

Berücksichtigung der Kopfposition bei der Erzeugung eines virtuellen Kopfhörers mittels Wellenfeldsynthese

Roman Franke^{1,4}, Günther Theile¹, Klaus Laumann^{1,3}, Dieter Leckschat²

¹ Institut für Rundfunktechnik GmbH, 80939 München

² Fachhochschule Düsseldorf, 40474 Düsseldorf

³ AG Technische Akustik, MMK, TU München, 80333 München

⁴ jetzt: Ruetz Technologies GmbH, 80807 München

Email: roman.franke@ruetz.de

Einleitung

Das auf der DAGA 2006 vorgestellte Audiowiedergabesystem "Binaural Sky" erzeugt unter Einsatz von Wellenfeldsynthese (WFS), Übersprechkompensation (engl.: Crosstalk Cancellation, XTC) und binauraler Raumsynthese (BRS) virtuelle Schallquellen. Das System besteht aus 22 Lautsprechern, die in einem Radius von 50 cm ringförmig über dem Kopf des Zuhörers angeordnet sind. Mit Hilfe eines Headtrackers werden rotatorische Kopfbewegungen erfasst und bei der Signalverarbeitung berücksichtigt, so dass die virtuellen Schallquellen bei Drehungen des Kopfes ortstabil wahrgenommen werden [1]. Kern des Systems ist das Verfahren zur Synthese eines "virtuellen Kopfhörers" für die Wiedergabe binauraler Signale über Lautsprecher.

Dieses Verfahren beruht auf bekannten Algorithmen zur Übersprechkompensation mit der Besonderheit, dass als Berechnungsgrundlage keine realen Lautsprecher, sondern per WFS erzeugte, fokussierte Schallquellen dienen. Dieses Konzept erlaubt es, einen statischen XTC-Filtersatz zu verwenden, wobei Kopfdrehungen des Zuhörers durch dynamische Positionsanpassung der fokussierten Schallquellen kompensiert werden.

Eine wesentliche Einschränkung besteht in der Tatsache, dass das System nur stabil ist, wenn sich der Hörer in der Mitte unter dem Array aufhält; der maximal mögliche Bewegungsradius wird mit etwa 7 cm angegeben.

Im Folgenden werden verschiedene Lösungsansätze zur Erweiterung dieses "sweet spot" diskutiert, um dem Zuhörer eine angemessene Bewegungsfreiheit einzuräumen.

Signalverarbeitung im Ausgangssystem

Die Synthese des virtuellen Kopfhörers wird durch die Open-Source-Faltungssoftware BruteFIR [2] umgesetzt.

Die Echtzeitfähigkeit des Systems wird garantiert, indem alle notwendigen FIR-Filter vorberechnet werden, wobei der Prozessor mit 2×22 Faltungsoperationen belastet wird.

Zunächst werden die sog. WFS-Driving-Functions zur Erzeugung von n Fokusquellen erstellt und deren HRTFs berechnet. Basierend auf diesen Ergebnissen können pro Fokusquelle zwei XTC-Filter erstellt werden. Fasst man die vorliegenden XTC-Filter und Driving-Functions in geeigneter Weise zusammen, erhält man für jeden Kopf-Azimuth in einer Auflösung von 1° , pro Eingangskanal und pro Lautsprecher (m , hier: $m=22$) einen so genannten Array-Filter (Abb. 1). Im Arbeitsspeicher des Systems müssen also

$360 \times 2 \times 22$ Filter zur Verfügung stehen. Entsprechend den vom Headtracker ermittelten Azimut-Daten wird ein Filtersatz von 2×22 FIR-Filtern für die Echtzeitsignalverarbeitung gewählt, vgl. [1].

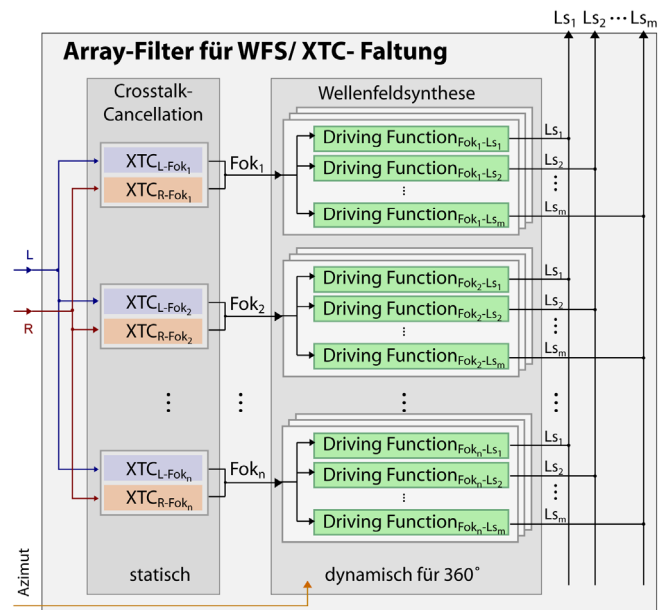


Abbildung 1: Theoretischer Signalverlauf der Array-Filter zur Erzeugung des virtuellen Kopfhörers für n Fokusquellen und m Lautsprecher.

Lösungsansätze

Um bei der Synthese des "virtuellen Kopfhörers" auch translatorische Bewegungen zu berücksichtigen, sind verschiedene Ansätze denkbar, die auf eine Positionsanpassung der fokussierten Schallquellen zielen.

Modellbasierter Ansatz

Eine Möglichkeit zur Anpassung des Systems besteht in der Auftrennung der zusammengefassten Array-Filter in zwei verschiedene Filterinstanzen. Der Signalverlauf entspräche dem in Abb. 1 dargestellten, wobei anstelle der vorberechneten Driving-Functions in Abhängigkeit der Positionsdaten Verzögerung (engl. delay) und Pegelanpassung nach Prinzipien der WFS in Echtzeit berechnet würden. Der Vorteil dieser Methode liegt in der verringerten Anzahl durchzuführender Faltungsoperationen ($2 \times n$), die restlichen Signalmanipulationen beschränken sich auf Verzögerung und Pegelanpassung. Der Nachteil der Methode liegt im Auftreten der hohen Anzahl von Signalen ($22 \times n$) im Zwischenschritt der Berechnung, weshalb eine

experimentelle Umsetzung am aktuellen System vorerst nicht weiter verfolgt wurde.

Datenbasierter Ansatz

Denkbar ist auch eine Beibehaltung des aktuellen Verfahrens, bei dem die Filter für alle möglichen Positionen vorberechnet werden. Eine Hörzone mit 30 cm Radius und einem Raster von 1 cm ergibt 2827 Positionen, an denen auch der Kopf-Azimuth berücksichtigt werden müsste. D.h. es sind $2827 \times 360 \times 2 \times 22 = 44\,779\,680$ Filter zu erstellen. Bei der zurzeit verwendeten Impulsantwortlänge von 256 Samples und einer Auflösung von 32 Bit entstünde ein Datensatz von ca. 40 GB. Dieser müsste vollständig im Arbeitsspeicher zur Verfügung stehen oder dynamisch verwaltet werden, was als eindeutiger Nachteil dieses Ansatzes angesehen werden kann.

Kombination von modell- und datenbasiertem Ansatz

Es wurde schließlich ein Verfahren umgesetzt, das für rotatorische Bewegungen weiterhin auf den datenbasierten Ansatz aufbaut und die Anpassung an translatorische Bewegungen – als Annäherung an den modellbasierten Ansatz – über die Implementierung steuerbarer Verzögerungen für jeden Lautsprecherkanal regelt.

Gleichung (1) zeigt die Berechnung dieser Delay-Zeiten: Zunächst wird für jeden Lautsprecher der Abstand l_m zur Kopfposition in allen drei Raumkoordinaten ermittelt. Der am weitesten entfernte Lautsprecher mit dem Abstand l_{max} wird als Referenz benutzt und ohne Verzögerung angesteuert. Für jeden weiteren Lautsprecher wird die Entfernungsdifferenz mit $l_{max} - l_m$ berechnet. Die Division dieses Wertes durch die Schallgeschwindigkeit c ergibt für den Lautsprecher m die Verzögerung $delay_m$, die in eine entsprechende Anzahl von Abtastwerten umgerechnet wird.

$$delay_m = \frac{l_{max} - l_m}{c} \times f_{Sample} \quad [s] \quad (1)$$

Bei schnellen Bewegungen des Zuhörers kann es durch große Unterschiede in den Verzögerungszeiten zu Sprüngen im Signal kommen. Um hörbare Artefakte zu verhindern, wurde zusätzlich eine "Rampenfunktion" implementiert, die den neuen Wert bei Überschreitung einer maximal zulässigen Unterschiedsschwelle schrittweise annähert.

Evaluierung

Zur objektiven Evaluierung der Anpassungen wurden Kunstkopfmessungen an verschiedenen Positionen innerhalb der Hörzone durchgeführt. Dabei wurde über den linken Eingangskanal des Systems ein Sinus-Sweep wiedergegeben und das resultierende Schallsignal an beiden Ausgängen des Kunstkopfes ausgewertet. Dabei sollte das Signal auf dem linken Kunstkopf-Kanal einen möglichst linearen Frequenzgang aufweisen und auf dem rechten Kanal möglichst gut unterdrückt werden. Anschließend wurde diese Messung für den rechten Eingangskanal mit entsprechend umgekehrten Erwartungen wiederholt.

Beispielhaft ist in Abb. 2 das Messergebnis für eine Vorwärtsbewegung des Zuhörers um 20 cm mit Anpassung

der Verzögerung einem Messergebnis ohne Anpassung der Verzögerung in Abb. 3 gegenüber gestellt.

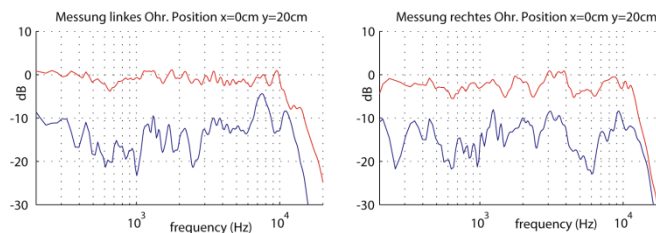


Abbildung 2: Übersprechkompensation des virtuellen Kopfhörers **mit** Delay-Anpassung nach einer Vorwärtsbewegung um 20 cm; rot: Nutzsignal, blau: Übersprechanteil

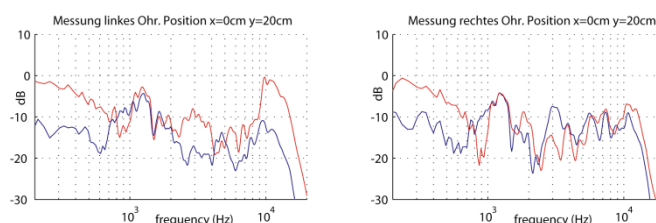


Abbildung 3: Übersprechkompensation des virtuellen Kopfhörers **ohne** Delay-Anpassung nach einer Vorwärtsbewegung um 20 cm; rot: Nutzsignal, blau: Übersprechanteil

An diesen Messungen ist deutlich zu erkennen, wie die Anpassung der Delay-Werte eine Kompensation des Übersprechens auch bei Positionen außerhalb der Mitte des Lautsprecher-Arrays verbessert. Abb. 3 zeigt, dass ohne Anpassung eine Kanaltrennung nicht mehr vorhanden ist, wohingegen in Abb. 2 bis zu einer oberen Grenzfrequenz von ca. 7 kHz eine Dämpfung des Übersprechanteils von mehr als 10 dB erreicht werden kann.

Ähnliche Ergebnisse zeigen sich bei seitlichen Bewegungen, allerdings mit einer leicht verringerten Übersprechkompensation, da die Lage der Array-Lautsprecher in Bezug auf die Ohren des Zuhörers nicht mehr symmetrisch ist.

Zusammenfassung

Es konnte gezeigt werden, dass sich durch die Korrektur von Verzögerungszeiten eine gute Anpassung des virtuellen Kopfhörers an translatorische Bewegungen des Zuhörers erreichen lässt. Damit konnte die Hörzone auf einen Radius von 30 cm erweitert werden. Diese Lösung stellt, wie zuvor bereits erwähnt, einen ressourcenschonenden Kompromiss dar. Interessant wäre ein Vergleich mit den ebenfalls vorgestellten reinen modell- bzw. datenbasierten Ansätzen.

Literatur

[1] Menzel, D.; Wittek, H.; Theile, G.; Fastl, H.: Binaurale Raumsynthese mittels Wellenfeldsynthese. Tagungsband DAGA 2006, Braunschweig, 20.-23.03., DEGA 2006

[2] BruteFIR, URL: <http://www.ludd.luth.se/~torger/brutefir.html>