An augmented acoustics demonstrator with realtime stereo up-mixing and binaural auralization

Exposé zur Masterarbeit

Raffael Tönges, #305907
rtoenges@mailbox.tu-berlin.de

20. Mai 2015

Betreuer: Prof. Dr. Stefan Weinzierl
Dr. Alexander Lindau

Technische Universität Berlin
Fachgebiet Audiokommunikation

1 Einleitung und Fragestellung

Musik wird häufig für die Wiedergabe über Lautsprecher in Stereoanordnung produziert. Die empfundene Räumlichkeit während der Wiedergabe ergibt sich aus dem Zusammenspiel mehrerer Komponenten der Übertragungskette.

Zunächst enthält eine Aufnahme den räumlichen Klang der Aufnahmeumgebung. Dieses aufgenommene Musikstück wird bei der Mischung / dem Mastering je nach Wunsch zusätzlich durch Effekte in seinem Raumklang verändert. Bei der Wiedergabe der Musik über Lautsprecher beeinflussen die räumlichen Eigenschaften der Wiedergabeumgebung, die verwendeten Lautsprecher und deren Aufstellung ebenfalls die Wahrnehmung der Räumlichkeit des wiedergegebenen Materials. Wird auf gewisse Konventionen für die Positionierung von Lautsprechern für die Stereowiedergabe [1] geachtet, entsteht eine virtuelle Bühne, die sich zwischen den beiden Lautsprechern aufspannt.

Die Morphologie des Rezipienten hat ebenfalls einen maßgeblichen Einfluss darauf, wie das Gehörte wahrgenommen wird. So wird der Frequenzgang des Signals durch Reflexionen, Abschattung/Dämpfung und durch Resonanzbildung des Körpers, Kopfes und der Außenohren moduliert.

Der Schall, der an einem Ohr ankommt, kommt auch an dem anderen Ohr an. Je nachdem wie die Quelle relativ zum Rezipienten positioniert ist, können die Signale an den beiden Ohren unterschiedlich laut, spektral gefärbt und verzögert sein.

Die Kopfhörerwiedergabe von Schallereignissen, die nicht binaural aufgenommen wurden oder nicht korrekt bezogen auf die realen Kopfbewegungen dynamisiert sind (z.B. klassische Stereosignale), führt oft zur Im-Kopf-Wahrnehmung der akustischen Szene.

Wird ein Musikstück mit *Head Related Transfer Functions* (HRTFs) gefiltert, kann der Eindruck einer Externalisierung erweckt werden, da dem Musikstück dabei die beim beidohrigen Hören von natürlichen Schallquellen auftretenden physikalischen Effekte per Filterung aufgeprägt werden. HRTFs enthalten die Filterfrequenzgänge, die durch die Oberkörper- und Kopfmorphologie im Freifeld aus im Idealfall jedem Einfallswinkel entstehen und auf ein akustisches Signal wirken.

HRTFs können entweder individuell für einem Rezipienten in einer speziellen Messumgebung ermittelt werden oder es können “fremde” HRTFs [2] verwendet werden. In dem Fall, dass nicht die eigenen HRTFs für die Filterung verwendet werden, müssen meist auditive Abbildungsfehler in Kauf genommen werden. Das können z.B. Lokalisations- oder Klangfarbenfehler sein.

Wie gut die Quelle als externalisiert (d.h. ausserhalb des Kopfes) wahrgenommen wird, hängt von einem Zusammenspiel mehrerer Faktoren (Richtungseigenschaften, Räumlichkeit, Dyna-

[1] Die Stereolautsprecher sollten in gleicher Entfernung und gleichem Winkel zum Hörer aufgestellt werden. Der gesamte Winkel zwischen den beiden Lautsprechern sollte nicht die Größe von 60° übersteigen.

[2] Generische HRTFs werden mit statischen Kunstköpfen oder neuerdings mit binauralen Messrobotern aufgenommen. Es handelt sich bei den Messrobotern um Nachbildungen des menschlichen Kopfes oder Oberkörpers. In den beiden Gehörgängen ist jeweils ein Mikrofon integriert. Die derart ermittelten HRTFs sind mit unterschiedlichem Erfolg für andere Personen benutzbar.

mik, etc.) ab und kann nicht generell vorhergesagt werden. Wie stark der Einfluss der einzelnen Faktoren ist, ist Gegenstand aktueller Forschung. Filtert man jedoch ausschließlich mit HRTFs und nimmt keine weiteren Anpassungen an einem Audiosignal vor, ist die wahrgenommene Externalisierung erfahrungsgemäß eher gering. Um die Externalisierung zu verbessern können statt HRTFs *Binaural Room Impulse Responses* (BRIRs) verwendet werden, die Schallinformationen eines bestimmten Raumes enthalten.

Durch geeignete räumliche Analysealgorithmen kann ein Stereosignal in direkte und räumliche Komponenten geteilt werden. Diese Komponenten können - unter zusätzlicher Verwendung künstlicher Raumschallinformationen (z.B. durch Filterung mit BRIRs) - zur Wiedergabe auf einer realen oder virtuellen (binauralen) Mehrkanalanlage aufbereitet werden; man spricht dabei von (räumlichem) *up-mixing*. Es wird also ein virtueller Raum um den Hörer herum geschaffen, der die akustischen Eigenschaften eines realen Raums simuliert. Beim *up-mixing* wird der räumliche Höreindruck eines Musikstücks gesteigert, indem aus dem virtuell generierte Mehrkanalstück mit künstlichen BRIRs binaurale Signale generiert werden (vgl. [1]).

1.1 Zielsetzung

Viele Menschen besitzen ein Smartphone und der Kauftrend ist steigend [3]. Die meisten Smartphones besitzen die Fähigkeit Musik wiederzugeben. In der Regel wird die Musik vom Smartphone mit einem Kopfhörer rezipiert.

Im Rahmen dieser Arbeit soll eine mobile application (kurz APP) für iOS Geräte (ab iOS 8.0) implementiert werden, welche Stereo-Audiodateien für die binaurale Wiedergabe in Echtzeit *up-mixed*. Ebenfalls wird ein Kopfhörer angepasst, um als Demonstrator in Verbindung mit der erwähnten APP zu fungieren. Dieser Kopfhörer soll zudem mit Mikrofonen an beiden Seiten des Kopfes ausgestattet werden, damit Umgebungsgeräusche wahlweise “durchgelassen” werden können. Somit soll eine variable Integration der wiedergegebenen Musik in die reale akustische Umgebung ermöglicht werden.

2 Stand der Forschung

Für die Bearbeitung der Fragestellungen sind hauptsächlich drei Forschungsgebiete zu betrachten. Zum einen ist die binaurale Aufbereitung von Audiomaterial ein wichtiger Aspekt, da die verwendeten Musikstücke in einem gewöhnlichen Stereoformat vorliegen. Ferner wird das *upmixing* betrachtet, da das Signal um weitere Kanäle virtuell ergänzt werden soll, um einen Schalleinfall aus allen Richtungen und somit die Integration der Musik in einen künstlichen Raum zu simulieren. Zusätzlich ist eine Auseinandersetzung mit der Technik für das “Durchlassen” von Audiosignalen mit Hilfe von an den Kopfhörer angebrachten Mikrofonen (*hear-through*) für die Bearbeitung der Fragestellung notwendig.

[3] Nach [2] wurden 2013 das erste Mal in Stückzahlen mehr Smartphones als Featurephones verkauft.

2.1 Binaurale Filterung

Die Annäherung zwischen der Wiedergabe aufgenommener Audiosignale oder speziell Musik über Kopfhörer an eine reale Hörsituation ist ein bekanntes Forschungsgebiet. In [3] werdendie Grundlagen der Binauraltechnologie beschrieben. Musik wird mit Hilfe von HRTFs gefiltert und über Kopfhörer wiedergegeben. Die derart gefilterte Musik kann klingen, als sei sie von einer Quelle im Freifeld wiedergegeben. Die Position der Quelle wird hierbei durch die Wahl des HRTF Filters bestimmt.

Die HRTFs unterscheiden sich für jedes Individuum. [4] und [5] haben ein System vorgestellt, wie die Messung der individuellen HRTFs sehr zeitsparend möglich ist. Das System steht an der Technischen Universität Berlin zur Verfügung.

Als Erweiterung zur statischen Binauralsynthese besteht die Möglichkeit der dynamischen Binauralsynthese. Hierbei wird die Kopfstellung (Rotation und ggf. Neigung) des Hörers mittels *head tracker* erkannt. Eine Quelle kann dann an einem festen Ort relativ zum Hörer positioniert werden, indem für die Filterung der Quelle die jeweils korrekten HRTFs gemäß ihres Winkels verwendet werden. Im Rahmen dieser Arbeit wird auf dynamische Binauralsynthese verzichtet, auch wenn dadurch Verschlechterungen der Externalisierung in Kauf genommen werden müssen. Der verwendete Algorithmus qualifiziert sich für die Anwendung auf mobilen Geräten, durch die Trennung der Berechnungen in zeitkonstante und zeitvariante Terme. Die zeitkonstanten Terme werden zu Beginn einmal berechnet. Bezüglich des Ortes ist keines dieser Berechnungen konstant. Der Rechenaufwand für den dynamischen Fall ist mit den heutigen Smartphones in Echtzeit nicht zu bewältigen.

2.2 Up-mixing

Up-mixing von Stereosignalen ist ein Verfahren, dass seinen Ursprung in der Filmindustrie hat. Stereospuren aus älteren Filmen sollten mit der Einführung von Surround-Systemen für die neue Technik aufbereitet werden. Das Stereosignal wird zunächst, wie in [6] und [7] beschrieben, analysiert und in eine Primärquelle und Ambientkomponenten zerlegt. Die Komponenten können, erweitert durch künstliche Reflexionen virtuell simulierter Wände, auf geeignete Weise auf eine beliebige Surround-Lautsprecherkonfiguration (5.1, 7.1, 10.2, etc.) verteilt werden. [1] liefert einen Ansatz wie *stereo-up-mixing* mit Konzertraumakustik und statisches binaurales Rendering mit BRIRs kombiniert und für mobile Geräte optimiert werden könnte.

2.3 Hear-through

In [8] werden zwei *Spatialized Augmented-Reality Audio* (SARA) Systeme vorgestellt. Die verwendeten Kopfhörer sind akustisch besonders durchlässig, so dass die Rezeption der akustischen Umgebung verglichen mit dem Hören ohne Kopfhörer wenig verändert wird. Die vorgestellten Systeme verändern beide marginal den Klang der akustischen Szene, die sie umgibt. Beide Systeme ermöglichen eine gute Lokalisation innerhalb des *Virtual Auditory Space* (VAS). Eine Abschottung von der akustischen Umwelt, ist mit diesen Systemen nicht möglich.

In [9] wird ein *hear-through headset* [1] beschrieben. Als beste Position für die Mikrofone wird

[1] Hear-through headset bezeichnet hier ein In-Ear Kopfhörer, bestückt mit jeweils einem Mikrofon pro Seite, welche

die Außenkante am blockierten Ohrkanal ermittelt, um eine gute Lokalisierung und geringe Klangverfärbung zu erzielen. Das vorgestellte Headset ermöglicht eine sehr gute Lokalisation in der horizontalen Ebene. Verschiebungen der Quelle in der vertikalen Ebene können jedoch nicht präzise lokalisiert werden. Das weist darauf hin, dass relevante spektrale Cues (Pinna-Cues) vom System kompromittiert werden.

Kommerziell erscheinen aktuell verschiedene Kopfhörermodelle, die bereits Mikrofone integriert haben, um zum einen aktiv durch Antischall ein *noise cancelling* durchzuführen [2] oder um binaural aufnehmen zu können [3].

3 Umsetzung

Das finale System, das im Rahmen dieser Arbeit realisiert werden soll, soll aus einem Hardware und Softwareteil bestehen. Der Softwareteil soll hierbei von größerer Gewichtung sein. Die Hardware wird als Demonstrator realisiert.

Die Software wird für das mobile Betriebssystem iOS 8 von Apple implementiert. Die Software beinhaltet einen regelbaren Zwei-Kanal-Mischer. Auf dem ersten Kanal können die Mikrofonsignale eingeblendet werden. Auf dem zweiten Kanal wird Musik wiedergegeben. Die Musik wird aus der Media Library des Geräts bezogen und liegt in den gängigen Formaten MP3 oder AAC vor.

Die Entfernung zwischen Musikquelle und Hörer kann über einen Schieberegler eingestellt werden.

Der nötige HRTF Datensatz wird im Format OpenDAFF der RWTH Aachen angelegt. Um den Datensatz zu adressieren, wird openDAFF an Objective-C für iOS angepasst und in das Projekt integriert.

Es soll ein Kopfhörer geschlossener Bauart verwendet werden. Das ist in der gewünschten Konfiguration sinnvoll, da zum einen eine akustische Abschottung möglich sein soll, falls der erste Kanal (Kanal für Umgebungsgeräusche) deaktiviert wird. Dies ist ein Vorteil im Vergleich zu den zuvor erwähnten SARA Systemen, da der Benutzer frei bestimmen kann, ob nur die Musik oder auch die Umgebungsgeräusche hörbar sind. Zum anderen muss der Schall, der vom Kopfhörer nach außen abgestrahlt wird möglichst gut gedämpft werden, um eine Rückkopplung mit den außen am Kopfhörer angebrachten Mikrofonen zu vermeiden. Der Kopfhörer ist kabelgebunden.

Die optimale Position für Mikrofone (abschließend mit dem Ohrkanal) kann bei einem Umbau handelsüblicher Kopfhörer nicht gewährleistet werden. Die Mikrofone werden auf beiden Kopfseiten auf der Außenseite des Kopfhörers angebracht. Es ist davon auszugehen, dass eine optimale Lokalisierung der “durchgelassenen” Signale mit dieser Konfiguration nicht möglich ist, da unnatürliche ITD und Klangverfärbungen durch die Position der Mikrofone auftreten. Es soll im Laufe der Arbeit eine Konfiguration mit in-ear Kopfhörern und eine weitere mit

gesprochene Sprache “durchlassen”.

[2] Z.B. *Bose QC 20* und *Harman/Kardon NC*

[3] Z.B. *Hooke Wireless 3D Audio Headphones* und *Roland CS-10EM*

circum-auralen Kopfhörern untersucht werden und die Vor- und Nachteile diskutiert werden. Beide Fälle werden mit kostengünstigen Kopfhörermodellen realisiert.

Da bei einer Montage der Mikrofone an der Außenseite der Kopfhörer die ITDs größer werden, sind nach [10] Lokalisierungsfehler zu erwarten, der mit dem Winkel zwischen Quelle und Hörer variiert. Klangverfärbungen werden bei beiden Demonstratorvarianten erwartet.

Die Mikrofone werden per Bluetooth mit dem iOS Gerät verbunden, um keine zusätzliche Hardware (Audiointerface, Adapter, etc.) an dem Smartphone anschließen zu müssen.

4 Evaluation

Sowohl die Wahrnehmung der Externalisierung des Musiksignals, wie auch die Fähigkeit einer richtigen Lokalisation der eingespeisten Umgebungsgeräusche sollen zunächst informell vom Autor evaluiert werden. Ferner sollen technische Evaluationen Aufschluss über die Qualität des Systems liefern.

Die Frequenzgänge der gefilterten Musiksignale sollen mit den Frequenzgängen der Eingangssignale verglichen und dabei der Einfluss des *up-mixing* Prozesses und der Binauralsynthese untersucht werden. Weiter sollen die vom System künstlich generierten BRIRs qualitativ (plausible spektrale und zeitliche Gestalt) und quantitativ (Bestimmung der erzielten raumakustischen Parameter) untersucht werden.

Die Komponententrennung zwischen Primärkomponente und Abientkomponente, die als Vorstufe des *up-mixing* vorgenommen wird, soll ebenfalls überprüft werden. Hier bestimmt sich die Qualität des Algorithmus durch seine Fähigkeit die Richtung der Primärquelle bei verschiedenen Direktschall-/Diffusschallverhältnissen zu bestimmen.

5 Zeitplan

Zeitraum	Aufgabe
Mai 2015	Literaturrecherche, Exposé, Sichtung der Vorarbeiten
Juni 2015	Implementierung Multichannel MP3 Player für iOS 8.0 und Dokumentation
Juli 2015	Implementierung OpenDAFF framework für iOS 8.0, Erstellung eines eigenen OpenDAFF Datensatzes und Dokumentation, Implementierung Realtime Filter in Multichannel MP3 Player und Dokumentation
August 2015	Implementierung Mikrofon - Smartphone Kommunikation, Hardwareanpassungen Mikrofon - Kopfhörer und Dokumentation
September 2015	Implementierung Up-mixing - Externalisierungs-Algorithmus und Dokumentation, Evaluation und Dokumentation
Oktober 2015	Dokumentation

Literaturverzeichnis

[1] Lee, Taegyu; Yonghyun Baek; Young-cheol Park; and Dae Hee Youn (2014): "Stereo Upmix-based Binaural Auralization for Mobile Devices." In: *IEEE Transactions on Consumer Electronics*, **60**(3), pp. 411–419.

[2] AES (2014): "AES Technical Committee. Audio for Telecommunications." Online. URL `http://www.aes.org/technical/at/`.

[3] Møller, H. (1992): "Fundamentals of binaural technology." In: *Applied Acoustics*, **36**, pp. 171–218.

[4] Fuß, Alexander (2014): *Entwicklung eines vollsphärischen Multikanalmesssystems zur Erfassung individueller kopfbezogener Übertragungsfunktionen*. Master's thesis, Technische Universität, Berlin.

[5] Fallahi, Mina (2014): *A system for the fast measurement of individual head-related transfer functions. Simulation and implementation of measurement algorithms*. Master's thesis, Technische Universität, Berlin.

[6] Goodwin, Michael M. (2008): "Geometric signal decompositions for spatial audio enhancement." In: *ICASSP*, pp. 409–4012.

[7] Baek, Yong-Hyun; Se-Woon Jeon; Young-Cheol Park; and Seok-pil Lee (2012): "Efficient Primary-Ambient Decomposition Algorithm for Audio Upmix." In: *133rd AES Convention, Convention Paper 8754*. San Francisco, USA.

[8] Martin, Aengus; Craig Jin; and André Van Schaik (2009): "Psychoacoustic Evaluation of Systems for Delivering Spatialized Augmented-Reality Audio." In: *Journal of the Audio Engineering Society*, **57**(12), pp. 1016–1027.

[9] Hoffmann, Pablo F.; Anders Kalsgaard Møller; Flemming Christensen; and Dorte Hammershøi (2014): "Sound localization and speech identification in the frontal median plane with a hear-through headset." In: *Forum Acusticum*.

[10] Woodworth, Robert S. and Harold Schlosberg (1962): *Experimental psychology", Holt, ,*. New York: Holt, Rinehard and Winston.