Aufnahme und Wiedergabe räumlicher Schallszenen: Fehleranalyse und Korrekturansatz

Matthieu Kuntz, Bernhard U. Seeber

Professur für Audio-Signalverarbeitung, Technische Universität München, Email: {matthieu.kuntz, seeber}@tum.de

Einleitung

Ambisonics erlaubt die weitgehend akkurate Aufnahme und Wiedergabe räumlicher Schallszenen. Diese reproduzierbare und kontrollierbare Wiedergabe ist für die Hörforschung äußerst interessant, da sie reale akustische Situationen ins Labor bringen, und somit für Hörtests in komplexeren Situationen verwendet werden kann. Die pegel- und zeitgetreue Wiedergabe der aufgenommenen Schallszenen ist jedoch schwierig. Dieser Beitrag behandelt einzelne Schritte dazu. Die Kette von der Aufnahme der Schallszenen mit einem Mikrofonarray bis zur Wiedergabe über ein Lautsprecherarray wird analysiert. Die auftretenden Fehler werden diskutiert und ein Korrekturansatz dafür vorgestellt.

Setup

Für die Aufnahme der Schallszenen wurde ein 36-kanäliges, zylindrisches Mikrofonarray verwendet. Mit einem Radius von 15.75 cm und einer Höhe von 57.20 cm kann es mit einem unendlich langen Zylinder angenähert werden. Damit können Schallszenen mit Higher-Order Ambisonics (HOA) 17. Ordnung horizontal aufgezeichnet werden. Die Grenzfrequenz dieses Arrays beträgt 11.3 kHz [1].

Für die Wiedergabe der Aufnahmen wurde der horizontale Ring des 'real-time Simulated Open Field Environment'(rtSOFE, Abb. 2), ein 36-kanäliges Lautsprecherarray, verwendet. Die im Rechteck angeordneten Lautsprecher wurden neben der Frequenz- und Phasenentzerrung auch laufzeitentzerrt, sodass sie ein virtuelles kreisförmiges Array bilden. Die virtuell kreisförmige Anordnung zielt auf eine Wiedergabe mit Ambisonics.

Simulation

Es wurden für die Simulation 2-dimensionale Schallszenen mit einer Quelle im Fernfeld betrachtet. Auch die Lautsprecher werden als Quellen im Fernfeld beschrieben, damit das Ambisonics-Verfahren angewendet werden kann. Die Abb. 3 zeigt ein Flowchart des Ablaufs, der hier kurz erkärt ist. Das Quellsignal wird mit Impulsantworten der Mikrofone gefaltet, um das aufgenommene Signal zu berechnen. Diese werden von dem Zeit- in den Frequenzbereich transformiert, wo mit Ambisonics die passenden Lautsprechersignale ausgerechnet werden. Im Zeitbereich werden die Lautsprechersignale addiert um das Signal am Mittelpunkt des Lautsprecherarrays auszurechnen.

In diesem Beitrag wurden nur Dirac-Impulse verwendet

Abbildung 1: Das Lautsprecherarray des rtSOFE. Verwendet wurde der horizontale Teil des Arrays. Foto von Bernhard Seeber.

als Quellsignale verwendet. So ließen sich das wiedergegebene Zeitsignal und der dazugehörige Frequenzgang leicht abbilden und analysieren. Eine einfache Faltung dieser Impulsantworten mit einem beliebigen Quellsignal reicht, um dessen Wiedergabe zu beobachten.

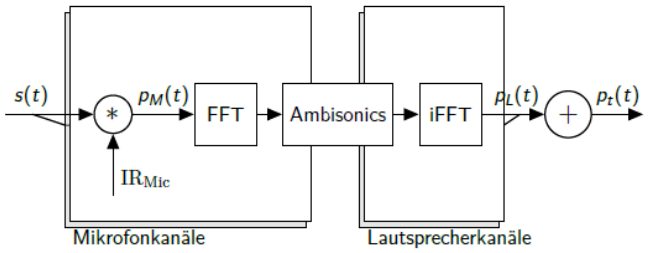

Abbildung 2: Simulation der Aufnahme und Wiedergabe räumlicher Schallszenen.

Berechnen der Mikrofonsignale

Um die zu erwartende Mikrofonsignale zu berechnen, wird das Polarkoordinatensystem $P(r, \phi)$ mit Ursprung im Zylindermittelpunkt. Es werden folgende kreisharmonische Funktionen definiert [2]:

$$Y_n^m(\phi) = \begin{cases} 1 & \text{für} \quad n = 0 \\ \sqrt{2}\cos(m\phi) & \text{für} \quad m = n \\ \sqrt{2}\sin(-m\phi) & \text{für} \quad m = -n \end{cases} \tag{1}$$

Der Schalldruck an der Oberfläche des Zylinders $p_t(kr, \phi)$ setzt sich aus der Summe von einer auftreffenden ebenen Welle $p_i(kr, \phi)$ und einer vom Zylinder ausgehenden

kreisförmigen Welle $p_s(kr, \phi)$ zusammen. Beide Wellen können als Summe von Kreisharmonischen beschrieben werden [3].

$$p_i(kr, \phi) = P_0 \sum_{n=0}^{+\infty} \mathrm{i}^n \mathrm{J}_n(kr) \sum_{m=\pm n} Y_n^m(\phi) Y_n^m(\phi_i), \quad (2)$$

$$p_s(kr, \phi) = -P_0 \sum_{n=0}^{+\infty} \mathrm{i}^n \left[\frac{\mathrm{J}'_n(ka)}{\mathrm{H}_n^{\prime(2)}(ka)} \mathrm{H}_n^{(2)}(kr) \right] \sum_{m=\pm n} Y(\phi) Y(\phi_i), \quad (3)$$

$$p_t(kr, \phi) = P_0 \sum_{n=0}^{+\infty} w_n(kR) \sum_{m=\pm n} Y_n^m(\phi) Y_n^m(\phi_i), \quad (4)$$

wo

$$w_n(kr) = \mathrm{i}^n \left[\mathrm{J}_n(kr) - \frac{\mathrm{J}'_n(kR)}{\mathrm{H}_n^{\prime(2)}(kR)} \mathrm{H}_n^{(2)}(kr) \right] \quad (5)$$

die frequenzabhängige Gewichtung der einzelnen Ordnungen beschreibt [2]. Abb. 3 und 4 zeigen die Ergebnisse für einzelne Mikrofone, für eine Quelle mit einem Schalldruckpegel von 70 dB SPL.

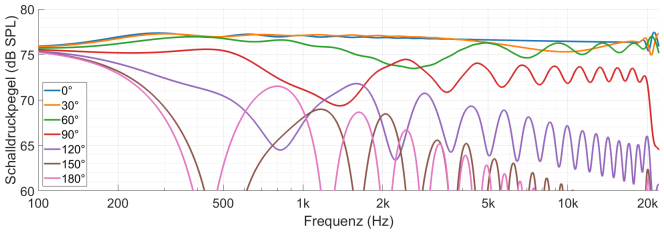

Abbildung 3: Berechneter Frequenzgang einzelner Mikrofone.

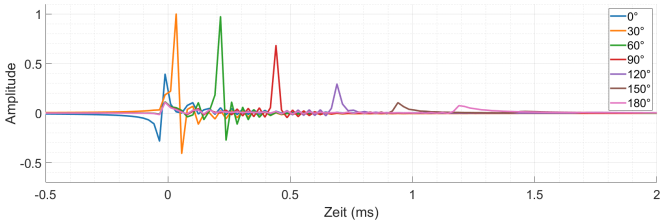

Abbildung 4: Berechnete Impulsantworten einzelner Mikrofone.

Vergleich mit Messungen

Die Mikrofonimpulseantworten wurden auch in einem reflexionsarmen Raum gemessen. Sie sind für ausgewählte Mikrofone auf Abb. 5 zu sehen.

Es treten, wie in der Simulation, Laufzeitunterschiede von 1.2 ms zwischen dem der Quelle zugewandten und dem gegenüberliegenden Mikrofon auf. Auch die Amplitudenunterschiede zwischen den Mikrofonen scheinen gut von der Simulation berechnet zu werden.

Der Unterschied zwischen den Messungen und der Berechnung ist in den Frequenzgängen (Abb. 6) deutlich zu sehen: Es wurde an den tiefen Frequenzen kein Staudruck von 6 dB gemessen. Das deutet darauf hin, dass die Beschreibung des Schalldrucks um den Zylinder als Summe von zwei Wellen an tiefen Frequenzen nicht ganz zutrifft.

Der Abfall der Frequenzgänge über 6 kHz liegt an den Mikrofonen. Es wurden Freifeldmikrofone im Array verbaut, die mit einem Tiefpassfilter die Reflektionen des Schalls auf der Membran des Mikrofons korrigiert, um den Einfluss des Mikrofons auf das Schallfeld aus den Messungen auszugleichen. Im Falle eines geschlossenen Mikrofonarrays ist diese Reflektion jedoch erwünscht, da das Array als voll-reflektierend angenähert wird.

Für die Berechnung der Mikrofonsignale in der Simulation wurden die gemessenen Impulsantworten gewählt.

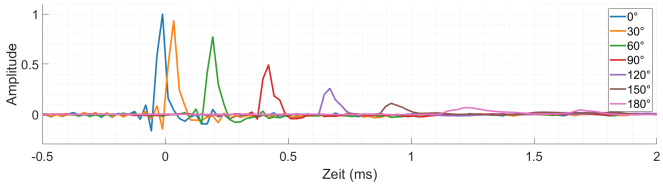

Abbildung 5: Gemessene Impulsantworten für verschiedene Mikrofone.

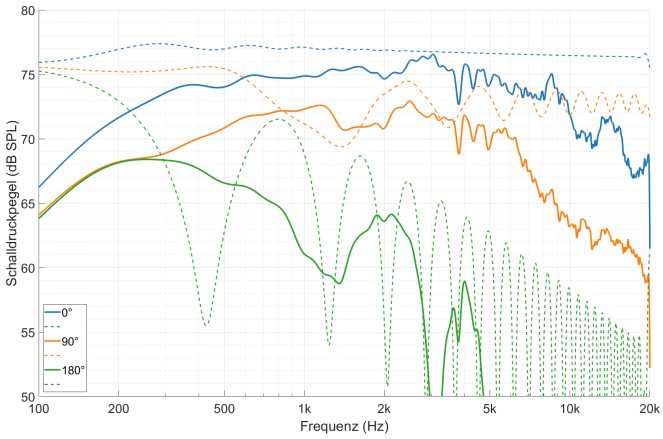

Abbildung 6: Gemessene Frequenzgänge für verschiedene Mikrofone. Die ausgerechneten Frequenzgänge sind als gestrichelte Linien dargestellt.

Ambisonics-Verfahren

Zur räumlichen Wiedergabe wurde HOA mit dem Standard Dekoder verwendet. Sie ist hier nochmal kurz zusammengefasst. Das Schallfeld um das zylindrische Array kann als gewichtete Summe von kreisharmonischen Funktionen beschrieben werden:

$$p_t(kr, \phi) = P_0 \sum_{n=0}^{+\infty} w_n(kR) \sum_{m=\pm n} b_n^m \cdot Y_n^m(\phi) \quad (6)$$

Diese Gleichung kann nun in Matrix-Form, $\mathbf{p} = \mathbf{Y_M} \cdot \mathbf{W} \cdot \mathbf{b}$, invertiert werden, um die sogenannten Ambisonics-Koeffizienten anhand der gemessenen Schalldrucke $\mathbf{p}$ zu bestimmen:

$$\mathbf{b} = \mathbf{W^{-1}} \cdot \mathbf{Y_M}^{+} \cdot \mathbf{p}, \quad (7)$$

mit

$$\mathbf{Y_M} = \begin{bmatrix} Y_0^0(\phi_1) & Y_1^{-1}(\phi_1) & Y_1^1(\phi_1) & \cdots & Y_n^n(\phi_1) \\ \vdots & \vdots & \vdots & & \vdots \\ Y_0^0(\phi_M) & Y_1^{-1}(\phi_M) & Y_1^1(\phi_M) & \cdots & Y_n^n(\phi_M) \end{bmatrix}$$

und

$$\mathbf{W} = \mathrm{diag}(w_0, w_1, w_1, \cdots, w_n, w_n).$$

$[]^+$ bezeichnet die Pseudoinverse einer Matrix, da die Matrix $\mathbf{Y_M}$ nicht quadratisch ist.

Die Winkel $[\phi_1, \cdots, \phi_M]$ bezeichnen die Azimuthwinkel der Mikrofone, $[\phi_1, \cdots, \phi_L]$ die der Lautsprecher. Bei dem verwendeten Setup sind diese zwei Vektoren, und daher auch $\mathbf{Y_M}$ und $\mathbf{Y_L}$ gleich. Die Lautsprechersignale werden anhand der Ambisonics-Koeffizienten und der Lautsprecherpositionen bestimmt:

$$\mathbf{s} = \frac{1}{N_L} \mathbf{Y_L} \cdot \mathbf{b}. \tag{8}$$

Ergebnisse der Simulation

Dieses Verfahren wurden mit den gemessenen Mikrofonimpulsantworten angewendet, um das wiedergegebene Schallfeld am Arraymittelpunkt zu bestimmen. Das Zeitsignal und der Frequenzgang sind auf Abbildungen 7 und 8 zu sehen. Die Energie des Impulses E dient als Maß der Qualität der Wiedergabe. Die Laufzeitunterschiede werden wie erwartet verzögert wiedergegeben. Dies führt zu einer höheren Energie im Signal, und einem tiefpassähnlichen Frequenzgang.

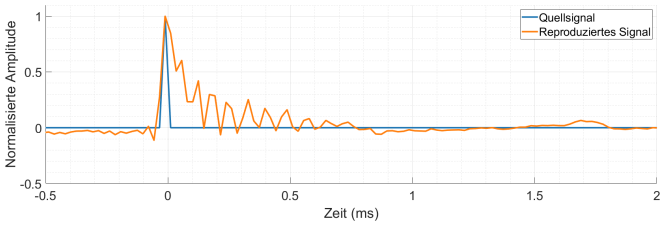

Abbildung 7: Simulierte Wiedergabe der gemessenen Impulse. $E = 3.2$

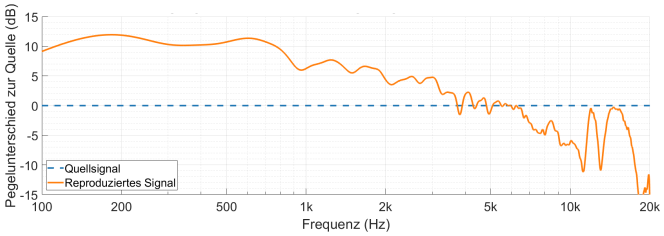

Abbildung 8: Frequenzgang der simulierten Wiedergabe der gemessenen Impulse. RMS-Fehler: 5.3 dB

Korrekturansatz

Um das wiedergegebene Zeitsignal zu verbessern, wird folgender Ansatz vorgeschlagen: Die Mikrofonsignale werden vor dem Ambisonics-Verfahren richtungsabhängig verzögert, um die Laufzeitunterschiede während der Aufnahme auszugleichen.

Die horizontale Position der Quelle wird mit einem GCC-PHAT basiertem Ansatz geschätzt. Mit dieser Information können die Laufzeitunterschiede zwischen den Mikrofonen ausgerechnet und mit einem einfachem Filter ausgeglichen werden. Das neue Zeitsignal ist deutlich näher an einem Impuls, mit einer Energie von 1.6 anstatt von 3.2. Der RMS-Fehler im Frequenzgang wird von 5.3 dB auf 2.8 dB verbessert.

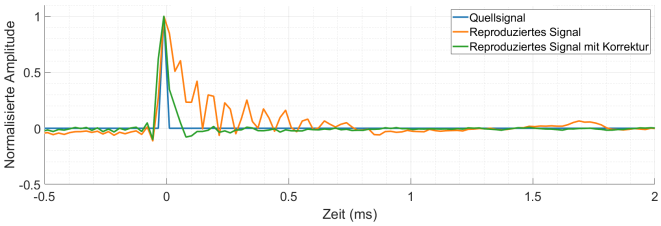

Abbildung 9: Simulierte Wiedergabe der gemessenen Impulse mit Laufzeitkorrektur $E = 1.6$

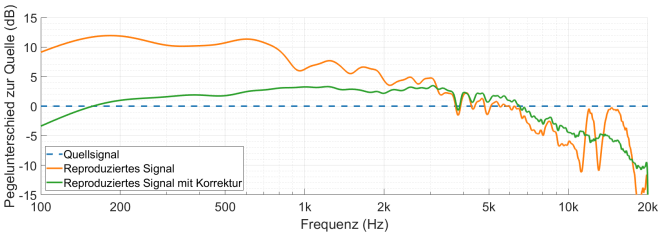

Abbildung 10: Frequenzgang der simulierten Wiedergabe der gemessenen Impulse mit Laufzeitkorrektur. RMS-Fehler: 2.8 dB

Lokalisierung

Ein entscheidender Punkt für Schallfeldwiedergabeverfahren, vor allem, wenn diese für Hörforschung verwendet werden, ist die Lokalisierung der wiedergegebenen Quellen durch Probanden. Ein einfaches Maß dafür ist der Energievektor r_E, der aus der Verstärkung der verschiedenen Lautsprecher g_i und deren Position ϕ_i bestimmt wird:

$$r_E = \frac{\sum_i g_i^2 \phi_i}{\sum_i g_i^2}, \quad ||r_E|| \in [0, 1]. \tag{9}$$

Wenn ein einziger Lautsprecher aktiv ist, ist der Betrag des Energievektors 1 und die virtuelle Quelle kann von Probanden optimal lokalisert werden. Desto niedriger der Betrag, desto diffuser ist das Schallfeld, was die Lokalisierung beeinträchtigt. Es wird also angestrebt, den Betrag des Energievektors zu maximieren. Ein Ansatz dazu ist zum Bespiel die sogenannte max-r_E Ambisonics Dekodierung.

Abbildung 11 repräsentiert die Lautsprecherverstärkung im Verhältnis zur Lautsprecherposition. Diese Verstärkung wurde aus den aufgenommenen, laufzeitkorrigierten Impulsantworten mit Ambisonics ausgerechnet. Der Betrag des Energievektors ist mit dieser Methode 0.61.

Durch ein Eingreifen vor dem Ambisonics-Verfahren, indem die Mikrofonsignale nicht nur richtungsabhängig

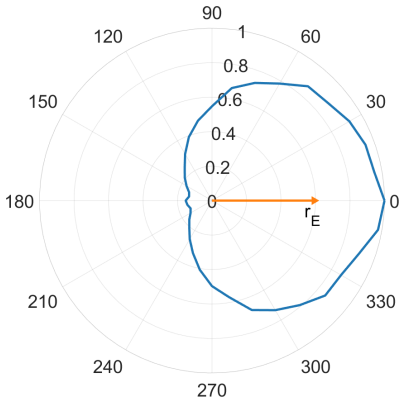

Abbildung 11: Lautsprecherverstärkung versus Winkel. $||r_E|| = 0.61$

verzögert, sondern auch gewichtet werden, kann der Betrag des Energievektors erhöht werden.

Das Beispiel auf Abbildung 12 wurde mit einer Gauß'schen Gewichtung der Mikrofonsignale berechnet. Die Standardabweichung der Gausskurve betrug 29 Grad (0.5 rad). So kommt der Betrag des Energievektors auf 0.91. Man bemerkt, dass die Richtung des Energievektors konstant bleibt. Auch das Zeitsignal und der Frequenzgang des wiedergegebenen Impulses verändern sich kaum, sodass die Qualität des wiedergegebenen Signals nicht beeinträchtigt wird. Dies setzt jedoch eine genaue Schätzung der Schalleinfallsrichtung voraus.

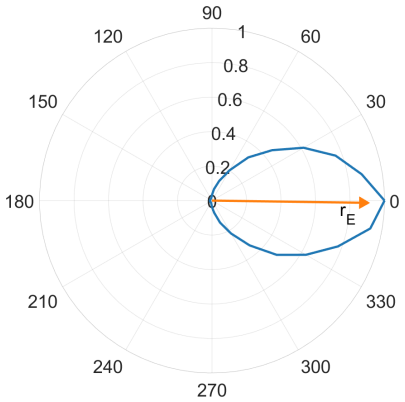

Abbildung 12: Lautsprecherverstärkung versus Winkel, mit Gewichtung der Mikrofonsignale. $||r_E|| = 0.91$

Weitere Forschung

Weitere Forschung beschäftigt sich mit der erweiterung der Laufzeitkorrektur auf Schallszenen mit mehreren Quellen und die Evaluierung von anderen Ambisonics-Dekodierungen, zum Beispiel max-r_E und Near-Field Compensated HOA [4].

Literatur

[1] Bihler, F. (**2017**). "Design and construction of a circular microphone array for recording and reproduction of acoustic scenes", Master's thesis, Professur für Audio-Signalverarbeitung, Technische Universität München.

[2] Parthy, A., Epain, N., van Schaik, A., and Jin, C. T. (**2011**). "Comparison of the measured and theoretical performance of a broadband circular microphone array", The Journal of the Acoustical Society of America **130** , 3827–3837.

[3] Morse, P. and Ingard, K. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York), p.360.

[4] Daniel, J. (**2003**). "Spatial sound encoding including near field effect: introducing distance coding filters and a viable, new ambisonic format", in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction.*