Entwicklung eines Dynamikprozessors basierend auf psychoakustischer Modellierung

Diplomarbeit

von

Christian Göttlinger

durchgeführt am

Institut für Elektronische Musik und Akustik

der Universität für Musik und darstellende Kunst Graz

Betreuer: Univ. Ass. DI Dr. Alois Sontacchi

Graz, Mai 2009

Zusammenfassung

Das Ziel eines Kompressors ist es, den dynamischen Umfang von Musikmaterial stufenlos mit möglichst geringer Wahrnehmbarkeit zu reduzieren. Um dies zu erreichen, werden zuerst die für Dynamikbearbeitung relevanten psychakustischen Aspekte beleuchtet sowie Stärken und Schwächen bisheriger Ansätze analysiert und diskutiert. An erster Stelle des hier vorgestellten Ansatzes steht eine möglichst gute Modellierung des Verlaufes der Lautheit inklusive auftretender zeitlicher Verdeckungseffekte. Hierfür wird die sehr effiziente „Warped-FFT“ in Verbindung mit einer Weiterentwicklung des nichtlinearen Innenohrmodells von Karjalainen („neuronales Rückkopplungsmodell“) verwendet. Eine statische Kompressorkurve dient zur Ermittlung der Abweichung der tatsächlichen Lautheit von dem gewünschten Soll. Der Regelvorgang zum Erreichen dieses Solls geschieht nun adaptiv zum jeweils psychoakustisch günstigsten Zeitpunkt. Um auch bei starken Amplitudenveränderungen zusätzlich die spektrale Balance beizubehalten, geschieht die weitere Verarbeitung bei der erweiterten „Multiband“-Version in einer lautheitsbewerteten Zeit-Frequenz-Umgebung. Die finale Bearbeitung des Audiosignals erfolgt mit Hilfe eines dynamischen Warped-Filters im Falle der Multiband-Variante und eines einzelnen dynamischen Verstärkungskoeffizienten im Falle der einfacheren, einkanaligen „Fullband“-Variante. Die Implementierung des erarbeiteten Ansatzes erfolgt in MATLAB® und wird durch einen informellen Hörversuch hinsichtlich seiner Leistungsfähigkeit im Vergleich mit herkömmlichen kommerziellen Lösungen untersucht.

Development of a Dynamics Processor Based on Psychoacoustical Modelling

A dynamic processor´s asset is to reduce the dynamic range of music or speech steplessly in a preferably imperceptible way. To achieve this aim, general psychoacoustic aspects plus pros and cons of contemporary approaches are discussed.

The first stage of the proposed approach requires an enhanced modelling of the course of the loudness including temporal masking effects. For this purpose the efficient "Warped-FFT" in conjunction with an advanced version of Karjalainen's inner ear model ("neural feedback model") is used. After that the "static function" determines the difference between the actual loudness and the target loudness. The following control process is adaptively conducted at psychoacoustically opportune moments. To retain spectral balance even at considerable alterations in amplitude, an additional "multiband"-version is proposed. In this case, all further calculations are carried out in a loudness-weighted time-frequency-domain. The final audio-processing is done with the aid of a dynamic "warped"-filter. The simpler "fullband"-alternative utilizes a single gain coefficient for this purpose.

The developed approach is implemented in MATLAB®. Finally an informal listening test is conducted to compare the performance of both versions with current commercial solutions.

Danksagung

Der erste Dank gilt meinen Eltern. Stets wusste ich euch hinter mir, auch bei unkonventionellen Entscheidungen. Eure Liebe und Unterstützung ermöglichten meine Ausbildung und letztlich damit auch diese Arbeit.

Der Verantwortliche für den wissenschaftlichen Tiefgang dieser Arbeit am IEM: der unermüdliche Dr. Alois Sontacchi. Nie um eine neue Sichtweise und Inspiration verlegen und trotz auch Ihrer vielfältigen Aufgaben hatten Sie stets Zeit für eine Sprechstunde oder zwei... Ich danke Ihnen vielmals für die sehr gute Betreuung.
Besonders wichtig für den Ansatz und die praktischen Erdung: Dr. Christoph Musialik von der Algorithmix GmbH. Ich danke Ihnen sehr für ihre großzügige Art, Wissen zu teilen und mit Ihrer Erfahrung einen Weg durch den großen Dschungel der Möglichkeiten und Hindernisse zu bahnen. Außerdem unverzichtbar bei bei Detailfragen zur Implementierung: Dr. Ulrich Hatje - Vielen Dank auch Ihnen.
Einen großes Dankeschön natürlich an alle enthusiastischen Probanden des Hörtests. Speziell hervorheben möchte ich in diesem Zusammenhang Daniel Hojka für seine Beispielsamples und das Diskutieren kniffeliger Fachfragen, Hannes Pomberger für das Bereitstellen der PEAQ-Software und Patrick Gampp für viele Kleinigkeiten.

Schließlich ein riesengroßes Dankeschön an die wunderbarste Frau der Welt, meine Freundin Nadine. Geduldig lebtest du sämtliche Höhen und Tiefen in den vergangenen Monaten mit durch und sorgtest für das richtige Wort oder die richtige Tat zur rechten Zeit.

Inhaltsverzeichnis

Kapitel 1

Einleitung

In fast allen modernen Musikstilen wird die Dynamik heute teils massiv verändert. Es fällt in der Tat sogar sehr schwer, unkomprimierte aktuelle Audioaufnahmen aus dem Jazz oder gar Pop-Bereich zu finden. Sieht man von ästhetischen Gründen ab, so dient der Einsatz eines Kompressors vor allem zum Anpassen an die verfügbare Dynamik. Bei analogen Übertragungskanälen oder Trägermedien war die Bearbeitung wegen dem eingeschränkten Dynamikbereich zwischen dem Grundrauschen und hörbaren Verzerrungen („Übersteuern“) unverzichtbar. Heute bestimmt eher die Wiedergabesituation den zur Verfügung stehenden Spielraum. Musik wird oft im Auto oder mit tragbaren Geräten und Kopfhörern in der Öffentlichkeit gehört. Der hier zur Verfügung stehende Dynamikbereich ist sehr gering. Im Auto herrschen zum Beispiel oft Geräuschkulissen von $60dB$ über der Hörschwelle vor. Abbildung 1.1 demonstriert diese Problematik. Um den Hörer nicht mit zu hoher Lautstärke zu belasten oder sogar Hörschäden zu verursachen, müssen nun die lauteren Passagen gedämpft werden. Die leisen Stellen hingegen sollten angehoben werden, damit sie nicht im Rauschen untergehen. Die nötige Korrektur der Lautheit kann beträchtlich sein. Symphonieorchester können durchaus $60dB$ Dynamik zwischen Pianissimo und Fortissimo entfalten, im Auto stehen je nach Geschwindigkeit teilweise weniger als $15dB$ zur Verfügung. In privaten Wohnungen ist auch oft die maximal mögliche Lautstärke ein einschränkender Faktor

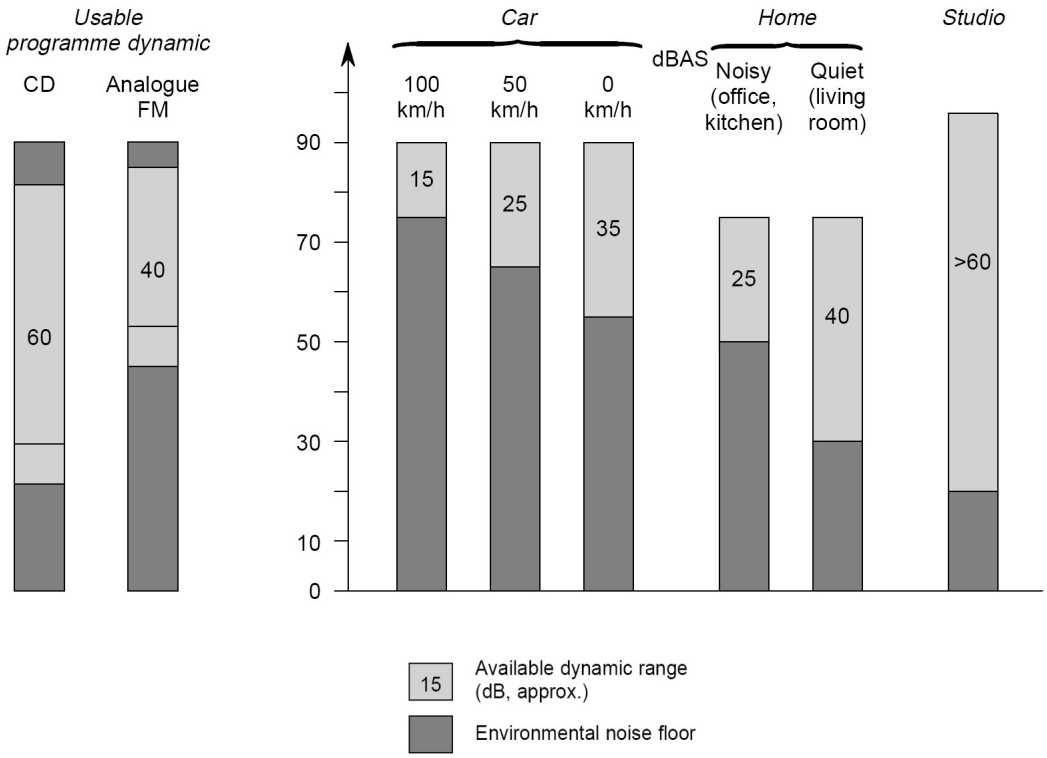

Abbildung 1.1: Verfügbare Dynamik über der Hörschwelle ($[dBAS] \rightarrow dB$ above Silence, über der Hörschwelle) verschiedener Abhörsituationen

(Abbildung 1.1). Musik soll die Nachbarn nicht stören, wird nebenbei gehört oder soll Gespräche durch plötzliche laute Stellen nicht unterbrechen.

Die Kunst besteht nun darin, diese Regelung möglichst unauffällig zu gestalten, das Klangbild und auch die musikalische Aussage in den deutlich reduzierten Dynamikbereich zu übertragen.

Hierfür werden heute verschiedene analoge wie digitale Dynamikprozessoren eingesetzt, die allesamt in einer langen Tradition von kontinuierlicher Weiterentwicklung oftmals durch „Trial and Error“ stehen. Erstaunlich ist hier die Kluft zwischen der angewandten praktischen Implementierung und der wissenschaftlich psychoakustischen Forschung. Der gegenwärtige kommerzielle Trend tendiert zum Beispiel zum Simulieren von komplexen analo-

gen „Vintage“-Geräten in digitalen Umgebungen.

Das Ziel dieser Arbeit ist es, mit Hilfe psychoakustischer Erkenntnisse und Gehörmodelle einen Ansatz zu entwerfen, der nur die Lautheit des Audiomaterials ohne Artefakte möglichst effizient verändert. Andere psychoakustische Parameter wie die Impulstreue, die spektrale Balance und deren Entwicklung über die Zeit oder auch Akzentuierungen sollen dabei nur zu einem Mindestmaß beeinflusst werden. Die durch die Verarbeitung hervorgerufene Verzögerung kann größere Werte annehmen, da ein direkter Einsatz im Musiklivebetrieb nicht vorgesehen ist. In den angedachten Wiedergabesituationen wie dem Auto, der Heimkinoanlage oder einer Studioumgebung ist dieses Delay nicht störend, beziehungsweise kompensierbar.

In Kapitel 2 werden darum die für die Wahrnehmung der Lautstärke wichtigen psychoakustischen Grundlagen und Erkenntnisse zusammengefasst. Anschließend erfolgt eine Analyse des bisher üblichen Vorgehens. Am Ende werden aus den beiden vorherigen Punkten Schlüsse gezogen und konkrete Anforderungen und gewünschtes Verhalten formuliert. Kapitel 3 schildert ausführlich den hier vorgeschlagenen Ansatz inklusive Implementierungsdetails. In Kapitel 4 wird der abschließende vergleichende Hörtest diskutiert.

Kapitel 2

Theorie und Situationsanalyse

2.1 Psychoakustische Grundlagen

In diesem Kapitel werden die für eine Dynamikbearbeitung relevanten psychoakustischen Grundlagen beleuchtet. Zuerst wird die physiologische Seite des Gehörs behandelt, anschließend werden einige Erkenntnisse aus Studien über die subjektive Wahrnehmung dargelegt. Eine kurze Übersicht über einige aktuelle Hörmodelle folgt. Das Kapitel schließt in einer Diskussion über die Verwendung dieser Erkenntnisse für die Dynamikkompression.

2.1.1 Physiologischer Aufbau des peripheren Gehörs

Der physiologische Aufbau des Ohres (Abbildung 2.1) selbst und die Funktion der einzelnen Komponenten sind inzwischen relativ gut erforscht. Im Folgenden werden die für die Wahrnehmung der Lautstärke und Dynamik nicht so wesentlichen Teile nur kurz abgehandelt. Interessierte Leser seien für genauere Informationen auf die ausführlichen Bücher von Gelfand [Gel04] für physiologische Details oder Moore [Moo03] für deren Modellierung verwiesen.

Primär kann der menschliche Hörsinn durch das periphere Gehör und die neuralen Verarbei-

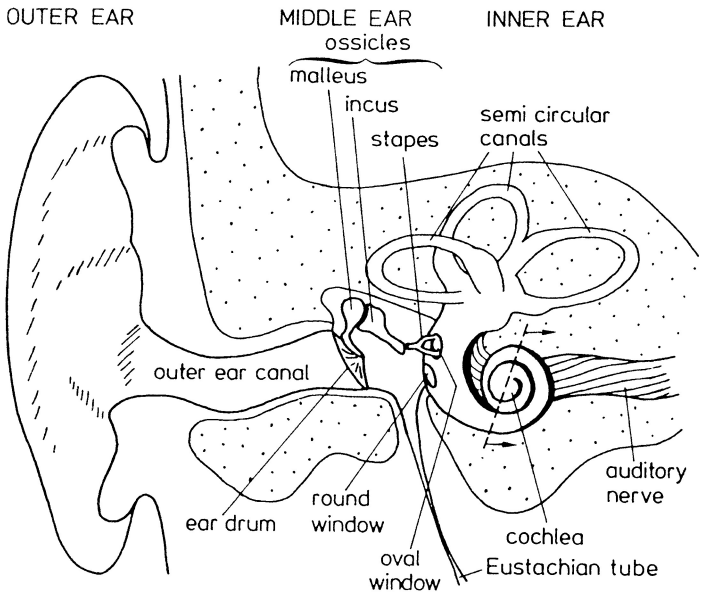

Abbildung 2.1: Schematischer Aufbau des Ohres [FZ07]

tungsstufen beschrieben werden. Das periphere Gehör gliedert sich in das Außen-, Mittel- und Innenohr. In folgendem werden die einzelnen Teile von außen nach innen erläutert.
Die wesentliche Aufgabe der Pinna des Außenohres besteht als Schalltrichter in der Impedanzanpassung und außerdem in der Verbesserung der räumlichen Lokalisation. Dies geschieht durch eine richtungsabhängige Filterung. Der Hörkanal trägt neben seiner Schutzfunktion (das empfindliche Mittel- und Innenohr befindet sich weiter innen im Kopf) auch als Resonator zu der erhöhten Empfindlichkeit des Ohres in dem für die Sprachverständlichkeit wichtigen mittleren und hohen Frequenzbereich bei. Abbildung 2.2 zeigt die deutliche Verstärkung zwischen 1 und 8 kHz. Das Mittelohr hat die Hauptfunktion, den Luftschall mittels eines Hebelsystems über das ovale Fenster auf das mit Flüssigkeit gefüllte Innenohr zu übertragen. Wegen der unterschiedlichen akustischen Impedanzen würde ansonsten ein Großteil des Luftschalles reflektiert und nicht übertragen werden. Das Hebelsystem besteht aus den drei Gehörknöchelchen (Hammer, Amboss und Steigbügel) und den beiden Mittelohrmuskeln. Einer der beiden, der „Musculus stapedius“ ist in der Lage, mit dem „Stapediusreflex“

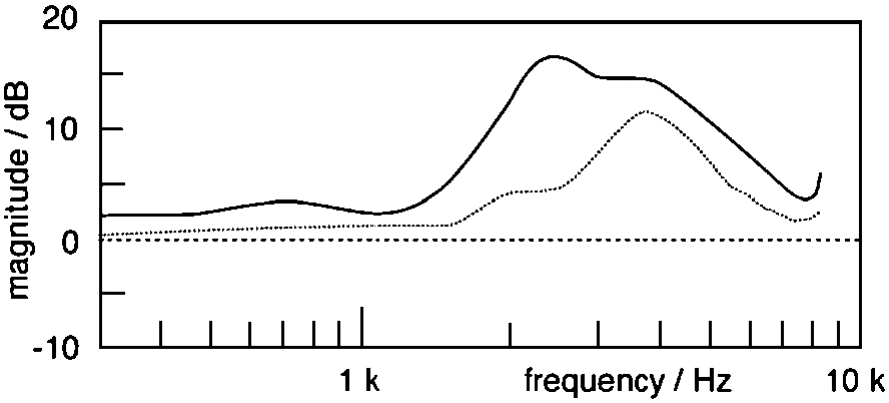

Abbildung 2.2: Übertragungsfunktion von einer frontalen Schallquelle (durchgezogene Linie) sowie vom Eingang des Ohrkanals (gestrichelte Linie) zum Trommelfell [Kar08]

die Übertragungsfähigkeit des Mittelohres zu verschlechtern (siehe Abbildung 2.3). Dies geschieht als Schutzfunktion bei hohen Schalldrücken und setzt zwischen etwa 75 (Rauschen) und 90 $dB\,SPL$ (sinusförmige Anregung) ein. (SPL: „Sound Pressure Level“: dB-SPL wird für Messungen auch nichtstatischer Signale verwendet, da es einen zeitlich lokalen und begrenzten Integrator über das Signal enthält. In der Praxis werden die ankommenden Messwerte quadriert und anschließend mit einem Tiefpass 1. Ordnung integriert und über die Zeit „verschmiert“. Als Nullpunkt dient die Hörschwelle.) Die Verstärkung des Mittelohres ist ebenfalls nicht über alle Frequenzen gleich gut. Tiefe und vor allem hohe Freqenzen werden schlechter übertragen. In Addition mit der Übertragungsfunktion des Außenohres liegt hier der Grund für die erhöhte Empfindlichkeit des Gehörs im mittleren Frequenzbereich. Die Eustachische Röhre dient lediglich zum Ausgleich des statischen Luftdrucks und zum Ableiten von Sekreten.

Im hinter dem Mittelohr liegenden Innenohr werden die akustischen Schwingungen in neuronale Impulse umgesetzt. Dies geschieht in der Hörschnecke, der Cochlea. Oft wird diese wie auch hier (Abbildungen 2.4 und 2.5) zu Demonstrationszwecken „ausgerollt“ skizziert. Sie besteht prinzipiell mechanisch aus einem vom ovalen Fenster zu Ihrem Ende (Apex)

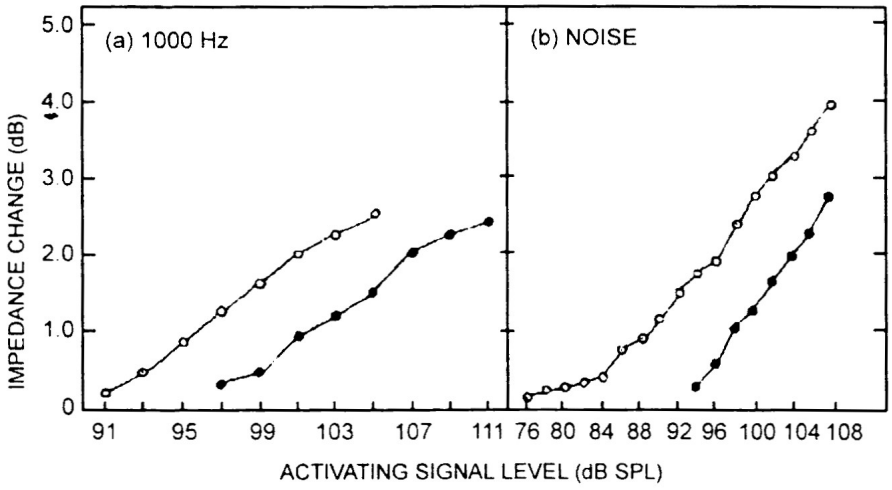

Abbildung 2.3: Akustischer Reflex bei Gesunden (offene Kreise) und Hörgeschädigten (gefüllte Kreise) [Gel04]

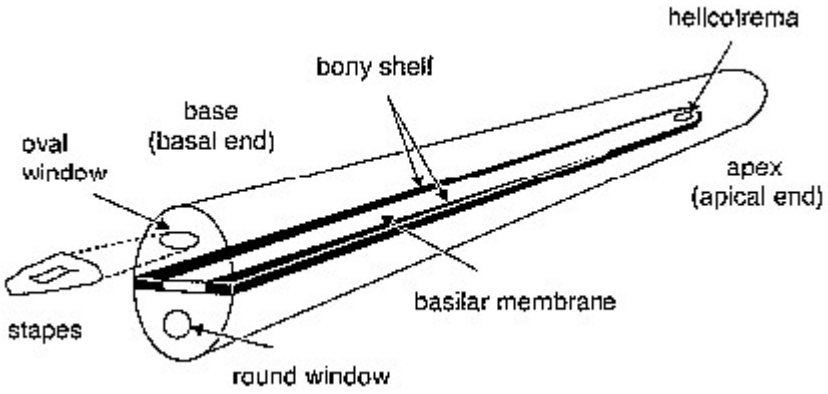

Abbildung 2.4: Skizze Cochlea linearisiert [Kar08]

hinführenden Kanal sowie einem von diesem wieder zum Mittelohr zurückführenden Kanal. Die Verbindung beider Kanäle am Apex geschieht durch das Helicotrema, die zurückführende Scala tympani endet am runden Fenster zum Mittelohr. Getrennt werden beide Kanäle durch die flexible Basilarmembran, auf der sich das eigentliche Hörorgan befindet. Der hinleitende Kanal ist durch die feine Reissnersche Membran noch in die positiv ionisierte Scala Media

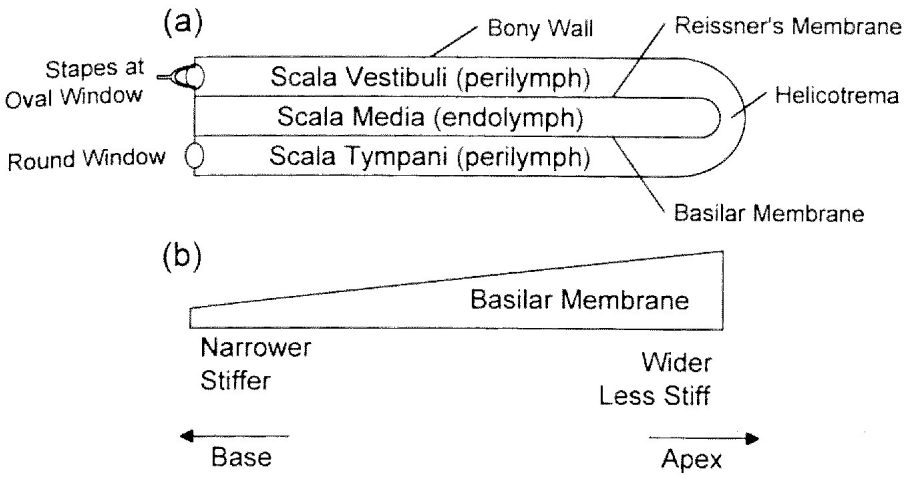

Abbildung 2.5: Cochlea Schema [Gel04]

und die neutrale Scala vestibuli unterteilt. Dies hat auf das mechanische System keinerlei Auswirkungen, die positive Ladung der in der Scala Media vorhandenen Endolymphe wird zum Auslösen der neuronalen Impulse benötigt.

Die vom Mittelohr übertragenen Schwingungen ziehen als Wanderwelle vom ovalen Fenster

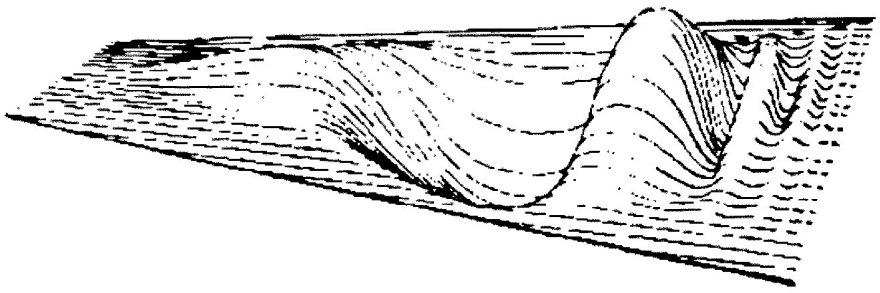

Abbildung 2.6: Basilarmembran mit Wanderwelle [Gel04]

zum Helicotrema und dann wieder zurück zum runden Fenster. Die elastische Basilarmembran wird dadurch wie in den Abbildungen 2.6 und 2.7 ausgelenkt.

Da die Scala Vestibuli und die Scala Media zum Helicotrema immer weiter werden und die Basilarmembran immer weniger steif (2.5), erreichen hohe Frequenzen am Beginn der Coch-

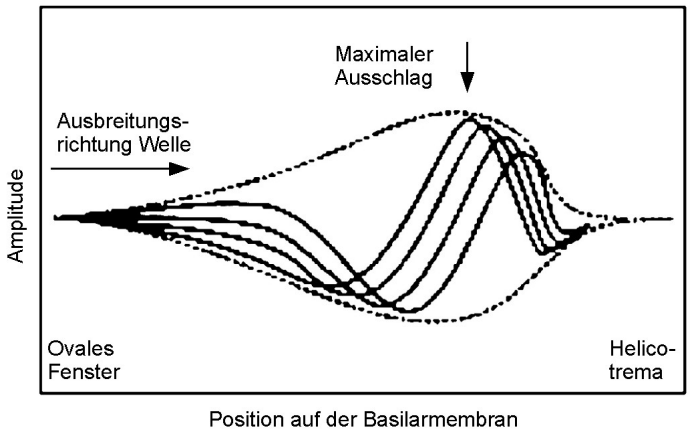

Abbildung 2.7: Basilarmembran mit Wanderwelle seitlich, nach[Kar08]

lea hohe Auslenkungen, tiefe Frequenzen am Ende. Es findet also eine gewisse Frequenz-Ort Umsetzung über die Basilarmembran statt.

Die Umsetzung der Basilarmembranbewegungen in neuronale Impulse geschieht nun mit

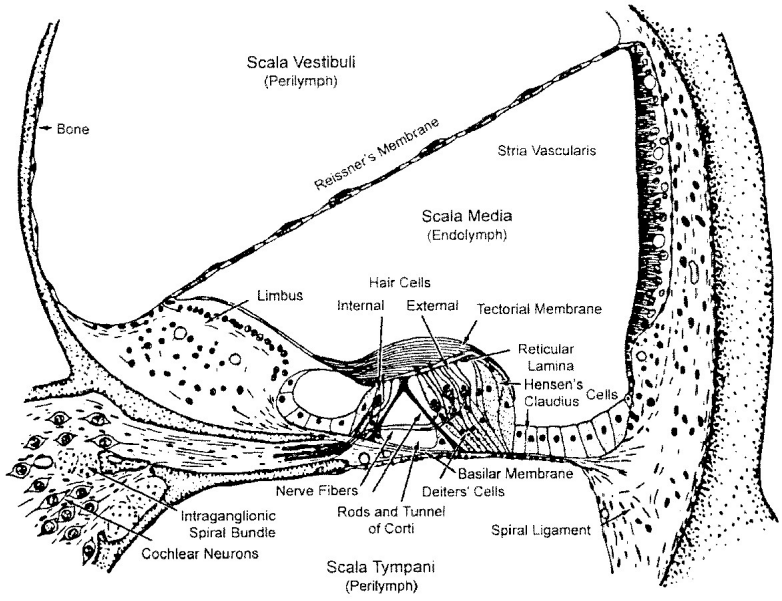

Abbildung 2.8: Cochlea Durchschnitt, Schema [Gel04]

dem Cortischen Organ auf der Basilarmembran. In ihm befindet sich longitudinal angeord-

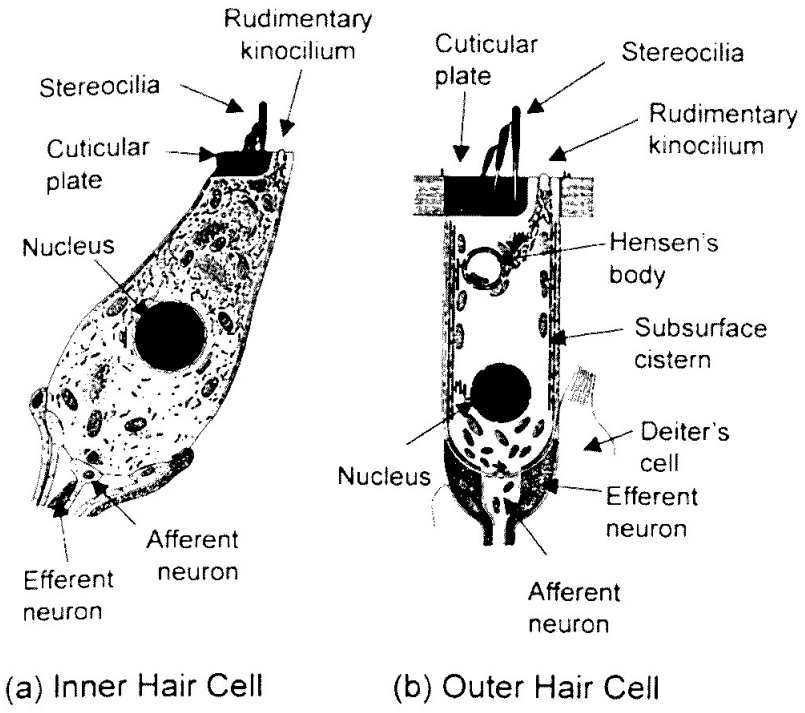

Abbildung 2.9: Innere und äußere Haarzellen, Schema [Gel04]

net eine Reihe „Innere“ und je nach Position 3 bis 5 Reihen „Äussere Haarzellen“. Insgesamt befinden sich laut [Gel04] etwa 12000 Äußere und 3500 Innere Haarzellen im Innenohr. Diese sind jeweils an der Spitze mit den sogenannten Stereocilia-Bündeln versehen, die in der Endolymphe schweben. Ohne genauer auf den mechanisch-chemischen Vorgang einzugehen kann allgemein gesagt werden, dass durch eine Seitwärtsbewegung der Stereocilia die mit den Haarzellen verbundenen affarenten Neuronen geladen werden. Wird eine jeweils spezifische Ladungsmenge erreicht, so geben diese Neuronen einen Impuls ab. Es kann also vereinfacht von einer Pulsfrequenzmodulation gesprochen werden. Je langsamer die maximale Feuerrate, desto größer ist der Dynamikbereich, aber desto schlechter natürlich auch die zeitliche Auflösung (siehe Abbildung 2.10). Der Großteil der Neuronen besitzt einen Dynamikbereich von $20-30\ dB$, einige sehr langsame auch mehr als 50 dB. Insgesamt befinden sich etwa 30000 affarente Neuronen nebst Nerv zum zentralen Nervensystem im Innenohr,

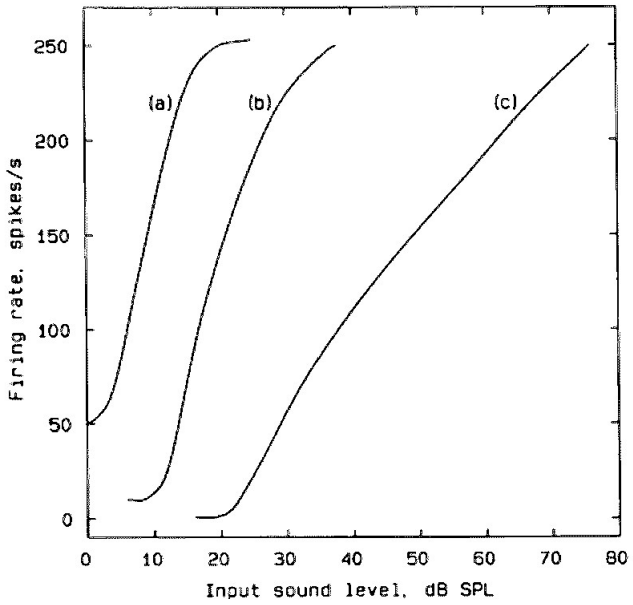

Abbildung 2.10: Neuronen mit a) hoher, b) mittlerer und c) niedriger Feuerrate [Moo03]

eine innere Haarzelle ist meist exklusiv mit etwa 10 Neuronen unterschiedlicher Feuerrate verbunden. Nur 5% der affarenten Neuronen sind mit den äußeren Haarzellen verbunden, also sind hier etwa 10 Haarzellen jeweils an einen Nerv gekoppelt. Ausserdem ist ein Nerv mit etwa $4 - 6$ äußeren Haarzellen quer verbunden. Die Lokalität und Auflösung der hier entstehenden Wahrnehmung ist also deutlich geringer. Während sich die Stereocilien der Inneren Haarzellen freischwebend in der Endolymphe befinden, stoßen die Spitzen der Äußeren Haarzellen an die an der Innenseite der Scala Media befestigen Tektorialmembran. In der Folge ist die Auslenkung der Stereocilien und damit die Feuerrate der Neuronen bei den Äußeren Haarzellen analog zum Schalldruck. Dieser verursacht die Auslenkung der Basilarmembran, was eine Scherbewegung zwischen dieser und der Tektorialmembran erzeugt. Die Stärke der Auslenkung wird so auf die Stereocilien übertragen. Die Auslenkung der Inneren Haarzellen ist im Gegensatz dazu analog zur Änderungsrate der Basilarmembranbewegung (differenzierende Wirkung). Ausserdem erfolgt bei Ihnen nach kurzer Zeit (etwa 20-60 ms) eine Synchronisation der Impulsabstände auf die Signalfrequenz. Abhängig von der Frequenz

eines Sinus feuert eine Zelle in immer gleichen Zeitabständen. Diese Abstände haben eine relativ eindeutige Beziehung zur Signalfrequenz.
Auffällig ist nun die klaffende Lücke zwischen der Dynamik der einzelnen Neuronen von größtenteils $20 - 30\ dB$ und der tatsächlichen Dynamik des Gehörs von mehr als 100 dB. Einerseits sind die Neuronen nach [Här99a] mit verschiedenen Offsets versehen (sichtbar auch in Abbildung 2.10), andererseits wird dies durch einen aktiven rückgekoppelten Verstärkungsprozess aufgeholt. Die Wanderwelle wird vor allem bei mittleren Lautstärken ($40 - 70\ dB$ über der Hörschwelle) in Ihrer Amplitude massiv verstärkt. Dies geschieht frequenzselektiv (also ortsselektiv auf der Basilarmembran) um die jeweilige Anregungsfrequenz herum. Verantwortlich hierfür sind die Äußeren Haarzellen. Deren Länge kann sich elektrochemisch schnell verändern. Die äußeren Haarzellen sind weiterhin deutlich dichter efferent (d.h. vom Zentralnervensystem hinausleitend, steuernd) mit Nerven verbunden als affarent. Ein bedeutender Teils des nichtlinearen Verhaltens des Gehörs findet also hier seinen Ursprung. Damit verbundene Erscheinungen sind die sich lautstärkenabhängig verändernde simultane und zeitliche Verdeckung und nichtlineare Lautstärkeneffekte. Betrachtet man den zeitlichen Verlauf der Dichte der Neuronenimpulse zum Beispiel von Sinusbursts (Abbildung 2.11) ganzer Haarzellenpopulationen, so ergibt sich der Effekt des Überschwingens beim Einsatz („Onset“) und des Unterschwingens beim Ende des Bursts („Offset“). Grund hierfür ist, dass die Verstärkung in dem betreffenden Frequenzbereich kurz vor dem Onset sehr hoch, bei Stille sogar maximal ist. Trifft nun der Burst auf diese hohe Verstärkung, so schlagen die meisten Hörzellen mit ihrer maximalen Feuerrate aus, insgesamt geht die Dichte der Impulse in die Sättigung. Die Verstärkung wird schnell reduziert, die Impulsdichte nimmt ab und pegelt sich auf einem konstanten Niveau ein. Verschwindet nun das Signal plötzlich, so ist die Verstärkung viel zu gering um das nun sehr viel leisere Signal noch in den Messbereich der meisten Neuronen zu heben. Die Impulse setzen fast komplett aus. Die Verstärkung wird nun wieder erhöht um den nun anliegenden Signalinhalt wieder in den Messbereich des Großteils der Neuronen zu bringen.

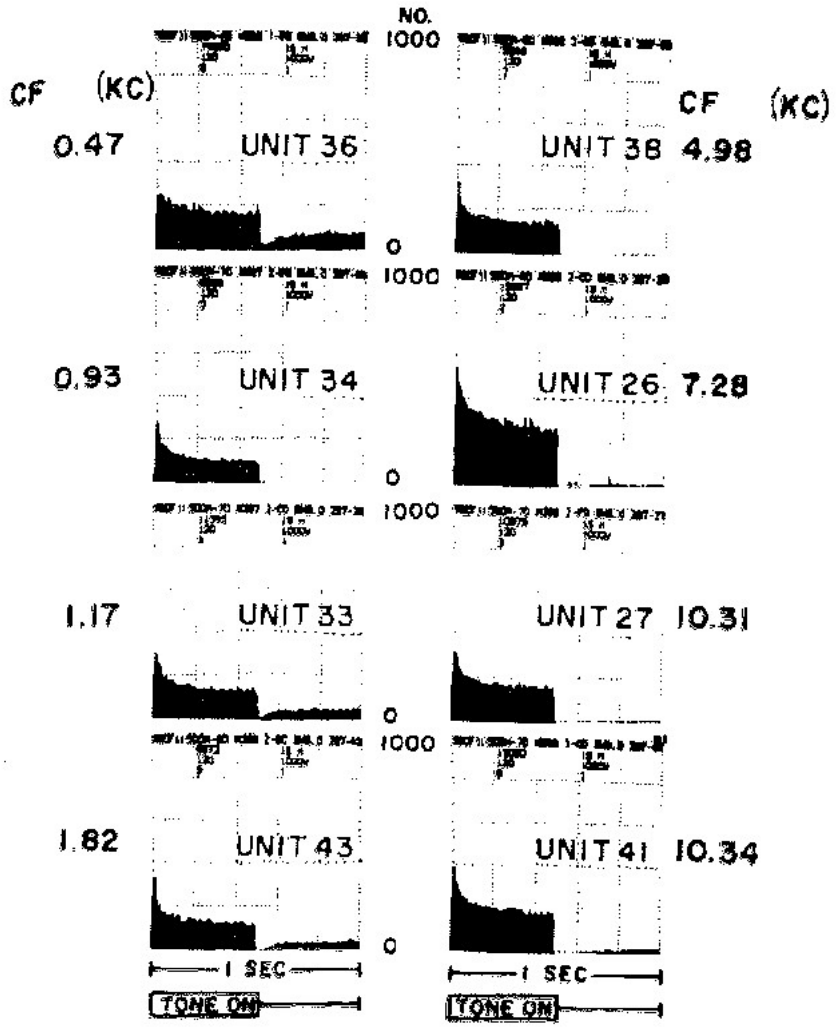

Abbildung 2.11: Neuronenfeuerdichte bei verschiedenen Zentralfrequenzen als Antwort auf einen Sinusburst, Schema [Gel04]

Systemtheoretisch handelt es sich hier somit um ein „dynamisch nichtlineares Verhalten“.

2.1.2 Erkenntnisse aus psychoakustischen Studien

Im Gegensatz zum vorhergehenden Kapitel geht es hier um die Zusammenfassung empirisch erworbener Erkenntnisse.

Lautstärkewahrnehmung

Nicht zuletzt im Rahmen von Lärmmessungen stellt sich immer wieder die Frage nach der tatsächlich empfundenen Lautstärke und weniger des physikalischen Schalldrucks. Die meisten Methoden sind ursprünglich für statische Quellen wie Rauschen oder Sinusschwingungen entwickelt worden und später teilweise für dynamische Vorgänge erweitert worden. Eine der Kernpunkte ist dabei die lautstärkenabhängige Gewichtung der Schalldrücke sowie die Beurteilung des Einflusses der Bandbreite des Signals. Die Kurven gleicher Lautstärke geben an, wie laut ein statischer Ton beliebiger Frequenz sein muss, um genauso laut wie ein entsprechender Ton bei 1kHz wahrgenommen zu werden. Die Einheit ist *Phon*. 1 *Phon* entspricht jeweils 1 dB bei $1kHz$. Eine der einfachsten Messmethoden ist die RMS-Mittelung

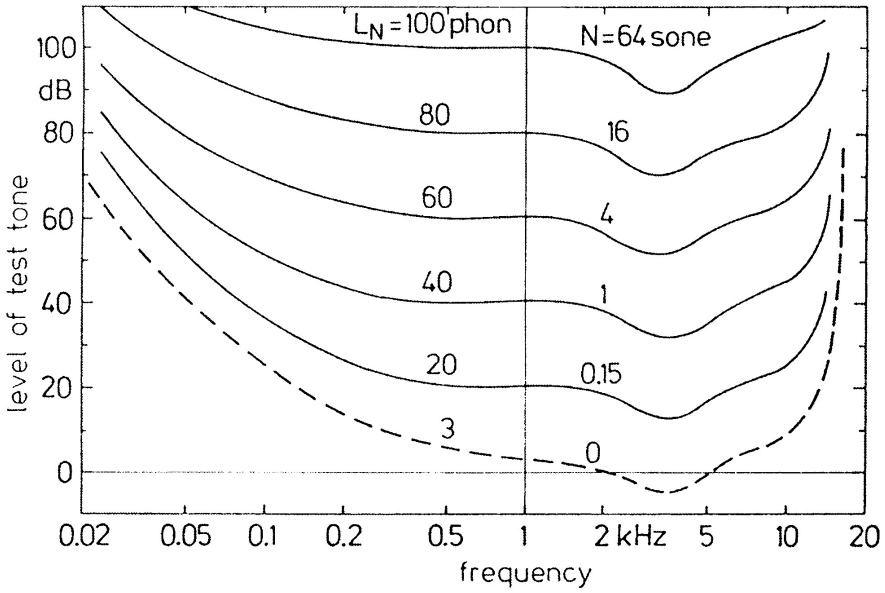

Abbildung 2.12: Kurven gleicher Lautheit für Sinustöne [FZ07]

des Schalldrucks über lange Zeit. Um eine ungefähr korrekte Gewichtung der Frequenzen zu erhalten, wird ein Bewertungsfilter entsprechend der ungefähren Lautstärkekategorie vorgeschaltet. Beispiel hierfür ist die im IEC 61672-Standard für Lautstärkemessungen definierte

A-Gewichtung ($40Phon$) oder die in [ITU06] aufgeführte RLB-Gewichtung.

Zwicker [FZ07] entwickelte bereits in den sechziger Jahren des vorigen Jahrhunderts ein bis heute in Abwandlungen verwendetes Lautheitsmodell, das die Frequenzgewichtung, die nichtlineare Skalierung der Lautstärke und die Bandbreitenbewertung der Signale bereits beinhaltet.
Zu Beginn wird das Signal durch eine Filterbank (alternative Implementierungen benutzen FFTs mit Summierung der jeweiligen Bänder, „Warped"-FFTs [Hau97] etc.) in die gehörrichtigen Frequenzgruppen („Barkbänder") unterteilt. Diese werden aus dem Konzept der kritischen Bandbreite abgeleitet, das gewissermaßen die frequenzielle Auflösungsfähigkeit des Gehörs beschreibt. Ein Effekt ist zum Beispiel, dass ein amplitudenmoduliertes Schmalbandrauschen lauter wahrgenommen wird, wenn es durch die Modulation spektral breiter als die kritische Bandbreite von einem Bark wird. Innerhalb der kritischen Bandbreite bleibt die Lautstärkenwahrnehmung in etwa gleich. Genau definiert ergibt sich dieser Effekt um die zentrale Frequenz des Signals. Die Breite der Filter erweitert sich bei hohen Lautstärken. Dies ist zum Beispiel mit den sogenannten Gammchirp-Filtern modellierbar (siehe z.B. [UIG+06]). Interessant ist in diesem Zusammenhang, dass jedes Barkband einen etwa $1.3mm$ langen Abschnitt auf der Basilarmembran und damit etwa 150 Inneren Haarzellen entspricht ([FZ07]). Mit Hilfe des quadratischen Mittels („Root Mean Square-RMS") und anschließender Quadrierung wird die durchschnittliche Anregung (Schallintensität) in den einzelnen Bändern während des Beobachtungszeitraumes geschätzt. Mit folgender Formel aus [FZ07] wird die Anregung in spezifische Lautheit N' umgerechnet:

$$N' = 0.08\left(\frac{E_{TQ}}{E_0}\right)^{0.23}\left[\left(0.5 + 0.5\frac{E}{E_{TQ}}\right)^{0.23} - 1\right]\frac{sone_G}{Bark} \tag{2.1}$$

E_{TQ} : Wahrnehmungsschwelle im jeweiligen Barkband
E_0 : Referenzintensität 0dB@1kHz
E : Aktuelle Anregung im jeweiligen Barkband

Mit der abschließenden Summierung über alle Bänder ergibt sich die Lautheit in Sone. Eine

z [Bark]	f_u [Hz]	f_o [Hz]	Δf_G [Hz]	f_m [Hz]
0	0	100	100	50
1	100	200	100	150
2	200	300	100	250
3	300	400	100	350
4	400	510	110	450
5	510	630	120	570
6	630	770	140	700
7	770	920	150	840
8	920	1080	160	1000
9	1080	1270	190	1170
10	1270	1480	210	1370
11	1480	1720	240	1600
12	1720	2000	280	1850
13	2000	2320	320	2150
14	2320	2700	380	2500
15	2700	3150	450	2900
16	3150	3700	550	3400
17	3700	4400	700	4000
18	4400	5300	900	4800
19	5300	6400	1100	5800
20	6400	7700	1300	7000
21	7700	9500	1800	8500
22	9500	12000	2500	10500
23	12000	15500	3500	13500
24	15500			

Abbildung 2.13: Die Kritischen Bänder in *Bark* nach Zwicker [Zöl05]

Verdopplung des Sone-Wertes entspricht auch einer Verdoppelung der empfundenen Lautheit. Ein Sinuston von 40 dB bei 1 kHz entspricht 1 Sone.

Um das Modell auch auf Signale mit nicht statischer Lautstärke anwendbar zu machen erfolgt die Schätzung der Lautheit laufend in möglichst kurzen Fenstern (10-50ms). Diese Ergebnisse werden abschließend mit einem Tiefpass über die Zeit integriert. Eine Daumenregel nach

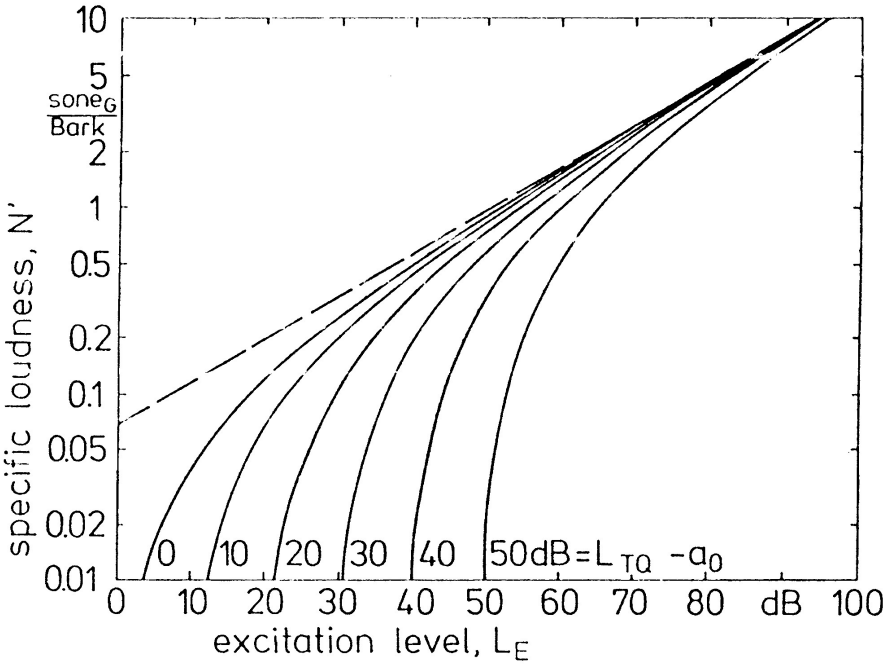

Abbildung 2.14: Die Spezifische Lautheit [FZ07]

Zwicker dafür ist, dass die Lautheit in Phon nach dem Signalbeginn linear mit etwa $10Phon$ pro Dekade Zeit(= 3 Phon pro Verdoppelung) ansteigt. Dies berücksichtigt natürlich nicht die tatsächlichen Sättigungserscheinungen nach gewisser Zeit. Für das exakte Zeitverhalten ist es besser die Zunahme der Nachverdeckung mit der Maskerlänge als Referenz zu nehmen (siehe weiter unten).

Für die Dauer des Lautstärkeeindruckes gibt es zwei verschiedene Größen [GM02]: die kurzfristige ("shortterm") und langfristige („longterm") Lautheit. Erstere bezieht sich auf die empfundene Lautheit einzelner Ereignisse wie etwa Silben, zweitere auf die Lautheit größerer Strukturen wie etwa ganzen Sätzen und bezieht somit auch ein gewisses Gedächtnis mit ein. Die kurzfristige Lautheitsempfindung korreliert eng mit dem Verdeckungseffekt ([Fas77b], [FZ07], [GM02]), weswegen Erkenntnisse über die Länge und Stärke der Verdeckung auf den Verlauf der Lautheit übertragen werden können.

Moore und Glasberg bauen in ihrem deutlich erweiterten Hörmodell ([MGB97], [MO98],

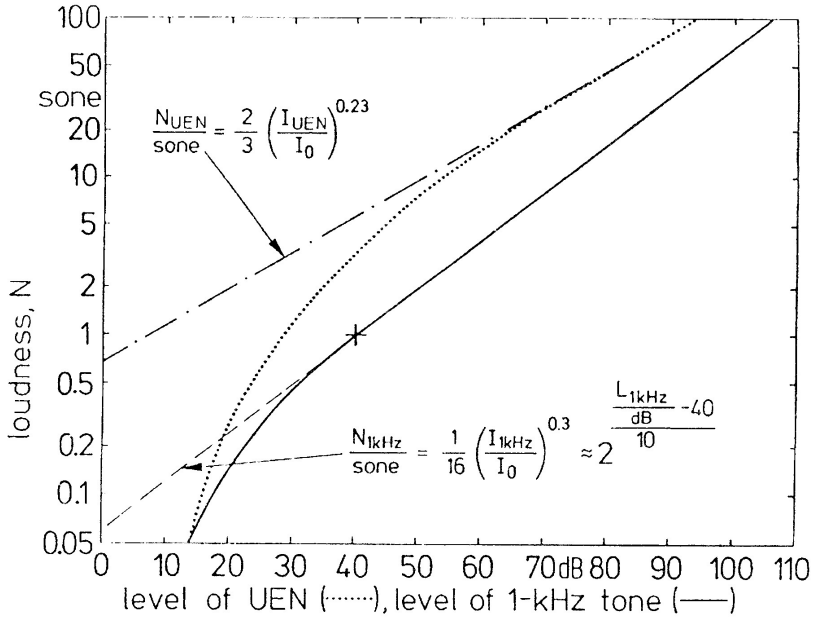

Abbildung 2.15: Vergleich der Lautheit von Rauschen und einem Sinuston [FZ07]

[GM02], [GM06], [MG07]) auf Zwickers Arbeit auf, verwenden jedoch neueste Erkenntnisse wie zum Beispiel die aus einer erneuten Ermittlung der kritischen Bandbreite hervorgegangenen „Equivalent Rectangular Bandwith"-Filter (ERB) statt der Barkfilter, oder eine neue, deutlich aufwendigere statisch nichtlineare Transformationsfunktion von Schalldruck zu spezifischer Lautheit [*Sone*]. Diese berücksichtigt auch sehr elaboriert simultane Verdeckungseffekte (mehr Details und Hintergründe hierzu zum Beispiel in [WVB86]). Die zeitliche Integration findet nun mit einem an die tatsächliche Verdeckung angelehnten Tiefpaßfilter (zwei Zeitkonstanten: $20ms$ für Onset, $50ms$ für Offset) vor der Summierung zur globalen Lautheit aus den Bändern statt. Einige Arbeiten ([OMV97], [POD06]) stellen fest, das eine statische Nichtlinearität am Ende nach der temporalen Integration und damit auch der Ermittlung der Nachverdeckungseffekte zur Simulation des nichtlinearen Verhaltens der Cochlea nicht exakt ist. Deswegen wird die Nichtlinearität auf eine möglichst kurz gefensterte Zwischengröße, hier „Instantaneous Loudness" genannt angewandt und erst anschlie-

ßend der genaue temporale Verlauf jedes Bandes ermittelt. Außerdem findet sich auch eine Erweiterung auf Binaurale Signale. Das Modell ist auch in der Lage, die Hörbarkeit von Signalen in Hintergrundgeräuschen relativ gut vorherzusagen. Als Nachteil ist seine Komplexität und der hohe Berechnungsaufwand zu nennen sowie die eingeschränkte Invertierbarkeit der Schalldruck-Sone Transformation. Einige wohl durch die dynamische Nichtlinearität in der Cochlea hervorgerufenen Erscheinungen wie zum Beispiel das sogenannte „Overshoot-Masking“ (mehr dazu im nächsten Kapitel) sind aber auch mit diesem Modell nicht exakt beschreibbar.

Zeiteigenschaften Gehör, Verdeckung

In der verlustbehafteten Audiokodierung („Mp3“) bilden Verdeckungsmodelle die Basis der Datenreduktion([Joh88b], [Joh88a], [PS00], [Zöl05]). Durch sie werden bei anliegendem Signal laufend die frequenzabhängigen Hörschwellen für additives Quantisierungsrauschen ermittelt. Ist dies getan, so wird das Signal in Frequenzbänder aufgeteilt und diese in der minimal nötigen digitalen Bitrate kodiert. Es wird dabei hauptsächlich die Simultanverdeckung benutzt, d.h. die Verdeckung von parallelen Ereignissen im Frequenzbereich durch die lauteren sogenannten Maskierer. In selteneren Fällen wird auch auf die Nachverdeckung, also die Verdeckung von leiseren Ereignissen zeitlich nach dem Maskierer oder sogar die Vorverdeckung von Ereignissen vor dem Maskierer zurückgegriffen ([GD97], [HC02], [ZK02], [WSKH05], [Gun07]).
In dieser Arbeit geht es hingegen in erster Linie um eine Bearbeitung im Zeitbereich, der Schwerpunkt liegt somit zuerst auf der Vor- und Nachverdeckung und dann erst der Simultanverdeckung.
Der prinzipielle Nutzen eines solchen Modells hierfür liegt auf der Hand: In verdeckten Bereichen lässt sich gut regeln beziehungsweise modulieren, da der Signalanteil auf den diese Bearbeitung erfolgt, auch nicht hörbar ist. Sofern nun keine Ereignisse in unverdeckte Bereiche verschoben werden (durch zu starke Amplitudenanhebung aus der Nachverdeckung oder

durch spektrale Verbreiterung mittels Amplitudenmodulation aus der Simultanverdeckung), bleibt die Bearbeitung dieser Anteile damit auch kaum hörbar bzw. verändert das wahrgenommene Bild weniger. In Betrachtung von Abbildung 2.16 lassen sich bereits ein paar

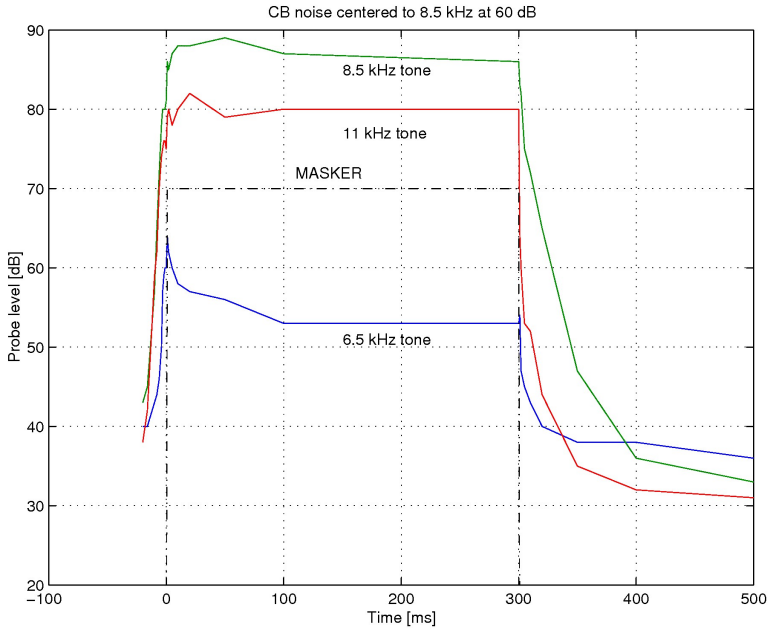

Abbildung 2.16: Temporale Verdeckung von Rauschen in Barkbandbreite mit der Zentrumsfrequenz von $8kHz$, Testtonlänge $= 1\ ms$ [Här99c], [Fas77a]

wesentliche Kennzeichen der temporalen Verdeckung erkennen. Als Maskierer diente hier breitbandiges Rauschen. Als Testton wurde ein $1ms$ langer Sinuston von $8kHz$ verwendet. Die notierte Lautstärke gibt an, ab wann dieser hörbar ist. Die Tatsache, dass die Lautstärke des Testtons oft sogar deutlich über der des Maskierers liegt, erklärt sich aus der kurzen Dauer des Testtones. Betrachtet man die Simultanverdeckung während der Maskierer anliegt, so fällt ins Auge, dass die Hörschwelle bei dem höheren Ton ($11kHz$) deutlich höher als bei dem tieferen Ton, der sich in gleichem Frequenzabstand unterhalb des Maskierers befindet, liegt. Tiefe Ereignisse verdecken also höhere besser als umgekehrt. Die frequenzabhängige Sensitivität des Ohres trägt in diesem hohen Frequenzbereich zusätzlich noch ein wenig bei (vgl.

Abbildung 2.12, maximal hier etwa $10dB$). Die höchste Verdeckung ergibt sich natürlich bei der Mittenfrequenz des Rauschens. Eine reproduzierbare Besonderheit lässt sich aber bei der Verdeckungskurve des tieferen Tones beobachten. Wird der Maskierer abrupt eingeschaltet, so ist die Wahrnehmungsschwelle kurz nach dessen Einsatz deutlich erhöht. Der Effekt klingt relativ schnell ab und wird in der englischsprachigen Literatur prägnant „Overshoot Masking“ genannt. Bis heute (siehe [KSE+09]) ist dazu trotz einiger Forschungsarbeit ([OM95], [HLK99], [Här99a]) noch keine eindeutige Erklärung und damit auch kein exaktes Modell gefunden worden. Mögliche Ursachen wären aktive Prozesse in der Cochlea oder auch neurale Vorgänge.

Nach dem Ende des Maskierers ist eine vergleichsweise langsam (*maximal* $200ms$) abfallende blinde Zone zu erkennen. Hierbei handelt es sich um die Nachverdeckung („Forward Masking“). Sie kommt bei allen gesunden Hörern mit verhältnismäßig geringer Streuung im Verlauf vor. Zu diesem Thema gibt es sehr umfangreiche Literatur (zum Beispiel [Moo93], [Moo96], [Fas76], [Fas77a], [Fas79], [Fas77b], [Oxe98], [JBL82]). Die generelle Kurvenform lässt sich gut durch exponentiell abfallende Prozesse in dem logarithmischen Maß *Phon* beschreiben (siehe auch [Nov99], [Gun07], [GA06]). Eine längere Dauer des Maskierers erhöht den Maskierungseffekt. Ein längerer Testton wird dagegen besser erkannt [OP00]. Auch hier treten einige Effekte zutage, die ihre Ursache in Nichtlinearitäten im Ohr haben ([MO98], [PO98], [OS06]). Die Kurvenform ist zum Beispiel bei sehr leisen ($< 35dB$) Maskierern und auch Signalen sehr flach und wird erst bei größeren Lautstärken konvex. Überraschend ist auf den ersten Blick die vordere Flanke des Maskierers. Bis etwa $20ms$ vor dem Onset wird der Testton von dem zeitlich später kommenden Rauschen verdeckt. Die Erklärungen für diese Vorverdeckung („Backward Masking“) sind vielfältig und reichen von verschiedenen Laufzeiten in der Cochlea [DS84] bis hin zu höher angesiedelten kognitiven Prozessen. Erwähnenswert ist, dass die ermittelten Kurvenverläufe auch bei ein und demselben Probanden eine große Streuung besitzen und dass sich durch Übung der Vorverdeckungseffekt verringern lässt ([Fas76], [Här99c]). Auch die Vorverdeckung nimmt mit der Länge des Mas-

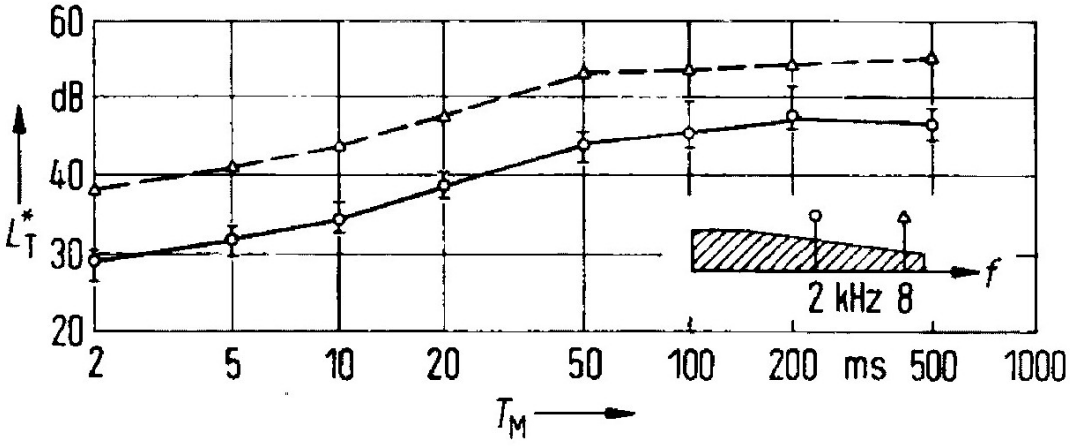

Abbildung 2.17: Nachverdeckung von Breitbandrauschen: Abhängigkeit von der Maskierlänge T_M; Frequenz des zu detektierenden Testtones: $2kHz$(Kreise) und $8kHz$(Dreiecke), Testtonlänge: $1ms$; Aufgenommen $10ms$ nach dem Ende des Rauschens [Fas76]

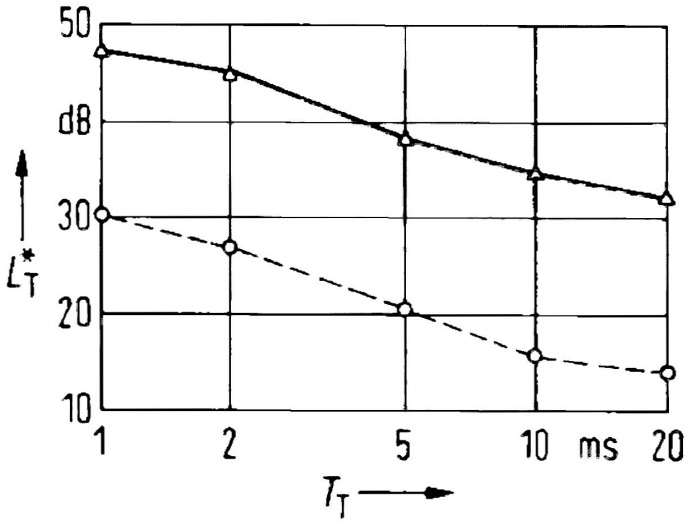

Abbildung 2.18: Nachverdeckung von Breitbandrauschen: Abhängigkeit von der Länge des Testtones T_T; Testtonfrequenz: $8kHz$; Aufgenommen $20ms$ nach dem Ende des Rauschens; Strichliert: Hörschwelle ohne Maskierer in Stille [Fas76]

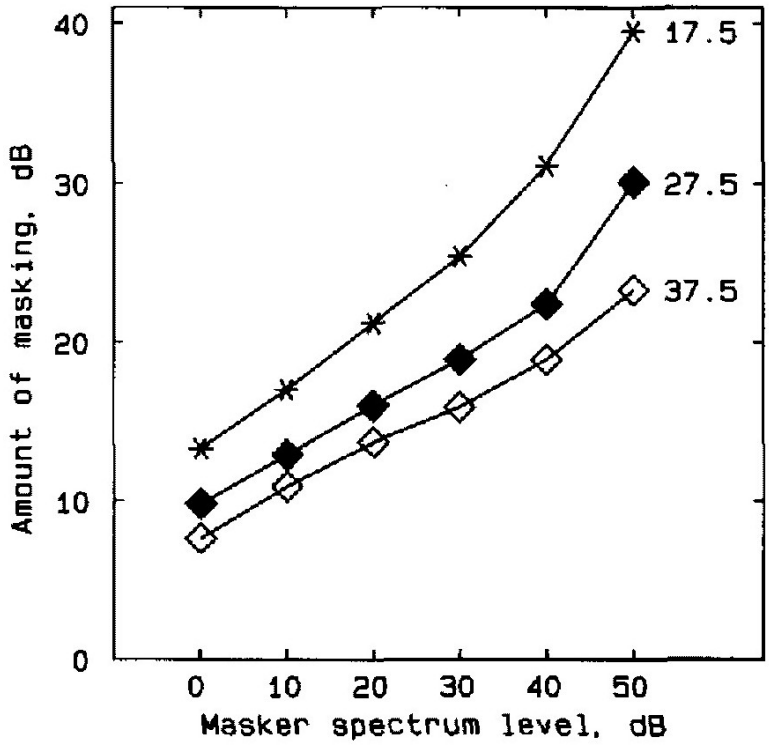

Abbildung 2.19: Verlauf der Kurvenform der Nachverdeckung in Abhängigkeit vom absoluten Pegel des Maskierers („Masker Spectrum Level"); Signalfrequenz: $4kHz$, drei Kurven zu unterschiedlichen Delays nach dem Ende des Maskierers [MO98]

kierers zu und der Länge des Testtones ab.

Die Vor- und Nachverdeckung verringert somit auch die mögliche zeitliche Auflösung. Deren Kennwerte wurden ergänzend in mehreren Studien untersucht [IP82], [Oxe97],[HF99].

Ein hier für die Kompression außerdem wichtiger Teil ist die Wahrnehmbarkeit von Modulationen. Die prinzipiellen Größenordnungen bei statischen Signalen sind in der Standardliteratur ([Moo03], [FZ07] [MHY+95]) gut aufgeführt. In [LB97] wird untersucht, ob die Wahrnehmungsschwellen von Amplitudenmodulationen kurz nach dem Einsatz eines Sinusträgersignals sich gegenüber dem eingeschwungenen Testfall verändern. Tatsächlich sind sie in der Nähe des Onsets deutlich höher. Je höher die Modulationsfrequenz, desto schneller

klingt dieser Effekt ab.

2.2 Implementierungspraxis von Dynamikprozessoren

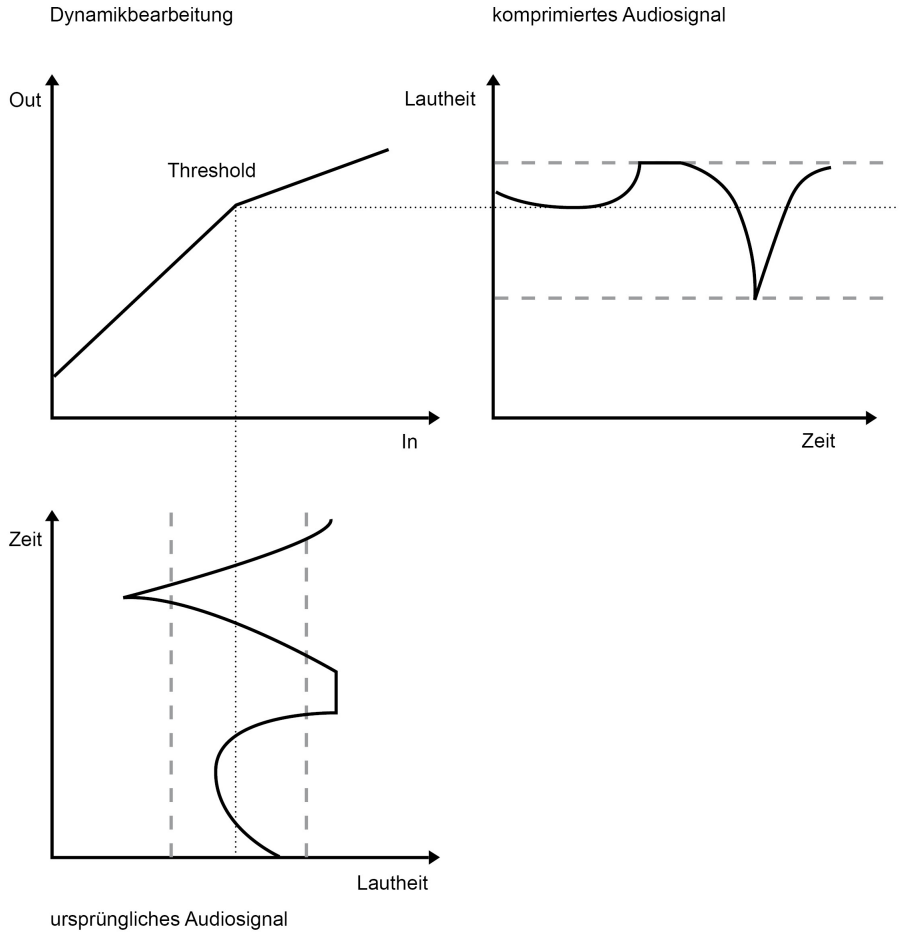

Abbildung 2.20: Grundsätzliches Prinzip der Dynamikkompression

Das grundsätzliche Vorgehen bei der Komprimierung ist in Abbildung 2.20 skizziert. Über einem Grenzwert, dem „Threshold“, wird die Steigerung der Lautheit im Vergleich zum Orginal reduziert. Lautheitserhöhungen im Orginal sollen nur mit dem Faktor $r = 1/R$ an die Ausgabe weitergereicht werden. Bei einer Ratio R von 2 der üblichen Nomenklatura

hat die Übertragungsgerade eine Steigung r von 1/2. Ein Lautheitszuwachs von 6 *Phon* im Orginal hätte in der Ausgabe nur eine Steigerung von 3 *Phon* zur Folge. Die praktische Umsetzung dieses Vorhaben wird in Folgendem diskutiert.

2.2.1 Allgemeiner Aufbau

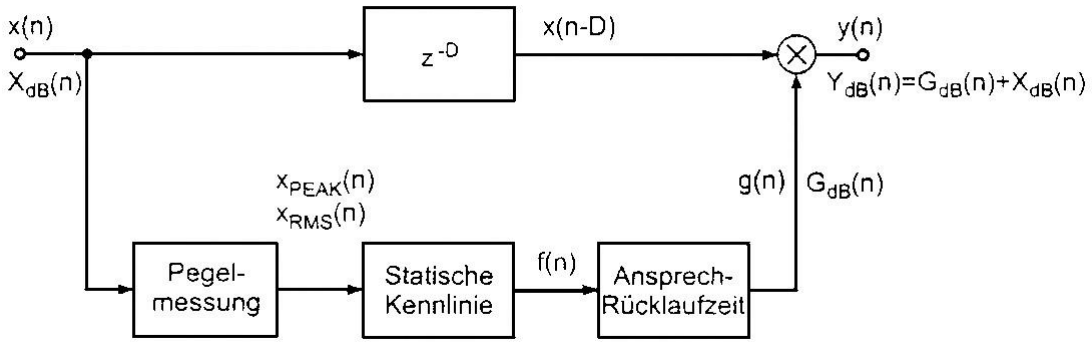

Abbildung 2.21: Prinzipieller Aufbau Kompressor [ZAA+02]

In Abbildung 2.21 wird das grundsätzliche Verarbeitungsschema der heute üblichen Kompressoren dargestellt. Das Eingangssignal wird mittels eines Steuerfaktors (auch Verstärkung, englisch „Gain“) derart moduliert, dass die Dynamik der Lautstärke verkleinert wird. Der Unterschied zwischen lauten und leisen Stellen wird verringert, indem bei lauten Passagen kurzzeitig die Verstärkung reduziert wird (siehe auch Abbildung 2.22). Der Steuerfaktor $g(n)$ („Gain“) wird in der sogenannten „Sidechain“ folgendermaßen ermittelt:
Am Beginn steht eine Pegelmessung. Soll eine Bearbeitung der Lautstärke vorgenommen werden, so wird hier ein RMS-Pegel ausgegeben, der diese möglichst genau approximiert. Dieser wird häufig analog zu einer RMS/SPL-Messung mit einem rekursiven Tiefpaß-Integrator 1. Ordnung nach einer Quadrierung des Signals ermittelt. Eine andere Auslegung zu einem so genannten „Limiter“ zur Beschränkung von kurzen Signalspitzen ist durch Benutzung sehr kurzer Integrationszeiten möglich.

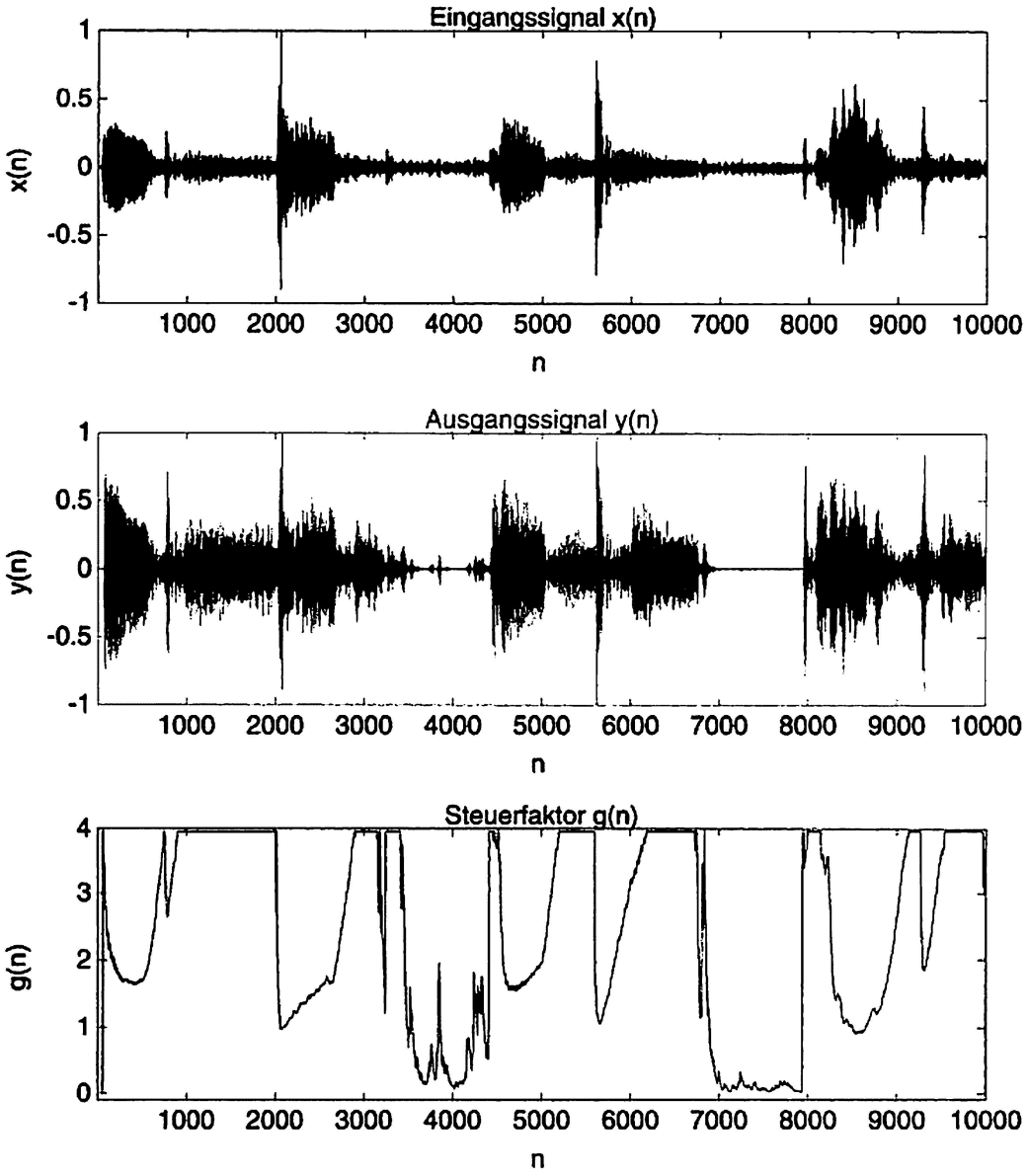

Abbildung 2.22: Zeitsignale $x(n)$, $y(n)$ und Steuerfaktor $g(n)$ [Zöl05]

In dem Modul „Statische Kennlinie“ wird nun die Abweichung des Ist-Wertes vom Soll-Wert ermittelt und diese Differenz als Ergebnis ausgegeben. Der Benutzer stellt die gewünschten Verhältnisse von Ist zu Soll meist mit Hilfe von zwei Größen, nämlich dem „Threshold“(CT)

und der „Ratio“(R) ein. Der Threshold dient der Bestimmung des Pegels, ab dem sich die Lautstärke des Ausgabesignales verringern soll. Die Ratio bestimmt die Steigung des Verhältnisses Ist/Soll ab dem „Threshold“. Bei einem direkten, harten Übergang von der Geraden unterhalb des Thresholds zu einer Geraden mit anderer Steigung oberhalb des Thresholds, spricht man auch von einem „Hard Knee“. Bei dem „Soft Knee“ ist der Übergang um den Threshold herum sanfter ausgeführt. Würde das nun gewonnene Korrektursignal

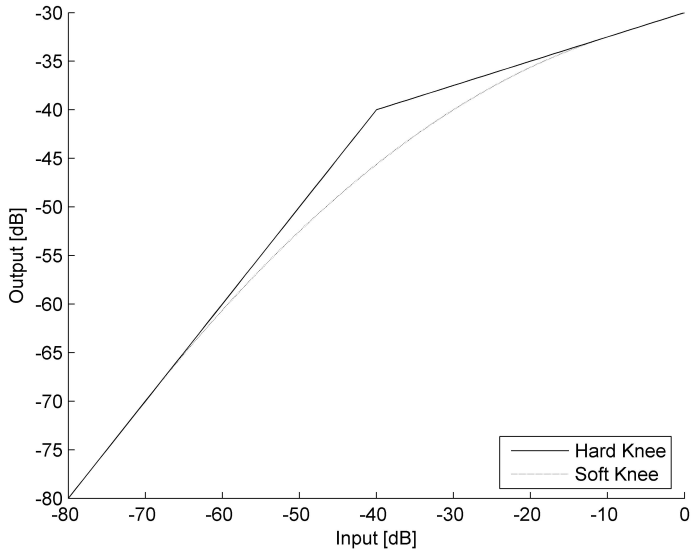

Abbildung 2.23: Vergleich Hard-/Softknee

direkt auf das Signal aufmultipliziert, so ergäben sich sehr unnatürliche und hörbare Signalveränderungen beim Überschreiten des Thresholds. Es wird deshalb noch mit einem rekursiven Tiefpass-Integrator 1. Ordnung geglättet. Dieser besitzt je nachdem, ob das Korrektursignal (und damit auch der Level des Eingangssignals) steigt oder fällt, jeweils eine andere Zeitkonstante. Die kürzere Ansprechzeit („Attack“, üblich sind 5-30ms) wird verwendet für ein langsames Reduzieren des ersten Impulses bei Transienten, die längere Rücklaufzeit („Release“, 50-400 ms) für ein behutsames Aufheben der Verstärkungsreduzierung während des Ausklangs.

Schlussendlich wird das eventuell um D Samples verzögerte Eingangssignal $x(n)$ mit dem Steuerfaktor $g(n)$ multipliziert. Die Verzögerung kompensiert die durch die rekursiven Tiefpässe in der Pegelmessung und dem Ansprech- und Rücklaufmechanismus auftretenden großen Gruppenlaufzeit der Einhüllenden (sehr tiefer Frequenzbereich). Ausserdem ermöglicht sie es, den Regelvorgang vor das tatsächliche Eintreffen eines Transienten zu legen und somit die Vorverdeckung auszunutzen. Sind die Impulse jedoch deutlich kürzer als die Verzögerung, so regelt das System vor dem Impuls und ist beim tatsächlichen Eintreffen desselben bereits wieder ausgeschwungen. Dieses Vorgehen ist darum nicht unproblematisch, die Verzögerung muss immer wieder aufs neue an wechselnde Eingangsignale angepasst werden.

2.2.2 Genauere Betrachtung und praktische Lösungen

Die Pegelmessung dient also zur Bestimmung des Ist-Zustandes, die Statische Kennlinie bestimmt die Abweichung vom Soll, und der Ansprech- und Rücklaufmechanismus bestimmt den zeitlichen Verlauf der Regelung.
Die Aufgabenverteilung ist leider nicht so vollkommen getrennt wie oben geschildert. Da der Kompressor primär die Lautstärke und nicht die zeitliche Feinstruktur verändern soll, müsste die Pegelmessung mit vergleichsweise großen Zeitkonstanten (Moore benutzt in seinem Lautstärkemessmodell [GM02] 20 ms für den Onset) oder einem entsprechend langen FIR-Fenster geschehen. Dies macht die Regelung aber relativ träge. Es können mehrere Millisekunden nach dem Onset eines scharfen Transienten vergehen, bis der gemessene Pegel den Threshold erreicht. Erst ab diesem Zeitpunkt schlägt das Korrektursignal aus und der Regelvorgang beginnt.
In Abbildung 2.24 ist ein solcher Onset skizziert. Gut erkennbar ist das verspätete Einsetzen der Regelung selbst unter idealen Bedingungen. Hier sind somit Millisekunden vergangen, in denen eine Pegelreduktion weniger wahrnehmbar durchführbar gewesen wäre (siehe Kapitel 2.1). Eine kürzere Zeitkonstante in der RMS-Ermittlung ist eine gangbare Lösung, jedoch wird die Balance zwischen verschiedenen Signalanteilen (harmonisch/statisch/transient.) da-

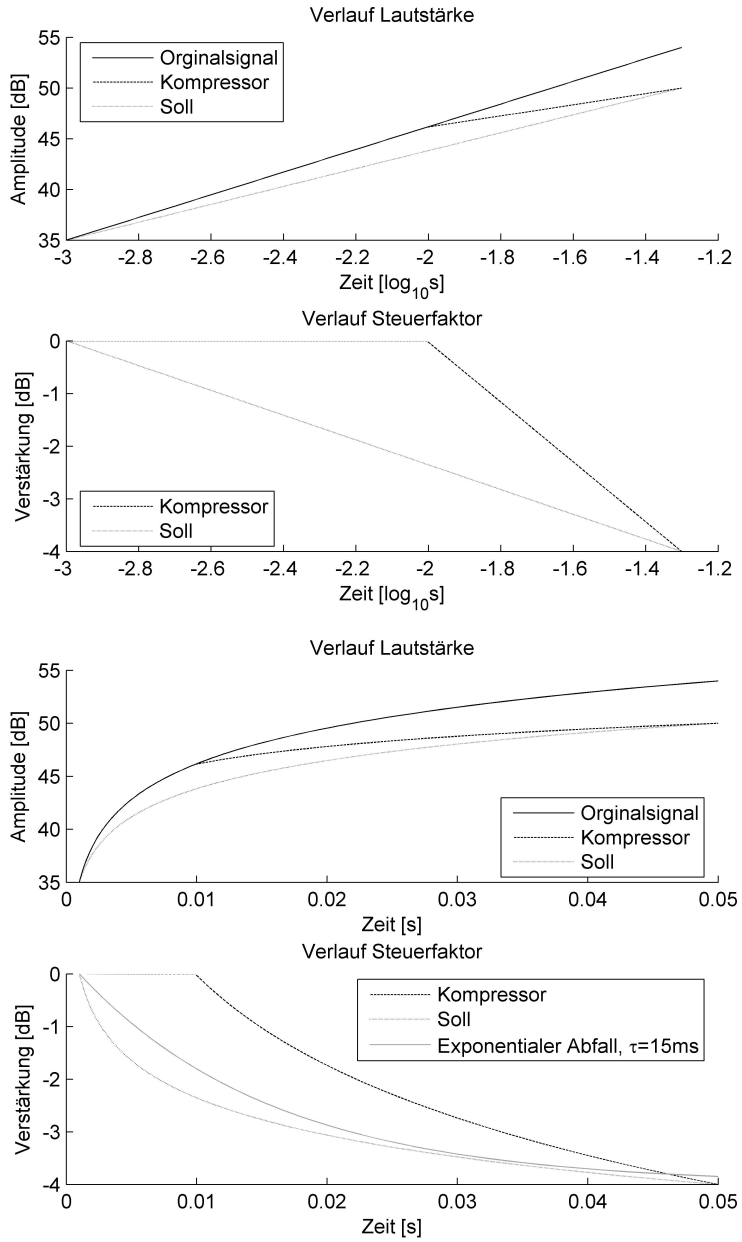

Abbildung 2.24: Onset Rauschen (von 36 auf 54 dB): Verlauf Lautstärke (Daten von Abbildung 2.17) und Steuerfaktor Kompressor (Threshold= 46 dB, Ratio= 2 : 1)

durch verändert. Bei einer Gitarre zum Beispiel wird das Verhältnis der Zupfgeräusche zu dem harmonischen Klang verändert. Bei ändernden Inhalt müssen die Einstellungen im-

mer wieder nachjustiert werden. Um diese Problematik zu beheben, sind bei modernen High-End-Kompressoren noch viele zusätzliche Funktionen wie etwa mehrstufige Release-Kurven, kombinierte Spitzenwert- und RMS-Pegelmesser, Hold-Funktionen und viele Einstellmöglichkeiten zur statischen Kurvenform eingebaut ([Saf03], [Saf07], [Mas08], [dbx08], [Ble69]). Ergänzend wird nach dem Kompressor meist noch ein Limiter zum Abfangen der durch die Verzögerung nicht bearbeiteten Transienten in Serie geschaltet. Dessen Regelkonstanten sind merklich kürzer, der Threshold dafür höher. Die optimale Einstellung dieser Geräte wird dadurch jedoch zunehmend schwerer, zusätzlich sind die Einstellungen (statische Kurve, dynamisches Verhalten, etc) nicht unabhängig voneinander. Anhand Abbildung 2.25 lassen sich diese Nachteile gut studieren. Es handelt sich hierbei um einen sehr kurzen Ausschnitt eines Jazztrios. Der Einsatz bei $0,3$ Sekunden ist eine singende, stehende Bassnote, der starke Impuls bei $0,475$ Sekunden ein Snareschlag. Beide unterscheiden sich lediglich durch die Zeitkonstante zur Ermittelung der RMS. Bei der kürzeren Zeitkonstante wird zum Beispiel der liegende Basston ab $0,3s$ stärker heruntergedrückt. Auffällig ist auch in beiden Fällen, wie spät die Verstärkungs-Reduktion beginnt. Der Einsatz des Basstones bleibt unbehandelt, der weitere eigentlich statisch stehende Ton wird dagegen verringert. Unschön ist auch die stufige Reduktion, die immer nach einer Wellenlänge erfolgt. Die Einhüllende ist nicht sauber extrahiert, es findet sich noch mindestens die Grundfrequenz in der Pegelanalyse wieder.

2.2.3 Auswirkungen von Kompression

Klangliche Auswirkungen

Wie im vorigen Kapitel angedeutet, kann die Anwendung eines Kompressors das Signal signifikant verändern. Der Einsatz ist in der Praxis nicht nur technischer Natur zur Anpassung der Wiedergabe an die verfügbare SNR, sondern wird in manchen Musikstilen wie der heutigen Popmusik auch massiv zur ästhetischen Gestaltung benutzt. Transienten können betont

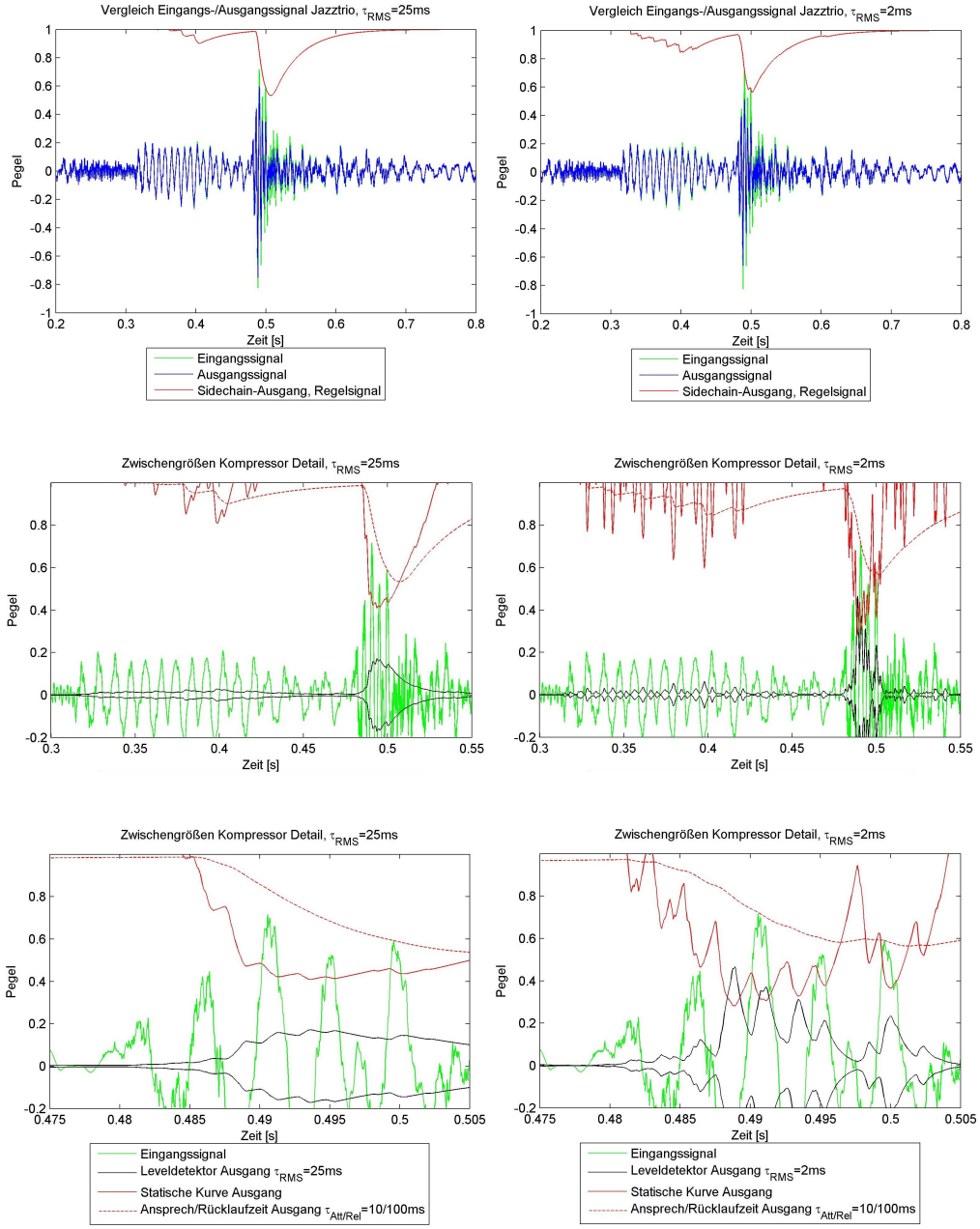

Abbildung 2.25: Vergleich Verschiedener Zeitkonstanten $\tau_{RMS} = 0,25(links)/0.02(rechts)$, Threshold= $-18dB$, Ratio=1/4

oder auch reduziert werden, was bei der prinzipiell üblichen aber unnatürlichen Nahmikrofonierung zum Beispiel von Gesang häufig getan wird. Bei einer solchen nahen Abnahme

aus unter einem Meter Entfernung sind die Konsonanten noch nicht durch die Raumimpulsantwort verschliffen und somit überdeutlich und scharf. Der Sänger wird mittels des Kompressors und des nachfolgendes Halles künstlich auf Distanz gebracht, ohne deutlich an Brillianz und Durchsichtigkeit zu verlieren. Weiterhin ist es möglich, bei geringer verfügbarer Dynamik laute und leise Ausdrucksformen zu nutzen. Die Hörerfahrung ermöglicht es dem Zuhörer aus dem Spektrum und seinem Verlauf zu entnehmen, ob die Artikulation laut oder leise ist. Ist die Dynamik selbst jedoch ein wichtiges ästhetisches Ausdrucksmittel wie in der klassischen Musik, so ist dies natürlich nicht sinnvoll.
Bemerkenswert sind die spektralen Auswirkungen von Kompression. Selbst ein einkanaliger Kompressor verändert den Klang teils deutlich. Komprimierte Sprecher zum Beispiel klingen nach der Kompression oft voluminöser aber manchmal auch schärfer. Der Effekt lässt sich dadurch erklären, dass bei gleicher durchschnittlicher Lautstärke, leisere Signalanteile angehoben werden und damit deren hoch- und tieffrequenten Anteile nun besser wahrgenommen werden. Aus den Kurven gleicher Lautstärke ist dieser Effekt gut ersichtlich.
Aus den Dolby Labs gibt es hierzu ein praktisch orientiertes Konzept, die ursprüngliche spektrale Balance bei Lautstärkenveränderungen und Dynamikbearbeitungen („Automatic Levelling) in Abhängigkeit von der Wiedergabelautheit beizubehalten [See07]. Hierfür wird das Signal mit Hilfe einer Filterbank in gehörrichtige Barkbänder zerlegt und die Lautheit in Sone ermittelt. Sämtliche Bearbeitungen geschehen nun auf diese Lautheit. Schließlich wird aus der Veränderung der Lautheit in Sone auf die Verstärkungsänderung der einzelnen Bänder in der linearen Domäne zurückgerechnet. Die Bänder werden mit diesen Gain-Faktoren wieder resynthetisiert.

Auswirkungen im Zeitbereich

Eine typische Auswirkung auf den zeitlichen Verlauf des Signals ist das sogenannte „Pumpen“. Er bezeichnet das hörbare Erhöhen der Verstärkung nach einer Reduktion durch einen

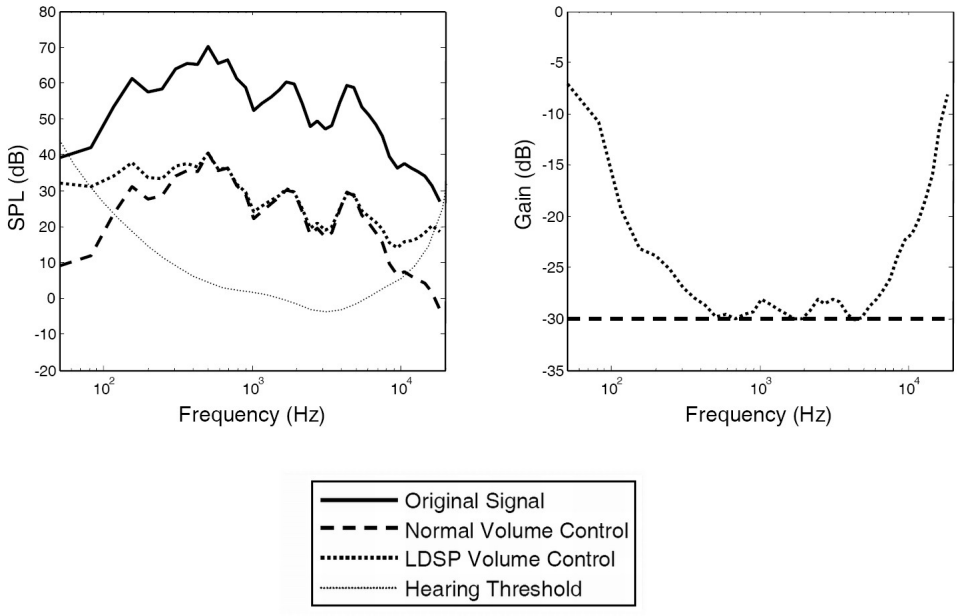

Abbildung 2.26: Lautheitskorregierte Pegelabsenkung [See07]

vorhergehenden Impuls. Prinzipiell entsteht er dadurch, dass die Rücklaufzeit zu lang gewählt ist, um den Regelprozess durch den Impuls zu verdecken. Andererseits ist er aber zu kurz, als dass die Pegeländerung pro Milisekunde unbemerkbar wird. Bei Sprache scheint eine unnatürliche Atmung zu entstehen, beim Vorhandensein von geräuschhaften Schallquellen im Hintergrund (z.B. Hihat, Becken, „Ambience“) scheinen diese nach einem Ereignis im Vordergrund aufzuklingen. Die mögliche Verkürzung der Rücklaufzeit ist abhängig von der Art der im Signal enthaltenen Instrumente und dem Anteil der Räumlichkeit. Bei einer fertigen Mischung mit vielen Elementen (z.B. einer Bigband) ist nur ein Kompromiss möglich. Bei vorhandenem starken Hallanteil und definierter Räumlichkeit besteht die Gefahr bei kurzen Attack- und Releasezeiten die wahrgenommene Balance zwischen Direktschall und Raumantwort zu verändern, da der lautere Direktschallanteil stärker komprimiert wird als der darauffolgende Nachhall. Wenn letzterer auftritt ist die verringerte Verstärkung bereits

wieder aufgehoben. Dies kann zum Beispiel zu einer kurzzeitigen Wahrnehmung von größerer Distanz nach starken und damit lautstärkenreduzierten Impulsen führen. Ist die Rücklaufzeit zu lang, ist die erhaltene Kompression sehr gering: die Verstärkung geht von einem Impuls zum nächsten kaum mehr auf den Ausgangswert zurück, sondern bleibt relativ statisch auf dem reduzierten Wert stehen.

Komodulation und spektrale Dekorrelation

Ein unangenehmer Seiteneffekt der Kompression mit einem Steuersignal über die ganze Frequenzbreite („Single Band Compressor“, „Fullband Compressor“) ist die Komodulation. Sie tritt beim Vorhandensein von mehr als einer eigenständigen Quelle im Audiosignal auf. Praktisch ist dies fast immer gegeben, da zumindest immer ein mehr oder weniger wahrnehmbares Grundrauschen auftritt. Nun bestimmt die jeweils lautere den Verlauf der gemessenen Lautstärke und damit des Regel- oder Modulationssignals. Finden beide nun gleichzeitig in jeweils gut voneinander unterscheidbaren und unverdeckten Frequenzbereichen statt, so wird der leisere hörbar mit dem lauteren mitgeregelt. Ein Beispiel wäre eine prominente Bassbegleitung zu einem hohen Solisten: Jedes mal, wenn der Bassist eine tiefe und sehr laute Note spielt, wird der Solist kurz leiser. Es entsteht ein nicht zu seiner Linie und Betonung passendes Flackern. Meist ist dies äußerst unmusikalisch und unnatürlich, vor allem wenn beide Streams eine andere Artikulation oder Betonung haben. Eine sehr offensichtliche Komodulation tritt unter Verwendung von Fullband-Kompressoren bei starken Impulsen im Bassbereich, zum Beispiel bei Basedrum-Anschlägen auf. Diese besitzen sehr viel Energie und dominieren damit auch spektral, werden aber bei normalen Abhörlautstärken leiser wahrgenommen (aus den Kurven gleicher Lautstärke ersichtlich). Bei üblicher RMS-Mittelung dominieren nun diese Tiefbässe das Pegelsignal, in der Folge wird alles kurz heruntergefahren. Das gesamte Signal wird also immer wieder ohne Grund im restlichen Audio teils massiv heruntergeregelt. Die Ursache erscheint aber vom Höreindruck eigentlich unwichtig. In der Praxis wird vor den Pegelmesser ein Hochpaßfilter ähnlich der dB(B)-Gewichtung eingefügt.

Diese Gewichtung stimmt aber nur bei einer gewissen Eingangslautheit (zum Beispiel 70 Phon) exakt. Hat das Signal eine große Dynamik, so führt diese Gewichtung zu falschen Abschätzungen bei deutlich leiseren (immer noch zu viel Bassanteil in der Pegelmessung) oder lauteren (zu wenig Bassanteil in der Pegelmessung) Stellen.
Diese Problematik kann generell durch Benutzung eines „Multi-Band-Kompressors“ umgangen werden. Dieser besteht nicht aus einem, sondern aus üblicherweise 3 bis 5 Regelkreisen, bei dem jeder für einen eigenen, abgetrennten Frequenzbereich zuständig ist. Diese einzelnen Bänder werden am Beginn mittels einer Filterbank getrennt. Die Bearbeitung geschieht unabhängig voneinander anhand jeweils eigener statischer Kennlinien sowie Ansprech- und Rücklaufzeiten. Letztendlich werden diese Bänder wieder zusammengesetzt. Das Problem der Komodulation ist in diesem Fall kaum mehr gegeben, außerdem kann eine spektrale Loudnessangleichung vorgenommen werden, d.h. die Bässe und Höhen können stärker komprimiert werden als die besser hörbaren Mitten. Zusätzlich zu der relativ schwierigen Bedienung (jedes Band muss gegebenenfalls einzeln optimiert werden) ergibt sich jedoch das gegenteilige Problem. Signale verlieren Ihren spektralen Kontrast und damit Ausdruck. Unterschiede zwischen laut (prominentere Obertonstruktur) und leise (stärkerer Abfall der Obertöne zu hohen Frequenzen hin) gespielten Instrumenten werden kleiner, da das Spektrum insgesamt flacher wird. Auch die Durchsichtigkeit oder Verständlichkeit (bei Sprache) kann leiden, da eigentlich zu einem Instrument gehörende Signalanteile in verschiedenen Frequenzbereichen nicht mehr so stark zusammen steigen und fallen.
Weitere Details zu diesem Thema finden sich in Artikeln von Moore und Stone ([MPS99], [SMAG99], [SM03], [SM04], [SM07], [SM08]). Sie beschäftigen sich mit der Optimierung von Kompressoren für Hörgeräte sowie Cochlea-Implantate. Die Beeinträchtigung der Sprachverständlichkeit in Präsenz von Hintergrundgeräuschen durch Komodulation (Einbandkompression) wird ebenso diskutiert und quantifiziert, wie die Verschlechterung derselben durch das Abflachen der frequenziellen Struktur in Multibandkompressoren.

Aliasing

Die meisten qualitativ hochwertigen Limiter, aber auch Kompressoren benutzen heutzutage ein mindestens zweifaches Oversampling, um Aliasing zu vermeiden. In [Map98] befindet sich hierfür eine Worst-Case-Abschätzung.
Hier nun ein weiterer Erklärungsversuch: Das Steuersignal verändert sich zum Zeitpunkt eines Impulses sehr schnell und sprunghaft von „1“ auf einen Wert „< 1“ wenn der Threshold überschritten wurde. Dies erzeugt kurzfristig ein breites Frequenzspektrum im Steuer-/Regelsignal. Der darauffolgende Glättungsfilter („Attack-Release“) ist normalerweise nur 1. Ordnung, dämpft also mit einem Abfall von lediglich $6dB$/Oktave selbst hohe Frequenzen nicht sehr stark. Dieses Steuersignal moduliert nun das Audiosignal. Die beiden sehr breiten Spektren werden im Frequenzbereich gefaltet, es entstehen Seitenbänder mit sehr hohen Frequenzen auch über der Abtastfrequenz. Diese werden wieder in das Spektrum zurückgespiegelt und erzeugen so störendes Aliasing.
Bei Multibandansätzen mit zwischenzeitlichem Downsampling (zum Beispiel mit Polyphasen-Filterbänken) gilt dies besonders. Die Abtastfrequenz ist hier niedriger, das Spektrum des Regelsignals beim Überschreiten des Thresholds aber ähnlich breit.
Beim Vorgang der Rauschunterdrückung scheint diese Problematik geringer zu sein. Der Regelvorgang geschieht dort in leisen, weniger prominenten Passagen. Also finden sich auch die Aliasing-Artefakte hauptsächlich dort.
Ein Mittel zur Verringerung des subjektiv hörbaren Aliasing wäre also ein Verschieben des Regelvorgangs zu leiseren oder vorverdeckten Zeitpunkten.

2.2.4 Anforderungen an einen neuen Ansatz

Aus den vorangehenden theoretischen Erläuterungen und Erfahrungen aus der Praxis lässt sich folgende Anforderungen zusammenstellen:

1. **Genaue Abschätzung der wahrgenommenen Lautheit**

Um möglichst nur die Lautheit verändern zu können, muss genau diese als Referenz erst einmal gewonnen werden. Nicht exakte Abschätzungen führen zu Regelungen zu falschen Zeitpunkten und falschen Intensitäten. Als Bonus sollte auch ein Übergang zwischen kurzfristiger und langfristiger Lautheit als Referenz möglich sein.

2. **Korrektur der spektralen Balance**

   Die spektrale Balance des ursprünglichen Signals soll beibehalten werden, bei angehobenen Stellen müssen also gemäß den Kurven gleicher Lautheit die Bässe und Höhen abgesenkt werden.

3. **Exaktes Zeitverhalten („Timing“) der Regelprozesse**

   Die Regelung soll bei Transienten genau an deren Beginn („Onset“) starten und an deren Ende dem durch die statische Kennlinie vorgegebenen Zielwert möglichst nahe kommen. Das Ende ist entweder durch den Übergang in einen eher statischen Zustand (zum Beispiel gehaltenen Töne) oder in einen Abfall der Lautheit (Ausklang, „Offset“) gekennzeichnet. Beim Ausklang gilt dasselbe Prinzip. In den statischen Phasen genügt eine Tiefpassfilterung durch den Ansprech- und Rücklaufmechanismus wie bisher. Dadurch werden die Amplitudenmodulationen so gering gehalten, dass sie nicht gehört werden.

4. **Regelverhalten möglichst unauffällig**

   Da der Startpunkt und Endpunkt des Regelvorgangs nun feststeht, stellt sich die Frage, wie dieser Weg zurückgelegt werden soll. Aus theoretischen Überlegungen erscheinen Exponentialkurven in der logrithmischen Domäne, also dB, am günstigsten. Dies gilt sowohl für den Onset (siehe auch Abbildung 2.24) wie auch den Offset. Da die Nachverdeckungskurven einen ähnlichen Verlauf haben, wird auch dann am meisten geregelt, wenn auch die Verdeckung am größten ist. Erfahrungen aus der Praxis konventioneller Kompressoren bestätigen diese These.

5. **Regelung auch in vorverdeckten Bereichen**

   Vor starken Impulsen mit großer Steigung kann relativ sicher von Vorverdeckung ausgegangen werden. Dieser unhörbare Bereich soll zur vorausschauenden Reduktion der Lautstärke genutzt werden.

6. **Vermeiden von Artefakten**

   Durch die digitale Implementierung entstehende Signalbeeinträchtigungen wie etwa starke Phasenverzerrungen durch Filterbänke, vor allem bei keiner Bearbeitung, oder Aliasing durch Modulationen sind auf jeden Fall zu vermeiden.

7. **Effizienz**

   Das System muss ausreichend effizient sein, um in Echtzeit auf einem Computer oder DSP implementiert werden zu können. Die Länge der Verzögerung zwischen Ein- und Ausgabe spielt dagegen keine Rolle.

Die ersten beiden Punkte machen prinzipiell die Bearbeitung in verschiedenen Frequenzbereichen, etwa durch eine Filterbank oder einen adaptiven Filter nötig. Verschiedene Audioereignisse haben im Frequenz- und im Zeitberich jeweils eine unterschiedliche Ausdehnung. Um in jedem Frequenzband auch eine optimale transiente Regelung nach den Punkten 3. bis 5. zu erreichen, müssen diese Bänder jeweils eine eigene Zeitsteuerung haben. Es ergibt sich somit folgendes Bild:
Die Zielvorgabe (Reduktion der Lautheit) wird global formuliert und anschließend gemäß der Kurven gleicher Lautstärke für jedes Band einzeln übersetzt. Die exakte zeitliche Steuerung zum Erreichen des Ziels wird in jedem Band unabhängig voneinander bestimmt.

Kapitel 3

Neuer Ansatz

In diesem Kapitel wird nun das sich aus den Forderungen des vorherigen Kapitels ableitende Konzept vorgestellt. Die beeinhaltenden Elemente werden diskutiert und teilweise ausführlich erläutert.

3.1 Übersicht

Dem hier vorgestellten Lösungsweg liegt wie bei heute üblichen Kompressoren eine Parallelstruktur zu Grunde. Die Analyse und das Erzeugen der Veränderungen läuft in einem getrennten Pfad. Das Ergebnis wird erst am Schluss mittels eines dynamischen Filters auf das Signal angewandt. Hierdurch werden mögliche Signaldegradierungen, zum Beispiel durch das Hin- und Zurückwandeln des Signals in den Zeit-Frequenzbereich mittels einer „Kurzzeit-Fouriertransformation“ (STFT) wie etwa beim Denoising vermieden. Außerdem ermöglicht dies das Heruntersetzen der Abtastfrequenz einzelner Analyseteile und spart so Rechenzeit.

Der erste Schritt im Parallelpfad ist die Ermittlung der spezifischen Lautheit $A_m(n)$ [1] der einzelnen Frequenzgruppen nach der Barkskala in Sone. In dieser Arbeit werden aus später

[1] m bezeichnet das Frequenzband, n die Zeit in Samples

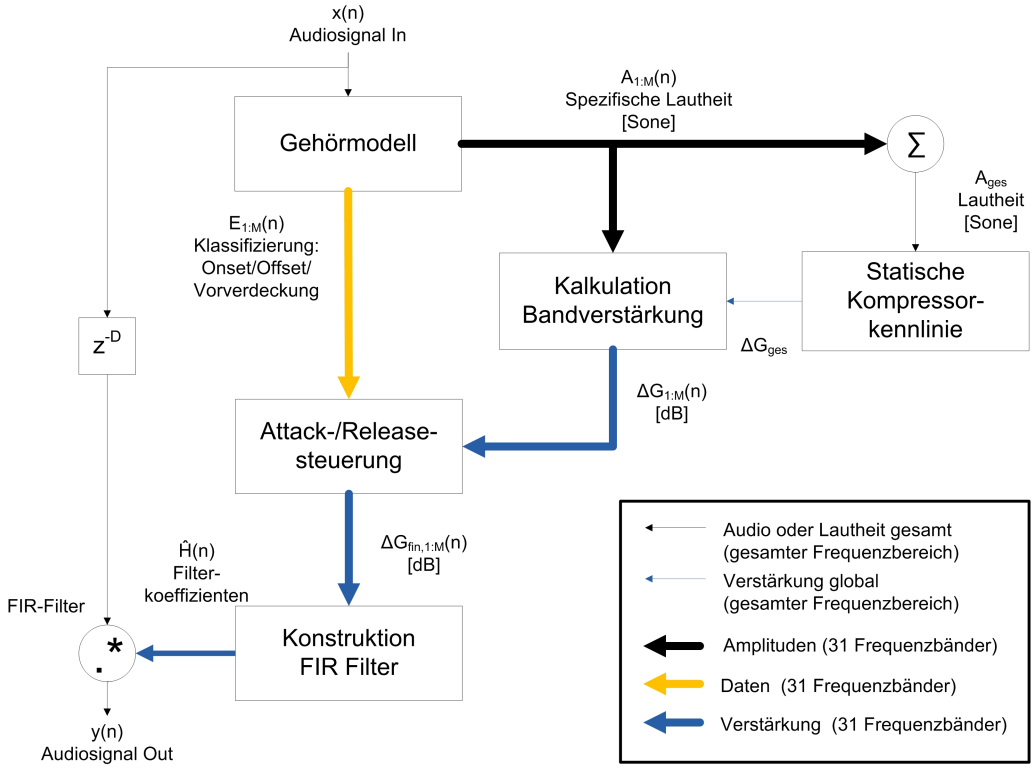

Abbildung 3.1: Gesamtübersicht über die Dynamikbearbeitungsumgebung

erläuterten Gründen 31 statt der üblichen 25 Bänder verwendet, die Auflösung ist also etwas größer. Werden diese addiert, so ergibt sich die kurzfristige Gesamtlautheit $A_{ges}(n)$(ebenfalls in Sone) zum jeweiligen Zeitpunkt.

Mit Hilfe der statischen Kompressorkennlinie wird nun ermittelt, um welchen linearen Faktor die gegenwärtige Lautheit vom vorgegebenen Ziel abweicht ($\Delta G_{Gesamt}(n)$). Aus diesem Faktor und der spezifischen Lautheit der einzelnen Bänder wird die zum jeweiligen Zeitpunkt nötige Steuer- und Korrekturverstärkung für die einzelnen Bänder $\Delta G_{1:M}(n)$ in dB errechnet. Die Attack-/Releasesteuerung geht wie in Kapitel 2.2.4 gefordert, deutlich weiter als Standardkompressoren. Sie passt ihr Regelverhalten mit Hilfe der Klassifizierungsinformationen E_m an die jeweilige Situation an.

Das Ergebnis $\Delta G_{fin,1:M}$ ist nun ein sich jedes Sample aktualisierender Vektor mit den

Verstärkungswerten in dB. Aus diesem Vektor werden nun FIR-Filterkoeffizienten generiert. Das zeitverzögerte Signal wird mit diesen gefiltert und ergibt die Ausgabewerte $y(n)$. Zwei alternative Filtervarianten sind in dieser Arbeit ausgeführt: ein Linear-Phase-Filter sowie ein sehr effizienter minimalphasiger Warped-Filter. Die Zeitverzögerung ist nötig, um die teils großen Latenzen aus der Warped-FFT, der Glättung sowie der Attack-/Releasesteuerung zu kompensieren.

Anmerkung: In den weiteren Erläuterungen der einzelnen Komponenten wird auf eine detaillierte Beschreibung der Zeitkorrektur durch Verzögerung verzichtet. Stattdessen werden zum besseren Verständnis einzelne Elemente akausal eingeführt. Dies ist problemlos kausal implementierbar, da immer eine fixe maximale Voraussicht etwa durch einen FIR-Filter vorliegt.

Generell ist zu erwähnen, dass jedes zu bearbeitende Audiosignal auf eine Abspiellautstärke normiert sein muss, da die Bearbeitung anhand dezidierter, absoluter Lautheitswerte stattfindet. Bei der Implementierung wird in Anlehnung an SMTPE-Richtlinien angenommen, dass 0 dB_{FS} einem Spitzenwert von 105 dB entsprechen. Dies entspricht einer maximalen Lautstärke vom 85 $dB_{SPL}(C)$ mit 20 dB Headroom.

Alternativ zur Multiband-Analyse und -Bearbeitung ist in dieser Arbeit auch noch ein einkanaliger („Single-Band“) Ansatz implementiert. Dieser benötigt natürlich keine Analyse-FFT und auch keinen Filter am Ende, da das gesamte Frequenzspektrum mit einem Verstärkungskoeffizienten geregelt wird. Die Einhüllende wird am Eingang durch die Amplitudenbestimmung des „Analytischen Signals“ gewonnen. Dieses entsteht durch Hinzufügen eines durch eine Hilberttransformation gewonnenen komplexen Teils. Anschließend wird genau wie beim Multibandansatz mit dem von-Hann-Fenster geglättet. Ansonsten ist das System identisch, es handelt sich lediglich immer um $M = 1$ Frequenzband.

3.2 Gehörmodell

Das Gehörmodell ist der zentrale Bestandteil der Dynamikbearbeitung, da sowohl die zeitliche Klassifizierung und Steuerung, wie auch die Ermittlung der statischen Zielvorgaben ihm nachgereiht sind.

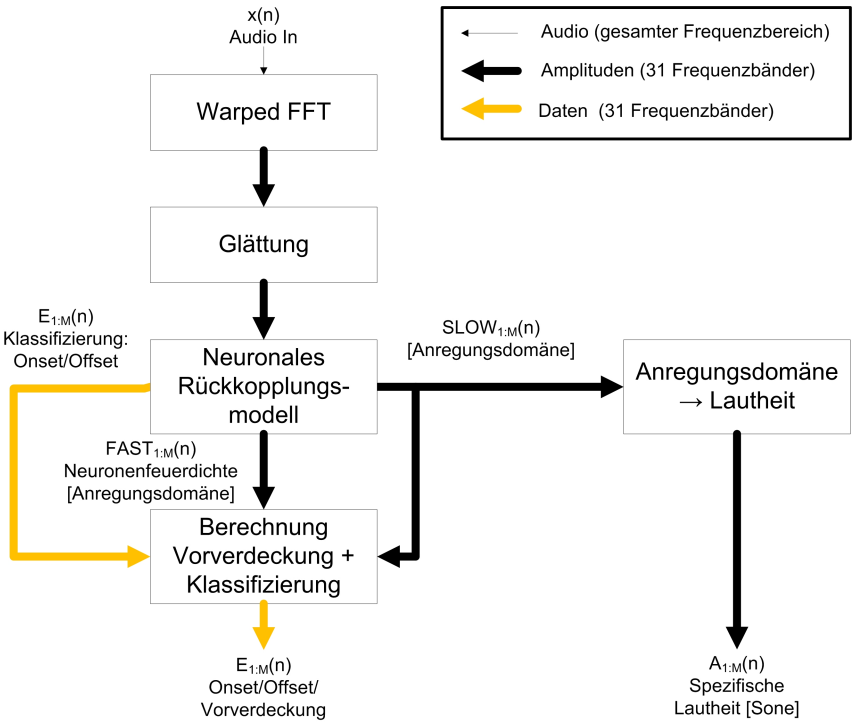

Abbildung 3.2: Übersicht über das Gehörmodell

Der erste Schritt ist die Aufteilung in Frequenzgruppen. Eine weit verbreitete Methode ist das Verwenden von Gammatonfiltern ([LAH01],[LHA01], [LBA06],[Kla08]). Hier wird statt diesem Filterbandansatz die sogenannte „Warped FFT“ verwendet. Der Grund ist folgender: Die direkte Analyse und Synthese durch Gammatonfilter wie zum Beispiel in [KK99a], [KK99b], [Bau97], [Bau02] verbietet sich, da die hier vorkommenden Amplituden- und Phasenverzerrungen die Forderung nach einer fehlerfreien Rekonstruktion nicht erfüllen. Eine

Parallelstruktur mit einer separaten Filterbank oder einem langem, zeitvarianten FIR-Filter zur Audiobearbeitung ist somit nötig. Eine Polyphasen-Filterbank ist fürs Erste hier zu empfindlich für Aliasing. Die auftretenden Modulationen durch das Steuersignal können sehr hohe Frequenzen beinhalten (siehe Kapitel 2.2.1) und somit die Signale in den Bändern spektral deutlich über deren Grenzen verbreitern. In Kombination mit einem Warped-FIR Filter ist eine Warped-FFT deutlich effizienter, weil eine Zwischenstufe (der Inhalt der Warped-Delay-Line) sowohl in der Analyse wie auch im Filter verwendet werden können. Durch das Warping kann der Filter von 512 auf 64 Koeffizienten verkürzt werden.
Nach der FFT und einer Glättung werden die Frequenzbänder zum Simulieren des zeitlichen Lautheitsverlauf durch ein Modell der neuronalen Rückkopplung geschickt. Dieses wurde in [Kar96] erstmals vorgestellt und bildet die aktiven Prozesse der Cochlea nach. Als Ausgabegröße steht in jedem Band die spezifische Lautheit in Sone zur Verfügung. Das transiente Verhalten (Verdeckung, Lautheitsverlauf) lässt sich nun sehr gut dem aus Hörtests bekannten angleichen (siehe auch Kapitel 3.2.2). Die für die transiente Steuerung wichtige Onset/Offset-Klassifizierung wird zuletzt noch um die Information, ob Vorverdeckung vorliegt, ergänzt.

3.2.1 Warped FFT, Glättung

Die Warped FFT weist statt der gleichmäßigen Frequenzbreite der Bänder eine unregelmäßige Bandbreite auf. Hier nimmt sie entsprechend dem menschlichen Gehör von hohen zu tiefen Frequenzen hin ab. Erreicht wird dies durch das Ersetzen der normalen Delays in der Speicherkette vor der Fensterung („Hann“) und der FFT durch Allpaßglieder. Diese haben zwar einen linearen Amplitudengang, aber dafür eine nichtlineare Phase und Gruppenlaufzeit - tiefe Frequenzen werden mehr verzögert als hohe. Da dieser Effekt zwischen jeder Speicherstelle auftritt, werden die Signalkomponenten mit niedriger Frequenz in größeren Zeitabständen als hohe abgetastet und erscheinen dadurch im anschließenden FFT-Spektrum bei höheren Bins („frequenzabhängiges Resampling“ siehe auch [HKS+00]).
Wie aus Abbildung 3.4 erkennbar, lässt sich die Frequenzaufteilung mit Allpässen erster

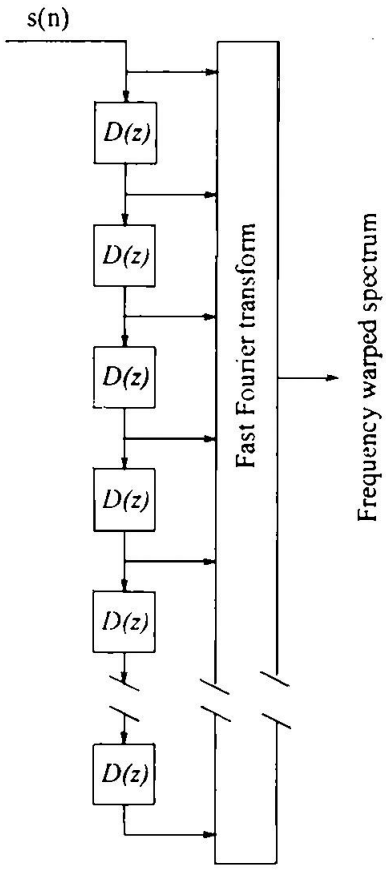

Abbildung 3.3: Eine Warped Delay Line vor einer FFT [HKS+00]

Ordnung sehr nah an die von Zwicker empfohlenen Barkbänder annähern. Die etwas höhere Auflösung von 31 (entspricht einer 64-Punkte FFT) statt 25 Bändern sorgt entsprechend den Erkenntnissen von Moore (ERB-Bänder) für eine etwas höhere Auflösung im Bassbereich. Die nicht mehr lineare Gruppenlaufzeit bei der Analyse entspricht ebenfalls relativ gut den Gruppenlaufzeiten in der Cochlea. Diese bandabhängige Verzögerung wird gleich im Anschluss an die FFT mit einer jeweils bandspezifischen Verzögerung kompensiert. Weitere Details finden sich in der Literatur([BCA+01], [HKS+00], [KA05], [MM01], [Mak03], [Mak06], [WSKH05], [ZAA+02]).

Im Anschluss an die FFT oder Filterung mit Gleichrichtung erfolgt ähnlich wie bei dem

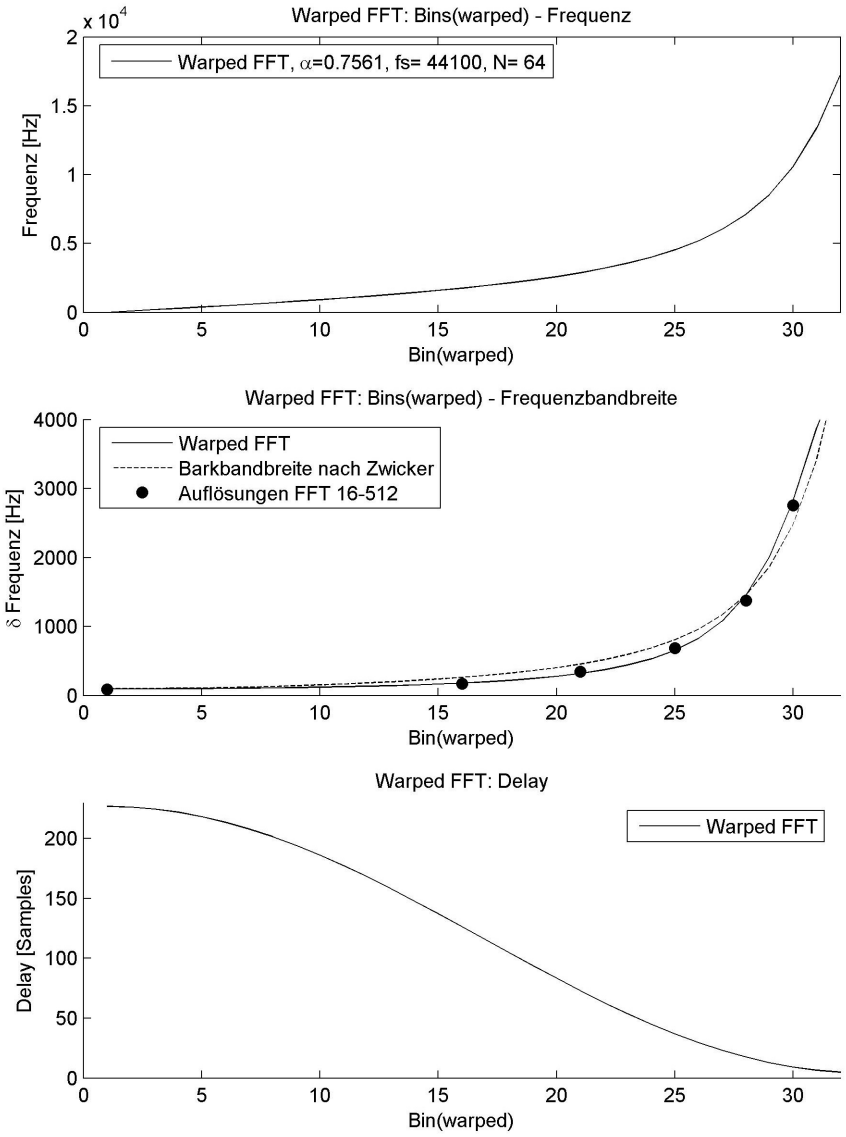

Abbildung 3.4: Frequenz- und Zeiteigenschaften der Warped FFT

anerkannten PEAQ-Modell ([TTB+00], [Opt01], [CSTT99]) eine Glättung mit einem von-Hann-Fenster von etwa 10 Millisekunden Länge. Dieses approximiert die Verschmierung bei der Umsetzung der Wanderwelle in die Quantität der Nervenimpulse durch über tausend asynchrone, unterschiedlich häufig feuernde Neuronen pro Barkband.

3.2.2 Neuronales Rückkopplungsmodell

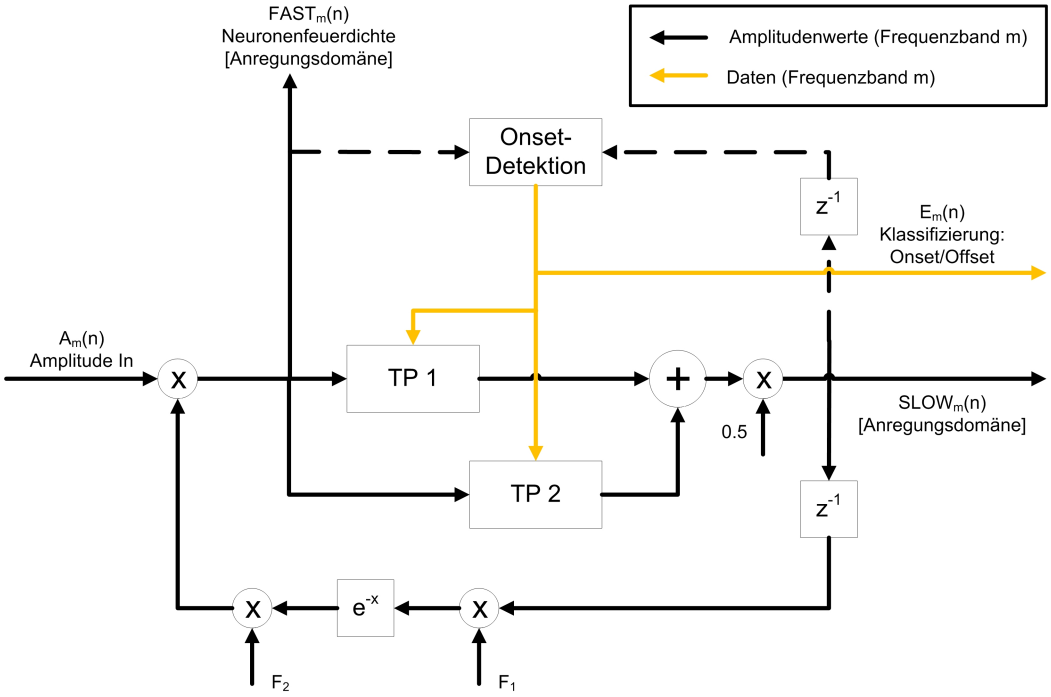

Abbildung 3.5: Das neuronale Rückkopplungsmodell, Schema eines Bandes

Das „neuronale Rückkopplungsmodell“ von Karjalainen modelliert wie bereits erwähnt, die aktiven Prozesse in der Cochlea. Es handelt sich im Wesentlichen um eine Automatische Verstärkungsregelung („Automatic Gain Control“, „AGC“), die das sehr dynamische Eingangssignal in dem deutlich kleineren Dynamikbereich der Neuronen festhält. Die Ausgabegröße $FAST$ entspricht dabei der Überlagerung aller Neuronenimpulse in einem kritischen Filterband und wird deshalb Neuronenfeuerdichte genannt. Das Synchronisieren der Neuronenfeuerrate auf die Signalfrequenz wird nicht berücksichtigt, da der Akzent der Umgebung auf transienten Vorgängen liegt. Die Größe $SLOW$ wird durch die beiden Tiefpaßfilter aus $FAST$ integriert. Sie ist eine Zwischengröße auf dem Weg zur empfundenen spezifischen Lautheit. Beide Größen werden zusammen zur „Onset/Offset“-Detektion (Einsatz/Ausklingen) verwendet. Ist $FAST(n)$ größer als $SLOW(n-1)$, so steigt $SLOW(n)$,

ein Onset liegt an. Im umgekehrten Fall wird der Zustand als Offset klassifiziert. Um die Detektion robuster gegen kleine Schwankungen zu machen, ist beim Umschalten eine Hysterese eingebaut. Für einen Wechsel der Zustände muss $FAST(n)$ das Integral $SLOW(n-1)$ deutlicher über- oder unterschreiten. Die Information wird auch benutzt, um in den beiden Tiefpässen zwischen den zwei verschiedenen Koeffizientensätzen für Onset oder Offset umzuschalten. Damit kann das Aufschwingen und Abschwingen durch die unterschiedlichen Zeitkonstantenpaare beschleunigt (Onset) oder verlangsamt (Offset) werden.
Der zeitliche Verlauf ähnelt stark den Nachverdeckungskurven: zuerst ein deutlicher exponentieller Abfall zum Nullpunkt, anschließend ein bemerkbares Abflauen des Gefälles zur statischen Ziellautstärke.

Optimierung des Modells

Sehr vorteilhaft ist, dass sich das System statisch unabhängig von den dynamischen Eigenschaften optimieren lässt. Das Ziel ist es, die beiden Verstärkungskoeffizienten F_1 und F_2 so zu bestimmen, dass das Verhältnis von Eingang $A_m(n)$ zur ausgegebenen $SLOW$-Größe dem Ergebnis von Zwickers Formel für die Spezifische Lautheit entspricht. Hierfür muss zuerst eine Beziehung zwischen der konstanten linearen Eingangsamplitude $A_m(n)$ und dem Ausgang $SLOW(n)$ berechnet werden. Diese Möglichkeit ist in der ursprünglichen Formulierung in [Kar96] und der daran anknüpfenden Implementierung von Härmä in der HUT-Ear-Toolbox ([Här99b] [Här99a]; [Här99c] [HLK99]) nicht ausgeführt. Dies wird hier nachgeholt:

$$FAST(n) = F_2 \cdot A(n) \cdot e^{-F_1 SLOW(n-1)} \tag{3.1}$$

$$SLOW(n) = 0.5[SLOW_1(n) + SLOW_2(n)] \tag{3.2}$$

$$SLOW_m(n) = \alpha_{m,up/down} SLOW_m(n-1) + (1 - \alpha_{m,up/down}) FAST(n) \tag{3.3}$$

$$m = 1, 2; \quad \text{falls Onset, verwenden von } \alpha_{m,up}, \text{ sonst } \alpha_{m,down}$$

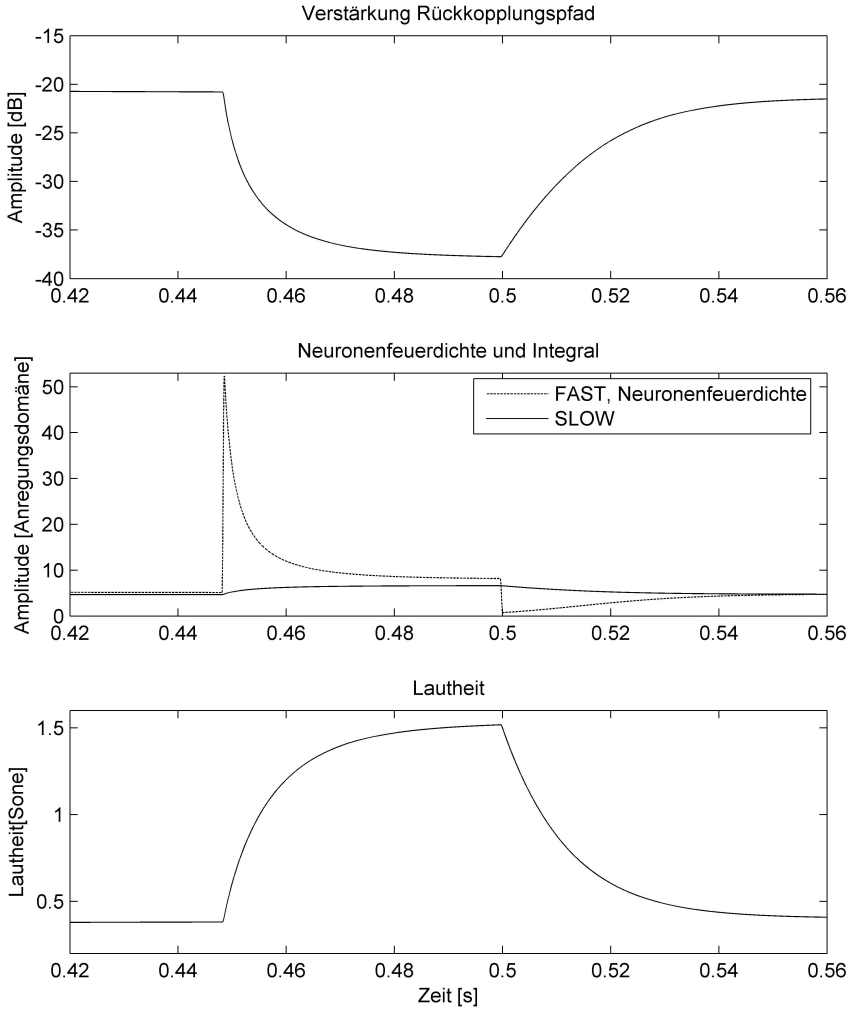

Abbildung 3.6: Zeitverhalten des neuronalen Rückkopplungsmodells, Sinus 4kHz, Reaktion auf einen Sprung von $35dB$ auf $56dB$ und zurück

Falls $A(n) = A = const \Rightarrow FAST(n) = SLOW(n) = const = SLOW$

$$\mathbf{A} = \mathbf{F_2} \cdot \mathbf{SLOW} \cdot \mathbf{e^{F_1 SLOW}} \tag{3.4}$$

$$\mathbf{A[dB]} = \mathbf{20} \cdot (\log_{\mathbf{10}} \frac{\mathbf{1}}{\mathbf{F_2}} + \log_{\mathbf{10}} \mathbf{SLOW} + \mathbf{F_1} \log_{\mathbf{10}}(\mathbf{e}) \mathbf{SLOW}) \tag{3.5}$$

Wie aus Formel 3.5 ersichtlich, tauchen die beiden Verstärkungsfaktoren nun getrennt voneinander in einzelnen Summanden auf. F_1 und F_2 lassen sich nun mit der Methode der

kleinsten Fehlerquadrate so bestimmen, dass der Verlauf von A zu $SLOW$ (mit einer kleinen Nachbearbeitung) dem Verlauf von A zur spezifischen Lautheit nach Formel 2.1 gleicht. Eine numerisch sehr stabile Umrechnung, bei dem sich entsprechend Karjalainens Entwurf die Rückkopplungsverstärkung F_2 in etwa im Bereich von 1 befindet, ist folgende:

$$N' = 0.08(\frac{E_{TQ}}{E_0})^{0.23}(\frac{SLOW}{3.3})^4\frac{sone_G}{Bark} \tag{3.6}$$

E_{TQ} : Wahrnehmungsschwelle
E_0 : Referenzintensität 0dB@1kHz

Das dynamische Verhalten oder auch die Trägheit des Systems wird von den beiden inte-

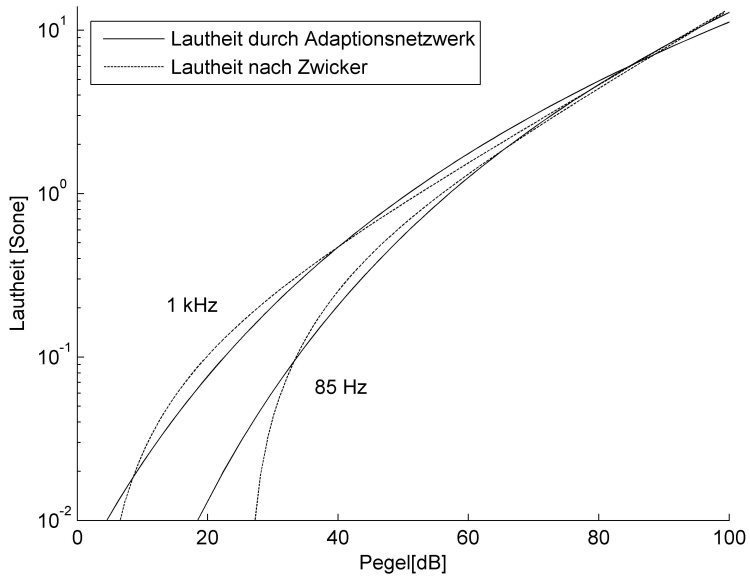

Abbildung 3.7: Vergleich der Lautheit nach dem neuronalen Rückkopplungsmodell und gemäß Zwickers Formel

grierenden Tiefpassfiltern (1. Ordnung) bestimmt. Für das Aus- und Einschwingen stehen jeweils 2 eigene Zeitkonstanten zur Verfügung, die benutzt werden, um das Ein- und Ausschwingverhalten an die in [Fas77a] ermittelten Nachverdeckungskurven anzugleichen.

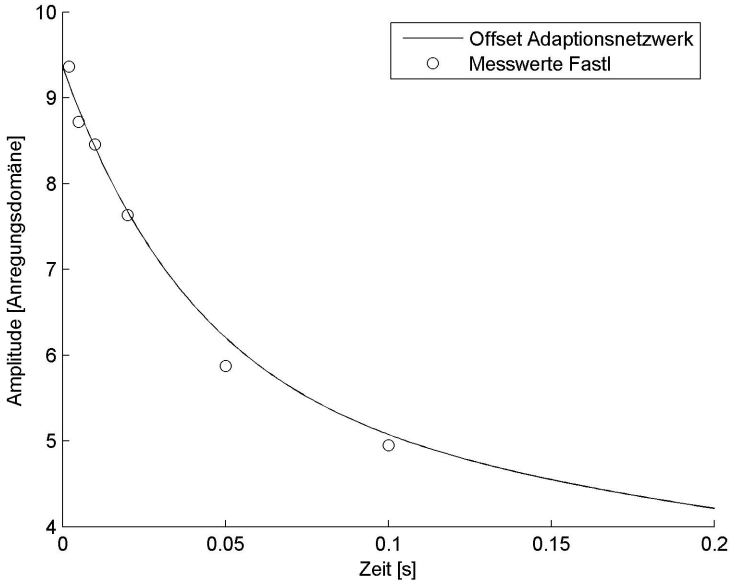

Abbildung 3.8: Offset/Ausschwingen des neuronalen Rückkopplungsmodells im Vergleich zu den psychoakustischen Messdaten aus [Fas77a]

3.2.3 Detektion der Vorverdeckung

Um die exakte Erklärung der Vorverdeckung werden durchaus noch einige Kontroversen geführt. In der hier benutzten konservativen Abschätzung berechnet sie sich aus der Differenz des aktuellen Zustandes der Neuronenfeuerdichte ($FAST(n)$)und eines gewichteten Mittels über die kommende Neuronenaktivität ($v(n)$). Gewissermaßen also eine Beurteilung der Bedeutung eines einzelnen Samples im Vergleich zu den zukünftigen. Überschreitet das Ergebnis nun das Integral $SLOW$, so ist dieser Bereich vorverdeckt. Zur Feineinstellung ist noch ein Gewichtungsfaktor α beigefügt, um die Detektionshäufigkeit zu erhöhen oder abzusenken. Standard ist $\alpha = 1$.

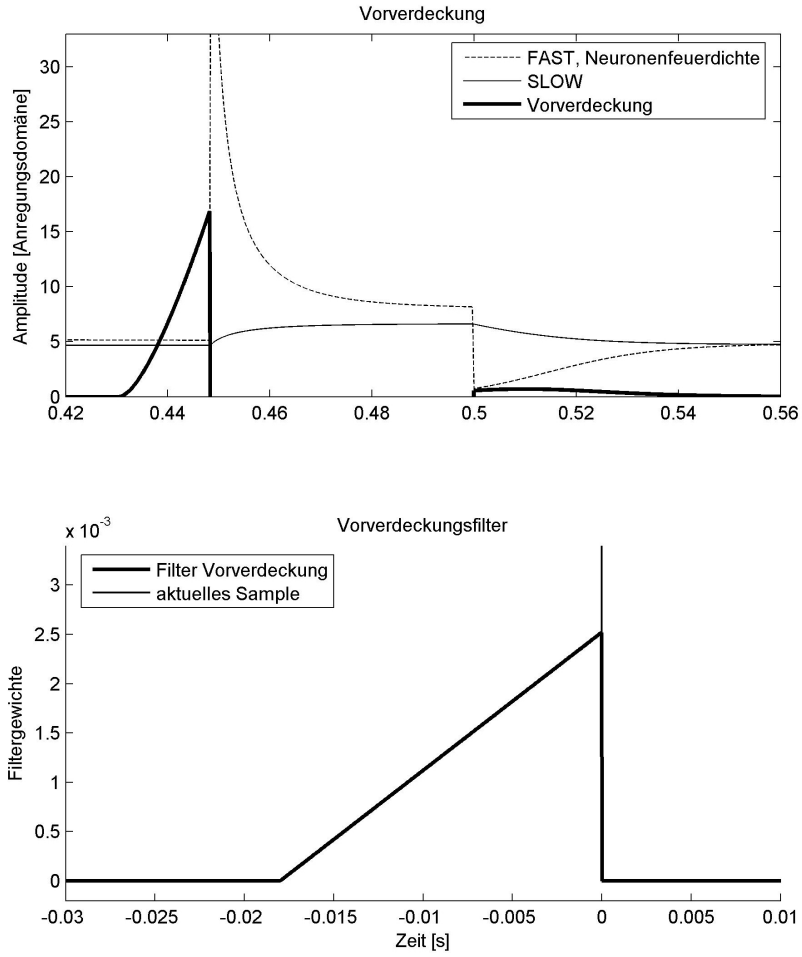

Abbildung 3.9: Funktionsweise Vorverdeckungsfilter

$$V_{fut}(n) = \frac{1}{1+2+\ldots+N_V} \cdot [FAST(n+N_v+1) + \\ +2 \cdot FAST(n+N_v) + \ldots + N_V \cdot FAST(n+1)] \tag{3.7}$$

$$V(n) = V_{fut}(n) - FAST(n) \tag{3.8}$$

$$V(n) > \alpha \cdot SLOW(n) \quad \Rightarrow \text{Sample verdeckt} \tag{3.9}$$

$V(n)$: Vorverdeckung [Anregungsdomäne]
$V_{fut}(n)$: Neuronenfeuerdichte Zukunft [Anregungsdomäne]
N_V : Länge Vorverdeckung in Samples
α : Gewichtungsfaktor für Verdeckungshäufigkeit

3.3 Statische Kompressorkennlinie, Zielvorgabe

3.3.1 Bestimmung des globalen Regelziels

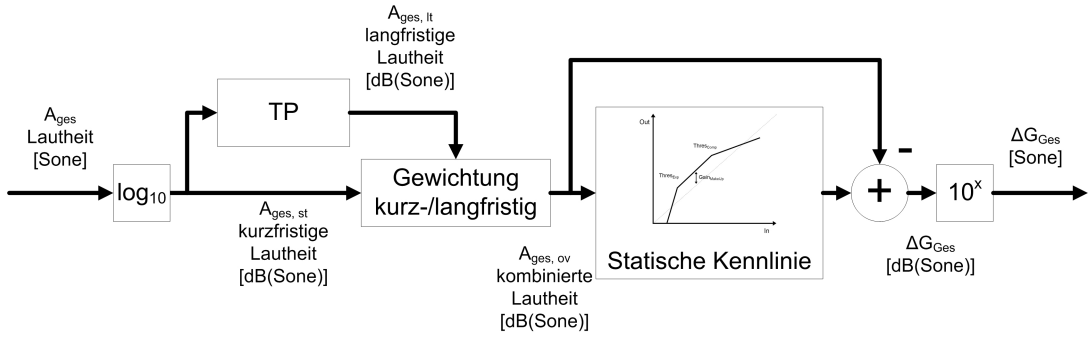

Abbildung 3.10: Gesamtsystem der Statischen Kennlinie, Zielvorgabe

Das die statische Kompressorkennlinie enthaltene System ist im Vergleich zu den State-of-the-Art Geräten noch etwas erweitert worden.

Wie üblich werden hier die internen Berechnungen in der logarithmischen Domäne ausgeführt. Da die Ausgangsgröße das zur Lautheitsempfindung lineare *Sone*-Maß ist, wird die Einheit fortan also $dB(Sone)$ bezeichnet. 6 dB Erhöhung bedeuten nun eine Verdoppelung der Lautheit und nicht des Schalldrucks des physischen Signals. Da im Großteil des dynamischen Bereichs des Ohres bei einer Schalldrucksteigerung von 10 dB eine Lautheitsverdoppelung erfolgt, ist das Verhältnis der Ratio r des neuen Kompressors zu der Ratio klassischer Prozessoren in etwa $\frac{10}{6}$. Die Steigung ist hier also geringer. Die statische Kennlinie selbst ist prinzipiell in der vorliegenden Implementation durchaus konventionell. Sie kalkuliert aus den tatsächlichen Lautheitswerten („In“) die gewünschten Sollwerte („Out“) in dB. Es existieren zwei Thresholds ($Thres_{Exp}$, $Thres_{Comp}$) mit jeweils einer zugehörigen Ratio (r_{Exp}, r_{Comp}). Üblicherweise sollen die Ausgangswerte unterhalb des $Thres_{Exp}$ im Vergleich zu den Eingangswerten kleiner sein, da es sich in diesem Bereich größtenteils um statisches Hintergrundrauschen handelt. In der statischen Kompressorkurve äussert sich dies mit einer mit der Steigung r_{Exp} stärker abfallenden Kurve. Je leiser ein Eingangswert, desto weiter ent-

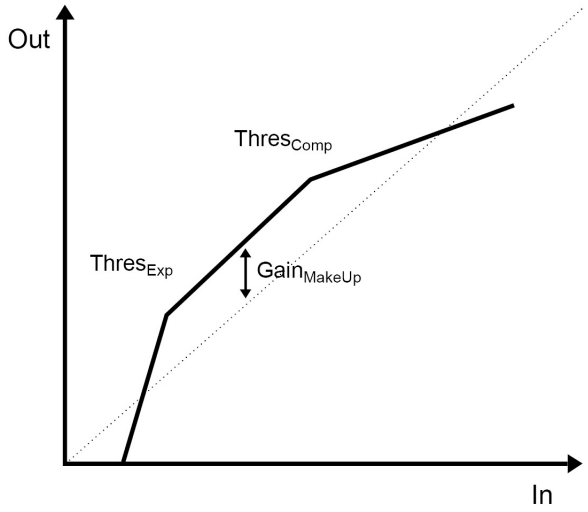

Abbildung 3.11: Statische Kompressorkennlinie

fernt ist er von $Thres_{Exp}$, desto kleiner wird auch sein Ausgabewert. Bei Eingangswerten, die größer sind als der obere Threshold $Thres_{Comp}$ verhält es sich ähnlich. Auch hier sollen die Ausgabewerte kleiner werden als die Eingabewerte, der Anstieg der Lautheit in der Ausgabe kleiner sein als in der Eingabe. Der Abstand zwischen beiden Werten wird umso größer, je weiter der Eingabewert von $Thres_{Comp}$ entfernt ist, desto größer er also im Vergleich zu diesem ist. Durch das Subtrahieren der ursprünglichen Lautheitswerte von den Sollwerten ergibt sich der zur Korrektur des Audiosignals verwendete Verstärkungsfaktor.

falls $A_{Ges,ov} > Thres_{Comp}$

$$A_{Ges,ref} = Thres_{Comp}(1 - r_{Comp}) + r_{Comp}A_{Ges,ov} + Gain_{MakeUp} \tag{3.10}$$

falls $Thres_{Comp} > A_{Ges,ov} > Thres_{Exp}$

$$A_{Ges,ref} = A_{Ges,ov} + Gain_{MakeUp} \tag{3.11}$$

falls $A_{Ges,st} < S_{in}$

$$A_{Ges,ref} = Thres_{Exp}(1 - r_{Exp}) + r_{Exp}A_{Ges,ov} + Gain_{MakeUp} \tag{3.12}$$

$$G_{Ges} = A_{Ges,ref} - A_{Ges,ov} \tag{3.13}$$

Als Erweiterung ist vor der eigentlichen Zielfindung in der Kennlinie noch eine Ermittlung der langfristigen Lautheit aus der kurzfristigen sowie eine Gewichtung der beiden eingefügt. Damit ist es möglich, den Akzent von einer Regelung (und damit auch Lautheitsangleichung) einzelner Ereignisse - etwa isolierter Buchstaben - auf größere Gruppen wie etwa Wörter oder Sätze zu lenken. Die langfristige Lautheit wird durch den bereits bekannten Tiefpaßintegrator mit umschaltbarer Zeitkonstante aus der kurzfristigen Lautheit berechnet. Die Zeitkonstante im Falle einer Steigung (Onset) beträgt hierbei jedoch 0, die langfristige Lautheit entspricht also in diesem Fall exakt der eingehenden kurzfristigen Lautheit. Im Falle eines Abfalls (Offset) beträgt die Zeitkonstante sinnvollerweise 0.4 bis 5 Sekunden, die Lautheit bleibt länger in Erinnerung und klingt deutlich langsamer ab.

Die oft gefundene Lösung, aus diesen beiden Größen wieder eine Referenz zu bilden, wäre ein

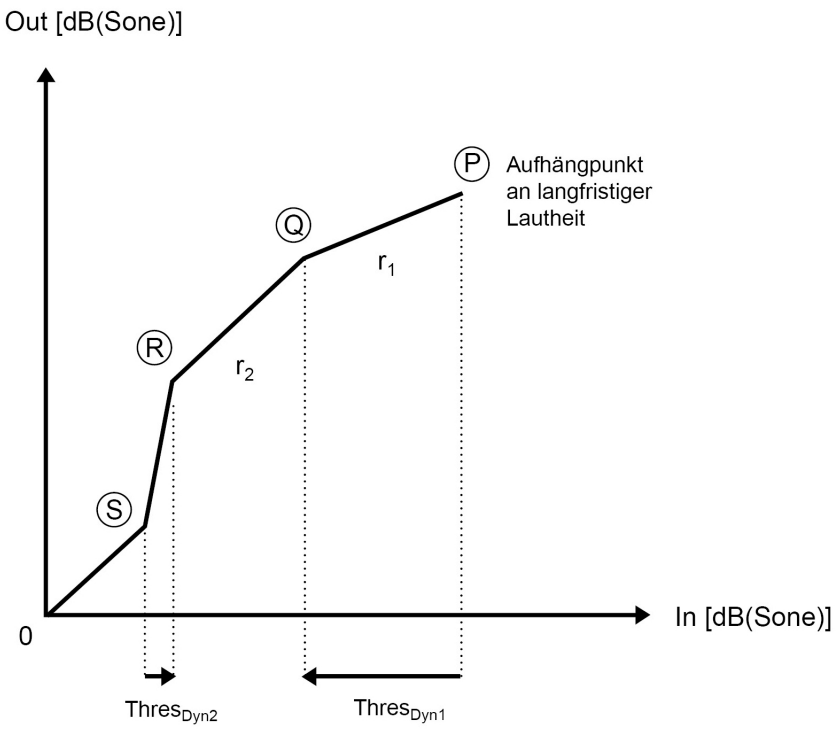

Abbildung 3.12: Gewichtung kurz-/langfristige Lautheit: Dynamische Kompression der kurzfristigen Lautheit mit der langfristigen Lautheit als Referenz

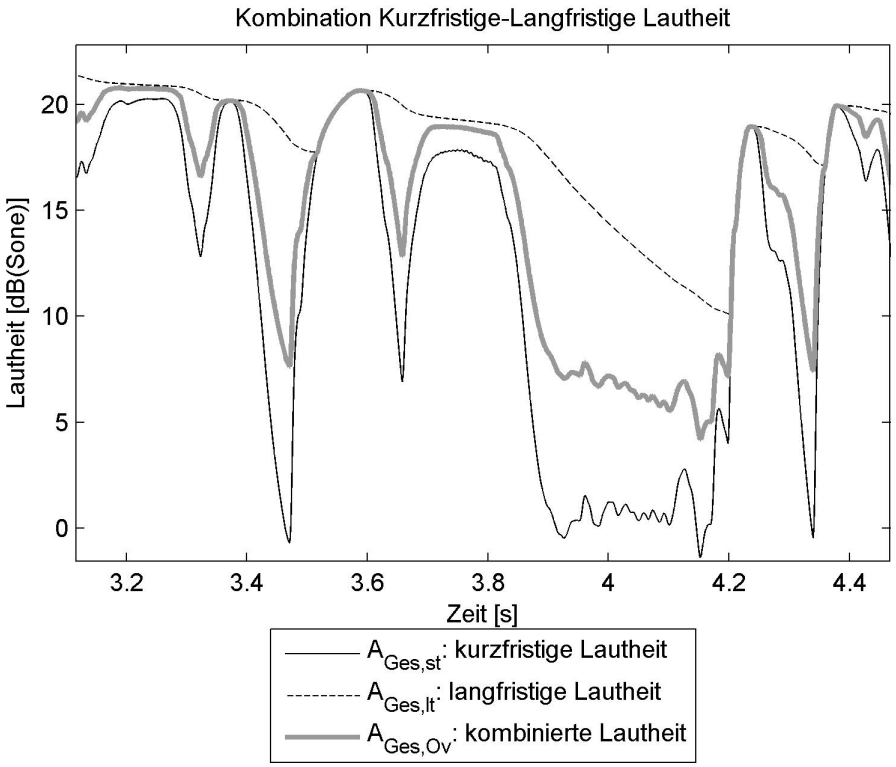

Abbildung 3.13: Gewichtung kurz-/langfristige Lautheit: Verlauf aller drei Größen bei einem kurzen Ausschnitt männlicher Sprache

Min-Max-Vergleich oder eine RMS-Addition mit einem einstellbaren Bias auf einer Größe, um die Betonungen zu ändern.

Die hier verwendete Lösung funktioniert etwas anders: Die langsam abklingende Erinnerung rückt um so mehr in den Vordergrund, je weniger prominent, also leiser alle nachfolgenden Ereignisse im Verhältnis zu diesem sind. Für die Implementation bedeutet dies: Die langfristige Lautstärke ist der relative Nullpunkt, die Bewegungen der kurzfristigen Lautstärke werden auf einer Geraden mit einer Steigung< 1 von dieser aus komprimiert. Je weiter die kurzfristige Lautstärke von der langfristigen entfernt ist, desto weniger fällt deren Veränderung und Verlauf ins Gewicht (Abbildung 3.12). Abbildung 3.13 zeigt als Beispiel die Lautheitsverläufe eines männlichen Sprechers. Die Aufteilung in zwei verschiedene Steigungen r_1 und

r_2 am $Thres_{Dyn1}$ dient zur Feinjustierung. Der $Thres_{Dyn2}$ wird auf den unteren Threshold $Thres_{Exp}$ der statischen Kurve aufgeschlagen. Ab hier beginnt die Rückkehr zur kurzfristigen Lautheit um weiterhin eine effiziente und schnelle Unterdrückung des Rauschens ab dem $Thres_{Exp}$ zu gewährleisten.

Bei jedem eingehende Sample sind folgende Arbeitsschritte zu berechnen:

1. Überprüfung

$$\begin{aligned} &\text{falls} \quad A_{Ges,lt} - (Thres_{Exp} - ThresDyn1) < \quad 2 \cdot Thres_{Dyn2} : \\ &\qquad A_{Ges,ref} = \quad AGes, st \\ &\text{sonst:} \\ &\qquad \textit{Berechnung von 2.+3.} \end{aligned} \tag{3.14}$$

2. Berechnung der Kurveneckpunkte

$$P_{in} = A_{Ges,lt} \tag{3.15}$$
$$P_{out} = A_{Ges,lt} \tag{3.16}$$
$$S_{in} = Thres_{Exp} \tag{3.17}$$
$$S_{out} = Thres_{Exp} \tag{3.18}$$
$$Q_{in} = A_{Ges,lt} + ThresDyn1 \tag{3.19}$$
$$Q_{out} = A_{Ges,lt} + r_1 \cdot ThresDyn1 \tag{3.20}$$
$$R_{in} = Q_{in} + ThresDyn2 \tag{3.21}$$
$$R_{out} = (R_{in} - Q_{in})r_2 + Q_{out} \tag{3.22}$$

3. Berechnung der Ausgabewerte

$$\text{falls} \quad A_{Ges,st} > Q_{in}$$

$$A_{Ges,ov} = \quad Q_{in}(1 - r_1) + r_1 A_{Ges,st} \tag{3.23}$$

$$\text{falls} \quad Q_{in} > A_{Ges,st} > R_{in}$$

$$A_{Ges,ov} = \quad (A_{Ges,st} - Q_{in})r_2 + Q_{out} \tag{3.24}$$

$$\text{falls} \quad R_{in} > A_{Ges,st} > S_{in}$$

$$A_{Ges,ov} = \quad \frac{S_{out} - R_{out}}{S_{in} - R_{in}} A_{Ges,st} + \frac{S_{out} - R_{out}}{S_{in} - R_{in}} R_{out} - R_{in} \tag{3.25}$$

$$\text{falls} A_{Ges,st} < S_{in}$$

$$A_{Ges,ov} = A_{Ges,st} \tag{3.26}$$

3.3.2 Bestimmung der lokalen Regelziele

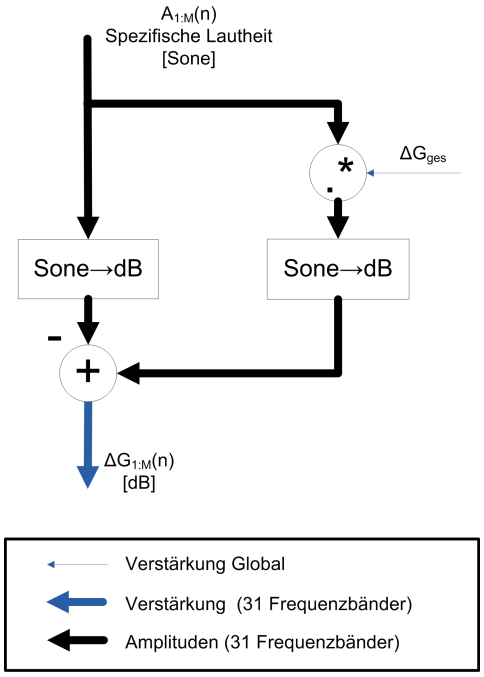

Abbildung 3.14: Berechnen der lokalen Regelziele pro Band $[dB]$ aus dem globalen Regelziel

Nachdem das globale Regelziel bestimmt worden ist, fehlt noch die Übersetzung dieses Ziels auf die einzelnen lokalen Bänder. Hier entsteht das gewünschte frequenzabhängige Regelverhalten. Jedes Band besitzt wegen der Kurven gleicher Lautheit eine eigene Funktion zur Rücktransformation von Lautheit [*Sone*] in Schalldruck [*dB*] (vgl. Abbildung 3.7). Da diese Funktion auch in sich deutlich nichtlinear ist, ist das *dB*-Äquivalent einer Pegeländerung in *Sone* abhängig vom Ausgangspegel. Je lauter ein Signal, desto weniger *dB* entsprechen einer Lautheitsänderung um einen konstanten Faktor.
Auch hier wird ein Soll ermittelt, in diesem Fall durch die Multiplikation des globalen Regelzieles mit der Spezifischen Lautheit des Bandes. Sowohl der Ist-Zustand wie auch der eben kalkulierte Soll-Zustand werden in ihre entsprechenden Schalldruckpegel in *dB* transformiert. Durch die Subtraktion des Ist-Pegels vom Soll wird die nötige Korrektur $\Delta G_m(n)$ ermittelt. Die Formel für die Umrechnung von *Sone* in *dB* für das jeweilige Band m ergibt sich aus der Optimierung des Hörmodells:

$$\begin{aligned} A_m(n)[dB] &= -20\log_{10} F_{2,m} + 20\log_{10} SLOW_m(n)[ExDom] - \\ &\quad -20\log_{10}(e)F_{1,m} \cdot SLOW_m(n)[ExDom] \qquad (3.27) \\ SLOW_m(n)[ExDom] &= \left(\frac{A_m(n)[Sone]}{Sf_m}\right)^{0.25} \cdot 3.3 \qquad (3.28) \\ Sf_m &= 0.08\left(\frac{E_{TQ,m}}{E_{0,m}}\right)^{0.23} \qquad (3.29) \end{aligned}$$

E_{TQ} : Wahrnehmungsschwelle
E_0 : Referenzintensität 0dB@1kHz
Sf_m : Skalierungsfakor

3.4 Die Attack-/Releasesteuerung

Ein Hauptanliegen dieser Arbeit ist es, den zeitlichen Verlauf der Regelung zu optimieren. In Kapitel 2.2.4 ist als Hauptforderung ein exakter Beginn und vorausschauendes Regeln

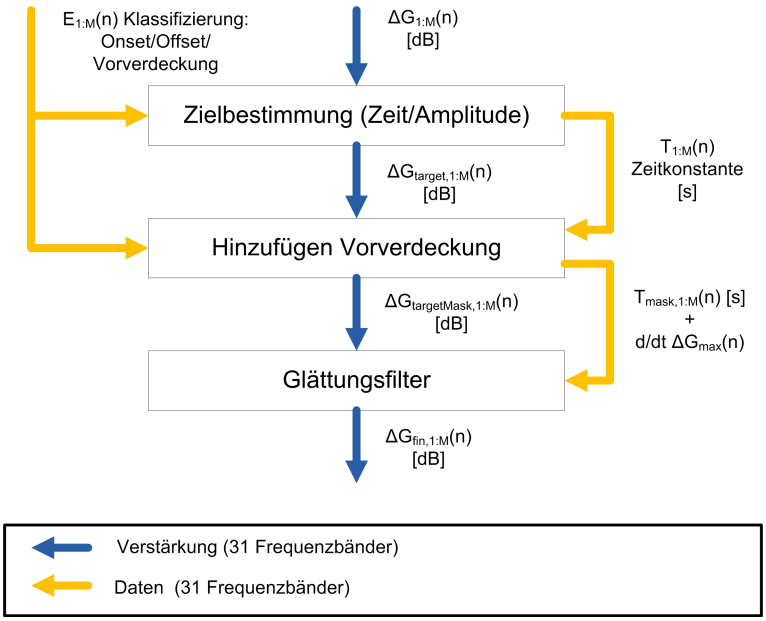

Abbildung 3.15: Aufbau der Attack/Release-Regelung

vermerkt. Zusätzlich soll die Vorverdeckung, falls sie auftritt, zur Verstärkungsreduktion benutzt werden. Das konkrete Regelverhalten des Systems lässt sich anhand von Abbildung 3.16 gut nachvollziehen. Findet ein Wechsel von Offset zu Onset oder umgekehrt statt, so wird innerhalb der vorgegebenen Regelzeiten vorausschauend der nächste Wechsel gesucht und der Wert zu diesem Zeitpunkt als Ziel $\Delta G_{m,target}$ gesetzt. Die vorgegebenen Regelzeiten nennen sich, an die üblichen Kompressoren angelehnt, ebenfalls Attack- und Releasezeit und sind auch in deren Größenordnung (etwa $20ms$ für Attack und $130ms$ für Release angesiedelt). Findet sich innerhalb der Attack-/oder Releasezeit kein Wechsel, so wird der Wert des sich um diese Zeitkonstante in der Zukunft befindlichen Samples als Ziel genommen. Auf diese Weise tastet sich die Vorschau sampleweise nach vorne und gibt Werte in der Zukunft als Ziel an. Trifft sie auf einen Statusumschwung, so bleibt die Zielsetzung $\Delta G_{m,target}$ auf dem Wert zu diesem Zeitpunkt stehen.

Parallel dazu wird die Zeit zwischen dem Umschwenken festgehalten. Sie wird später als Zeitkonstante $T_m(n)$ für den nachgeschalteten Glättungsfilter verwendet. Aus $T_m(n)$ wird

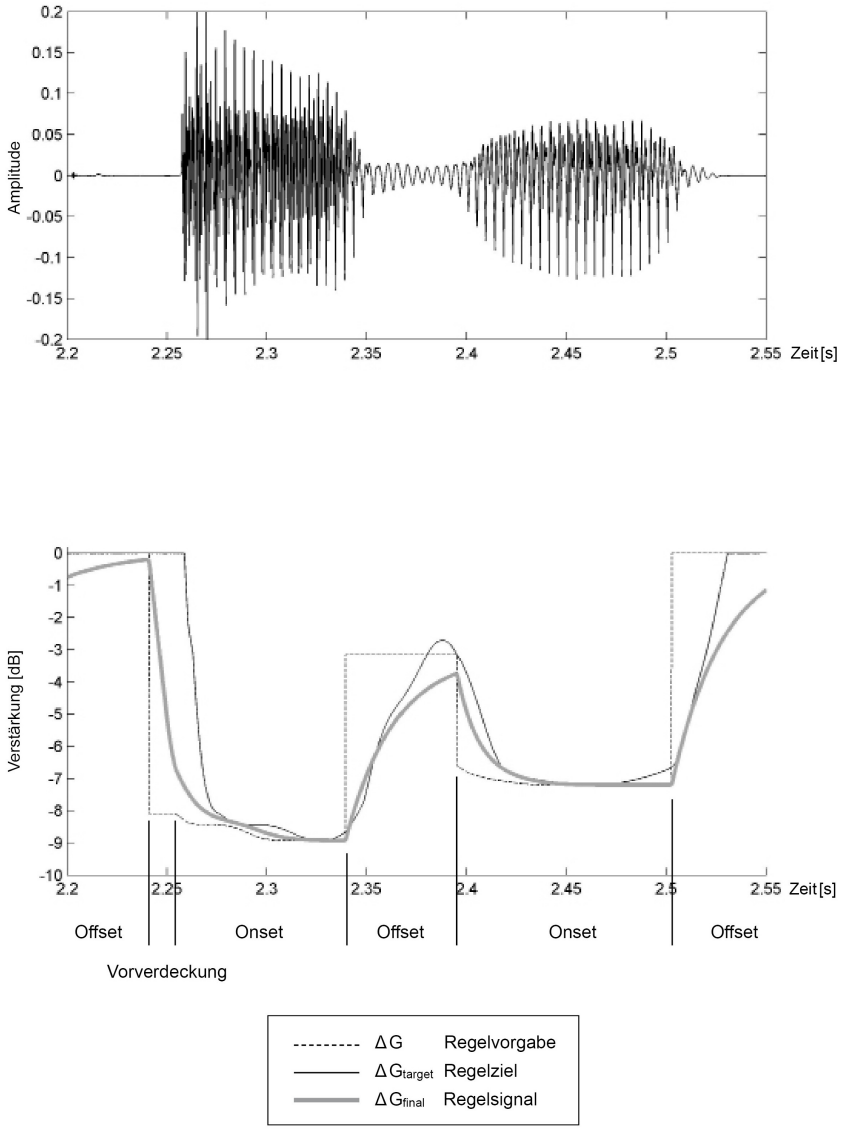

Abbildung 3.16: Beispiel (Sprache) des vorausschauenden adaptiven Attack-Release Mechanismusses, Signal + Steuersignalwerte eines Single-Band-Kompressors

der Filterkoeffizient α des Glättungsfilters derart berechnet, dass bei einem Abschwingen der Filter nach $T_m(n)$ Sekunden seinen Zielwert am Eingang zu 90% erreicht hat. Der vorgegebene Zielwert hier ist $\Delta G_{m,target}$, der so geglättete Ausgang $\Delta G_{fin,m}(n)$. Falls kein neues Ereignis gefunden wurde, gleicht die Zeitkonstante der Vorschau (Attack-/Releasezeit). Das

ist gewünscht, in einem solchen Fall ohne neuen Onset/Offset ist das Signal statisch und der Filter unterdrückt störende Modulationen.
Zu erwähnen wäre noch die leichte Verspätung der Onsets und Offsets (zum Beispiel bei 2.4 Sekunden). Diese stammt von der in der Detektion (Kapitel 3.2.2) inkludierten Hysterese. Diese Verzögerung ist aber unabhängig von den statischen Vorgaben, vergleichsweise gering und gleicht den Verschmierungseffekt des symmetrischen von-Hann-Fensters vor dem neuronalen Rückkopplungsmodell aus.
Bei dem deutlichen Einsatz bei 2.25 Sekunden (Abbildung 3.16) lässt sich die Ausnutzung der Vorverdeckung begutachten. Statt wie übliche Kompressoren den sehr großen Regelweg von $8dB$ hörbar nach dem Einsatz zurückzulegen („wegdrücken“), wird die Lautstärkenkorrektur hier vor den Einsatz gelegt. Der Weg selbst wird auf einer Geraden zurückgelegt, statt wie bei der sonstigen Regelung auf einer e-Funktion. Grund ist, dass bei letzterer der größte Teil des Regelweges in den ersten Momenten zurückgelegt wird, die Vorverdeckung dagegen anfangs flach verläuft und zum Onset hin immer steiler wird. Diese lineare Kurvenform wird erreicht, indem im Glättungsfilter zusätzlich eine Beschränkung der Steigung von einem Abtastzeitpunkt zum nächsten eingebaut wird. Wird die Steigung auf einen konstanten Maximalwert beschränkt und die Zeitkonstante kurz gewählt, so verläuft die Kurve linear. Da der Start- und Endzeitpunkt der Vorverdeckung ebenso bekannt ist wie der zurückzulegende Weg, lässt sich diese Steigung leicht bestimmen:

$$|d/dt\Delta G_{max}| = \left|\frac{\Delta G_{target}(n_{end}) - \Delta G_{target}(n_{start})}{n_{end} - n_{start}}\right| \tag{3.30}$$

Der Glättungsfilter 1.Ordnung ist nach dem üblichen Schema aufgebaut, besitzt jedoch zeitvariable Koeffizienten. Abweichend von der Norm (63%) bezieht sich die Zeitkonstante auf die Zeit, nach der der Filter 90% des Endwertes erreicht.

$$\Delta G_{fin,m(n)} = \alpha_m(n)\Delta G_{fin,m(n-1)} + (1 - \alpha_m(n))\Delta G_{targetMask,m}(n) \tag{3.31}$$

$$\alpha_m(n) = 0.1^{\frac{1}{f_s T_{m(n)}}} \tag{3.32}$$

Hinsichtlich der praktischen Implementierung lässt sich sagen, dass in diesem Modul die meiste Verzögerung anfällt. Die oben beschriebene Vorschau ist natürlich nichts anderes als eine Verzögerung des gerade zu bearbeitenden Samples. Die Vorschau für die kurze Attackzeit beträgt etwa $20ms$, für die längere Releasezeit dagegen sinnvollerweise mindestens $20ms$ bis zu etwa $200ms$. Die Vorverdeckungdetektion und Regelung schlägt noch einmal mit etwa $20ms$ zusätzlich zu Buche. Die gesamte Verzögerungszeit des Systems beträgt also mindestens $40ms$ bis $220ms$ zuzüglich der durch die Implementierung bedingten Verzögerungen (Buffering).

Abschließend noch ein Beispiel zur unabhängigen zeitlichen Steuerung der einzelnen Bänder. Es handelt sich um zwei gezupfte Bassnoten mit einem anschließenden Snareschlag. Drei Bänder wurden exemplarisch herausgegriffen:

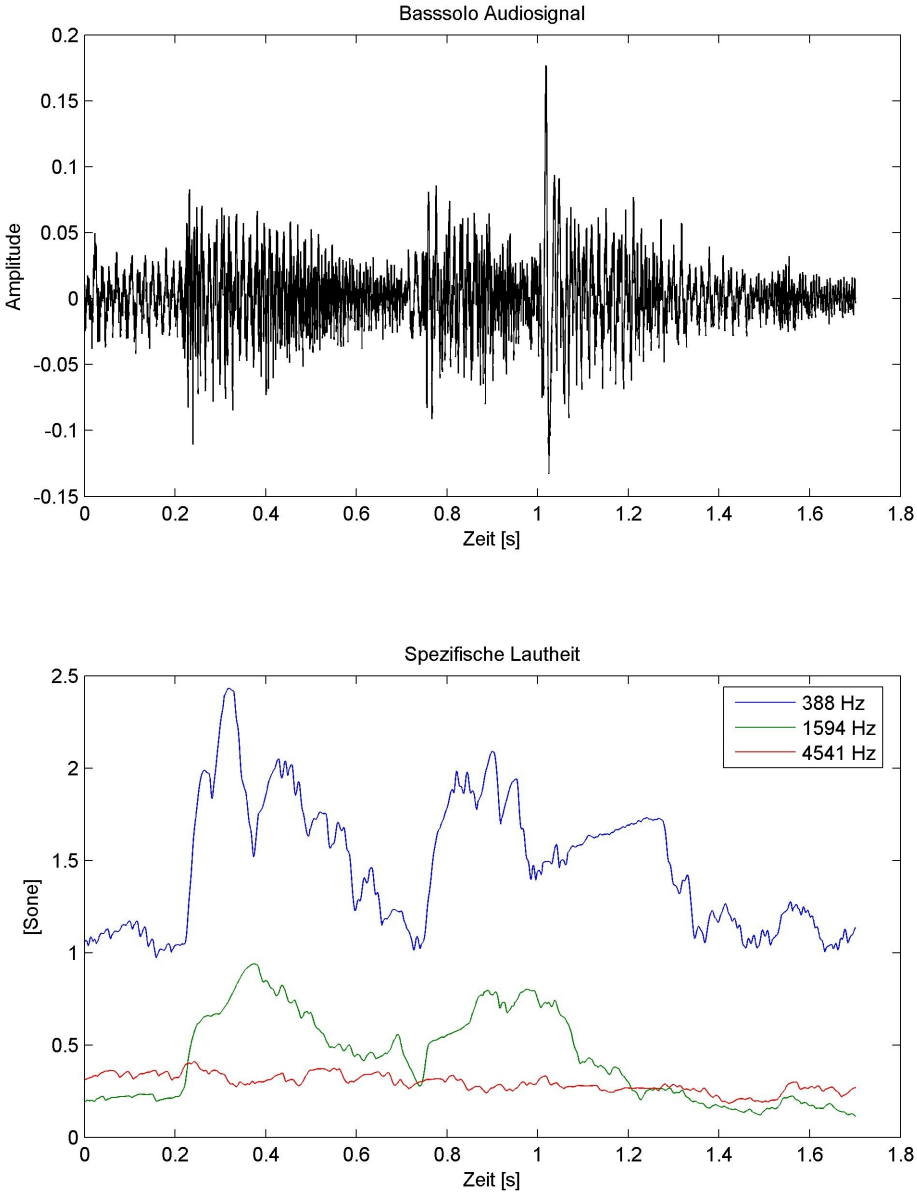

Abbildung 3.17: Übersicht über die Schritte zur Gewinnung des bandspezifischen Regelsignals: zwei Bassnoten mit darauf folgendem Snareschlag; Amplitude + spezifische Lautheit

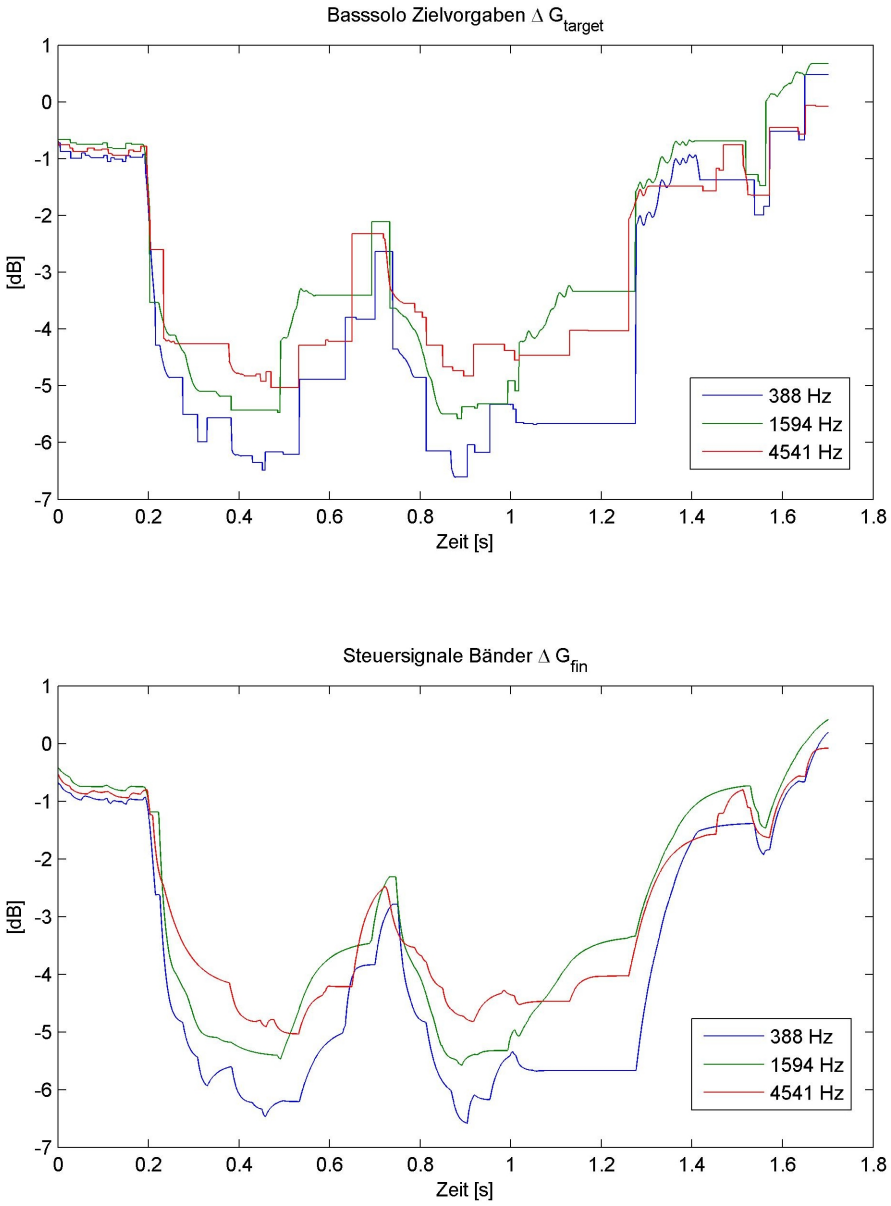

Abbildung 3.18: Übersicht über die Schritte zur Gewinnung des bandspezifischen Regelsignals: zwei Bassnoten mit darauf folgendem Snareschlag; Zielvorgaben ΔG_{target} + Steuersignale ΔG_{fin}

3.5 Filterkonstruktion und Audiobearbeitung

Die Bearbeitung des Audiosignals erfolgt nun abschließend mit einem noch zu konstruierenden FIR-Filter. Hier sind zwei Varianten möglich: eine linearphasige Variante mit 512 Koeffizienten bei $44,1kHz$ Samlefrequenz sowie einer minimalphasigen mit lediglich 64 Koeffizienten. Dieser besitzt bei tiefen Frequenzen in etwa die gleiche Auflösung wie der bedeutend längere und damit rechenaufwändigere Linear-Phase-Filter.
Der Grund, wieso hier auf die Verwendung einer Polyphasen-Filterbank verzichtet wird, sind die unbedingt zu vermeidenden Aliasing-Artefakte. In dieser Implementierung soll erst einmal eine klanglich optimale Referenz aufgestellt werden, bevor weitere Effizienzbestrebungen vorgenommen werden können. Wie in Kapitel 2.2.3 beschrieben sind bereits gängige Fullband-Kompressoren anfällig für Aliasing-Artefakte. In einer Polyphasen-Filterbank verschärft sich dies weiter: schon bei einer statischen Bearbeitung der Bandsignale zwischen Analyse und Synthese wird die Aliasing-Auslöschungsbedingung der Filterbank ungültig [Vai93]. Selbst bei einem großzügigen, nicht kritschen Downsampling der Filterbank besteht hier eine gewisse Gefahr, dass dies geschieht. Stattdessen wird der FIR-Filter am Ausgang hier noch für jedes Sample neu berechnet, man könnte diese Operation also als dynamischen Filter bezeichnen. Bei der Verwendung des Warped-Filters lässt sich ausserdem der Inhalt der Warped-Delay-Line der Analyse-FFT im Gehörmodell benutzen, es fallen also keine weiteren Rechenzyklen für das Warping an.

3.5.1 Minimalphasiger Warped-Filter

Beim Warped FIR-Filter wird der Filter zuerst als ganz normaler Minimum-Phase-FIR-Filter mit einer Länge von 64 Samples erzeugt. Als Vorlage dienen hier die Steuerfaktoren der einzelnen Barkbänder in dB. Dieser kurze Filter kann nun nicht direkt mit den verzögerten Eingangssamples gefaltet werden, da ja in diesem Fall die Frequenzen im Gegensatz zur Analyse weiterhin gleichmäßig über π verteilt und nicht gewarpt wären. Er wird stattdes-

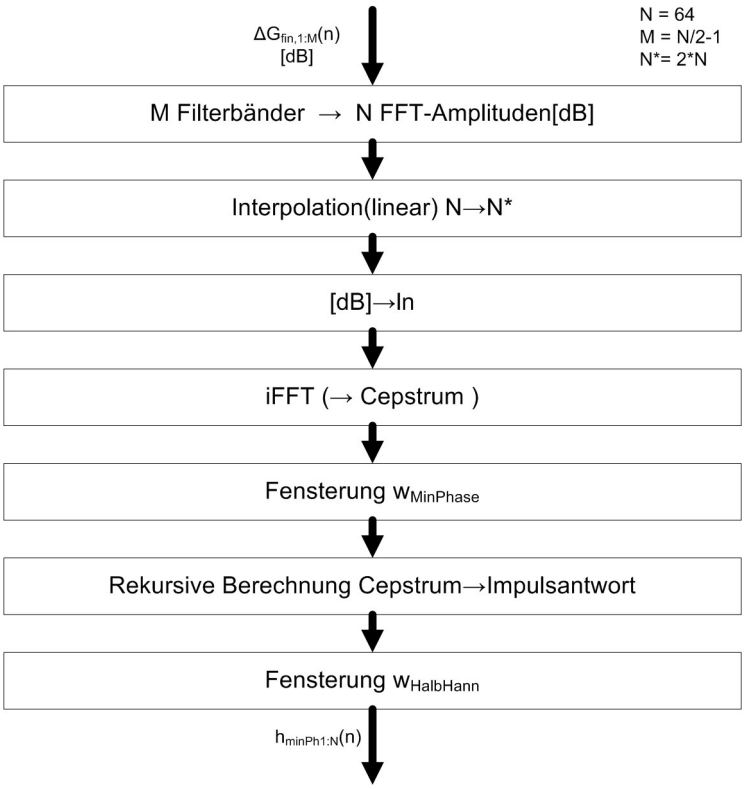

Abbildung 3.19: Konstruktion des minimalphasigen Warped-Filters aus den Steuerfaktoren der Bänder

sen mit dem ebenfalls verzögerten Inhalt der Warped-Delay-Line der FFT-Analyse gespeist. Diese wird jedes Sample rekursiv neu berechnet, es müssem also immer alle 64 Stellen des Inhalts gespeichert werden. Abgesehen von diesem Speicherbedarf benötigt diese Vorgehensweise aber deutlich weniger Rechenzyklen als ein ungewarpter Filter gleicher Auflösung im Bassbereich.

Um im Zustand eines spektral flachen Filters (alle Frequenzen sind auf $0dB$) keine Phasenverzerrungen beim Warped-FIR-Filter zu erhalten, muss der erste Koeffizient der einzige „gesetzte“ sein. Alle anderen werden von mit Allpaßfiltern verzögerten Werten gespeist, es entsteht gezwungenermaßen eine Dispersion sobald diese zum Ergebnis beitragen. Diese Bedingung erfüllt definitiv kein Linear-Phase-Filter mit Verzögerung, aber auf jeden Fall ein

minimalphasiger Filter. Auch in Verbindung mit dem Warping bleibt der Filter minimalphasig (siehe Abbildung 3.21).
Um den Filter hierfür zu konstruieren, wird das in [PL06] beschriebene Verfahren mit Hilfe des Cepstrums verwendet. Abbildung 3.19 skizziert die Schritte:
Zuerst werden die $M = 31$ Filterbänder wieder zu den $n = 64$ FFT-Bändern ergänzt. Die Werte 0 und $N/2$ sind dabei 0 zu setzen, dazwischen stehen die Filterbänder. Die Werte von $N/2 + 1$ bis $N - 1$ werden mit den spiegelverkehrt angeordneten Filterbändern aufgefüllt.
Der nun ideal vorgegebene Frequenzverlauf kann mit den vorhandenen Fenster von 64 Bins aber nicht perfekt approximiert werden. Es tritt das zum Beispiel auch in [OSB99] anhand eines Tiefpasses erklärte Gibbssche Phänomen auf. Die Impulsantwort wird mit dem Rechteck-Fenster im Zeitbereich multipliziert. Im Frequenzbereich bedeutet dies eine Faltung des gewünschten Frequenzverlaufs dem des Rechteckfensters. Vor allem bei starken Änderungen (wie etwa einem perfekten Tiefpass- oder Notchfilter) von einem Bin zum nächsten ergibt sich eine deutliche Welligkeit im Spektrum. Dies ist hier sehr ungünstig, besser wäre ein glatterer Verlauf mit eventuell geringerer Trennschärfe. Das übliche Verfahren bei vorgegebener idealer Impulsantwort wäre eine Fensterung mit einem längeren und glatterem Fenster (z.B. von-Hann). Da der Frequenzverlauf direkt eingegeben wird, wird dieser Vorgang durch das interpolieren zusätzlicher Stützpunkte simuliert. Hier genügte eine lineare Interpolation um den Faktor 2. Es wird also immer ein Zwischenwert zwischen den Bins errechnet werden.
Anschließend folgt die Umrechnung von dB zum natürlichen Logarithmus:

$$\ln(\Delta G_{fin}) = \frac{\Delta G_{fin}[dB]}{20} \cdot \ln 10 \tag{3.33}$$

Mit einer anschließenden iFFT erfolgt die Transformation in das reale Cepstrum. Das Cepstrum $\hat{g}$ wird nun mit diesem Fenster multipliziert:

$$w_{Cep}(k) = \begin{cases} 1 & \text{falls } k = 0,\ \frac{\dot{N}}{2} \\ 2 & \text{falls } 1 < k \leq \frac{\dot{N}}{2} - 1 \\ 0 & \text{falls } \frac{\dot{N}}{2} + 1 < k \leq \dot{N} - 1 \end{cases}$$

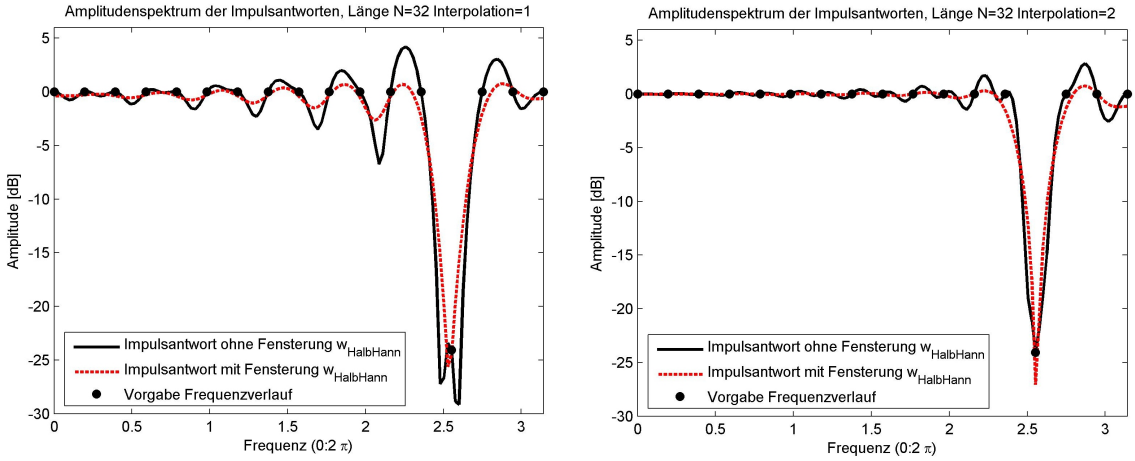

Abbildung 3.20: Vergleich der Amplitudenspektra einer generierten Impulsantwort ohne zusätzliche Interpolation (links) und mit (rechts); hier zu Demonstrationszwecken jeweils mit der endgültigen Länge von 32 Samples nach der letzten Fensterung mit $w_{HalbHann}$

Die minimalphasige Impulsantwort kann anschließend direkt aus dem Cepstrum in den Zeitbereich rekursiv berechnet werden [PL06]:

$$h(l) = \begin{cases} e^{\hat{g}(0)} & \text{falls } l = 0 \\ \sum_{k=1}^{l} (\frac{k}{l}) \hat{g}(k) h(l-k) & \text{falls } l = 1 \ldots N-1 \end{cases}$$

$\hat{g}$: Cepstrum der Impulsantwort, Ausgangsgröße
h : Impulsantwort, Zielgröße

Um Artefakte wegen des abrupten Endes zu vermeiden wird die Impulsantwort zuletzt noch mit einem halbierten (nur der rechte, fallende Teil) von-Hann-Fenster multipliziert. Andere Fenster, wie etwa das Tukey-Fenster wären auch möglich.

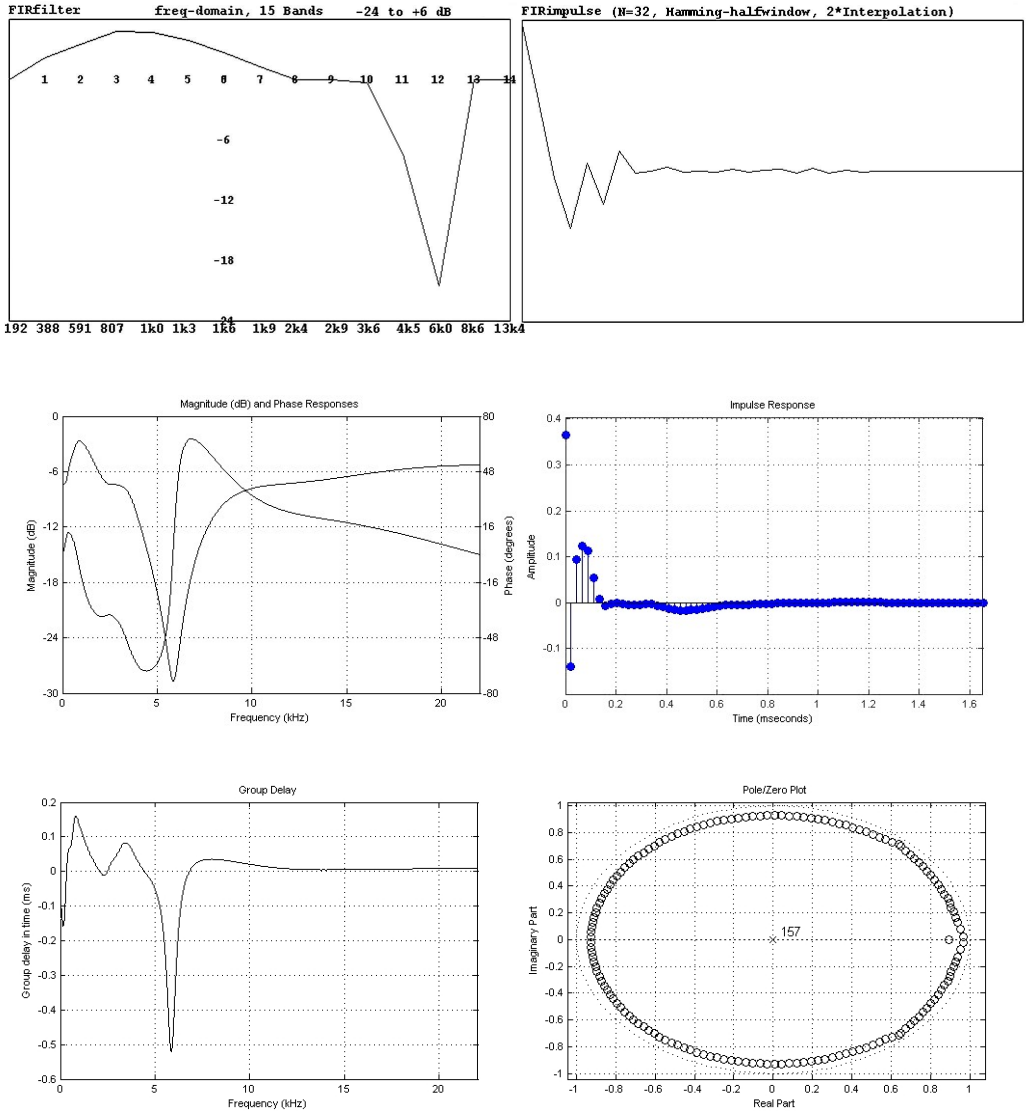

Abbildung 3.21: Beispiel eines konstruierten Warped-Minimum-Phase-FIR-Filters, $M = 15$ Filterbänder, $N = 32$ Samples Länge; Analyse durch Messen der Impulsantwort am Systemausgang

3.5.2 Linearphasiger Filter

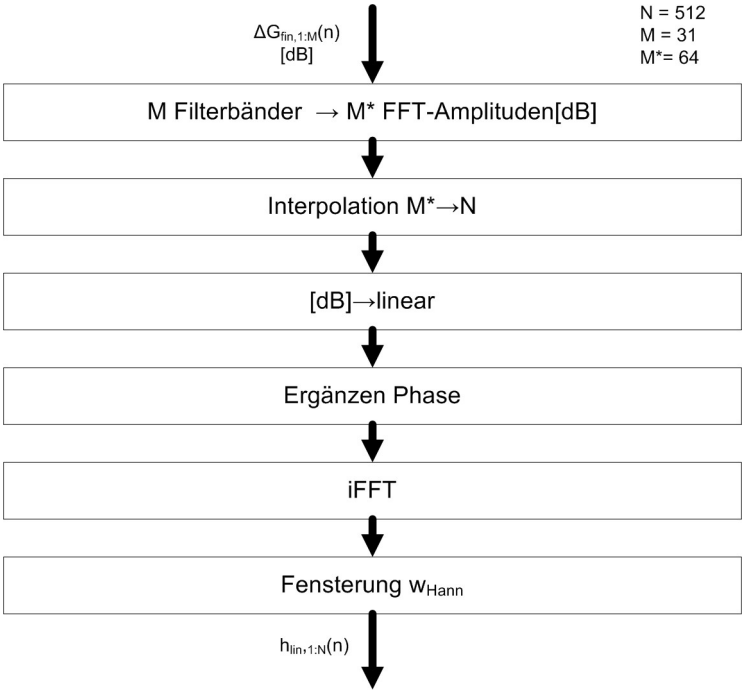

Abbildung 3.22: Konstruktion des linearphasigen Filters aus den Steuerfaktoren der Bänder

Der linearphasige Filter ist wegen der höheren Koeffizientenzahl rechenaufwendiger in der Erstellung und auch in der Anwendung. Als Vorteile schlagen bei ihm die bessere Impulstreue und auch der geringere Speicheraufwand zu Buche. Bei dem Warped-Filter werden vergleichsweise dazu 64 Speicherstellen pro Sample Verzögerungsausgleich mit der Warped-Delay-Line belegt.

Die Erstellung linearphasiger Filter im Frequenzbereich ist relativ trivial. Zuerst werden wie im minimalphasigen Fall die 31 Filterbänder in ein FFT-Amplitudenspektrum umgestellt und ergänzt. Anschließend werden die nichtlinear aufgeteilten Barkbänder mittels Interpolation in die gleichmässig frequenzdiskretisierte „normale“ Frequenzdomäne übertragen. Um die gleiche Auflösung im Tiefbassbereich zu erhalten, sind nun deutlich mehr Koeffizienten (512 statt 64) nötig. Die Amplitudenkurve wird anschließend potenziert ([dB] zu [linear]) und mit

der richtigen Phase versehen [Sal98]:

$$\theta(k) = \begin{cases} -k(\frac{(N-1)}{N}\pi) & \text{falls } 0 < k \leq \frac{N}{2} - 1 \\ 0 & \text{falls } k = \frac{N}{2} \\ (N-k)\frac{(N-1)}{N}\pi & \text{falls } \frac{N}{2} + 1 < k \leq N - 1 \end{cases}$$

Dann wird die $iFFT$ ausgeführt und damit die Impulsantwort erzeugt. Eine abschließende Fensterung mit einem hier nun vollständigen von-Hann-Fenster verringert auch hier Welligkeit der Impulsantwort im Frequenzbereich in Folge des Gibbsschen Phänomens.

Kapitel 4

Hörtest

Abbildung 4.1: Der Testaufbau

Abschließend wurde ein informeller Hörtest vorgenommen, um die Wirkung des neuen Kompressors zu überprüfen und seine Leistung einzuordnen. Hierfür wurde sowohl die einkanalige Fullband-Version wie auch die mehrkanalige Variante des Algorithmus mit einem aktuellen Studiokompressor-Plugin, dem „Sonnox Dynamics“ verglichen. Dessen Algorithmen sind

auch in den Sony-Oxford Rundfunkmischpulten enthalten und gelten klanglich als sehr neutral und transparent.

4.1 Versuchsaufbau

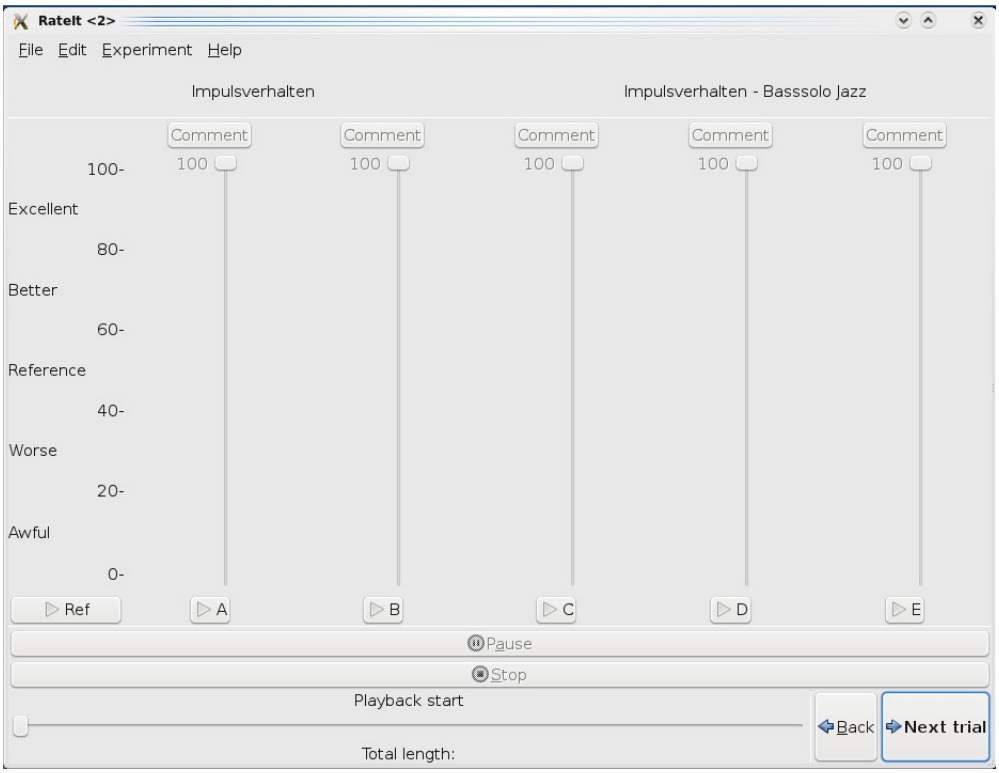

Abbildung 4.2: Die Bedienoberfläche des Hörtests

Die grundsätzliche Vorgangsweise des Versuchs ist stark an den bekannten „MUSHRA“-Test angelehnt. MUSHRA steht dabei für „Multiple Stimuli with Hidden Reference and Anchor“. Es werden dem Probanden pro Versuch mehrere, hier 5, verschieden bearbeitete Beispiele desselben Musikausschnittes zum selbstständigen Vergleich angeboten. Ziel ist es, die verschiedenen Bearbeitungen jeweils auf einer Skala von 0 bis 100 einzuordnen. Eines davon ist die Referenz selbst („Hidden Reference“). Diese soll per Definition mit 50 gewertet werden und dient später zum Prüfen der Zuverlässigkeit der Probanden. Einzelne Proben können

aber auch besser als die Referenz, also > 50, bewertet werden. Ein Anker, also ein sehr schlecht klingendes Beispiel wurde angeboten, aber in weiterer Konsequenz nicht verwendet. Die Positionen der einzelnen Bearbeitungen wechseln von Proband zu Proband, um Absprachen zu verhindern.
Der Test selbst fand kopfhörerbasiert im Produktionsstudio des Instituts für Elektronische Musik und Akustik der Universität für Musik und darstellende Kunst Graz statt. Bei dem Kopfhörer handelte es sich um einen Elektrostaten der Firma STAX (SR-007 in normalen Ausgangskanal des Vorverstärkers SRM-007t). Die allgemeine Wiedergabelautstärke der Orginaldateien in $dB-SPL$ (kurze Zeitkonstante, RLB-Gewichtung) wurde auf $76dB$ festgelegt. Diese Lautstärke wurde garantiert durch eine wiederholte Kalibrierung des Wiedergabeapparates durch einen Bruel&Kjaer Kunstkopf nebst Normschallquelle. Die Testoberfläche „Rate-It“ wurde auf einem Laptop mit einem RME Fireface 800 als Ausgabequelle betrieben. Der hauptsächlich durch den Laptoplüfter hervorgerufene Störpegel betrug am Platz des Hörers etwa $32dB$ linear. Mit der hier anwendbaren $dB-A$ Gewichtung wird dieser Störpegel sogar noch geringer. Somit stehen mindestens $44dB$ Dynamik zur Verfügung, genug um die Testsamples auch unkomprimiert praktisch störungsfrei und voll hörbar wiederzugeben.

4.2 Hörbeispiele

Es wurden 3 Testreihen zu den Fragestellungen Gesamteindruck, Impulsverhalten und Klang durchgeführt. Die Probanden waren 14 Studenten im Alter von 21 bis 31 Jahren. Der überwiegende Teil studiert den Studiengang Elektrotechnik-Toningenieur und verfügt mindestens über musikalische Grundkenntnisse. Etwa die Hälfte ist häufig an audiotechnischen Produktionen beteiligt oder besitzt Erfahrung als Testhörer. Die 6 Audiobeispiele umfassen verschiedene Quellen und Stile und sollen somit möglichst unterschiedliche Anforderungen an einen Dynamikprozessor stellen. Alle besitzen eine relativ große Dynamik und sind nicht oder minimal mit Kompressoren vorbearbeitet. Dies bezieht sich vor allem auf den

stilgemäß zurückhaltenden, gestalterischen Einsatz von Kompressoren auf einzelnen Instrumenten (Snare, Basedrum...) im ersten Beispiel. Die Länge der Beispiele beträgt jeweils 5 bis 10 Sekunden und umfasst immer mindestens eine musikalische Phrase. Alle Beispiele bis auf die Jazzstücke stammen von der „Sound Quality Assessment Material"-CD der Europäischen Rundfunkunion EBU.

Das erste Beispiel ist aus dem Jazzbereich: Es handelt sich um ein eher leises Kontrabasssolo, begleitet von einem flächig auf den Becken spielendem Schlagzeug mit gelegentlichen Snare- und Basedrum-Akzenten sowie einem dezent begleitenden Klavier. Es stammt aus einer laufenden Liveproduktion mit Grazer Jazzstudenten. Dieses Beispiel ist die geradezu klassische Konfiguration für den Komodulationseffekt. Die Gefahr ist groß, dass laute Basspassagen und vor allem Basedrumschläge die „Ambience" durch die Becken mitmodulieren.
Das zweite Exempel lenkt den Fokus auf die transienten Eigenschaften. Es handelt sich um ein Kastagnettenpattern mit definierter Räumlichkeit.
Als nächstes eine Aufnahme mit außergewöhnlich hohem Dynamikumfang aus dem Jazzbereich, ein Klaviertrio. Das Klavier soliert mit Bassbegleitung. Dazwischen kommen vier sehr plötzliche und prominente Snareschläge. Es stammt von der CD „Ray Bryant: Through the Years Vol 1" und ist eine Stereo-Direktaufnahme ohne Mastering-Kompression.
Das erste klassische Beispiel besteht aus einer gezupften Gitarre solo. Das zweite ist ein Ausschnitt der Arie „Der Hölle Rache kocht in meinem Herzen" aus der „Zauberflöte". Der dynamische Umfang ist sehr groß. Kniffelig sind die versetzten Einsätze des Orchesters und des Soprans im Vordergrund. Auch hier sind Komodulationen zu befürchten.
Der letzte Test ist eine deutsche Sprecherin von der SQAM-CD [1].
In einem ersten Verarbeitungsschritt wurden alle Dateien auf $76dB - SPL$ normiert. Anschließend kam die Bearbeitung mit dem jeweiligen Kompressionsalgorithmus. Um auch den

[1] Aus dem Klang lässt auf eine Mikrofonabnahme in relativ kurzer Entfernung sowie eine kurze Nachhallzeit schließen

gewünschten Verstärkungsgewinn richtig einstufen zu können, werden die Ausgabesamples schließlich auf die gleiche maximale Aussteuerung angehoben. Referenz ist hier das unbearbeitete Sample.
Zu den Verarbeitungsdetails:
In allen Kompressoren wurde der Threshold sowie die Kurve angeglichen. Ersterer befindet sich bei $73dB$, die Ratio der statischen Kompressorkurve betrug 4 bei dem Sonnox-Kompressor und wegen der Arbeit in der Lautheitsdomäne 2.4 bei den hier vorgeschlagenen Algorithmen. Bei allen wurde, um Vergleichbarkeit zu gewährleisten, ein Hard-Knee verwendet. Insgesamt also eine relativ starke Kompression, damit laufend eine Korrektur nötig ist. Um die Dauer des Gesamtdurchganges eines Probanden unter im Schnitt 40 Minuten zu halten, wurde bei dem Sample „Kastagnetten“ die Klangbewertung sowie bei den Samples „Klassische Gitarre“ und “Zauberflöte“ die Bewertung des Impulsverhaltens weggelassen. Weiterhin wurden die Probanden mit einer kurzen Übungsphase an die Oberfläche und die Höraufgabe herangeführt. Der Einführungsdurchgang bestand aus einem einzelnen, später nicht mehr verwendeten männlichen Sprachsample von der „SQAM“-CD dessen Bearbeitungen ebenso wie die späteren wirklichen Tests mit der Oberfläche bewertet werden mussten.

4.3 Ergebnisse

Die erste Erkenntnis aus dem Hörtest war die Feststellung, dass eine Beurteilung der Dynamikverarbeitung offensichtlich nicht trivial ist. Vor allem bei der Beurteilung des Impulsverhaltens ergab sich bei einer Auswertung mit allen 14 Probanden eine sehr große Streuung - sinnvolle Schlüsse waren kaum möglich.
Bei genauerer Begutachtung wies eine relativ große Gruppe von 4 Personen bei der Beurteilung der Referenz eine vergleichsweise große Streuung beziehungsweise mehr als 3 deutliche Ausreißer auf. Das Herausnehmen dieser Gruppe (Probanden 2,3,7,12) verringert die Unschärfe deutlich. Alle weiteren Auswertungen sind nun ohne diese Probanden. Auffällig

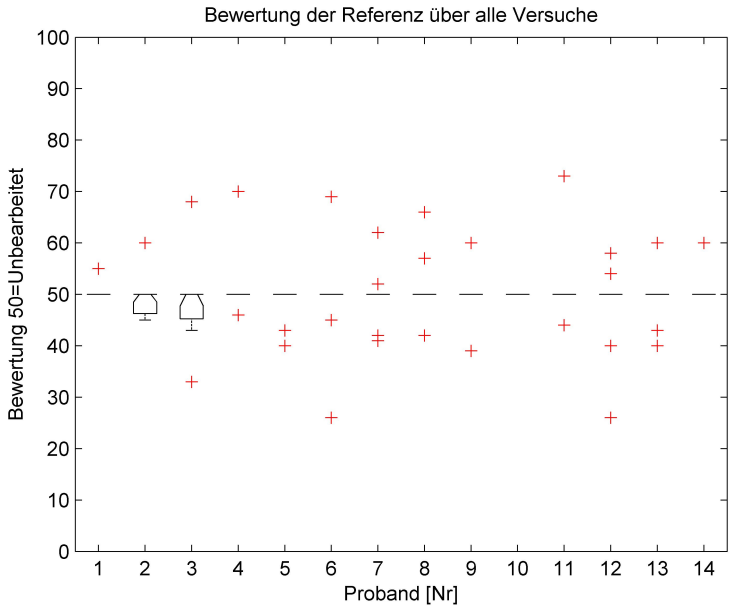

Abbildung 4.3: Bewertung der Referenz in allen Versuchen durch jeweilige Probanden

ist, dass fast nur Probanden übrig blieben, die häufig selbst im Rahmen von Audioproduktionen oder Lifebeschallungen mit Dynamikprozessoren hantieren oder regelmäßig an Hörtests partizipieren.

Ein Proband war bei späterer Nachprüfung sogar in der Lage, die verschiedenen Kompressoren wiederholt richtig zu erkennen. Die Mehrzahl der anderen Probanden äußerte sich in der Nachbesprechung über Auffälligkeiten und besondere Eindrücke. Die meisten anschließenden Folgerungen stammen aus diesen Gesprächen.

In den Plots ist der Median mit seinem Konfidenzintervall (als Kerben) und der Interquartilsabstand (die „Box“). Die dünnen Linien unterhalb und oberhalb der Box werden „Whisker“ genannt und bezeichnen das Konfidenzintervall von 95% in dem sich alle Bewertungen befinden. Treten deutliche Ausreißer auf, so werden sie als Stern eingetragen.

Beim ersten Beispiel, dem Basssolo (Abbildung 4.4), zeigt sich fast erwartungsgemäß eine Präferenz für den Multibandansatz. Bei den beiden anderen bemängelten einige Probanden

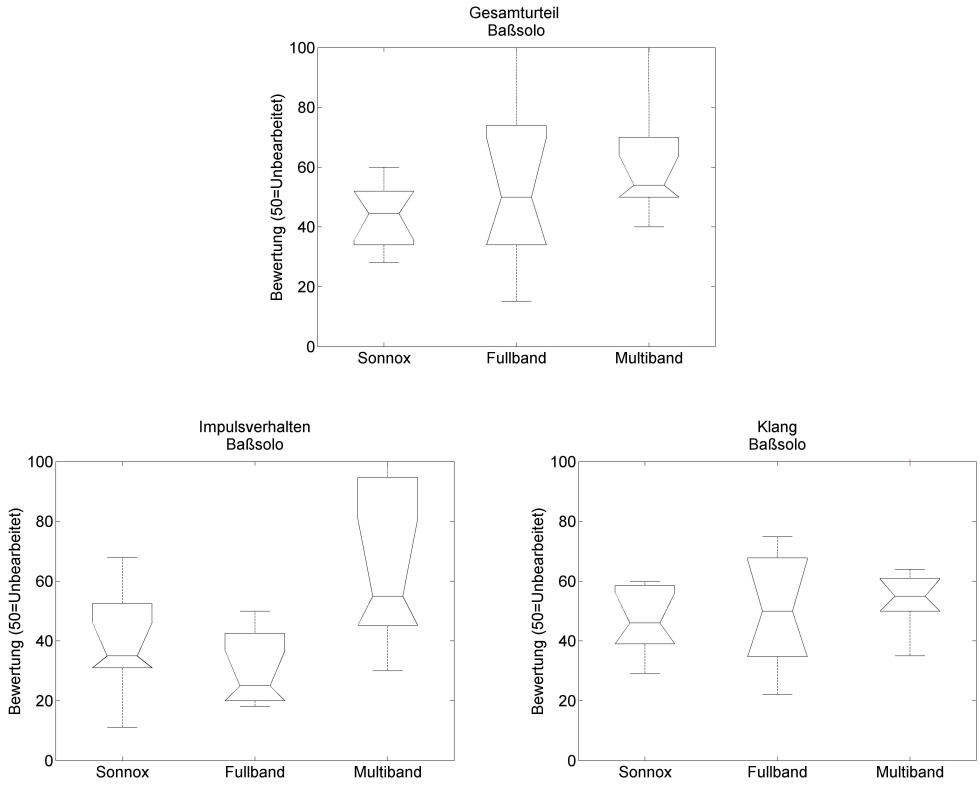

Abbildung 4.4: Ergebnisse des Hörtests für das Sample „Basssolo“; Gesamteindruck, Impulsverhalten und Klang

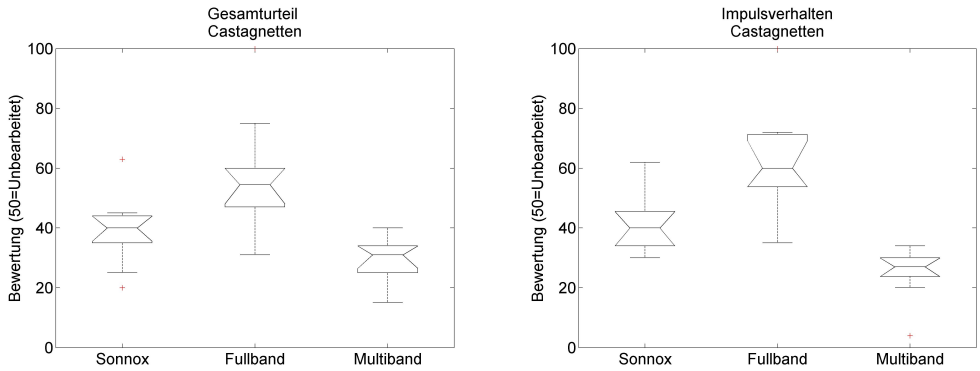

Abbildung 4.5: Ergebnisse des Hörtests für das Sample „Kastagnetten“,; Gesamteindruck und Impulsverhalten

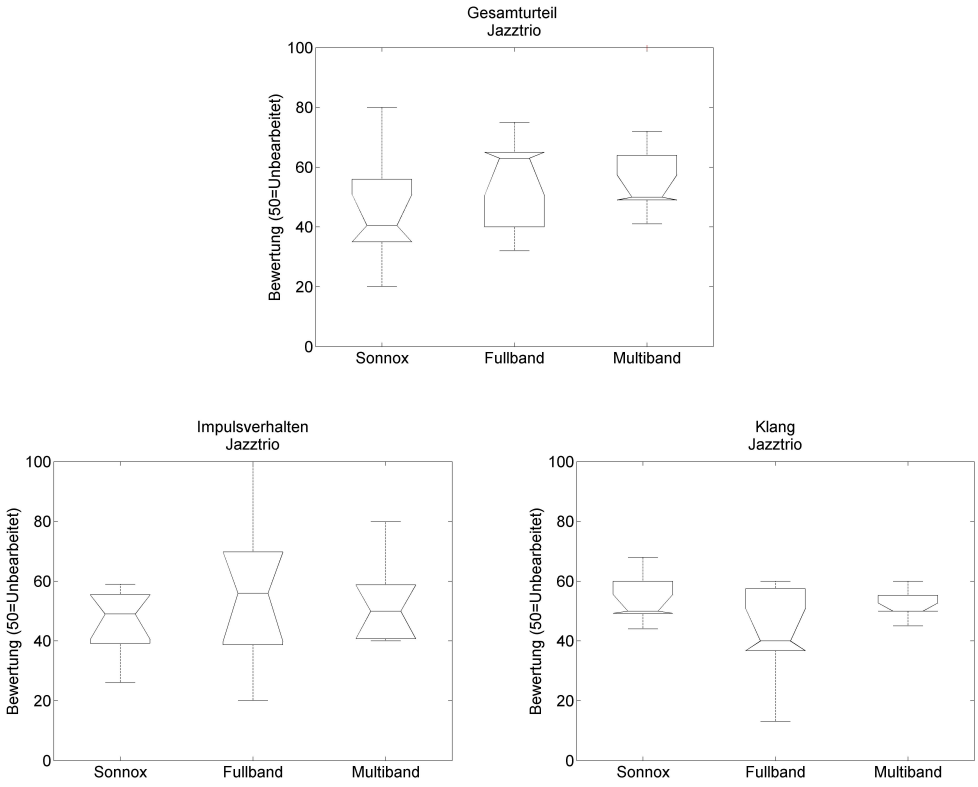

Abbildung 4.6: Ergebnisse des Hörtests für das Sample „Jazztrio“; Gesamteindruck, Impulsverhalten und Klang

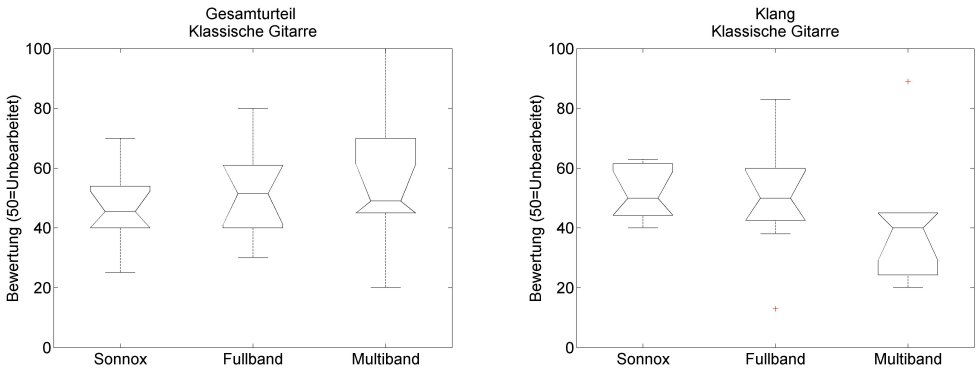

Abbildung 4.7: Ergebnisse des Hörtests für das Sample „Klassische Gitarre“; Gesamteindruck und Klang

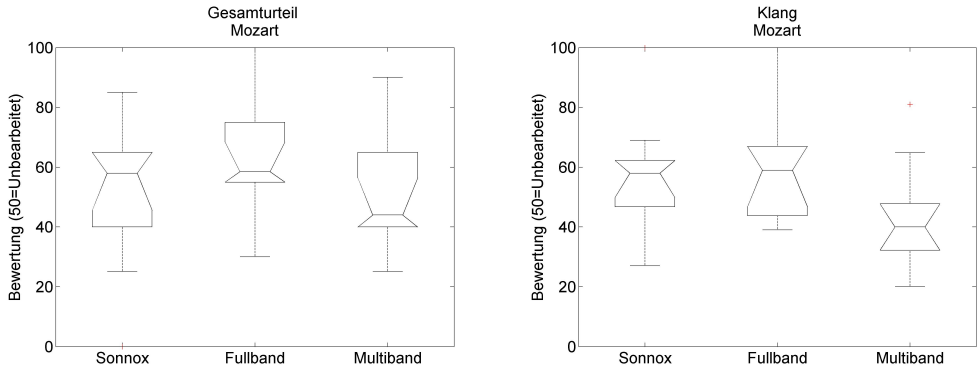

Abbildung 4.8: Ergebnisse des Hörtests für das Sample „Zauberflöte“; Gesamteindruck und Klang

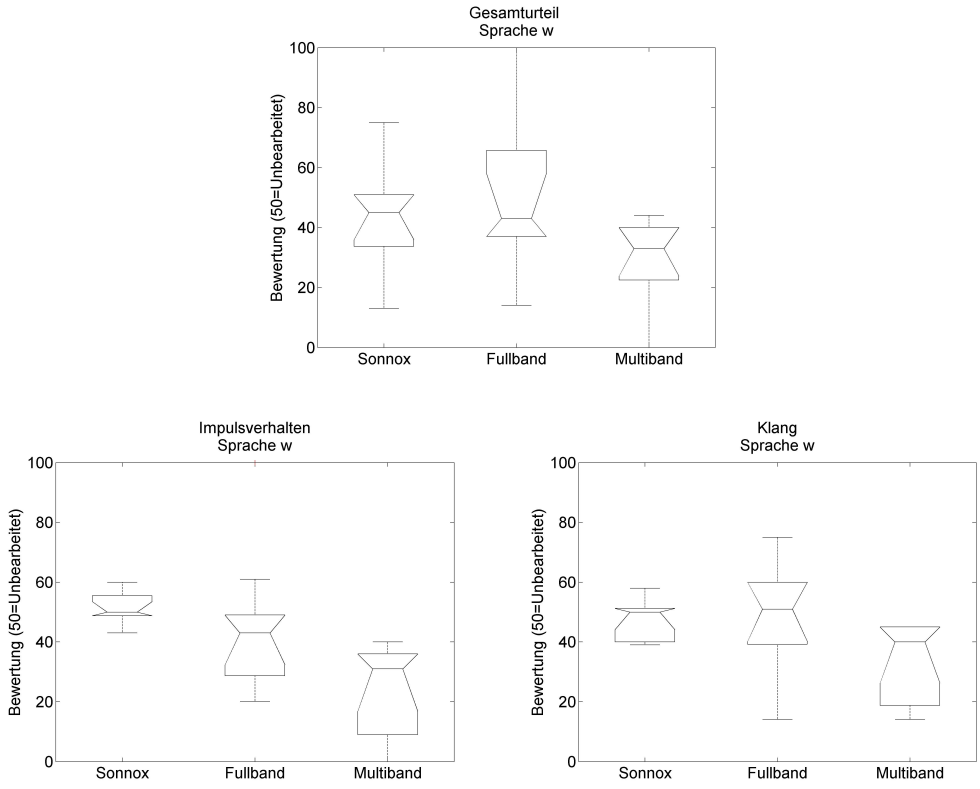

Abbildung 4.9: Ergebnisse des Hörtests für das Sample „Sprache weiblich“; Gesamteindruck, Impulsverhalten und Klang

das deutliche „Wegdrücken“ des gesamten Klangbilds bei Basedrumimpulsen oder exponierten Bassnoten. Dies schlägt sich deutlich im Impulsverhalten nieder.
Das Beispiel der Kastagnetten (Abbildung 4.5) zeigt ein deutliches Ergebnis zu Gunsten des Fullband-Ansatzes. Die „Boxes“ sind hier auch eindeutig voneinander getrennt, die Aussage ist somit relativ verlässlich. Der Sonnox-Kompressor fällt hier ab, der Multiband-Kompressor sogar deutlich.
Beim Jazztrio (Abbildung 4.6) fällt die Abstufung schon schwerer. Vor allem in der dezidierten Bewertung des Impulsverhaltens lassen sich die drei Algorithmen kaum voneinander unterscheiden. Die Konfidenzintervalle der Mediane des Fullband- und des Multiband-Ansatzes überschneiden sich, während der Sonnox-Kompressor doch etwas abfällt
Aus Abbildung 4.7 lässt sich nun kaum mehr ein Favorit bestimmen, die Konfidenzintervalle der Mediane überschneiden sich. Interessant ist, dass bei diesem Test 3 Probanden äußerten, zwar Unterschiede gehört zu haben, aber diese nicht als „besser“ oder „schlechter“ einordnen wollten. Dies ist insofern bemerkenswert, als auch alle 3 Kompressoren genau so gut beurteilt werden wie das Orginal, obwohl durch die Kompression ein Gaingewinn zustande kam.
Bei dem Ausschnitt aus der Zauberflöte (Abbildung 4.8) ist die größte Dynamik zu bewältigen. Durch das starke Aufholen der leisen Bereiche ändert sich das Klangbild durchaus wahrnehmbar. Um so erstaunlicher ist das schlechte Abschneiden des Multibandprozessors. Dieser sollte eigentlich genau in solchen Situationen die klangliche Balance konservieren. Auf Nachfragen wurde das deutliche frequenzabhängige Aufholen der Mitten und Höhen bei kurzen Pausen der Solistin bemängelt. Tatsächlich scheint das System im Hochtonbereich zu träge zu regeln.
In der Beurteilung bei Sprache (Abbildung 4.9) scheint nun erstmals der Sonnox-Kompressor die 1. Wahl zu sein. Die Bewertung des Gesamteindrucks und des Klanges zeigt wegen der starken Überschneidungen der Mediankonfidenzintervalle keinen signifikanten Unterschied zwischen dem Full-Band-Algorithmus und dem Sonox-Kompressor. Erstaunlich ist die signifikant bessere Bewertung des Impulsverhaltens des Sonnox-Kompressors. Dies steht ein wenig im Wiederspruch mit den bisherigen Beobachtungen. Der Multibandansatz mit seinem

weniger brilliantem und sonoren Klangbild wird dagegen abgewertet.

4.4 PEAQ-Test

Feature	Basssolo	Kastagnetten	Git klassisch	Jazztrio	Zauberflöte	Sprache w
Hörtest	MB (54)	FB (54.5)	FB (51.5)	FB (63)	FB (58.5)	So (45)
- Gesamt	FB (50)	So (40)	MB (49)	MB (50)	So (58)	FB (43)
	So (44.5)	MB (31)	So (45.5)	So (40.5)	MB (44)	MB (33)
Objective Difference	MB (-2.49895)	FB (-0.584621)	FB (-1.4636)	So (-2.76749)	MB (-3.52005)	So (-0.742361)
Grade	So (-3.00503)	MB (-3.1985)	So (-1.77429)	MB (-2.90985)	So (-3.53891)	MB (-1.12975)
(AV)	FB (-3.17122)	So (-3.76291)	MB (-2.1718)	FB (-3.55028)	FB (-3.55028)	FB (-3.30575)
Distortion	MB (-0.607499)	FB (1.4398)	FB (0.401893)	So (-0.901738)	MB (-2.09574)	So (1.211321)
Index	So (-1.1963)	MB (-1.47574)	So (0.100757)	MB(-1.07319)	So (-2.14264)	MB (0.747488)
(AV)	FB (-1.43342)	So (-2.90945)	MB (-0.279716)	FB(-2.2195)	FB (-2.17177)	FB (-1.65424)
RMS Modulation	MB (135.926)	FB (131.56)	FB (142.038)	MB (135.569)	FB (139.833)	FB (153.97)
Difference	FB (136.782)	MB (148.229)	So (147.669)	FB (136.977)	MB (143.292)	So (160.147)
	So (141.351)	So (400.071)	MB (152.623)	So (137.465)	So (145.172)	MB (194.581)
Averaged Linear	MB (1.49618)	FB (0.682617)	MB (1.15184)	MB (2.5617)	FB (79.2963)	So (1.62137)
Distortions	So (3.97064)	So (1.30607)	So (1.33185)	So (3.20947)	MB (127.933)	MB (2.26959)
	FB (5.19511)	MB (1.9641)	FB (1.39173)	FB (36.0455)	So (142.951)	FB (9.21125)
Harmonic Structure	FB (0.179773)	So (0.153997)	So (0.320727)	So (0.265906)	So (0.138401)	FB (0.287436)
of Error over Time	So (0.194717)	FB (0.185388)	FB (0.332202)	FB (0.271787)	FB (0.146859)	So (0.29407)
(FFT)	MB (1.16937)	MB (1.84895)	MB (0.724898)	MB (1.51177)	MB (1.33684)	MB (0.991174)
RMS Noise Loudness	MB (6.1543)	So (5.12078)	FB (6.86198)	So (5.77373)	FB (6.02914)	So (13.8411)
Asym	FB (6.40811)	FB (6.2271)	So (7.18654)	MB (6.27476)	MB (6.26954)	FB (15.3016)
	So (6.51756)	MB (6.86198)	MB (7.73866)	FB (6.64965)	So (6.34093)	MB (20.1515)
Segmental Noise	FB (-8.65244)	FB (-10.3831)	FB (-8.4153)	So (-9.5105)	FB (-7.20962)	FB (-8.06952)
Masking Ratio	So (-7.78936)	So (-9.14373)	So (-6.74705)	FB (-9.20166)	So (-5.5604)	So (-7.65969)
(FFT)	MB (-7.34519)	MB (-5.70915)	MB (-4.48321)	MB (-7.78283)	MB (-5.09431)	MB (-5.2677)

Abbildung 4.10: Auswertung des PEAQ-Tests: dunkelgrau unterlegt bei wahrscheinlich geringem Zusammenhang des jeweiligen Features; Ergebnisse des Hörtests(Gesamturteil): hellgrau unterlegt, falls sich die Konfidenzintervalle des Median überschneiden; So - kommerzieller Sonnox Kompressor, FB - Fullband Kompressor, MB - Multibandkompressor

Die Hörbeispiele wurden abschließend noch mit dem „Perceptual Quality Assessment for Digital Audio“(PEAQ)-Test beurteilt. Dieser Test wurde für die vor allem für die Evaluierung verlustbehafteter Kodierungsalgorithmen für Breitbandaudio entwickelt und enthält ein elaboriertes Hörmodell. Um eine möglichst hohe Zeitauflösung zu erhalten, wurde der

Filtebankbasierte „Advanced“- Modus benutzt. Als Ergebnis gibt das System zwei alternative Gesamtbewertungen (Objective Difference Grade, ODG sowie Distortion Index, DI) aus. Diese werden aus 5 „Features“ also ermittelten Signaleigenschaften gebildet.
Die absoluten Werte vor allem des ODG zeigen insgesamt eine deutliche Abwertung. In den Hörtests befand sich meist aber zumindest ein Algorithmus in der Nähe der Referenz. Die schlechte Absolutbewertung des PEAQ findet sich also nicht in den Hörtests wieder, der Test ist damit für die Bewertung von Kompressionsalgorithmen nicht geeignet. Betrachtet man die relative Reihung der Algorithmen untereinander, so scheint die Bewertung nur in manchen Fällen mit den Ergebnissen des Hörtests übereinzustimmen. Bei den meisten der zugrundeliegenden Features scheint gleiches zu gelten: Die Eigenschaften „Segmental Noise Masking Ratio“ und „Rms Noise Loudness Asym“ dienen zur Beurteilung von hinzugefügtem (Quantiesierungs-)Rauschen bzw. weggelassener Signalanteile und die Wahrnehmbarkeit dieser Fehler. Da kein Rauschen hinzugefügt wird, laufen diese beiden Kennzeichen fast vorhersehbar ins Leere. Die Aussagekraft des Features „Harmonic Structure of Error over Time“ scheint gering. Als prinzipiell aussichtsreich, jedoch letztendlich nicht eindeutig wäre die Beurteilung der „Averaged linear Distortions“ zu nennen. Einzig bei dem quadratischen Mittel der „Modulation Difference“ scheint es eine wahrnehmbare Ähnlichkeit in der Reihung zum Hörtest gegeben. Dies ließe sich als vorsichtige Bestätigung der These werten, dass es wesentlich für die Dynamikbearbeitung sei, möglichst unhörbar zu regeln.

4.5 Weitere Schlüsse und Interpretationen

Insgesamt lässt sich sagen, dass sich der Single-Band-Ansatz dem kommerziellen Sonnox auf alle Fälle ebenbürtig erwiesen hat und beim Vorhandensein ausgeprägter Transienten sogar besser abschneidet. Der Multibandansatz zeigt durch seine besseren Abschätzungen der Lautheit im Bassbereich und unabhängige Steuerung der Bänder in Komodulationsaufgaben seine Stärken. Das Impulsverhalten bei deutlichen und vor allem hochfrequenten Transien-

ten führt jedoch oft zur deutlichen Abwertung. Der Vergleich des Warped-Filters mit dem Linear-Phase-Filter zeigte in einem kurzen späteren Vergleich zwar eine leichte Besserung, aber nicht in dem festgestellten Ausmaß. Es scheint nicht unwahrscheinlich, dass die Filter und Verdeckungszeiten (im Lautheitsmodell) entgegen den Ergebnissen aus Hörversuchen zu lang gewählt wurden. Das praktisch optimierte PEAQ-Modell verfügt zwar generell über sehr kurz angesetzte Nachverdeckungszeiten, diese nehmen zu hohen Frequenzen aber noch einmal deutlich ab: von $50ms$ bei $100Hz$ zu $4ms$ bei hohen Frequenzen [CSTT99]. Die meisten kommerziellen Multiband-Kompressoren besitzen zudem in hohen Frequenzbändern über deutlich kürzere Attack- und Releasezeiten (sogar im einstelligen Milisekundenbereich) als in den Bässen und Mitten.

Ein weiterer Schluss wäre die Feststellung, dass die gewünschte Beibehaltung der spektralen Balance mittels des Multibandansatzes bei der Kompression hinter eine möglichst ideale Zeitsteuerung zurücktritt.

Im Hinblick auf weitere Hörversuche in diesem Bereich lässt sich wohl begründet schließen, dass als Probanden nur noch wirklich geübte Hörer in Frage kommen. Zukünftige Aufgabenstellungen könnten sich analog der Versuchsreihen für Sprache in ([MPS99], [SMAG99], [SM03], [SM04], [SM07], [SM08]) mit der Durchsichtigkeit und Wahrnehmbarkeit auch bei lauten Hintergrundkulissen beschäftigen.

Kapitel 5

Ausblick

Durch den erfolgreichen Einbau eines Hörmodells zur Analyse und Regelung ist die Brücke zwischen psychoakustischer Theorie und praktischen Einsatz geschlagen. Die Konstruktion des Kompressors führte zu einer Frequenz-Zeit-Bearbeitungsumgebung mit adaptivem zeitlichen Regelverhalten. Dieses selbst stellt bereits in der einkanaligen Breitbandimplementierung im Vergleich zu einem kommerziellen „State-of-the-Art“-Kompressor eine teils spürbare Verbesserung dar. Die mehrkanalige Vollversion erwies sich dagegen bei dichtem und komplexem Material mit mehreren parallelen Ereignissen als vorteilhaft, zeigte jedoch im Falle von hochfrequenten Impulsen Verbesserungsbedarf. Mit einer weiteren frequenzabhängigen Optimierung der Glättungs- und Verdeckungszeiten wie im PEAQ-Modell sowie der Integration der Simultanverdeckung durch asymetrische Gehörfilter, lässt sich dies aber sicherlich verbessern.

Künftige Entwicklungsschritte lassen sich relativ einfach erkennen: Die Umgebung ist auch für andere Aufgaben einsetzbar. Insbesondere wäre eine Modifikation für Denoisingaufgaben interessant, da sich das unnatürliche abrupte Abschneiden („Gate“) von Hallfahnen oder Onsets durch die vorausschauende Regelung mit dieser Umgebung verhindern ließe.
Durch eine Erweiterung zu einem binauralen Modell wie zum Beispiel in [Kar96] oder [MG07]

und davor platziertem räumlichen Modell ließe sich das Verfahren auf beliebige mehrkanalige Wiedergabesituationen erweitern. Dies wäre im Vergleich zu den sich auf den Markt befindlichen händisch optimierten Surroundlösungen ein vielversprechender neuer Ansatz.
Zu guter Letzt steckt in dem neuronalen Rückkopplungsmodell nach Ansicht des Autors ebenfalls noch deutliches Potenzial. Durch eine Weiterentwicklung ließen sich eventuell einige nichtlineare Erscheinungen der Cochlea gut nachvollziehen. Durch eine Verknüpfung der einzelnen Bänder über den Rückkopplungspfad (die Verstärkung wirkt im Ohr breiter zurück als die Ursache, [FZ07]) ließen sich möglicherweise bisher nur aufwendig beschreibbare bandübergreifende Effekte wie die „Two-Tone-Suppression“ oder das „Overshoot-Masking“ modellieren. In der Audiologie könnten explizit Hörschäden an den Inneren („Eingang“) oder Äußeren Haarzellen („Rückkopplungspfad“) modelliert werden.

Abbildungsverzeichnis

Literaturverzeichnis

[Bau97] Baumgarte, F.: A physiological ear model for specific loudness and masking. In: *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1997, S. 4 pp.–

[Bau02] Baumgarte, F.: Improved audio coding using a psychoacoustic model based on a cochlear filter bank. In: *IEEE Trans. Speech Audio Process.* 10 (2002), Oct., Nr. 7, S. 495–503. http://dx.doi.org/10.1109/TSA.2002.804536. – DOI 10.1109/TSA.2002.804536

[BCA+01] Bellini, A. ; Cibelli, G. ; Armelloni, E. ; Ugolotti, E. ; Farina, A.: Car cockpit equalization by warping filters. In: *IEEE Trans. Consum. Electron.* 47 (2001), Nr. 1, S. 108–116. – ISSN 0098–3063

[Ble69] Blesser, B.: Audio dynamic range compression for minimum perceived distortion. In: *IEEE Trans. Audio Electroacoust.* 17 (1969), Nr. 1, S. 22–32. – ISSN 0018–9278

[CSTT99] Colomes, Catherine ; Schmidmer, Christian ; Thiede, Thilo ; Treurniet, William C.: Perceptual Quality Assessment for Digital Audio: PEAQ-The New ITU Standard for Objective Measurement of the Perceived Audio Quality. In: *17th AES International Conference: High-Quality Audio Coding*, 1999, S. 337–351

[dbx08] DBX PROFESSIONAL PRODUCTS (Hrsg.): *AutoVelocity Dynamics and the dbx 160SL - White Paper*. v1. 8760 South Sandy Parkway, Sandy, Utah 84070: dbx Professional Products, 2008

[DS84] DOLAN, Thomas G. ; SMALL, Arnold M.: Frequency effects in backward masking. In: *The Journal of the Acoustical Society of America* 75 (1984), Nr. 3, 932-936. `http://dx.doi.org/10.1121/1.390540`. – DOI 10.1121/1.390540

[Fas76] FASTL, Hugo: Temporal Masking Effects: I. Broad Band Noise Masker. In: *ACUSTICA* 35 (1976), Nr. 5, S. 287–302

[Fas77a] FASTL, Hugo: Temporal Masking Effects:II. Critical Band Noise Masker. In: *ACUSTICA* 36 (1977), Nr. 5, S. 317–331

[Fas77b] FASTL Hugo: Subjective duration and temporal masking patterns of broadband noise impulses. In: *The Journal of the Acoustical Society of America* 61 (1977), Nr. 1, 162-168. `http://dx.doi.org/10.1121/1.381277`. – DOI 10.1121/1.381277

[Fas79] FASTL, Hugo: Temporal Masking Effects:III. Pure Tone Masker. In: *ACUSTICA* 43 (1979), S. 282–293

[FZ07] FASTL, Hugo ; ZWICKER, Eberhard ; HUANG, Thomas S. (Hrsg.) ; SCHROEDER, Manfred R. (Hrsg.) ; KOHONEN, Teuvo (Hrsg.): *Psychoacoustics: Facts and Models - Third Edition*. Springer, 2007

[GA06] GUNAWAN, T. S. ; AMBIKAIRAJAH, E.: A New Forward Masking Model and its Application to Speech Enhancement. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2006* Bd. 1, 2006. – ISSN 1520–6149, S. I–I

[GD97] Garnero, B. ; Drygajlo, A.: Perceptual speech coding using time and frequency masking constraints. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-97* Bd. 2, 1997, S. 1363–1366 vol.2

[Gel04] Gelfand, Stanley A. ; Gelfand, Stanley A. (Hrsg.): *Hearing.* 4th Edition. Marcel Dekker, 2004

[GM02] Glasberg, Brian R. ; Moore, Brian C.J.: A Model of Loudness Applicable to Time-Varying Sounds. In: *JAES* 50 (2002), S. 331–342

[GM06] Glasberg, Brian R. ; Moore, Brian C.J.: Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. In: *The Journal of the Acoustical Society of America* 120 (2006), Nr. 2, 585-588. `http://dx.doi.org/10.1121/1.2214151`. – DOI 10.1121/1.2214151

[Gun07] Gunawan, Teddy Surya: *Audio Compression and Speech Enhancement using Temporal Masking Models*, The University of New South Wales, Diss., 2007

[Hau97] Hauenstein, M.: A computationally efficient algorithm for calculating loudness patterns of narrowband speech. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-97* Bd. 2, 1997, S. 1311–1314 vol.2

[HC02] Huang, Yuan-Hao ; Chiueh, Tzi-Dar: A new audio coding scheme using a forward masking model and perceptually weighted vector quantization. In: *IEEE Trans. Speech Audio Process.* 10 (2002), Nr. 5, S. 325–335. `http://dx.doi.org/10.1109/TSA.2002.800559`. – DOI 10.1109/TSA.2002.800559. – ISSN 1063–6676

[HF99] Heinz, Michael G. ; Formby, C.: Detection of time- and bandlimited increments and decrements in a random-level noise. In: *The Journal of the Acoustical Society*

of America 106 (1999), Nr. 1, 313-326. http://dx.doi.org/10.1121/1.428039. – DOI 10.1121/1.428039

[HKS+00] Härmä, Aki ; Karjalainen, Matti ; Savioja, Lauri ; Välmäki, Vesa ; Laine, Unto K. ; Jyri Huopaniemi: Frequency-Warped Signal Processing for Audio Applications. In: *J. Audio Eng. Soc.* 48 (2000), November, Nr. 11, S. 1011–1031

[HLK99] Härmä, Aki ; Laine, Unto K. ; Karjalainen, Matti: On the utilization of overshoot effects in low-delay audio coding. In: *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing* Bd. 2, 1999, S. 893–896

[Här99a] Härmä, Aki: Low-level auditory modeling of temporal effects. In: Okuno, H. G. (Hrsg.): *16th Int. Joint Conf. on Artificial Intelligence, Workshop on Computational Auditory Scene Analysis*, 1999, S. 1–9

[Här99b] Härmä, Aki: *Psychoacoustic temporal masking effects with artificial and real signals*. 1999. – Hearing Seminar at HUT, 1999

[Här99c] Härmä, Aki: Temporal Masking Effects: Single Incidents / HUT. 1999 (1). – Forschungsbericht. – Internal Report

[IP82] Irwin, R. J. ; Purdy, Suzanne C.: The minimum detectable duration of auditory signals for normal and hearing-impaired listeners. In: *The Journal of the Acoustical Society of America* 71 (1982), Nr. 4, 967-974. http://dx.doi.org/10.1121/1.387578. – DOI 10.1121/1.387578

[ITU06] ITU: Algorithms to measure audio programme loudness and true-peak audio level / Recommendation ITU-R BS.1770 / International Telecommunication Union (ITU). 2006 (ITU-R BS.1770). – Forschungsbericht. – Recommendation

[JBL82] Jesteadt, Walt ; Bacon, Sid P. ; Lehman, James R.: Forward masking as a function of frequency, masker level, and signal delay. In: *The*

Journal of the Acoustical Society of America 71 (1982), Nr. 4, 950-962. http://dx.doi.org/10.1121/1.387576. – DOI 10.1121/1.387576

[Joh88a] Johnston, James D.: Estimation of Perceptual Entropy Using Noise Masking Criteria. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1988. ICASSP-88.*, 1988, S. 2524 – 2527

[Joh88b] Johnston, James D.: Transform Coding of Audio Signals Using Perceptual Noise Criteria. In: *IEEE Journal on Selected Areas in Communications* 6 (1988), February, Nr. 2, S. 314–323

[KA05] Kates, James M. ; Arehart, Kathryn Hoberg: Multichannel Dynamic-Range Compression Using Digital Frequency Warping. In: *EURASIP Journal on Applied Signal Processing* 18 (2005), S. 3003–3018

[Kar96] Karjalainen, M.: A binaural auditory model for sound quality measurements and spatial hearing studies. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-96* Bd. 2, 1996, S. 985–988 vol. 2

[Kar08] Karjalainen, Matti: *Communications Acoustics, Lecture Slides Spring 2008.* 2008. – Helsinki University of Technology

[KK99a] Kubin, G. ; Kleijn, Bastiaan W.: On speech coding in a perceptual domain. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '99* Bd. 1, 1999, S. 205–208 vol.1

[KK99b] Kubin, G. ; Kleijn, W. B.: Multiple-description coding (MDC) of speech with an invertible auditory model. In: *Proc. IEEE Workshop on Speech Coding*, 1999, S. 81–83

[Kla08] Klapuri, A.: Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model. In: *IEEE Transactions on Audio, Speech, and Language Processing* 16 (2008), Nr. 2, S. 255–266. http://dx.doi.org/10.1109/TASL.2007.908129. – DOI 10.1109/TASL.2007.908129. – ISSN 1558–7916

[KSE+09] Keefe, Douglas H. ; Schairer, Kim S. ; Ellison, John C. ; Fitzpatrick, Denis F. ; Jesteadt, Walt: Use of stimulus-frequency otoacoustic emissions to investigate efferent and cochlear contributions to temporal overshoot. In: *The Journal of the Acoustical Society of America* 125 (2009), Nr. 3, 1595-1604. http://dx.doi.org/10.1121/1.3068443. – DOI 10.1121/1.3068443

[LAH01] Lin, L. ; Ambikairajah, E. ; Holmes, W.: Log-magnitude modelling of auditory tuning curves. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)* Bd. 5, 2001, S. 3293–3296 vol.5

[LB97] Lee, Jungmee ; Bacon, Sid P.: Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration. In: *The Journal of the Acoustical Society of America* 101 (1997), Nr. 6, 3688-3693. http://dx.doi.org/10.1121/1.418329. – DOI 10.1121/1.418329

[LBA06] Lin, Andrew ; Berber, Stevan ; Abdulla, Waleed: An investigation of non-uniform bandwidths auditory filterbank in audio coding. In: *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, 2006, S. 360–365

[LHA01] Lin, L. ; Holmes, W. H. ; Ambikairajah, E.: Auditory filter bank inversion. In: *Proc. IEEE International Symposium on Circuits and Systems ISCAS 2001* Bd. 2, 2001, S. 537–540 vol. 2

[Mak03] Makur, A.A.: Computational Schemes for Warped DFT and its Inverse. In: *IEEE Transactions on* 1 1 (2003), Volume PP,&a href='/xpl/tocpreprint.jsp?isnumber=4358591&punumber=8919'¿Forthcoming&, S. 1–1. http://dx.doi.org/10.1109/TCSI.2008.921023. – DOI 10.1109/TCSI.2008.921023. – Accepted for future publication Circuits and Systems I: Regular Papers

[Mak06] Makur, Anamitra: Fast Computation of WDFT and its Application in Image Compression. In: *Proc. TENCON 2006. 2006 IEEE Region 10 Conference*, 2006, S. 1–4

[Map98] Mapes-Riordan, Dan: A Worst-Case Analysis for Analog-Quality (Alias-Free) Digital Dynamics Processing. In: *105. AES Convention*, 1998, S. 1–23

[Mas08] Massenburg, George ; George Massenburg Labs (Hrsg.): *GML 8900 Dynamik Range Controller User Reference*. 1. P.O. Box 1366, Franklin, TN 37065: George Massenburg Labs, December 2008. http://www.massenburg.com/cgi-bin/ml/8900ref.html

[MG07] Moore, Brian C.J. ; Glasberg, Brian R.: Modeling binaural loudness. In: *The Journal of the Acoustical Society of America* 121 (2007), Nr. 3, 1604-1612. http://dx.doi.org/10.1121/1.2431331. – DOI 10.1121/1.2431331

[MGB97] Moore, Brian C.J. ; Glasberg, Brian R. ; Baer, Thomas: A Model for the Prediction of Thresholds, Loudness, and Partial Loudness. In: *JAES* 45 (1997), S. 224–240

[MHY+95] Moore, Brian C.J. ; Hartmann, William Morris ; Yates, Graeme K. ; Palmer, Alan R. ; Plack, Christopher J. ; Carlyon, Robert P. ; Eddins, David A. ; Green, David M. ; Hall, Joseph W. ; Grose, John H. ; Mendoza,

Lee ; Houtsma, Adrianus J.M. ; Grantham, D. Wesley ; Stern, Richard M. ; Trahiotis, Constantine ; Darwin, C.J. ; R.P. Carlyon ; Handel, Stephen ; Moore, Brian C.J. (Hrsg.): *Hearing*. Academic Press, 1995

[MM01] Makur, A. ; Mitra, S.K.: Warped discrete-Fourier transform: Theory and applications. In: *IEEE Trans. Circuits Syst. I* 48 (2001), Nr. 9, S. 1086–1093. http://dx.doi.org/10.1109/81.948436. – DOI 10.1109/81.948436. – ISSN 1057–7122

[MO98] Moore, Brian C.J. ; Oxenham, Andrew J..: Psychoacoustic Consequences of Compression in the Peripheral Auditory System. In: *Psychological Review* 105 (1998), S. 108–124

[Moo93] Moore, Brian C.J.: Characterization of Simultaneous, Forward and Backward Masking. In: *12th International AES Conference: The Perception of Reproduced Sound*, 1993, S. 22–33

[Moo96] Moore, Brian C.J.: Masking in the Human Auditory System. In: *Collected Papers on Digital Audio Bit-Rate Reduction*. AES, 1996, S. 9–19

[Moo03] Moore, Brian C.J. ; Moore, Brian C.J. (Hrsg.): *An Introduction to the Psychology of Hearing, 5th Ed.* Academic Press, 2003

[MPS99] Moore, Brian C.J. ; Peters, Robert W. ; Stone, Michael A.: Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. In: *The Journal of the Acoustical Society of America* 105 (1999), Nr. 1, 400-411. http://dx.doi.org/10.1121/1.424571. – DOI 10.1121/1.424571

[Nov99] Novorita, B.: Incorporation of temporal masking effects into bark spectral distortion measure. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '99* Bd. 2, 1999, S. 665–668 vol.2

[OM95] Oxenham, Andrew J. ; Moore, Brian C.J.: Overshoot and the "severe departure" from Weber's law. In: *The Journal of the Acoustical Society of America* 97 (1995), Nr. 4, 2442-2453. http://dx.doi.org/10.1121/1.411965. – DOI 10.1121/1.411965

[OMV97] Oxenham, Andrew J. ; Moore, Brian C.J. ; Vickers, Deborah A.: Short-term temporal integration: Evidence for the influence of peripheral compression. In: *The Journal of the Acoustical Society of America* 101 (1997), Nr. 6, 3676-3687. http://dx.doi.org/10.1121/1.418328. – DOI 10.1121/1.418328

[OP00] Oxenham, Andrew J. ; Plack, Christopher J.: E¡ects of masker frequency and duration in forward masking: further evidence for the in£uence of peripheral nonlinearity. In: *Hearing Research* 150 (2000), S. 258–266

[Opt01] Opticom: List of corrections of the ITU-R Recommendation BS.1387 / Opticom. 2001 (1). – Forschungsbericht. – PEAQ Corrections

[OS06] Oxenham, Andrew J. ; Simonson, Andrea M.: Level Dependence of Auditory Filters in Nonsimultaneous Masking as a Function of Frequency. In: *Journal of the Acoustic Society of America* 119 (2006), January, Nr. 1, S. 444–453

[OSB99] Oppenheim, Alan V. ; Schafer, Ronald W. ; Buck, John R. ; Robbins, Tom (Hrsg.): *Discrete-Time Signal Processing, 2nd Ed.* Prentice-Hall, 1999

[Oxe97] Oxenham, Andrew J.: Increment and decrement detection in sinusoids as a measure of temporal resolution. In: *The Journal of the Acoustical Society of*

America 102 (1997), Nr. 3, 1779-1790. http://dx.doi.org/10.1121/1.420086. – DOI 10.1121/1.420086

[Oxe98] Oxenham, Andrew J.: Temporal integration at 6 kHz as a function of masker bandwidth. In: *The Journal of the Acoustical Society of America* 103 (1998), Nr. 2, 1033-1042. http://dx.doi.org/10.1121/1.421229. – DOI 10.1121/1.421229

[PL06] Pei, S. C. ; Lin, H. S.: Minimum-Phase FIR Filter Design Using Real Cepstrum. In: *IEEE Trans. Circuits Syst. II* 53 (2006), Nr. 10, S. 1113–1117. http://dx.doi.org/10.1109/TCSII.2006.882193. – DOI 10.1109/TCSII.2006.882193. – ISSN 1549–7747

[PO98] Plack, Christopher J. ; Oxenham, Andrew J.: Basilar-membrane nonlinearity and the growth of forward masking. In: *The Journal of the Acoustical Society of America* 103 (1998), Nr. 3, 1598-1608. http://dx.doi.org/10.1121/1.421294. – DOI 10.1121/1.421294

[POD06] Plack, Christopher J. ; Oxenham, Andrew J. ; Drga, Vit: Masking by Inaudible Sounds and the Linearity of Temporal Summation. In: *The Journal of Neuroscience* 26 (2006), August, Nr. 34, S. 8767–8773

[PS00] Painter, T. ; Spanias, A.: Perceptual coding of digital audio. In: *Proc. IEEE* 88 (2000), Nr. 4, S. 451–515. http://dx.doi.org/10.1109/5.842996. – DOI 10.1109/5.842996. – ISSN 0018–9219

[Saf03] Safe Sound Audio (Hrsg.): *P1 Audio Processor - White Paper*. 1. Leeds: Safe Sound Audio, May 2003

[Saf07] Safe Sound Audio (Hrsg.): *Dynamics Toolbox Audio Processor - White Paper*. 1. Leeds: Safe Sound Audio, November 2007

[Sal98] Salous, S.: The design of linear phase FIR filters using the IDFT. In: *IEEE Trans. Educ.* 41 (1998), Nr. 3, S. 229–231. http://dx.doi.org/10.1109/13.704553. – DOI 10.1109/13.704553. – ISSN 0018–9359

[See07] Seefeldt, Alan: Loudness Domain Signal Processing. In: *Proceedings of the 123th AES Convention*, 2007, S. 1–15

[SM03] Stone, Michael A. ; Moore, Brian C.J.: Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task. In: *The Journal of the Acoustical Society of America* 114 (2003), Nr. 2, 1023-1034. http://dx.doi.org/10.1121/1.1592160. – DOI 10.1121/1.1592160

[SM04] Stone, Michael A. ; Moore, Brian C. J.: Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. In: *The Journal of the Acoustical Society of America* 116 (2004), Nr. 4, 2311-2323. http://dx.doi.org/10.1121/1.1784447. – DOI 10.1121/1.1784447

[SM07] Stone, Michael A. ; Moore, Brian C.J.: Quantifying the effects of fast-acting compression on the envelope of speech. In: *The Journal of the Acoustical Society of America* 121 (2007), Nr. 3, 1654-1664. http://dx.doi.org/10.1121/1.2434754. – DOI 10.1121/1.2434754

[SM08] Stone, Michael A. ; Moore, Brian C.J.: Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task. In: *The Journal of the Acoustical Society of America* 123 (2008), Nr. 2, 1063-1076. http://dx.doi.org/10.1121/1.2821969. – DOI 10.1121/1.2821969

[SMAG99] Stone, Michael A. ; Moore, Brian C.J. ; Alcántara, José I. ; Glasberg, Brian R.: Comparison of different forms of compression using wearable digital

hearing aids. In: *The Journal of the Acoustical Society of America* 106 (1999), Nr. 6, 3603-3619. http://dx.doi.org/10.1121/1.428213. – DOI 10.1121/1.428213

[TTB+00] Thiede, Thilo ; Treurniet, William C. ; Bitto, Roland ; Schmidmer, Christian ; Sporer, Thomas ; Beerends, John G. ; Colomes, Catherine: PEAQ-The ITU Standard for Objective Measurement of Perceived Audio Quality. In: *JAES* 48 (2000), S. 3–29

[UIG+06] Unoki, Masashi ; Irino, Toshio ; Glasberg, Brian ; Moore, Brian C. J. ; Patterson, Roy D.: Comparison of the roex and gammachirp filters as representations of the auditory filter. In: *The Journal of the Acoustical Society of America* 120 (2006), Nr. 3, 1474-1492. http://dx.doi.org/10.1121/1.2228539. – DOI 10.1121/1.2228539

[Vai93] Vaidyanathan, P.P. ; Vaidyanathan, P.P. (Hrsg.): *Multirate Systems and Filter Banks*. Prentice Hall, 1993

[WSKH05] Wabnik, S. ; Schuller, G. ; Kramer, U. ; Hirschfeld, J.: Frequency warping in low delay audio coding. In: *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)* Bd. 3, 2005. – ISSN 1520–6149, S. iii/181–iii/184 Vol. 3

[WVB86] Widin, Gregory P. ; Viemeister, Neal F. ; Bacon, Sid P.: Effects of forward and simultaneous masking on intensity discrimination. In: *The Journal of the Acoustical Society of America* 80 (1986), Nr. 1, 108-111. http://dx.doi.org/10.1121/1.394170. – DOI 10.1121/1.394170

[ZAA+02] Zoelzer, Udo ; Amatriain, Xavier ; Arfib, Daniel ; Bonada, Jordi ; Poli, Giovanni De ; Dutilleux, Pierre ; Evangelista, Gianpaolo ; Keiler, Florian ; Loscos, Alex ; Rocchesso, Davide ; Sandler, Mark ; Serra, Xavier ;

Todoroff, Todor ; Zölzer, Udo (Hrsg.): *DAFX - Digital Audio Effects*. John Wiley & Sons, Ltd, 2002

[ZK02] Zakharenko, Alexander V. ; Kowalguin, Yurii: Improving Perceptual Coding of Wideband Audio Signal When Taking into Consideration of Temporal Masking. In: *21st International AES Conference: Architectural Acoustics and Sound Reinforcement*, 2002, S. 235–237

[Zöl05] Zölzer ; Zölzer (Hrsg.): *Digitale Audiosignalverarbeitung*. Springer, 2005