

Zahlentheoretisch motivierte Optimierung von Mikrofonpositionen für einen virtuellen Kunstkopf

Eugen Rasumow¹, Matthias Blau¹, Simon Doclo², Dirk Püschel³, Martin Hansen¹, Steven van de Par²,
Volker Mellert²

¹ Institut für Hörtechnik und Audiologie, Jade Hochschule, D-26121 Oldenburg

² Carl von Ossietzky Universität Oldenburg, D-26121 Oldenburg

³ Akustik Technologie Göttingen, D-37073 Göttingen

Einleitung

Um kopfbezogene Übertragungsfunktionen des Menschen (**Head Related Transfer Function, HRTF**) in akustische Messungen mit einzubeziehen, werden oft Kunstköpfe benutzt. Traditionelle Kunstköpfe sind Nachbildungen menschlicher Köpfe mit mittleren anthropometrischen Geometrien und eingelassenen Ohren. Dadurch erfahren die an den Mikrofonen ankommenden Schallsignale eine charakteristische frequenz- und richtungsabhängige Verfärbung, ähnlich wie beim Menschen. Diese frequenzabhängigen Richtcharakteristiken können jedoch auch mithilfe von Mikrofonarrays angenähert werden. Dabei werden N Mikrofone in einer festen Konstellation angeordnet und für jedes Mikrofon ein Filter $\underline{\mathbf{w}}(\omega)$ berechnet, sodass die Summe der gefilterten Mikrofonensignale am besten den vorgegebenen Richtcharakteristiken entspricht. Ein Mikrofonarray hat gegenüber herkömmlichen Kunstköpfen den Vorteil, dass es flexibel z.B. an verschiedene Anatomien angepasst werden kann, auch wenn eine Aufnahme schon gemacht ist. Die relative Position der einzelnen Mikrofone hat jedoch einen großen Einfluss auf die Anpassgenauigkeit bei der Optimierung der Filter und die Robustheit des Ergebnisses bei eventuellen Positionsungenauigkeiten der Mikrofone.

Least Squares Optimierung

Für die Berechnung der optimalen komplexen Filtergewichte $\underline{\mathbf{w}}(\omega)$ wurde eine schmalbandige Least-Squares Kostenfunktion $J(\underline{\mathbf{w}}(\omega))$ zwischen der gewünschten Richtcharakteristik $\underline{D}(\omega, \Theta_i)$ und der aus den Gewichten und dem steering vector $\underline{\mathbf{d}}(\omega, \Theta_i)$ resultierenden Richtcharakteristik $\underline{H}(\omega, \Theta_i) = \underline{\mathbf{w}}^H(\omega) \underline{\mathbf{d}}(\omega, \Theta_i)$ aufgestellt.

$$J(\underline{\mathbf{w}}(\omega)) = \sum_{i=1}^P |H(\omega, \Theta_i) - D(\omega, \Theta_i)|^2 \quad (1)$$

$$\underline{\mathbf{w}}(\omega) = \underline{\underline{\mathbf{Q}}}^{-1}(\omega) \underline{\mathbf{a}}(\omega) \quad (2)$$

$$\underline{\underline{\mathbf{Q}}}(\omega) = \sum_{i=1}^P \underline{\mathbf{d}}(\omega, \Theta_i) \underline{\mathbf{d}}^H(\omega, \Theta_i) \quad (3)$$

$$\underline{\mathbf{a}}(\omega) = \sum_{i=1}^P \underline{\mathbf{d}}(\omega, \Theta_i) D^*(\omega, \Theta_i) \quad (4)$$

Die Variable P gibt die Anzahl der diskreten Richtungen Θ_i der Richtcharakteristiken an. Die Kostenfunktion $J(\underline{\mathbf{w}}(\omega))$ wurde, in Abhängigkeit des Filtervektors mit Einträgen für alle verwendeten Mikrofone, für jede Frequenz ω einzeln optimiert. Die Frequenzgewichte $\underline{\mathbf{w}}(\omega)$ zur Minimierung der Kostenfunktion $J(\underline{\mathbf{w}}(\omega))$ können nach [1] durch Gleichung 2-4 berechnet werden. Für die

Berechnung der steering vectors $\underline{\mathbf{d}}(\omega, \Theta_i)$ wurden dabei omnidirektionale Mikrofone (d.h. die Mikrofone beeinflussen sich nicht gegenseitig) und die Ausbreitung im Fernfeld angenommen.

Random-Sampling Positionierung

Für die folgenden Simulationen wurde eine Fläche von 10cm x 10cm für mögliche Mikrofonpositionen angenommen. Voraussetzung für die skizzierte Optimierung ist, dass die Positionen der Mikrofone und damit der steering vector $\underline{\mathbf{d}}(\omega, \Theta_i)$ bekannt sind. Somit wurden für 10000 unterschiedliche, zufällig angeordnete Mikrofonanordnungen die optimalen Filtergewichte berechnet und die Mikrofonanordnung mit dem geringsten Least-Squares-Fehler gewählt. Diese Optimierung wurde weiterhin für verschiedene Mikrofonanzahlen N mit $4 \leq N \leq 24$ durchgeführt.

Golomb Positionierung

Der Abstand zwischen zwei Mikrofonen bei einem Mikrofonarray bestimmt die frequenzabhängige Keulbreite in der Richtcharakteristik. Um eine optimale Anpassung an die variierende Richtcharakteristik $\underline{D}(\omega, \Theta_i)$ zu gewährleisten, muss in jede Richtung die größtmögliche Varianz an Abständen zwischen verschiedenen Mikrofonen gegeben sein. Basierend auf diesem Gedanken bietet es sich an, für das Gitter der möglichen Mikrofonpositionen Redundanzen zu minimieren. Bezogen auf eine Dimension würde ein Golomb-Lineal eine bestmögliche Frequenzauflösung bieten, da dieser Maßstab keinen Abstand zwischen zwei Mikrofonen mehr als einmal enthält. Um diesen Maßstab auf mehrdimensionale Anordnungen zu erweitern, wurde in [2] ein Verfahren entwickelt, bei dem ein Golomb-Lineal der Ordnung M errechnet und daraus ein Gitter aufgespannt wird. Auf diesem Gitter kann zu Beginn der erste der M^2 (2D) bzw. M^3 (3D) möglichen Knotenpunkte zufällig ausgewählt werden. Im Folgenden können weitere, zufällig gewählte Knotenpunkte auf ihre Legitimität geprüft werden. Ein Knotenpunkt gilt als legitim, wenn die Bedingung erfüllt ist, dass das durch den aktuellen Knotenpunkt entstandene Muster sich in dieser Form nicht in der Anordnung wiederholen darf. Diese Bedingung kann überprüft werden, indem das aktuelle Knotenpunktmuster (inklusive zu testendem Knotenpunkt) reproduziert und entlang aller möglichen Rasterpunkte und Dimensionen des verwendeten Gitters verschoben wird. Während beim Verschieben 0 alle Knotenpunkte des Original-Knotenpunktmusters über den Knotenpunkten des verschobenen Knotenpunktmusters liegen, dürfen sich bei jedem weiteren Verschieben nie an

mehr als einem Ort die Knotenpunkte beider Muster überlagern. Ist diese Bedingung erfüllt, wird der zu testende Knotenpunkt in das Knotenpunktmuster aufgenommen. Andernfalls wird dieser Knotenpunkt verworfen und ein weiterer Knotenpunkt getestet. Dadurch, dass die Knotenpunkte randomisiert ausgewählt werden, können bei fester Ordnung M unterschiedliche Anordnungen mit unterschiedlich vielen legitimen Knotenpunkten entstehen. Für die folgenden Simulationen wurde darauf geachtet, dass die resultierende Anordnung bei $N = 24$ eine möglichst gute räumliche Verteilung der Mikrofone beinhaltet. Diese Methode der Mikrofonpositionierung ist viel weniger rechenintensiv als die Random Sampling Methode.

Vergleich beider Positionierungsverfahren

Mithilfe beider Positionierungsverfahren wurden die optimalen Mikrofonpositionen (im Sinne des kleinsten Least-Squares Fehlers) für die Mikrofonanzahlen $N = 4, 6, 8, 10, 12, 14, 16, 18, 19, 20, 22, 23, 24$ separat berechnet. Weiterhin wurde zur Analyse von $\underline{w}(\omega)$ ein zufälliger Positionierungsfehler jedes einzelnen Mikrofons von $|\Delta x_{max}|, |\Delta y_{max}| \leq 1$ mm in $\underline{d}(\omega, \Theta_i)$ berücksichtigt, um ein realistisches Szenario zu schaffen und die Mikrofonanordnungen bezüglich ihrer Robustheit untersuchen zu können. Zur Analyse der Fehler wurde der Quotient aus gewünschter und resultierender Richtcharakteristik in kritischen Bändern (mit 50% Überlapp) gemittelt.

$$\delta_{CB}(\Theta_i) = \frac{1}{M} \sum_{\omega=\omega_1}^{\omega_M} \frac{H(\omega, \Theta_i)}{D(\omega, \Theta_i)}$$

M gibt dabei die Anzahl der Frequenzbins, ω_1 die unterste und ω_M die oberste Frequenz im jeweiligen kritischen Band (CB) an. Daraus lässt sich das Fehlermaß $E(CB)$ für jede Mikrofonanzahl N berechnen.

$$E(CB) = \frac{1}{P} \sum_{i=1}^P \left| 20 \lg(|\delta_{CB}(\Theta_i)|) \text{ dB} \right|$$

Abbildung 1 zeigt den über alle P diskreten Einfallrichtungen Θ_i gemittelten Fehler $E(CB)$ der Anpassung an HRTF-Daten (Horizontalebene, subject 1005, Auflösung von 15°) der IRCAM-Datenbank unter Verwendung der Random-Sampling (oben) und der Golomb Positionierung (unten). Anhand Abbildung 1 ist deutlich zu erkennen, dass die Golomb-Anordnung der Mikrofone im allgemeinen geringere Fehler bei der Anpassung ermöglicht. Besonders für größere Mikrofonanzahlen mit ca. $N \geq 12$ ermöglicht die Golomb-Anordnung eine robustere Anpassung im Sinne des gezeigten Fehlermaßes. Dieser Effekt wird größer bei größeren Positionierungsfehlern und geringer, wenn keine Positionsfehler einbezogen werden. Darüber hinaus ermöglicht die Golomb Positionierung für $N \geq 12$ ein breiteres Frequenzspektrum robuster Übertragung. Hierbei ist ebenfalls zu beachten, dass die Golomb Positionierung mit $N = 24$ Mikrofonen, durch ihren sukzessiven Charakter, bereits alle Mikrofonanordnungen mit $N < 24$ beinhaltet. Für die Random-Sampling Positionierung hingegen, kann bei festem Aufbau nicht die Mikrofonanzahl adaptiv variiert werden. Bei Mikrofonanzahlen $N \leq 12$ ergibt die Golomb Positionierung eine etwas schlechtere Performance als die

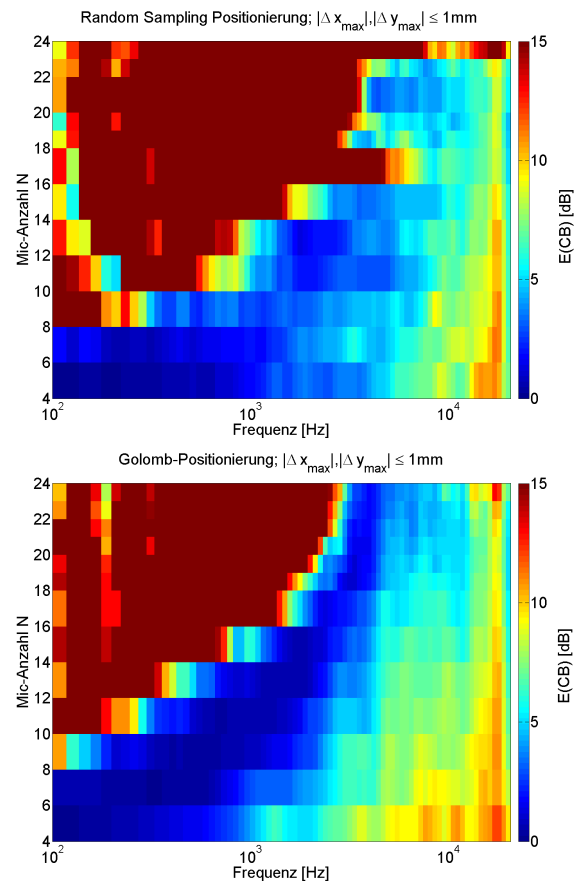


Abbildung 1: Fehler $E(CB)$ bei Verwendung der Random-Sampling (oben) und der Golomb Positionierung (unten). Auf der x-Achse ist logarithmisch die Frequenz in Hz und auf der y-Achse die Mikrofonanzahl aufgetragen. Fehler $E(CB) \geq 15$ dB werden einheitlich in braun dargestellt.

Random-Sampling Positionierung. Dieses kann daran liegen, dass bei der Golomb Positionierung lediglich bei $N = 24$ eine geeignete räumliche Verteilung der Mikrofone kontrolliert wurde.

Fazit

Die Golomb Positionierung ermöglicht eine weniger rechenintensive Berechnung geeigneter Mikrofonpositionen als die Random-Sampling Positionierung. Ein weiterer Vorteil der Golomb Positionierung ist die zum großen Teil bessere Robustheit bezüglich Positionierungsfehlern und die variable Mikrofonanzahl bei festem Aufbau.

Danksagung

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 17080X10 gefördert.

Literatur

- [1] Doclo S. & Moonen M., Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. IEEE Transaction on signal processing, 51(10):2511-2526, October 2003.
- [2] Sender- bzw. Sensoranordnung zur Erzielung optimierter Richtcharakteristik für Sende- oder Empfangsvorrichtungen in zwei und drei Dimensionen, DE 10 2010 012 388 A1