



**Technische Universität Berlin**

Fakultät I - Geisteswissenschaften

Institut für Sprache und Kommunikation

Fachgebiet Audiokommunikation

Exposè

Masterarbeit

**Binaural Scrambler - Design und  
Implementierung eines Werkzeugs zur  
Echtzeitmanipulation binauraler Signale**

Edgar Arnold

Matr.-Nr.: 336714

29. April 2015

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

# Inhaltsverzeichnis

<b>1 Abstract</b>	<b>3</b>
<b>2 Einleitung</b>	<b>3</b>
2.1 Motivation . . . . .	3
2.2 Stand der Forschung . . . . .	4
<b>3 Anforderungsanalyse</b>	<b>5</b>
3.1 Anwendungsszenarien . . . . .	6
3.2 Anforderungen . . . . .	8
<b>4 Methoden</b>	<b>9</b>
4.1 Entwicklungsumgebung . . . . .	12
<b>5 Evaluation</b>	<b>13</b>
<b>6 Arbeits- und Zeitplan</b>	<b>13</b>
<b>Literatur</b>	<b>15</b>

# 1 Abstract

Die DFG-Forschergruppe *Simulation and Evaluation of Acoustical Environments (SE-ACEN)* widmet sich der Virtualisierung und Evaluation akustischer Umgebungen. Das in [1] präsentierte standardisierte Vokabular *Spatial Audio Quality Inventory (SAQI)* erlaubt eine differenzierte auditive Bewertung virtueller akustischer Umgebungen. Das *SAQI* umfasst 48 Deskriptoren (*SAQI*-Qualifier) zur qualitativen Beschreibung binauraler Audiosignale. Diese Wahrnehmungsaspekte ermöglichen differenzierte Vergleiche der virtuellen spatialen Umgebungen mit realen, virtuellen oder imaginären Referenzumgebungen im Rahmen von Nutzertests. Für die Durchführung und Evaluation der Versuche wird das Softwarewerkzeug *WhisPER* [2] verwendet. Um die Reliabilität [3] von Ergebnissen perceptiver Versuche zu gewährleisten, wird ein über die Stichprobe homogener Grad an Expertise bzgl. der Einschätzung binauraler Signale anhand von Qualitätsmerkmalen angestrebt. Diese Homogenität kann durch ein adäquates Training der Versuchspersonen erzielt werden. Im Rahmen der geplanten Masterarbeit soll daher die softwaretechnische Planung und Implementierung eines Werkzeugs zur Echtzeitmanipulation binauraler Audiosignale in Hinblick auf die 48 *SAQI*-Qualifier durchgeführt werden, mit dessen Hilfe Versuchspersonen auf Hörversuche vorbereitet, bzw. Ankerstimuli für Hörversuche erstellt und in Echtzeit verändert werden können.

## 2 Einleitung

### 2.1 Motivation

Das in [1] vorgestellte Vokabular dient dazu, virtuelle akustische Umgebungen nach 48 (nicht zwingend disjunkten) perceptiven Qualitätsaspekten differenziert zu bewerten. Das Vokabular umfasst die folgenden 8 übergeordneten Kategorien, welche jeweils 3 bis 10 Aspekte beinhalten :

- *Klangfarbe, Tonalität, Geometrie, Raum, Zeitverhalten, Dynamik, Artefakte, Allgemeines*

Die Bewertung der *Virtual Acoustic Environment (VAE)*s soll dabei von Probanden vorgenommen werden, welche aufgrund ihrer Erfahrung und/oder intensiven Trainings

sehr gut mit den Wahrnehmungsqualitäten vertraut sind.

Um die Auftrittswahrscheinlichkeit von subjektspezifischen Bewertungsfehlern zu minimieren, soll nun eine Trainingsumgebung entwickelt werden, mit deren Hilfe ein Reiz entsprechend eines Qualitätsmerkmals manipuliert werden kann.

In [4] wird zwischen zufälligen und systematischen Bewertungsfehlern unterschieden. Dabei sind die zufällig vorkommenden Fehler als mehr oder minder zufällige Streuung in den erhobenen Bewertungen erkennbar und durch Verwendung erwartungstreuer Trendschätzer (z.B. des Mittelwerts) behandelbar. Systematische Fehler, wie z.B. verschiedene Saklierungstendenzen sind dagegen sowohl schwieriger zu identifizieren als auch nachträglich gezielt zu behandeln (üblws. durch Zentralisierung, Re-Skalierung). Die systematischen Fehler werden auch *bias* genannt. *Biases* können probandenspezifisch auftreten oder auch innerhalb einer Stichprobe. So etwa der *Range Equalization-Bias*, welcher das Phänomen beschreibt, dass innerhalb einer (untrainierten) Stichprobe bei der Bewertung von Stimuli stets auf den gesamten Wertebereich der Skale abgebildet wird. Die Bewertungen aufgrund unterschiedlicher Stichproben können deshalb nur valide miteinander verglichen werden, wenn die Probanden ausreichend mit den Qualitätsmerkmalen vertraut sind und/oder Ankerstimuli [5] den perzeptiven Definitionsbereich ergänzen.

Neben einer solchen Trainingsumgebung soll das Programm, der sogenannte *Binaural Scrambler*, auch die Möglichkeit der Generierung von Ankerstimuli bieten. Diese sollen durch additive Überlagerung der von den SAQI-Qualifiern abhängigen Signalverarbeitungen in Echtzeit synthetisiert sein. Die Auswahl und Ausprägung der jeweiligen SAQI-Qualitäten innerhalb eines solchen Ankerstimulus' soll preset-basiert erfolgen und über das *Open Sound Control (OSC)*-Protokoll kommuniziert werden.

## 2.2 Stand der Forschung

Die für die für die Synthese von binauralen VAEs verwendeten binauralen Signale werden durch die relative Lage der Schallquelle zum Kopf des Hörers in Kugelkoordinaten (Radius, Azimuth, Höhenwinkel), die individuelle Morphologie des Hörapparats (Kopfmaße) [6] sowie durch die Raumakustik bestimmt, letzteres zumindest in dem Falle, dass BRIRs *Binaural Room Impulse Response (BRIR)s* vorliegen (näheres in [7]). Ein am Fachgebiet Audiokommunikation entwickeltes Verfahren, durch welches

mithilfe eines entlang der Achsen justierbaren Kunstkopfes für diskrete Positionen die *BRIR* gemessen werden kann, wird in [8] vorgestellt. Ein binaurales Signal wird nach [9] abhängig von der Kopfposition dynamisch mittels Faltung eines nachhallfreien, monophonen Eingangssignals mit der zuvor erhobenen *Head Related Impulse Response (HRIR)* bzw. *BRIR* auralisiert. Die Lokalisationswahrnehmung ist dabei durch die Ausprägung bestimmter Signaleigenschaften, wie z.B. der *Interaural Time Difference (ITD)*, der *Interaural Level Difference (ILD)* und der *Spectral Cues (SC)* bestimmt [6].

Die zu implementierende Anwendung ist dem Bereich *Technische Gehörbildung* zuzuordnen. In der Vergangenheit bildeten Hochschulkurse wie das *Timbre Solfege* [10], ein Lehrgang an der Chopin Academy, Warschau, die primäre Fortbildungsoption. Tonmeister wurden dort in einem Zeitraum von 3 Jahren in allen Aspekten der Audioqualität geschult. Im Gegensatz dazu ist *Golden Ears* [11], eine Anleitung zur technischen Gehörbildung, zum Selbststudium bestimmt. Das Handbuch besteht aus 4 Teilwerken mit jeweils 2 Audio-CDs und dazugehörigen Testbögen, welche den Audioingenieur instruieren, die auf das Hörmaterial angewandten Equalizereinstellungen oder Signalverarbeitungstypen (z.B. Hall, Kompression, Verstärkung) zu identifizieren.

Kontemporative Übungsplattformen im Bereich Gehörbildung liegen meist in Form einer interaktiven digitalen Anwendung vor. Vorteile der softwarebasierten Lösungen gegenüber den konventionellen Methoden sind nach [12] bspw. die individuelle Anpassung des Programms an den Kenntnisstand des Nutzers, die Beschleunigung des Lernprozesses durch multimodales Feedback (textlich, visuell, auditiv), sowie die Möglichkeit der automatischen Erhebung von Nutzerdaten für eine anschließende statistische Auswertung. Vertreter aus dem Bereich der computergestützten Gehörbildung werden in [12] und [13] vorgestellt.

### 3 Anforderungsanalyse

Die geplante Anwendung *Binaural Scrambler* soll sowohl als Trainingswerkzeug zum Selbststudium der Wahrnehmungsqualitäten, als auch zur Generierung von Ankerstimuli dienen.

### 3.1 Anwendungsszenarien

Dazu soll es ermöglicht werden, dynamische (abhängig von Kopfposition der Versuchsperson) binaurale Signale in Echtzeit zu manipulieren, ohne Kenntnis und Bearbeitung der *HRIR/BRIR*. Ebenso soll die offline-Generierung von *OSC*-Presets im *Binaural Scrambler* bewerkstelligt werden. Die Zuordnung von Signalverarbeitungsalgorithmen zu Ankerstimuli soll aus *WhisPER* heraus möglich sein. Die Auswahl der gewünschten Typen und Ausprägungen der Wahrnehmungsqualitäten wird dabei mithilfe des *OSC*-Protokolls zum *Binaural Scrambler* übertragen. Sie erfolgt presetbasiert (für eine Qualitätsverkettung) oder qualitätsspezifisch (für einzelne Qualitäten). Der Austausch der binauralen Signale in Echtzeit kann bspw. durch die Anwendung von Transportprotokollen (z.B. *OSC*) für den Fernzugriff oder lokal mit Dateiströmen realisiert werden. Die möglichen Anwendungsszenarien werden zur besseren Übersicht mithilfe von Diagrammen visualisiert.

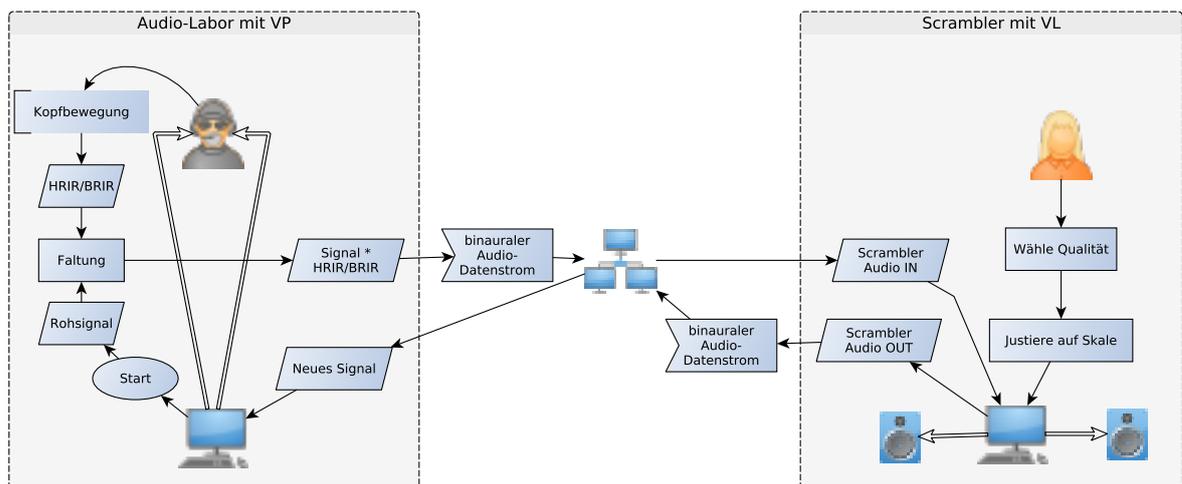


Abbildung 1: Anwendung zur Echtzeitmanipulation für Testzwecke

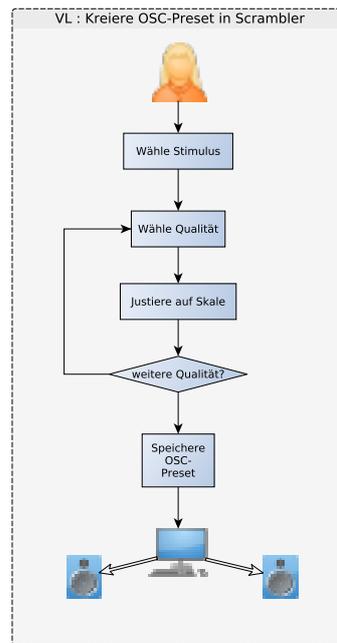


Abbildung 2: Anwendung zur Erstellung von *OSC*-Presets für *WhisPER* offline

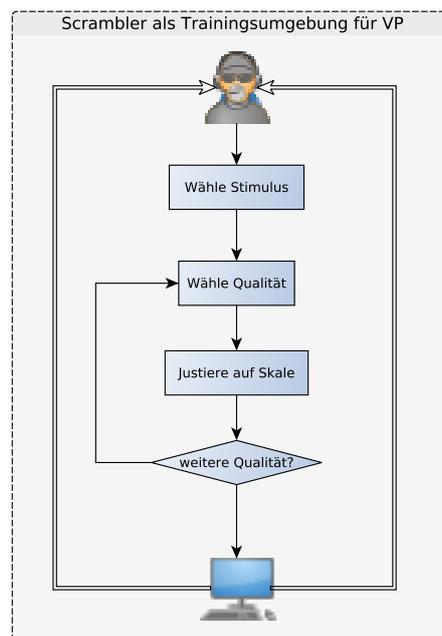
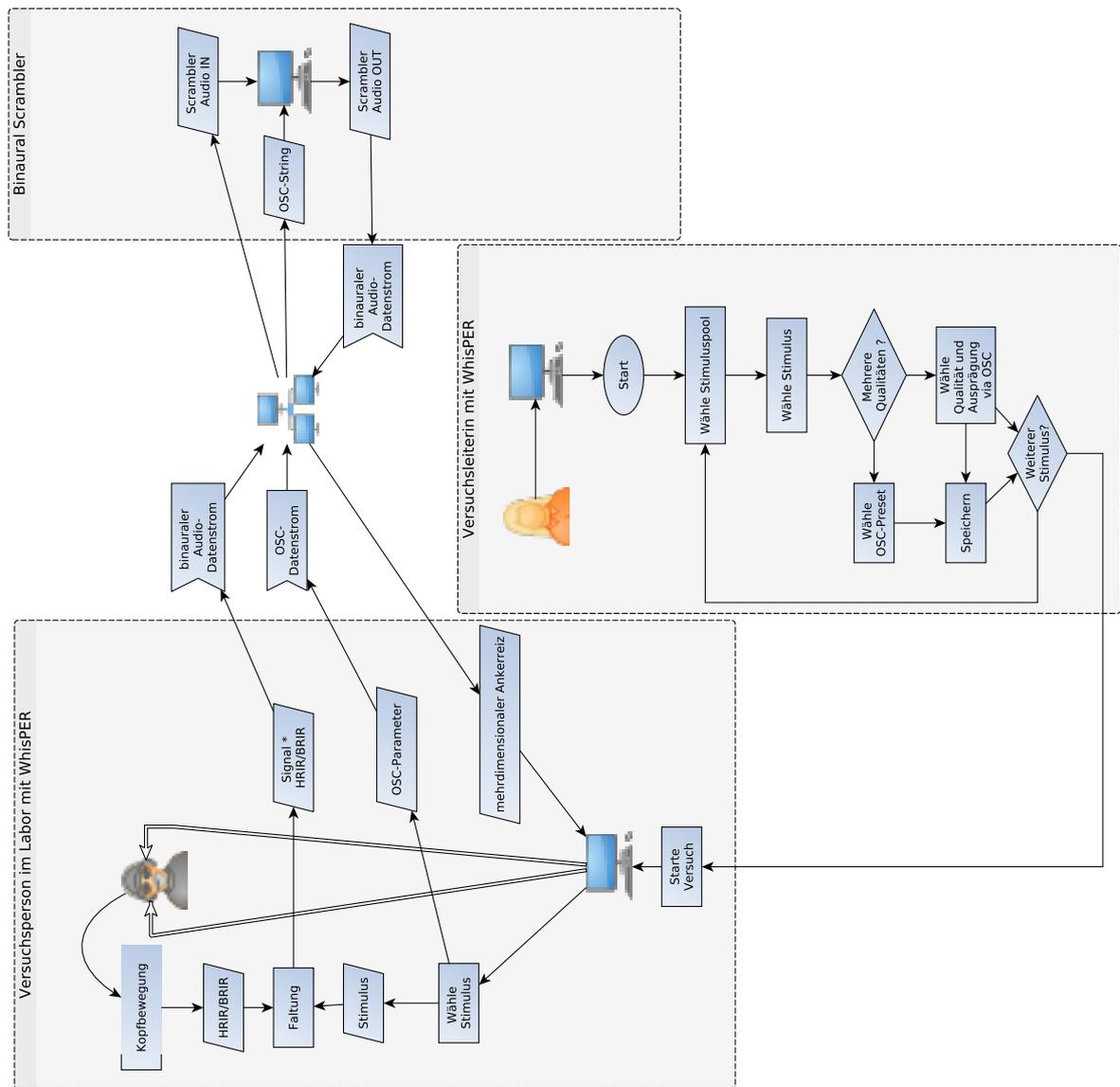


Abbildung 3: Anwendung als Trainingsumgebung

Abbildung 4: Anwendung zur Manipulation von Ankerstimuli in *WhisPER*

### 3.2 Anforderungen

Neben der Umsetzung der echtzeitfähigen Signalmanipulationsalgorithmen für möglichst viele der 48 Wahrnehmungsqualitäten, sollen auch nicht-funktionale Anforderungen an die Software berücksichtigt werden. Das Softwaredesign soll dabei wesentliche Kriterien wie Wartbarkeit (Entwicklungsaufwand), Erweiterbarkeit, Nutzerfreundlichkeit und Plattformunabhängigkeit erfüllen. Darüber hinaus soll das Programm fernsteuer-

bar und reliabel sein, d.h. für gleiche Eingaben sollen stets gleiche Resultate ausgegeben werden. Desweiteren soll das System mit Hilfe von freier Software umgesetzt werden. Der gewählte Softwaretechnologiestack (Programmiersprache, Rahmenwerk, Hilfswerkzeuge) sollte möglichst alle genannten Kriterien erfüllen und gleichzeitig eine ansprechende Performanz, selbst für komplexe Anwendungsfälle mit parallelen Programmfäden, bieten.

## 4 Methoden

Mit dem im Rahmen dieser Arbeit zu implementierenden Werkzeug soll es nun möglich sein, binaurale Signale in Echtzeit qualitätsspezifisch zu manipulieren. Die *HRIR* bzw. *BRIR* wird dem Werkzeug dabei nicht bekanntgegeben, d. h., die Bearbeitung erfolgt direkt auf den binauralen Signalen.

Die Basis der Signalverarbeitungsalgorithmen für die Synthese der *SAQI*-Qualifier bilden die Erkenntnisse aus [14]. Der Aufwand für die Implementierung einer Qualitätsmanipulationen ist dabei abhängig von der jeweiligen Qualitätskategorie.

Tabelle 1: Einteilung der Aspektkategorien nach Schwierigkeitsgrad der Effekturnsetzung

<i>Basis-Effekte</i>	<i>Experimentelle Effekte</i>	<i>Intelligente Effekte</i>
Dynamik	Allgemeine Qualitäten	Geometrie
Zeitverhalten	Artefakte	Raum
	Klangfarbe	
	Tonalität	

*Basis-Effekte* sind dabei solche, die erwartungsgemäß mithilfe von einschlägigen Lehrbüchern realisiert werden können. Für jene existiert bereits eine umfangreiche Sammlung an Berechnungsvorschriften. Als Beispiel sei die *SAQI*-Qualität *Nachechos* (Kategorie **Zeitverhalten**) genannt, welche als Kammfilter mit Vorwärtsspeisung umgesetzt werden können (wenn Rückspeisung, Überdeckung gleich 0 sind, siehe Abbildung 1).

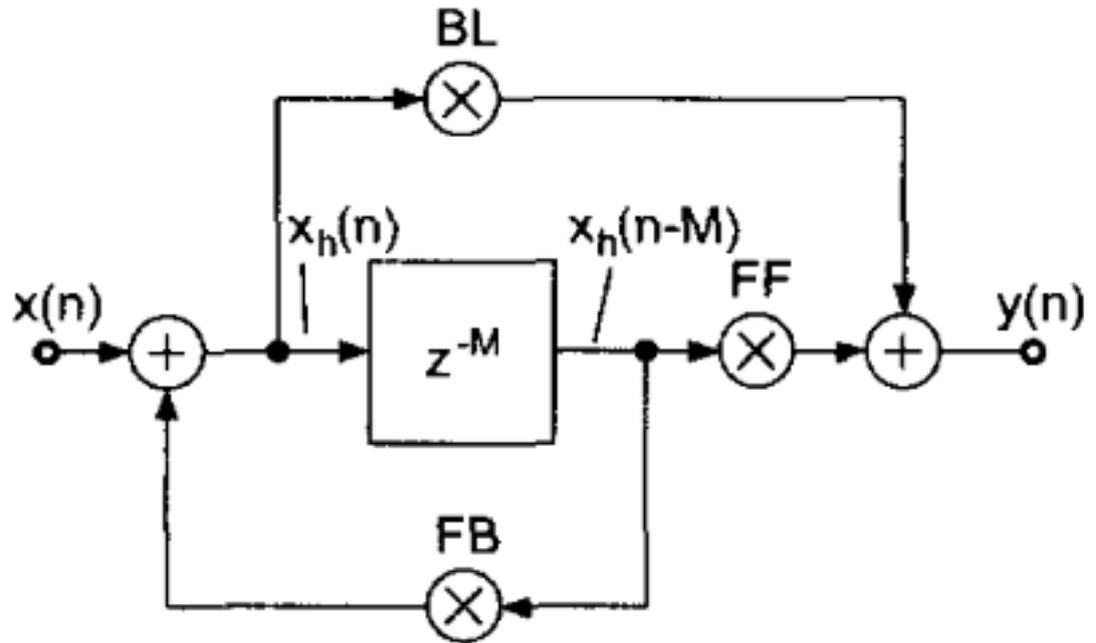


Abbildung 5: Kammfilter mit Überdeckung (BL), Vorwärtsspeisung (FF), Rückspeisung (FB) aus [14]

Die Umsetzung der *Experimentellen Effekte* bedeutet voraussichtlich mehr Forschungsaufwand. Da für die Qualitäten aus diesen Kategorien mehrere Umsetzungsstrategien denkbar sind, gilt es bei der Implementierung eine Kombination geeigneter Effekte zu finden, welche der gewünschten auditiven Veränderung am ehesten entspricht. Als Beispiel sei die *Metallische Klangfarbe* (Kategorie **Klangfarbe**) genannt, welche z.B. durch den Einsatz eines resonierenden Kammfilters, eines Ringmodulators oder einer Kombination der genannten realisiert werden kann. Die optimale Methode soll iterativ bestimmt werden.

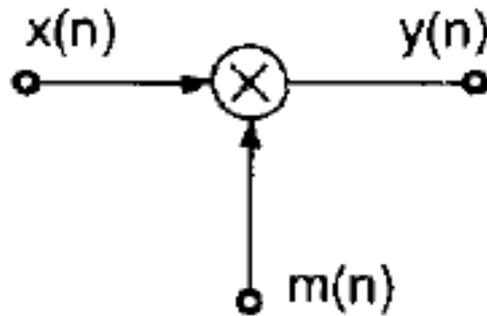


Abbildung 6: Ringmodulator nach [14] : Multiplikation im Zeitbereich

Manipulationen aus dem Bereich **Intelligente Effekte** erfordern während der Veränderung der jeweiligen Qualität idealerweise eine simultane Schätzung der aktuell wahrgenommenen Schallquellenposition. Um den räumlichen Höreindruck zu verändern, ohne dabei den realistischen Wertebereich zu verlassen, muss die Position der Schallquelle im virtuellen Raum approximiert werden, bevor diese durch den *Binaural Scrambler* editiert wird. Für die Umsetzung dieser Schätzungen kann man sich kontemporärer Algorithmen aus dem Bereich des *maschinellen Lernens* bedienen. Für die Approximation der Schallquellenposition durchläuft der Algorithmus bspw. eine Lernphase, in der die Abbildung von Schallquellenpositionen auf die obigen Lokalisationsparameter als globale Wissensbasis abgespeichert wird. Mögliche Verfahren zur Extraktion der *ITD* aus binauralen Signalen sind in [7] zusammengetragen. Unter der Berücksichtigung der in Echtzeit geschätzten Werte für *ITD,ILD,SC* und mithilfe von Interpolationsverfahren kann dann die perzipierte Schallquellenposition je nach Ausprägung der *SAQI*-Qualität verändert werden. Diese Verfahren dienen überdies dazu, den Wertebereich einer Manipulation einzuschränken, um Verzerrungen, bspw. hervorgerufen durch Clipping, zu vermeiden.

Im Folgenden wird die Vorgehensweise bei der Wahl einer geeigneten Softwaretechnologie für die Umsetzung der Bearbeitungsumgebung vorgestellt.

## 4.1 Entwicklungsumgebung

Aus den Anforderungen ergeben sich eine Reihe an Unterscheidungsmerkmalen der zu evaluierenden Technologien, wie z.B. der Installationsaufwand auf verschiedenen Betriebssystemen (Plattformunabhängigkeit), die Simplizität der Netzwerkkommunikation (Fernsteuerbarkeit, z.B. via *OSC* von *WhisPER*), Entwicklungsaufwand (Erweiterbarkeit und Wartbarkeit), Intuitivität und Freiheit bei der Erstellung der grafischen Nutzeroberfläche (Nutzerfreundlichkeit), Eignung für den professionellen Einsatz (Reliabilität), sowie die Qualität des Compilers und der Laufzeitumgebung (Performanz).

Folgende Kandidaten sind bei der Recherche in den engeren Auswahlkreis der möglichen Technologien gelangt :

- **Juce** DSP-Rahmenwerk und Codegenerationsumgebung für die Programmiersprache *C++* [15]
- **Pure Data (pd)** Visuelle Programmiersprache- und umgebung für komplexe Signalflusszenarien [16]
- **Web Audio API** DSP-Rahmenwerk für die Programmiersprache *Javascript* [17]

Die Bewertung der Kandidaten der technologischen Rahmenwerke anhand der Kriterien wird in der Tabelle 2 veranschaulicht. Je besser ein Rahmenwerk das jeweilige Kriterium erfüllt, desto mehr Punkte erhält es in der Tabelle.

Tabelle 2: Punktevergabe anhand der Evaluationskriterien

Kriterium	<i>Juce</i>	<i>pd</i>	<i>Web Audio</i>
Performanz	X X X X X	X X X X	X X X
Reliabilität	X X X X X	X X X X	X X
Entwicklungsaufwand	X X	X X X X	X X X
GUI-Komponenten	X X X X	X X X	X X X X X
Installationsaufwand	X X X X	X X X X	X X X
Netzwerkkommunikation	X X X X	X X X X	X X X X X

Es lässt sich resümieren, dass die Wahl tendenziell auf *Juce* oder *pd* fällt, da die Reliabilität für Echtzeitanwendungen von essentieller Bedeutung ist, die *Web Audio API* diese aufgrund der browserabhängigen Compilerqualität jedoch nicht gewährleisten kann.

Die letztendliche Wahl wird zu Beginn der Arbeit nach der Umsetzung einer prototypischen *OSC*-Kommunikationsumgebung in den beiden favorisierten Technologien erfolgen.

## 5 Evaluation

Die Anzahl der zu implementierenden Signalbearbeitungsmethoden soll durch ein zeitliches Abbruchkriterium (z.B. 60% der Gesamtbearbeitungszeit) begrenzt werden. Dabei sollen die einzelnen Methoden in Reihenfolge aufsteigender Komplexität implementiert werden (d.h. beginnend mit *basal* über *experimentell* zu *intelligent*). Die Restzeit (abzüglich Verschriftlichungsdauer) soll für eine formale perzeptive Evaluation (auditive Skalierung ausgewählter Ankerreize nach *SAQI*-Methodik) aller bis dahin implementierten Methoden verwendet werden. Die Fertigstellung des Gesamtprojekts soll im Rahmen einer teils bereits parallel anlaufenden Anschlussarbeit (X. Presser, cand. MSc AKT) erreicht werden.

## 6 Arbeits- und Zeitplan

Zeitraum	Ziele
April - Mai	Softwaredesign und Literaturrecherche
Mai - Juli	Implementierung der GUI, Basiseffekte
August - Oktober	Versuchsplanung/durchführung, optional : Implementierung Experimenteller Effekte

## Abkürzungsverzeichnis

**BRIR** Binaural Room Impulse Response.

**HRIR** Head Related Impulse Response.

**ILD** Interaural Level Difference.

**ITD** Interaural Time Difference.

**OSC** Open Sound Control.

**pd** Pure Data.

**SAQI** Spatial Audio Quality Inventory.

**SC** Spectral Cues.

**SEACEN** Simulation and Evaluation of Acoustical Environments.

**VAE** Virtual Acoustic Environment.

## Literatur

- [1] Lindau, Alexander (2014): „Spatial Audio Quality Inventory (SAQI). Test Manual.” In: *Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin*.
- [2] Ciba, Simon; Wlodarski, André und Maempel, Hans-Joachim (2009): „Whisper – A New Tool for Performing Listening Tests.” In: *Audio Engineering Society Convention 126*. S. 1–100. URL <http://www.aes.org/e-lib/browse.cfm?elib=14945>.
- [3] Bortz, J. und Döring, N. (2006): *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler*. Springer-Lehrbuch. Physica-Verlag. URL <http://books.google.de/books?id=13GbPUYAUHsC>.
- [4] Zielinski, Slawomir; Rumsey, Francis und Bech, Søren (2008): „On Some Biases Encountered in Modern Audio Quality Listening Tests-A Review.” In: *J. Audio Eng. Soc.*, 56(6) S. 427–451. URL <http://www.aes.org/e-lib/browse.cfm?elib=14393>.
- [5] International Telecommunications Union, Switzerland, Geneva (2014): *Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems*.
- [6] Nicol, R. (2010): *Binaural Technology*. Monographs of the American Ethnological Society. Audio Engineering Society. URL <http://books.google.de/books?id=bjpBMwEACAAJ>.
- [7] Lindau, Alexander; Estrella, Jorgos und Weinzierl, Stefan (2010): „Individualization of Dynamic Binaural Synthesis by Real Time Manipulation of ITD.” In: *Audio Engineering Society Convention 128*. URL <http://www.aes.org/e-lib/browse.cfm?elib=15385>.
- [8] Lindau, Alexander und Weinzierl, Stefan (2007): „FABIAN – Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden.”
- [9] Lindau, Alexander; Hohn, Torben und Weinzierl, Stefan (2007): „Binaural Resynthesis for Comparative Studies of Acoustical Environments.” In: *Audio Engineering Society Convention 122*. URL <http://www.aes.org/e-lib/browse.cfm?elib=14017>.

- 
- [10] Miskiewicz, Andrzej (1992): „Timbre Solfege: A Course in Technical Listening for Sound Engineers.” In: *J. Audio Eng. Soc.*, 40(7/8) S. 621–625. URL <http://www.aes.org/e-lib/browse.cfm?elib=7041>.
- [11] Moulton, David (1992): *Golden Ears Manual*. KIQ Productions.
- [12] Quesnel, René und Wieslaw R. Woszczyk (1994): „A Computer-Aided System for Timbral Ear Training.” In: *Audio Engineering Society Convention 96*. URL <http://www.aes.org/e-lib/browse.cfm?elib=6376>.
- [13] Olive, Sean (2001): „A New Listener Training Software Application.” In: *Audio Engineering Society Convention 110*. URL <http://www.aes.org/e-lib/browse.cfm?elib=9960>.
- [14] Zölzer, Udo (2002): *DAFX - Digital Audio Effects*. John Wiley & Sons.
- [15] ROLI Ltd. (2015): „JUICE Framework.” <http://www.juce.com/>. Accessed: 2015-02-15.
- [16] Puckette, Miller (2015): „Pure Data.” <http://puredata.info/>. Accessed: 2015-02-15.
- [17] Mozilla Corp. (2015): „Web Audio API spec.” [https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Audio\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Audio_API). Accessed: 2015-02-15.