Intelligente Steuerung eines Koinzidenzmikrofons in Mehrsprecheranwendungen

Marco Riemann, Martin Opitz

AKG Acoustics GmbH, 1230 Wien, Österreich, Email: marco.riemann@harman.com, martin.opitz@harman.com

Einleitung

Mikrofonarrays bieten die Möglichkeit die Position einer im Raum befindlichen Schallquelle zu orten. In dieser Arbeit wird ein Algorithmus angewendet, der zur Schallquellenlokalisation eine koinzidente Mikrofon-Anordnung [1] verwendet. Diese besteht aus drei Druckgradientenwandlern und einem Druckempfänger (CMT Array). Die Druckgradientenwandler haben eine zylindrische flache Bauform mit einer Höhe von 3,4mm und einem Durchmesser von 13,5mm. Die Schalleintrittsöffnungen der Wandler sind auf der oberen Seite der Zylinder positioniert. Anders als bei herkömmlichen Mikrofonkapseln liegt die Haupt-Empfindlichkeitsachse parallel zur oberen Seite des Zylinders. Die drei Druckgradientenwandler sind in einer Ebene so angeordnet, dass sich die Haupt-Schalleintrittsöffnungen in einem Winkel von je 120° zueinander befinden. Durch die enge Bauweise entstehen keine Laufzeitdifferenzen. Zur Schallquellenortung werden ausschließlich die Pegeldifferenzen der drei Mikrofonkapseln verwendet.
Bei dem hier eingesetzten Richtungserkenner handelt es sich um das sogenannte Similarity Verfahren [1], einen Minimum Distanz Klassifikator, der blockweise aus den Mikrofonsignalen extrahierte Merkmalsvektoren mit einer Merkmalsdatenbank vergleicht und so die Bestimmung der örtlichen Position einer Schallquelle erlaubt.

Für Anordnungen mit mehreren Sprechern rund um einen Tisch wurde ein Idealverhalten des Mikrofons mit der Bezeichnung Beamforming Modus (BFM) definiert. Beim BFM Verfahren erfolgt eine flexible Umschaltung bzw. Überblendung zwischen omnidirektionaler und unidirektionaler Charakteristik. Sobald genau ein Sprecher aktiv ist, wird auf „unidirektional“ geschaltet, wenn mehrere Sprecher simultan sprechen sowie in Sprachpausen wird die „omni“-Schaltung aktiviert.
Im Fall der Schaltung auf „unidirektional“ in eine gewünschte Richtung wird nach [2] das B-Format berechnet und daraus ein Gradientensignal mit der gewünschten Haupt-Schalleinfallsrichtung ermittelt. Um dieses Idealverhalten mit realen Signalen zu überprüfen, wurde in MATLAB ein entsprechender Algorithmus abgebildet.

Audioaufzeichnungen und Dialogerstellung

Es wurden Audioaufzeichnungen in einem Abhörraum der AKG Acoustics GmbH in Wien mit einem männlichen Sprecher und einer weiblichen Sprecherin angefertigt. Der Abhörraum hat eine Größe von 7,65x5,6x3,2m^3. Detaillierte Informationen zu dem Raum sind in [3] beschrieben. Um eine möglichst hohe Reproduzierbarkeit der einzelnen Audioaufzeichnungen und des Schallfeldes sicherzustellen, wurden die erwähnten Sprachsequenzen und weißes Rauschen mit einem künstlichen Mund (Brüel&Kjaer 2238) wiedergegeben und mit dem CMT Array auf einem Harddisk-Recordingsystem nacheinander aufgezeichnet. In Abbildung 1 sind die Punkte, an denen der künstliche Mund positioniert wurde, mit den Indizes MP1 bis MP9 gekennzeichnet.

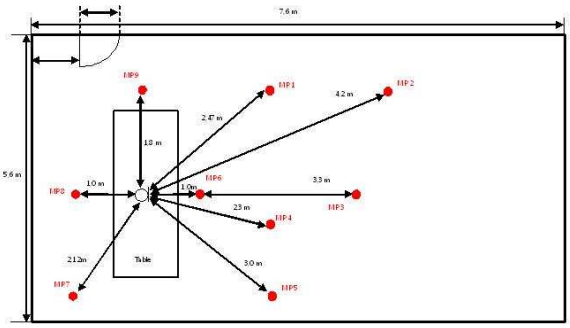

Abbildung 1: AKG Demo Raum Grundriss mit dazugehörigen Messpunkten. Links ist die Position des Tisches mit dem CMT-Mikrofon darauf eingezeichnet.

Für die im Folgenden beschriebenen Beispiele wurden Dialoge mit einem speziellen auf Basis von MATLAB erstellten Dialog Synthesizer erstellt. Es wurden verschiedene Dialoge sowohl mit Pausen als auch mit Sprachsignalüberlappungen erzeugt entsprechend Gesprächen, die mit mehr oder weniger Gesprächsdisziplin geführt werden. Zu den Sprachsequenzen wurde ein annähernd diffuses weißes Rauschen dazu gemischt, welches, wie oben erwähnt, ebenfalls mit dem CMT Array aufgezeichnet wurde.

Fehlermaß

Ziel des BFM ist es ungewünschte Artefakte zu unterbinden, die durch das Hin- und Herschwenken des Beams entstehen können. Dies ist insbesondere dann gegeben, wenn mehrere Schallquellen sich zeitlich überlagern. In Abbildung 2 ist die Auswertung des oben beschriebenen synthetisierten Dialoges abgebildet. Im Folgenden wird nur der Azimutwinkel zur Richtungsschätzung herangezogen. Zur Auswertung der Fehlermaße wird das Signal in Blöcke von ca. 20ms Dauer zerlegt. An oberster Stelle in Abbildung 2 ist das erste Gütemaß abgebildet welches Bestandteil des Similarity Algorithmus ist. Hierbei handelt es sich um die Prozentangabe der Blöcke, innerhalb der der absolute Winkelfehler kleiner als ein vorgegebener Grenzwert ist. ACC5=90% bedeutet, dass 90% aller Blöcke einen Fehler in der Azimutwinkeldetektion kleiner gleich 5° aufweisen, ACC15 und ACC30 gelten äquivalent für Toleranzbereiche von 15° und 30°. Weitere Fehlermetriken sind der mittlere absolute Fehler in Grad "mean absolute error" (MAE), sowie

"root mean square error (RMSE)". Darunter befindet sich der mit 1 gekennzeichnete Bereich, in dem die Wellenform der synthetisierten Sprachsequenz, die vom Richtungserkenner detektierte Richtung und der wahre Azimutwinkel abgebildet sind. Die vom Richtungserkenner detektierte Richtung wird durch die blaue strichpunktierte Kurve symbolisiert. Die rote Kurve entspricht der wahren Richtung. Unter dem Bereich 1 befindet sich der mit 2 gekennzeichnete Verlauf des Similarity Gewichtungsfaktors b0, der in [1] beschrieben ist. Die Variable b0 ist auf das Intervall [0,1] beschränkt. Je höher der Wert b0 ist, desto höher ist die Wahrscheinlichkeit, dass der momentane aus dem CMT Signal extrahierte Merkmalsvektor mit einem Vektor innerhalb der Merkmalsdatenbank übereinstimmt. Der Wert b0 wird gemeinsam mit einem empirisch ermittelten Schwellwert für die Ermittlung des BFM Modus verwendet und mit einem gleitenden Mittelwertbilder geglättet, um allzu häufige Umschaltvorgänge zu vermeiden. Die mit der Glättung einhergehende geringere Steilheit der Kurvenanstiege im b0-Verlauf führt allerdings zu leichten Verzögerungen bei der Modus-Umschaltung, zu sehen an der Balkengrafik im Bereich 3 von Abbildung 2.

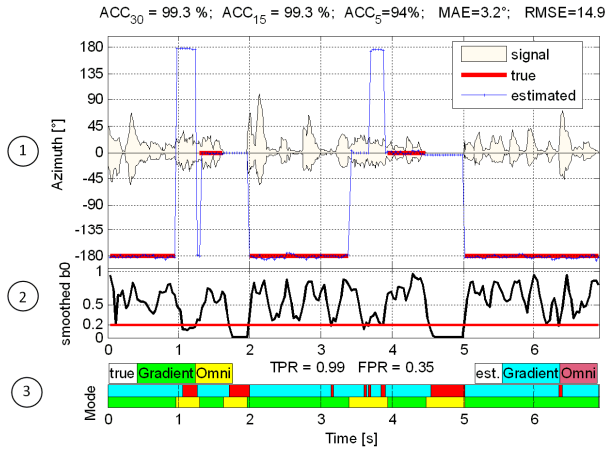

Abbildung 2: Oben (1) : Die Wellenform des Signals mit dem wahren (rot) und dem errechneten Azimutwinkel (blau). Mitte(2) : Die geglättete Kurve von b0 zur Ermittlung des Beamformer-Modus (schwarz) und der Schwellwert (rot). Unten (3) : Die Balken-Graphen mit dem wahren (grün, gelb) und dem errechneten (Cyan, rot) Beamforming Modus. Weiters sind die True Positive Rate (TPR) und die False Positive Rate (FPR) in Zahlen angegeben.

Als Fehlergröße für die Ermittlung des BFM wird die sogenannte „Receiver Operating Characteristic“ (ROC) verwendet. Bei der Fehlerauswertung wird nun blockweise der ermittelte BF-Modus mit dem idealen Modus verglichen und es werden dabei Werte aus 4 Klassen nach dem folgenden Schema ermittelt:

Wahrer Modus	gradient	omni	gradient	omni
Errechneter Modus	gradient	omni	omni	gradient
Klasse	TP	TN	FN	FP

Jeder Block kann bei der Berechnung in eine dieser 4 Klassen eingeordnet werden. True Positive bedeutet dabei das richtige Erkennen des Gradientenmodus, True Negative das richtige Erkennen des Omni-Modus. Ein Block wird als False Negative gewertet, wenn Omni-Modus erkannt, aber der wahre Wert gradient ist, und False Positive bedeutet wahrer Wert Omni-Modus und erkannter Modus gradient.
Aus diesen Werten werden dann die True Positive Rate (TPR) und die False Positive Rate (FPR) folgendermaßen errechnet:

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}.$$

Ziel ist es , dass TPR nahe bei 1 liegt, FPR soll nahe bei 0 sein. In Abbildung 2 sieht man, dass ein gewünschter Gradientenmodus sehr gut erkannt wird (TPR=0.99) , ein gewünschter Omni-Modus in 35% der Fälle aber fälschlicherweise durch eine Gradientenschaltung abgebildet wird.
Durch Anpassung des Schwellwertes für b0 sowie durch Hinzufügen von weißem Rauschen wie oben erwähnt werden die Werte von TPR und FPR beeinflusst. Es zeigte sich, dass durch Erhöhung der b0-Schwelle von 0,2 auf 0, 4 die Werte von TPR und FPR sanken. Die optimalen Werte von b0 können erst nach Durchführung geeigneter Hörversuche ermittelt werden.
Einen ähnlichen Effekt kann man bei der Hinzufügung des Rauschsignals beobachten. Ausgehend vom ungestörten Fall wurde das SNR auf einen Wert bis zu 10dB SNR verringert. Der TPR-Wert nahm bis auf 0.5 deutlich ab, was ungewünscht ist, da die gewünschte Gradientenschaltung nur in 50% der Frames nachgebildet wurde. Der FPR-Wert fiel bis auf 0.02 ab, was als positiv zu vermerken ist, da der gewünschte Omni-Modus tatsächlich erreicht wurde.

Ausblick

Die gefundenen Ergebnisse sollen mit entsprechenden Hörtests verglichen werden. Besonderes Interesse gilt dabei Gesprächsverläufen mit Überlappungen der Sprachsignale.

Danksagung

Diese Arbeit wurde mit Unterstützung der Europäischen Union innerhalb des 7. Rahmenprogramms für Forschung und Entwicklung (FP7/2007-2011), CompanionAble Project (grant agreement n. 216487) durchgeführt.

Literatur

[1] Freiberger, K.: Developement and Evaluation of Source Localization Algorithms for Coincident Microphone Arrays, Diploma Thesis, Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria, April 2010

[2] Gerzon, M.: "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound" presented at the 50th AES Convention London, March 1975

[3] Breunhölder, R.: "Planung eines Multimedia-Demoraumes für Surround Anlagen und Kopfhörer mit Tieftonoptimierung", Diplomarbeit, Fachhochschulstudiengang, Technikum Wien, April 2004