Wiedergabe von Schallfeldern: Analyse der interaurale Merkmale

Matthieu Kuntz, Bernhard U. Seeber

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

Einleitung

Virtuelle akustische Umgebungen mit Freifeldwiedergabe des Schalls ermöglichen es, bekannte und kontrollierbare akustische Szenen über Lautsprecher zu synthetisieren und so die Nachteile einer HRTF-basierten Wiedergabe zu beheben [1,2,3]. Sie können in der Hörforschung, sowohl für allgemeine psychoakustische Forschung als auch für die Entwicklung von Hörgerätealgorithmen verwendet werden [3]. Einige dieser Hörgerätestrategien bauen auf Kodierung der interauralen Merkmale auf [4]. Dadurch ist natürlich die korrekte Wiedergabe der interauralen Merkmale und Evaluation dieser äußerst wichtig.

Es ist bekannt, dass Ambisonics die Lokalisierung beeinflusst und dass dies unter anderem von der Ambisonicsordnung, dem verwendeten Dekoder und der Zuhörerposition abhängt [5,6]. Die synthetisierte Quelle ist breiter [7], die interauralen Merkmale sind seitlich nicht unbedingt korrekt wiedergegeben [8] und das Sprachverständnis ist davon beeinflusst [9].

Ein gängiger Ansatz, um die Lokalisierung, die mit dem einfachen *basic* Dekoder außerhalb des Sweetspots und für höhere Frequenzen erreicht wird, zu verbessern, ist die Anwendung von modifizierten Dekodern, *max* r_E [10] und *in-phase* [11]. Bei *max* r_E wird der Energievektor $\overrightarrow{r_E}$ des Schallfeldes maximiert, der an hohen Frequenzen die Lokalisierung maßgeblich beeinflusst [12]. Bei *in-phase* nimmt der Lautsprecherpegel mit Abstand zur Richtung der virtuellen Quelle ab. Dieser unterdrückt die Nebenkeulen der Ambisonics Panning-Funktion komplett und wirkt dem Präzedenzeffekt entgegen [13].

Daher wurde empfohlen, bei tiefen Frequenzen und Zuhörerpositionen in der Nähe des Sweetspots den *basic* Dekoder zu verwenden und mit steigender Frequenz und Verschiebung entgegen der Mitte erst *max* r_E und dann *in-phase* zu verwenden [13]. Allerdings zeigten Frank *et al.* [5], in einem Lokalisierungsexperiment, dass der *in-phase* Dekoder im Mittelpunkt und an einer Seitenposition des verwendeten Lautsprecherarrays die höchsten Fehler aufwies.

In diesem Beitrag werden die interauralen Merkmale einer mit verschiedenen Dekodern synthetisierten Schallquelle bei $\varphi_S = 0°$ dargestellt und analysiert.

Setup

Simulation

In diesem Beitrag wird ein rechtshändiges Kartesisches Koordinatensystem [x,y,z] verwendet, wo z nach unten zeigt. Der Azimuthwinkel φ wird als Winkel zwischen x und y definiert, sodass der Winkel im Uhrzeigersinn zunimmt.

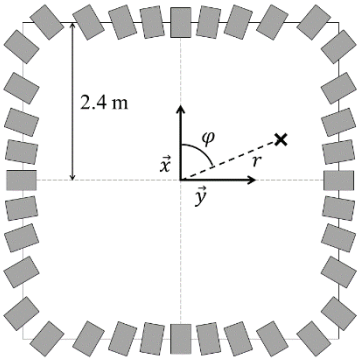

Abbildung 1: Simuliertes Lautsprecherarray. Die Lautsprecher sind in Winkelabständen von 10° montiert und pegel-, laufzeit- und phasenentzerrt.

Das simulierte Lautsprecherarray besteht aus $L = 36$ Lautsprechern, die auf einem viereckigen Rahmen mit einer Seitenlänge von 4,8 m montiert und auf dessen Mittelpunkt gerichtet sind (Aufbau der rtSOFE der TUM [14]). Die Lautsprecher sind pegel-, laufzeit- und phasenentzerrt, sodass sie effektiv in einem Kreis mit Radius 2,4 m platziert sind. Sie sind im gleichmäßigen Winkelabstand positioniert ($\varphi_{LS,1} = 0°, \varphi_{LS,2} = 10°, \dots, \varphi_{LS,36} = 350°$), vgl. Abb. 1. Das Lautsprecherarray ermöglicht eine 2D Ambisonicsordnung $\mathrm{N} = 17$.

Berechnung der Lautsprechersignale

Es wurde eine virtuelle Schallquelle im Fernfeld, mit einem Winkel $\varphi_\mathrm{S} = 0°$ zum Lautsprecherarray virtuell wiedergegeben. Um die Ambisonicssignale $\boldsymbol{\chi}(\mathrm{t})$ der Quelle $\mathrm{s}(\mathrm{t})$ auszurechnen, wird das Signal richtungsabhängig enkodiert. Der Enkodierungsvektor $\mathbf{y}(\varphi_\mathrm{S})$ ergibt sich aus den Fourier-Koeffizienten einer Dirac-Verteilung auf einem Kreis, bis zur Ordnung N [15]:

$$\mathbf{y}(\varphi_\mathrm{S}) = \frac{1}{\sqrt{\pi}}\left[\frac{1}{\sqrt{2}}, \cos(\varphi_\mathrm{S}), \sin(\varphi_\mathrm{S}), \dots, \cos(\mathrm{N}\varphi_\mathrm{S}), \sin(\mathrm{N}\varphi_\mathrm{S})\right]^\mathrm{T},$$

$$\boldsymbol{\chi}(\mathrm{t}) = \mathbf{y}(\varphi_\mathrm{S}) \times \mathrm{s}(\mathrm{t}) \,. \qquad (1)$$

Bei dem *basic* Dekoder werden die Ambisonicssignale an den Lautsprecherpositionen φ_LS abgetastet und richtungsabhängig gewichtet, um die Lautsprechersignale $\mathrm{s}_\mathrm{LS}(\mathrm{t})$ zu berechnen:

$$\mathrm{s}_\mathrm{LS}(\mathrm{t}) = \sqrt{\frac{2\pi}{\mathrm{L}}}\left[\mathbf{y}(\varphi_{\mathrm{LS},1}), \dots \mathbf{y}(\varphi_{\mathrm{LS,L}})\right]^\mathrm{T} \times \boldsymbol{\chi}(\mathrm{t}) \qquad (2)$$

Um Nebenkeulen in der Ambisonics Panning-Funktion zu reduzieren, die durch die begrenzte Ordnung in der Zerlegung der Dirac-Verteilung entstehen, können die Ambisonicssignale vor dem Abtasten an den Lautsprecherpositionen ordnungsabhängig gewichtet werden. Die Dekoder *max* r_E und *in-phase* verwenden unterschiedliche Gewichte, um jeweils den Energievektor $\overrightarrow{r_E}$ zu maximieren oder die

Tabelle 1: Gewichte für *basic*, *max* r_E und *in-phase* Dekoder [13]

Dekoder	*basic*	*max* r_E	*in-phase*
w_n	1	$\cos\left(\frac{n\pi}{2(N+1)}\right)$	$\frac{(N!)^2}{(N-n)!\,(N+n)!}$

Nebenkeulen zu vermeiden. Die Gewichte, die diese Dekoder verwenden, sind in Tabelle 1 eingetragen. Gleichungen (1) und (2) werden als Matrixgleichungen umgeschrieben und die diagonale Gewichtungsmatrix **W** wird eingefügt:

$$\mathbf{W} = \mathrm{diag}(w_0, w_1, w_1, w_2, \dots, w_N, w_N),$$

$$s_{LS}(t) = \sqrt{\frac{2\pi}{L}}\left[\mathbf{y}(\varphi_{LS,1}), \dots \mathbf{y}(\varphi_{LS,36})\right] \times \mathbf{W} \times \boldsymbol{\chi}(t). \quad (3)$$

Abbildung 2 stellt die Pegel der Lautsprechersignale $s_{LS}(t)$ für eine Wiedergabe eines Pseudorauschen von 70 dB SPL dar. Dort ist deutlich zu sehen, wie die physische Quellenbreite zunimmt, der Einfluss der seitlichen Lautsprecher aber verringert wird.

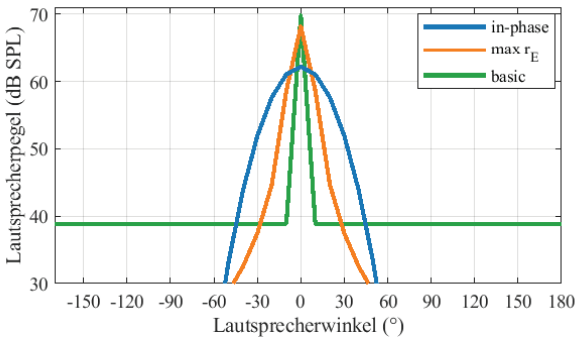

Abbildung 2: Pegel der Lautsprechersignale $s_{LS}(t)$ für eine Wiedergabe eines 70 dB SPL Pseudorauschen mit den verschiedenen Ambisonics-Dekoder.

Berechnung der Ohrsignale

Mit einer HRTF Datenbank mit einer Auflösung von 1°, gemessen mit einem Lautsprecher in 2 m Entfernung [16] wurden die Ohrsignale auf einer Linie von der Mitte des Lautsprecherarrays nach rechts (in positive y-Richtung), alle 2 cm ausgerechnet, bis zu einer Verschiebung von 2 m. Aus den binauralen Signalen wurden interaurale Pegelunterschiede (ILD), Zeitunterschiede (ITD), Kohärenz (IC) und Korrelation (IACC) ausgerechnet.

Die ITDs wurden aus den tiefpassgefilterten Signalen (1,5 kHz) als Verzögerung des Maximums der Korrelation ausgerechnet. Die ILDs wurden in Bark-Bändern als Differenz zwischen den Pegeln ausgerechnet. Eine breitbandige ILD wurde durch Mittelung der Pegel über die Bark-Bänder 3 bis 23 (200 Hz – 14 kHz) ermittelt. Positive ILDs und ITDs entstehen durch eine Schallquelle an der rechten Seite des Zuhörers ($\varphi_S > 0°$). Die IC und die IACC wurden ebenfalls in Bark-Bändern ausgerechnet und über die Bark-Bänder 3 bis 11 (200 Hz - 1,5 kHz) gemittelt.

Ergebnisse

Die Abbildungen 3-6 vergleichen die interauralen Merkmale die an verschiedenen Positionen entstehen, wenn eine Quelle an $\varphi_S = 0°$ mit Ambisonics 17. Ordnung synthetisiert wird. Als Vergleich werden die Merkmale einer physischen Quelle an ($x = 2{,}4\ m, y = 0\ m, \varphi_S = 0°$) dazugezeichnet, wie es bei einem *nearest loudspeaker mapping* der Fall wäre.

Interaurale Zeitunterschiede

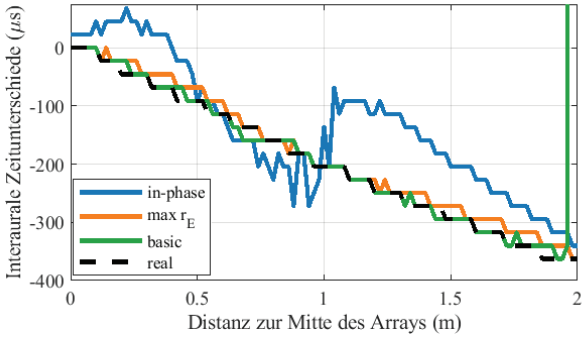

Abbildung 3: Interaurale Zeitunterschiede (ITD) für verschiedene Ambisonics-Dekoder und einer realen Quelle.

In Abbildung 3 ist zu sehen, dass die ITDs der echten Schallquelle von den *basic* und *max* r_E Dekoder sehr gut reproduziert werden. Der *in-phase* Dekoder scheint in der Nähe der Mitte robuster gegen kleine Positionsänderungen (die z.B. bei stehenden Probanden schnell entstehen) zu sein und die Richtung des Schalleinfalls ($\varphi_S = 0°$) im Sinne einer ebenen Welle in den ITDs ohne Parallaxverschiebung korrekt wiederzugeben. An den lateralen Positionen, für Distanzen zur Mitte über 50 cm ist der Verlauf der ITDs variabler, was für positionsabhängige Fehler bei der Lokalisierung spricht.

Interaurale Pegelunterschiede

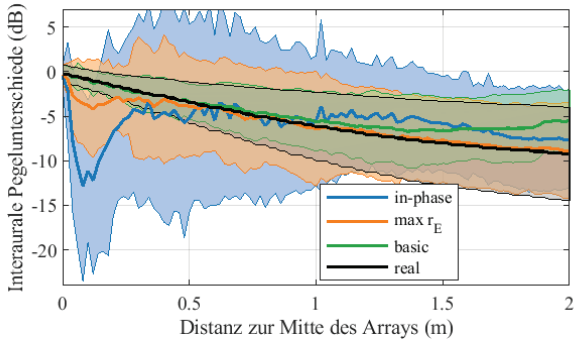

Abbildung 4: Interaurale Pegelunterschiede (ILD) für verschiedene Ambisonics-Dekoder und einer realen Quelle, gemittelt über Bark-Bänder. Der gefärbte Bereich zeigt für jede Kurve die Standardabweichung der ILD über die Bark-Bänder an.

Abbildung 4 zeigt die ILDs als Funktion der seitlichen Abweichung der Hörerposition relativ zur Mitte des Lautsprecherarrays auf. Insbesondere für die *max* r_E und *in-phase* Dekoder fällt die große Standardabweichung der ILDs über die Frequenzgruppen hinweg auf. So ist mit einer höheren wahrgenommenen Quellenbreite und dementsprechend größeren Fehlern bei der Lokalisierung von Schallquellen zu rechnen

Die ILDs nehmen bei einer realen Quelle mit der Distanz zur Mitte stetig ab, bis auf -9 dB. Die ILDs sind mit dem *basic* Dekoder bis zu 1 m kaum davon zu unterscheiden. Ab 1 m

Verschiebung bleiben sie konstant und nehmen dann wieder zu, was am Einfluss der seitlichen Lautsprecher liegt.

Dies ist bei dem *max* r_E Dekoder nicht zu sehen, da der Pegel der seitlichen Lautsprecher vernachlässigbar ist, vgl. Abb. 2. Der *in-phase* Dekoder weist ab 30 cm Verschiebung im Mittel eine ziemlich konstante ILD von ungefähr -5 dB auf. Das spricht für eine von der Verschiebung unabhängige Lokalisierung, aber aufgrund der großen Standardabweichung auch für Probleme mit der Quellenbreite und Ortbarkeit.

Interessant ist der Einbruch der ILDs für Verschiebungen bis zu 30 cm, sowohl für den *in-phase* als auch für den *max* r_E Dekoder.

Interaurale Kohärenz

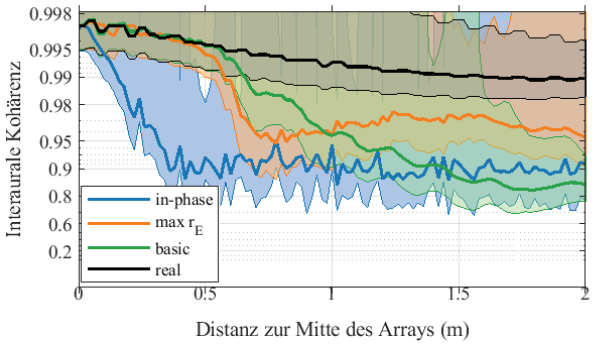

Abbildung 5: Interaurale Kohärenz (IC) für verschiedene Ambisonics-Dekoder und einer realen Quelle, gemittelt über Bark-Bänder. Der gefärbte Bereich zeigt die Standard-abweichung der IC über die Bark-Bänder.

Bei einer realen Quelle im Freifeld ist die Kohärenz (IC) wie zu erwarten sehr hoch, vgl. Abb. 5. Der *in-phase* Dekoder weist schon für geringe Verschiebungen eine deutlich niedrigere IC mit mehr Standardabweichung über die Frequenzgruppen auf. Ab 50 cm bleibt die IC fast konstant um 0,9. Die *basic* und *max* r_E Dekoder weisen jeweils bis zu 60 cm und 50 cm eine ähnliche IC wie eine reale Quelle auf, fallen danach aber schnell wieder ab. Die IC des *max* r_E Dekoders bleibt für Verschiebungen über 1 m um 0,95, die des *basic* Dekoders fällt ab auf ähnliche Werte wie des *in-phase* Dekoders.

Interaurale Korrelation

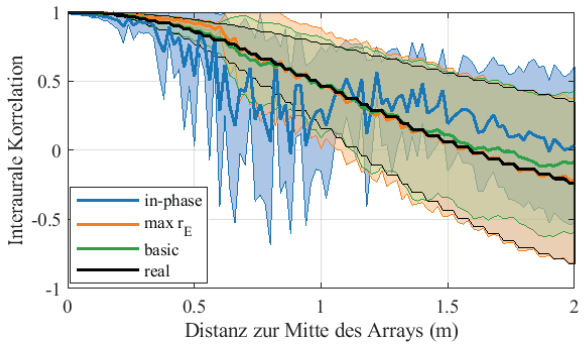

Abbildung 6: Interaurale Korrelation (IACC) für verschiedene Ambisonics-Dekoder und einer realen Quelle, gemittelt über Bark-Bänder. Der gefärbte Bereich zeigt die Standardabweichung der IACC über die Bark-Bänder.

Abbildung 6 zeigt die IACC der seitlichen Abweichung der Hörerposition relativ zur Mitte des Lautsprecherarrays auf. Bei diesem Merkmal sind alle drei Dekoder im Mittel und in ihrer Standardabweichung sehr ähnlich. Der Grund für die hohe Standardabweichung sind die Kammfiltereffekte, die frequenz- und positions-abhängig sind. So ist immer in einigen Frequenzgruppen die IACC negativ (bis zu -1 bei 180° Phasenverschiebung zwischen den Ohrsignalen) und daher die Standardabweichung dementsprechend hoch. Interessant ist, dass der *in-phase* Dekoder eine leicht geringere Standardabweichung aufweist. Dies liegt daran, dass die Ohrsignale weniger korreliert sind und die IACC in den mittleren und hohen Bark-Bändern sich tendenziell zwischen ±0,5 befindet.

Diskussion

Es stellt sich heraus, dass sowohl der *basic* als auch der *max* r_E Dekoder die interaurale Merkmale einer realen Quelle im Freifeld (vor allem ITDs und IACC) im Mittel gut synthetisieren können. Ab einer seitlichen Translation von ungefähr 50 cm wird der Effekt der seitlichen Lautsprecher mit dem *basic* Dekoder in einer niedrigeren IC und niedrigeren ILDs (Abbildungen 4 und 5) sichtbar. Ab 1 m Verschiebung kann der *max* r_E Dekoder diese Merkmale im Mittelbesser wiedergeben. In der Nähe der Mitte schwanken die ILDs mit dem *max* r_E Dekoder jedoch stark, was die Lokalisierung der Quelle im Sweetspot beeinträchtigt, wenn sich der Kopf des Probanden leicht bewegt. Das könnte auch eine weitere Quellenbreite erzeugen.

Die geringe Standardabweichung in den ILDs über den Bark-Bändern im *basic* Dekoder deutet auf eine niedrige Quellbreite und eine bessere Lokalisierbarkeit der synthetisierten Quelle hin. Der *max* r_E Dekoder weist für Verschiebungen über 1,3 m eine vergleichbare Standard-abweichung in den ILDs auf. Allerdings ist sie für kleinere Verschiebungen erheblich höher. Auch das sollte zu einer erhöhten Quellenbreite führen.

Der *in-phase* Dekoder weist ab 30 cm Verschiebung zur Mitte fast konstante über die Frequenzgruppen gemittelte ILDs auf, vgl. Abb. 3. Das deutet auf eine konsistente Lokalisierung außerhalb des Sweetspots hin, die nicht von der seitlichen Verschiebung beeinflusst wird. Allerdings ist die Standard-abweichung der ILDs über die Frequenzgruppen extrem hoch, was eine präzise Lokalisierung sehr erschwert. Die IC ist vergleichsweise niedrig und mit großer Standardabweichung, dass eine hohe Diffusität im Schallfeld anzeigt. Rakerd und Hartmann zeigten, dass bei einer niedrigeren IC die Lokalisierung weniger von den ITDs und stärker von den ILDs abhängt [17]. Das spricht für eine konstante Wahrnehmung außerhalb des Sweetspots, da die ILDs und IC für größere Verschiebungen entgegen der Mitte relativ konstant bleiben. Auch wenn die Lokalisierung und wahrgenommene Quellenbreite vergleichsweise schlechter synthetisiert werden als bei den anderen Decodern, kann diese Konsistenz erwünscht sein, e.g. bei künstlerisch motivierten Wiedergaben, wo Klangfarbe und Natürlichkeit der Wiedergabe wichtiger als eine präzise Lokalisierung sind.

Anhand der simulierten interauralen Merkmale lassen sich Zonen definieren, in denen die Wiedergabe akzeptabel ist. Für

den *basic* Dekoder ist der Radius ungefähr 60 cm. Dies reicht für Experimente mit einem Probanden, auch wenn dieser sich im wiedergegebenen Schallfeld bewegen darf. Falls ein größerer Sweetspot benötigt sein sollte (zum Beispiel für Demos, wo mehrere Personen verteilt im Lautsprecherarray stehen), kann der *max* r_E Dekoder verwendet werden.

Die Ergebnisse aus Frank *et al.* [5] werden in dieser Simulation auch beobachtet: die interauralen Merkmale, die mit dem *in-phase* Dekoder entstehen deuten auf eine schlechte Lokalisierung sowohl in der Mitte als auch an seitlichen Positionen im Lautsprecherarray.

Es wurde in diesem Beitrag nicht auf den Präzedenzeffekt eingegangen. Es ist angesichts der Lautsprecherpegel und dem laufzeitentzerrten Lautsprecherarray zu erwarten, dass der *basic* Dekoder anfälliger als die *max* r_E und *in-phase* Dekoder ist. Ab welcher Verschiebung von der Mitte des Lautsprecherarrays dies ein Problem für die Lokalisierung wird, ist noch offen.

Der Effekt der Nahfeldkompensation (NFC) für Ambisonics (*near-field corrected higher order Ambisonics* [18]) wurde nicht in diesem Beitrag aufgenommen. Erste Ergebnisse zeigen ähnliche Ergebnisse. Der *in-phase* Dekoder weist große Standardabweichungen und größere Fehler im Mittel der Merkmale auf. Die *basic* und *max* r_E Dekoder geben die interauralen Merkmale in der Nähe des Sweetspots gut wieder, für größere Verschiebungen tendieren die Merkmale zu denen des normalen Ambisonicsverfahren ohne NFC.

Zusammenfassung

In diesem Beitrag wurde eine virtuelle akustische Umgebung simuliert, in der mit HRTFs Ohrsignale eines Kunstkopfes ausgerechnet wurden, um die interauralen Merkmale für verschiedene seitliche Translationen von der Mitte des Lautsprecherarrays aus visualisieren zu können. Für das simulierte Lautsprecherarray mit einer Quellenposition voraus schneidet der *basic* Dekoder am besten für geringe Translationen ab. Für größere Translationen liefert der *max* r_E Dekoder die besten Ergebnisse. Der *in-phase* Dekoder scheint aufgrund der erwarteten Probleme in der Ortbarkeit, Quellenbreite und Diffusität des Schallfelds nicht für Hörforschung geeignet zu sein.

Danksagung

Die SOFE wurde durch das BMBF Bernstein Center for Computational Neuroscience, 01GQ1004B, finanziert.

Literatur

[1] Wefers, F., Pelzer, S., Bomhardt, R.., Müller-Trapet, M., Vorländer, M., Audiotechnik des aixCAVE Virtual Reality-Systems. Fortschritte der Akustik. Dt. Ges. f. Akustik (2015), 467-470.

[2] Favrot, S., Buchholz, J., A virtual auditory environment for investigating the auditory signal processing of realistic sounds. Journal of the Acoustical Society of America (2008), 123(5):3935.

[3] Seeber, B., Kerber, S., Hafter. E., A System to Simulate and Reproduce Audio-Visual Environments for Spatial Hearing Research. Hearing Research (2010), 260(1-2): 1-10.

[4] Monaghan, J., Seeber, B., A method to enhance the use of interaural time differences for cochlear implants in reverberant environments. Journal of the Acoustical Society of America (2016), 140(2):1116-1129

[5] Frank, M., Zotter, F., Sontacchi, A., Localization Experiments Using Different 2D Ambisonics Decoders. 25th Tonmeistertagung – VDT International Convention (2008)

[6] Stitt, P., Bertet, S., Van Walstijn, M., Off-centre Localisation Performance of Ambisonics and HOA For Large and Small Loudspeaker Array Radii. Acta Acustica united with Acustica (2014), 100(5):937-944

[7] Zotter, F., Frank, M., All-Round Ambisonics Panning and Decoding. Journal of the Audio Engineering Society (2012), 60(10):807-820

[8] Stitt, P., Bertet, S., Van Walstijn, M. Extended Energy Vector Prediction of Ambisonically Reproduced Image Direction at Off-Center Listening Positions. Journal of the Audio Engineering Society (2016), 64(5):299-310

[9] Ahrens, A. Characterizing auditory and audio-visual perception in virtual environments. PhD Thesis, DTU Health Technology (2019)

[10] Daniel, J., Rault, J.-B., Polack, J.-D., Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions.Proceedings of the AES 105th Convention (1998), 4795

[11] Malham, D., Experience with Large Area 3D Ambisonic Sound Systems. Proceedings of the Institute of Acoustics (1992) 209-215

[12] Gerzon, M., Optimum Reproduction Matrices for Multispeaker Stereo. Journal of the Audio Engineering Society (1992), 40(7-8):571-589

[13] Daniel, J., Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un context multimédia. PhD Thesis, Université de Paris 6 (2001)

[14] Seeber, B., Clapp, S., Interactive simulation and free-field auralization of acoustic space in the rtSOFE. Journal of the Acoustical Society of America (2017), 141(5):3974.

[15] Zotter, F., Frank, M., Ambisonics: A Practical 3D Audio Theory for Recording. Springer Topics in Signal Processing, 2019.

[16] Wierstorf, H., Geier, M., Raake, A., Spors, S., A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances. Engineering Brief presented at the AES 130th Convention (2011).

[17] Rakerd, B., Hartmann, W., Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise. Journal of the Acoustical Society of America (2010), 128(5):3052-3063.

[18] Daniel, J., Spatial sound encoding including near field effect: introducing distance coding filters and a viables, new Ambisonics format. Proceedings of theAES 23rd International Conference (2003) 1-15.