Mehrkanalsignalverarbeitung zur Bestimmung der Kopfausrichtung in geschlossenen Räumen

Philipp Roeske[1], Joerg Bitzer[1], Uwe Simmer[1]

[1] *Institut für Hörtechnik und Audiologie (IHA), FH Oldenburg, Ofenerstr.16, 26121 Oldenburg (www.hoertechnik-audiologie.de)*

1 Einleitung

In vielen Fällen ist eine genaue Kenntnis der momentanen Position eines Sprechers in gegebener Raumgeometrie Voraussetzung für robuste Algorithmen zur mehrkanaligen digitalen Sprachsignalverarbeitung. Nur so ist es möglich, Sprache möglichst unverzerrt mit vorhandenen Sensor(Mikrofon-)anordnungen aufzunehmen. Die fehlende Kenntnis über die Ausrichtung des Kopfes kann jedoch auch bei perfekter räumlicher Lokalisation zu Signalverzerrungen führen, da die akustische Abschattung des menschlichen Körpers mit steigender Frequenz für hohe Dämpfung sorgt. Dies gilt insbesondere dann, wenn die Ausrichtung der Sensoren der freien Bewegung des Sprechers in seiner Umgebung folgen soll. Um Algorithmen testen und vergleichen zu können, wurde eine Simulationsumgebung basierend auf der Spiegelquellenmethode von Allen und Berkley implementiert. Eine Erweiterung des Modells ermöglicht zudem die Richtcharakteristik des Mundes am menschlichen Kopf der simulierten Quelle aufzuprägen, um so eine realistischere Akustik zu erhalten. Für die Auswertung der Daten wurden zunächst verschiedene bestehende Algorithmen zur Sprecherlokalisation implementiert. Darauf basierend erfolgt durch Modifikation und Erweiterung eine Schätzung der Kopfausrichtung. Dieser Artikel stellt erste Ergebnisse unserer Arbeit vor.

2 Sprecherlokalisation (SL)

Um die Qualität existierender SL-Algorithmen beurteilen zu können, wurden verschiedene Verfahren implementiert. Im Folgenden werden zwei Algorithmen vorgestellt, da diese in den Untersuchungen gute Ergebnisse zeigten. Die detaillierten Untersuchungen zur SL sind in [Roe07] nachzulesen.

2.1 GCC mit Phasentransformation (GCC-PHAT)

Die Gewichtung des Kreuzleistungsdichtespektrums (KLDS) $\Phi_{X_iX_l}(n)$ zweier Sensorsignale mit dessen Betragsspektrum führt zum Phasenspektrum. Durch dieses pre-whitening des KLDS ergibt die inverse DFT (*Discrete Time Fourier Transform*) näherungsweise in einem

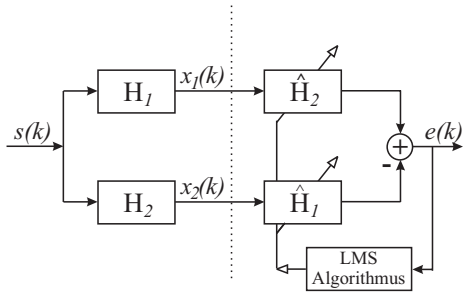

Abbildung 1: Blockschaltbild des AEDA. Die adaptiven Filter $\hat{H}_1$ und $\hat{H}_2$ sollen die tatsächlichen Übertragungsfunktionen des jeweils anderen Kanals ausgleichen.

$\delta-$Impuls.

$$R_{x_ix_l}(\kappa) = \frac{1}{N}\sum_{n=0}^{N-1}\frac{\Phi_{X_iX_l}(n)}{|\Phi_{X_iX_l}(n)|}e^{j2\pi n\kappa/N} \quad (1)$$

$$= \frac{1}{N}\sum_{n=0}^{N-1}e^{j2\pi n(\kappa-\tau_{il}f_s)/N} \quad (2)$$

$$= \delta(\kappa-\tau_{il}f_s) \quad (3)$$

Wobei τ_{il} die Verzögerung zwischen den Sensoren m_i und m_l und f_s die Abtastrate ist. Diese Eigenschaft erleichtert die Schätzung des zeitlichen Versatzes und damit τ_{il}.

2.2 AED

Der AED-Algorithmus (AEDA) [BYE00] bestimmt die Laufzeit zwischen den Sensoren durch Schätzung der Verzögerung der Direktschallanteile. Dies erfolgt adaptiv mittels *Constrained Frequency-Domain Least Mean Squares* (CFLMS). Abb.1 zeigt das Blockschaltbild des Algorithmus. Die Impulsantworten $h_{1,2}(k)$ weisen ähnlich den GCC-Verfahren Spitzen auf, deren relativer Abstand der Laufzeit τ entspricht.

2.2.1 AED-PHAT

Eine während der Untersuchungen entwickelte Erweiterung des AEDA ergab zudem eine Verbesserung. Hierbei wurden die adaptierten Filter $\hat{H}_1(n)$ und $\hat{H}_2(n)$ phasentransformiert, um wie bei den GCC-Verfahren eine möglichst schmale und in der Amplitude maximale Spitze zu erzielen.

2.3 Ergebnisse

Aus den geschätzten Laufzeiten τ_{il} wurden mittels zweier Methoden die Koordinaten $[x_s, y_s, z_s]$ berechnet. Diese

waren das *One-Step-Least-Squares* (OSLS) [BYE00] sowie die *Triangulation*. Da letzteres Verfahren die besseren Resultate erzielte, werden nun kurz drei Ergebnisse des SL-Experimentes bei der Nachhallzeit $T_{60} = 500ms$ und einem SNR von 15dB vorgestellt. Gezeigt ist der Simulationsraum in der Draufsicht.

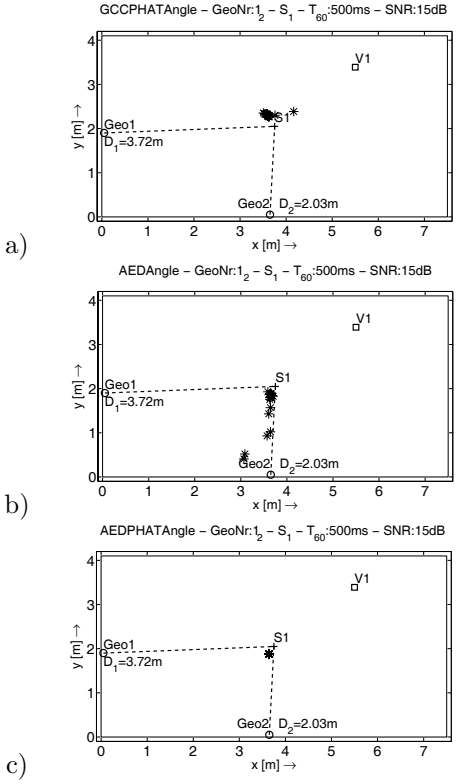

Abbildung 2: Ergebnisse des SL-Experimentes bei $T_{60} = 500ms$ und 15dB SNR. a) GCC-PHAT, b) AED, c) AED-PHAT. S_1 zeigt die Sprecher-, V_1 die Störgeräuschquelle.

3 Sprachrichtungserkennung

Die Untersuchungen zur Sprachrichtungserkennung basieren darauf, dass die beim Sprechen erzeugten Schallwellen frequenz- und richtungsabhängig gedämpft werden. Ausgenutzt wird dabei zum einen das Abstrahlverhalten des menschlichen Kopfes und zum anderen die Abschattung des Quellsignals durch den Kopf beim Empfang.

Dies führt dazu, dass in Sprecherrichtung ab $f \geq 500$Hz mehr Energie abgestrahlt wird, als beispielsweise seitlich oder hinter dem Sprecher. Reflexionen, die aus dieser Richtung auf einen Kunstkopf (Kemar) treffen, weisen somit unterschiedliche Intensitäten bei hohen Frequenzen auf. Dies lässt auch darauf schließen, dass Freifeldsensoranordnungen eine schlechtere Erkennungsleistung erwarten lassen.

3.1 Untersuchungen

Die an den Sensoren empfangenen Signale $x(k)$ wurden blockweise mittels einer Bark-Filterbank in mehrere Frequenzbänder zerlegt. In den Bändern zwischen 2kHz und 8kHz erfolgte eine Berechnung der Energie. Es zeigte sich, dass, in jeweils der Sprecherrichtung, empfangsseitig in

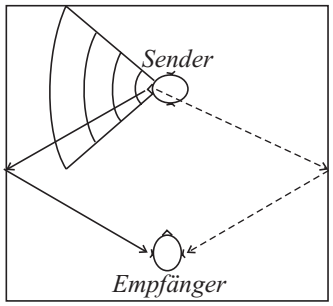

Abbildung 3: Versuchsaufbau: Sender: menschl. Sprecher, Empfänger: Kemar

bestimmten Bändern mehr Energie relativ zum anderen Sensor auftritt. Somit scheint es möglich, dies als ein Kriterium zur Sprachrichtungserkennung zu nutzen. Die Signalverarbeitung zeigt Abb.4.

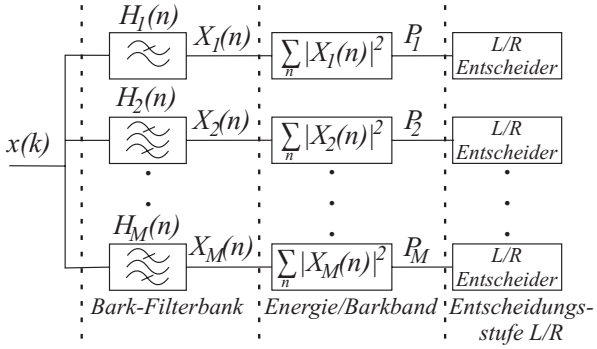

Abbildung 4: Blockschaltbild der Signalverarbeitung.

4 Zusammenfassung

Die Sprecherlokalisation in geschlossenen Räumen zeigt gute Ergebnisse. Die Sprachrichtungserkennung ist jedoch bereits bei geringem Nachhall eine große Herausforderung. Der hier gezeigte Ansatz kann als ein Kriterium zur Sprachrichtungserkennung genutzt werden. Weitere Untersuchgen sind notwendig, um die vom Menschen genutzten Hinweise zur Sprachrichtungsschätzung auch technischen Systemen zur Verfügung zu stellen.

Literatur

[BOS05] Brutti, A., Omologo, M. und Svazier, P.: *Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays.* Interspeech, Seiten 2337–2340, 2005.

[BYE00] Benesty, J., Yiteng, H. und Elko, G.: *Microfone Arrays for Video Camera Steering.* In: Benesty, J. (Herausgeber): *Acoustic Signal Processing for Telecommunication*, Kapitel 11, Seiten 239–259. Berlin, Heidelberg, New York, 2000.

[Roe07] Roeske, P.: *Sprecherlokalisation in gestörter Umgebung,* Diplomarbeit. 2007.