Abschlussarbeit zum Toningenieursprojekt

Implementierung eines Modells zur Messung und Darstellung binauraler Lokalisation

Stefan Richardt

stefan_richardt@web.de

Graz, 19.02.2010

Betreuer:

Dr. Franz Zotter

UNIVERSITÄT FÜR MUSIK UND DARSTELLENDE KUNST GRAZ · AUSTRIA

Institut für elektronische Musik und Akustik

Inhaltsverzeichnis

1 Einleitung und Aufgabenstellung

Implementierung eines Modells zur Messung und Darstellung binauraler Lokalisation

Zur Beurteilung akustischer Situationen im Hinblick auf ihre Lokalisierbarkeit bzw. Lokalisationsschärfe bei der Richtungswahrnehmung im Gehör bedarf es eines beträchtlichen Aufwands, da es um aussagekräftige Ergebnisse zu erzielen eine Mindestanzahl an Versuchspersonen benötigt.

Ziel der Projektarbeit ist es, eine praktikable Applikation zu erarbeiten, die ein schnelles Abschätzen der menschlichen Lokalisation ermöglicht. Durch ein geeignetes Modell der Hörwahrnehmung speziell der Lokalisation, sollen mit einem Kunstkopf aufgenommene akustische Ereignisse im Sinne ihrer Lokalisationswahrscheinlichkeit bewertet und graphisch dargestellt werden. Aus einer Vielzahl bereits vorhandener Modelle der menschlichen Hörwahrnehmung, gilt es nun ein geeignetes auszuwählen, dieses zu modifizieren und zu erweitern.

Das erarbeitete Modell wird je nach Versuchsaufbau eine Schätzung der tatsächlichen Lokalisation sein und in Detailfragen reale Hörversuche nicht ersetzen können. Zur Veranschaulichung der Ergebnisse und zur schnellen Auswertung soll die Applikation aussagekräftige und leicht verständliche Grafiken generieren.

2 Modell nach Faller und Merimaa

Als Grundlage meiner Arbeit dient ein Modell von Christof Faller und Juha Merinaa, vorgestellt im Artikel „Source localization in complex listening situations: Selection of binaural cues based on interaural coherence“. Das Modell orientiert sich am menschlichen Ohr, wobei versucht wird den Signalfluss nachzubilden, sodass das Modell möglichst viele, für die Lokalisation wichtige Effekte beinhaltet. Gegliedert ist das Modell in drei Teile 1. *Auditory periphery* 2. *Binaural processor* 3. *Higher model stages*.

Die Arbeit von Faller und Merinaa eignet sich gut als Grundlage, da das Modell nur bis zu einem bestimmten Punkt implementiert ist und die erhaltenen Testergebnisse umfassend und ausführlich dokumentiert sind. Es wird klar wo die Grenzen des Modells liegen, und wo ein sinnvolles Weiterarbeiten möglich ist.

2.1 Modellbeschreibung

2.1.1 Auditory periphery

Das Eingangssignal soll dem Signal am Ohr des Hörers entsprechen. Um dies zu erreichen nimmt man das Testsignal mit einem Kunstkopf auf oder simuliert ein binaurales Signal. Dies ist mittels *head related transfer functions* (HRTFs) oder *binaural room impulse responses* (BRIRs) möglich, indem man das Testsignal mit dem gewünschten Paar HRTFs oder BRIRs filtert. HRTFs und BRIRs sind winkelabhängige Übertragungsfunktion von Kopf, Schulter und Außenohr, wobei BRIRs zusätzlich eine Rauminformation enthalten. Schlussendlich liegt ein binaurales Stereosignal vor.

Kernstück des Modells ist die Simulation des Innenohrs. Um die Frequenzanalyse der Basilarmembran nachzubilden wird zuerst eine Gammatone – Filterbank verwendet, die sowohl vom linken als auch vom rechten Signal durchlaufen wird. Jedes resultierende kritische Band wird nun nach einem Modell von Bernstein (1999) bearbeitet. Die folgenden Bearbeitungsschritte dienen zur Simulation der neuralen Übertragung. Die Einhüllende eines jeden Signals wird mittels Hilberttransformation errechnet und mit 0.23 potenziert. Die komprimierte Einhüllende wird mit dem ursprünglichem Signal punktweise multipliziert. Danach wird eine Halbwellengleichrichtung durchgeführt. Das so erhaltene Signal, wird wiederum quadriert und durch einen Tiefpass vierter Ordnung mit einer Grenzfrequenz von 425 Hz gefiltert. Verwendet wird die frei verfügbare Matlab – Toolbox von Slaney (1998) und Akeroyd (2001), in der das oben beschriebene Modell Bernsteins implementiert ist.

Vor dem Bearbeitungsschritt der neuronalen Übertragung fügt Faller den Signalen Rauschen hinzu, um die Ungenauigkeiten des Gehörs zu simulieren. Das unkorrelierte Gaußsche Rauschen hat Faller zuvor mit den selben Gammatone-Filtern bearbeitet, wie das jeweilige kritische Band. Das Rauschen hat in jedem Frequenzband einen Pegel, der in der Nähe der Hörschwelle liegt.

2.1.2 Binaural processor

Ziel ist es nun die i*nteraural time difference* (ITD), die *interaural level difference* (ILD) und die *interaural coherence* (IC) zu bestimmen, mit Hilfe derer auf die Lokalisation geschlossen werden kann. Der binaurale Prozessor berechnet nun für alle vorliegenden Signale, d.h. für jedes kritische Band, die normalisierte Kreuzkorrelation. Die fortlaufende Berechnung wird rekursiv durchgeführt, was eine einfache Realisierung eines exponentiell ausklingenden Zeitfensters mit einer bestimmten Dauer T ermöglicht.

Aus Gründen der Übersichtlichkeit wurde die Variable z für das jeweilige Band bei den Formeln nicht mit angegeben. Die Berechnungen müssen jedoch für jedes Band separat ausgeführt werden. Die für den *binaural processor* verwendeten Formeln folgen den Beschreibungen:

$$\gamma(n,m)=\frac{a_{12}(n,m)}{\sqrt{a_{11}(n,m)\cdot a_{22}(n,m)}}$$ Kohärenz

hierfür benötigt man Kreuz – und Autokorrelation:

$$a_{12}(n,m)=\alpha\, x_1(n-max\{m,0\})\cdot x_2(n-max\{-m,0\})+(1-\alpha)\, a_{12}(n-1,m)$$ Kreuzkorrelation

$$a_{11}(n,m)=\alpha\, x_1^2(n-max\{m,0\})+(1-\alpha)\, a_{11}(n-1,m)$$ Autokorrelation

$$a_{22}(n,m)=\alpha\, x_2^2(n-max\{-m,0\})+(1-\alpha)\, a_{22}(n-1,m)$$ Autokorrelation

n	Sample	m	Verschiebung in Samples
x_1	Eingangssignal links	x_2	Eingangssignal rechts
α	Zeitkonstante	$T=\frac{1}{\alpha f_s}$	Dauer der Fensters

Die Eingangssignale x_1 und x_2 werden nach obigen Formeln korreliert. Aus der Kreuz – bzw. Autokorrelation lässt sich $\gamma(n,m)$ die Kohärenz in Abhängigkeit des Zeitpunktes n und der Verschiebung m berechnen. Die Kohärenz entspricht einer normierten Autokorrelation, deren Wertebereich zwischen [0,1] liegen, wobei 0 für vollständig unkorreliert und 1 für eine vollständige Übereinstimmung der Signale steht.

Der Bereich der Verschiebung m kann auf [-1,1] ms eingeschränkt werden; in Samples $m/f_s \in [-1,1]ms$. Die Einschränkung ist möglich, da eine Verzögerung des Eingangssignals von einem zum anderen Ohr von mehr als 1 ms aus anatomischen Gründen nicht überschritten werden kann.

Ebenso ist es sinnvoll die Signallänge und damit die Datenmenge zu beschränken. Die von C.Faller und J.Merimaa untersuchten Signale haben ein Länge von maximal einer Sekunde, meist deutlich weniger.

Schwieriger gestaltet sich Wahl der Fensterlänge T. Studien kommen zu dem Ergebnis, dass ein zweiseitiges Fenster der Länge 20 – 40 ms (Kollmeier und Gilkey, 1990) in Bezug auf die menschlichen binauralen Wahrnehmung am sinnvollsten wäre. Da ein zweiseitiges Fenster den Präzedenzeffekt auch nicht simulieren kann, ist ein kürzeres, einseitiges Fenster ausreichend und wird von C.Faller und J.Merimaa mit 10 ms festgelegt.

Aus der berechneten Kohärenz wird kann ITD und IC wie folgt berechnet werden:

$$\tau(n) = arg\max_m \gamma(n,m)$$ ITD Interaural Time Difference

$$c_{12}(n) = \max_m \gamma(n,m)$$ IC Interaural Coherence

Sowohl der Index des Maximums als auch das Maximum selbst beziehen sich auf alle m, bei fortlaufendem n. In Abhängigkeit der ITD wird nach folgenden Formel die ILD bestimmt:

$$\Delta L(n) = 10\log_{10}\frac{L_2(n,\tau(n))}{L_1(n,\tau(n))}$$ ILD Interaural Level Difference

$$L_1(n,m) = \alpha x_1^2(n-max\{m,0\}) + (1-\alpha)L_1(n-1,m)$$

$$L_2(n,m) = \alpha x_2^2(n-max\{-m,0\}) + (1-\alpha)L_2(n-1,m)$$

$$E(n) = L_1(n,\tau(n)) + L_2(n,\tau(n))$$ Signalenergie

Als Ergebnis liegt nun für jedes einzelne Band die Zeitdifferenz ITD, die Pegeldifferenz ILD und die Kohärenz IC in Abhängigkeit der Zeit vor { $\tau(n)$ $\Delta L(n)$ $c_{12}(n)$ }. Sie bilden die Grundlage für alle weiteren Analysen. ITD und ILD sind jedoch nur unter bestimmten Bedingungen korrekt. Um falsche bzw. abweichende Ergebnisse auszuschließen, kann die IC als ein Maß für die Gültigkeit von ILD und ITD verwendet werden. Die sogenannten *higher model stages* beschäftigen sich genau mit dieser Thematik.

2.1.3 Higher model stages

Als Grundlage für alle weiteren Analyse dienen ausschließlich ITL, ILD und IC { $\tau(n)$ $\Delta L(n)$ $c_{12}(n)$ }. Im Folgendem wird für $\tau(n)$ und $\Delta L(n)$ festgelegt, welche der ermittelten Daten überhaupt korrekt sein können. Mit Hilfe der Kohärenz $c_{12}(n)$ können somit alle nicht relevanten bzw. falschen Werte ausgeschlossen werden, sodass bei einem Zusammenfassen der Ergebnisse, die Genauigkeit erheblich erhöht werden kann.

Um ein exaktes Ergebnis zu erhalten, muss es sich bei ITL und ILD um sogenannte *free field cues* handeln. Das heißt, die errechneten Werte der ITL und ILD beruhen auf einem Eingangssignal einer einzigen Quelle, aufgenommen unter Freifeldbedingungen. Diese Werte, ohne jeden Raumeinfluss lassen einen eindeutigen Schluss auf den genauen Winkel der Quelle zu. Enthält das Eingangssignal Raumreflexionen oder mehrere Quellen, so spricht man nicht mehr von *free field cues.* Die resultierenden ITDs und ILDs verlieren an Aussagekraft, da dies zu einer Verwischung bzw. zum Springen der Ergebnisse führt. Mittels IC versucht man nun ungültige Werte auszuschließen, um für ITD, ILD möglichst die Qualität der *free field cues* zu erreichen. Es stellt sich also die Frage: Wie kann ich mit Hilfe der IC über die Korrektheit der Ergebnisse entscheiden?

Die Kohärenz $c_{12}(n)$ kann ausschließlich Werte zwischen Null und Eins annehmen. Besteht das Eingangssignal beispielsweise aus mehrere aufeinander folgenden Quellen, so lässt sich das Verhalten der Kohärenz in zwei Phasen unterteilen.

1. Die Ergebnisse resultieren aus dem Direktschall von nur einer Quelle, die Kohärenz geht gegen eins. $c_{12}(n) \approx 1$.
2. Zwei Quellen spielen gleichzeitig, die Ergebnisse stehen für keine der beiden Richtungen und schwanken stark. $c_{12}(n) < 1$

Deshalb führt man die sogenannte *cue selection* ein. Die Ergebnisse von ITD und ILD haben nur Gültigkeit wenn ein gewisser Schwellwert der IC überschritten wird $c_{12}(n) > c_0$, $\Delta L(n), \tau(n) | c_{12}(n) > c_0$. Da bei der ersten Wellenfront $c_{12}(n) \approx 1$ und bei Reflexionen $c_{12}(n) < 1$ können die dadurch entstehenden Ungenauigkeiten ausgeschlossen werden. Die Kohärenz fungiert also als Indikator für ITD und ILD in Bezug auf die gewünschten *free field cues*.

Wählt man nun für jedes Frequenzband eine geeignete Schwelle c_0 , so kann man die Ergebnisse einfach selektieren. Dies stellt jedoch ein Problem dar, da $c_{12}(n)$ vom Signal selbst, der Richtung und vom Frequenzband abhängt. Je höher man die Schwelle c_0 legt, desto näher liegen die Ergebnisse an den *free field cues*. Auf der anderen Seite gehen bei einer hohen Schwelle c_0 auch Informationen verloren, im ungünstigsten Fall liegen alle Werte unter der Schwelle, was den vollkommenen Informationsverlust bedeuten würde.

Beim Menschen wird von einer sich langsam anpassenden Schwelle ausgegangen, ein adaptive c_0 wäre dementsprechend wünschenswert. Aufgrund des beträchtlichen Aufwands, dass das Erarbeiten und Implementieren eines adaptiven Schwellwertes mit sich bringen würde, haben C.Faller und J.Merimaa im Zuge ihrer Arbeit darauf verzichtet.

2.2 Testergebnisse

Fokus des herangezogenen Artikels liegt auf dem Testen der *higher model stages*, also auf der Auswahl der Daten. Getestet wurde mit verschiedenen Signale und verschiedenen Randbedingungen. Es wurden Signale verwendet, bei denen die Testergebnisse an Versuchspersonen durch psychophysikalischen bzw. psychoakustischen Studien bereits bekannt sind.

Die Ergebnisse des Modells wurden sowohl ohne als auch mit *cue selection* verglichen. Wichtige Maße bei der Beurteilung, unter Berücksichtigung der *free field cues* τ_ϕ und ΔL_ϕ sind die systematische Abweichung und die Standartabweichung:

$$b_r = |E\{\tau(n)\} - \tau_\phi| \qquad \text{systematische Abweichung}$$

$$b_{(\Delta L)} = |E\{\Delta L(n)\} - \Delta L_\phi|$$

$$\sigma_{\tau} = \sqrt{E\{(\tau(n) - E\{\tau(n)\})^2\}}$$ Standartabweichung

$$\sigma_{(\Delta L)} = \sqrt{E\{(\Delta L(n) - E\{\Delta L(n)\})^2\}}$$

Es ist hinzuzufügen, dass nur einzelnen Bänder betrachtet werden. Der Zusammenhang der Bänder wird nicht weiter untersucht. Ebenso wurde der Schwellwert c_0 individuell gewählt, um die erhaltenen Ergebnisse bestmöglich sichtbar zu machen. Dargestellt werden die Ergebnisse durch Wahrscheilichkeitsdichtefunktionen der ITL und ILD, sogenannten PDFs (*PropabilityDensityFunction*) oder durch Histogramme von ITD und ILD mit bzw. ohne *cue selction*. Folgende Situationen und Signale wurden getestet und mit den vorliegenden Ergebnissen aus den obig erwähnten Studien verglichen:

1. Ein Sprachsignal wird nacheinander aus drei oder fünf unabhängige Quellen unter Freifeldbedingungen gespielt. Zu erkennen ist, dass sich die optimale c_0 mit steigender Frequenz nach oben verschiebt. Es zeigt sich, dass die Ergebnisse mit *cue selction* weit weniger verschwommen sind als ohne.

2. *Click-train and noise*: Gaußsches Rauschen von 0° kommend fungiert als Ma skierer, während von einer anderen Positionen aus fünf Impulse pro Sekunde gespielt werden. Dieses Szenario wird für verschiedene Winkel und verschiedene Lautstärkenverhältnisse von Impuls zu Maskierer -3, -9 und -21 dB durchgeführt. Der Versuch wurde von Good und Gilkey 1996 mit Versuchspersonen durchgeführt und eignet sich daher gut zum Vergleich. Das Ergebnis ist stark frequenzabhängig, für 2kHz stimmen Ergebnisse weitgehend mit Hörversuchen von Good und Gilkey überein, nur bei -3 und -9 dB ist eine Lokalisation möglich, ansonsten nicht. Auch hier sind die Ergebnisse mit *cue selection* gut. Die These von Brasch (2003), dass ILDs weniger verlässlich sind als ITDs, wenn ein Maskierer verwendet wird, kann ebenfalls im Modell bestätigt werden. Die genaueren Ergebnisse liefert die ITD.

3. Der Präzedenzeffekt oder Haas-Effekt kann mittels zweier aufeinander folgender Impulse getestet werden. Bei einem Abstand der beiden gleich lauten Impulse von 1 – ca. 10 ms sollte die Richtungsinformation der des ersten Impulses entsprechen. Unter 1 ms wird die Richtungsinformation zusammengefasst, ab ca. 10 – 20 ms kann man zwei unabhängige Quellen wahrnehmen. Die Tests am Modell haben eine weitgehende Übereinstimmung mit einer Studie Litovskys ergeben, allerdings nur wenn c_0 optimal gewählt wurde, dass heißt wenn c_0 für verschiedene Abstände der Impulse modifiziert wurde.

4. Ein 500 Hz Sinuston mit einer Anstiegsrate von 0, 5 und 50 ms wird aus einem Winkel von 30° wiedergegeben. Rakerd und Hartman haben 1986 herausgefunden, dass bei einer hinzukommenden Reflexion, weder der Winkel der ursprünglichen Quelle noch der der Reflexion korrekt lokalisiert werden kann. Die Differenz hängt maßgeblich von der Anstiegsrate ab. Bei verhältnismäßig klein gewählter c_0 stimmen die Abweichungen des Modells für 0 ms und 5 ms mit denen der Studie überein. Für eine langsamere Anstiegsrate stimmen die Ergebnisse nicht mehr mit denen des Experiments überein.

5. Ein Sprachsignal mit Azimuth 30° bzw. zwei Sprachsi gnale des selben Sprechers mit Azimuth +/-30° werden gleichzeitig in einem Raum mit einer Nachhallzeit von 1.4 s und 2.0 wiedergegeben. Die Lokalisation unter diesen Bedingungen ist sehr schwierig. Ohne *cue selection* ist durch die Fülle an Reflexionen nur ein geringer bis kein Informationsgewinn über die Richtung der Quelle möglich. Die einzelnen Frequenzbänder verhalten sich außerdem sehr verschieden. Mit *cue selection* ist in bestimmten kritischen Bändern eine genau Lokalisation möglich. Auch die Unterscheidung der zwei Quellen ist möglich.

2.3 Fazit

Christof Faller und Juha Merimaa (2004) befassen sich in ihrer Arbeit umfassend mit dem Testen des obig beschrieben Modells. Mittels *cue selection*, basierend auf interauraler Kohärenz ist es für eine Vielzahl teilweise komplexer Situation möglich auf korrekte ITDs und ILDs zu schließen. Würde man von ITDs und ILDs also auf die entsprechenden Winkel schließen, so entspricht das Modell zu großen Teilen der menschlichen Lokalisation.

Ein großes Problem an den präsentierten Testergebnissen stellt der händisch modifizierte Schwellwert c_0 dar. Ein fixer c_0 für jedes kritische Band würde nicht für alle Versuche solch gute Ergebnisse liefern. Es kann gezeigt werden, dass weitgehend korrekte ITDs und ILDs prinzipiell möglich sind, die Entwicklung eines adaptiven Schwellwertes ist jedoch Thema für eine weitere Arbeit.

Faller und Merimaa beschränken sich auf die Berechnung der ITDs und ILDs. Zur vollständigen Lokalisation, also zur Angabe eines Winkels bzw. einer Wahrscheinlichkeitsdichtefunktion der Winkel fehlen mehrere Verarbeitungsschritte. ITDs und ILDs liegen für alle kritischen Bänder vor. Zur korrekten Bewertung ist eine Berücksichtigung der Lautheit in den einzelnen Bändern erforderlich. ITDs und ILDs haben außerdem in verschiedenen Bändern unterschiedlichen Einfluss auf die Lokalisation. So trägt im tieffrequenten Bereich die ILD kaum oder gar nicht zur Lokalisation bei, während sie im Hochfrequenten den wesentlichen Anteil ausmacht. ITDs und ILDs sollten also auch in dieser Hinsicht bewertet werden. Des weiteren kann man den Zusammenhang der einzelnen Bänder untersuchen.

Nach all diesen Betrachtungen könnten ILDs und ITDs zusammengefasst, und in einer Tabelle die entsprechenden Winkel ausgelesen werden.

3 Implementierung, Modifikation und Erweiterung des Modells

3.1 Realisierung

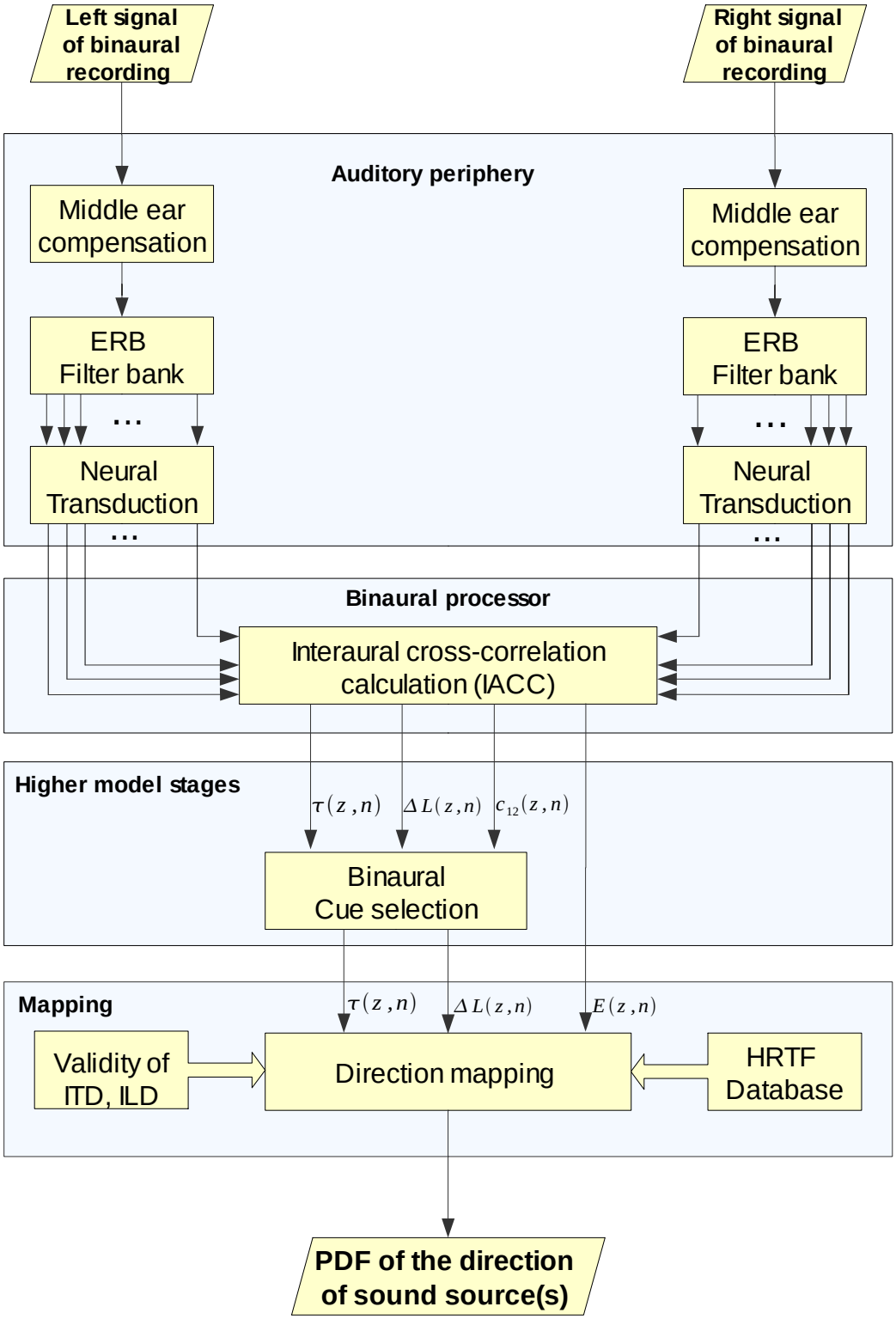

Abb. 1 Strukturdiagramm gesamt

Das implementierte Modell lässt sich in vier Teile unterteilen. *Auditory peripherie*, *binaural processor* und *higher modell stages* basieren auf dem Modell von von Faller und Merimaa. Es wurde lediglich ein Mittelohrfilter hinzugefügt. Im vierten Abschnitt, dem *mapping* wird die Datenmenge stark reduziert und mit einem Referenzdatensatz verglichen. Das letztendliche Ergebnis stellt die Auftrittswahrscheinlichkeit der Schallquelle in Abhängigkeit des Winkels dar. Eine einfache und leicht verständliche graphische Darstellung und Bewertung der Ergebnisse ist somit möglich.

Als Eingangssignal wurde HRTF gefiltertes Rauschen verwendet. Der verwendete Datensatz HRTF´s stammt von der Firma Kemar und ist im Internet frei verfügbar. Betrachtet wurde die Ebene für eine Elevation von Null Grad. Für den Azimuth lagen 72 HRTF – Paare vor, was einer Auflösung von fünf Grad entspricht. Die *Samplerate* beträgt 44100 Hz, die Impulsantworten wurden mit 512 samples gespeichert. Durch einfaches Falten eines Eingangssignals mit den dementsprechenden HRTF´s kann somit ein binaurales Signal erzeugt werden.

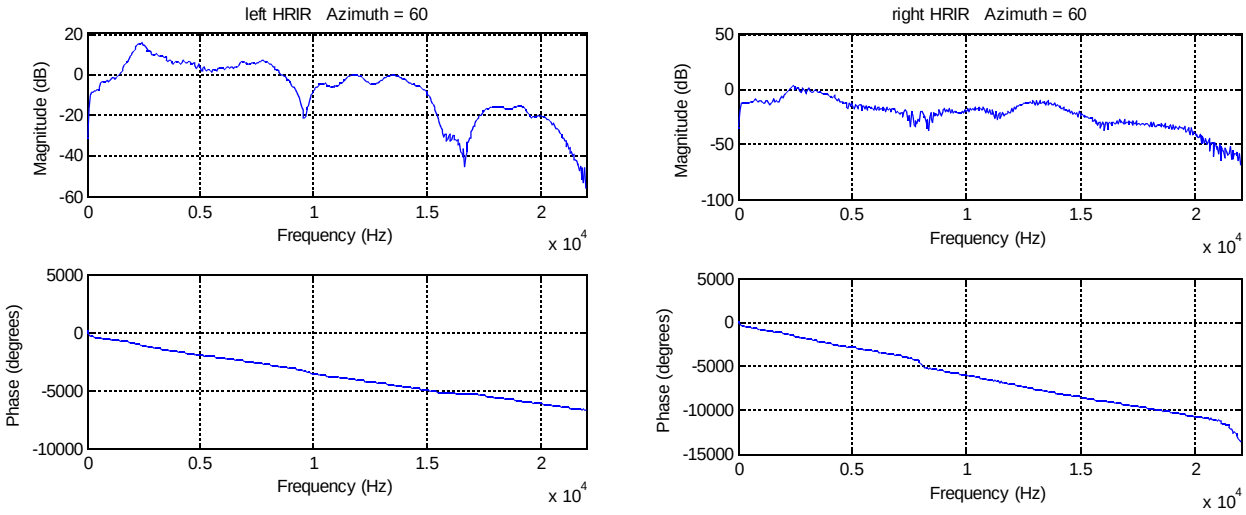

Abb. 2 linker und rechter Filter für 60° (Winkel ent gegen Uhrzeigersinn)

3.1.1 Auditory periphery

Wie im Strukturdiagramm ersichtlich, besteht die *Auditory periphery* aus drei Abschnitten, dem Mittelohrfilter, der ERB – Filterbank und der neuronalen Übertragung . Der Mittelohrfilter wird als lineares Filter angenommen und wurde als minimalphasiges IIR-Filter zwanzigster Ordnung implementiert. Die genauen Daten wurden der Diplomarbeit „Untersuchung von psychoakustischen Prinzipien für automatische Mixdown-Algorithmen“ von Philipp Aichinger entnommen. Funktion des Mittelohrfilters ist die Gewichtung des Eingangssignals entsprechend der menschlichen Wahrnehmung. Zusammen mit dem Außenohrfilter und der Innenohrübertragung sollte der errechnete Schalldruckpegel der menschlichen Wahrnehmung entsprechen. Dementsprechend liegt der größte Anteil der Signalenergie im Bereich von 1000 – 4000 Hz.

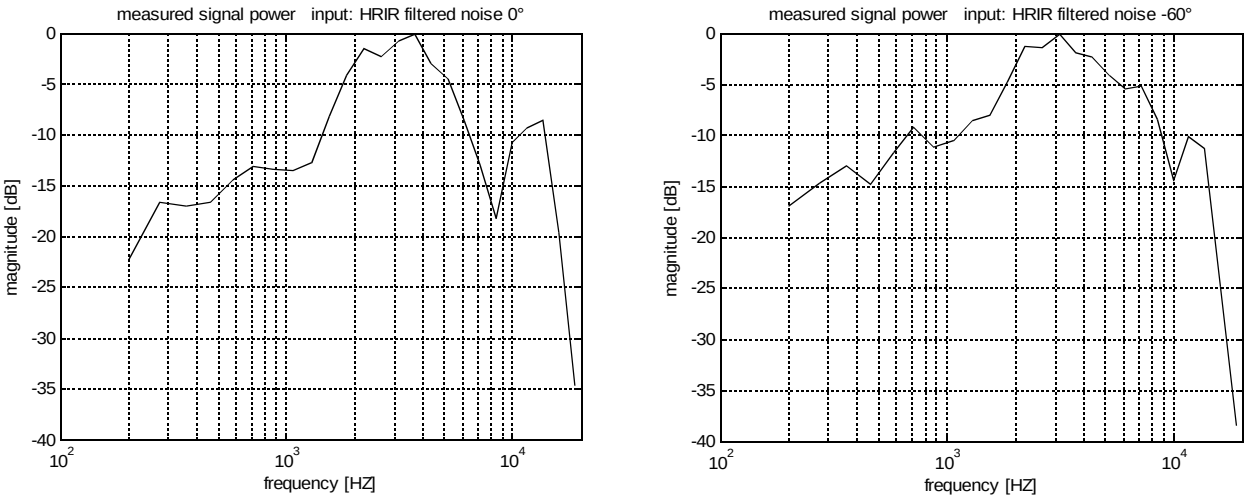

Abb. 3 Signalerengie nach auditory periphery HRTF gefiltertes weißes Rauschen bei 0°/ -60°

Da die berechnete Energie unserem tatsächlichem Empfinden in Annäherung entspricht, kann die Energie als ein Maß der Relevanz einzelner Frequenzbänder genutzt werden.

Nach dem Mittelohrfilter ist eine ERB – Filterbank implementiert. ERB steht als Abkürzung für *equivalent rectangular bandwidth*, also äquivalente Rechteckbandbreite. Die ERB – Filter sind so konstruiert, dass die Energiemenge eines jeden Filters bei Anregung mit weißem Rauschen gleich ist. Die ERB – Skala ist phsychoakustisch motiviert und entspricht in Annäherung der Barkskala. Implementiert wurde eine Filterbank mit 25 Filtern wobei die niedrigste Mittenfrequenz bei 200 Hz und die höchste bei 18.8 kHz liegt.

Das Ausgangssignal besteht folglich aus 25 Bandpasssignalen, welche in der nächste Stufe parallel bearbeitet werden.

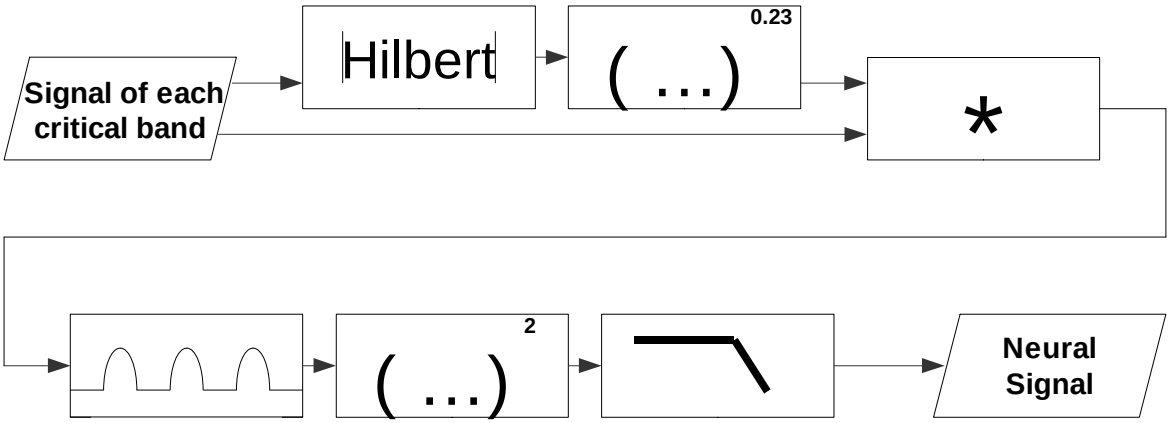

Abb. 4 Signalflussgraph der neuronalen Übertragung

Den dritten Abschnitt der *autitory periphery* bildet das Modell zur Simulation der neuronalen Übertragung. Für eine genaue Beschreibung siehe Abschnitt 2.1.1. Zur Implementierung habe ich die frei verfügbare *Binaural auditory processing toolbox for MATLAB* von Akeroyd verwendet. Sie beinhaltet sowohl die ERB-Filterbank als auch die neuronale Übertragung.

Als Ausgang der a*utitory periphery* erhält man für links und rechts jeweils 25 Signale im Zeitbereich, die der neuronalen Übertragung im jeweiligem Frequenzband entsprechen.

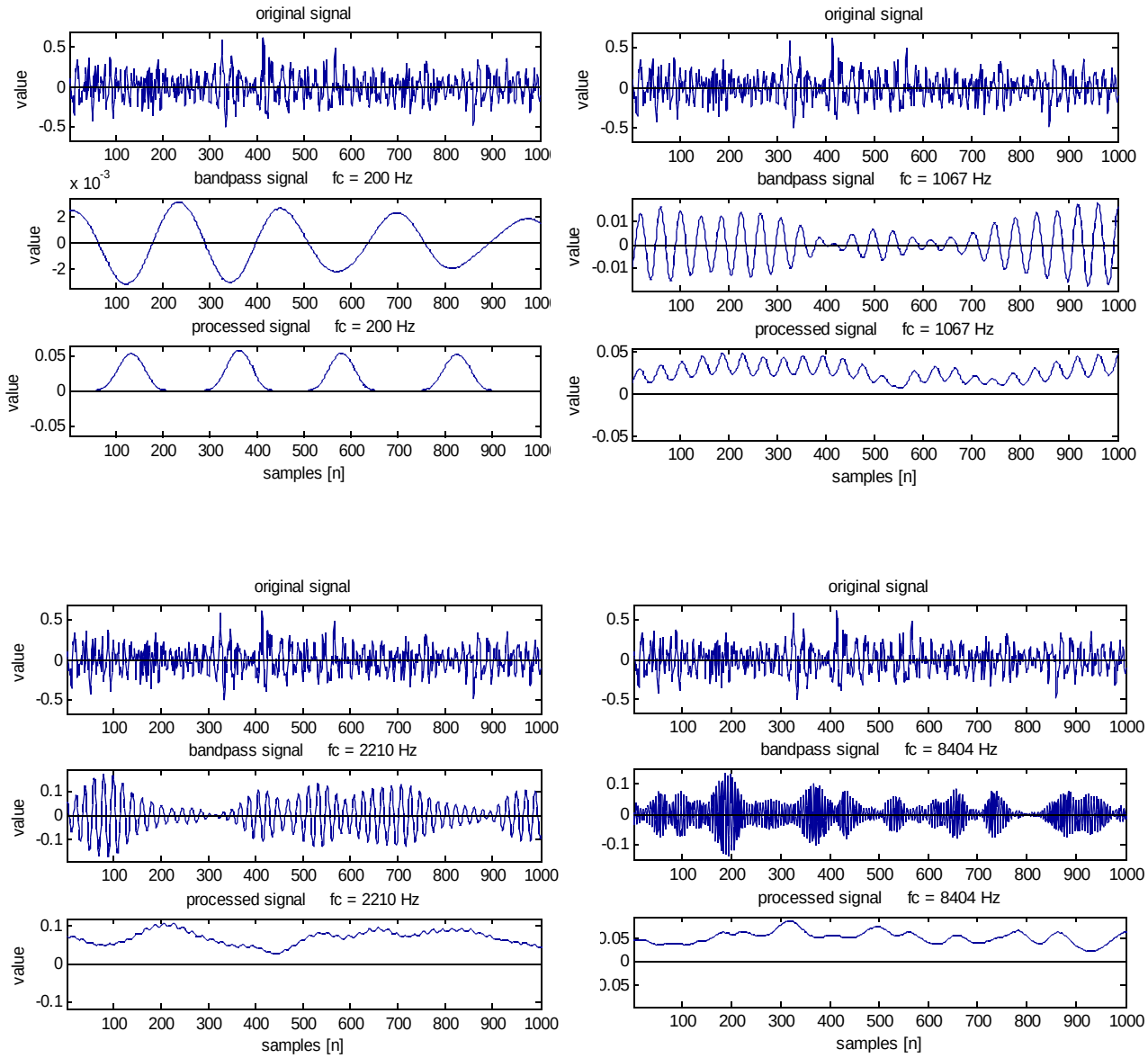

Abb. 5 neuronale Übertragung bei 200 Hz, 1967 Hz, 2210 Hz und 8404Hz

Ohne die Simulation der neuronalen Übertragung wäre eine Korrelationsanalyse, wie sie später durchgeführt wird, nicht möglich. Vergleicht man die Bandpass – Signale höherer Frequenzen mit den bearbeiteten Signalen, so ist dies leicht zu erkennen.

Der verwendete Tiefpassfilter ist 4. Ordnung und hat eine Grenzfrequenz von 420 Hz. Deshalb unterscheidet sich die Signalform des bearbeiteten Signals gerade in hohen Bändern stark vom Eingangssignal. Es ist zu erkennen, dass es die Energie und der Verlauf der Einhüllenden im Signal erhalten bleibt.

3.1.2 Binaural Processor

Der *binaural processor* berechnet die Zeitdifferenz ITD $\tau(z,n)$, Leveldifferenz ILD $\Delta L(z,n)$, Kohärenz IC $c_{12}(z,n)$ und die Signalenergie $E(z,n)$. Die genaue Beschreibung der Berechnung ist im Abschnitt 2.1.2 zu finden. Das exponentiell abklingende Fenster wurde mit 10ms festgelegt.

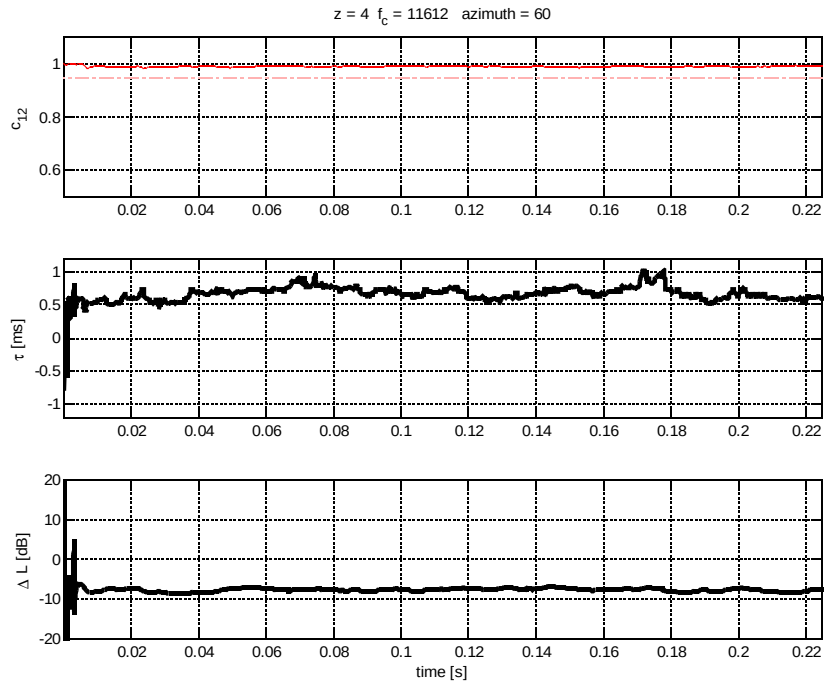

Abb. 6

$c_{12}(n)$, $\tau(n)$ und $\Delta L(n)$

$c_0 = 0.95$

Eingangssignal:

HRTF – gefiltertes Rauschen azimuth 60° f = 11.612Hz

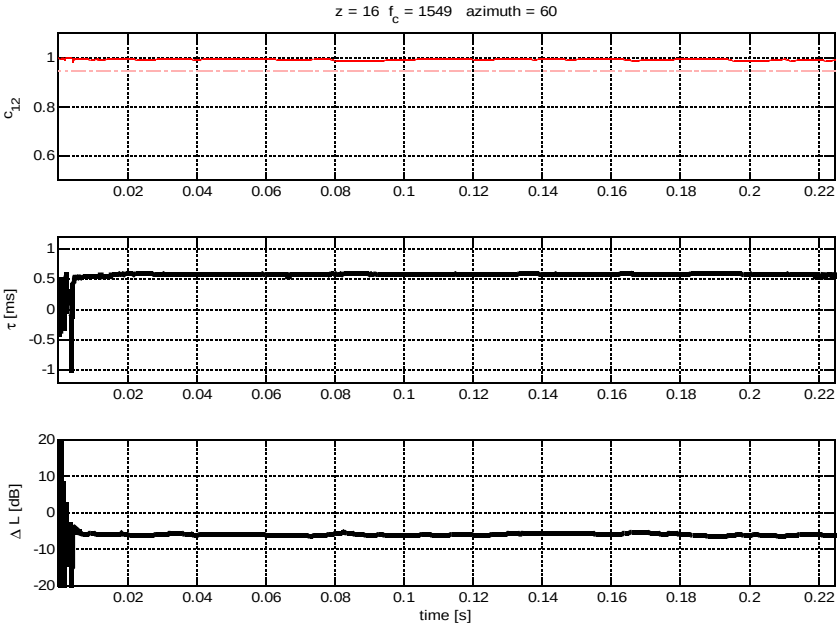

Abb. 7

$c_{12}(n)$, $\tau(n)$ und $\Delta L(n)$

$c_0 = 0.95$

Eingangssignal:

HRTF – gefiltertes Rauschen azimuth 60° f = 1.549Hz

3.1.3 Higher Model Stages

Die *higher model stages* dienen zum Ausschließen ungültiger Signale. Wie in Abschnitt 2.1.3 ausführlich erläutert, wird hier eine sogenannte *cue selection* eingeführt. Die Ergebnisse von ITD und ILD haben nur Gültigkeit wenn eine gewisser Schwellwert der IC überschritten wird $c_{12}(n) > c_0$, $\Delta L(n), \tau(n) | c_{12}(n) > c_0$. Nicht korrekte $\tau(z,n)$ und $\Delta L(z,n)$ werden im Programm NaN gesetzt und später einfach gelöscht.

Ein Beispiel: Die berechnete $\tau(z,n)$ bei einem Eingangssignal für ein Band z mit einer Länge von 10.000 samples besitzt aufgrund der gewählten Schwelle $c_0 = 9.5$ 8.000 gültige samples oder bei $c_0 = 9.0$ 9.700 gültige samples. Bei einer späteren Zusammenfassung der Ergebnisse über alle n werden nur die gültigen berücksichtigt.

Der gewählte Schwellwert im Programm beträgt für alle z $c_0 = 9.5$. Eine frequenzabhängiges Festlegen der Schwelle wäre ohne weiteres möglich und könnte bei sinnvoller Wahl die Testergebnisse maßgeblich verbessern. Da $c_{12}(n)$ von dem Signal selbst, der Richtung und vom Frequenzband abhängt, wurde im Zuge dieser Arbeit auf eine frequenzselektive Wahl von c_0 verzichtet.

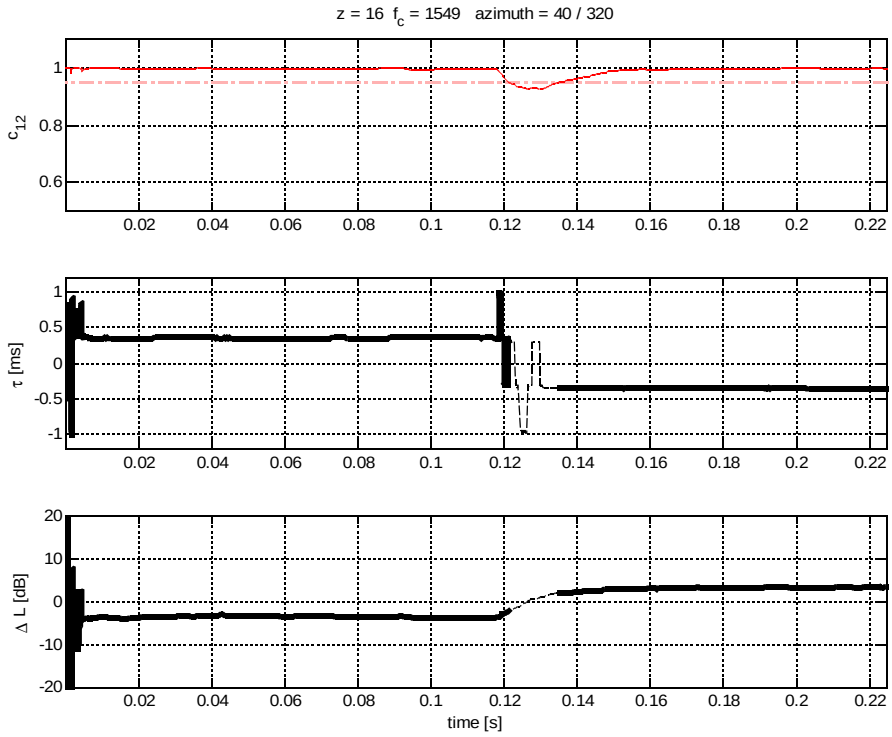

Abb. 8

$c_{12}(n)$, $\tau(n)$ und $\Delta L(n)$

$c_0 = 0.95$

Eingangssignal:

HRTF – gefiltertes Rauschen mit wechselndem azimuth 40° / 320°

f = 1.549Hz

3.2 Mapping

3.2.1 Funktionsweise

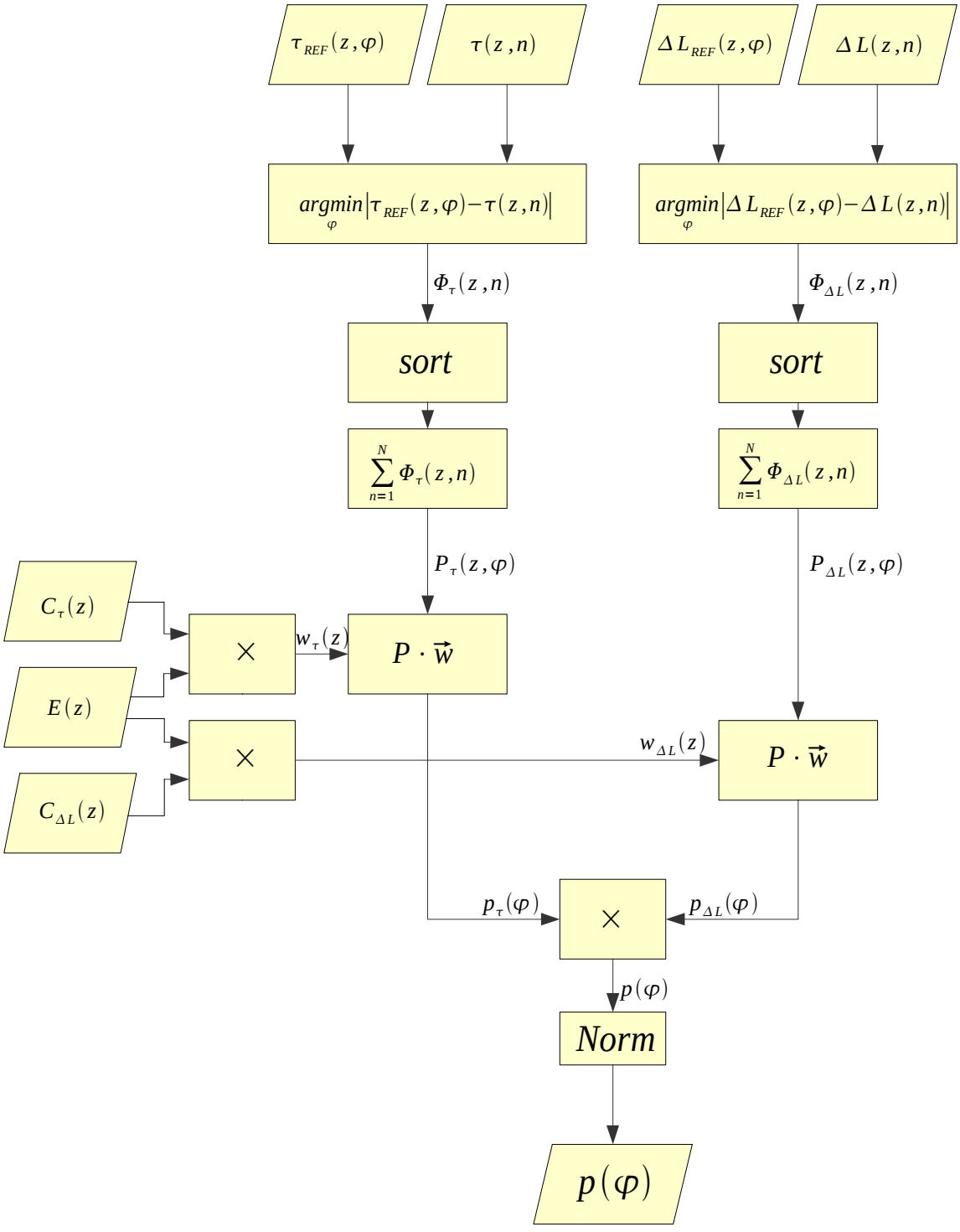

Abb: 9 Strukturdiagramm Mapping

Das *mapping* ist eine Abbildungsvorschrift, wobei die die Auftrittswahrscheinlichkeit der menschlichen Lokalisation in Abhängigkeit vom Winkel anhand der zuvor berechneten Zeit – und Pegeldifferenzen bestimmt wird.

Die Eingangsdaten werden in mehreren Schritten auf einen Ausgangsvektor mit 72 Einträgen reduziert. Dieser stellt die Wahrscheinlichkeit einer Schallquelle in Abhängigkeit des Azimuth – Winkels dar, wobei die Auflösung 5° beträgt. Die eigentlichen Eingangsdaten $\tau(z,n)$ und $\Delta L(z,n)$, werden mit einem Referenzdatensatz $\tau_{REF}(z,\varphi)$ und $\Delta L_{REF}(z,\varphi)$ verglichen. Für jeden Zeitpunkt n wird im entsprechendem Frequenzband sowohl für τ als auch für ΔL der Referenzwinkel ausgewählt, der die geringste Abweichung aufweist.

Die resultierenden Matrizen $\Phi_{\tau}(z,n)$ und $\Phi_{\Delta L}(z,n)$ beinhalten somit den zugeordneten Winkel im jeweiligem Band z zu jedem Zeitpunkt n. Da eine Auftrittswahrscheinlichkeit in Abhängigkeit vom Winkel dargestellt werden soll, werden die Daten nach Winkeln sortiert und über alle Zeitpunkte aufsummiert.

In einem einzelnem Band wird folgendes durchgeführt: $\Phi_{\tau}(n) \rightarrow p_{\tau}(\varphi)$

Aus einem Vektor der Signallänge N, indem zu jeweiligem Zeitpunkt der zugeordnete Winkel gespeichert ist wird ein Vektor mit 72 Einträgen (einen für jeden möglichen Winkel) indem die Anzahl der zugeordneten Zeitpunkte steht. Diese Zahl stellt stellt also die Häufigkeit einer Detektion eines jeden Winkels dar. Dieser Verarbeitungsschritt wird in jedem Band ausgeführt, sodass sich die Matrizen $P_{\tau}(z,\varphi)$ und $P_{\Delta L}(z,\varphi)$ ergeben.

Um eine allgemeine Aussage zu treffen, fehlt lediglich das Zusammenfassen der einzelnen Frequenzbänder. Um die Bänder sinnvoll zu gewichten werden drei Eingangsvariablen herangezogen. Zum einen wird die Relevanz von τ und ΔL allgemeinen betrachtet. Die Lokalisation mittels Laufzeitdifferenzen τ findet im tieffrequenzen Bereich statt, während die Lokalisation mittels Pegeldifferenzen ΔL im höherfrequenten Bereich zum tragen kommt. $C_{\tau}(z)$ und $C_{\Delta L}(z)$ sind Funktionen, die frequenzabhängige Faktoren zur Gewichtung der Bänder zur Verfügung stellen. $C_{\tau}(z)$ und $C_{\Delta L}(z)$ ergeben in Summen immer eins. Übergangsfrequenz und Breite des Übergans sind frei wählbar, wobei ein Übergang zwischen 500 und 1000 Hz sinnvoll ist. Zusätzlich wird die Energie $E(z)$ betrachtet. Wie in Punkt 3.1.1 beschreiben stellt diese ein Maß für empfundene Lautstärke des Eingangssignals dar. Zusammen mit $C_{\tau}(z)$ und $C_{\Delta L}(z)$ ergeben sich die beiden Vektoren $w_{\tau}(z)$ und $w_{\Delta L}(z)$ zur Gewichtung von τ und ΔL .

Die ach der Gewichtung erhaltenen Vektoren $p_{\tau}(\varphi)$ und $p_{\Delta L}(\varphi)$ müssen nun nur noch zusammengefasst und normiert werden. Das Zusammenfassen kann sowohl additiv als auch multiplikativ geschehen, wobei sich in der Praxis gezeigt hat das die multiplikative Verknüpfung in den meisten Fällen die bessere Wahl ist, da nur bei Übereinstimmung der Ergebnisse von τ und ΔL Werte > 0 zu erwarten sind.

Des weiteren gibt es die Möglichkeit das Übersprechen einzelner Bänder zu simulieren. Es kann im Programm separat für τ und ΔL aktiviert werden. In diesem Fall ist das Übersprechen der Referenzdaten ebenso implementiert. Da die Pegeldifferenz ΔL bei hohen Frequenzen starke Schwankungen aufweist gibt es die Möglichkeit alle Bänder ab einer bestimmten Frequenz zusammenzufassen. Die Frequenz ist frei wählbar, wobei Werte zwischen 10 und 15 kHz Sinn machen.

3.2.2 Referenzdaten

Die Referenzdatenbank beinhaltet $\tau_{REF}(z,\varphi)$ und $\Delta L_{REF}(z,\varphi)$, das heißt einen Referenzwert τ und ΔL für jeden möglichen Winkel in jedem Frequenzband. Gewonnen werden die Werte, indem bekannte Testsignale vom Algorithmus analysiert werden. Als Eingangssignal wurde ein mit den Kemar HRTFs gefiltertes weißes Rauschen benutzt, die Länge betrug 10.000 samples. Für jeden mögliche Winkel wurde dass entsprechende Paar HRTF´s genommen, mit dem Rauschen gefaltet und der Algorithmus wurde durchlaufen. Der Algorithmus gab nun für jedes der 10.000 samples ein errechnetes τ und ein ΔL aus. Es wurde nun über alle Zeitpunkte gemittelt wobei die ersten 1000 samples zur Berücksichtigung des Einschwingvorgangs nicht in die Rechnung mit einbezogen wurden. Die Werte die sich für τ ergaben wurden auf die ganzzahlige samples gerundet.

Für τ ergaben sich vor der Mittlung über alle Zeitpunkte folgende Ergebnisse:

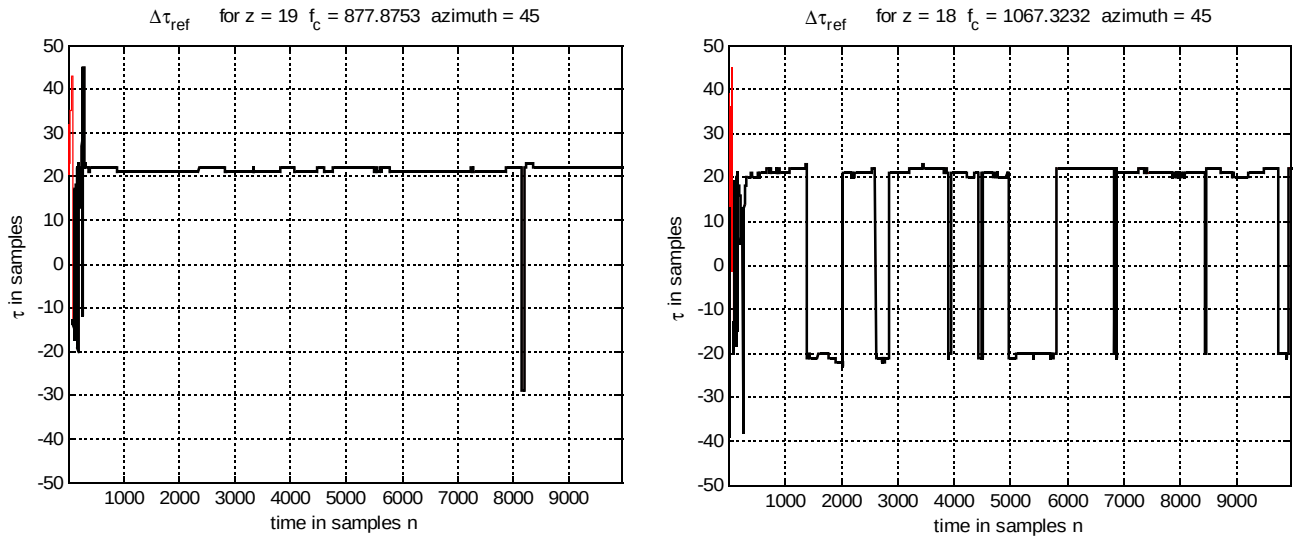

Abb. 10 errechnete Referenzdaten für $\tau(n)$ Daten für welche gilt: $c_{12}(n) < c_0$
werden bei der Mittlung nicht beachtet und sind rot gekennzeichnet

Die obigen Bilder wurden bewusst ausgewählt um ein bestimmtes Problem zu verdeutlichen. Bei einem Winkel von 45° sollte die Verzögerung etwa 0,5 ms b zw. 22 samples betragen. Die korrekte Verzögerung wird sowohl bei 877 Hz als auch bei 1067 Hz erkannt, es ist jedoch zu beobachten dass τ um ca. 40 samples abweichen kann. Diese Doppeldeutigkeit kommt dadurch zustande dass, eine Verzögerung von 1ms in etwa der Laufzeit einer Wellenlänge bei 1067 Hz entspricht.

$$\frac{1}{1067\,Hz} = 0.93\text{ms}$$

Es ergibt sich der Fall, dass das Signal am einem Ohr um 180° phasengedreht eintrifft und der Algorith mus nicht entscheiden kann aus welcher Richtung das Signal als erstes kam. Diese Doppeldeutigkeit kommt jedoch nicht vom Algorithmus selbst sondern ist durch die Physiologie des Menschen vorgegeben. Das Richtungshören beim Menschen kann nur bis zu einer bestimmten Grenze mittels Laufzeitverzögerung funktionieren. Das Programm kann den korrekten Laufzeitunterschied bis zum Frequenzband von 716 Hz weitgehend fehlerfrei detektieren.

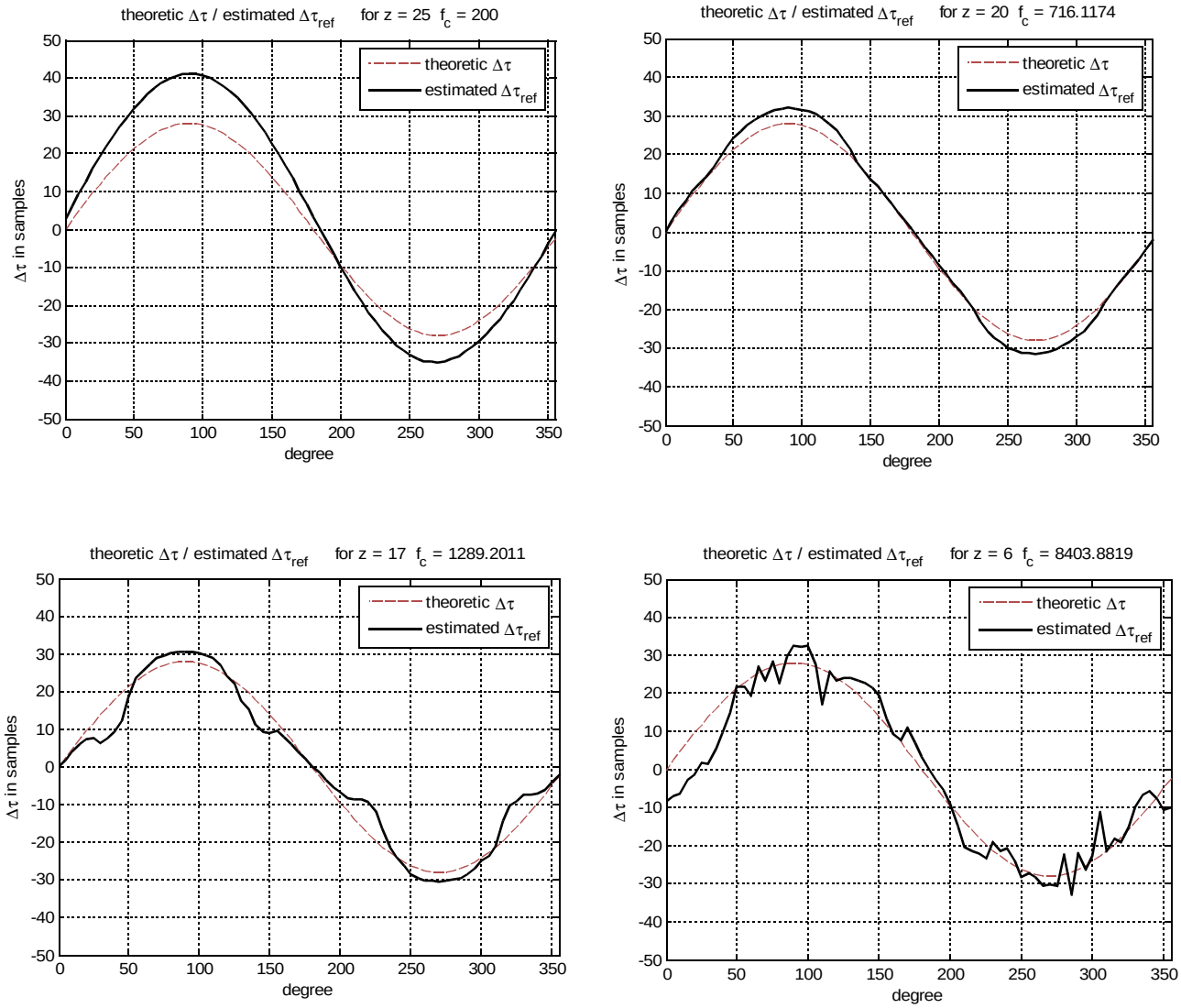

Abb. 11 berechnetes τ_{ref} Darstellung über alle Winkel bei 200 Hz, 716 Hz, 1289 Hz und 8403 Hz

Die gemittelten Werte über alle Zeitpunkte, dargestellt über alle Winkel, entsprechen bis zur Grenze von 716 Hz in Annäherung den theoretisch errechneten Werten. Problematisch ist, dass die errechnete Kurve bei tieferen Frequenzen von den analytisch bestimmten Daten abweicht. Diese Abweichung nimmt mit sinkender Frequenz zu. Die Laufzeitdifferenz τ ist im tieffrequenten Bereich zu groß. Grund dafür sind die verwendeten HRTFs. In Abhängigkeit von Winkel und Frequenz kann die Laufzeitdifferenz τ bis zu einem Faktor von zwei zu groß sein.

Im Extremfall, bei 100 Hz kommt es im Bereich von Winkeln um die 90° zu einer Detektion von $\tau = 45\,samples$. Das Programm kann Laufzeitunterschiede $\tau > 45\,samples$ bzw. $\tau > 1\,ms$ nicht erkennen, da durch den Kopf des Menschen die Laufzeitdifferenz auf 1ms begrenzt ist. Es wird somit für mehrere Winkel ein Maximum von 45 samples detektiert. Die Kurve wird also, ähnlich wie bei einem *Limiter*, abgeschnitten. Dies ist auch der Grund, weshalb eine minimale Mittenfrequenz von 200Hz gewählt wurde.

Dieses fehlerhafte Verhalten ließ sich sowohl bei den Kemar HRTFs als auch bei den HRTFs vom IEM beobachten. Für die Berechnung von τ bzw. für die Zeitverzögerung von binauralen Signalen ist die Phase des Filters ausschlaggebend. Wie in Abb. 2 ersichtlich ist die Phase der Filter nur annähernd linear, die Gruppenlaufzeit also nicht konstant. Die Verzögerung kann im tieffrequenten Bereich nicht genau genug durch den Filter abgebildet werden. Da die Genauigkeit im Bereich < 0.1 ms liegen sollte, wirken sich die resultierenden Fehler stark auf das Ergebnis aus. Ein Beispiel: Die Phase des Kemar Filtersatzes für 45° beträgt bei 86 Hz:

$$\varphi_L = -167° \qquad \varphi_R = -191°$$

Die Zeitdifferenz berechnet sich wie folgt:

$$\frac{\Delta\varphi}{360} \cdot \frac{1}{86} = 0.78\,ms \quad \text{bzw.} \quad \frac{\Delta\varphi}{360} \cdot \frac{1}{86} \cdot 44100 = 34\,samples$$

Das Ergebnisse liegt ca. 50% über dem korrekten Wert. Dieses fehlerhafte Verhalten ließ sich sowohl bei den Kemar HRTFs als auch bei den HRTFs vom IEM beobachten.Die Impulsantworten der Kemar HRTFs wurden mit 512 samples gespeichert, die des IEM mit 128 samples, beide mit einer *samplingrate* von 44.100. Bei der Verwendung der HRTFs vom IEM war der Effekt jedoch deutlich stärker ausgeprägt, was vermutlich auf die Länge der Impulsantwort zurückzuführen ist. Dies ist letztendlich auch der ausschlaggebende Grund für die Verwendung der Kemar HRTF´s in dieser Arbeit.

Die Referenzdaten für ΔL wurden aus folgenden Signalen gewonnen:

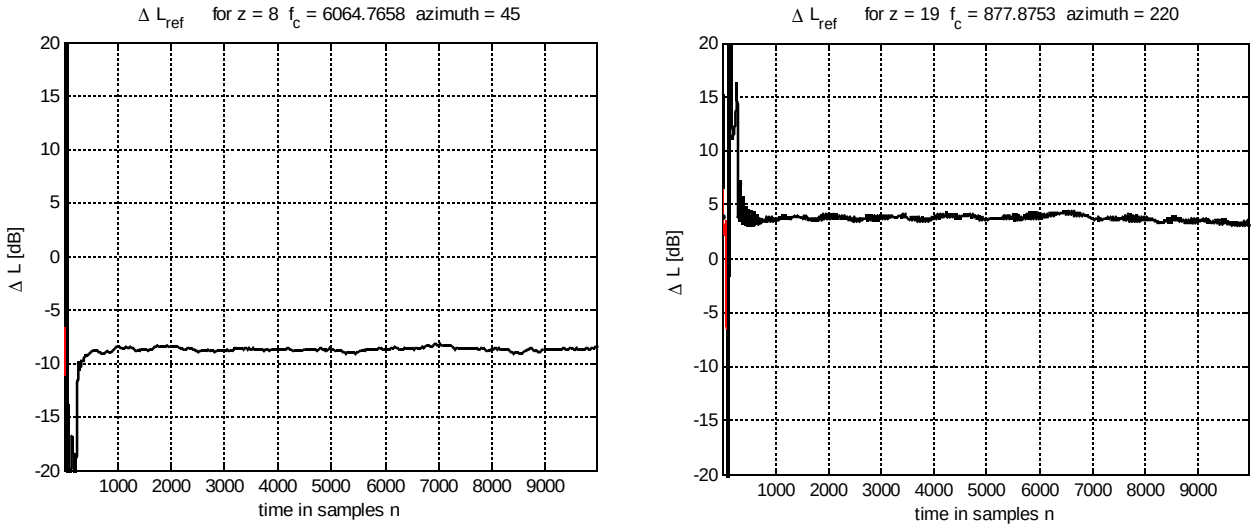

Abb. 12 errechnete Referenzdaten für $\Delta L_{REF}(n)$ Daten für welche gilt: $c_{12}(n) < c_0$
werden bei der Mittlung nicht beachtet und sind rot gekennzeichnet

Wie bereits erwähnt werden die Daten für $n > 1000$ gemittelt. Für Frequenzbänder kleiner 877 Hz wird der Fenstereffekt sichtbar. Da Pegeldifferenzen erst ab einer bestimmten Frequenz maßgeblich zur Lokalisation beitragen und deshalb unterhalb einer bestimmten Grenze auch vom Programm nicht berücksichtigt werden, stellt dies kein Problem dar.

Die über die Zeit gemittelten Referenzdaten sind in den folgenden Abbildungen dargestellt. Der realen Wahrnehmung entsprechend, treten beim Betrachten der Pegeldifferenz Mehrdeutigkeiten auf.

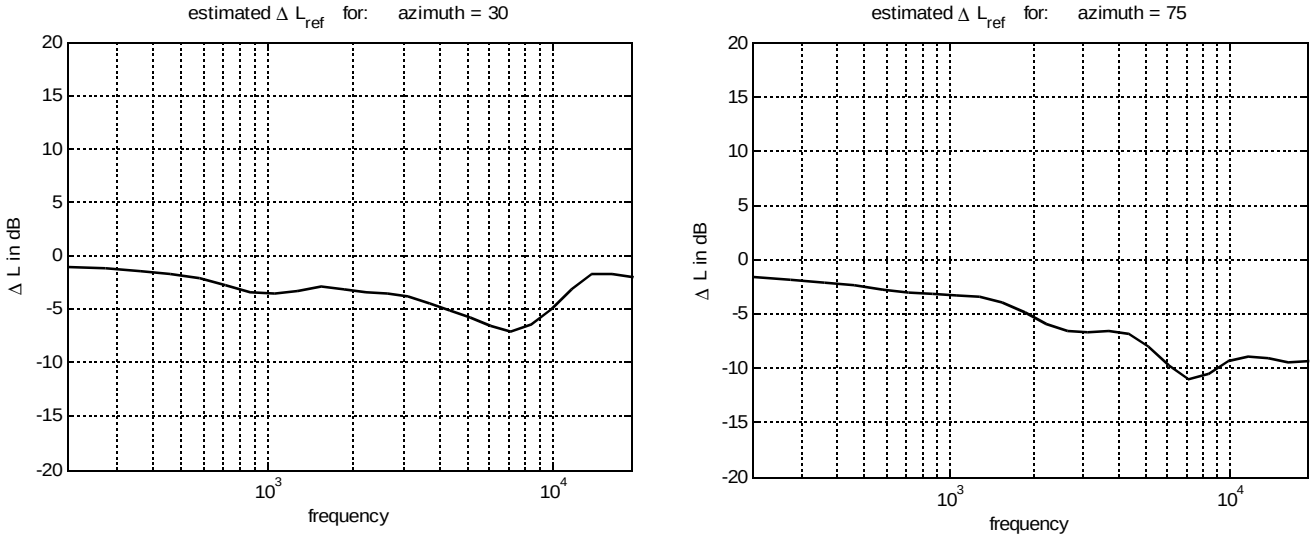

Abb. 13 berechnetes ΔL_{ref} Darstellung über alle Frequenzen bei 30° und 75°

Darstellung der Referenzdaten ΔL über alle Winkel bei einer bestimmten Frequenz.:

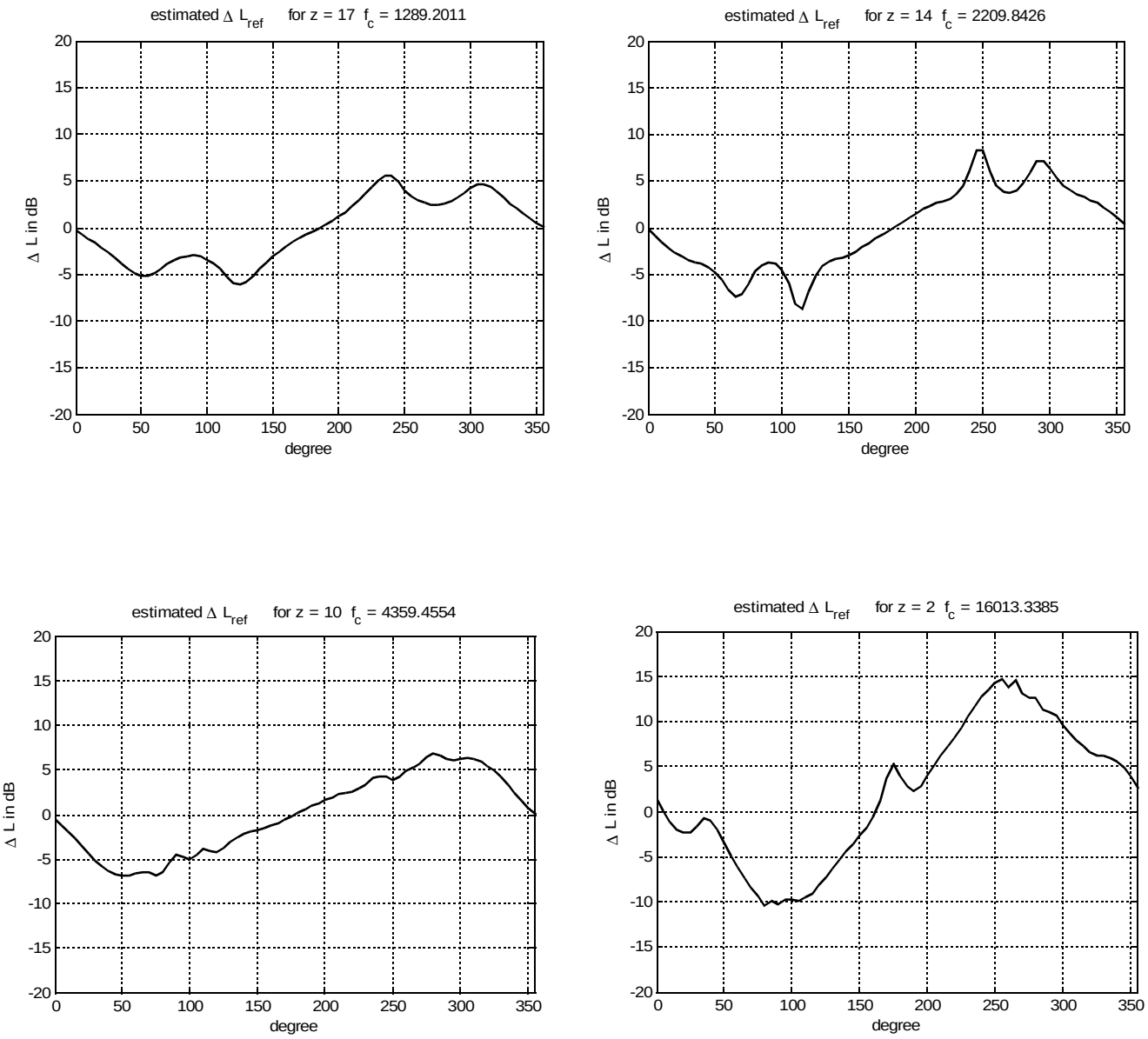

Abb. 14 berechnetes ΔL_{ref} Darstellung über alle Winkel für 1289 Hz, 2209 Hz, 4359 Hz und 16013 Hz

3.3 Graphische Darstellung der Ergebnisse

3.3.1 Ideale binaurale Signale

Ein ideales binaurales Testsignal lässt sich einfach mit der Faltung der entsprechenden HRTFs realisieren. Wie auch bei der Erstellung der Referenzdatenbank wurden die Kemar HRTFs verwendet. Dass die selben HRTFs verwendet wurden, trägt maßgeblich zu guten Testergebnisse bei. Des weiteren sind die Signale trocken und frei von jeglichen Störungen. Die Grafiken stellen die Ergebnisse an unterschiedlichen Verarbeitungsschritten während und nach dem *mapping* dar. Bei allen folgenden Graphiken steht schwarz für das Maximum und weiß für Null.

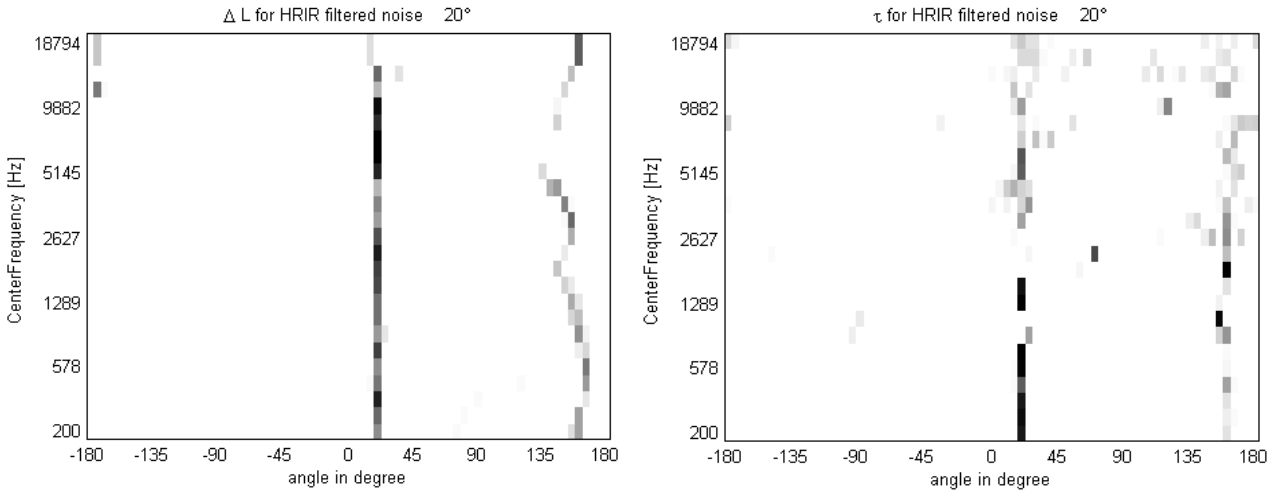

Abb. 15 Auftrittswahrscheinlichkeit von ΔL und τ für ein Eingangssignal von 20°

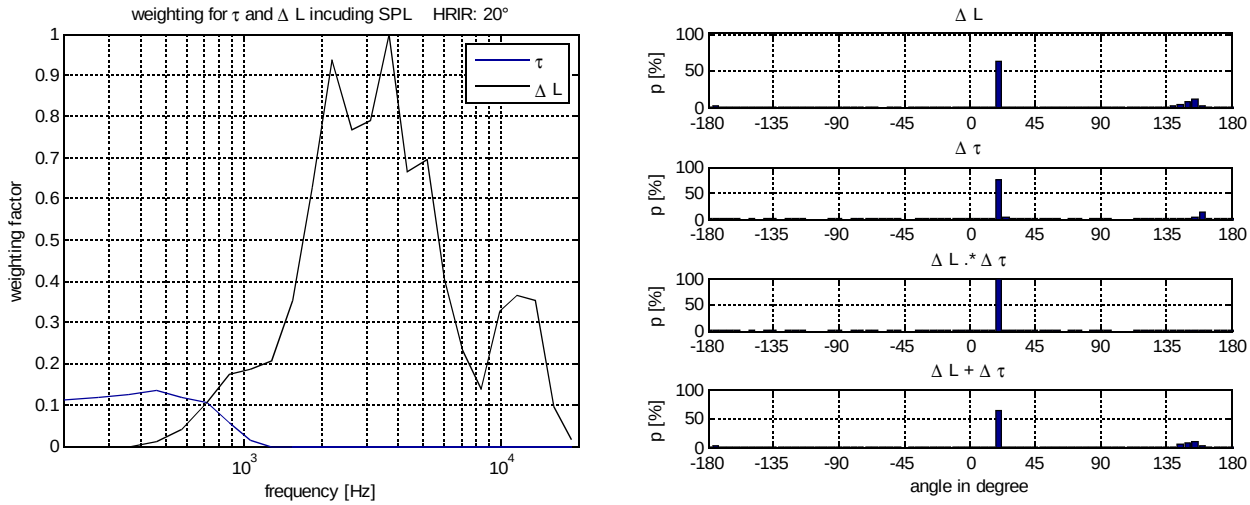

Abb. 16 Gewichtungsfunktion für ΔL und τ für ein Eingangssignal von 20° und Histogramme für ein Eingagngssignal von 20°

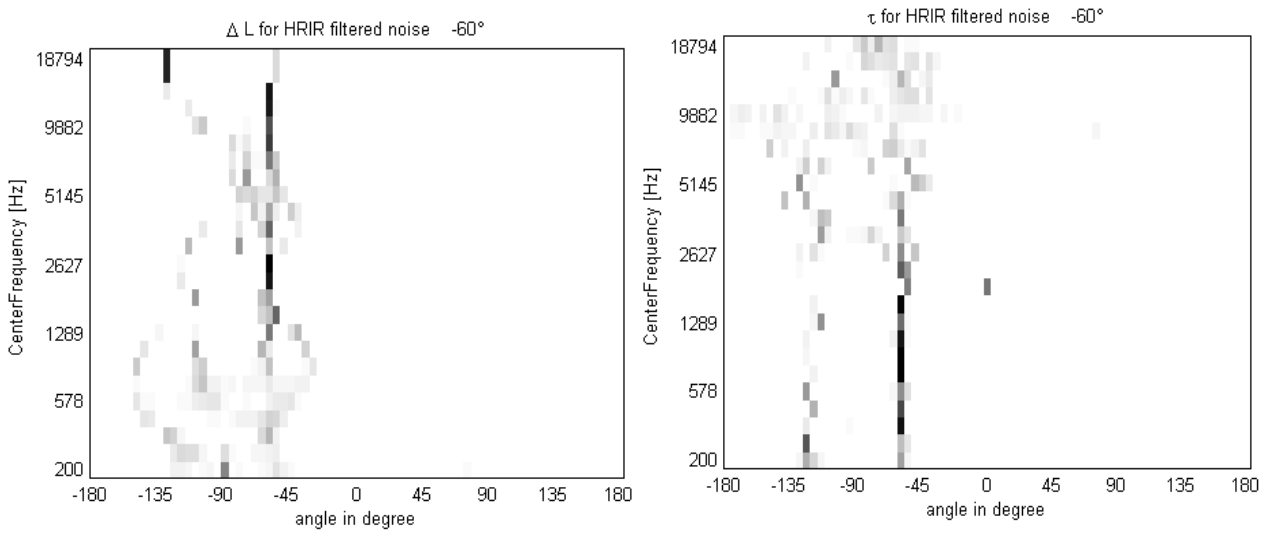

Abb. 17 Auftrittswahrscheinlichkeit von ΔL und τ für ein Eingangssignal von 60°

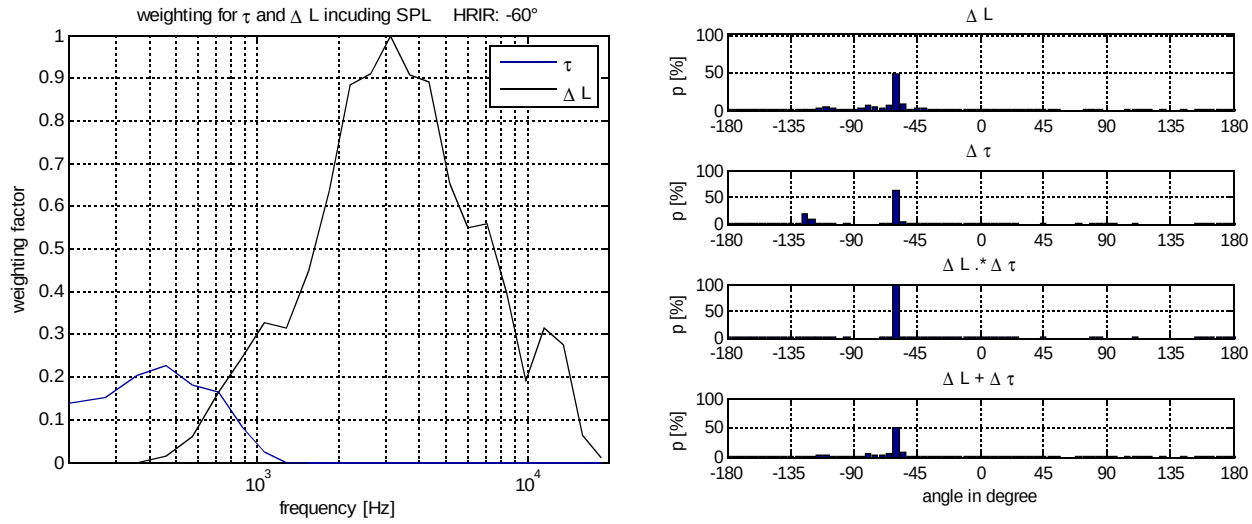

Abb. 18 Gewichtungsfunktion für ΔL und τ für ein Eingangssignal von 60° und Histogramme für ein Eingagngssignal von 60°

Abbildungen 15 und 17 zeigen die Zuordnung von ΔL und τ eines simulierten Signals mit Azimut: 20° bzw. 60°. Bei beiden Signalen ist zu erkennen, dass die Zuordnung zu großen Teilen korrekt ist, jedoch in bestimmten Bereichen abweicht. ΔL ist bei hohen Frequenzen der zuverlässigere Wert, während τ im tieffrequenten Bereich bessere Ergebnisse liefert. Aus diesem Grund werden die Werte, wie in Punkt 3.2.1beschrieben, gewichtet. Die Gewichtungsfaktoren sind in Abbildung 16 und 18 ersichtlich. Mit ihnen werden die Daten für alle Frequenzen zusammengefasst und es entstehen die Histogramme, welche ebenfalls in Abbildung 16 und 18 zu sehen sind.

Die Histogramme für ΔL und τ wurden dabei einmal additiv und einmal multiplikativ verknüpft, wobei die multiplikative Verknüpfung meist die besseren Ergebnisse erzielt.

Bei der Multiplikation fallen Winkel die bei ΔL oder τ nicht detektiert wurden, heraus. Gegensätzliche Aussagen von ΔL und τ fließen anders als bei der Addition nicht ins Ergebnis mit ein.

Die folgenden Bilder zeigen alle 72 Histogramme zusammengefasst. Es wird die Zuordnungswahrscheinlichkeit eines Eingangssignal mit dem Winkel x zum berechnetem Winkel y dargestellt.

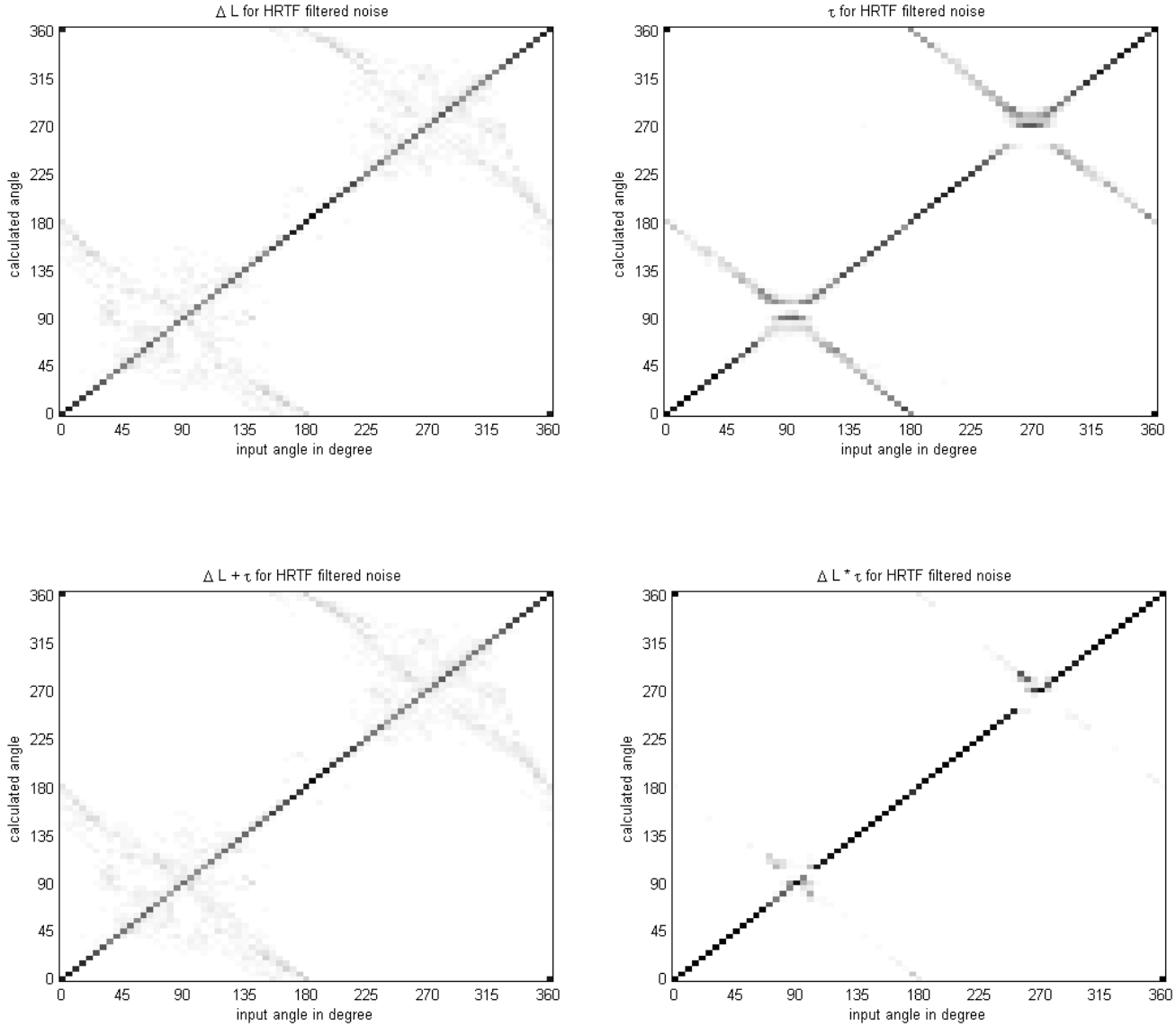

Abb. 19 zusammengefasste Daten mit Gewichtungsfunktion, Übergangsfrequenz 700 Hz

Es wurde für alle 72 möglichen Richtungen das entsprechende Paar HRTF ausgewählt und überprüft wie genau der Algorithmus dieses ideale Signal zuordnen kann. Die Daten wurden in diesem Fall mit der Energie und der Überblendfunktion, welche eine Frequenz von 700 Hz hatte, gewichtet.

Zum Vergleich nun die Ergebnisse ohne jegliche Gewichtung.

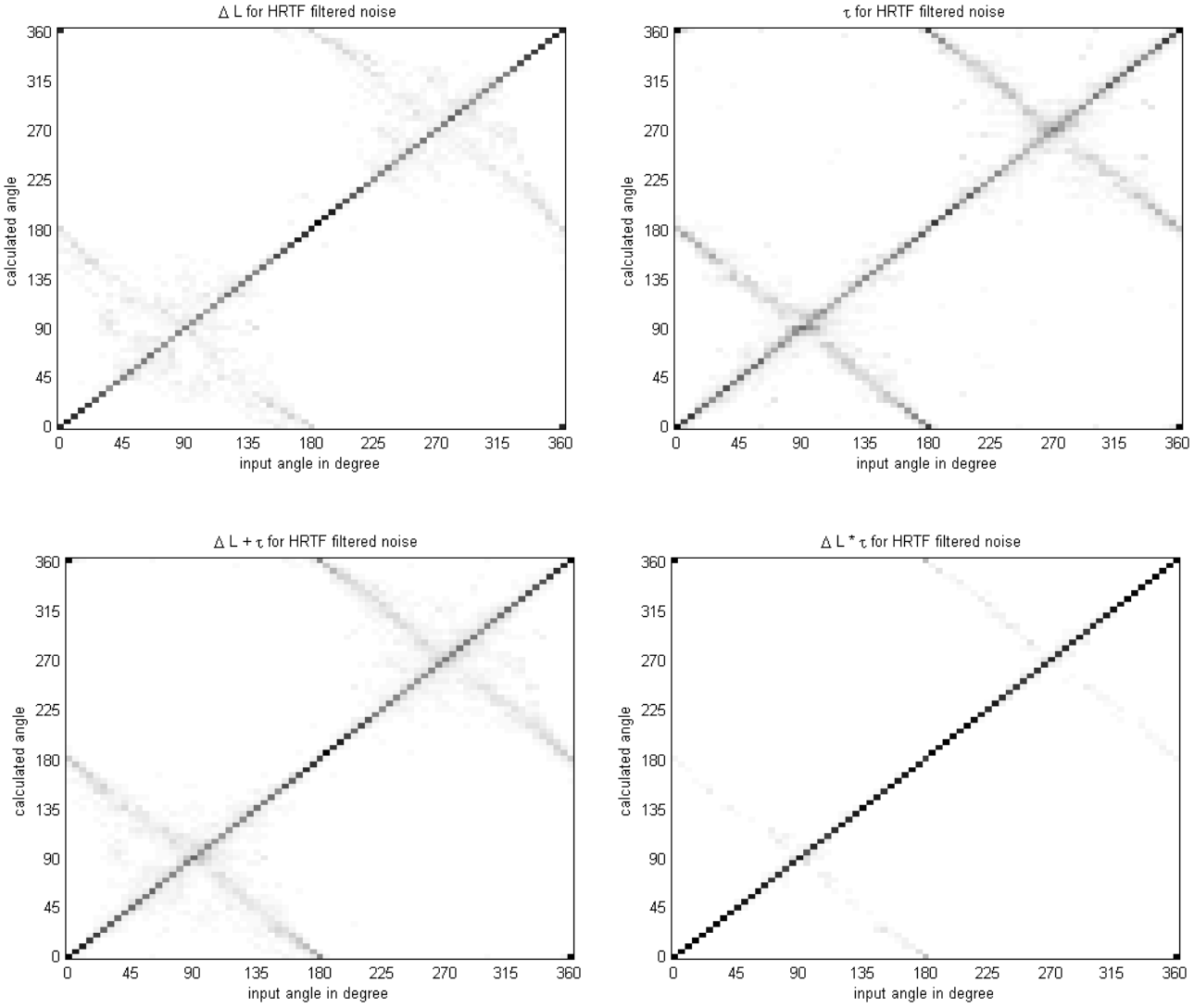

Abb. 20 zusammengefasste Daten ohne Gewichtungsfunktion

Interessant ist, dass die Ergebnisse ohne Gewichtung teilweise besser sind.

Vergleicht man sie mit Abbildung 19 kann man bei ca. 100° und 260° ist bei einer multiplikativer Verkn üpfung eine falsche Zuordnung zu beobachten. Der Grund dafür wird in Abbildung 21 zu erkennen. Während bei 90° Eingang die Daten korrekt erkannt werden liegt bei 100° die Zuordnung von τ um 5° daneben. Bei 100° wird nichts detektiert, was bei der darauf folgenden Multiplikation zur Folge hat, dass die korrekten Werte der ΔL nicht berücksichtigt werden. Die Wahrscheinlichkeit beim eigentlich korrekten Wert geht gegen Null. Dies entspricht sicher nicht der menschlichen Wahrnehmung. Das Problem tritt jedoch nur bei der Multiplikation auf, so dass sich die Frage stellt ob die Addition nicht doch die bessere Möglichkeit einer Verknüpfung von τ und ΔL darstellt.

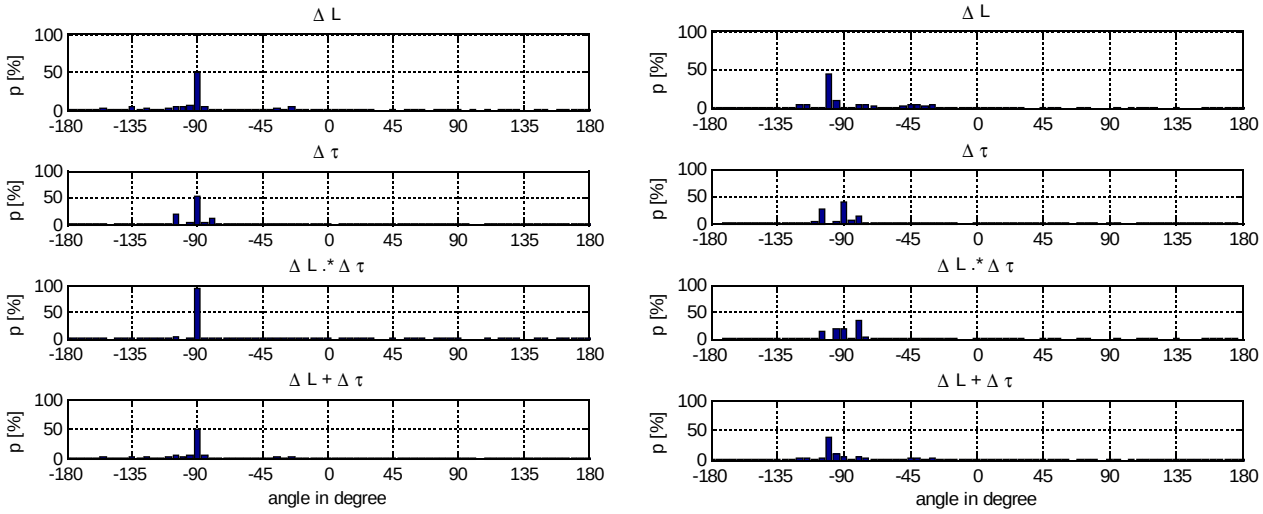

Abb. 21 Histogramme für ein Eingangssignal von -90° u nd -100°

Ohne Gewichtsfunktion gibt es diesen Effekt nicht. Dies ist durch eine größere Streuung der Daten zu erklären, da sowohl für τ als auch für ΔL alle Frequenzbänder berücksichtigt werden. Es sei darauf hingewiesen, dass der Effekt bei realen Signalen sehr wahrscheinlich nicht auftreten wird. Die erzielte Genauigkeit für τ und ΔL ist weitaus geringer, die Werte variieren stärker und im Bereich um den tatsächlichen Wert gibt es keine Ergebnisse die gegen Null gehen.

3.3.2 Reale binaurale Signale

Neben der Reihe an Tests mit idealen Signalen, wurde der Algorithmus auch mit realen Signalen getestet. Hierfür wurden verschiedene Signale wie weißes Rauschen, Bandpassrauschen, Musik und Sprache verwendet. Aufgenommen wurde mit sogenannten Original – Kopf – Mikrofonen (OKM), da dies ein einfacher und schneller Weg ist binaurale Aufnahmen zu generieren. Die Mikrofone sehen aus wie handelsübliche Walkmankopfhörer und werden einfach in die Ohren gesteckt. Die HRTFs entsprechen denen des Anwenders und variieren leicht in Abhängigkeit von Körperhaltung, Kleidung etc. In einem Abstand von 2.10 m wurden die über einen Studiomonitor abgespielten Signale mit den OKMs aufgenommen. Auf einen Testaufbau mit mehren Quellen gleichzeitig wurde verzichtet. Die Winkeländerung wurde über eine Änderung der Sitzposition in 15° Sc hritten realisiert. Die Nachhallzeit des Raumes betrug im Mittel 1.2 s.

Abb. 22 binaurale Aufnahme realen Bedingungen mittels Original – Kopf – Mikrofonen

Die Genauigkeit der Ergebnisse variiert sehr stark mit dem Azimuth. Während Winkel um die 0° oder um die 180° im Vergleich gut erkannt werden können, kommt es bei 90° bzw. 270° zu größeren Abweichungen. Keinen wesentlichen Einfluss auf die Ergebnisse hatte die Beschaffenheit des Signals. Ein Unterschied bzw. eine eindeutige Tendenz zwischen weißen Rauschen und Musik beispielsweise konnte nicht festgestellt werden.

Abbildung 23 und 24 zeigen die Resultate für eine Schallquelle aus – 75° Richtung mit weißem Rauschen.

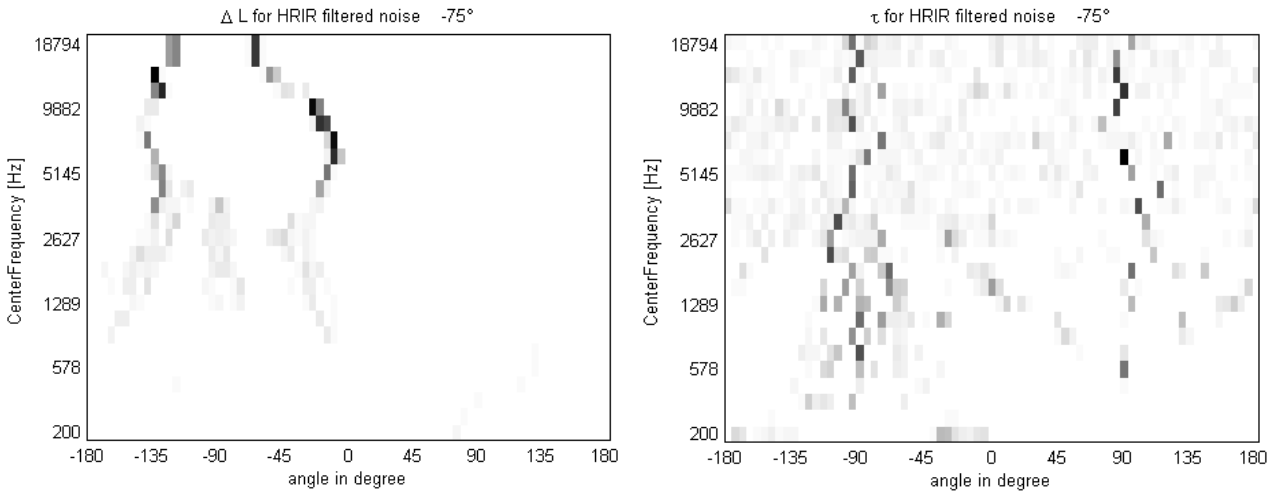

Abb. 23 Auftrittswahrscheinlichkeit von ΔL und τ für ein Eingangssignal: weißes Rauschen aus -75° Richtung

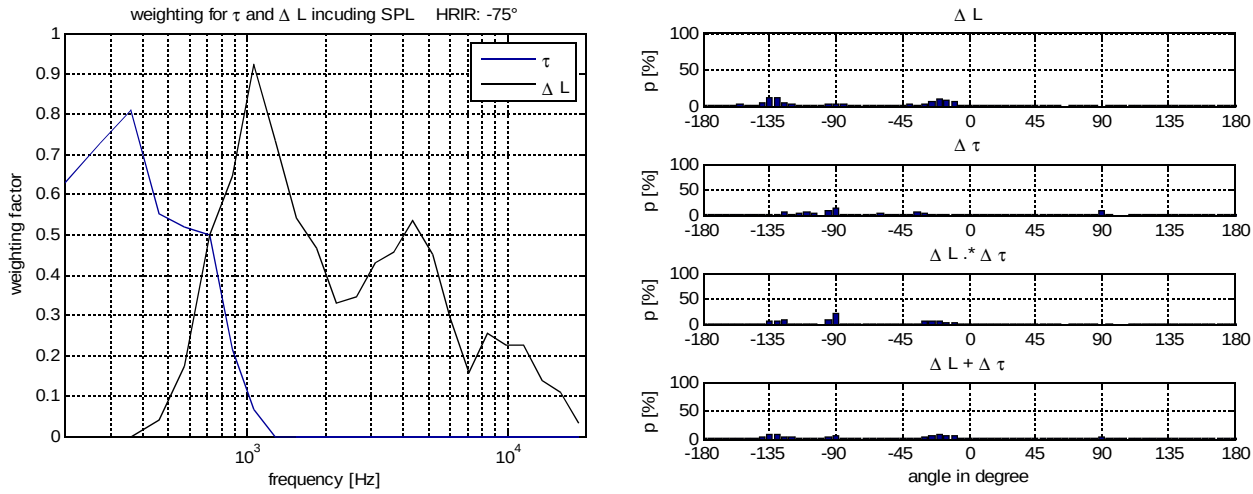

Abb. 24 Gewichtungsfunktion für ΔL und τ und Histogramme für ein Eingangssignal: weißes Rauschen aus -75° Richtung

Abbildung 25 und 26 zeigen die Resultate für eine Schallquelle aus 0° Richtung mit Popmusik , die Länge des untersuchten Ausschnitts betrug 0.5 Sekunden.

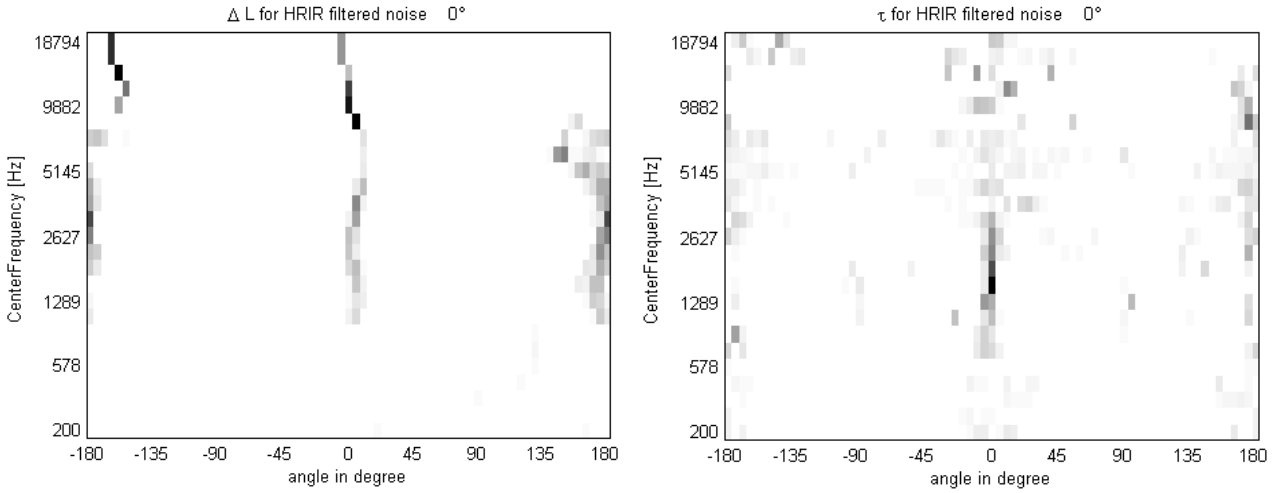

Abb. 25 Auftrittswahrscheinlichkeit von ΔL und τ für ein Eingangssignal: Popmusik aus 0° Richtung

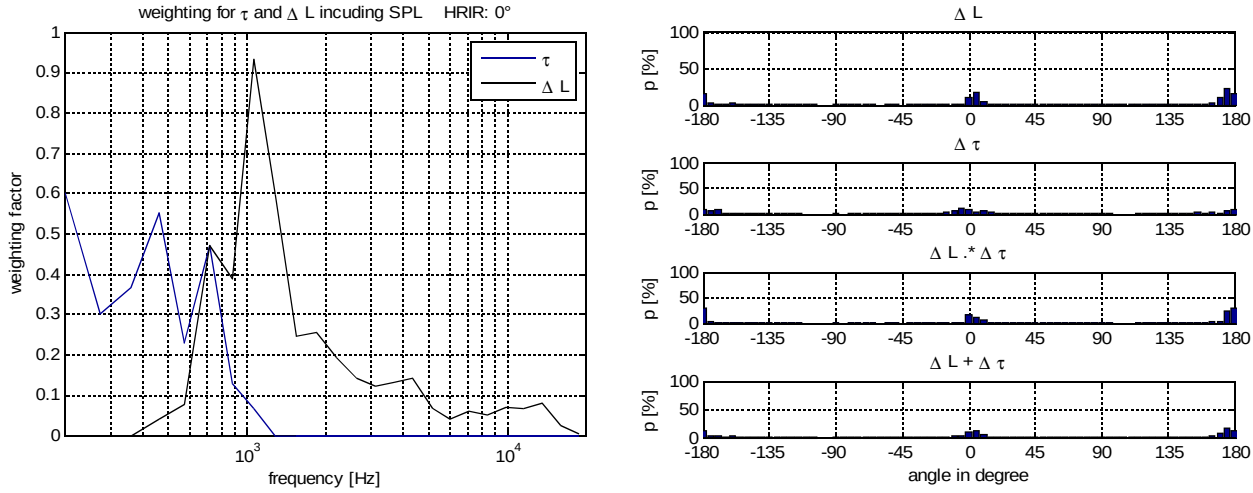

Abb. 26 Gewichtungsfunktion für ΔL und τ und Histogramme für ein Eingangssignal: Popmusik aus 0° Richtung

Die Lokalisation in halliger Umgebung ist sehr schwierig. C.Faller und J.Merimaa können nur mit Einschränkungen und mit der Wahl des geeigneten Schwellwertes im jeweiligen Frequenzband die korrekten ΔL und τ bestimmen. Dies ist jedoch nicht für alle Bänder möglich. Außerdem ist die Streuung deutlich höher als im Freifeld.

Beim erstellten Modell werden alle Bänder berücksichtigt und eine händische Wahl des Schwellwertes kann nicht vorgenommen werden. Dementsprechend fallen die Ergebnisse weitaus weniger eindeutig aus, als die der idealen Signale. Teilweise kommt es zu Fehlern, vor allem nimmt jedoch die Lokalisationsschärfe ab. Während die Abweichung bei – 75° ca. 15° beträgt , ist die Lokalisation bei 0° weitestgehend korrekt . Die Verwechselung von vorne und hinten entspricht bei nicht bekanntem Testsignal der menschlichen Wahrnehmung. In Abb. 23 ist zu erkennen , dass die Berechnung bzw. die Zuordnung der Pegeldifferenzen fehlerhaft ist. Grund dafür sind die Mehrdeutigkeiten der Daten (siehe Abb. 14). Problematisch hierbei ist die eindeutige Zuordnung beim Vergleich mit den Referenzdaten, da Ergebnisse, die nicht die geringst mögliche Abweichung aufweisen, nicht beachtet werden.

Wie schon erwähnt waren signifikante Unterschiede der Ergebnisse für verschiedene Signale nicht festzustellen, wobei extrem schmalbandige Signale nicht untersucht wurden.

Ein Problem beim Messaufbau stellen die verschiedenen HRTFs von Messung und Datenbank dar. Idealerweise würde man zum Erstellen der Datenbank dieselben HRTFs wie die für die Messung verwenden, was jedoch nicht im Rahmen der Möglichkeiten lag. Unter den Voraussetzungen gleicher HRTFs wäre ein ausführlicheres Testen sinnvoll.

4 Fazit und Ausblick

Es ist festzuhalten, dass die Berechnung der Lokalisation, die der menschlichen Wahrnehmung nahe kommt möglich ist. Wie sehr die Ergebnisse des Algorithmus der menschlichen Wahrnehmung im Detail entsprechen bleibt zu untersuchen.

Eine notwendige Erweiterung ist die Einführung eines adaptiven Schwellwertes, da ohne einen geeigneten Schwellwert die Ergebnisse bei komplexeren Versuchsaufbauten nicht ausreichend aussagekräftig sind. Ebenso gilt es das oben beschriebene Problem der zu großen Verzögerungen der HRTFs im tieffrequenten Bereich zu lösen.

Des weiteren hätte man die Möglichkeit mit Hilfe zunächst einfacher Versuche den Algorithmus zu optimieren. Sinnvoll erscheint ein Versuchsaufbau, der bereits mit Testpersonen durchgeführt wurde, sodass ein Vergleich der Ergebnisse leicht möglich ist. So könnte man beim *mapping* den Entscheidungsprozess nach hinten verlagern. Ergebnisse, die bei der Differenzbildung mit den Referenzdaten kein absolutes Minimum aufweisen, könnten so trotzdem zum letztendlichen Ergebnis beitragen. Ebenso erscheint eine Untersuchung der Auswirkung der Gewichtungsfunktion sinnvoll.

Darüber hinaus wäre z.B. die Untersuchung der Lokalisationsschärfe bei Ambisonic-systemen denkbar , wobei eine Erweiterung in die Elevation Voraussetzung wäre.

Quellen

Faller, C & Merimaa, J (2004) „source localisation in complex listening situations: Selection of binaural cues based on interaural coherence“ AES

Tarkanen, M (2008) „A Binaural Auditory Model for Evaluating Quality Aspects in Reproduced Sound“

Willert, V & Eggert, J & Adamy, J & Körner Edgar (2006) „A Probabilistic Model for Binaural Sound Localization“ IEEE

Pulkki, V & Hirvonen T (2005) „Lokalization of Virtual Sources in Multichannel Audio Reproduction“ IEEE

Marchand, S & Vialard, A (2009) „The Hough Transformation for Binaural Source Lokalization“ DAFx

Venegas, R & Lara, M & Correa, R. & Floody, S. (2006) „Spatial Sound Lokalization Model using neural Network“ AES

Breebaart J. (2001) „Binaural Processing model based on contralateral inhibition“ Acoustical Society of America