation. Es werden die wesentlichen Vor- und Nachteile dieser drei Ansätze aufgeführt. Zum Abschluss werden die Herausforderungen beleuchtet, welche noch zu meistern sind, um zu einer universell einsetzbaren echten Vollduplex–Kommunikation zu gelangen. Dies sind einerseits die begrenzte Adaptionsfähigkeit heutiger Echokompensationsalgorithmen an größere Räume. Andererseits führen nichtlineare Verzerrungen, welche durch kleine Lautsprecher, begrenzte Bauvolumina, Begrenzungseffekte in elektronischen Bauelementen und vibrierende Gehäusebauteile verursacht werden, zu einer grundsätzlichen Begrenzung der Echodämpfungsleistung.

Literatur

[1] HÄNSLER, E. und G. SCHMIDT: *Acoustic Echo and Noise Control - a practical approach*. Wiley, 2004.

[2] HÄNSLER, E. und G. SCHMIDT: *Topics in Acoustic Echo and Noise Control*. Springer, 2006.

[3] ITU-T G.167: *Acoustic echo controllers*, 1993.

[4] STENGER, A.: *Kompensation akustischer Echos unter Einfluss von nichtlinearen Audiokomponenten*. Doktorarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2000.

[3] NLP – nonlinear processor