Evaluation eines Algorithmus zum Rendern von Distanz und Nähe bei binauraler Kopfhörerwiedergabe

Simone Füg[1], Jan Plogsties[1], Felix Fleischmann[1], Mikko-Ville Laitinen[2]
[1] *Fraunhofer Institut für Integrierte Schaltungen IIS, Erlangen, Deutschland, Email: simone.fueg@iis.fraunhofer.de*
[2] *Nokia Technologies, Espoo, Finnland*

Einleitung

Das menschliche Gehör ist in der Lage, Entfernungen von Schallquellen wahrzunehmen und genau abzuschätzen. Vor allem für Anwendungen im Bereich der virtuellen Akustik ist es daher für ein realistisches Zusammenspiel von visuellem und auditivem Eindruck wichtig, neben der Richtung einer Schallquelle auch eine der Quelle zugeordnete Distanz reproduzieren zu können.

Bekannte Wahrnehmungs-Merkmale für die auditive Distanzwahrnehmung sind unter anderem Schallintensität, das Verhältnis aus Direktschall- und Nachhallenergie (*Direct-to-Reverberant Energy Ratio*, D/R-Verhältnis), spektrale Eigenschaften (Dämpfung der hohen Frequenzen bei großen Entfernungen) sowie interaurale Differenzen, die besonders bei sehr kleinen Entfernungen die Distanzwahrnehmung unterstützen [1]. Im Gegensatz zur Erzeugung eines Richtungseindrucks ist die gezielte Steuerung der Distanzwahrnehmung unter Verwendung von räumlicher Audiowiedergabe oder Rendering-Systemen (beispielsweise binauraler Kopfhörerwiedergabe) bislang wenig erforscht. Für einige Ansätze sei auf die folgenden Veröffentlichungen verwiesen: [2, 3, 4].

Die vorliegende Arbeit vertieft einen Ansatz zur Erzeugung eines Entfernungseindrucks, der in einer früheren Arbeit für Lautsprecherwiedergabe erprobt wurde [5]. Dabei wird ein Eingangssignal nach Pegel- und Kohärenzanpassung sowie Anwendung eines Panning-Algorithmus aus mehreren Richtungen wiedergegeben. In der vorliegenden Arbeit wird nun untersucht, inwiefern sich das in [5] vorgestellte Verfahren auf binaurale Kopfhörerwiedergabe übertragen lässt.

Verfahren zur Steuerung des Distanzeindrucks

Das folgende Kapitel beschreibt in einem kurzen Überblick das dieser Arbeit zugrundeliegende Verfahren zum Distanzrendering. Für eine detailliertere Beschreibung sei auf [5] verwiesen.

Für die gezielte Erzeugung einer Distanz- oder Nähewahrnehmung eines Audio-Objektes werden zwei weitere zusätzliche Audio-Objekte erzeugt, die räumlich versetzt zum Originalobjekt gerendert werden. Dies ist in Abbildung 1 dargestellt. Die zusätzlichen Objekte sind in ihrer Position zum Originalobjekt um +30° und -30° Azimut verschoben und ein Pair-Wise Amplitude Panning [6] wird zum Rendern der Objekte verwendet. Die drei gerenderten Objektinstanzen (siehe Abbildung 1) werden entsprechend des gewünschten Distanzeindrucks in ihrer Phasenkohärenz und ihrer interauralen Kohärenz manipuliert. Zusätzlich wird das relative Pegelverhältnis der drei Objektinstanzen beeinflusst.

Der zugrunde gelegte Zusammenhang zwischen Kohärenz, Pegel und Entfernungseindruck basiert auf Kunstkopfmessungen (Cortex Manikin Mk1) in verschiedenen Räumen und mit verschiedenen Abständen zwischen Kunstkopf und Lautsprechern.

Es ergibt sich eine frequenz- und distanzabhängige interaurale Kohärenzfunktion zwischen den Ohrsignalen, sowie eine ebenfalls distanz- und frequenzabhängige Phasenkohärenz. Die Abschätzung der Phasenkohärenz wird mithilfe eines psychoakustischen Modells nach [7] durchgeführt. Sowohl die interaurale Kohärenz als auch die Phasenkohärenz nehmen mit steigender Distanz zwischen Quelle und Mikrofon ab, wie in Abbildung 2 gezeigt.

Für die gezielte Beeinflussung des Distanzeindrucks wird wie folgt vorgegangen: Soll das Originalobjekt als hinter

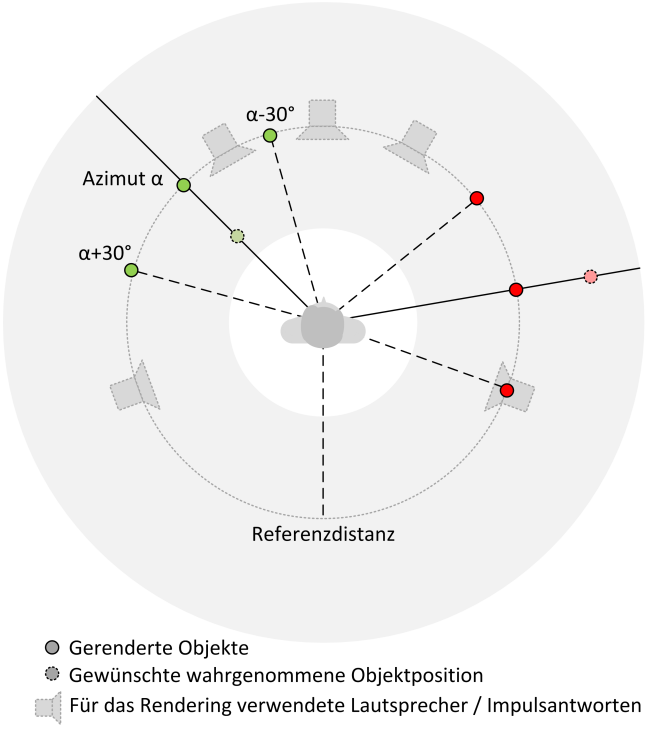

Abbildung 1: Rendering von Objekten mit Distanz- oder Näheeindruck über Verwendung zusätzlicher Objekte bei +30° und -30° Azimut in Bezug zum Originalobjekt.

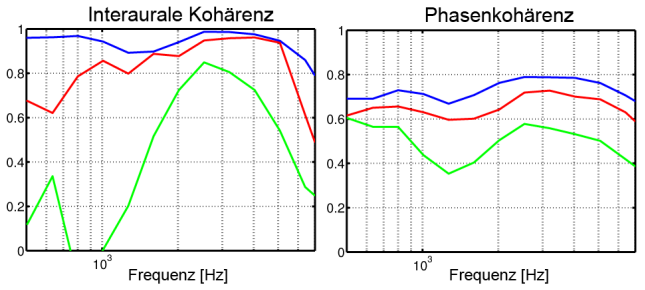

Abbildung 2: Interaurale Kohärenz sowie Phasenkohärenz (unter Verwendung des psychoakustischen Modell nach [7]) eines Hörraums jeweils als Funktion der Frequenz für drei unterschiedliche Entfernungen (blau: 0.575 m, rot: 1.15 m und grün: 2.3 m). Die interaurale Kohärenz und die Phasenkohärenz nehmen mit steigender Entfernung ab.

den Lautsprechern befindlich wahrgenommen werden (Abbildung 1, rot), so werden Dekorrelationsfilter [8, 9] verwendet um die Signale der zusätzlichen Objekte zu manipulieren, sodass die Phasenkohärenz abgesenkt wird. Die Dekorrelationsfilter werden über frequenzabhängige randomisierte Delays erzeugt und haben somit einen zufälligen Phasengang. Die dekorrelierten Signale werden über mehrere verteilte Lautsprecher in einem 5.1-Setup wiedergegeben, sodass sich auch die interaurale Kohärenz verringert:

$$\begin{aligned} y_c(n) &= g_{dry,f}(d_1) \cdot x(n), \\ y_l(n) &= g_{dec}(d_1) \cdot (x \star h_1)(n), \\ y_r(n) &= g_{dec}(d_1) \cdot (x \star h_2)(n). \end{aligned} \tag{1}$$

y_c, y_l, und y_r sind die Signale des originalen Audio-Objekts, sowie der zwei zusätzlichen Objekte. x ist das Eingangssignal, n ein diskreter Zeitindex, und d_1 ist die gewünschte Distanz zwischen 0 und 1, wobei 0 für die Lautsprecherdistanz und 1 für eine Verdopplung der Lautsprecherdistanz steht.

h_1 und h_2 sind unabhängige Dekorrelationsfilter, und $\{\star\}$ beschreibt eine Faltungsoperation. $g_{dry,f}$ und g_{dec} sind entfernungsabhängige Gewichtungsfaktoren, die den Pegel der drei Objekte anpassen. Die Gewichte basieren auf den aus den Messungen abgeleiteten Kurven und werden wie folgt bestimmt:

$$\begin{aligned} g_{dry,f}(d_1) &= 1 - 0.5 \cdot d_1, \\ g_{dec}(d_1) &= 0.25 \cdot d_1. \end{aligned} \tag{2}$$

Soll sich das Objekt vor den Lautsprechern befinden (Abbildung 1, grün), so werden folgende Signale erzeugt:

$$\begin{aligned} y_c(n) &= g_{dry,n}(d_2) \cdot x(n), \\ y_l(n) &= g_{coh}(d_2) \cdot x(n), \\ y_r(n) &= g_{coh}(d_2) \cdot x(n). \end{aligned} \tag{3}$$

Der Wert d_2 entspricht dabei der gewünschten Distanz zwischen 0 (Lautsprecherdistanz) und 1 (Halbierung der Lautsprecherdistanz). g_{coh} ist ein Gewichtungsfaktor für die zusätzlichen seitlich versetzen Objekte, $g_{dry,n}$ für das Originalobjekt. Die Gewichtungsfaktoren werden wie folgt bestimmt:

$$\begin{aligned} g_{dry,n}(d_2) &= 1, \\ g_{coh}(d_2) &= (2/(2 - d_2) - 1) \cdot 0.6. \end{aligned} \tag{4}$$

Evaluation des Verfahrens für binaurale Kopfhörerwiedergabe

In dieser Arbeit wird untersucht, inwiefern sich das in [5] vorgestellte Verfahren auf binaurale Kopfhörerwiedergabe übertragen lässt. Dazu wurde ein Hörtest mit binauraler Kopfhörerwiedergabe in Anlehnung an [5] durchgeführt. Dabei wurden möglichst identische Bedingungen geschaffen: Der selbe Raum (90 m^2 Hörraum mit einer mittleren Nachhallzeit von 0.4 Sekunden bei Frequenzen über 125 Hz), sowie die selben zwei Testsignale wurden verwendet. Die Benutzeroberfläche der Hörtestsoftware war ebenfalls identisch. Analog zum Lautsprechertest wurde vor dem Hörer ein akustisch transparenter Vorhang positioniert. Der Testaufbau ist in Abbildung 3 dargestellt. Die Lautsprechersymbole in der Abbildung entsprechen den Richtungen und Distanzen der für die Binauralisierung verwendeten binauralen Raumimpulsantworten (*Binaural Room Impulse Responses*, BRIRs).

Erzeugung der Stimuli

Für die Erzeugung der binauralen Kopfhörersignale wurden zwei kurze Testsignale (Ausschnitte aus Pop-Musik „*Music*“ und männlicher Sprache „*Speech*“) mit dem Panning-Algorithmus für die Azimut-Richtungen 0°, +15° und -15° (jeweils bezogen auf das Originalobjekt) auf ein virtuelles 5.1-Setup gerendert und für die Beeinflussung des Distanzeindrucks mit dem erläuterten Algorithmus bearbeitet. Dabei wurden ausgehend von einer Referenzdistanz von 2.3 m rechnerisch Signale für Distanzen von 1.15 m, 1.73 m, 3.45 m und 4.6 m erzeugt.

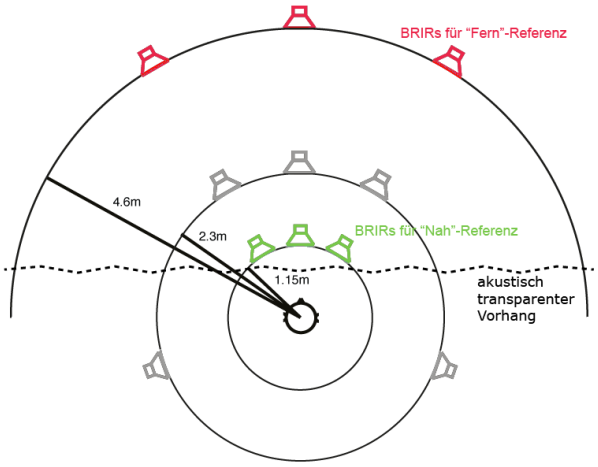

Abbildung 3: Verwendeter Testaufbau für den Hörtest. Die dargestellten Lautsprechersymbole entsprechen den Positionen, an denen die verwendeten BRIRs gemessen wurden.

Die Pegel der Signale wurden entsprechend der Formeln 5 und 6 energieerhaltend normalisiert, sodass die Amplitude der Signale nicht zur Distanzbeurteilung herangezogen werden konnte.

$$
\begin{aligned}
g_{norm,f} &= \sqrt{(g_{dry,f}(d_1))^2 + 2 \cdot (g_{dec}(d_1))^2}, \\
\hat{g}_{dry}(d_1) &= \frac{g_{dry,f}(d_1)}{g_{norm,f}}, \\
\hat{g}_{dec}(d_1) &= \frac{g_{dec}(d_1)}{g_{norm,f}}.
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
g_{norm,n} &= \sqrt{(g_{dry,n}(d_2))^2 + 2 \cdot (g_{coh}(d_2))^2}, \\
\hat{g}_{dry}(d_2) &= \frac{g_{dry,n}(d_2)}{g_{norm,n}}, \\
\hat{g}_{coh}(d_2) &= \frac{g_{coh}(d_2)}{g_{norm,n}}.
\end{aligned}
\tag{6}
$$

Anschließend wurden die erzeugten Signale mit einem Set aus gemessenen BRIRs (Kunstkopfaufnahmen mit einem Cortex Manikin Mk1 aus dem verwendeten Hörraum) aus den Richtungen 0°, ±30° und ±110° (entsprechend der Lautsprecheranordnung im virtuellen 5.1 Setup und der Lautsprecheranordnung in [5]) und der Referenzdistanz von 2.3 m gefaltet, um binaurale Kopfhörersignale zu erzeugen. Zusätzlich wurden Referenzsignale berechnet, indem eine Faltung der gerenderten 5.1-Testsignale ohne Anwendung des Distanzrenderings mit binauralen Raumimpulsantworten aus 1.15 m (Nah-Referenz) und 4.6 m (Fern-Referenz) durchgeführt wurde.

Für die Messung der BRIRs wurden die gleichen Lautsprecher verwendet, die auch in [5] für den Hörtest zur Evaluation des Verfahrens für Kopfhörerwiedergabe genutzt wurden. Die Richtungen und Distanzen der BRIRs entsprechen den in Abbildung 3 dargestellten Lautsprechersymbolen.

Hörtestdurchführung

Zehn Hörer nahmen am Hörtest teil. Die Bewertung der zu testenden Signale erfolgte jeweils durch eine relative Distanzbewertung auf einer Skala zwischen 0 und 100 in Bezug auf Nah-Referenz (Skalenwert 20) und Fern-Referenz (Skalenwert 80). Die Referenzen wurden offen dargeboten und ihre Lage auf der Skala war bekannt.

Die zu bewertenden Konditionen wurden in einem Multi-Stimulus-Test präsentiert, sodass ein paralleler Vergleich untereinander und mit den Referenzen möglich war. Dabei wurden neben den durch das Distanzrendering erzeugten Signalen auch die Signale aus der Referenzdistanz von 2.3 m dargeboten, die durch reine Faltung erzeugt wurden. Jedem Hörer wurden Signale aus der 0°-Richtung präsentiert, zwischen +15° und -15° wurde randomisiert zwischen den Testpersonen variiert, wobei sichergestellt war, dass jede dieser beiden Richtungen von gleich vielen Testpersonen bewertet wurde. Die Richtung der Referenzen entsprach jeweils der Richtung der zu bewertenden Signale.

Ergebnisse und Diskussion

Die Ergebnisse des Hörtests sind in Abbildung 4 als Boxplots mit Median und Interquartilabständen dargestellt. Den Signalen ist für die Abbildung ein Distanzfaktor zwischen 0.5 und 2 in Bezug auf die Referenzentfernung von 2.3 m zugeordnet. Für die Auswertung wurden die Richtungen von +15° und -15° kombiniert.

Die Hörtestergebnisse zeigen, dass die Reihenfolge der mittleren wahrgenommenen Entfernungsunterschiede zu der Reihenfolge der für die Berechnung der Signale angenommenen Distanzen korrespondiert: Signale, die durch das Rendering weiter entfernt klingen sollen, werden entfernter wahrgenommen, als Signale, denen eine kleinere Distanz aufgeprägt wurde. Es ist also möglich, eine relative Entfernungsstaffelung von Objekten mit dem Verfahren auch für binaurale Kopfhörerwiedergabe zu erzeugen.

Es zeigt sich zudem eine Komprimierung des Wahrnehmungsbereichs durch die zwei vorgegebenen Referenzen: Die Hörer bewerteten die mit dem Distanzrendering-Verfahren erzeugten Signale tendenziell nicht so nah und nicht so fern wie die Referenzsignale.

Signale, die einem Distanzfaktor von 1 zugeordnet sind (also Signale, bei denen das Distanzrendering nicht aktiv war und die durch eine reine Faltung mit den Impulsantworten aus 2.3 m Entfernung erzeugt wurden), werden in ihrer Distanzwahrnehmung etwa mittig zwischen der Nah- und der Fern-Referenz eingeordnet. Die Bewertung dieser Signale zeigt trotz der Erzeugung durch eine reine Faltung eine recht große Varianz, die mit der Varianz der mithilfe des Distanzrenderings erzeugten Signale vergleichbar ist. Die Signalverarbeitung des Distanzrenderings verschlechtert also nicht die Genauigkeit oder Sicherheit der Distanzbewertung.

Die in diesem Hörtest gemachten Beobachtungen für binaurale Kopfhörerwiedergabe finden sich in ähnlichem Maße auch in der Auswertung des Lautsprecher-Hörtests in [5] wieder.

Zusammenfassung

Das Ziel dieser Arbeit war es, zu überprüfen, ob ein Distanzrendering-Verfahren, das für Lautsprecher-Wiedergabe entwickelt wurde [5], sich auch für binaurale Kopfhörerwiedergabe eignet. Es sollte getestet werden, ob sich mit den Verfahren auch bei binauraler Kopfhörerwiedergabe der Distanz- und Näheeindruck des Hörers gezielt steuern lässt.

Für die gezielte Beeinflussung der Distanzwahrnehmung wurde dazu ein Eingangssignal mithilfe eines Panning-Algorithmus auf ein virtuelles 5.1-Setup unter Verwendungen von binauralen Raumimpulsantworten einer Referenzdistanz gerendert und zudem Pegel und Kohärenz der einzelnen Ausgangssignale verändert. Für eine gewünschte wahrgenommene Distanz größer einer Referenzdistanz wurden zusätzlich Dekorrelationsfilter eingesetzt, um die interaurale Kohärenz und die Phasenkohärenz zu beeinflussen. Für eine gewünschte wahr-

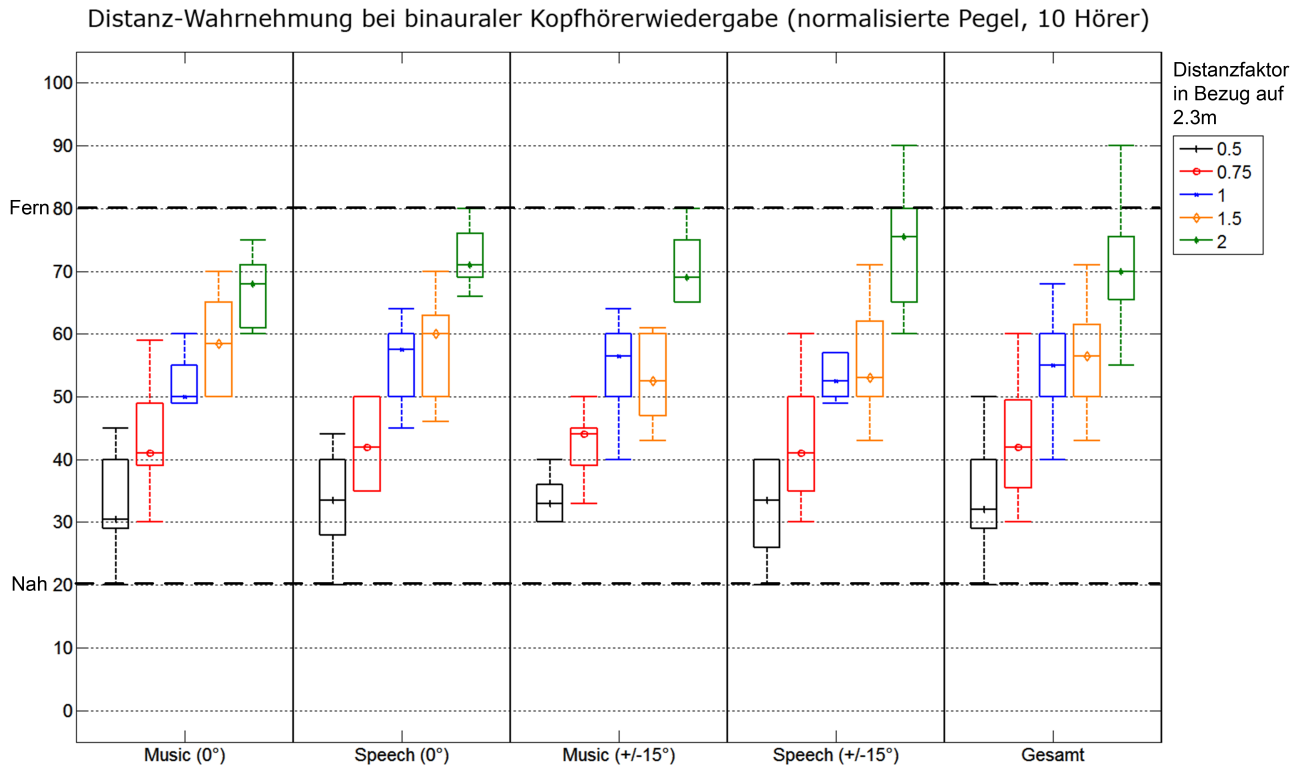

Abbildung 4: Hörtestergebnisse unter Verwendung von zwei Testsignalen (*Music*, *Speech*). Die Testsignale wurden aus verschiedenen Richtungen wiedergegeben. Jede Spalte zeigt die Ergebnisse für ein Testsignal für eine Richtung und verschiedene Distanzen, sowie ganz rechts zusammengefasst über alle Signale und Richtungen. Die Distanzen werden durch einen Faktor relativ zu einer Referenzdistanz von 2.3 m dargestellt.

genommene Distanz kleiner der Referenzdistanz wurde eine kohärente Wiedergabe mit Pegelanpassung genutzt.

Die Ergebnisse eines durchgeführten Hörtests zeigen, dass sich das Verfahren aus [5] auch für binaurale Kopfhörerwiedergabe eignet. Die Distanzwahrnehmung von Audio-Objekten kann gezielt beeinflusst werden. Die mit dem Distanzrendering-Verfahren gerenderten Objekte werden näher oder ferner wahrgenommen, als Signale, die durch eine reine Faltung mit Raumimpulsantworten einer bestimmten Messdistanz erzeugt wurden.

Der Bereich der wahrgenommenen Distanz ist gegenüber der idealen Distanz gestaucht, die erzeugten Signale werden also etwas weniger nah beziehungsweise weniger fern empfunden als ideale Referenzsignale. Die Reihenfolge der mittleren wahrgenommenen Entfernungsunterschiede korrespondiert zu der Reihenfolge der für die Berechnung der Signale angenommenen Distanzen, eine relative Entfernungsstaffelung von Objekten kann also mit dem Verfahren erzeugt werden.

Literatur

[1] Zahorik, P., Brungart, D.S., Bronkhorst, A.W.: *Auditory distance perception in humans: A summary of past and present research.* Acta Acustica with Acustica Vol. 91 (2005), 409-420.

[2] Gerzon, M.A.: *The design of distance panpots.* 92nd AES Convention, Wien, März 1992.

[3] Kim S., Lee, Y.W., Kim Y.: *3D audio depth rendering for enhancing an immersion of 3DTV.* 131st AES Convention, New York, Oktober 2011.

[4] Füg, S., Werner, S., Brandenburg, K.-H.: *Controlled auditory distance perception using binaural headphone reproduction - Modification of measured BRIRs.* 27. Tonmeistertagung, Köln, November 2012.

[5] Laitinen, M.-V. et al: *Auditory distance rendering using a standard 5.1 loudspeaker layout.* 139th AES Convention, New York, Oktober 2015.

[6] Blumlein, A.D.: *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems - British patent specification 394325, 1931.* J. Audio Eng. Soc. Vol. 6, No. 2 (1958).

[7] Laitinen, M.-V., Disch, S., Pulkki, V.: *Sensitivity of human hearing to changes in phase spectrum.* J. Audio Eng. Soc. Vol. 61 (2013), 860-877.

[8] Kendall, G.S.: *The decorrelation of audio signals and its impact on spatial imagery.* Computer Music Journal Vol. 19, No. 4 (1995), 71-87.

[9] Laitinen, M.-V. et al: *Reproducing applause-type signals with directional audio coding.* J. Audio Eng. Soc. Vol. 59 (2011), 29-42.