

---

# Funktionelle Mechanismen der auditorischen Kodierung von Träger- und Einhüllendenperiodizität beim Menschen

Alexandra Stein

---



München 2004



---

# Funktionelle Mechanismen der auditorischen Kodierung von Träger- und Einhüllendenperiodizität beim Menschen

Alexandra Stein

---

Dissertation  
an der Fakultät für Biologie  
der Ludwig–Maximilians–Universität  
München

vorgelegt von  
Alexandra Stein  
aus Wichita Falls, Texas, USA

München, den 14.04.2004

Erstgutachter: Prof. Dr. Gerhard Neuweiler

Zweitgutachter: Prof. Dr. Benedikt Grothe

Tag der mündlichen Prüfung: 22.06.2004

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>v</b>
<b>Abbildungsverzeichnis</b>	<b>ix</b>
<b>Abkürzungsverzeichnis</b>	<b>xi</b>
<b>Zusammenfassung</b>	<b>xiii</b>
<b>1 Periodizität im Träger oder in der Einhüllenden von Schallsignalen</b>	<b>1</b>
1.1 Stimuli mit Variationen in der Feinstruktur . . . . .	2
1.1.1 Iterated Rippled Noise . . . . .	2
1.2 Stimuli mit Variationen in der Einhüllenden . . . . .	5
1.2.1 Sinusförmige Amplitudenmodulation . . . . .	5
1.2.2 Amplitudenmodulation mit Rechteckmodulatoren . . . . .	8
1.2.3 Sparse Noise . . . . .	9
<b>2 Interaktion zwischen Tonhöhen- und Modulationswahrnehmung</b>	<b>11</b>
2.1 Einleitung . . . . .	12
2.2 Der Effekt von Trägerperiodizität auf die Wahrnehmungsschwellen von Amplitudenmodulation . . . . .	14
2.2.1 Paradigma . . . . .	14
2.2.2 Probanden . . . . .	15
2.2.3 Stimuli und Versuchsaapparatur . . . . .	15
2.2.4 Ergebnisse . . . . .	17
2.3 Der Effekt von Einhüllendenperiodizität auf die Wahrnehmung von Trägerperiodizität . . . . .	18
2.3.1 Paradigma . . . . .	18
2.3.2 Probanden . . . . .	18
2.3.3 Stimuli und Versuchsaapparatur . . . . .	18
2.3.4 Ergebnisse . . . . .	18
2.4 Diskussion . . . . .	19
2.5 Modell Simulationen der Experimente . . . . .	20
2.5.1 Das Modulations-Filterbank-Modell . . . . .	21

2.5.2	Das Autokorrelations-Modell . . . . .	29
2.6	Theoretische Zusammenhänge zwischen Modulationsfiltern und Autokorrelation . . . . .	36
2.7	Zusammenfassung und Ausblick . . . . .	38
<b>3</b>	<b>Modulationsmaskierung mit unterschiedlichen Maskierertypen</b>	<b>43</b>
3.1	Einleitung . . . . .	43
3.2	Methoden . . . . .	48
3.2.1	Paradigma . . . . .	48
3.2.2	Probanden . . . . .	48
3.2.3	Stimuli und Versuchsapparatur . . . . .	48
3.3	Ergebnisse . . . . .	49
3.4	Diskussion . . . . .	50
<b>4</b>	<b>Modulationsdetektion mit aperiodischen Modulatoren</b>	<b>53</b>
4.1	Einleitung . . . . .	53
4.2	Methoden . . . . .	54
4.2.1	Paradigma . . . . .	54
4.2.2	Stimuli und Versuchsapparatur . . . . .	55
4.2.3	Probanden . . . . .	56
4.3	Ergebnisse . . . . .	56
4.4	Diskussion . . . . .	57
<b>5</b>	<b>Unterscheidungsschwellen für spektral aufgelöstes und spektral nicht aufgelöstes IRN</b>	<b>63</b>
5.1	Einleitung . . . . .	63
5.2	Messung von Unterscheidungsschwellen für IRN-Stimuli . . . . .	65
5.2.1	Methoden . . . . .	65
5.2.2	Ergebnisse . . . . .	67
5.2.3	Diskussion . . . . .	69
5.3	Ermittlung des Abschwächungsfaktors . . . . .	70
5.3.1	Stimuli und Paradigma . . . . .	70
5.3.2	Ergebnisse und Diskussion . . . . .	71
5.4	Unterscheidungsschwellen von IRN-Stimuli mit verändertem gain . . . . .	73
5.4.1	Stimuli und Paradigma . . . . .	73
5.4.2	Ergebnisse . . . . .	75
5.4.3	Diskussion . . . . .	75
<b>6</b>	<b>Periodizitätsdetektion: Ein Mechanismus oder zwei?</b>	<b>79</b>
6.1	Ein physiologisch basiertes Tonhöhenmodell . . . . .	80
6.1.1	Die Ausgeprägtheit von Tonhöhe in simulierten Chop-S Neuronen . . . . .	80
6.1.2	Periodizitäten unterhalb der Tonhöhengrenze . . . . .	82
6.1.3	Ein Tonhöhenmechanismus . . . . .	83

6.2 Kodierung der Einhüllenden . . . . .	84
<b>Literaturverzeichnis</b>	<b>85</b>
<b>Lebenslauf</b>	<b>91</b>
<b>Publikationen</b>	<b>93</b>
<b>Danksagung</b>	<b>95</b>
<b>Ehrenwörtliche Erklärung</b>	<b>97</b>



# Abbildungsverzeichnis

1.1	Berechnung von IRN . . . . .	2
1.2	Autokorrelationsfunktionen von GN und IRN . . . . .	3
1.3	IRNs mit unterschiedlichen Verzögerungen . . . . .	4
1.4	Sinus-amplitudenmoduliertes Rauschen (SAMN) . . . . .	6
1.5	Modulationsseitenbänder . . . . .	7
1.6	Modulationsspektrum von GN . . . . .	7
1.7	Rechteckmodulator . . . . .	8
1.8	Magnitudenspektren von Rechteck- und Sinusmodulator . . . . .	8
1.9	Zeitsignal, Magnitudenspektrum und Einhüllendenspektrum von Sparse Noise . . . . .	10
2.1	Zeitliche und spektrale Eigenschaften von SAMN , IRN und amplitudenmoduliertem IRN . . . . .	16
2.2	Ergebnisse von Experiment I . . . . .	17
2.3	Ergebnisse von Experiment II . . . . .	19
2.4	Schematische Darstellung des Modulationsfilterbank-Modells . . . . .	22
2.5	Modulationsdetektion mit dem MFB Modell . . . . .	23
2.6	Einhüllendenspektren für SAMN und IRN AM . . . . .	25
2.7	SAMN und IRN im Modulations-Filterbank Modell . . . . .	26
2.8	Trägerperiodizitätsdetektion mit dem MFB Modell . . . . .	27
2.9	Modellvorhersagen nach ersetzen der MFB durch einen 8 Hz Tiefpass . . . . .	28
2.10	Schematische Darstellung des Autokorrelations-Modells . . . . .	30
2.11	Modulationsdetektion mit dem AC-Modell . . . . .	31
2.12	Grundlage zur Detektion von Amplitudenmodulation im AC-Modell . . . . .	32
2.13	Modulationsdetektion im AC-Modell mit einem optimalem Detektor als Entscheidungsmodul . . . . .	33
2.14	Trägerperiodizitätsdetektion mit dem AC-Modell . . . . .	34
2.15	Trägerperiodizitätsdetektion im AC-Modell mit einem optimalem Detektor als Entscheidungsmodul . . . . .	35
2.16	Vergleich von Modulationsfilter und Autokorrelator: Impulsantwort und Übertragungsfunktion . . . . .	37
2.17	Vergleich zwischen U-förmigem Modulationsfilter und Autokorrelation . . . . .	39
3.1	Vergleich der unterschiedlichen Maskierersituationen . . . . .	45

3.2	Zeitfunktionen von Rechteckmodulatoren mit unterschiedlichem Tastverhältnis . . . . .	46
3.3	Wellenformen von Maskieren unterschiedlicher Tastverhältnisse bei gleicher effektiver Modulationstiefe der Grundkomponente . . . . .	47
3.4	Modulationsmaskierung mit verändertem Tastverhältnis des Maskierers . .	50
3.5	Modulatoren mit unterschiedlichem Tasverhältnis aber gleichem Effektivwert	51
4.1	Ergebnis der Modulationsdetektion mit aperiodischen Modulatoren . . . .	56
4.2	Werte des 4. Moments . . . . .	59
4.3	Graphische Darstellung der Werte für das 4. Moment . . . . .	60
4.4	Modulationsspektrum von Schmalbandrauschen . . . . .	61
4.5	Vergleich der Modulationsspektren von verschiedenen SNs . . . . .	62
5.1	Autokorrelationsfunktion von IRN mit $F_0=90,9$ Hz bei einer Stimulusdauer von 25 ms . . . . .	66
5.2	$F_0$ DLs für IRNs mit Abschwächungsfaktor eins . . . . .	68
5.3	Umrechnung von Tonhöhen-Ausgeprägtheit (AU) zu IRN <i>gain</i> (dB) . . . .	71
5.4	Ergebnisse für die Ermittlung des $g$ -Wertes . . . . .	72
5.5	$F_0$ DLs für IRNs mit Abschwächungsfaktor $g<1$ . . . . .	74

# Abkürzungsverzeichnis

2AFC	<i>two-alternative forced-choice</i> (engl.), Auswahlverfahren; Methode zur Ermittlung von Schwellenwerten bei Wahrnehmungsversuchen
AC	Autokorrelation
AC1	erstes Maximum in der Autokorrelationsfunktion
ACF	Autokorrelationsfunktion
AM	Amplitudenmodulation
BF	Beste Frequenz, d.h. die optimale Anregungsfrequenz
CF	Charakteristische Frequenz
CN	<i>cochlear nucleus</i> (engl.), <i>Nucleus cochlearis</i>
CR	<i>chopping rate</i> (engl.), intrinsische Oszillationsfrequenz eines Chopper Neurons
dB	Dezibel
ED	Euklidische Distanz
ERB	<i>equivalent rectangular bandwidth</i> (engl.), Äquivalente rechteckige Bandbreite, psychoakustisches Maß für die Breite der auditorischen Filter
F0	Grundfrequenz eines harmonisch aufgebauten Signals
FFT	<i>Fast Fourier Transformation</i> (engl.), Algorithmus zur Berechnung eines Frequenzspektrums
GN	Gauss'sches Rauschen
IR	<i>impulse response</i> (engl.), Impulsantwort, entspricht der Übertragungsfunktion eines Filters im Zeitbereich

IRN	<i>iterated rippled noise</i> (engl.), iteriertes Kosinusrauschen
LLP	<i>lower limit of pitch</i> (engl.), die untere (Frequenz-)Grenze des Bereichs der Tonhöhenwahrnehmung
MFB	Modulations-Filterbank
RN	<i>rippled noise</i> (engl.), Kosinusrauschen
SACG	Summenautokorrelogramm
SAMN	<i>sinusoidally amplitude-modulated noise</i> (engl.), sinusförmig amplitudenmoduliertes Rauschen
SBR	Schmalbandrauschen
SN	Sparse Noise
SPF	<i>spike probability function</i> (engl.), Wahrscheinlichkeitsfunktion einer Nervenfasers, spikes zu generieren
SPL	<i>sound pressure level</i> (engl.), Schalldruckpegel
TMTF	<i>temporal modulation transfer function</i> (engl.)
VCN	<i>ventral cochlear nucleus</i> (engl.), ventraler Bereich des <i>Nucleus cochlearis</i>

# Zusammenfassung

Die Verarbeitung von Schall bietet dem Menschen eine wichtige Grundlage, seine Umwelt zu erfassen, sich darin zu orientieren und zu kommunizieren. Dabei wirkt permanent eine enorme Anzahl an akustischen Reizen auf das Ohr ein. Um diese Vielzahl an Eindrücken sinnvoll zu ordnen, kommt es zur Bildung auditorischer Objekte. Wichtige Parameter, die dazu beitragen, sind z.B. Tonhöhe und Amplitudenmodulation. Anders als die Erforschung neuronaler Verarbeitung sensorischer Eindrücke, versucht die Psychophysik, in diesem speziellen Fall die Psychoakustik, einen funktionalen Zusammenhang zwischen den physikalischen Parametern eines Schallsignals und dessen Wahrnehmung zu ergründen. Von zunehmender Bedeutung sind Modellvorstellungen der auditorischen Verarbeitung. Solche psychoakustischen Modelle können bei der Interpretation von Verhaltensmustern sehr hilfreich sein, sowie interessante Fragestellungen für weiterführende Versuche aufwerfen.

Diese Arbeit befasst sich mit der Verarbeitung von Periodizität in Schallsignalen beim Menschen. Jedes Schallsignal kann mathematisch in zwei Teile zerlegt werden: In die Einhüllende, die sich nur langsam über die Zeit verändert, und den Träger, der rasche Veränderungen über die Zeit aufweist. Periodizität kann sowohl im Träger als auch in der Einhüllenden auftreten. Amplitudenmodulation ist eine, in der Regel periodische, Veränderung der Einhüllenden. Trägerperiodizität bewirkt eine Tonhöhenwahrnehmung.

Um den funktionellen Mechanismen der auditorischen Kodierung von Träger- und Einhüllendenperiodizität nachzugehen, werden in dieser Arbeit vier unterschiedliche psychoakustische Experimente dargestellt. Das erste Experiment, beschrieben in Kapitel zwei, befasst sich mit der Interaktion zwischen Tonhöhen- und Modulationswahrnehmung. Es wurde untersucht, ob und in welchem Ausmaß die Wahrnehmung von Amplitudenmodulation durch die Gegenwart von Trägerperiodizität, die zur Tonhöhenwahrnehmung führt, beeinflusst wird und wie es sich im umgekehrten Falle verhält. In einem Auswahlverfahren (*two-alternative, forced choice, 2AFC*) wurden Detektionsschwellen entweder für Amplitudenmodulation oder für Trägerperiodizität ermittelt. Dabei stellte sich heraus, dass die Präsenz von Trägerperiodizität die Wahrnehmung für Amplitudenmodulation bei sämtlichen Testfrequenzen gleichermaßen beeinträchtigt. Dagegen werden die Detektionsschwellen für Trägerperiodizität nur bei den tiefen Testfrequenzen durch eine zusätzliche Amplitudenmodulation verringert. Die Tonhöhenwahrnehmung spielt dabei eine entscheidende Rolle.

Die Versuche wurden mit zwei unterschiedlichen psychoakustischen Modellen simuliert, die sich funktionell sehr ähnlich sind, da beide eine zeitliche Periodizitätsanalyse der Aktivität im Hörnerv vornehmen. Obwohl beide Modelle unterschiedliche physiologische

Mechanismen der Periodizitätswahrnehmung implementieren, waren letztendlich beide in der Lage, die empirischen Daten zumindest qualitativ korrekt zu rekonstruieren.

Das dritte Kapitel geht der Frage nach, auf welche Art und Weise das auditorische System die Eigenschaften der Einhüllenden extrahiert. Eine Arbeit von Ewert (2002) legt nahe, dass eine Strategie der spektralen Zerlegung der Einhüllenden vorgenommen wird. Viemeister et al. (2004) unterstreicht die Relevanz der lokalen zeitlichen Eigenschaften der Modulatorwellenform. Es werden Hörversuche dargestellt, die eine Brücke zwischen den beiden genannten Arbeiten schlagen. Die Ergebnisse sprechen für eine temporale Analyse der Modulatorwellenform. Ausschlaggebend für die Wahrnehmung ist nach diesen Ergebnissen ein Produkt aus lokaler Modulationstiefe und der Dauer, während die Modulationstiefe gleichbleibend ist.

Das vierte Kapitel nimmt sich der Wahrnehmung aperiodischer Modulatoren an. *Sparse Noise* (Hübner und Wiegrebe, 2003) kann durch Modulation mit einem aperiodischen Modulator erzeugt werden. Außerdem bietet der Stimulus die Möglichkeit, den Grad der Einhüllendenfluktuation selektiv zu variieren. Aufgrund der Einhüllendeneigenschaften von Sparse Noise (SN) könnte es ein hilfreicher Stimulus sein, um die Mechanismen der Einhüllendenkodierung besser verstehen zu lernen. Um die Wahrnehmung von SN zu charakterisieren, wurden Detektionsschwellen in Abhängigkeit von der Einhüllendenfluktuation gemessen. Der entscheidende Parameter für die Detektion von SN ist die physikalische Rauigkeit.

Schließlich befasst sich Kapitel fünf wieder mit einem Aspekt der Tonhöhenwahrnehmung. Eine der herrschenden Kontroversen der Tonhöhenforschung ist die Frage ob das auditorische System zwei unterschiedliche Mechanismen zur Tonhöhenextraktion anwendet oder nur einen einzigen. Vertreter der erstgenannten Möglichkeit unterscheiden danach, ob die spektralen Komponenten eines komplexen Signals von der Peripherie spektral aufgelöst werden oder nicht. Die Ausprägtheit der Tonhöhe ist für spektral nicht aufgelöste Signale deutlich geringer. Iteriertes Kosinusrauschen (*iterated rippled noise*, IRN) bietet durch dessen Algorithmus die Möglichkeit, die Ausprägtheit der Tonhöhe stufenlos zu variieren. Insofern ist IRN gut geeignet, um zu überprüfen, ob Unterschiede in den Wahrnehmungsschwellen von spektral aufgelösten/nicht aufgelösten Signalen durch die unterschiedliche Ausprägtheit der Tonhöhe bedingt sind. Das würde der Theorie der zwei Tonhöhenmechanismen widersprechen. Es wurden Unterscheidungsschwellen für IRN-Grundfrequenzen in Abhängigkeit von der Signaldauer gemessen. Die Stimuli waren bandpass-gefiltert, so dass sie entweder spektral aufgelöst werden konnten oder nicht. Diese wurde mit Unterscheidungsschwellen der gleichen Art verglichen, die mit spektral aufgelösten IRN-Stimuli erhoben wurden, deren Ausprägtheit der Tonhöhe einem spektral nicht aufgelösten IRN entsprach. Die Schwellen von IRNs gleicher Auflösbarkeit waren einander ähnlicher als die Schwellen von IRNs gleicher Tonhöhenausprägung. Insofern konnte die Eingangsüberlegung nicht bestätigt werden. Es wird erläutert, warum dieses Ergebnis dennoch nicht automatisch die Theorie eines einzigen Mechanismus widerlegt.

Zu Beginn werden grundlegende physikalische und wahrnehmungsbezogene Eigenschaften der angewandten Stimuli erläutert.

# Kapitel 1

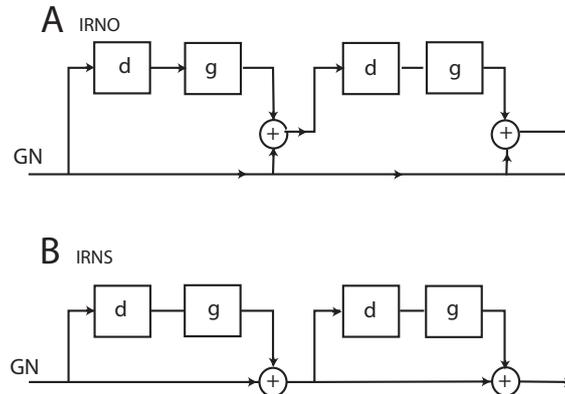
## Periodizität im Träger oder in der Einhüllenden von Schallsignalen

Jedes Schallsignal lässt sich mathematisch in unterschiedliche Bestandteile zerlegen. Das Fourier-Theorem z.B. besagt, dass jedes komplexe Schallsignal sich in eine Summe von einzelnen Sinustönen zerlegen lässt. Das bedeutet, dass sich durch eine Transformation nach Fourier der spektrale Gehalt eines Signals extrahieren lässt. Durch Frequenz-Orts-Transformation in der Cochlea, bei der unterschiedliche Frequenzen an einem jeweils diskreten Ort entlang der Hörschnecke abgebildet werden, findet letztendlich eine Transformation ähnlich dem Fourier-Theorem statt.

Nach einem von Hilbert aufgestellten Theorem lässt sich jedes Schallsignal mathematisch als das Produkt aus der sogenannten *Einhüllenden*, oder *Hüllkurve*, und der sogenannten *Feinstruktur* beschreiben. Die Einhüllende, auch Modulation genannt, verändert sich nur sehr langsam über die Zeit. Die Feinstruktur, auch Träger genannt, ist dagegen sehr raschen Veränderungen entlang der Zeitachse ausgesetzt. Sowohl die Einhüllende als auch die Feinstruktur eines Signals beinhalten wertvolle Informationen zur Bildung von auditorischen Objekten. Die Bildung von auditorischen Objekten wiederum ist essentiell, um die Flut von akustischen Reizen, die permanent auf den Menschen einwirken, in einer sinnvollen Art und Weise zu sortieren.

Smith et al. (2002) haben in eindrucksvollen Experimenten gezeigt, dass die Struktur der Einhüllenden besonders wichtig bei der Sprachwahrnehmung ist, wohingegen die wesentliche Information zur Wahrnehmung von Tonhöhe und zur Lokalisation von Schallen, in der Feinstruktur bzw. dem Träger enthalten ist. Auch bei der Weiterverarbeitung von Schallen im Gehirn, findet eine Unterscheidung von Einhüllender und Feinstruktur statt. In Bereichen des Hirnstamms werden Modulation und Träger teilweise von unterschiedlichen Neuronengruppen kodiert (z.B. Joris und Yin (1995); Langner (1992); Rhode et al. (1983); Yin und Chan (1990)).

Diese Arbeit befasst sich mit den funktionellen Mechanismen der Kodierung von Periodizität im Träger und in der Einhüllenden. Dazu wurden verschiedene psychoakustische Verhaltenstests durchgeführt. Die angewandten, Schallsignale, im Folgenden Stimuli genannt, leiten sich alle vom Gauss'schen Rauschen (GN) ab. Durch unterschiedliche Ver-



**Abbildung 1.1:** Schematische Darstellung des Prozesses aus verzögern ( $d$ ) - abschwächen ( $g$ )- addieren ( $\oplus$ ), zur digitalen Herstellung von iterated rippled noise. (A) beschreibt den Zyklus des *add-original* Prozesses bei dem jeweils das Originalrauschen zu der verzögerten Kopie zurückaddiert wird. (B) beschreibt den Zyklus des *add-same* Prozesses bei dem jeweils das Eingangssignal zu der verzögerten Kopie zurückaddiert wird

änderungen des Rauschens wird dem Stimulus Periodizität aufgeprägt. Je nach Art der Veränderung, lassen sich die verwendeten Stimuli im wesentlichen in zwei Gruppen einteilen: Stimuli mit Variationen im Träger einerseits und Stimuli mit Variationen in der Einhüllenden andererseits.

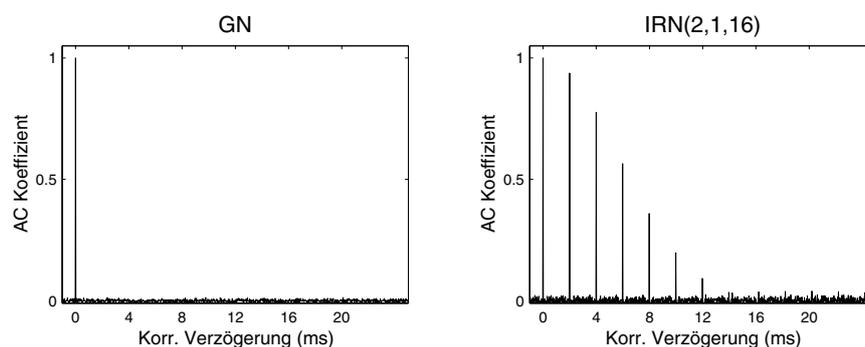
## 1.1 Stimuli mit Variationen in der Feinstruktur

### 1.1.1 Iterated Rippled Noise

*Iterated rippled noise* (IRN) ist ein tonales Rauschen, das von unkorreliertem Rauschen, wie z.B. GN, abgeleitet wird. Zuerst soll auf die Herstellung von IRN eingegangen werden, um im Anschluss die daraus resultierenden zeitlichen und spektralen Eigenschaften zu erläutern.

Durch verzögerte Selbstaddition von GN erhält man ein Kosinusrauschen (engl.: *rippled noise*, RN). Der genaue Vorgang dabei ist folgender: Ein zeitlich definiertes (jedoch beliebig langes) Stück Gauss'sches Rauschen (oder Weisses Rauschen) wird zunächst kopiert. Dann wird die Kopie um einen bestimmten Betrag zeitlich verzögert (*delay*). Diese verzögerte Kopie erfährt dann (wahlweise) eine Abschwächung, durch Multiplikation mit einem Abschwächungsfaktor. Das zeitlich verzögerte und abgeschwächte Rauschen wird abschließend zu dem nicht verzögerten Rauschen zurückaddiert. Wiederholt man diesen Vorgang mehrere Male, erhält man iteriertes Kosinusrauschen (engl.: *iterated rippled noise*, IRN).

Man unterscheidet zwei Arten von IRN, je nach Art ihrer Berechnung. Die entsprechenden Schaltkreise sind in Abb. 1.1 dargestellt. In Abb. 1.1A wird zu dem verzögerten Rauschen jedesmal, dasselbe ursprüngliche (originale) nichtverzögerte Rauschen addiert.



**Abbildung 1.2:** Autokorrelationsfunktionen von GN auf der rechten Seite und IRN(2,1,16) auf der linken Seite. Man sieht dass IRN Maxima bei  $d$  und dessen Vielfachen zeigt, GN dagegen völlig unkorreliert ist. Bei einer Verzögerung von 0 ist der Korrelationsgrad immer 1.

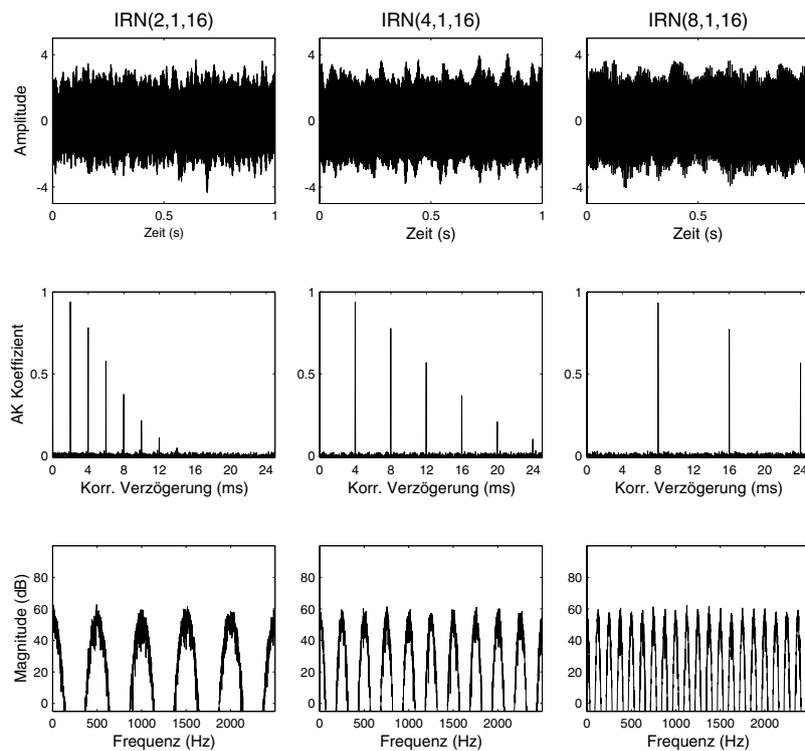
Daher wird dieses Verfahren als „add original“ bezeichnet, der resultierende Stimulus als IRNO. Das in Abb. 1.1B dargestellte Verfahren unterscheidet sich dadurch, dass das verzögerte und abgeschwächte Rauschen jeweils zu Eingang des vorangegangenen Verzögerungsprozesses addiert wird. Dieses Verfahren wird daher als „add same“ bezeichnet, der resultierende Stimulus als IRNS. In den Experimenten der vorliegenden Arbeit wird ausschließlich IRNS verwendet und wird im Folgenden nur als IRN bezeichnet. Dieser Prozess wird einige Male wiederholt. Die wichtigsten Parameter, durch die ein IRN sich definiert, sind die Verzögerung ( $d$ ), der Abschwächungsfaktor ( $g$ ) und die Anzahl der Iterationen ( $n$ ). Daraus leitet sich die Notation für IRN ab:  $\text{IRN}(d,g,n)$ .

In einer natürlichen Umgebung ist man oft mit vergleichbaren Schallereignissen konfrontiert. Grob gesehen, ist jede Reflexion eine verzögerte und abgeschwächte Version des Originalschalls. Das Rauschen eines Wasserfalls z.B. wird durch eine nahegelegene Felswände reflektiert. Somit erreicht sowohl das „originale Rauschen“ das Ohr, wie auch die Reflexion von der Wand, also eine verzögerte und abgeschwächte Kopie. Die Schaltkreise aus Abb. 1.1 simulieren quasi diese Situation. Die Reflexion verleiht dem Rauschen eine schwache Tonhöhe oder zumindest eine „spektrale Koloration“ (Bilsen und Ritsma, 1969). Die veränderte Wahrnehmung bei diesem Phänomen, beruht auf einer Veränderung der Feinstruktur.

Auch IRN erhält durch dessen veränderte Feinstruktur, im Verhältnis zu GN eine spektrale Koloration. Der Grund dafür liegt in den physikalischen Parametern, die diese Veränderung mit sich bringt. Durch den Prozess des Verzögerns, Abschwächens und Addierens, ist das Rauschen nicht mehr unkorreliert. Je öfter dieser Vorgang iteriert wird, desto höher wird der Korrelationsgrad (Yost, 1996b). Berechnet man die Autokorrelation von IRN sieht man, dass sich in der Autokorrelationsfunktion (ACF) Maxima bei der zeitlichen Verzögerung und deren Vielfachen ergeben (Abb.1.2). Aus der ACF kann man die Periodizität eines gegebenen Stimulus entnehmen. Die Orte der Maxima geben die Periodendauer an und die Höhe der Maxima, sprich der Korrelationskoeffizient, gibt die Ausprägtheit der Periodizität auf einer Skala von 0 bis 1 an. Der Korrelationsgrad, durch den der Grad an Periodizität ausgedrückt wird, wird bei IRN zum einen über die Anzahl der Iterationen

( $n$ ) bestimmt. Mit zunehmender Iterationszahl konvergiert der Korrelationsgrad des ersten Maximums in der ACF gegen eins. Zum anderen hat auch der Abschwächungsfaktor ( $g$ ) Einfluss auf die Ausprägtheit der Periodizität. Ist  $g=1$ , hat das IRN den höchsten, durch  $n$  bestimmten, Korrelationsgrad. Ist dagegen  $g=0$ , bleibt das in den Verzögerungsprozess eingehende Rauschen praktisch unverändert und der Korrelationsgrad bleibt gleich 0 (Yost, 1996a,b).

Periodizität in einem Schallsignal beeinflusst auch dessen Wahrnehmung. Sie verleiht dem Signal Tonalität. Beim Anhören eines IRN, nimmt man eine Tonhöhe wahr, welche jeweils dem Reziproken der Verzögerungszeit entspricht. Betrachtet man das Magnitudenspektrum eines IRN Signals wird deutlich warum.



**Abbildung 1.3:** Zeitsignale, Autokorrelationsfunktionen und Magnitudenspektren von IRNs mit unterschiedlichen Verzögerungszeiten. Die linke Spalte zeigt IRN(2,1,16), die mittlere Spalte IRN(4,1,16) und die rechte Spalte IRN(8,1,16). Die obere Reihe zeigt die Zeitsignale. Die Autokorrelationsfunktionen (ACFs), die in der mittleren Reihe dargestellt sind, zeigen Maxima bei der jeweiligen Verzögerung und deren Vielfachen. Die Magnitudenspektren weisen ein eindeutiges Kammfilterprofil auf. Die spektralen Maxima befinden sich bei dem Reziproken der Verzögerung und dessen Vielfache.

In Abb. 1.3 werden IRNs mit unterschiedlichen Verzögerungszeiten dargestellt. Die obere Reihe zeigt jeweils das Zeitsignal, die mittlere die ACFs und die untere Reihe das Magnitudenspektrum. Die ACFs weisen Maxima an der jeweiligen Verzögerungszeit und deren Vielfachen auf (vgl. auch Abb. 1.2). Die Magnitudenspektren zeigen ein deutliches Kammfilterprofil, das man englisch als „*rippled spectrum*“ bezeichnet. Daher leitet sich auch der Name ab: *iterated*, da der Prozess aus verzögern, abschwächen und addieren iteriert wird; *rippled*, wegen des Kammfilterprofils; *noise* von Rauschen. Die Maxima des Spektrums befinden sich jeweils bei  $1/d$  und dessen Vielfachen. Ein solches Kammfilterprofil ist vergleichbar mit einem Harmonischen Komplex. Ein Harmonischer Komplex ist zusammengesetzt aus Sinustönen einer sogenannten Grundfrequenz und deren Vielfachen. Die wahrgenommene Tonhöhe eines solchen Komplexes entspricht jeweils der Grundfrequenz. Ebenso verhält es sich bei IRN. Das erste Maximum im Kammfilterprofil von IRN befindet an der Stelle des Reziproken von  $d$ . Deshalb wird die wahrgenommene Tonhöhe durch das Reziproke der Verzögerung bestimmt.

Auch die Ausgeprägtheit der Periodizität, angegeben durch den Korrelationsgrad in der ACF (Ordinate, Abb. 1.3, mittlere Reihe), beeinflusst ebenfalls die Wahrnehmung. Die Höhe erste Maximums in der ACF (AC1) bestimmt die Ausgeprägtheit der Tonhöhe von IRN (Yost, 1996b). Je höher der Korrelationsgrad von AC1, desto ausgeprägter ist die Tonhöhe des IRN. Andersherum: Je geringer also der Korrelationsgrad ist, desto rauschhafter klingt das IRN.

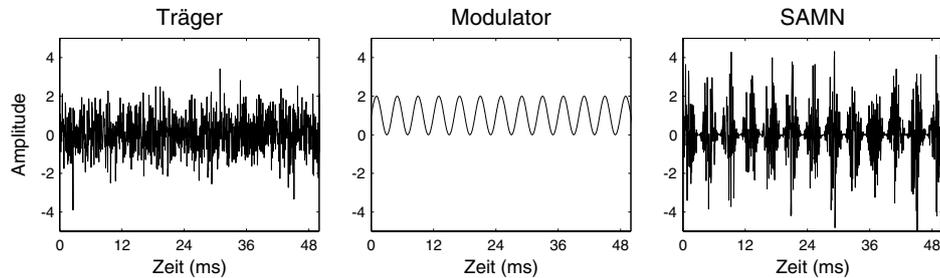
## 1.2 Stimuli mit Variationen in der Einhüllenden

Variationen in der Einhüllenden einzuführen ist eine relativ einfache Sache, die vielerlei unterschiedliche Möglichkeiten bietet. Man rufe sich noch einmal das Theorem von Hilbert ins Gedächtnis: jedes Schallsignal lässt sich durch das Produkt von Träger und Modulator (bzw. Feinstruktur und Einhüllende) definieren. Genau nach diesem Prinzip wird bei der Berechnung dieser Art von Stimuli vorgegangen. Man multipliziert einen Träger, z.B. Gauss'sches Rauschen, mit einem beliebigen Modulator, z.B. einem Sinuston. Da diese Multiplikation die Amplitude des Trägers verändert wird, bezeichnet man diesen Vorgang als Amplitudenmodulation (AM) bezeichnet. Im folgenden Abschnitt werden unterschiedliche Modulatoren, die für die vorliegende Arbeit relevant sind, vorgestellt.

### 1.2.1 Sinusförmige Amplitudenmodulation

Hier wird der Träger mit einer Sinusfunktion multipliziert. Ob der Träger ein Rauschen oder etwa selbst ein Sinuston ist, ist prinzipiell egal.

Diese Form der Amplitudenmodulation entspricht letztendlich der Einführung von Lautstärkeschwankungen des Trägers. Zur Veranschaulichung sind in Abb. 1.4 die Zeitfunktionen eines Rauschträgers, des Sinusmodulators und deren Produkt dargestellt. Das Produkt aus Sinusmodulator und Rauschträger nennt man sinusamplitudenmoduliertes Rauschen (SAMN). Betrachtet man die Amplitudenwerte des Modulators in Abb. 1.4 (mitt-



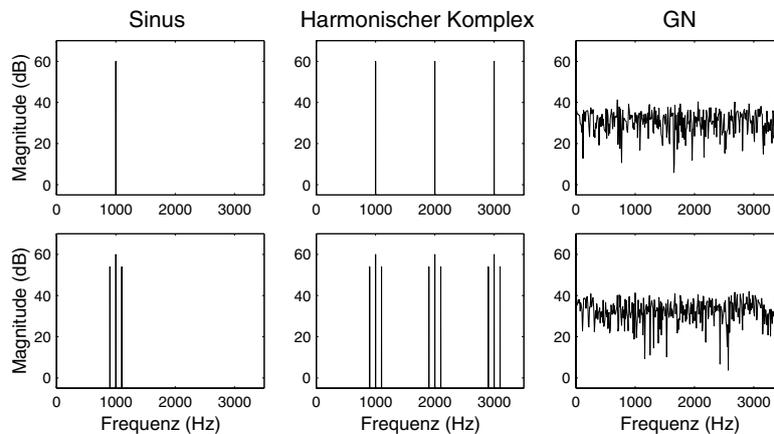
**Abbildung 1.4:** Zeitfunktionen eines Trägers (links), des Sinus-Modulators (mitte) und des zusammengesetzten Signals (rechts). Der Träger ist in diesem Beispiel GN und der Modulator hat eine Frequenz von 250 Hz. Der Modulator muss stets um +1 verschoben werden, damit er keine negativen Amplitudenwerte aufweist, bevor er mit dem Träger multipliziert wird. Andernfalls würde sich die Einhüllendenfrequenz verdoppeln (siehe Text).

leres Feld), sieht man, dass sie zwischen 0 und 2 verlaufen und nicht, wie für Sinustöne üblich, zwischen -1 und 1. Das hat folgenden Grund: Der Träger und der Modulator werden Punkt für Punkt miteinander multipliziert. Die Punkte der negativen Halbwelle des Modulators würden durch die Multiplikation mit negativen Amplitudenwerten des Trägers alle positiv. In der Konsequenz würde sich die Frequenz der Einhüllenden verdoppeln.

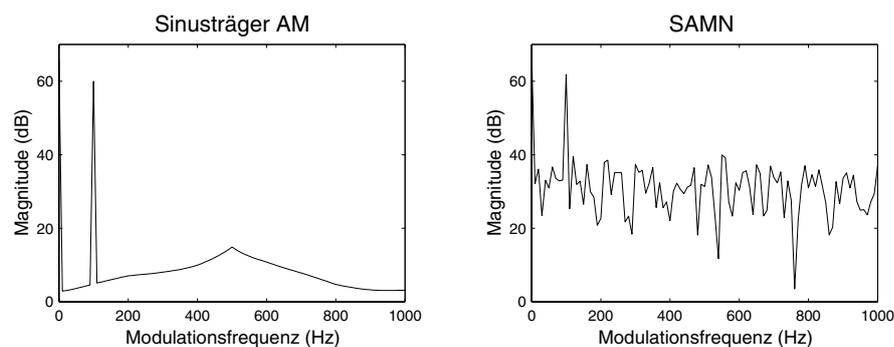
Neben der Lautstärkeschwankung über die Zeit hat eine AM auch Auswirkungen auf das Spektrum des Stimulus. Das lässt sich am einfachsten demonstrieren, wenn man einen Sinuston oder einen Harmonischen Komplex als Träger nimmt. Im Spektrum produziert eine Amplitudenmodulation sogenannte Seitenbänder. Diese liegen bei  $cf - mf$  und  $cf + mf$ . Dabei entspricht  $cf$  der Trägerfrequenz und  $mf$  der Modulationsfrequenz.

In Abb. 1.5 sind die Magnitudenspektren der Träger (obere Reihe) und der amplitudenmodulierten Signale (untere Reihe) abgebildet. Man sieht, dass die AM rechts und links, entlang der Frequenzachse, neben den spektralen Maxima des Trägers Seitenbänder produziert. Da das Spektrum von Gauss'schem Rauschen weiss ist, also alle Frequenzen enthält, fallen die Modulationsseitenbänder natürlich zusammen mit den bereits vorhandenen spektralen Maxima des Trägers und sind nicht gesondert erkennbar. Um die Modulationsfrequenz von SAMN zu ermitteln bzw. darzustellen, kann mittels einer Hilberttransformation die Einhüllende extrahiert werden, um daraus das Frequenzspektrum zu ermitteln.

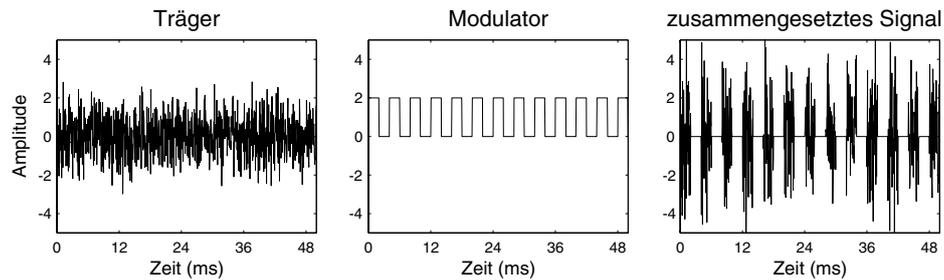
Auf diese Weise erhält man das *Modulationsspektrum*. Die Modulationsspektren des amplitudenmodulierten Sinusreintons und des SAMN sind in Abb. 1.6 zu sehen. In beiden Graphiken ist ein deutliches Maximum an Stelle der Modulatorfrequenz zu beobachten.



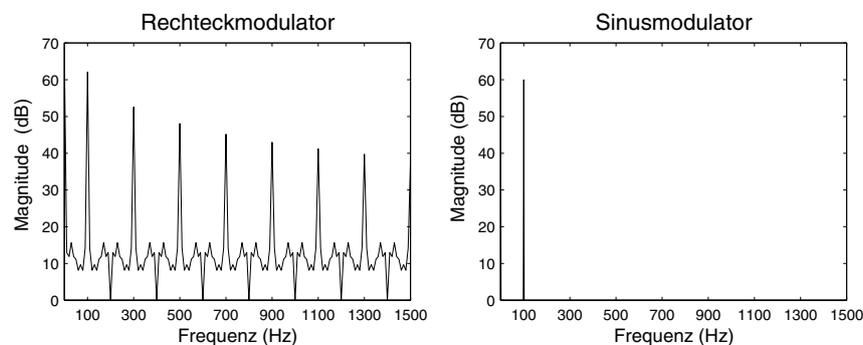
**Abbildung 1.5:** Hier werden unterschiedliche Träger mit einem Sinusmodulator multipliziert, um darzustellen, welche spektralen Veränderungen eine Sinusmodulation hervorruft. In der oberen Reihe sind die Magnitudenspektren der Träger abgebildet: links ein Sinus, in der Mitte ein harmonischer Komplex und rechts ein GN. In der unteren Reihe sind die Magnitudenspektren der AM-Signale dargestellt. Die Modulationsfrequenz ( $mf$ ) beträgt 100 Hz. Man sieht, dass die Modulation spektrale Seitenbänder verursacht. Der Abstand zum spektralen Maximum des Trägers beträgt immer  $\pm mf$ . Da ein GN bereits sämtliche Frequenzen enthält, sind auch im Magnitudenspektrum von SAMN keine prominenten Seitenbänder zu erkennen.



**Abbildung 1.6:** Modulationsspektren von einem SAMN (rechts) und einem amplitudenmodulierten Sinusträger (links). Man sieht in beiden Fällen, dass sich der Träger auch in der Hilbert Einhüllenden widerspiegelt. Im linken Feld sieht man eine leichte Erhöhung bei 500 Hz, der Frequenz des Sinusträgers. Im rechten Feld sieht man eine Art Grundrauschen im Modulationsspektrum, welches vom GN Träger hervorgerufen wird. In beiden Fällen hebt sich jedoch das, durch die AM verursachte Maximum sehr deutlich vom Hintergrund ab, der die Wahrnehmung nicht maßgeblich beeinflusst.



**Abbildung 1.7:** Zeitfunktionen des GN-Trägers (links), des Rechteckmodulators (Mitte) und des zusammengesetzten Signals (rechts). Wie der Sinusmodulator muss auch der Rechteckmodulator um  $+1$  verschoben werden, damit er keine negativen Werte enthält. Sonst wäre, durch Multiplikation mit dem Träger die Einhüllendenfrequenz doppelt so hoch wie die Frequenz des Modulators. Die Frequenz des Modulators beträgt hier 250 Hz, das entspricht einer Periode von 4 ms.



**Abbildung 1.8:** Links ist das Magnitudenspektrum eines Rechteckmodulators abgebildet und rechts das Magnitudenspektrum eines Sinusmodulators mit gleicher Frequenz (hier: 100 Hz). Zwar besitzt der Rechteckmodulator die gleiche temporale Periodizität der Sinusmodulator, unterscheidet sich jedoch sehr deutlich im Spektrum.

## 1.2.2 Amplitudenmodulation mit Rechteckmodulatoren

Anstelle einer Sinusfunktion kann u.A. auch ein Rechteckimpuls als Modulator dienen. Bei einer Rechteckfunktion sind alle Werte der positiven Halbwelle gleich eins und alle Werte der negativen Halbwelle gleich  $-1$ .

Abbildung 1.7 zeigt im linken Feld die Zeitfunktionen des Rauschträgers, im mittleren Feld die eines Rechteckmodulators und im rechten Feld die Zeitfunktion eines rechteckig-amplitudenmodulierten Rauschens. Wie im mittleren Feld zu sehen ist, muss natürlich auch der Rechteckmodulator derart verschoben werden, dass er selbst keine negativen Amplitudenwerte besitzt. Man sieht sofort dass die Lautstärkeschwankungen sehr scharfkantig sind. Darin liegt der wesentliche Unterschied zur Modulation mit einer Sinusfunktion, denn diese scharfkantigen Übergänge haben wiederum Auswirkungen auf das Frequenzspektrum der Einhüllenden.

Während ein Sinusmodulator, der ja einem Sinusreinton entspricht, nur ein Frequenzmaximum in im Magnitudenspektrum hat, produzieren die scharfkantigen Übergänge des

Rechteckimpulses Harmonische im Spektrum. Das Magnitudenspektrum eines Rechteckmodulators enthält Maximum an der Modulationsfrequenz plus den *ungeraden* Harmonischen. Auch wenn ein Sinus- und ein Rechteckmodulator die gleiche temporale Periodizität in der Einhüllenden erzeugen, haben beide einen deutlich unterschiedlichen spektralen Gehalt (Abb. 1.8).

### 1.2.3 Sparse Noise

Das Sparse Noise (SN) lässt sich am leichtesten beschreiben als ein Weisses Rauschen, dass über den Zeitverlauf immer wieder unterbrochen ist, von Lücken mit variabler Länge. Die maximale Lückendauer ist dabei definiert. SN kann auch durch Amplitudenmodulation erzeugt werden. Der Träger ist auch dabei Gauss'sches Rauschen. Um den Modulator digital zu generieren, wird zunächst eine Reihe von Nullen erzeugt, deren Länge sich aus der Stimulusdauer und der Abtastrate errechnet. Dann werden zufallsverteilt einige Nullen durch eins ersetzt. Der Abstand zwischen den Abtastpunkten deren Wert eins ist, ist ein beliebiger Wert zwischen 0 und der maximalen Lückendauer. Die maximale Lückendauer wird in der Anzahl an Abtastpunkten definiert. Der Index hierfür heisst *lsep* (length of separation) und hat einen ganzzahligen Wert zwischen 1 und 12. Die Formel für die maximale Lückendauer lautet:

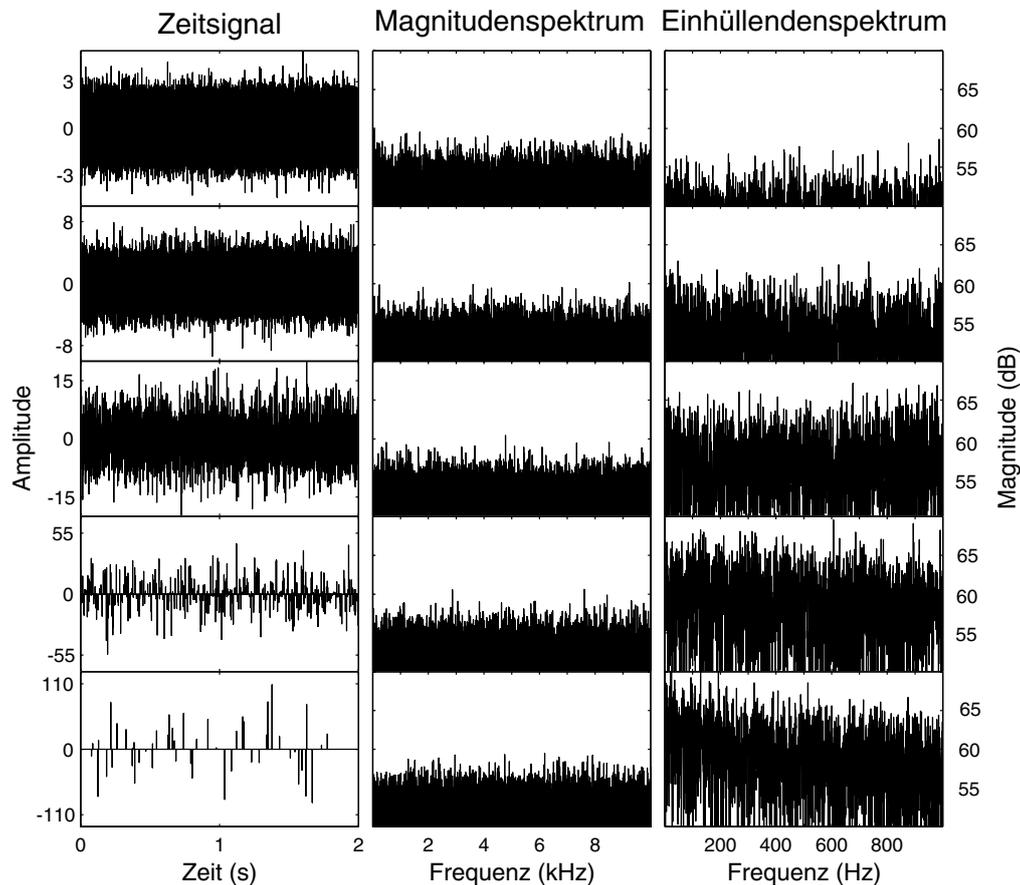
$$\text{max. Lückendauer} = 2^{(lsep)} - 1 \text{ Abtastpunkte} \quad (1.1)$$

Multipliziert man dann diesen Modulator mit dem GN erhält man Sparse Noise. Die maximale Lückendauer ist also durch die Variable *lsep* und der Abtastrate definiert. Da die einzelnen Lücken eine beliebige Dauer von 0 bis zur Maximaldauer haben können, entspricht die mittlere Lückendauer die halben maximalen Lückendauer.

Im Unterschied zu den vorherigen Modulatoren, ist der SN-Modulator aperiodisch. Das hat auch Auswirkungen auf das Spektrum. Während periodische Modulatoren, wie der Sinus- oder der Rechteckmodulator, ein harmonisches Frequenzspektrum aufweisen, ist das Einhüllendenspektrum eines Sparse Noise (SN) auch aperiodisch.

Abbildung 1.9 zeigt eine Reihe von SN Stimuli mit steigender max. Lückendauer von oben nach unten. Im Zeitsignal sieht man, dass die Anzahl der Amplituden abnimmt, dafür die Höhe der Amplituden deutlich ansteigt (man beachte die Beschriftung der Ordinate). Sowohl Magnituden- als auch Einhüllendenspektrum sind für alle Lückendauern weiss, sprich sie beinhalten alle Frequenzen. Im Einhüllendenspektrum allerdings, erhöht sich mit steigender max. Lückendauer die Energie der einzelnen Komponenten. Darüberhinaus wird mit steigender max. Lückendauer ein zunehmender Tiefpass im Modulationspektrum sichtbar.

Mit steigender maximaler Lückendauer wächst auch die Unruhe der Signals. Während sich ein SN mit niedriger max. Lückendauer wie homogenes Rauschen anhört, ähnelt ein SN mit steigender max. Lückendauer immer mehr einem Raschelgeräusch bis hin zu einem Knistern.



**Abbildung 1.9:** Zeitsignal (rechte Spalte), Magnitudenspektrum (mittlere Spalte) und Einhüllendenspektrum (linke Spalte) von Sparse Noises mit steigender maximaler Lückendauer (LD). Von oben nach unten: 0; 0,06; 0,62; 5,1 und 40,94 ms. Betrachtet man die Zeitsignale, sieht man dass mit steigender max. LD die Einzelamplituden weniger werden, dafür aber immer höhere Werte aufweisen (man beachte die Beschriftung der Ordinate). Da die Signale alle den gleichen Effektivwert besitzen, verteilt sich die Gesamtenergie auf weniger Einzelamplituden die dadurch natürlich höhere Werte aufweisen. Das Magnitudenspektrum dagegen verändert sich nicht. Weder in der Frequenzverteilung noch in der Lautstärke der einzelnen Komponenten. Die Frequenzverteilung im Einhüllendenspektrum ist ebenfalls gleichbleibend. Allerdings werden nimmt die Energie der einzelnen Komponenten deutlich zu mit steigender max. LD und es bildet sich ein zunehmendes Tiefpassverhalten heraus.

## Kapitel 2

# Interaktion zwischen Tonhöhen- und Modulationswahrnehmung

Der Inhalt dieses Kapitels wurde beim *Journal of the Acoustic Society of America* zur Veröffentlichung eingereicht und befindet sich derzeit im Review-Prozess.

Im vorherigen Kapitel wurde beschrieben, dass sich jedes Schallsignal, durch eine Hilbert-Transformation in die Einhüllende, auch Modulation genannt, und den Träger, auch Feinstruktur genannt, unterteilen lässt. Es wurde auch dargestellt, durch welche Veränderungen Periodizität in der Einhüllenden bzw. im Träger hervorgerufen wird. Dieses Kapitel befasst sich mit den perzeptionellen Interaktionen zwischen Trägerperiodizität und Einhüllendenperiodizität. In Abschnitt 1.1.1 wurde erläutert, dass Periodizität im Träger eine Tonhöhenwahrnehmung hervorruft. Aufgrund dieser eindeutigen Tonhöhe, die ein IRN hervorruft, fallen Versuche zur Wahrnehmung von IRN oft unter den Begriff der Tonhöhenwahrnehmung, anstatt dass man von Trägerperiodizitätswahrnehmung spricht. Die Wahrnehmung von Amplitudenmodulation (kurz: Modulationswahrnehmung) ist ein weites und viel untersuchtes Feld in der Psychoakustik. Amplitudenmodulation (AM) bedeutet immer eine Veränderung der Einhüllenden. Eine AM bringt in der Regel eine Periodizität in der Einhüllenden mit sich.

Tonhöhen- und Modulationswahrnehmung können also beide auf der Wahrnehmung von Periodizitäten beruhen. Aufgrund dieser Annahme und aufgrund von Funktionsweisen führender psychoakustischer Modelle zur Tonhöhen- und zur Modulationswahrnehmung, stellt sich die Frage ob eventuell ein neuronaler Mechanismus existiert, der Periodizität an sich erfasst. Ist dies der Fall, werden folgende Fragen aufgeworfen: Sind es zwei unterschiedliche Mechanismen? Einer zur Kodierung von Trägerperiodizität und einer zur Kodierung von Einhüllendenperiodizität? Oder gibt es einen *gemeinsamen* Mechanismus? Um sich dieser Frage anzunähern wurden zunächst psychoakustische Experimente durchgeführt, die untersuchen, ob es Interaktion in der Wahrnehmung zwischen beiden Arten der Periodizität gibt. Sollte das simultane Vorkommen von Träger- und Einhüllendenperiodizität keinen Einfluss auf die Wahrnehmung haben, im Vergleich zur Situation, in der nur eine der beiden Arten von Periodizität gegeben ist, wäre das ein Hinweis darauf, dass es getrennte

Verarbeitungswege gibt. Im Umkehrschluss kann eine Interaktion in der Wahrnehmung als Hinweis auf einen gemeinsamen Verarbeitungsmechanismus gedeutet werden.

Psychophysische Modelle sind ein etabliertes Mittel sich der Funktionsweise der sensorischen Verarbeitung anzunähern. Sie bieten den Vorteil einer schematischen Betrachtung statt einer zu sehr detaillierten Untersuchung z.B. der neuronalen Grundlagen. Obwohl diese ab einem gewissen Zeitpunkt essentiell ist, sind jedoch die Modellansätze hilfreich, um sich einmal ein grobes Bild von funktionalen Mechanismen zu verschaffen. Insofern werden auf der Suche nach Zusammenhängen zwischen Wahrnehmung von Einhüllenden- und Trägerperiodizität auch die bestehenden Modellansätze betrachtet.

## 2.1 Einleitung

Im Laufe der letzten Jahrzehnte entwickelten sich die meisten psychoakustischen Modelle dahingehend, die periphere Vorverarbeitung des Außen-, Mittel- und Innenohrs relativ detailliert zu berücksichtigen. Da eine relativ gute Vorstellung dieser peripheren Vorverarbeitung existiert sind auch die bestehenden Modelle untereinander sehr ähnlich aufgebaut.

In Bezug auf die Tonhöhen- und Modulationsverarbeitung gibt es unterschiedliche Modelle, die sich jeweils ausschließlich mit der Tonhöhen- *oder* Modulationsverarbeitung befassen.

Ein sehr erfolgreiches Modell in der Vorhersage von Tonhöhenwahrnehmungsphänomenen wurde von Meddis und Hewitt (1991a,b) eingeführt. Dieses Modell basiert auf der von Licklider (1951) vorgeschlagenen Annahme, das auditorische System wende ein Verfahren ähnlich einer Autokorrelation an, um die Periodizität eines Stimulus zu extrahieren. Meddis und Hewitt erweiterten diese Annahme indem sie das Ergebnis der Autokorrelation in ihrem Modell entlang einer tonotopen Frequenzachse aufsummierten. Folglich steht am Ausgang eine rein zeitliche Repräsentation der Periodizität eines gegebenen Stimulus. Dieses Modell ist in der Lage, unterschiedlichste Phänomene der Tonhöhenwahrnehmung zu reproduzieren bzw. vorherzusagen. Dazu zählt u.a. virtuelle Tonhöhe oder die Dominanz des mittleren Frequenzbereichs. Das Modell wurde in den folgenden Jahren weiterentwickelt (Meddis und O'Mard, 1997) und ist nun auch in der Lage weitere Wahrnehmungsphänomene wie die Tonhöhenempfindung von harmonischen Komplexen mit alternierenden Phasen der einzelnen Harmonischen, sowie die Sensitivität für Frequenzabweichungen einzelner Harmonischen nachzubilden.

Ein weiteres Modell, das hier genauer betrachtet werden soll, wurde von Dau et al. (1997a,b) entwickelt, um die Wahrnehmung von Amplitudenmodulation zu simulieren. Dieses Modell basiert auf der Theorie von Houtgast (1989), nach der die Einhüllende eines Stimulus ebenfalls einer spektralen Zerlegung unterzogen wird, ähnlich der Dekomposition im Innenohr durch die cochleären Filter. Houtgast konnte eine Frequenzselektivität bezüglich der Einhüllenden in psychoakustischen Experimenten demonstrieren (Houtgast, 1989). Als funktionale Implementation der Frequenzselektivität für die Einhüllende haben Dau et al. (1997a,b) eine Modulationsfilterbank entwickelt. Gleich dem oben beschriebenen Modell zur Tonhöhenwahrnehmung, beinhaltet dieses Modell ebenfalls eine detaillierte Implemen-

tierung der peripheren Verarbeitung, die der Modulationsfilterbank vorausgeht. Zahlreiche Phänomene der Amplitudenmodulationswahrnehmung werden mit diesem Modell erfolgreich simuliert. Dazu gehören u.a. die Frequenzselektivität für die Einhüllende sowie die Modulationsdetektion für schmalbandiges Rauschen.

Obwohl die beiden genannten Modelle unterschiedliche Ziele verfolgen, ist das zugrunde liegende Konzept recht ähnlich. Beide Modelle beinhalten eine Gammaton-Filterbank sowie eine funktionale Simulation der mechano-neuralen Transduktion der Inneren Haarzellen in der Cochlea. In beiden Modellen folgt dieser Stufe ein Modul, das die temporale Periodizität der Feuerwahrscheinlichkeit der Hörnervfaser extrahiert. In einem Fall in Form einer Modulationsfilterbank implementiert, im anderen Fall in Form eines Autokorrelators. Zwar liegen die wahrnehmbaren Periodizitäten von Tonhöhe und Amplitudenmodulation durchaus in unterschiedlichen Regionen (Krumbholz et al., 2000; Viemeister, 1979), dennoch gibt es einen großen Überlappungsbereich zwischen 30 und 1000 Hz.

Sollten sowohl Trägerperiodizität als auch Amplitudenmodulation (also Einhüllendenperiodizität) zeitlich kodiert werden, ist es nicht abwegig einen gemeinsamen Verarbeitungsmechanismus anzunehmen. Denn es gibt in der Tat Hinweise darauf, dass die Wahrnehmung von Trägerperiodizität und Amplitudenmodulation breitbandiger Rauschträger nicht unabhängig voneinander sind. Burns und Viemeister (1976) sowie auch Burns (1981) zeigten, dass sinus-amplitudenmoduliertes Rauschen (SAMN) eine tonhafte Wahrnehmung hervorruft und mit SAMN Signalen auch eine Melodie erzeugt werden kann. Die im Folgenden beschriebenen Experimente sollen die Interaktion von Träger- und Einhüllendenperiodizität von breitbandigen Stimuli in der Wahrnehmung untersuchen.

Ein typischer Stimulus mit Einhüllendenperiodizität ist SAMN. Der Träger von SAMN, hier Gauss'sches Rauschen (GN), ist aperiodisch. Die Periodizität dieses Stimulus liegt somit ausschließlich in der Einhüllenden. Die Periodendauer wird durch das Reziproke der Modulationsfrequenz bestimmt.

Im Gegensatz dazu ist das sogenannte *iterated rippled noise* (IRN) ein Stimulus mit ausgeprägter Trägerperiodizität. Diese Trägerperiodizität spiegelt sich zu einem gewissen Grad auch in der Einhüllenden wider. Eine Reihe psychoakustischer Experimente haben jedoch gezeigt, dass die Wahrnehmung von IRN von der Periodizität des Trägers und nicht von der Periodizität der Einhüllenden dominiert wird (Wiegrebe und Patterson, 1999). IRN wird ebenfalls von GN abgeleitet. IRN wird erzeugt indem man ein Stück Rauschen verzögert, das verzögerte Stück mit dem Abschwächungsfaktor (*gain*) multipliziert und wieder zu sich selbst zurück addiert. Dieser Prozess wird mehrere Male iteriert. Das erzeugt eine Periodizität im Stimulus. Das Magnitudenspektrum stellt sich als Kammspektrum dar, mit Maxima an Stelle des Reziproken der Verzögerungsdauer sowie ganzzahlige Vielfache davon (siehe Abb. 1.3). Insofern ist die Grundfrequenz ( $F_0$ ) eines IRN Stimulus das Reziproke der Verzögerung. Dieser Wert entspricht auch der wahrgenommenen Tonhöhe eines IRN Stimulus. Im Folgenden wird in der Regel auf die  $F_0$  Bezug genommen.

Somit liegen zweierlei periodische, breitbandige Stimuli vor. Dabei wird die Wahrnehmung von SAMN von der Einhüllenden dominiert, die Wahrnehmung von IRN dagegen wird vom Träger dominiert. Darüberhinaus lässt sich der Grad der Periodizität beider Stimuli stufenlos verändern. Im Falle von SAMN geschieht dies über Veränderung der Mo-

dulationstiefe. Im Falle von IRN hingegen wird der Abschwächungsfaktor, englisch *gain* ( $g$ ), in der Verzögerungsschleife variiert: ist  $g$  gleich eins, erhält der Stimulus den maximalen Grad an Periodizität, je nach Anzahl der Iterationen (vgl. 1.1.1). Bei einem  $g$  von Null bleibt das GN am Ausgang der Verzögerungsschleife unverändert.

Im Folgenden werden nun zwei psychoakustische Experimente präsentiert. Das erste Experiment untersucht die Fragestellung, in wiefern die Wahrnehmung von Einhüllendenperiodizität durch die gleichzeitige Anwesenheit von Trägerperiodizität beeinflusst wird. In anderen Worten, wie verändern sich die Schwellen zur Detektion einer Amplitudenmodulation wenn der aperiodische Träger durch einen Träger mit ausgeprägter Periodizität ersetzt wird? Das zweite Experiment untersucht die Fragestellung, in wiefern die Wahrnehmung von Trägerperiodizität durch die gleichzeitige Anwesenheit von Einhüllendenperiodizität beeinflusst wird. Das bedeutet, wie verändern sich die Schwellen zur Detektion von Trägerperiodizität, wenn die Stimuli gleichzeitig amplitudenmoduliert sind.

## 2.2 Experiment I: Der Effekt von Trägerperiodizität auf die Wahrnehmungsschwellen von Amplitudenmodulation

Um festzustellen ob, und in welchem Ausmaß die Präsenz von Trägerperiodizität die Wahrnehmung von Amplitudenmodulation beeinflusst, wurden Hörversuche entwickelt, in denen die Probanden die Aufgabe hatten eine sinusförmige Amplitudenmodulation zu detektieren. Der Träger war entweder GN, also aperiodisch, oder IRN, also quasi-periodisch. Wurde ein IRN Träger verwendet, wurde die Verzögerung derart gewählt, dass sich eine IRN  $F_0$  ergab, die gleich der Modulationsfrequenz war, aus folgendem Grund: IRN zeigt im Magnitudenspektrum harmonisch angeordnete Maxima. Ist nun die Modulationsfrequenz gleich der IRN  $F_0$ , fallen die spektralen Seitenbänder, die von der Amplitudenmodulation produziert werden (siehe Abschnitt 1.2.1), genau auf die Maxima des IRN Spektrums. Dies soll gewährleisten, dass im Stimulus keine zusätzlichen spektralen Inhalte entstehen, die dem Probanden als Entscheidungskriterium zur Verfügung stehen. Die Modulationsdetektion mit einem GN-Träger ist eine Wiederholung der klassischen *temporal modulation transfer function* (TMTF) die von Viemeister (1979) veröffentlicht wurde.

### 2.2.1 Paradigma

Die Detektionsschwellen für die Amplitudenmodulation wurden mittels eines sogenannten adaptiven *two-alternative forced-choice* (2AFC) Paradigmas untersucht. Dem Probanden werden 2 Intervalle vorgespielt, von denen eines das Signal beinhaltet. Entsprechend wird dieses als Signalintervall bezeichnet, das andere als Standardintervall. Nachdem zwei Intervalle präsentiert wurden, muss der Proband eine Entscheidung treffen, welches Intervall das Signalintervall war. Nach jeder Entscheidung hat der Proband ein Feedback erhalten, ob die Antwort richtig war. Der adaptive Parameter in diesem Experiment war die

Modulationstiefe. Dieser Wert wurde nach einer *one-up three-down* Regel verändert. Das bedeutet, dass, nachdem drei Mal in Folge das Signalintervall korrekt identifiziert wurde, die Modulationstiefe schrittweise um einen bestimmten Betrag verringert wurde. Nach einem Fehler wurde sie wieder um den selben Betrag erhöht. Eine Veränderung des adaptiven Parameters, nach drei aufeinanderfolgenden korrekten Antworten bzw. nach einem Fehler, wird als Umkehrpunkt bezeichnet. Zu Anfang jeder Schwellenmessung wurde die Modulationstiefe jeweils um 5 dB verändert. Nach dem dritten Umkehrpunkt verringerte sich die Schrittweite auf 2 dB. Danach wurde die Modulationstiefe nur noch um 1 dB verändert. Insgesamt wurden pro Schwellenmessung 11 Umkehrpunkte aufgenommen. Der Mittelwert der letzten 6 Umkehrpunkte ergab die Schwelle. Diese Vorgehensweise ermittelt den Wert der 79,4% Marke einer Psychometrischen Funktion (Levitt, 1971). Das bedeutet, dass bei diesem Wert des adaptiven Parameters, das Signalintervall zu 79,4% korrekt identifiziert werden kann. Für jede Signalkonfiguration wurde der Mittelwert von drei Messungen als Schwelle für die jeweilige Versuchsperson gewertet. Von allen zu testenden Modulationsfrequenzen wurde für jede Schwellenmessung jeweils eine zufällig ausgewählt. Den Probanden wurde eine Trainingsphase gewährt, um stabile Werte zu erreichen (d.h. der Standardfehler zwischen drei aufeinanderfolgenden Durchgängen pro Signalkonfiguration ist nicht größer als 2 dB).

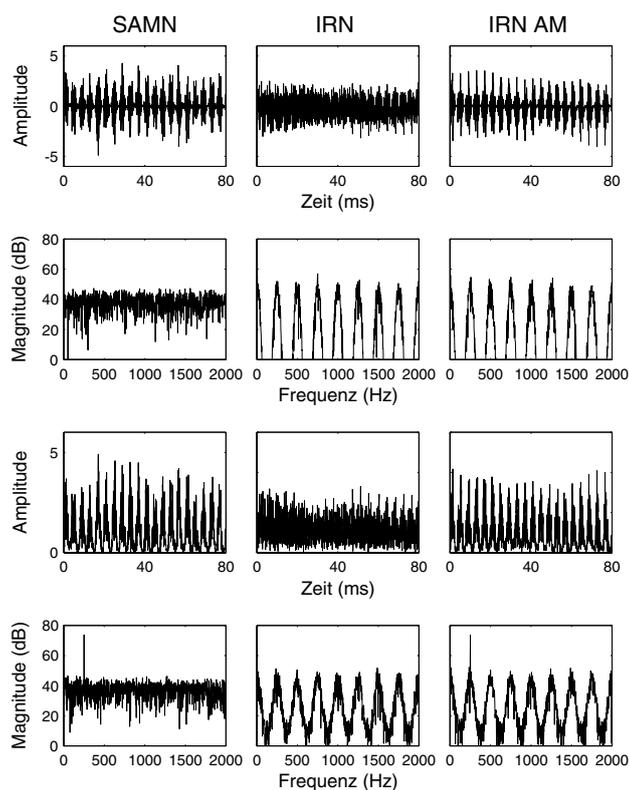
### 2.2.2 Probanden

Es beteiligten sich vier Versuchspersonen im Alter von 25 bis 32 Jahren an diesem Experiment. Die Teilnahme erfolgte freiwillig und unentgeltlich. Alle Probanden verfügten über ein klinisch normales Hörvermögen.

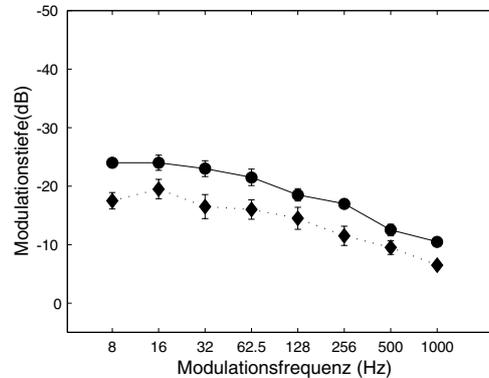
### 2.2.3 Stimuli und Versuchsausrüstung

Der Träger war entweder aperiodisches GN oder quasi-periodisches IRN und wurde mit einem sinusförmigen Modulator mit variabler Modulationstiefe multipliziert. Die Modulationsfrequenzen lagen im Bereich von 7,8125 bis 1000 Hz, in Abständen von einer Oktave. Der IRN Träger wurde in der *add-same* (siehe Abschnitt 1.1.1) Konfiguration berechnet (Yost, 1996a). Der Abschwächungsfaktor  $g$  des IRN betrug eins und die Anzahl der Iterationen war 16. Die  $F_0$  der IRN-Stimuli entsprach stets der jeweiligen Modulationsfrequenz. Abbildung 2.1 stellt einen Überblick über die Zeitfunktionen und Magnitudenspektren der hier verwendeten Stimuli dar.

Die Stimulusdauer betrug 500 ms, inklusive kosinusförmiger An- und Abstiegsflanken von je 20 ms Dauer. Die Stimuli wurden auf einem IBM kompatiblen PC mit einer Abtastrate von 44,1 kHz generiert und anschließend auf einen Tucker-Davis-Technologies AP2 Array Prozessor übergeben. Ein Tucker-Davis-Technologies DA 3-2 D/A-Wandler, wandelte das digitale Signal in ein analoges Signal um. Alle Stimuli wurden mit einem Yamaha M35 2/4 Channel Power Amplifier verstärkt und passierten im Anschluss einen passiven Endabschwächer, der auf 40 dB eingestellt war. Der mittlere Ausgabepegel lag bei 60 dB SPL. Der Ausgabepegel wurde bei jeder Präsentation im Bereich von  $\pm 6$  dB SPL randomisiert,



**Abbildung 2.1:** Zeitliche und spektrale Eigenschaften von SAMN (linke Spalte), IRN (mittlere Spalte) und amplitudenmoduliertem IRN (rechte Spalte). In der obersten Reihe sind die Zeitfunktionen der Signale abgebildet, in der zweiten Reihe die Magnitudenspektren. Die dritte Reihe zeigt die Hilbert-Einhüllende, die vierte die Magnitudenspektren der Einhüllenden, auch Modulationspektren genannt. Die IRN Verzögerung beträgt  $1/\text{Modulationsfrequenz}$ . Dadurch fallen die spektralen Seitenbänder der AM auf die spektralen Maxima im Kammfilterprofil des IRN (vierte Reihe, rechts). Ein Vergleich der Magnitudenspektren von IRN und IRN AM zeigt, dass durch AM keine merkliche Veränderung eintritt.



**Abbildung 2.2:** Ergebnisse der Modulationsdetektion. Die Kreise und die durchgezogene Linie repräsentieren die Daten, die mit einem GN-Träger erzielt wurden. Die Rauten und die gepunktete Linie stellen die Ergebnisse mit einem IRN-Träger dar. Die Detektionsleistung nimmt mit steigender Modulationsfrequenz ab. Der IRN-Träger verschlechtert die Detektionsleistung gleichermaßen (um ca. 5 dB) über den gesamten gemessenen Bereich.

um etwaige systematische Pegelunterschiede als Entscheidungskriterium auszuschließen. Die Ausgabe der Stimuli erfolgte diotisch über AKG K240DF Kopfhörer. Die Probanden wurden einzeln in einer reflexionsarmen und schallisolierten Hörkabine getestet. Antworten wurden über ein graphisches Interface auf einem Touchscreen abgegeben.

### 2.2.4 Ergebnisse

In Abbildung 2.2 ist die Modulationstiefe an der Detektionsschwelle als Funktion der Modulationsfrequenz dargestellt. Die Grafik stellt die durchschnittliche Leistung der Versuchspersonen dar. Die Fehlerbalken zeigen den Standardfehler zwischen den Probanden. Die Kreissymbole repräsentieren die Schwellen für den GN Träger, also die Detektionsschwellen für Amplitudenmodulation in *Abwesenheit* einer Trägerperiodizität. Die Rauten repräsentieren die Schwellen für den IRN Träger, also die Detektionsschwellen für Amplitudenmodulation in *Anwesenheit* einer Trägerperiodizität.

Ist der Träger GN, verschlechtert sich mit zunehmender Modulationsfrequenz die Fähigkeit der Probanden, die Amplitudenmodulation zu detektieren. Das geht einher mit Ergebnissen von Eddins (1993, 1999); Ewert und Dau (2000); Viemeister (1979). Ist der Träger IRN, bleibt der Verlauf der Kurve zwar gleich, die Werte sind allerdings insgesamt um durchschnittlich 5 dB schlechter. Dieses Ergebnis deutet darauf hin, dass die Sensitivität für Amplitudenmodulation durch die gleichzeitige Anwesenheit von Trägerperiodizität beeinträchtigt wird.

## 2.3 Experiment II: Der Effekt von Einhüllendenperiodizität auf die Wahrnehmung von Trägerperiodizität

Dieses Experiment ist im Prinzip das Gegenexperiment zu Experiment I. Hier hatten die Probanden die Aufgabe, Trägerperiodizität zu detektieren. Da man den Grad der Periodizität bei IRN mittels des Abschwächungsfaktors stufenlos variieren kann, bietet dieser Stimulus die Möglichkeit, eine Schwelle für die Wahrnehmung von Trägerperiodizität zu quantifizieren. Um den Effekt von Einhüllendenperiodizität hierbei zu testen, wurden die Schwellen einmal ohne und einmal mit zusätzlicher Amplitudenmodulation gemessen.

### 2.3.1 Paradigma

Die folgenden Messungen wurden ebenfalls nach dem oben beschriebenen 2AFC Verfahren ermittelt. Der adaptive Parameter war der Abschwächungsfaktor (gain,  $g$ ) des IRN. Die Schrittweiten mit denen dieser Parameter verändert wurde, betragen 10 dB nach den ersten beiden Umkehrpunkten, 5 dB für die Umkehrpunkte drei bis fünf und 2 dB nach den letzten sechs Umkehrpunkten. Für jede Schwellenmessung wurde eine der zu testenden IRNF0s zufällig ausgewählt. Die Schwelle der einzelnen Versuchspersonen war der gemittelte Wert aus drei Wiederholungen.

### 2.3.2 Probanden

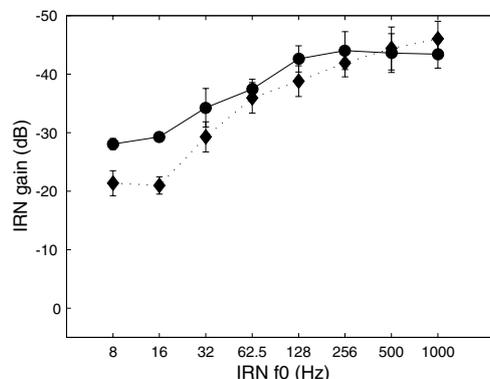
Vier Versuchspersonen im Alter zwischen 28 und 35 Jahren nahmen an diesem Experiment teil. Die Teilnahme erfolgte freiwillig und unentgeltlich. Alle Versuchspersonen verfügten über ein klinisch normales Hörvermögen.

### 2.3.3 Stimuli und Versuchsausrüstung

Der Signalstimulus war IRN und der Standardstimulus war GN. Die IRNF0s lagen zwischen 7,8125 und 1000 Hz (das entspricht IRN Verzögerungen von 128 bis 1 ms), in Abständen von einer Oktave. Die Anzahl der Iterationen betrug stets 16. Der adaptive Parameter in diesem Experiment war  $g$  und wurde somit im Laufe der Schwellenmessung variiert. Wurden Signal- und Standardstimulus zusätzlich amplitudenmoduliert, entsprach die Modulationsfrequenz der IRNF0, aus den gleichen Gründen, wie sie im vorangegangenen Experiment beschrieben wurden. Die Modulationstiefe lag bei 0 dB (100%). Das technische Equipment sowie der Ausgabepegel der Stimuli ist identisch zu Experiment I.

### 2.3.4 Ergebnisse

Abbildung 2.3 zeigt den  $g$ -Wert an der Detektionsschwelle als Funktion der IRNF0. Die Werte stellen die durchschnittliche Leistung der Versuchspersonen dar und die Fehlerbalken entsprechen dem Standardfehler zwischen den einzelnen Probanden. Die Kreissymbole



**Abbildung 2.3:** Ergebnisse der Trägerperiodizitätsdetektion. Die Kreise und die durchgezogene Linie repräsentieren die Daten, die ohne zusätzliche AM erlangt wurden. Die Rauten und die gestrichelte Linie stellen die Werte mit AM dar. Im Gegensatz zur Modulationsdetektion erhöht sich die Detektionsleistung für Trägerperiodizität mit zunehmender IRN F0, die das Pendant zur Modulationsfrequenz darstellt. Darüberhinaus beeinflusst die AM die Detektionsleistung ausschließlich im niederfrequenten Bereich.

repräsentieren die Detektionsschwellen von Trägerperiodizität in *Abwesenheit* einer Amplitudenmodulation, sprich Einhüllendenperiodizität. Die Karos repräsentieren die Detektionsschwellen in *Anwesenheit* von Amplitudenmodulation.

Sind die Stimuli nicht amplitudenmoduliert, verbessert sich zunächst die Fähigkeit Trägerperiodizität zu detektieren mit zunehmender IRN F0 und erreicht ein Plateau bei einer F0 von 125 Hz. In Anwesenheit von Einhüllendenperiodizität sind die Detektionsschwellen für F0s von 31,25 Hz und darüber nahezu unverändert. In dem tieferliegenden Bereich sind die Schwellen allerdings stark beeinträchtigt.

## 2.4 Diskussion

Die experimentellen Daten zeigen, dass die Wahrnehmung von Amplitudenmodulation durch die Anwesenheit von Trägerperiodizität deutlich beeinträchtigt wird, und zwar für den gesamten getesteten Frequenzbereich. Wohingegen die Anwesenheit von Amplitudenmodulation die Wahrnehmung von Trägerperiodizität nicht maßgeblich beeinträchtigt, solange die IRN F0 mindestens 31,25 Hz beträgt. Im sehr tiefen Frequenzbereich (IRN F0 von 7,8125 und 15,625 Hz) ist eine deutliche Herabsetzung der Detektionsschwelle zu beobachten.

In der Einleitung wurde darauf hingewiesen, dass sowohl IRN als auch SAMN eine Tonhöhenwahrnehmung hervorrufen und es ist anzunehmen, dass Tonhöhe sowie die Ausgeprägtheit der Tonhöhe die wichtigsten Kriterien waren, nach denen die Probanden ihre Entscheidungen über das Signalintervall trafen. Jedoch wird Tonhöhe erst ab einer Fre-

quenz von ca. 30 Hz wahrgenommen (Krumbholz et al., 2000; Pressnitzer et al., 2001). Im Folgenden werden die Ergebnisse im Hinblick auf Tonhöhenwahrnehmung diskutiert. Dazu wird der Frequenzbereich um den es hier geht in zwei Bereiche unterteilt. Auf der einen Seite steht der Bereich, der eine Tonhöhenwahrnehmung produziert. Dieser liegt zwischen 30 und 1000 Hz. Frequenzen unterhalb von 30 Hz können keine Tonhöhe produzieren. Diese 30 Hz Marke, genannt *lower limit of pitch* (LLP), ist auch genau die Grenze, an der die Detektionsleistung für Trägerperiodizität divergiert (siehe Abb. 2.3). Solange das IRN eine Tonhöhenwahrnehmung hervorruft, wird die Detektion von Trägerperiodizität durch die Anwesenheit von Einhüllendenperiodizität *nicht* beeinträchtigt.

Für IRN F0s unterhalb des LLP verliert das IRN seine Tonalität, stattdessen wird ein repetitives Element wahrgenommen, das den Lautstärkeschwankungen, die durch die AM hervorgerufen wird, sehr ähnelt. Es ist gut vorstellbar, dass die Probanden Schwierigkeiten hatten zu unterscheiden, ob diese periodischen Schwankungen von der Periodizität des Trägers oder der Einhüllenden hervorgerufen wird. Eine Befragung der Probanden im Anschluss an die Messreihen bestätigen diese Annahme. Dieser Annahme zur Folge würde man erwarten, dass, unterhalb des LLP, die Detektionsschwellen für Trägerperiodizität deutlich ansteigen, wenn die Stimuli zusätzlich amplitudenmoduliert sind. Genau dies ist in den Daten zu beobachten. Diese Tatsache wäre auch eine mögliche Erklärung für den Abfall der Detektionsleistung für Amplitudenmodulation unter Anwesenheit von Trägerperiodizität im niederfrequenten Bereich (unterhalb von 30 Hz) in Experiment I. Allerdings beeinträchtigt der periodische Träger die AM-Detektionsschwellen auch im Tonhöhenbereich. Psychoakustische Experimente zur Frequenzunterscheidung mit IRN (Krumbholz et al., 2000) und SAMN (Formby, 1985) weisen darauf hin, dass IRN eine wesentlich ausgeprägtere Tonhöhe erzeugt als SAMN. Nimmt man an, dass Tonhöhe das vorrangige Entscheidungskriterium der Probanden war, würde die schwach ausgeprägte Tonhöhe des SAMN von der stark ausgeprägten Tonhöhe des IRN überlagert werden, und die Detektionsschwellen für Amplitudenmodulation mit dem periodischen Träger (IRN) würden ansteigen. Dieser Trend ist ebenfalls in den Daten zu beobachten (siehe Abb. 2.2). Für die Detektionsschwellen für Trägerperiodizität im Tonhöhenbereich würde diese Annahme folgendes bedeuten: Die schwach ausgeprägte Tonhöhe des SAMN kann die Wahrnehmung der wesentlich stärker ausgeprägten Tonhöhe des IRN nicht wesentlich beeinträchtigen. Das entspricht genau den ermittelten psychoakustischen Daten (siehe Abb. 2.3).

## 2.5 Modell Simulationen der Experimente

Das Modulations-Filterbank-Modell (Dau et al., 1997a,b) und das Autokorrelationsmodell (AC-Modell) zur Tonhöhenwahrnehmung (Meddis und Hewitt, 1991a,b) wurden in der Einleitung bereits kurz erläutert. Im folgenden Abschnitt werden die beiden psychoakustischen Experimente in den jeweiligen Modellen simuliert. Sowohl das AC-Modell als auch die Modulationsfilterbank (MFB) extrahieren letztendlich die temporale Periodizität der Feuerwahrscheinlichkeit in der Hörnervfaser. Dennoch ist das AC-Modell bislang nur zur Simulation von Tonhöhenphänomenen herangezogen worden, d.h. zur Erfassung von Trä-

gerperiodizität. Mit der MFB wurde bislang nur Amplitudenmodulationswahrnehmung simuliert, sprich es wurde bislang nur Einhüllendenperiodizität mit diesem Modell erfasst.

Hier soll nun getestet werden, ob die Modelle in der Lage sind die experimentell gewonnenen Daten zu reproduzieren. Das Hauptaugenmerk liegt dabei in der Frage: Können die Modelle auch die Ergebnisse reproduzieren, die außerhalb ihres Bereiches liegen?

### 2.5.1 Das Modulations-Filterbank-Modell

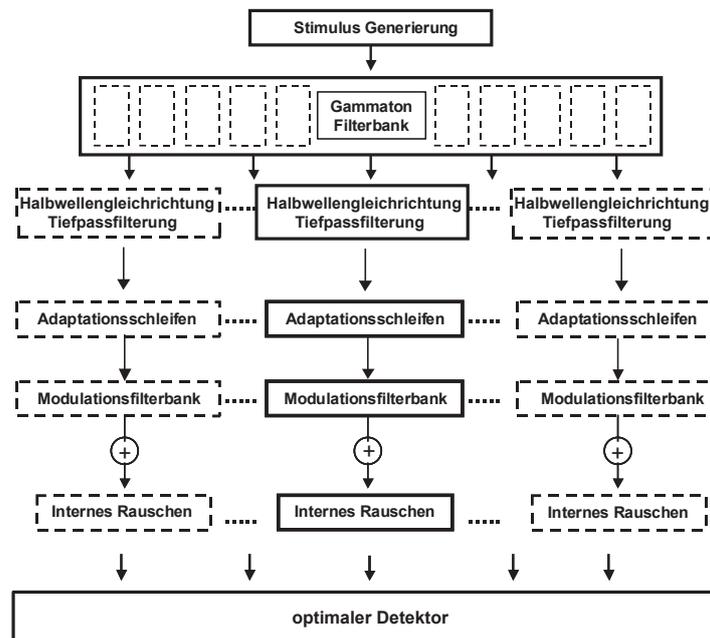
Das von Dau et al. (1997a,b) entwickelte Modell implementiert eine Modulations-Filterbank, um die Einhüllende von eingehenden Stimuli zu analysieren. Die folgenden Simulationen wurden mit einer MatLab-basierten Version des 1997 veröffentlichten Modells durchgeführt. Die Software wurde von Dr. Stephan Ewert zur Verfügung gestellt und die Simulationen wurden mit seiner Unterstützung durchgeführt.

#### Der Aufbau des Modells

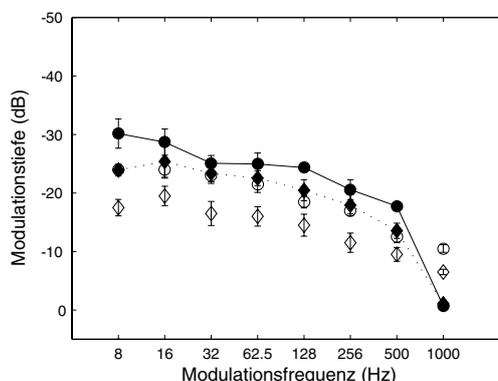
Ein Blockdiagramm dieses Modells ist in Abb. 2.4 zu sehen. Der erste Verarbeitungsschritt in diesem Modell simuliert die Frequenz-Orts-Transformation der Cochlea. In der Cochlea produzieren Sinusschwingungen unterschiedlicher Frequenzen eine maximale Resonanz an einem definierten Ort entlang der Cochlea. Insofern hat jeder Ort entlang der Cochlea eine Bandpasscharakteristik. Um dies zu simulieren, wird hier die von Patterson et al. (1992) entwickelte Gammaton-Filterbank implementiert. Der Ausgang jedes einzelnen Filters entlang der Gammaton-Filterbank wird halbwellengleichgerichtet und dann tiefpassgefiltert, bei einer Eckfrequenz von 1000 Hz, um die mechano-neurale Transduktion zu simulieren. Darauf folgen Adaptationsschleifen (Münkner, 1993), die Adaptationsvorgänge in den Inneren Haarzellen nachbilden.

Der Ausgang der Adaptationsschleifen entspricht nun der Aktivität der Hörnervfasern. An jedem Kanal, die Menge ist durch die Anzahl der Filter in der Gammatonfilterbank gegeben, hängt im Anschluss an die Adaptationsschleifen eine weitere Filterbank, die Modulations-Filterbank.

Für die Simulationen der vorliegenden Arbeit hatte die Gammatonfilterbank 30 Bandpassfilter, deren Mittenfrequenzen äquidistant entlang einer ERB-Achse von 100 bis 8000 Hz aneinandergereiht waren. Die Modulations-Filterbank (MFB) beinhaltete elf Modulationsfilter. Die Modulationsfilter waren Resonanzfilter erster Ordnung, mit einem Q-Wert von 2 (Dau et al., 1997a,b). Die Modulationsfilter, mit Mittenfrequenzen von 10 bis 1000 Hz, waren äquidistant entlang einer logarithmischen Frequenzachse aneinandergereiht und überlappten an den 3 dB - Punkten. Im Anschluss an die MFB wird noch *internes Rauschen* addiert, was einem physiologischen Grundrauschen entspricht. Abschließend wird der Ausgang dieser Stufe mit einem Entscheidungsmodul evaluiert. Das Entscheidungsmodul in diesem Modell ist ein *optimaler Detektor* (Dau et al., 1996). Diese Methode basiert auf dem Vergleich zwischen der internen Repräsentation des eingehenden Signals und einer zuvor errechneten Vergleichsvorlage, dem *template* ( $T$ ). Da es auch in Messungen mit Versuchspersonen geschieht, dass der Hörer sich erst „inhören“ muss, wird zunächst mal



**Abbildung 2.4:** Schematische Darstellung des Modulationsfilterbank-Modells. Die Gammatonfilterbank simuliert die Bandpasscharakteristik der Basilmembran. Die anschließende Halbwellengleichrichtung und Tiefpassfilterung simuliert die mechano-neuronale Transduktion. Die Adaptationsschleifen sollen Adaptationsvorgängen in den inneren Haarzellen Rechnung tragen. Darauf folgt die Modulationsfilterbank. Schließlich wird noch das interne Rauschen addiert. Daraus resultiert die interne Repräsentation des Stimulus, die dann durch den optimalen Detektor ausgewertet wird.



**Abbildung 2.5:** Simulationsergebnis der Modulationsdetektion mit dem MFB Modell. Die ausgefüllten Kreise mit den durchgezogenen Linien repräsentieren die Vorhersagen für einen GN Träger. Die ausgefüllte Rauten und die gepunktete Linie stellen die Vorhersagen für einen IRN Träger dar. Die offenen Symbole repräsentieren die entsprechenden experimentellen Daten. Der Trend der empirischen Daten wird durch die Simulation recht gut wiedergegeben. Allerdings wird der Einfluss von IRN auf die Detektionsschwellen vom Modell unterschätzt (hier im Mittel ca. 3 dB im Gegensatz zu 5 dB). Für 1000 Hz Modulationsfrequenz liegt die Vorhersage (-1 dB) weit unter der gemessenen Schwelle (-11 dB)

die interne Repräsentation des überschwellig dargebotenen Signalstimulus  $R(sig)$  errechnet sowie die interne Repräsentation des Standardstimulus  $R(std)$ . Die Differenz dieser Beiden stellt dann das *template* dar.

$$T = R(sig) - R(std) \quad (2.1)$$

Die Energie des *templates* wird auf eins normiert, d.h. der Effektivwert wird gleich eins gesetzt. Das interne Rauschen wurde derart festgelegt, dass die vorhergesagte Schwelle im Bereich einer 16-64 Hz Amplitudenmodulation eines GN Trägers um die -25 dB betrug. Dieser Wert wurde dann für alle simulierten Konditionen beibehalten.

Die Simulation lief nach dem selben adaptiven Verfahren (2AFC), wie es auch in den psychoakustischen Experimente angewandt wurde. Sowohl Standard- als auch Signalstimulus wurden, nachdem sie die MFB passierten, mit dem *template* kreuzkorreliert. Das Intervall mit dem höheren Korrelationswert wurde als Wahlintervall gewertet. Stimmt dieses mit dem Signalintervall überein, wurde die Entscheidung als korrekt gewertet, wenn nicht, wurde sie als falsch gewertet. Es wurden für jede Signalkonfiguration je drei Schwellenwerte ermittelt. Deren Mittelwert galt als Ergebnis der Simulation.

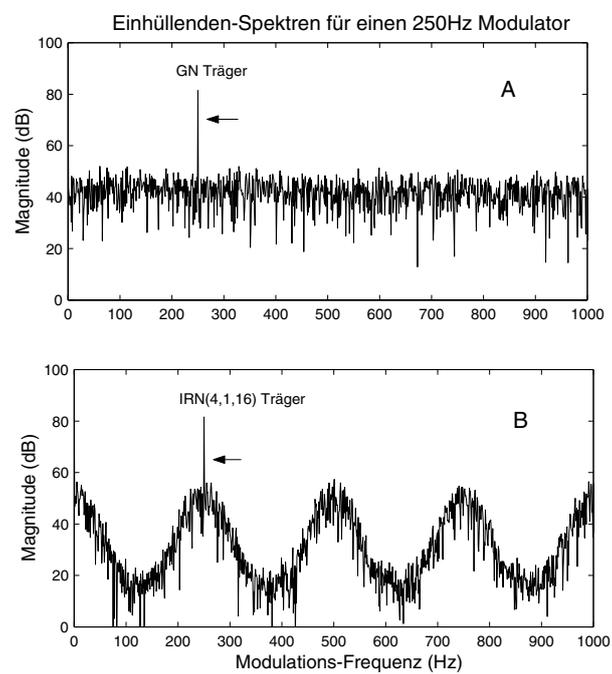
## Ergebnisse und Diskussion

**Simulation von Experiment I: Detektion von Einhüllendenperiodizität** In Abb. 2.5 sind die Ergebnisse der Simulation von Experiment I dargestellt. Die ausgefüllten Kreise und die durchgezogene Linie repräsentieren die Detektionsschwellen für Amplitudenmodulation mit GN als Träger. Die ausgefüllten Rauten und die gepunktete Linie symbolisieren die Schwellen für einen IRN-Träger. Für Modulationsfrequenzen zwischen 8 und 500 Hz und GN als Träger, spiegelt die Simulation die gemessenen Daten relativ gut wider, auch wenn die Empfindlichkeit des Modells insgesamt etwas größer ist. Bei einer Modulationsfrequenz von 1000 Hz dagegen, wird die Detektionsleistung vom Modell stark unterschätzt. In den experimentellen Daten liegt die Schwelle bei ca. -11 dB, in der Simulation nur bei -1 dB. Das Anheben der Schwellen durch den periodischen Träger IRN, wird ebenfalls im Bereich zwischen 8 und 500 Hz Modulationsfrequenz (bzw. IRN F0) relativ gut wiedergegeben. Jedoch werden im Modell die Schwellen im Schnitt nur ca. 3 dB angehoben, wogegen die experimentellen Daten einen Einbruch von ca 5 dB zeigen. Bei der höchsten Modulationsfrequenz von 1000 Hz zeigt das Modell keinen Unterschied im Bezug auf den Träger. Eine mögliche Erklärung dafür wäre die spektrale Begrenzung der auditorischen Filter in der Gammaton-Filterbank. Der höchste auditorische Filter hat eine Mittenfrequenz von 8000 Hz, d.h. er hat eine ERB von 889 Hz. Das ist möglicherweise zu schmal, um eine Amplitudenmodulation von 1000 Hz zu übertragen. Im Experiment wurden Frequenzen bis zu 20 kHz übertragen, somit hatten die Hörer die Möglichkeit, die Information über die AM mittels sehr hochfrequent abgestimmter auditorischer Filter zu bekommen, denn Filter in diesem Bereich sind breiter als 1000 Hz. Ein zusätzliches Problem könnte die Tatsache darstellen, dass nur ein einziger Filter entlang der MFB breit genug wäre, um eine solche Modulationsfrequenz zu erfassen, wohingegen es bei allen anderen Modulationsfrequenzen auf jeden Fall mehr als einer sind.

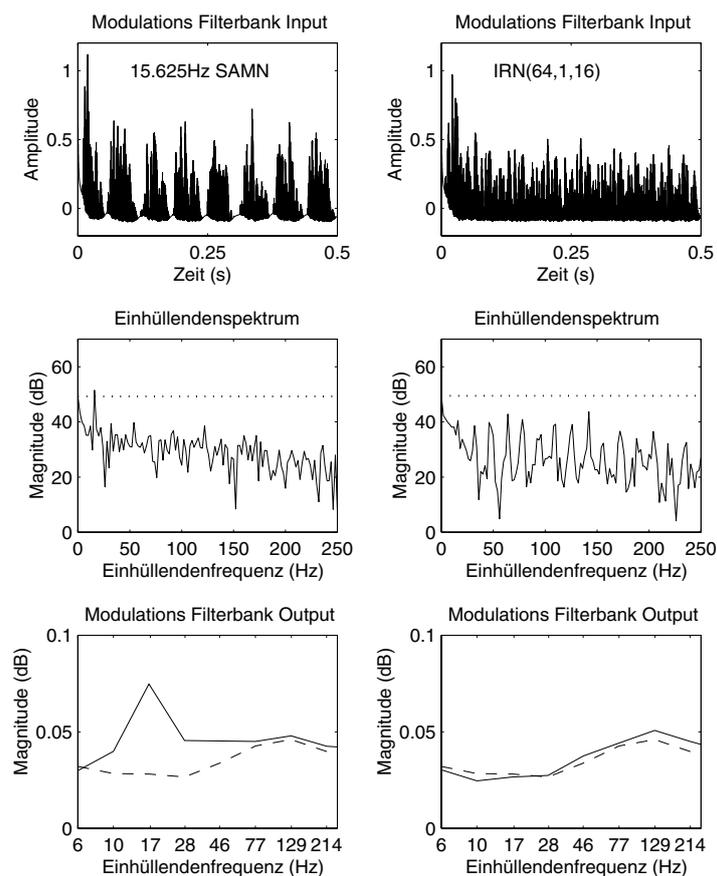
Es ist allerdings nicht ganz klar, warum das Modell nur eine 3 dB Verschlechterung der Schwellen bei einem IRN-Träger errechnet, im Gegensatz zu 5 dB in den empirischen Daten.

Wie im Magnitudenspektrum, zeigt IRN im Modulationsspektrum ebenfalls ein Kammfilterprofil, dessen erstes Maximum an der Stelle der Grundfrequenz (Abb. 2.6B) liegt. GN hingegen zeigt ein flaches Modulationsspektrum (Abb. 2.6A). Es könnte also sein, dass daher das IRN die Detektionsschwelle für Amplitudenmodulation anhebt, da das Modulationsmaximum (Pfeile in Abb.2.6) vor einem GN Hintergrund besser hervorsticht. Jede Filterbank mit Bandpassfiltern einer bestimmten Breite hat jedoch eine begrenzte Auflösung. Das heißt dass ein Filter, dessen Bandbreite größer ist als der Abstand zwischen zwei (oder mehr) spektralem Maxima, diese nicht mehr sauber abbilden kann. Dadurch sind sie am Ausgang eines solchen Filters nicht mehr erkennbar. Für die Modulationsfilter, im Bezug auf IRN, bedeutet dies dass das Kammfilterprofil, je nach Bandbreite der Modulationsfilter, geglättet wird.

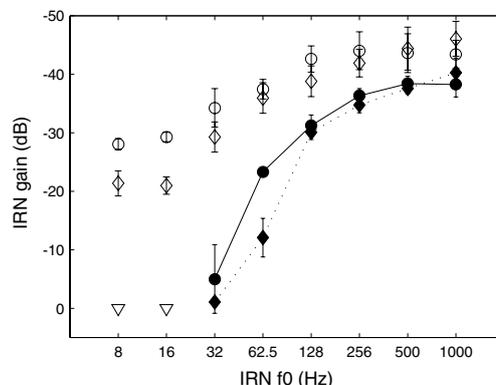
Daher würde man erwarten, dass der Einfluss eines IRN-Trägers davon abhängt, inwieweit das Kammfilterprofil am Ausgang der Modulationsfilter erhalten bleibt. Abb. 2.7 zeigt die interne Repräsentation eines 15,625 Hz SAMN (linken Spalte) und eines entsprechenden



**Abbildung 2.6:** Einhüllendenspektren für einen GN-Träger (A) und einen IRN(4,1,16)-Träger (B) mit einem 250 Hz Sinus amplitudenmoduliert mit einer Modulationstiefe von 0 dB. Die AM produziert ein einzelnes Maximum im Einhüllendenspektrum (Pfeile). Im Falle des IRN-Trägers sitzt dieses Maximum einem, durch das Kammfilterprofil des IRN bereits gegebenem, auf.



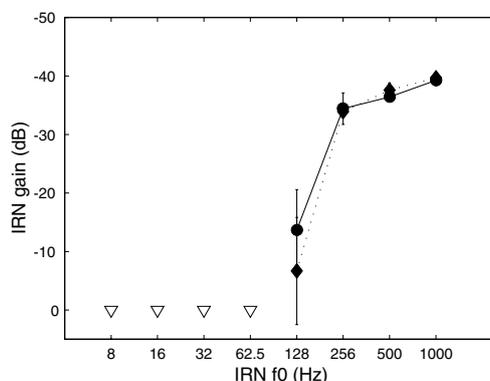
**Abbildung 2.7:** Verarbeitung eines SAMN und eines IRN mit einer Modulationsfrequenz bzw.  $F_0$  von 15.625 Hz im Modulations-Filterbank Modell. Die oberste Reihe zeigt die Spike-Wahrscheinlichkeit einer Hörnervfaser die auf 1 kHz abgestimmt ist, welche den Input in das MFB Modell darstellt. Die mittlere Reihe zeigt die dazugehörigen Einhüllendenspektren. Die gepunktete Linie repräsentiert die DC-Komponente. Zu beachten ist, dass SAMN ein einzelnes Maximum aufweist, IRN dagegen ein harmonisch strukturiertes Einhüllendenspektrum. Die unterste Reihe zeigt die Magnitude des Modell Outputs als Funktion der Modulationsfrequenz. Die gestrichelte Linie repräsentiert den Output des Modells für GN. Es zeigt sich, dass, während das einzelne Maximum des SAMN sehr gut erhalten ist, die multiplen Maxima des IRN nicht aufgelöst werden können. Daher ist der Output für ein IRN dem Output für GN sehr ähnlich.



**Abbildung 2.8:** Ergebnis zur Trägerperiodizitätsdetektion mit dem MFB Modell. Die ausgefüllten Kreise mit der durchgezogenen Linie repräsentieren die Detektionsvorhersagen ohne AM, die ausgefüllten Rauten und die gestrichelte Linie die Vorhersagen mit AM. Die offenen Symbole repräsentieren die entsprechenden experimentellen Daten. Die umgekehrten Dreiecke symbolisieren die Messpunkte, bei denen das Modell nicht in der Lage war, einen Wert vorherzusagen. Während es bei höheren F0s ( $\geq 125$  Hz) noch in der Lage ist, die empirischen Werte recht gut nachzubilden, weichen die Vorhersagen mit abnehmender F0 immer mehr ab (vgl. Abb. 2.3).

IRN ( $d = 64$ ms, rechte Spalte) in einem auditorischen Filter mit der Mittenfrequenz von 1000 Hz. So liegt das Signal vor, bevor es in der Modulationsfilterbank verarbeitet wird. Deren Modulationspektren sind in der mittleren Reihe abgebildet und in der unteren Reihe die jeweiligen Ausgänge der MFB. Die gestrichelte Linie in den beiden untersten Feldern repräsentiert jeweils den Ausgang eines unmodulierten GN. Die durchgezogene Linie im rechten unteren Feld entspricht dem IRN am Ausgang der MFB. Vergleicht man diese beiden Linien müsste man annehmen, dass es für die MFB in diesem Falle keinen Unterschied macht ob der Träger GN oder IRN ist. Im linken unteren Feld repräsentiert die durchgezogene Linie den Ausgang der MFB in Antwort auf ein SAMN. Die gestrichelte Linie stellt wieder den Ausgang eines unmodulierten GN dar. Anhand der durchgezogenen Linie sieht man, dass die Amplitudenmodulation von 15,625 Hz am Ausgang gut erhalten ist. Das ist insofern logisch, als dass eine derartige Amplitudenmodulation nur ein einzelnes Maximum im Modulationsspektrum produziert und somit immer erfasst wird, unabhängig von der Bandbreite des Modulationsfilters. Dass das Modell dennoch eine Erhöhung der Schwelle von ca. 3 dB vorhersagt, deutet darauf hin, dass im Mittel doch ein gewisser Unterschied zwischen GN und IRN am Ausgang der MFB existiert.

**Simulation von Experiment II: Detektion von Trägerperiodizität** Das Modell war für die Simulation dieses Experiments exakt gleich aufgebaut, wie im ersten Teil. Auch das interne Rauschen hatte den gleichen Wert.



**Abbildung 2.9:** Hier wurde die MFB durch einen 8 Hz Tiefpassfilter ersetzt. Das hat zur Folge, dass das Modell nur noch das audiospektrale Erregungsmuster auswerten kann. Die Vorhersagen für  $F_0s \geq 250$  Hz sind kaum verändert. Das deutet darauf hin, dass die Detektion von Trägerperiodizität im MFB Modell hauptsächlich auf audiospektralen Hinweisen, die durch das Kammfilterprofil des IRNs gegeben sind, beruht.

In Abb. 2.8 sind die Vorhersagen des Modells zu sehen. Die ausgefüllten Kreissymbole mit der durchgezogenen Linie repräsentieren die Detektion von Trägerperiodizität ohne Amplitudenmodulation. Die ausgefüllten Rauten mit der gepunkteten Linie symbolisieren die Detektion von Trägerperiodizität mit Amplitudenmodulation. Für den Bereich der höheren IRN  $F_0s$ , ab 125 Hz, spiegelt die Vorhersage des Modells die empirisch gewonnenen Daten recht gut wider. Für tieferliegende IRN  $F_0s$ , gleichbedeutend mit längeren Verzögerungen, sind die vorhergesagten Schwellen wesentlich höher als die gemessenen. IRN mit  $F_0s$  von 8 und 16 Hz kann das Modell überhaupt nicht von GN unterscheiden (umgekehrte Dreiecke in Abb. 2.8). Ob die Stimuli zusätzlich amplitudenmoduliert sind oder nicht, scheint keinen großen Unterschied zu machen. Lediglich bei einer IRN  $F_0$  von 64 Hz zeigt das Modell deutliche Unterschiede. Da die Modulationsmittelfrequenzen bis 10 Hz runtergehen, und für tiefere Frequenzen ein Modulations-Tiefpassfilter implementiert ist, ist es auf den ersten Blick verwunderlich, dass es nicht in der Lage ist IRN  $F_0s$  von 16 bzw. 8 Hz zu detektieren.

Daher wurde der Rolle, die die Modulationsfilter hier spielen, näher auf den Grund gegangen. Dazu wurde die MFB durch einen Tiefpassfilter mit der Eckfrequenz 8 Hz ersetzt.

Das hat zur Folge, dass dem Modell sämtliche Informationen über Modulation vorenthalten bleiben. Es könnte bestenfalls Modulation unterhalb von 8 Hz erkennen, jedoch nicht selektiv. Betrachtet man die Vorhersagen dieses reduzierten Modells (Abb. 2.9), stellt man fest, dass sich gegenüber dem MFB Ansatz im Bereich oberhalb von 125 Hz nichts ändert. Das deutet darauf hin, dass diese Ergebnisse nicht auf Modulationsinformationen beruhen, sondern dass es sich stattdessen um audiospektrale Hinweise handelt. Zwischen 125 und

32 Hz scheinen dagegen sehr wohl Modulationsinformationen herangezogen zu werden. Ohne diese kann das Modell für 32 und 64 Hz IRN F0 gar keinen Unterschied zwischen GN und IRN feststellen, bei 125 Hz F0 sind die Schwellen deutlich höher. In Abschnitt 2.5.1 wurde bereits festgestellt, dass die MFB das Kammfilterprofil eines IRN für niedrige ( $< 31,25$  Hz) F0s nicht auflösen kann. Dieses Phänomen spielt natürlich auch hier eine Rolle. In Abb. 2.7, im rechten unteren Feld, kann man gut erkennen, dass am Ausgang der MFB kaum Unterschied zwischen GN (gestrichelte Linie) und IRN (durchgezogene Linie) zu erkennen ist. Das Kammfilterprofil des IRN ist völlig geglättet.

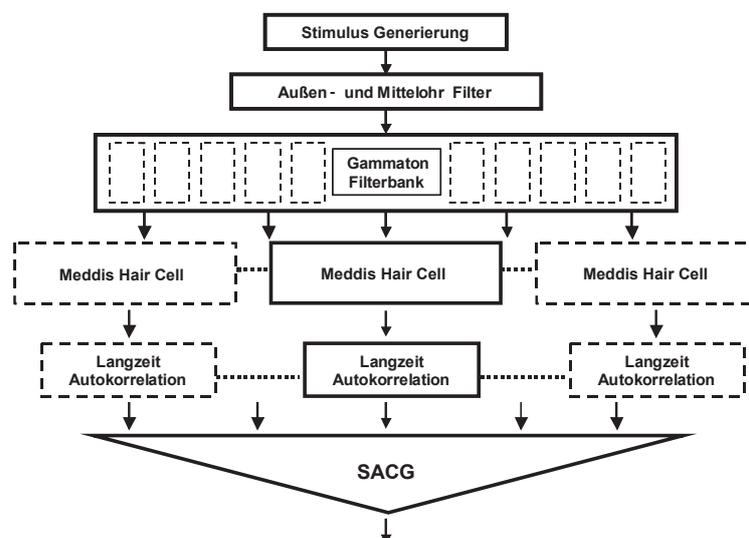
Zusammenfassend lässt sich feststellen, dass die Modulationsfilterbank in ihrer derzeitigen Form nicht in der Lage ist, die harmonische Struktur des Einhüllendenspektrums von Stimuli mit Trägerperiodizität aufzulösen.

## 2.5.2 Das Autokorrelations-Modell

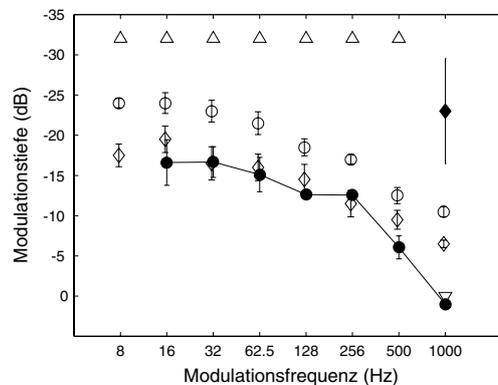
Das Autokorrelations-Modell (AC-Modell) ist eines der erfolgreichsten Tonhöhenmodelle, das eine Vielzahl an Wahrnehmungsphänomenen korrekt vorhersagen kann. Dieses Modell basiert auf einer Autokorrelation der Feuerwahrscheinlichkeit (SPF - *spike probability function*) der Hörnervfasern. Eingeführt wurde das Modell von Meddis und Hewitt (1991a,b). Einige Jahre später wurde das Modell nochmals, an bis dato aktuelleren Erkenntnissen, getestet (Meddis und O'Mard, 1997).

### Der Aufbau des Modells

Zuerst werden die Stimuli mit einer Außen- und einer Mittelohr-Transferfunktion verrechnet. Anschließend passieren die Signale eine 30-Kanalige Gammaton-Filterbank (Patterson et al., 1992). Die Mittenfrequenzen lagen zwischen 100 und 800 Hz, wie auch beim MFB-Modell. Die mechano-neurale Transduktion war in Form einer Reihe simulierter Innerer Haarzellen implementiert, die die Halbwellenrichtung sowie kompressive und adaptive Effekte modellieren (Abb. 2.10) Am Ausgang dieses letzten Moduls steht die Feuerwahrscheinlichkeit der Hörnervfaser als Funktion der Zeit und des auditorischen Filters. Der hauptsächliche Unterschied zur Modulations-Filterbank ist die Art und Weise, mit der nun die Periodizität aus dieser Funktion extrahiert wird. Das MFB-Modell knüpft eine komplette Filterbank (die Modulationsfilter) an jeden auditorischen Kanal an. Das AC-Modell vollzieht eine zeitvariante Autokorrelation der SPF der einzelnen auditorischen Kanäle, d.h. die Autokorrelation wird innerhalb eines exponentiellen Zeitfensters berechnet, das entlang der Zeitachse verschoben wird. Die Zeitkonstante wurde für die, in dieser Arbeit vorgestellten, Simulationen auf 100 ms festgesetzt. In den ursprünglichen Arbeiten (Meddis und Hewitt, 1991a,b) betrug die Zeitkonstante 2,5 ms. In nachfolgenden Anwendungen des Modells, wurde sie jedoch wesentlich verändert. Bei Signalen deren Grad an Periodizität unverändert entlang der Zeitachse bleibt, ermöglicht eine größere Zeitkonstante eine stabilere Schätzung der Tonhöhe und deren Ausprägtheit (Meddis und Hewitt, 1991b; Wiegrebe, 2001). Der maximale Verzögerungswert und somit auch die längste Periode, die der Autokorrelator erfassen kann, wurde auf 30 ms festgesetzt. Diese 30 ms entsprechen



**Abbildung 2.10:** Schematische Darstellung des Autokorrelations-Modells. Zunächst wird der eingehende Stimulus mit einer Außen- und Mittelohr Übertragungsfunktion multipliziert. Daran schließt sich eine Gammatonfilterbank an, die die Bandpasscharakteristik der Basilarmembran simuliert. Das Modul der *Meddis Hair Cell* simuliert die Übertragungseigenschaften der inneren Haarzellen. Diese beinhalten Halbwellenrichtung, Kompression, Tiefpassfilterung sowie Adaptationsvorgänge. Deren Ausgang wird dann einer Langzeit-Autokorrelation unterzogen. Am Ende wird dann über alle Kanäle ein Summenautokorrelationsgramm errechnet. Die Entscheidung über das Signalintervall erfolgt über die Euklidische Distanz der eingehenden, zu vergleichenden Stimuli.



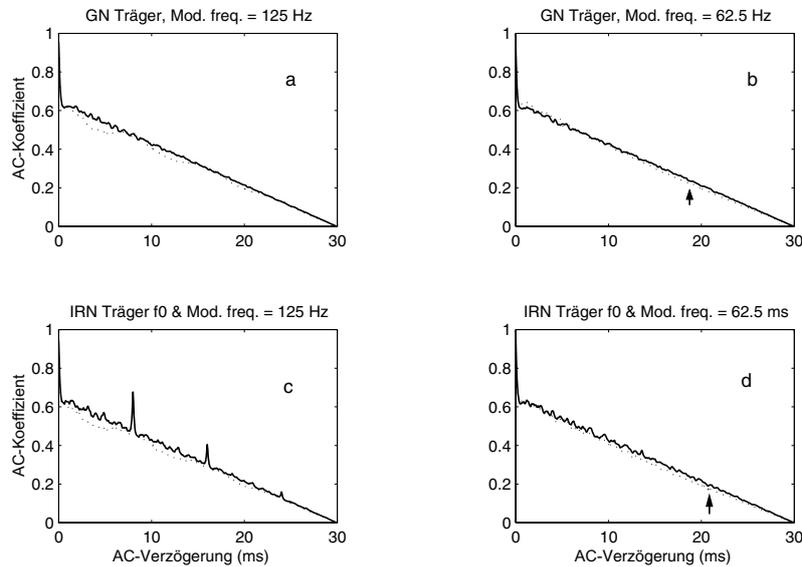
**Abbildung 2.11:** Simulationsergebnis der Modulationsdetektion mit dem AC-Modell. Die Kreise und die durchgezogene Linie repräsentieren die Vorhersagen mit einem GN-Träger. Das abwärtsgerichtete Dreieck sagt aus, dass für diesen Punkt keine Vorhersage (für einen GN Träger) ermittelt werden konnte. Wie in den empirischen Daten nimmt auch hier die Fähigkeit, AM zu detektieren, mit zunehmender Modulationsfrequenz ab. Die Modellvorhersagen für einen IRN Träger werden zum einen mit der Raute, zum anderen mit den aufwärtsgerichteten Dreiecken symbolisiert. Die aufwärtsgerichteten Dreiecke sollen darstellen, dass bei diesen Modulationsfrequenzen, kein sinnvoller Wert vorhergesagt werden konnte, d.h. sie lagen weit unter -50 dB.

der tiefsten wahrnehmbaren Tonhöhe (Krumbholz et al., 2000; Pressnitzer et al., 2001). Die Autokorrelationsfunktionen aller Audio-Frequenzkanäle werden aufsummiert zu einem Summenautokorrelogramm (SACG). Die Position sowie die Höhe der Maxima im SACG sollen dann Auskunft über die Tonhöhe bzw. deren Ausprägtheit geben.

Aktuelle Simulationen mit dem AC-Modell haben ergeben, dass die Änderungen in der Wahrnehmung proportional zur Euklidischen Distanz zwischen den SACGs zweier Stimuli zu sein scheint. Daher wird in dieser Simulation gemäß Pressnitzer et al. (2001), die Euklidische Distanz (ED) als Entscheidungsmechanismus herangezogen. Um das Signalintervall korrekt zu identifizieren, musste die ED der beiden SACGs einen bestimmten Schwellenwert überschreiten. Dieser Schwellenwert bestimmte somit auch die Empfindlichkeit des Modells allgemein. Festgelegt wurde dieser Wert mittels einer Kalibrierungs-Simulation. Ermittelt wurde die ED, mit der die Detektion von Trägerperiodizität bei IRN mit einer  $F_0$  von 250 Hz korrekt vorhergesagt wurde. Diese ED war das Schwellenkriterium für sämtliche Signalkonfigurationen. Die simulierte Schwellenermittlung erfolgte nach demselben Verfahren wie im psychoakustischen Experiment (siehe Abschnitt 2.2.1)

## Ergebnisse und Diskussion

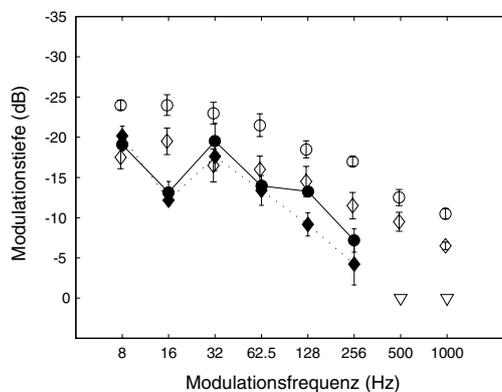
**Simulation von Experiment I: Detektion von Einhüllendenperiodizität** Die Schwellenwerte zur Modulationsdetektion, wie sie das AC-Modell errechnet, sind in Abb.



**Abbildung 2.12:** Grundlage zur Detektion von Amplitudenmodulation im AC-Modell: Hier werden die SACGs in Antwort auf den Standard- und den Signalstimulus (gepunktete und durchgezogene Linie) für eine Modulationsfrequenz von 125 Hz (linke Spalte) und 15,625 Hz (rechte Spalte) gezeigt. In der oberen Reihe war der Träger GN, in der unteren Reihe war der Träger IRN. Die AM des Trägers erzeugt eine sinusförmige Modulation des SACG (gepunktete Linien). Selbst wenn die Modulationsperiode, die dem Reziproken der Modulationsfrequenz entspricht, länger ist als die maximale Autokorrelationsverzögerung (rechte Spalte), kann das Modell dies erkennen (Pfeile). Man beachte die Rauigkeit/Unruhe in der SACG Grundlinie bei IRN (untere Reihe). Darin liegt der Grund für die unrealistischen Vorhersagen zur AM-Detektion mit einem IRN Träger (siehe Text).

2.11 dargestellt. Die vorhergesagten Schwellen für einen GN Träger liegen relativ nah bei den gemessenen Schwellen. Das Modell ist lediglich etwas unempfindlich. Da auch dieses Modell, gleich dem MFB-Modell, nicht in der Lage ist, eine Amplitudenmodulation von 1000 Hz zu detektieren (umgekehrtes Dreieck in Abb. 2.11), scheint die Limitation durch die auditorischen Filter (der auditorische Filter mit der höchsten Mittenfrequenz ist nicht breit genug, um eine 1000 Hz Modulation durchzulassen; vgl. Abschnitt 2.5.1) eine plausible Erklärung zu sein. Was auf den ersten Blick verwunderlich erscheint, ist dass Modulationsfrequenzen von 7,8125 oder auch 15,625 Hz detektiert werden können - wurde doch die maximale Verzögerungszeit auf 30 ms ( $\approx 33$  Hz) festgelegt.

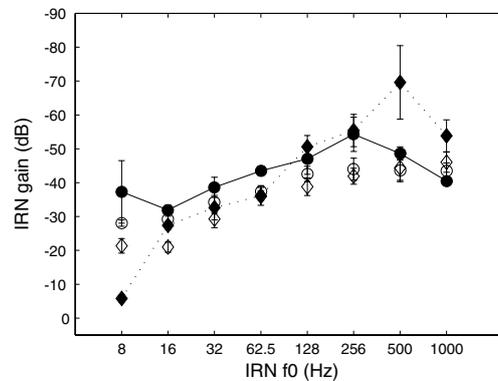
Das liegt daran, dass der Sinusmodulator eine sinusförmige SACG-Modulation hervorruft (vgl. gepunktete Linien in Abb. 2.12a,c), was dazu führen kann, dass ein ausreichender Unterschied zwischen den SACGs des Signal- und Standardintervalls herrscht, selbst wenn die Periode des Modulators länger ist als die maximale Verzögerungszeit (vgl. Pfeile in 2.12 b,d).



**Abbildung 2.13:** In dieser Simulation von AM-Detektion wurde statt der ED ein optimaler Detektor als Entscheidungsmodul eingesetzt. Hiermit werden für beide Träger (GN: durchgezogene Linie und Kreise; IRN: gepunktete Linie und Rauten) sinnvolle Werte ermittelt. Allerdings geht dies zu Lasten der Sensitivität für höhere Modulationsfrequenzen. Die Schwellenerhöhung durch den IRN-Träger kann mit diesem Ansatz auch nicht reproduziert werden.

Ist der Träger IRN, sind die Vorhersagen des Modells viel zu sensitiv. Die aufrechten Dreiecke in Abb. 2.11 sollen demonstrieren, dass die errechneten Schwellen weit jenseits von sinnvollen Werten liegen. Bei näherer Betrachtung der Funktionsweise des Modells, stellt man fest, dass die durch den IRN Träger induzierten Veränderungen im SACG nicht allein auf die deutlichen Maxima beschränkt sind, sondern sich ebenfalls in der Feinstruktur der SACG Grundlinie widerspiegeln. Die SACG Grundlinie zeigt ein sehr viel instabileres Verhalten in Antwort auf ein IRN als dies bei GN der Fall ist. Diese Instabilität wird dadurch verursacht, dass eine bestimmte Verzögerungszeit (entsprechend des *delays* des IRN) stark überrepräsentiert ist. Das hat zur Folge, dass alle anderen Periodizitäten mit einer geringeren Zufälligkeit auftauchen. Daher ist die Euklidische Distanz zwischen Standard- und Signalintervall selbst dann noch größer als der festgelegte Schwellenwert, wenn die Modulationstiefe weitaus geringer ist als -30 dB. Allerdings ist es in der Wahrnehmung höchst unwahrscheinlich, dass das auditorische System bei der Erfassung von Periodizität den Fokus weniger auf die überrepräsentierte Periode legt als auf die Veränderungen in der Statistik anderer Perioden. Dieses Problem könnte möglicherweise umgangen werden, wenn ein anderer Entscheidungsmechanismus implementiert wird.

Der optimale Detektor, der im MFB-Modell angewendet wird, verwendet eine Vergleichsvorlage, die aus 16 überschwelligen Präsentationen des Signals abgeleitet wird. Diese Vorlage wird dann mit dem Signal- und dem Standardintervall kreuzkorreliert. Durch die Mittelung über 16 Präsentationen, kommt die Feinstruktur des SACG in der Vorlage nicht zum Tragen.

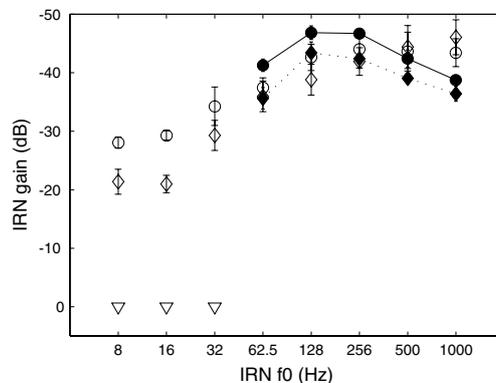


**Abbildung 2.14:** Vorhersagen des AC-Modells zur Trägerperiodizitätsdetektion. Die ausgefüllten Kreise und die durchgezogene Linie zeigen die Vorhersagen ohne AM. Die ausgefüllten Rauten und die gepunktete Linie zeigen die Vorhersagen mit AM. Die offenen Symbole repräsentieren die entsprechenden experimentellen Daten. Ohne AM ist auch hier der Trend der experimentellen Daten (Verbesserung der Detektion mit ansteigender IRN F0) einigermaßen erhalten. Das Modell ist wiederum in der Lage, Trägerperiodizitäten, die dem Reziproken der IRN F0 entsprechen, zu detektieren, die länger sind als die maximale Autokorrelationsverzögerung. Der Einfluss von AM auf die Detektionsschwellen wird nicht korrekt vorhergesagt. Das liegt z.T. wieder am Entscheidungsmodul (siehe Text).

In Abb. 2.13 sind nun die Schwellen dargestellt, wie sie das AC-Modell, unter Verwendung des optimalen Detektors als Entscheidungsmechanismus, vorhersagt. Erwartungsgemäß sind die Instabilitäten, die von der SACG Feinstruktur herrührten nicht mehr vorhanden. Das ermöglicht dem Modell sinnvolle Schwellenwerte für Amplitudenmodulation vorherzusagen, egal ob der Träger GN oder IRN ist. Vergleicht man die Werte des GN Trägers mit denen des IRN Trägers, muss man feststellen, dass auch dieser Ansatz die Minderung in der Detektionsleistung nicht in dem Maße reproduzieren kann, wie sie in den experimentellen Daten gefunden wurde. Die Minderung beträgt lediglich 1 bis 3 dB, im Gegensatz zu 5 dB in den Experimenten.

**Simulation von Experiment II: Detektion von Trägerperiodizität** Sämtliche Parameter, sowie die allgemeine Empfindlichkeit des Modells wurden beibehalten. Es wurde zunächst wieder die Euklidische Distanz als Entscheidungskriterium eingesetzt.

Abbildung 2.14 zeigt das Ergebnis der Simulation. Auch wenn die Schwellen insgesamt etwas zu niedrig ausfallen, spiegelt sich dennoch der allgemeine Trend der empirischen Daten wider. Allerdings gibt es auch signifikante Abweichungen: Bei einer Träger F0 von 500 oder 1000 Hz sind die Schwellen *mit* Amplitudenmodulation (Rauten) deutlich niedriger als *ohne* Amplitudenmodulation (Kreise). Wird ein IRN amplitudenmoduliert, produziert dies im SACG ausgeprägtere Maxima höherer Ordnung. Da das Modell empfindlich für diese Maxima höherer Ordnung ist, sagt es verbesserte Detektionsleistung für diese Kondition



**Abbildung 2.15:** Wiederholung der Simulation zur Trägerperiodizitätsdetektion. Die ED als Entscheidungsmodul, wurde hier ersetzt durch einen optimalen Detektor. Der Vergleich mit Abb. 2.14 macht den großen Einfluss deutlich, den das Entscheidungsmodul auf die Vorhersagen des Modells hat. Auch dieser Ansatz ist nicht in der Lage den Einfluss von AM auf die Detektion von Trägerperiodizität korrekt vorherzusagen.

voraus. Werden die Maxima ab der 5. Ordnung unterdrückt, erhält man ungefähr die gleichen Schwellenwerte mit und ohne zusätzliche Amplitudenmodulation. Da die maximale Verzögerungszeit der Autokorrelation auf 30 ms festgelegt ist, macht sich dieser Effekt bei längeren Perioden nicht bemerkbar.

Überraschend ist auch hier, dass Trägerperioden länger als 30 ms detektiert werden können. Wieder liegt der Grund in der SACG Feinstruktur. Eine lange Trägerperiode hat eine signifikante Veränderung der Statistik der kürzeren Perioden zufolge. Diese Veränderungen wertet das Modell aus und kann dadurch Perioden detektieren, die eigentlich außerhalb der festgelegten Grenze liegen. Für die Wahrnehmung würde das bedeuten, dass der Hörer auf Fluktuationen in der Ausgeprägtheit kurzer Perioden achten würde um lange zu detektieren. Was wiederum äußerst unwahrscheinlich sein dürfte.

Wiederholt man die Simulation mit dem optimalen Detektor als Entscheidungsmodul, welcher die Unterschiede in der Feinstruktur nicht mehr verfügbar macht, können Perioden länger als die maximale Verzögerungszeit auch nicht mehr detektiert werden (Abb. 2.15). Darüberhinaus sind auch die Unterschiede zwischen amplitudenmoduliert und nicht amplitudenmoduliert, wie sie im Falle einer IRN F0 von 500 und 1000 Hz zu sehen waren, nicht mehr gegeben. Die Maxima höherer Ordnung, die diesen Effekt verursachen, sind Teil der Vorlage des optimalen Detektors. Da die Vorlage mit den Stimuli beider Intervalle verglichen wird und die Maxima höherer Ordnung sowohl in der Vorlage als auch in den SACGs beider Intervalle vorhanden sind, können diese nicht mehr zur Signaldetektion beitragen.

## 2.6 Theoretische Zusammenhänge zwischen Modulationsfiltern und Autokorrelation

Das Autokorrelations-Modell und der Modulationsfilter-Ansatz wurden bislang immer als zwei völlig unabhängige Modell-Typen betrachtet. Die vorangegangenen Abschnitte haben jedoch gezeigt, dass durchaus Gemeinsamkeiten bestehen. Sie verfolgen das gleiche Ziel: beide extrahieren Periodizität aus den SPFs (*spike probability functions* - Feuerwahrscheinlichkeiten der Hörnervfaser). Beide liefern, sofern dasselbe Entscheidungsmodul (optimaler Detektor) implementiert wird, relativ ähnliche Ergebnisse. Der hauptsächlichste Unterschied zwischen den Modellen ist der Mechanismus der die Periodizität extrahiert. Das eine Mal sind es Modulationsfilter, das andere Mal ein Autokorrelator. In diesem Abschnitt sollen die mathematischen Zusammenhänge beider Ansätze beschrieben werden. Dazu werden die Begriffe *Impulsantwort* und *Faltung* kontextbezogen beschrieben.

Beschallt man einen Filter mit einem unendlich kurzen Impuls mit (theoretisch) unendlich hoher Amplitude, erhält man dessen *Impulsantwort*. Die Impulsantwort charakterisiert die Filtereigenschaften im Zeitbereich. Daraus folgt, dass ein Filter mit der Impulsantwort  $g(t)$  auf ein Eingangssignal  $x(t)$  mit einem definierten Ausgangssignal  $y(t)$  reagiert, und zwar nach folgender Gleichung:

$$y(t) = \int_{-\infty}^{\infty} x(t) * g(t - \tau) d\tau \quad (2.2)$$

Dieses Integral wird als *Faltungsintegral* bezeichnet. Die durch diese Gleichung definierte Verknüpfung von  $g(t)$  und  $x(t)$  wird folglich als *Faltung* bezeichnet. In einfachen Worten: Das Eingangssignal, gefaltet mit der Impulsantwort des Filters, ergibt das Ausgangssignal.

$$y(t) = x(t) * g(t)$$

Sprich:  $x$  gefaltet  $g$

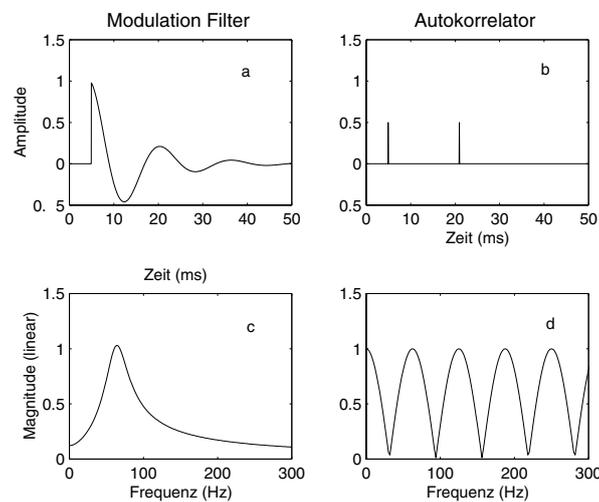
Im Frequenzbereich wird ein Filter charakterisiert durch die *Übertragungsfunktion*. Die Übertragungsfunktion ( $G(\omega)$ ) ist die FOURIER-Transformierte der Impulsantwort. Eine Faltung im Zeitbereich kann durch eine Multiplikation im Frequenzbereich ersetzt werden. Folgende Gleichungen sind also äquivalent:

$$y(t) = x(t) * g(t) \quad (2.3)$$

$$Y(\omega) = X(\omega) \cdot G(\omega) \quad (2.4)$$

$X(\omega)$  und  $Y(\omega)$  sind dabei die Frequenzspektren des Eingangs- bzw. Ausgangssignals.

In zeitdiskreten Systemen, wie digitale Filter sie darstellen, wird das Faltungsintegral zur Faltungssumme:



**Abbildung 2.16:** Hier werden ein Modulationsfilter mit der Mittenfrequenz 62,5 Hz und ein Autokorrelator mit entsprechender Verzögerung ( $1/f=16$  ms) verglichen. Oben links ist die IR des Modulationsfilters abgebildet. Oben rechts die IR, die ein Ausgangssignal erzeugt, ähnlich einem Autokorrelator mit entsprechender Verzögerung. Die untere Reihe zeigt die entsprechenden Übertragungsfunktionen. Während der Modulationsfilter eine Bandpasscharakteristik zeigt, kann der Autokorrelator mit einer U-förmigen IR ausgedrückt werden, die in einer Kammfilter-ähnlichen Transmission resultiert.

$$y[t] = \sum_{i=-\infty}^{+\infty} x[i] \cdot g[t - i] \quad (2.5)$$

Das Ausgangssignal eines Modulationsfilters lässt sich im Zeitbereich als eine Faltung der SPF (Eingangssignal) mit der Impulsantwort des Filters beschreiben.

Die Impulsantwort (IR) eines Modulationsfilters mit einer Mittenfrequenz von 62,5 Hz ist in Abb. 2.16a zu sehen. Abbildung 2.16c zeigt die Übertragungsfunktion dieses Modulationsfilters, die eine deutliche Bandpasscharakteristik aufweist. Die Autokorrelationsfunktion (ACF) eines Signals für eine bestimmte Verzögerung ist mathematisch definiert als eine Summation des Produkts von Signal und einer verzögerten Kopie, entlang der Zeitachse. Für eine spezifische Verzögerung wird sie daher oft mit einer Kammfilteranalyse des Stimulusspektrums assoziiert. Die entsprechende Übertragungsfunktion einer ACF mit der Verzögerung 16 ms ist in Abb. 2.16d zu sehen. Würde man einen derartigen Filter mit einem unendlich kurzen Impuls nun beschallen, ergäbe sich die in Abb. 2.16b dargestellte IR. Diese ist unipolar mit zwei Maxima, deren Abstand der Verzögerung entspricht.

Die Faltung eines Signals ist definiert als die Summation eines Signals mit dessen verzögerten Kopie. Damit unterscheidet sich der mathematische Vorgang deutlich von der Auto-

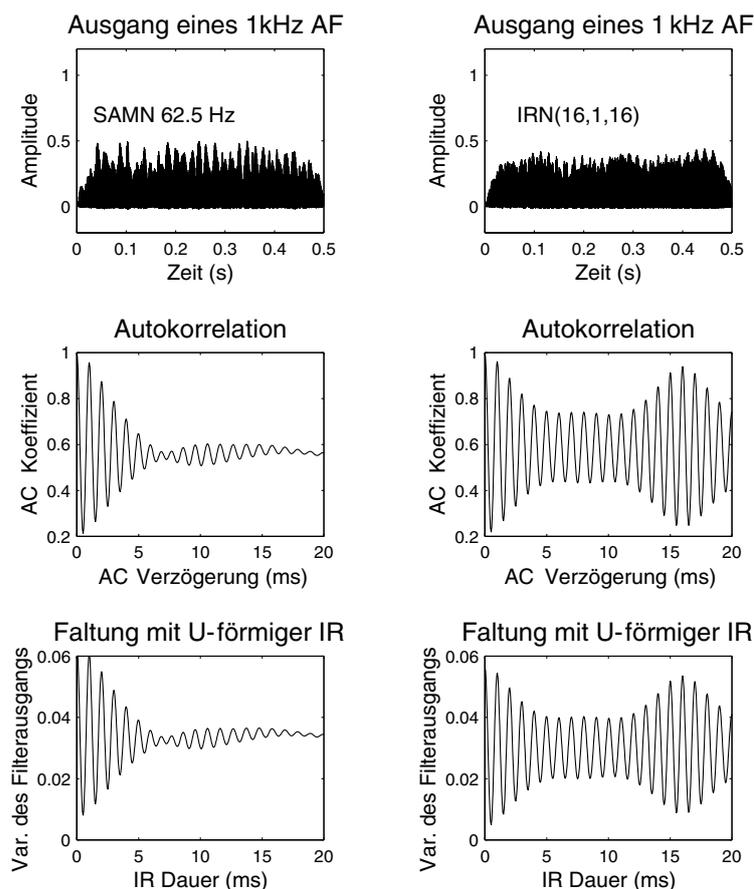
korrelation. Bei der Autokorrelation wird das Signal zunächst mit dessen verzögerten Kopie Punkt für Punkt multipliziert. Erst anschließend werden diese Produkte aufsummiert. Berechnet man allerdings die Varianz des Filterausgangs, liegen wieder zwei mathematisch sehr ähnliche Funktionen vor, da sich die Varianz als Summe der Abweichungsquadrate definiert. Erst werden erst die Abweichungsprodukte gebildet (Multiplikation), die anschließend aufsummiert werden. In Abb. 2.17 werden jeweils von einem 62,5 Hz SAMN, in der linken Spalte und einem IRN(16,1,16), in der rechten Spalte, die SPFs (oberste Reihe), die Autokorrelationen der SPFs (mittlere Reihe) und die Varianz des Ausgangssignals nach einer Faltung mit einer U-förmigen IR. Die Verzögerung mit der die Autokorrelation durchgeführt wurde entspricht dem Abstand der Maxima in der IR. Vergleicht man das Ergebnis der Autokorrelation mit der Varianz des Filterausgangs mit entsprechender U-förmigen Impulsantwort, lässt sich unschwer erkennen, dass beide qualitativ das gleiche Ergebnis liefern.

Da ein Autokorrelator als Filter implementiert werden kann, beschränkt sich der wesentliche Unterschied zwischen einem Autokorrelator und einem Modulationsfilter in der Übertragungsfunktion (Abb.2.16, untere Reihe). Mit dieser kammfilterartigen Übertragungsfunktion ist das AC-Modell in der Lage das Kammfilterprofil im Einhüllendenspektrum von IRN (vgl. Abb. 2.6) aufzulösen, wohingegen die Bandpassübertragungsfunktion der Modulationsfilter zu breit abgestimmt ist um dies zu bewerkstelligen.

## 2.7 Zusammenfassung und Ausblick

Die vorgestellten Experimente haben die Interaktion zwischen der Wahrnehmung von Amplitudenmodulation, die Einhüllendenperiodizität erzeugt, und der Wahrnehmung von Trägerperiodizität untersucht. Die Ergebnisse zeigen, dass die Empfindlichkeit für Amplitudenmodulation beeinträchtigt wurde, wenn der GN-Träger durch einen IRN Träger ersetzt wurde, der die gleiche Periodizität wie die Einhüllende besaß. Dieser Effekt war unabhängig von der Modulationsfrequenz. Die Empfindlichkeit für Trägerperiodizität wurde durch Amplitudenmodulation, die die gleiche Periodizität in der Einhüllenden erzeugte, nur dann beeinträchtigt, wenn die  $F_0$  des entsprechenden IRN (und somit auch die Modulationsfrequenz) an oder unterhalb des *lower limit of pitch*, der unteren Tonhöhenwahrnehmungsgrenze, lag.

Anschließend wurden die Experimente in einem aktuellen Modell zur Modulationswahrnehmung und in einem aktuellen Modell zur Tonhöhenwahrnehmung simuliert. Zwar waren beide Modelle in der Lage, den generellen Trend der empirischen Daten zu reproduzieren, jedoch wurden bei genauerer Betrachtung Mängel in beiden Modellen deutlich. Die Modulationsfilterbank ist durch die relativ breit abgestimmten Filter und deren DC-Sensitivität nicht in der Lage, ein harmonisches Einhüllendenspektrum aufzulösen. Daher zeigt sich in den Vorhersagen ein zu geringer Einfluss des IRN Trägers auf die Modulations-sensitivität. Die Detektion von Trägerperiodizität wird hauptsächlich über audio-spektrale Hinweise vermittelt. Modulationsempfindlichkeit spielte hier nur eine untergeordnete Rolle und beschränkte sich auf  $F_0$ s um die 62,5 Hz. Da das Autokorrelationsmodell über alle



**Abbildung 2.17:** Vergleich zwischen einem U-förmigem Modulationsfilter und einer Autokorrelation. Der Input in die hier abgebildete Periodizitätsanalyse ist die SPF am Ausgang eines auditorischen Filters mit einer Mittenfrequenz von 1 kHz, als Antwort auf ein 62,5 Hz SAMN (linke Spalte) bzw. ein IRN(16,1,16) (rechte Spalte). Die mittlere Reihe zeigt das Ergebnis einer Autokorrelation. Die unterste Reihe zeigt die Varianz des Ausgangssignals bei einer Faltung der SPF mit einer U-förmigen Impulsantwort. Der Abstand der Maxima in der IR entspricht der Verzögerung der Autokorrelation

Frequenzkanäle mittelt, stehen diesem Modell keine audio-spektralen Hinweise zur Verfügung. Somit kann dieses Modell, im Gegensatz zur Modulationsfilterbank, nur rein zeitliche Hinweise bearbeiten. Das Autokorrelationsmodell leidet hauptsächlich unter seinem Entscheidungsmodul. Ersetzt man das Maß der Euklidischen Distanz mit einem optimalen Detektor, können die Vorhersagen verbessert werden. Jedoch ist auch das Autokorrelationsmodell nicht in der Lage, die Beeinträchtigung der Modulationsdetektion durch einen IRN Träger zu reproduzieren. Die Detektion von Trägerperiodizität beschränkt sich auf F0s über 33 Hz.

Ein direkter Vergleich beider Modelle zeigt, dass der grundlegende Unterschied die Impulsantwort bzw. die Übertragungsfunktion ist, mit der die Spike-Wahrscheinlichkeitsfunktion verrechnet wird. Modulationsfilter haben eine Bandpass Übertragungsfunktion, während ein Filter, der ein Ausgangssignal ähnlich einer Autokorrelation produziert, eine Kammfilter Übertragungsfunktion besitzt. Man kann sagen, dass, während die aktuelle Version des MFB-Modells sich auf spektrale Verarbeitung in der Gammatonfilterbank stützen muss, um spektral aufgelöste Harmonische eines IRN-Stimulus extrahieren zu können, das AC-Modell dies auf rein zeitlicher Basis kann, mittels kammfilterartigen „Modulationsfilter“. Bislang unterscheiden sich die besprochenen Modelle noch in der Art und Weise, wie die periphere Verarbeitung implementiert ist und in ihrem Entscheidungsmodul. Letzteres wurde in dieser Arbeit bereits angeglichen. Im nächsten Schritt müsste eine einheitliche Implementierung der Peripherie stattfinden, so dass sich die beiden Modellansätze nur noch durch das jeweilige Modul unterscheiden, welches die Aktivität der Hörnervfaser auswertet. Also entweder eine Modulationsfilterbank oder ein Autokorrelator mit SACG. Damit könnte ein noch unmittelbarer Vergleich vorgenommen werden.

Zum aktuellen Zeitpunkt scheint es so zu sein, als könnten beide Modelltypen qualitativ korrekte Vorhersagen treffen. Jedoch verfolgen sie in ihrer Implementation unterschiedliche physiologische Mechanismen der Periodizitätswahrnehmung: Der eine basierend auf spektraler, der andere basierend auf zeitlicher Verarbeitung. Eine endgültige Befürwortung des einen oder anderen Modelltypus erfordert elektrophysiologische Untersuchungen der Periodizitätsdetektion in der Hörbahn.

Die nächste Verarbeitungsstation nach dem Hörnerv ist der Nucleus Cochlearis (CN). Der ventrale Teil dieses Kerns (VCN) besitzt u.a. Neurone, die aufgrund ihres Antwortcharakters als Chopper Neurone bezeichnet werden. Chopper Neurone zeigen ein repetitives Antwortmuster auf einen Stimulus. Unabhängig von ihrer charakteristischen Frequenz (CF), die das Neuron antreibt überhaupt zu antworten, hat ein Chopper Neuron eine bevorzugte intrinsische Oszillation, d.h. die repetitiven Spikes folgen in einem bestimmten Abstand voneinander. Der Abstand zwischen zwei Spikes heißt *inter spike interval*, ISI. Chopper Neurone haben also ein bevorzugtes ISI. Neuere Untersuchungen des ventralen CN (VCN) deuten darauf hin, dass Chopper Neurone eine Rolle bei der Periodizitätsdetektion spielen könnten. Winter et al. (2001) und Wiegrebe und Winter (2001) konnten in elektrophysiologischen Experimenten zeigen, dass ein Chopper Neuron sein Antwortmuster mit der Stimulusperiodizität von IRN synchronisieren kann, wenn diese im Bereich der intrinsischen Oszillationsrate liegt. Darüberhinaus konnten diese Neurone ihr Antwort-

muster auch dann synchronisieren, wenn die diese ein Vielfaches der Stimulusperiode war. Das entspricht einer Kodierung von Subharmonischen und könnte somit ein physiologisches Korrelat zu einem Autokorrelator darstellen.

Chopper Neurone sind in der Lage, neben der Trägerperiodizität, auch AM zu kodieren. In elektrophysiologischen Messungen wurden Chopper Neurone gefunden, die ein Bandpasstuning für Amplitudenmodulation zeigten, das an die Übertragungsfunktionen der Modulationsfilter erinnert (Frisina et al., 1990a,b). Auch in diesem Falle lag die CF des Neurons und die „beste Modulationsfrequenz“ weit auseinander. Hewitt et al. (1992) haben ein Computermodell von Sustained Chopper (Chop-S) Neuronen im VCN entwickelt. Die Autoren waren in der Lage mit diesem Modell Vorhersagen, mit guter Übereinstimmung zu *in vivo* gewonnenen Daten, zu produzieren.

Im Hinblick auf Periodizitätsdetektion verdienen die Chopper Neurone im VCN daher besondere Beachtung bei weiteren physiologischen Experimenten.



# Kapitel 3

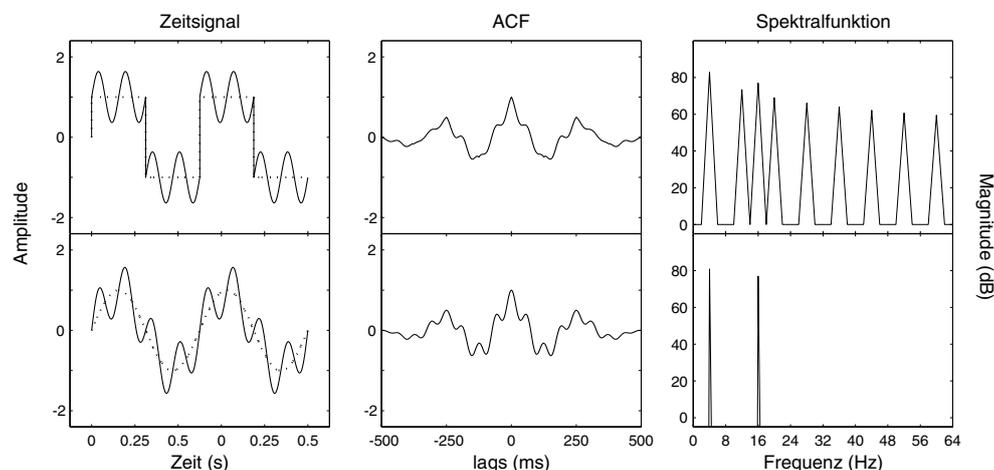
## Modulationsmaskierung mit unterschiedlichen Maskierertypen

### 3.1 Einleitung

Um der Frage auf den Grund zu gehen, ob das auditorische System bei der Kodierung der Einhüllenden eher die Strategie einer spektralen Dekomposition, wie es die Modulationsfilterbank macht, oder eine Analyse der temporalen Periodizität, im Sinne einer Autokorrelation, verfolgt, wurden von Ewert (2002) Experimente zur Modulationsmaskierung durchgeführt. Maskierungsversuche sind eine gängige Methode, um Selektivität zu untersuchen. Auf diese Weise wurde auch die Frequenzselektivität des Innenohrs psychoakustisch ermittelt.

Das Prinzip von Frequenzgruppen hat Fletcher (1940) im Audio-Bereich etabliert. Er hatte Detektionsschwellen für Sinustöne gemessen in Gegenwart eines bandpass-gefilterten Rauschens als Maskierer, welches jeweils um die zu detektierende Frequenz zentriert war. Mit steigender Bandbreite des Maskierers stiegen die Detektionsschwellen zunächst an und blieben dann unverändert. Die Bandbreite, ab der kein weiterer Anstieg der Schwelle zu beobachten ist, wird als *kritische Bandbreite* oder als *Frequenzgruppe* bezeichnet. Da mit steigender Bandbreite des Rauschens auch die Gesamtleistung des Rauschens steigt und ab einer bestimmten Bandbreite die Schwellenwerte unverändert blieben, schloss der Autor - in Anlehnung an Helmholtz (1863) - dass sich das Innenohr wie eine Reihe von Bandpassfiltern verhält. Um einen Ton zu detektieren wird der Ausgang eines Bandpassfilters mit entsprechender Mittenfrequenz evaluiert. Entscheidend für die Schwelle ist dabei das Signal/Rausch Verhältnis am Ausgang des Filters. Es kann demnach nur der Anteil des Rauschens, der in den selben Filter fällt wie das Signal, dieses maskieren. Im Audio-Bereich wurden diese Annahmen in das „*power spectrum model of masking*“ (Patterson und Moore, 1986) umgesetzt. Die Breite der Filter, also auch die Breite der Frequenzgruppen, ist abhängig von der Mittenfrequenz. In erster Näherung lässt sich sagen, dass die Bandbreite solcher Filter ca. 11 % der Mittenfrequenz beträgt.

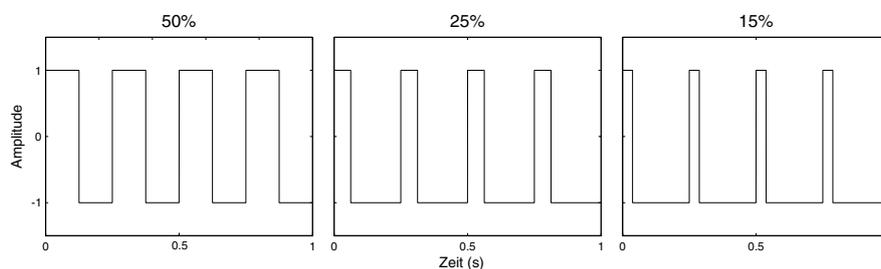
Neuere Arbeiten deuten darauf hin, dass auch die Einhüllende einem ähnlichen Verarbeitungsmechanismus unterzogen werden könnte. Dazu gehören unter anderem psychoakustische Arbeiten zur Modulationsdetektion mit schmalbandigen Rauschträgern (Dau et al., 1997a), Modulationsmaskierung mit schmalbandigem Rauschen als Maskierermodulator (Ewert und Dau, 2000; Ewert et al., 2002; Houtgast, 1989) sowie unterschiedliche physiologische Daten (z.B. von Frisina et al. (1990a,b) oder Schulze und Langner (1997)). In der Arbeit von Ewert (2002) wurde die Modulationsdetektionsschwelle einer sinusförmigen Signalmodulation bei 16 Hz in Gegenwart von zwei unterschiedlichen Maskierermodulationstypen gemessen. Beide Maskierer hatten die gleiche Frequenz von 4 Hz, unterschieden sich aber in der Zeitfunktion. Einer der Modulatoren war ein Sinus-, der andere ein Rechteckmodulator. Abbildung 3.1 vergleicht beide Maskierungssituationen miteinander. In der oberen Reihe ist die Maskierermodulation eine Rechteckfunktion, in der unteren Reihe ein Sinus. Die beiden linken Felder zeigen die Zeitfunktionen der zusammengesetzten Modulatoren, die gepunkteten Linien stellen den Maskierer alleine dar. Die mittleren Felder zeigen die Autokorrelationsfunktionen (ACF) der Einhüllenden, die rechten Felder die Einhüllendenspektren. Beide Modulatoren besitzen die gleiche zeitliche Periodizität von 250 ms (siehe ACFs) jedoch unterschiedliche Modulationsspektren. Während eine sinusförmige Amplitudenmodulation im Modulationsspektrum lediglich ein Maximum an Stelle der Modulationsfrequenz zeigt, produziert die Rechteckmodulation zusätzliche spektrale Maxima bei den ungeraden Harmonischen der Modulationsfrequenz (vgl. auch Abschnitt 1.2.1 und 1.2.2). Nimmt man an, die Einhüllende wird einer spektralen Zerlegung unterzogen, wäre zu erwarten, dass der Rechteckmodulator die sinusförmige Signalmodulation stärker maskiert als der Sinusmodulator. Eine hypothetische (Modulations-) Filterbank würde den Ausgang eines auf die Signalmodulationsfrequenz (16 Hz) abgestimmten Modulationsfilters betrachten. Die harmonischen Komponenten im Spektrum des 4-Hz Rechteckmaskierers würden folglich die Detektion der 16-Hz Signalmodulation erschweren, da sie in den Übertragungsbereich des Filters fallen. Im Vergleich würde beim Sinusmaskierer nur der Teil der 4-Hz Modulation, die den Bandpassfilter bei 16 Hz noch passiert zur Maskierung der Signalmodulation beitragen. Nimmt man dagegen an, dass lediglich die zeitliche Periodizität, unabhängig von der spezifischen Wellenform der Modulation, extrahiert wird, wäre zu erwarten, dass beide Maskierertypen die Signalmodulation in ungefähr gleichem Ausmaß maskieren. Ewert (2002) hat gezeigt, dass für einen breitbandigen Rauschträger der Rechteckmodulator ein stärkeres Maskierungsverhalten zeigt als der Sinusmodulator. Die Detektionsschwelle für die sinusförmige Signalmodulation in Gegenwart eines Rechteckmaskierers lag um ca. 3 dB über der Detektionsschwelle in Gegenwart eines Sinusmaskierers, wenn beide Maskierer die gleiche effektive Modulationstiefe besaßen. Diese Ergebnisse scheinen ein Hinweis für eine spektrale Zerlegung der Einhüllendenwellenform im auditorischen System darzustellen. Modellsimulationen dieser Versuche haben gezeigt, dass eine Verarbeitung auf Basis einer Autokorrelation die Ergebnisse nicht rekonstruieren kann; der Ansatz einer frequenzselektiven Filterbank im Modulationsbereich (EPSM, *envelope power spectrum model*, Ewert und Dau (2000)) dagegen schon (Ewert, 2002). Das EPSM wertet hierbei nur die Modulationsleistung am Ausgang des Modulationsfilters aus, der auf die Signalmodulationsfrequenz zentriert ist. Es handelt sich also um ein rein modula-



**Abbildung 3.1:** Vergleich der unterschiedlichen Maskierungssituationen. In der oberen Reihe ist der Maskierermodulator eine Rechteckfunktion, in der unteren Reihe ein Sinus. Die Maskierermodulationsfrequenz beträgt 4 Hz, die Signalmodulationsfrequenz 16 Hz. Die beiden linken Felder zeigen die Zeitfunktionen der addierten Modulatoren. Die gepunkteten Linien stellen den Maskierer alleine dar. Die mittleren Felder zeigen die Autokorrelationsfunktionen der Einhüllenden. Die rechten Felder die jeweiligen Spektren der zusammengesetzten Modulatoren. (Nach Ewert (2002))

tionsspektrales Modell, dass jede zeitliche Information, die prinzipiell am Ausgang eines Modulationsfilters noch enthalten sein kann (vgl. MFB-Modell von Dau et al. (1997a,b) aus Abschnitt 2.5.1), vernachlässigt.

Viemeister et al. (2004) haben Experimente zur Verarbeitung der Einhüllenden durchgeführt, bei denen die Rolle der zeitlichen Struktur der Signale im Vordergrund stand. Der Fokus lag allerdings nicht auf der Periodizität der Stimuli, sondern auf den lokalen zeitlichen Eigenschaften der Einhüllenden. Auch hier wurden Modulationsmaskierungsversuche mit rechteckförmigen Maskierermodulatoren und sinusförmiger Signalmodulation durchgeführt. In ihrem experimentellen Ansatz wurde systematisch die lokale zeitliche Struktur der Maskierermodulatoren verändert. Dabei wurde der Einfluss der „lokalen Modulationstiefe“ in den Tälern des Maskierers auf die Detektionsschwellen bei Modulationsmaskierung untersucht. Abbildung 3.2 stellt die Zeitfunktionen von Rechteckmodulatoren gleicher Frequenz aber unterschiedlichem Tastverhältnis dar (50%, 25% und 15%). Ein Tastverhältnis von 50% entspricht der schon bekannten Rechteckfunktion, bei der die beiden Phasen des Signals die gleiche Dauer besitzen. Mit abnehmendem Tastverhältnis verbreitert sich die untere Phase (das Tal) im Kurvenverlauf. Bei Viemeister et al. (2004) wurden die verschiedenen Rechteckwellenformen so gewählt, dass die Magnitude der Grundkomponente

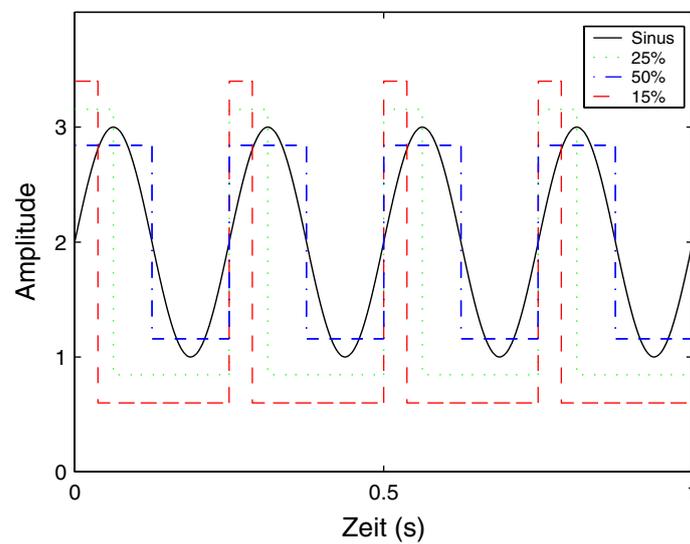


**Abbildung 3.2:** Zeitfunktionen von Rechteckmodulatoren mit gleicher Frequenz aber unterschiedlichem Tastverhältnis. Die Frequenz ist in allen Fällen 4 Hz. Mit abnehmendem Tastverhältnis verbreitert sich das Tal im Kurvenverlauf. Alle Wellenformen haben den gleichen Effektivwert.

(also diejenige die der Signalmodulationsfrequenz entspricht<sup>1</sup>), den gleichen effektiven Modulationsgrad von -9 dB für alle Tastverhältnisse aufweist. In diesem Fall, verringert sich der Wert des Maskierermodulators in der unteren Phase mit abnehmendem Tastverhältnis, wie durch die farbigen Linien in Abb 3.3 dargestellt. Die grundlegende Beobachtung in Viemeister et al. (2004) war, dass die Maskierwirkung mit abnehmendem Tastverhältnis geringer wurde. Dieses Ergebnis widerspricht der Theorie einer rein spektralen Zerlegung und Maskierung im Einhüllendenbereich. Wäre die Magnitude der Grundkomponente für die Maskierung bestimmend, hätten sich identische Schwellen für alle Maskierer gezeigt, da diese Komponente bei allen gleich war. Dies kann nur durch eine Auswertung lokaler zeitlicher Eigenschaften der Modulationswellenform erklärt werden, nämlich durch eine Auswertung der „lokalen“ Modulationstiefe während der unteren Phase (den Tälern) der Maskierermodulation. Hier besitzt die Signalmodulation eine scheinbar höhere lokale Modulationstiefe, je tiefer der Wert der Maskiererwellenform in den Tälern ist. Der Träger ist sozusagen stärker abgeschwächt. Dieses Ergebnis widerspricht der Theorie einer rein spektralen Zerlegung und Maskierung im Einhüllendenbereich.

In den Experimenten von Viemeister et al. (2004) war die Frequenz der Signalmodulation gleich der Frequenz der Maskierermodulation. In diesem Kapitel werden nun unterschiedliche Frequenzen für Signal- und Maskierermodulator verwendet. Dazu wurden die gleichen Frequenzen für Signal- und Maskierermodulation gewählt, wie in der Arbeit von Ewert (2002). Auf diese Weise soll überprüft werden, in wie weit die Ergebnisse von Ewert (2002), dessen Versuche nur mit einem Tastverhältnis von 50 % durchgeführt wurden, von der lokalen Modulationstiefe abhängen. Gleichzeitig soll getestet werden, ob sich die Beobachtungen von Viemeister et al. (2004) bestätigen, wenn die Frequenzen für Signal- und Maskierermodulation unterschiedlich sind. Mit einer Maskierermodulationsfrequenz von 4 Hz und einer Signalmodulationsfrequenz von 16 Hz, ist gewährleistet, dass beide Modu-

<sup>1</sup>Eine Rechteckfunktion enthält im Leistungsspektrum mehrere Frequenzen (siehe auch 1.8). Bei den von Viemeister et al. (2004) gewählten Tastverhältnissen, nehmen die Komponenten mit abnehmendem Tastverhältnis zu.



**Abbildung 3.3:** Wellenformen der von Viemeister et al. (2004) verwendeten Modulatoren. Die Frequenz ist in allen Fällen 4 Hz, die Tastverhältnisse sind der Legende zu entnehmen. Die Magnitude der Grundkomponente (4 Hz) ist für alle Maskierer gleich. Dadurch verringert sich, mit abnehmendem Tastverhältnis, der Wert des Maskierers in der unteren Phase der Zeitfunktion.

lationen in unterschiedliche hypothetische Modulationsfilter fallen. Durch das harmonische Verhältnis würden die zusätzlichen spektralen Komponenten des Rechteckmaskierers die Signalmodulation verstärkt maskieren (gegenüber einem sinusförmigen Maskierer). Diese Messreihe wird, im Vergleich zur Arbeit von Viemeister et al. (2004), um einen Rechteckmodulator mit 85%igem Tastverhältnis erweitert. Die Rechteckfunktionen mit den Tastverhältnissen 15 und 85% sind invers zueinander. Daher besitzen beide das gleiche Leistungsspektrum und sollten bei einer rein modulationsspektralen Analyse (wie im EPSM, Ewert und Dau (2000)) zu identischen Maskierwirkungen führen. Um die Abhängigkeit von der Modulationsfrequenz zu überprüfen, wurde das Experiment mit den Maskierer-/Signalmodulationsfrequenzen von 32 und 128 Hz wiederholt.

## 3.2 Methoden

### 3.2.1 Paradigma

Die Detektionsschwellen für die Amplitudenmodulation wurden in einem adaptiven 3AFC Verfahren ermittelt. Dies ist analog zu dem 2AFC Paradigma, welches in Abschnitt 2.2.1 erläutert wurde. Es unterscheidet sich nur dadurch, dass in diesem Verfahren jeweils drei statt zwei Intervalle präsentiert wurden. Jedes der drei Intervalle enthielt die Maskierermodulation. Eines enthielt zusätzlich die Signalmodulation. Dieses galt es zu erkennen. Der adaptive Parameter war die Modulationstiefe der Signalmodulation. Er wurde nach einer *three-down one-up* Regel variiert (siehe Abschnitt 2.2.1). Für eine Schwellenmessung wurden elf Umkehrpunkte aufgenommen. An den ersten drei Umkehrpunkten wurde die Modulationstiefe um 5 dB verändert, dann um 3 dB und an den letzten sechs Umkehrpunkten wurde die Modulationstiefe nur noch um 1 dB verändert. Der Mittelwert der letzten sechs Umkehrpunkte ergab die Detektionsschwelle (79,4% korrekt auf der psychometrischen Funktion) für einen Durchgang (Levitt, 1971). Als Schwelle pro Proband wurde der Mittelwert von drei Messungen für eine bestimmte Signalkonfiguration gewählt. Der Proband erhielt nach jeder Entscheidung ein visuelles Feedback über die Richtigkeit seiner Antwort.

### 3.2.2 Probanden

An den Versuchen nahmen 4 Personen im Alter zwischen 25 und 35 Jahren teil. Die Teilnahme erfolgte freiwillig und unentgeltlich. Alle Teilnehmer verfügen über ein klinisch normales Hörvermögen.

### 3.2.3 Stimuli und Versuchsapparatur

Träger der Amplitudenmodulation war stets GN. Die Frequenz der Maskierermodulation lag zunächst bei 4 Hz und die zu detektierende Signalmodulation entsprechend einem Verhältnis von 1:4 bei 16 Hz. In der Wiederholung wurden die Modulationsfrequenzen auf

32 und 128 Hz erhöht. Die Signalmodulation war immer sinusförmig. Als Maskierermodulatoren wurden ein Sinus sowie Rechteckfunktionen mit einem Tastverhältnis von 85, 50 oder 15% getestet. Der Effektivwert des Maskierers wurde jeweils auf -9 dB festgelegt. Die Modulationstiefe des Signalmodulators wurde nach dem adaptiven Verfahren variiert. Für das Signalintervall wurden der maskierende Modulator und der Signalmodulator addiert:

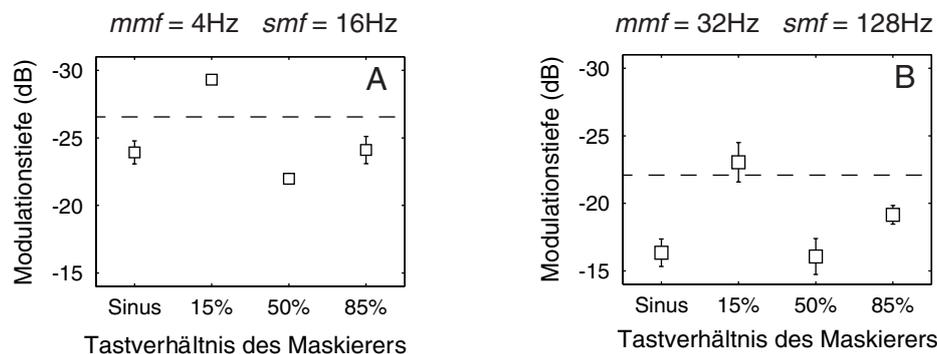
$$x(t) = \{[m(t) + md \times m_s(t)] + 1\}n(t) \quad (3.1)$$

In dieser Gleichung ist  $x(t)$  der zu präsentierende Stimulus,  $m(t)$  der Maskierer und  $m_s(t)$  die Signalmodulation. Die Modulationstiefe ( $md$ ) ist in dieser Gleichung ein linearer Wert. Die Modulationstiefe in dB entspricht  $20 \times \log md$ . Der zusammengesetzte Modulator wird mit dem Träger  $n(t)$  multipliziert. Die Stimulusdauer betrug 500 ms inklusive kosinusförmiger An- und Abstiegsflanken von je 20 ms. Die Abtastrate lag bei 50 kHz. Die Stimuli wurden auf einem IBM-kompatiblen PC mit MatLab generiert und anschließend auf einen Tucker-Davis-Technologies System 3 RP 2 *Enhanced Real-Time Processor* übergeben. Die analoge Ausgabe der Stimuli erfolgte über den integrierten D/A-Wandler und einen Tucker-Davis-Technologies System 3 HB 7 Kopfhörerverstärker, dessen Endabschwächung auf -12 dB eingestellt war. Der mittlere Ausgabepegel betrug 64 dB SPL. Es wurde eine Pegelrandomisierung von  $\pm 6$  dB SPL verwendet um auszuschließen, dass Versuchspersonen systematisch auftretende Pegelunterschiede durch die Signalmodulation als Entscheidungskriterium heranziehen können. Die Ausgabe der Signale erfolgte diotisch über AKG K240DF Kopfhörer. Die Probanden wurden einzeln in einer reflexionsarmen und schallisolierten Hörkabine getestet. Die Antworten wurden über ein graphisches Interface auf einem Touchscreen abgegeben.

### 3.3 Ergebnisse

Die Ergebnisse dieser Experimente sind in Abb. 3.4 grafisch dargestellt. Das linke Feld (**A**) zeigt die Ergebnisse für die Modulationsfrequenzen 4/16 Hz, das rechte Feld (**B**) die Ergebnisse für 32/128 Hz. Die einzelnen Quadrate zeigen die Modulationstiefe der Signalmodulation in dB an deren Detektionsschwelle an. Jedes Symbol repräsentiert dabei einen anderen Maskierermodulator. Dargestellt sind die Mittelwerte der Ergebnisse aller vier Probanden. Die Fehlerbalken stellen den Standardfehler zwischen den Versuchspersonen dar. Die gestrichelte Linie zeigt die Detektionsschwelle für die Signalmodulation alleine an. Die Werte auf der Y-Achse sind in umgekehrter Reihenfolge aufgetragen. Das bedeutet, dass eine niedrigere Schwelle entlang der Ordinate „höher“ angesiedelt ist (entsprechend einer besseren Detektionsleistung).

Für die Signal/Maskierer-Kombination von 16 und 4 Hz (Abb. 3.4 A) sind die Schwellen für den Sinusmaskierer und die 85% Rechteckfunktion in etwa gleich (-23,9 und -24,1 dB). Die Schwelle für den 50%-Maskierer liegt etwas darüber (-22 dB). Für beide Signal/Maskierer-Kombinationen liegt die Schwelle in Gegenwart des 15%-Maskierers unter der Referenzschwelle (Signalmodulation alleine), dargestellt durch die gestrichelte Linie. Bei 128 Hz ist dies sehr viel schwächer ausgeprägt. Bei einer Signalmodulation von 128 Hz



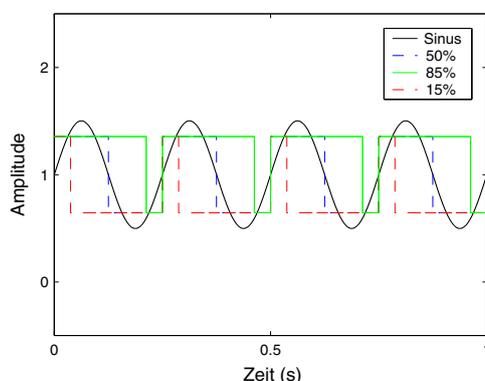
**Abbildung 3.4:** Ergebnis der Modulationsmaskierung mit Maskierern unterschiedlicher Tastverhältnisse. Entlang der Ordinate ist die Modulationstiefe der Signalmodulation an der Detektionsschwelle aufgetragen. Entlang der Abzisse die unterschiedlichen Maskierertypen. Die Prozent-Angaben beziehen sich auf das Tastverhältnis des maskierenden Rechteckmodulators. In **A**, betrug die Maskierermodulationsfrequenz ( $mmf$ ) 4 Hz, während die Signalmodulationsfrequenz ( $smf$ ) 16 Hz betrug. In **B**, lagen die Modulationsfrequenzen bei 32 Hz für die Maskierer und 128 Hz für die Signalmodulation. Es handelt sich hierbei um die Mittelwerte der Ergebnisse der einzelnen Probanden. Die Fehlerbalken stellen den Standardfehler zwischen den Probanden dar. Die gestrichelten Linien stellen Detektionsschwelle für die Signalmodulation ohne Maskierer dar.

und 32 Hz Maskierermodulation sind die Schwellen insgesamt niedriger. In dieser Situation sind die Schwellen für den Sinus- und den 50%-Maskierer nahezu identisch (-16,3 und -16,0 dB). Die Maskierwirkung des 85%-Modulators liegt etwas darunter (-19,2 dB). Generell ist die Detektionsleistung in Anwesenheit des Rechteckmaskierers mit einem Tastverhältnis von 15 % immer am besten.

### 3.4 Diskussion

Vergleicht man die vorliegenden Daten mit den eingangs zitierten Arbeiten (Ewert, 2002; Viemeister et al., 2004) lässt sich folgendes feststellen: Analog zu der Arbeit von Ewert (2002), maskiert auch hier die Rechteckfunktion mit einem Tastverhältnis von 50 % stärker als der Sinusmodulator, wenn man die vergleichbare 4/16-Hz Kombination betrachtet. Da die Maskierer die gleiche effektive Modulationstiefe (-9 dB) besaßen, kann dieses Ergebnis nicht mit Unterschieden im Leistungsspektrum der Einhüllenden erklärt werden, wenn man eine Selektivität für Einhüllendenfrequenz außer Acht lässt. Denn bei einer solchen Verarbeitung wird der Ausgang der Filters betrachtet, der auf die Signalmodulation abgestimmt ist. Das aktuelle Experiment zeigt allerdings, dass die Rechteckfunktionen mit den Tastverhältnissen 15 und 85 % ebenfalls in unterschiedlichem Ausmaß maskieren, nämlich mit einer Differenz von 5 dB. Da diese beiden Funktionen invers zueinander sind und somit ein identisches Modulationsleistungsspektrum besitzen, dürften sich nach der Theorie der rein spektralen Zerlegung der Einhüllenden keine Unterschiede im Maskierungsgrad ergeben.

Da die Maskierer in diesen Versuchen alle den gleichen Effektivwert besaßen, ergaben sich auch keine Unterschiede in der lokalen Modulationstiefe der Maskiererwellenformen



**Abbildung 3.5:** Zeitfunktionen der verwendeten Modulatoren. Der Effektivwert der Signale ist für alle Funktionen gleich, daher ändert sich auch die lokale Modulationstiefe nicht (Die Minimal- und die Maximalamplitude der Rechteckfunktionen sind gleich im Gegensatz zu 3.3).

(siehe Abb. 3.5). Es hätte also auch im Sinne einer Auswertung der *lokalen* Modulationstiefe in den Tälern, wie von Viemeister et al. (2004) vorgeschlagen, kein Unterschied zwischen den Modulatoren mit den Tastverhältnissen 15 und 85 % auftreten dürfen. Denn auch die maximale lokale Modulationstiefe ist, aufgrund des inversen Verhältnisses beider Funktionen, in jedem Fall gleich.

Die zeitliche Struktur bzw. der zeitliche Verlauf der Maskiererwellenform ist offensichtlich von Bedeutung. Es erscheint plausibel, dass die Länge der Täler eine Rolle spielt. Was nicht ganz zu dieser Überlegung passt, ist die Tatsache, dass eine Rechteckfunktion mit einem Tastverhältnis von 85 % etwas weniger maskiert als mit einem Tastverhältnis von 50 %. Hierzu muss erwähnt werden, dass die Daten für den 85%-Maskierer später im Verlauf der Experimente erhoben wurden und ein Trainingseffekt nicht auszuschließen ist.

In den Ergebnissen der 32/128 Hz Kombination (siehe Abb. 3.4 B) zeigt sich dieses Verhältnis zwischen den Schwellen für die Maskierer mit den Tastverhältnissen 50 und 85 % wieder. Dennoch widerspricht diese Tatsache der Überlegung, dass die Breite der Täler eine Rolle spielt, nicht per se. In Abschnitt 2.4 wurde erwähnt dass für Perioden ab ca 30 Hz eine Tonhöhe wahrgenommen wird (Krumbholz et al., 2000; Pressnitzer et al., 2001). Diese ist bei 32 Hz noch recht schwach ausgeprägt, aber eine Modulationsfrequenz von 128 Hz produziert bereits eine deutliche Tonhöhenempfindung. Insofern ist es plausibel, dass die Tonhöhe als Entscheidungskriterium von den Probanden herangezogen wurde (persönliche Befragung nach Abschluss der Messreihe bestätigte diese Annahme). Nimmt man an, dass bei einer Modulationsfrequenz von 32 Hz die zeitlichen Lautstärkeschwankungen gerade noch als solche wahrgenommen werden, oder als Rauigkeit des Signals, während eine Modulationsfrequenz von 128 Hz eine deutliche Tonhöhenwahrnehmung hervorruft, könnte es sein, dass diese beiden Signaleigenschaften als getrennte auditorische Objekte gewertet

werden. Dieses Perzept der Tonhöhe ist kontinuierlicher Natur und könnte daher vor einem ebenfalls kontinuierlichen Hintergrund leichter herauszuhören sein. Die beiden Rechteckimpulse mit 15 und 85 % Tastverhältnis bieten einen relativ konstanten Hintergrund. Sie wirken wie ein kontinuierliches Signal, das ca. alle 30 ms kurz unterbrochen wird (nur 15 % Berg bzw. Tal). Der Sinus- und der 50 % Maskierer stellen mit ihren ausgeprägteren Veränderungen entlang der Zeitachse einen „unruhigeren“ Hintergrund dar, der womöglich ein effektiveres Störsignal darstellt. Eventuell könnte das auch die Diskrepanz zwischen der Maskierwirkung der 50 und 85 % Rechteckfunktionen in der 4/16 Hz Kombination erklären. Um das Konzept der Bindung von Informationen (*informational binding*) in diesem Zusammenhang zu klären, wäre es hilfreich die Messreihen auf weitere Tastverhältnisse auszuweiten. Rechteckfunktionen mit den Tastverhältnissen von 25 und 75 % z.B. sind lohnenswert, da sie ebenfalls invers zueinander sind, also identische Frequenzspektren und gleiche lokale Modulationstiefen besitzen und in ihrem Ausmaß der Veränderlichkeit entlang der Zeitachse genau zwischen den hier getesteten Tastverhältnissen liegen. In jedem Fall machen die vorliegenden Ergebnisse deutlich, dass der zeitliche Verlauf der Maskierer eine wichtige Rolle beim Grad der Verdeckung innehaben.

# Kapitel 4

## Modulationsdetektion mit aperiodischen Modulatoren

### 4.1 Einleitung

In Abschnitt 1.2.3 wurde der Stimulus *Sparse Noise* (SN) bereits kurz eingeführt. Zur Wiederholung: SN kann am besten beschrieben werden, als ein weisses Rauschen, das über den Zeitverlauf immer wieder von Lücken variabler Länge unterbrochen ist. Je höher die maximale Lückendauer, desto stärker ist der Fluktuationsgrad der Einhüllenden und damit auch die Unruhe des Signals. Die maximale Lückendauer errechnet sich nach der Formel 1.1.

Ursprünglich wurde SN entwickelt um, bei der Fledermaus *Megaderma lyra*, Verhaltensversuche zur Detektion von Raschelgeräuschen durchzuführen (Hübner und Wiegrebe, 2003). *M. lyra* gehört zu den sogenannten *gleaning bats*, die sich dadurch auszeichnen, dass sie ihre Beute überwiegend von einem Untergrund oder der Vegetation aufgreifen (Neuweiler, 1993). Insofern sind beutegenerierte Raschelgeräusche, wie z.B. das Laufen einer Maus über getrocknetes Laub, notwendig um ein Jagdverhalten zu initiieren. Außerdem müssen diese Tiere in der Lage sein die beutegenerierten Geräusche von Störgeräuschen in der Umgebung zu differenzieren. Da diese Raschelgeräusche und deren natürliche Maskierer, wie z.B. durch Wind bewegtes Laub, sich spektral sehr ähneln, könnte die zeitliche Struktur von entscheidender Bedeutung für die Erkennung der Beutegeräusche sein. Mit SN hatten die Autoren einen rauschhaften Maskierer an der Hand, dessen zeitliche Struktur sich selektiv verändern lässt, durch die Veränderungen der zeitlichen Lücken. Es zeigte sich, dass die Detektion der Raschelgeräusche (eine Aufnahme von sich bewegenden Mehlwürmern, Larven des *Tenebrio molitor*), sich monoton verbesserte, wenn die zeitlichen Lücken des Maskierers eine Länge von ca. 0,3 ms überschritten.

Statt zeitliche Lücken in einen Abschnitt Gauss'schem Rauschens (GN) einzuführen, kann SN auch dadurch erzeugt werden, indem man GN mit einem aperiodischen, transienten Modulator multipliziert (siehe 1.2.3). Im Gegensatz zu den periodischen Sinusmodulatoren oder Rechteckfunktionen, deren Modulationsspektren eine harmonische Struktur

aufweisen, ist das Modulationsspektrum von SN, wie der Modulator selbst, aperiodisch. Unabhängig von der Lückendauer ergibt sich für jeden SN-Stimulus ein weisses Modulationsspektrum. Insofern stellt ein solcher Modulator keinen kritischen Test für psychoakustische Modelle, deren Verarbeitung der Einhüllendenstruktur auf einer selektiven Filterbank beruht, wie die Modulationsfilterbank von Dau et al. (1997a,b) oder das EPSM von Ewert und Dau (2000), dar. Dennoch ist es interessant wie solche aperiodischen Modulationen im auditorischen System des Menschen verarbeitet werden. Denn mit der Lückendauer verändert sich auch der Fluktuationsgrad der Einhüllenden. Das *vierte Moment* (Hartmann und Pumplin, 1988) quantifiziert den Grad an Einhüllendenfluktuation und wird nach folgender Formel berechnet:

$$M_4 = \frac{\frac{1}{T} \int_0^T x^4(t) dt}{\left(\frac{1}{T} \int_0^T x^2(t) dt\right)^2} \quad (4.1)$$

Dabei ist  $x(t)$  das Zeitsignal und  $T$  die Signallänge. Das vierte Moment steigt monoton mit der maximalen Lückendauer des SN.

Da bislang noch keine Versuche mit solchen aperiodischen Modulatoren beim Menschen durchgeführt wurden, dienen die Hörversuche, die in diesem Kapitel beschrieben werden, dazu die Wahrnehmung von SN zu charakterisieren. Es wurden psychometrische Funktionen ermittelt zur Detektion von SN für verschiedene Stimulusbandbreiten mit unterschiedlichen unteren Eckfrequenzen.

## 4.2 Methoden

### 4.2.1 Paradigma

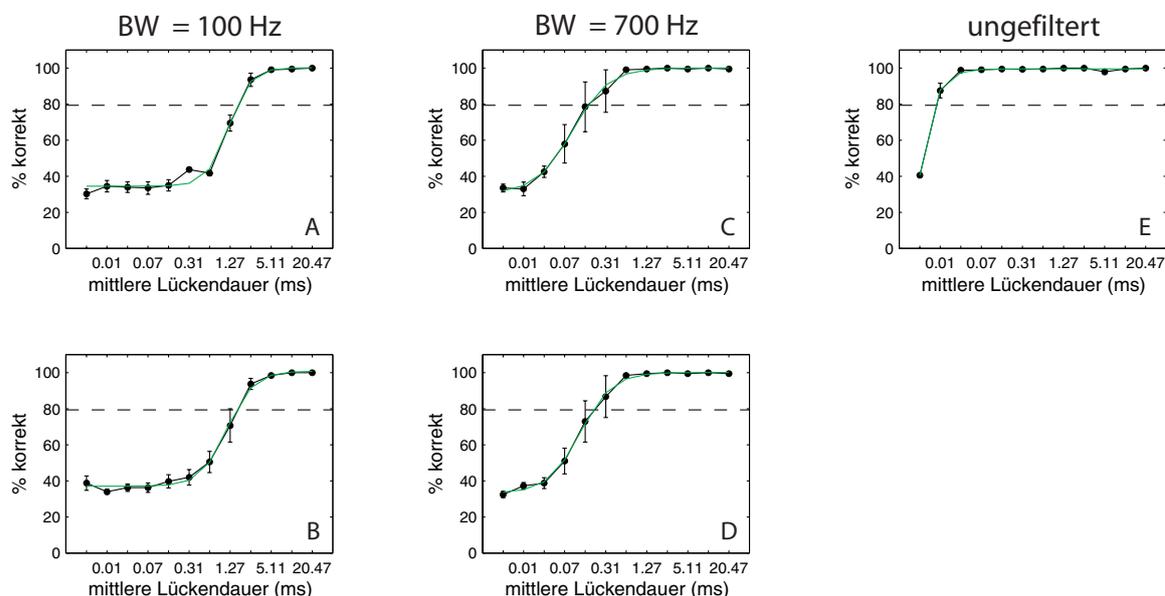
In diesem Experiment galt es das SN von GN zu unterscheiden. Dazu wurde ein *three-alternative forced choice* (3AFC) Verfahren angewandt. Wie in den Experimenten aus Kapitel 3, wurden auch hier drei Intervalle vorgespielt. In zwei Intervallen wurde GN präsentiert, in einem das SN. Die Aufgabe des Probanden lag darin, das Intervall zu erkennen, welches das SN beinhaltete. Anderes als in den vorangegangenen Experimenten wurde hier kein adaptives Verfahren angewendet. Stattdessen wurde eine psychometrische Funktion ermittelt. Das bedeutet, dass pro Signalkonfiguration eine bestimmte Anzahl von Einzelentscheidungen aufgenommen wurde und anschließend der Prozentsatz der richtigen Antworten errechnet wurde. Es wurden zwölf Punkte entlang der psychometrischen Funktion gemessen, für 12 verschiedene maximale Lückendauern. In einer Sitzung wurden 22 Entscheidungen pro max. Lückendauer aufgenommen. Das waren insgesamt 264 Entscheidungen pro Sitzung. Die Präsentation erfolgte in randomisierter Reihenfolge. Pro Person wurden 3 Sitzungen absolviert. Daraus ergeben sich 66 Einzelentscheidungen für jede max. Lückendauer. Die Detektionsschwelle für SN wurde, in Anlehnung an adaptive Verfahren,

auf 79,4 % richtige Antworten festgelegt. Um diesen Wert zu ermitteln wurde eine Fitfunktion auf die Daten angewandt und deren Schnittpunkt mit der 79,4 % Marke ermittelt.

### 4.2.2 Stimuli und Versuchsausrüstung

Die relevanten Stimuli in diesem Experiment sind GN und SN. Die Stimuli wurden, vor jeder Präsentation neu, auf einem IBM-kompatiblen PC generiert. Für das SN wurde GN mit einem entsprechendem Modulator multipliziert. Um den Modulator zu erzeugen wird entlang einer Zeitachse, die sich aus der Abtastrate und der Stimulusdauer ergibt, an beliebigen Abtastpunkten, die Amplitude gleich eins gesetzt, an den übrigen Punkten bleibt die Amplitude gleich Null. Der Abstand zwischen zwei positiven Amplitudenwerten ist ebenfalls beliebig, längstens jedoch  $2^{lep} - 1$  Abtastpunkte. Die Abtastrate lag bei 50 kHz. Daraus ergab sich eine max. Lückendauer von 0 bis 40,94 ms, wobei sich die Schrittweite von einem Messpunkt zu nächsten jedes Mal verdoppelt. Da die Lücken einen beliebigen Wert zwischen 0 ms und der Maximaldauer annehmen können, ist sie im Mittel 1/2 mal der max. Lückendauer. Die Stimulusdauer betrug 1 s inklusive kosinusförmiger An- und Abstiegsflanken von je 20 ms Länge. Der digitale Effektivwert war für alle Signale gleich. Die Stimuli wurden bandpass-gefiltert mit einem Butterworth-Filter 4. Ordnung, d.h. mit einer Flankensteilheit von 24 dB/Oktave. Es gab vier unterschiedliche Filterkonditionen. Die Eckfrequenzen betragen: (1) 1000-1100 Hz (2) 7000-7700 Hz (3) 1000-1700 Hz (4) 7000-7100 Hz. Somit ergeben sich zwei unterschiedliche Bandbreiten mit jeweils unterschiedlichen Eckfrequenzen. Die Bandbreite auditorischer Filter beträgt ca. 11%. Dieser Bandbreite entsprechen in etwa die Konditionen (1) und (2). Kondition (3) ist wesentlich breiter als ein auditorischer Filter in diesem Bereich (700 statt ca. 110 Hz) und Kondition (4) entsprechend wesentlich schmaler (100 statt ca. 770 Hz). Anschließend wurden noch Diskriminierungsschwellen für eine fünfte Bedingung (5) ohne jegliche Bandpassfilterung gemessen.

Das digitale Signal wurde auf einen Tucker-Davis-Technologies System 3 RP 2 *Real-Time Processor* übergeben. Die analoge Ausgabe der Stimuli erfolgte über den integrierten D/A-Wandler und einen Tucker-Davis-Technologies System 3 HB 7 Kopfhörerverstärker, dessen Endabschwächung auf -18 dB eingestellt war. Für die Signale ohne Bandpassfilterung mussten die Amplituden des digitalen Signals reduziert werden, da sonst die maximale Ausgangsspannung des D/A Wandlers überschritten worden wären, was eine Übersteuerung des Signals zur Folge gehabt hätte. Dies wurde über die Einstellung der Endabschwächung am Kopfhörerverstärker ausgeglichen, so dass sämtliche Signale mit dem gleichen Pegel präsentiert wurden. Die Ausgabe der Signale erfolgte über AKG K240 DF Kopfhörer mit einem mittlerem Pegel von 64 dB SPL  $\pm$  6 dB. Die Pegelrandomisierung diente auch in diesem Experiment dazu, dass eventuell systematisch auftretende Pegelunterschiede als Entscheidungskriterium von der Versuchsperson nicht herangezogen werden können.



**Abbildung 4.1:** Ergebnisse der SN Detektion in Abhängigkeit von der Bandbreite (BW) und der Eckfrequenzen der gefilterten Signale. **A.** 1000-1100 Hz **B.** 7000-7100 Hz **C.** 1000-1700 Hz **D.** 7000-7700 Hz **E.** ungefiltert. Die gestrichelte Linie repräsentiert die 79,4% Marke. Die Kurve der Fitfunktion ist in grau eingezeichnet. Die mittlere Lückendauer, ab der bandpassgefiltertes SN von GN signifikant unterschieden werden kann, ist bei höherer Bandbreite des Filters geringer (vgl. A/C und B/D), unabhängig von den Eckfrequenzen. Ungefiltertes SN, kann selbst bei der geringsten möglichen mittleren Lückendauer noch detektiert werden (siehe E).

### 4.2.3 Probanden

An diesen Experimenten nahmen vier Versuchspersonen im Alter von 22 bis 31 Jahren teil. Die Versuchspersonen bekamen einen Stundenlohn als Vergütung. Alle Teilnehmer verfügten über ein klinisch normales Hörvermögen.

## 4.3 Ergebnisse

Die Ergebnisse der SN Detektion sind in Abb. 4.1 zu sehen. Die gestrichelten Linien markieren den Wert von 79,4% korrekte antworten. In Kurve der Fitfunktion ist grau eingezeichnet. Die beiden linken Felder zeigen die Ergebnisse, die mit einem 100 Hz breitem Signal erzielt wurden, die mittleren Felder die Ergebnisse für 700 Hz breite Signale. In den oberen Feldern war die untere Eckfrequenz jeweils 1000 Hz, in Unteren lag sie bei 7000 Hz. Im rechten Feld war das Signal ungefiltert hat somit eine Bandbreite von >20 kHz. Die Fä-

higkeit SN von GN zu unterscheiden verbessert sich mit steigender Bandbreite des Signals, unabhängig von den Eckfrequenzen. Daher wurden die Werte an den Schnittpunkten mit der 79,44 % Marke für die gleiche Bandbreite gemittelt. Während bei einer Stimulusbandbreite von 100 Hz die Schwelle bei einer mittleren Lückendauer von 1,7 ms liegt, beträgt sie bei einer Bandbreite von 700 Hz nur noch 0,2 ms. Steigt die Bandbreite auf  $>20$  kHz, reicht eine mittlere Lückendauer von 0,02 ms aus um SN von GN unterscheiden.

## 4.4 Diskussion

Die Ergebnisse zeigen, dass die mittlere Lückendauer, ab der SN von GN diskriminiert werden kann, mit steigender Bandbreite des Signals abnimmt. Das ist prinzipiell im Einklang mit Theorien zur Lückendetektion von schmalbandigem Rauschen. Wenn ein auditorischer Filter breiter ist, als die Bandbreite des Rauschens, wird das zeitliche Muster der Fluktuationen durch den Filter kaum verändert. Das Nachschwingen der Filter füllt die Lücke nur zum Teil. Die Schwelle für die Lückendetektion hängt im wesentlichen davon ab, ob die Lücke mit den inherenten Fluktuationen des Rauschens verwechselt werden kann (Moore, 1997). Je höher die Bandbreite eines Stimulus ist, desto schneller sind die Fluktuationen im Zeitsignal. Das würde bedeuten, dass mit steigender Bandbreite des Rauschens, die Lückendetektionsschwellen sinken. Vergleicht man in Abb. 4.1 die Felder B und D mit den Eckfrequenzen 7000 - 7100 Hz (B) und 7000 - 7700 Hz (D), lässt sich diese Theorie wiedererkennen. Da die Bandbreite eines auditorischen Filters ca. 11 % der Mittenfrequenz entspricht, kann davon ausgegangen werden, dass für diesen Bereich die Signalbandbreite geringer ist. Ist die Bandbreite des Rauschens größer als die des auditorischen Filters, sind die Fluktuationen am Ausgang des Filters geringer als am Eingang. Das hätte eine Beeinträchtigung der Lückendetektion zur Folge und würde bedeuten, dass die Bandbreite des auditorischen Filters die Schwellen der Lückendetektion limitiert, würde nur der Ausgang eines einzelnen Filters betrachtet werden. Allerdings konnten verschiedene psychoakustische Arbeiten zeigen, dass die Lückendetektionsschwellen auch dann noch mit steigender Signalbandbreite sinken, wenn die Signalbandbreite die Bandbreite des auditorischen Filters übersteigt (Eddins et al., 1992; Shailer und Moore, 1983, 1985). Daraus kann man schlussfolgern, dass das auditorische System, bei der Auswertung von zeitlichen Lücken in einem Signal, über mehrere auditorische Filter integriert. Denn diese zeitliche Lücke wird synchron am Ausgang aller, durch das Signal angeregten, Filter abgebildet (Moore, 1997). Auch dieser Effekt ist in den Ergebnissen dieses Experiments zu beobachten (vgl. Abb. 4.1 A und B). Sind die Eckfrequenzen 1000 - 1100 Hz (A), ist die Bandbreite des Signals  $\leq$  der Bandbreite des auditorischen Filters. Sind die Eckfrequenzen 1000 - 1700 Hz (B), ist die Bandbreite des Signals  $\geq$  der Bandbreite der auditorischen Filters. Die Ergebnisse zeigen für A eine höhere Schwelle als für B (1,7 ms im Vergleich zu 0,2 ms).

In der Arbeit von Eddins et al. (1992) wird ein Vergleich angestellt von eigenen Daten zur Lückendetektion und denen aus anderen Veröffentlichungen (Buus und Florentine, 1985; Shailer und Moore, 1983, 1985). Es wird festgestellt, dass in allen Arbeiten die Lückendetektionsschwellen mit steigender Stimulusbandbreite sinken und dass dieser

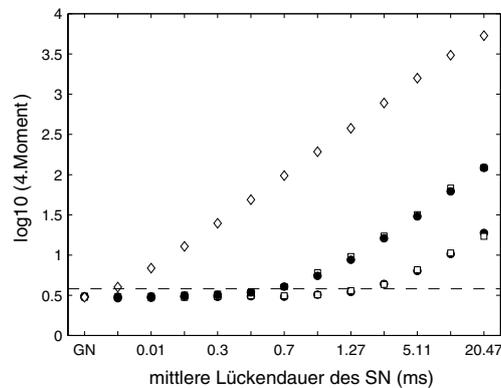
Effekt unabhängig ist von der Mittenfrequenz des Stimulus. Diese Beobachtung ist im Einklang mit den vorliegenden Ergebnissen zur SN-Diskriminierung. Ein wesentlicher Unterschied zwischen den Ergebnissen der genannten Arbeiten und denen des vorliegenden Experiments ist, dass die Schwellen, der SN-Diskriminierung deutlich geringer sind (vgl. z.B. Eddins et al. (1992):  $>20$  ms bei  $BW=100$  Hz, hier 1,7 ms; ca. 10 ms bei  $BW=800$  Hz, hier 0,2 ms bei  $BW=700$  Hz). In einer Arbeit von Snell et al. (1994) wurden bei einer Signalbandbreite von 12 kHz immerhin Schwellen von 2,22 ms gemessen. Allerdings wurden in den zitierten Arbeiten nur einzelne Lücken in ein Stück Rauschen eingefügt. SN dagegen enthält zahlreiche Lücken von unterschiedlicher, nach oben begrenzter Dauer. Es wäre denkbar, dass ein Redundanzphänomen auftritt. Dass also von zahlreichen Lücken ausreichend viele detektiert werden, um SN von GN zu unterscheiden. Ähnlich dem Prinzip, dass über mehrere auditorische Filter Information integriert wird. Viemeister und Wakefield (1991) haben ein Modell der „multiple looks“ vorgestellt, dass quasi eine Reihenschaltung von mehreren Beobachtungszeiträumen mit jeweils sehr kurzen Integrationszeiten (3 ms) vorschlägt. Diese „Einzelbeobachtungen“ werden dann in einem Kurzzeit-Speicher gesammelt, der eine eigene Abklingcharakteristik besitzt sowie eine eigene Integrationszeit, die natürlich wesentlich länger ist als die der einzelnen Beobachtungszeiträume. Dieser Speicher kann als eine Art Vektor der Einzelbeobachtungen betrachtet werden. Innerhalb jedes einzelnen Beobachtungszeitraum besteht eine gewisse Wahrscheinlichkeit dass das Signal die Detektionsschwelle überschreitet. Verlängert sich das Signal, erhöht sich die Anzahl der Einzelbeobachtungen und somit auch die Wahrscheinlichkeit dass in einem dieser Zeiträume das Signal überschwellig ist (Viemeister und Wakefield, 1991). Die Signaldauer der SN-Detektion betrug eine Sekunde. Die hohe Anzahl der Lücken (im Vergleich zur klassischen Lückendetektion) gepaart mit einem Mechanismus der im Stil des „multiple-looks-Modells“ funktioniert, bieten eine gute Basis dafür, dass für breitbandige Signale ein SN mit einer mittleren Lückendauer von nur 0,01 ms gegenüber GN detektiert werden kann.

Wenn man die Wahrnehmungsgröße, die SN hervorruft, qualitativ beschreiben möchte, ist dies mit dem Begriff „Rauigkeit“ am besten getan. Rauigkeit ist eine Wahrnehmungsgröße die Schallschwankungen anzeigt (Terhardt, 1998). Die Rauigkeit der Signale war das Entscheidungskriterium, welches von allen Versuchspersonen, laut Befragung, überwiegend herangezogen wurde. Bei sehr langen Lückendauern kann die Wahrnehmung in ein „Knistern“ übergehen. Diese Grenze hängt von der Bandbreite des Signals ab. Man muss an dieser Stelle deutlich machen, dass der Begriff „Rauigkeit“ quasi doppelt belegt ist. Einmal als physikalischer Parameter, das andere Mal als psychoakustischer Parameter, also die Wahrnehmungsgröße. So wie die *Lautstärke*, angegeben in dB SPL, einen physikalischen Parameter darstellt und die *Lautheit*, angegeben in sone, die psychoakustische Empfindungsgröße bezeichnet, gibt es für die Rauigkeit nur diesen einen Begriff.

Mit Hilfe des vierten Moments (Hartmann und Pumplin, 1988), welches sich nach der Formel 4.1 errechnet, lässt sich die Rauigkeit, im physikalischen Sinne, quantifizieren. In Anlehnung an Grunwald et al. (2004) wurde der dekadische Logarithmus des vierten Moments herangezogen. In Tabelle 4.2 sind diese Werte für die unterschiedlichen Signale angegeben. Sie sind nach Filterbedingung in unterschiedliche Spalten aufgeteilt. Die Lückendauer nimmt von oben nach unten zu. Der Bereich, unterhalb der jeweiligen Detektions-

Mitt. LD (ms)	1000-1100	7000-7100	1000-1700	7000-7700	ungefiltert
GN	0.4823	0.4899	0.4814	0.4760	0.4761
0	0.4646	0.4744	0.4824	0.4787	0.6021
0,01	0.4698	0.4728	0.4832	0.4881	0.8384
0,03	0.4792	0.4754	0.4904	0.5035	1.1070
0,07	0.4803	0.4827	0.5014	0.5159	1.3943
0,15	0.4880	0.4944	0.5341	0.5428	1.6887
0,31	0.4823	0.4949	0.6092	0.6063	1.9878
0,63	0.5098	0.5055	0.7437	0.7814	2.2850
1,27	0.5425	0.5579	0.9425	0.9813	2.5764
2,55	0.6431	0.6347	1.2094	1.2363	2.8904
5,11	0.8014	0.8188	1.4824	1.5036	3.1992
10,23	1.0144	1.2069	1.7915	1.8327	3.4841
20,47	1.2764	1.2335	2.0847	2.0886	3.7274

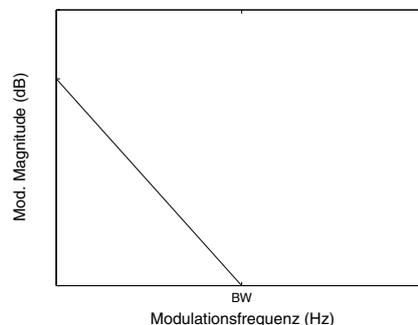
**Abbildung 4.2:** Werte des 4. Moments. In der linken Spalte sind die mittleren Lückendauer des SN angegeben. GN steht für Gauss'sches Rauschen, von dem es das SN zu unterscheiden galt. In der obersten Reihe sind jeweils die Eckfrequenzen des Bandpassfilters angegeben. Der Bereich in dem eine Diskriminierung von SN und GN möglich ist, ist grau hinterlegt.



**Abbildung 4.3:** Graphische Darstellung der Werte für das 4. Moment. Entlang der X-Achse sind die mittleren Lückendauern des SN aufgetragen, entlang der Y-Achse der dekadische Logarithmus des vierten Moments. Die Kreise stellen die Werte für eine untere Eckfrequenz von 1 kHz dar, die Quadrate diejenigen für eine untere Eckfrequenz von 7 kHz. Dabei ist die Bandbreite bei den offenen Symbolen 100 Hz, bei den geschlossenen 700 Hz. Die Rauten zeigen die Werte für die ungefilterten Signale an. Die gestrichelte Linie markiert den Schwellenwert, ab der eine Diskriminierung zwischen SN und GN nicht mehr möglich ist. Man sieht, dass bis zu dieser Grenze die Rauigkeit mit sinkender mittlerer Lückendauer linear abnimmt und sich darunter kaum noch verändert.

schwelen, ist grau hinterlegt. Anhand der Zahlenwerte, von unten nach oben in der Tabelle betrachtet, lässt sich erkennen, dass im Detektionsbereich die Rauigkeit mit abnehmender Lückendauer linear abnimmt. Oberhalb der Schwelle, sind die Veränderungen minimal. Um dies zu verdeutlichen ist das vierte Moment in Abhängigkeit von der Lückendauer in Abb. 4.3 graphisch dargestellt. Hier wird besonders deutlich, dass das Maß für die Rauigkeit sehr gut mit den gemessenen Schwellen korreliert. Da es aber schwer vorstellbar ist, dass das auditorische System eine Größe wie das vierte Moment „errechnet“ und auswertet, werden nun die Ergebnisse im Hinblick auf die psychoakustische Wahrnehmungsgröße der Rauigkeit diskutiert.

Im psychoakustischen Sinne ist Rauigkeit eine Wahrnehmungsgröße, die Schallschwankungen anzeigt: „Als Rauigkeit wird die schnarrende Hörqualität bezeichnet, welche sich aus relativ raschen Fluktuationen ergibt“ (Terhardt, 1998). Dabei handelt sich vornehmlich um Fluktuationen in der Einhüllenden. Beispiele für alltägliche Schalle die Rauigkeit hervorrufen sind ein fliegender Käfer oder das summen eines Elektromotors. Rauigkeit entsteht durch Modulationsfrequenzen von ca. 15-300 Hz. Laut Zwicker und Fastl (1999) muss das Modulationsspektrum Frequenzen in diesem Bereich aufweisen, damit Rauigkeit entsteht. Aus diesem Grund sind auch die meisten Schmalbandrauschen (SBR) als rau einzustufen, selbst wenn sie nicht amplitudenmoduliert sind, denn durch den Bandpass ergibt sich im Modulationsspektrum ein Tiefpass (Abb.4.4).

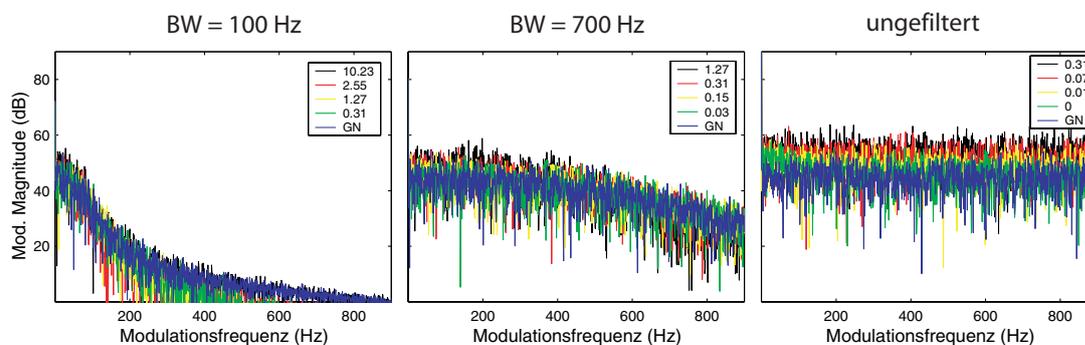


**Abbildung 4.4:** Idealisiertes Modulationsspektrum Schmalbandrauschens. BW markiert die Bandbreite des Signals.

Das höchste Maß an Rauigkeit wird, mit einer Modulationsfrequenz von 70 Hz erreicht. Das gilt sowohl für amplitudenmodulierte Reintöne ab einer Trägerfrequenz von 1 kHz als auch für amplitudenmoduliertes breitbandiges Rauschen (Zwicker und Fastl, 1999). Die Rauigkeit von unmoduliertem SBR ist für Bandbreiten um die 100 Hz besonders groß. Das liegt daran, dass die mittlere Folgefrequenz der Einhüllendenmaxima, ca.

$$\text{mittlere } f_{\text{mod}} = 0,64 \times \Delta f_{\text{SBR}}. \quad (4.2)$$

ist (Aures, 1985). Somit ist die mittlere Modulationsfrequenz für ein 100 Hz breites SBR ca. 64 Hz. In Anbetracht dessen ist es plausibel, dass die Detektionsschwelle bei einer Signalbandbreite von 100 Hz am höchsten ist (vgl. 4.1A,B). Hier ist die Rauigkeit der Stimuli, also auch die des Referenzstimulus (GN), durch die Bandbreite maximal ausgeprägt. Interessant ist, dass hier die mittlere Lückendauer die gerade noch detektiert werden konnte (im Versuch) 2,55 ms ist. Das entspricht ungefähr der Schwelle zur Lückendetektion bei Breitbandrauschen (Forrest und Green, 1987). Es wäre denkbar, dass in dieser Situation Rauigkeitsunterschiede keine ausreichende Grundlage mehr bieten um SN und GN zu unterscheiden und hier das zeitliche Auflösungsvermögen die Schwelle bestimmt. Grundsätzlich nimmt die Rauigkeit mit dem Modulationsgrad bzw. der Modulationstiefe ab. Mit abnehmender Modulationstiefe, verringert sich auch die Modulationsmagnitude an der entsprechenden Frequenz. Abbildung 4.5 zeigt die Modulationsspektren von SN mit unterschiedlichen Lückendauer und GN im Vergleich. In einem Feld sind jeweils die Spektren für Signale gleicher Bandbreite zusammengefasst. Den Legenden ist die mittlere Lückendauer in ms zu entnehmen. Bei Betrachtung des rechten Feldes (ungefilterte Signale) erkennt man dass die Modulationsmagnitude mit sinkender Lückendauer stetig abnimmt. In diesem Feld zeigt die gelbe Kurve das Modulationsspektrum für das SN an der Detektionsschwelle an. Bei einer Signal-Bandbreite von 700 Hz ist dies nur noch für Modulationsfrequenzen bis ca. 300 - 350 Hz der Fall, wenn auch weniger augenscheinlich. Beträgt die Bandbreite nur 100 Hz, scheinen die Unterschiede fast vollständig zu verwischen. Im Bereich bis ca. 70 Hz



**Abbildung 4.5:** Vergleich der Modulationsspektren von SNs mit unterschiedlicher mittlerer Lückendauer (siehe Legende, Angaben in ms). Die Bandbreite der Stimuli ist oberhalb der Graphen angegeben. Das Modulationsspektrum an der Schwelle wird in dem linken und dem mittleren Feld durch die rote Kurve angezeigt. Im rechten Feld durch die gelbe Kurve. Während bei den ungefilterten Signalen die Modulationsmagnitude deutlich abnimmt mit zunehmender Lückendauer, sind die Unterschiede mit abnehmender Bandbreite deutlich weniger ausgeprägt.

kann man zumindest noch Unterschiede zwischen dem SN an der Schwelle (rote Kurve) sowie dem GN (blaue Kurve) ausmachen. Berechnet man die mittlere Modulationsmagnitude zwischen 15-300 Hz und vergleicht die Werte von GN und dem SN an der jeweiligen Schwelle beträgt der Unterschied ca. 1,5-2,5 dB, mit Ausnahme der 100 Hz breiten Stimuli. Eine Magnitudendifferenz von 2 dB ist im Modulationsbereich eine wahrnehmbare Größe, insofern sind diese Unterschiede plausibel.

Klar ist jedoch, dass die Entscheidungen auf Grund von relativen Rauigkeitsunterschieden getroffen wurden. Am eindrucksvollsten wird das durch das vierte Moment gezeigt. Für Untersuchungen zur psychoakustischen Rauigkeit ist SN sicher ein lohnenswerter Stimulus.

# Kapitel 5

## Unterscheidungsschwellen für spektral aufgelöstes und spektral nicht aufgelöstes IRN

### 5.1 Einleitung

Die Suche nach Mechanismen der Tonhöhenwahrnehmung ist schon seit langer Zeit ein kontrovers diskutiertes Gebiet der auditorischen Forschung. Da Tonhöhe kein physikalisch definierter Parameter ist, sondern eine subjektive Empfindungsgröße, und die Tatsache dass physikalisch unterschiedliche Signale die gleiche Tonhöhe erzeugen, erschwert die Suche nach den Mechanismen. Eine Kontroverse ist die Frage, ob es einen oder mehrere Mechanismen gibt, die Tonhöhe eines Signals zu extrahieren.

Ein harmonischer Komplex (HC) besteht z.B. aus einer Grundfrequenz  $F_0$  und einer beliebigen Anzahl an ganzzahligen Vielfachen dieser Frequenz. Die wahrgenommene Tonhöhe eines solchen Komplexes entspricht der  $F_0$  unabhängig davon, ob diese Komponente physikalisch vorhanden ist oder nicht. Die niedrigen Komponenten werden dabei von der Peripherie des auditorischen Systems spektral aufgelöst, die höherfrequenten Komponenten dagegen nicht. Zwei Teiltöne eines Komplexes gelten als spektral aufgelöst, wenn sie in unterschiedliche auditorische Filter fallen, d.h wenn die Frequenz-Differenz dieser beiden Komponenten größer ist als die Bandbreite des durch sie angeregten auditorischen Filters. Carlyon und Shackleton (1994) haben drei Kriterien genannt, die ein HC erfüllen muss, um als spektral aufgelöst zu gelten: (i) Eine Phasenverschiebung jeder 2. Komponente um  $90^\circ$  bewirkt eine Verschiebung der wahrgenommenen Tonhöhe um eine Oktave (im Gegensatz zu der Situation wenn alle Komponenten dieselbe Phase haben). Dies tritt bei HCs, die spektral nicht aufgelöst sind, nicht auf. (ii) sämtliche  $F_0$  Unterscheidungsschwellen, die in dieser Arbeit ermittelt wurden, waren für nicht aufgelöste HCs höher als für aufgelöste

HCs. (iii) Die Anzahl der Harmonischen, die innerhalb der -10 dB Punkte<sup>1</sup> des auditorischen Filters liegen, dessen Mittenfrequenz in das Passband des entsprechenden HCs fällt, ist für spektral nicht aufgelöste Komplexe höher als drei und für aufgelöste Komplexe niedriger als zwei.

Das Tonhöhenmodell von Meddis und Hewitt (1991a,b) (siehe Abschnitt 2.5.2) z.B. berechnet die wahrgenommene Tonhöhe eines HC mit einem einzigen Mechanismus. Es errechnet die Autokorrelation des neuronalen Anregungsmusters des Hörnervs (*spike-probability-funktion*, SPF) und bestimmt somit die Grundfrequenz (F0) des Komplexes. Dabei ist es egal, ob die Teiltöne spektral aufgelöst werden oder nicht. Dass ein Mechanismus ausreichend ist, stellen u.a. Carlyon und Shackleton (1994) in Frage. In dieser Arbeit werden psychoakustische Versuche beschrieben, bei denen die Probanden die F0s von zwei simultan dargebotenen Komplexen vergleichen mussten. In einem Auswahlverfahren wurden in jedem Intervall zwei HCs simultan dargeboten und die Probanden mussten herausfinden, in welchem Intervall die beiden HCs unterschiedliche F0s hatten. Die Ergebnisse zeigten, dass eine bessere Leistung erzielt wurde, wenn *beide* Komplexe entweder aus spektral aufgelösten oder spektral nicht aufgelösten Komponenten bestanden. Die Autoren werteten ihre Ergebnisse als Hinweis dafür, dass es zwei unterschiedliche Mechanismen für die Extraktion der F0 gibt: Einen für spektral aufgelöste, einen für spektral nicht aufgelöste Harmonische. Um diese Idee weiter zu verfolgen, wurden in einer Arbeit von Plack und Carlyon (1995) gerade wahrnehmbare Grundfrequenzunterschiede (F0 DL - *fundamental frequency difference limen*) ermittelt, die zur Diskriminierung von F0s harmonischer Komplexe nötig waren, in Abhängigkeit von der Signaldauer. Die F0 DLs wurden sowohl für Komplexe mit spektral aufgelösten Harmonischen gemessen, als auch für Komplexe mit spektral nicht aufgelösten Harmonischen. Ihre Ergebnisse zeigten, dass während die F0 DLs der Erstgenannten relativ unabhängig von der Signaldauer waren, die Schwellen bei Letzteren deutlich ansteigen sobald die Signaldauer kürzer als 100 ms war. In der Tatsache, dass sich die Auswirkungen der Signaldauer auf die F0 DLs in beiden Bedingungen deutlich unterscheiden, sehen die Autoren ein weiteres Indiz dafür, dass das auditorische System zwei unterschiedliche Mechanismen zur Extraktion der F0 anwendet. Ungeachtet in ihrer Diskussion bleibt jedoch die Tatsache, dass spektral nicht aufgelöste Harmonische eine wesentlich schwächer ausgeprägte Tonhöhe produzieren als spektral aufgelöste. Es ist plausibel anzunehmen, dass die F0 DLs sich erhöhen, wenn die Tonhöhe nur schwach ausgeprägt ist, da man annehmen würde, dass eine schwächer ausgeprägte Tonhöhe die Identifikation erschwert. Das wiederum würde die Unterscheidungsfähigkeit beeinträchtigen (Houtsma und Smurzynski, 1990).

In diesem Kapitel wird eine Reihe von Experimenten beschrieben, die der Frage nachgehen, ob die Divergenz in der Auswirkung von Signaldauer auf die F0 DLs durch die verringerte Ausgeprägtheit der Tonhöhe (PS von engl. *pitch strength*) bei spektral nicht aufgelösten Harmonischen verursacht wird. An dieser Stelle soll nochmals daran erinnert

---

<sup>1</sup>Die -10 dB Punkte entsprechen denjenigen Frequenzpunkten entlang den abfallenden Flanken eines Bandpassfilters, an dem die Magnitude um 10 dB gegenüber der Magnitude der Mittenfrequenz abgeschwächt ist.

werden, dass die PS von IRN mittels des Abschwägungsfaktors  $g$  stufenlos variiert werden kann (Yost, 1996b). Darüber hinaus weist IRN ein Kammfilterprofil im Magnitudenspektrum auf (vgl. Abb. 1.3), welches eine Tonhöhenempfindung hervorruft, die durch das erste Maximum (entsprechend der  $F_0$ ) definiert ist. Insofern ist IRN ein Stimulus, der von den spektralen Eigenschaften her einem HC sehr ähnlich ist und zusätzlich die Möglichkeit bietet, die PS stufenlos zu variieren.

In den hier vorgestellten Experimenten werden Unterscheidungsschwellen zwischen zwei IRN Stimuli gemessen. Die Tonhöhe von IRN ist durch das Reziproke der Verzögerungszeit  $d$  festgelegt. Die  $F_0$  von IRN entspricht also  $1/d$  (siehe 1.1.1). Um den Vergleich zwischen den hier vorgestellten Experimenten mit anderen Arbeiten zu erleichtern, wird im Folgenden die jeweilige  $F_0$  der IRN Stimuli anstelle von  $d$  angegeben. Im ersten Teil wurden  $F_0$  DLs für IRN Stimuli mit einer mittleren  $F_0$  von 90,9 oder 250 Hz (entsprechend einem  $d$  von 11 bzw. 4 ms) gemessen. Dies geschah jeweils in Abhängigkeit von der Signaldauer. Die Stimuli wurden bandpassgefiltert, so dass sie entweder spektral aufgelöst oder nicht aufgelöst waren.

In einem zweiten Teil wurde ein Vergleichs- bzw. Angleichungsexperiment durchgeführt, das jenen Wert für  $g$  ermittelt hat, mit dem ein spektral aufgelöstes IRN eine ungefähr gleich ausgeprägte Tonhöhe besitzt wie ein spektral nicht aufgelöstes. Im letzten Teil wurden die IRNs mit diesem  $g$ -Wert berechnet und erneut  $F_0$  DLs gemessen, diesmal nur für spektral aufgelöste Stimuli. Ist die Ausgeprägtheit von Tonhöhe tatsächlich die Ursache für die Divergenz in der Auswirkung der Signaldauer auf  $F_0$  DLs von spektral aufgelösten und spektral nicht aufgelösten HCs, sollten die  $F_0$  DLs aus Teil III und die aus Teil I, für spektral nicht aufgelöstes IRN, annähernd dieselben sein.

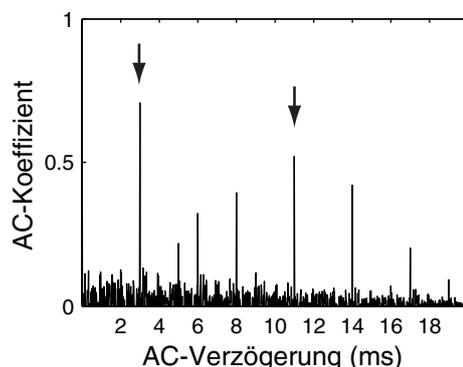
## 5.2 Messung von Unterscheidungsschwellen für IRN-Stimuli

Zunächst wurden die  $F_0$  DLs für IRN mit  $F_0=90,9$  Hz und IRN mit  $F_0=250$  Hz in unterschiedlichen Filterbedingungen gemessen, so dass die Signale entweder spektral aufgelöst wurden oder nicht. Ziel ist es, festzustellen wie klein der Unterschied der  $F_0$ s ( $\Delta F_0$ ) in Abhängigkeit von der Signaldauer sein darf, so dass die Versuchsperson noch eindeutig das Signal mit der höheren Tonhöhe identifizieren kann.

### 5.2.1 Methoden

#### Probanden

An diesen Versuchen beteiligten sich vier Probanden im Alter zwischen 25 und 32 Jahren. Die Versuchspersonen waren in allen drei Teilversuchen dieselben. Die Teilnahme erfolgte freiwillig und teilweise unentgeltlich. Teilweise wurde ein Stundenlohn gezahlt. Alle verfügten über ein klinisch normales Hörvermögen. Jeder Teilnehmer absolvierte eine Trainingsphase bis sich die Werte stabilisierten.



**Abbildung 5.1:** Autokorrelationsfunktion von IRN mit  $F_0=90,9$  Hz bei einer Stimulusdauer von 25 ms. Diese  $F_0$  entspricht einer IRN-Verzögerung von 11 ms. Man erkennt dass neben dem Maximum bei einer AC-Verzögerung von 11 ms, noch weitere Maxima vorhanden sind. Ein sehr deutliches bei 3 ms. An dieser Stelle ist der Korrelationskoeffizient sogar höher als bei 11 ms und verändert daher maßgeblich die wahrgenommene Tonhöhe.

### Stimuli und Versuchsausrüstung

Die IRN Stimuli wurden nach der in Abschnitt 1.1.1 erläuterten Methode berechnet. Das geometrische Mittel der Verzögerungszeiten ( $ds$ ) zweier zu vergleichender IRNs war entweder 4 oder 11 ms, entsprechend den  $F_0$ s von 250 oder 90,9 Hz. Die Anzahl der Iteration lag bei 32 und der Abschwächungsfaktor ( $g$ ) betrug eins. Alle Stimuli wurden bandpassgefiltert. Es gab drei unterschiedliche Filterbedingungen: LOW, MID und HIGH mit den Eckfrequenzen 125-625 Hz, 1375-1875 Hz und 3900-5400 Hz. Demzufolge waren die IRNs mit  $F_0=90,9$  Hz nur in der Filterbedingung LOW spektral aufgelöst. IRNs mit  $F_0=250$  Hz konnten in der LOW und in der MID Filterbedingung spektral aufgelöst werden. Die Bandpassfilter wurden als digitale Butterworthfilter 4. Ordnung, mit einer Flankensteilheit von 24 dB/Oktave, in MatLab implementiert. Die Abtastrate in diesem Experiment betrug 200 kHz. Da der Mensch äußerst empfindlich für Tonhöhenunterschiede ist und dadurch die gemessenen  $F_0$  DLs sehr gering waren, war es nötig eine sehr hohe Abtastrate anzuwenden. Bei einer Abtastrate von 200 kHz/s entspricht der Schritt von einem Abtastpunkt zum nächsten  $5 \mu s$ . Zeitdifferenzen, die geringer sind, können nicht mehr sauber abgebildet werden. Umgerechnet auf die  $\Delta F_0$ s bedeutet das, dass  $F_0$ -Unterschiede bis zu 0,04 % mit dieser Abtastrate erfasst werden können.

Für jede der insgesamt sechs Bedingungen (zwei  $F_0$ s mit je drei verschiedenen Filtern) wurde der  $F_0$  DL in Abhängigkeit von der Stimulusdauer gemessen. Die Stimulusdauern betragen 50, 100 und 200 ms für IRN mit einer  $F_0$  von 90,9 Hz. Für IRN mit einer  $F_0$  von 250 Hz waren die Stimuli 25, 50, 100 und 200 ms lang. Alle Signale beinhalteten kosinusförmige An- und Abstiegsflanken von je 5 ms. Da IRN mit einer  $F_0$  von 90,9 Hz bei einer Signaldauer von 25 ms zusätzliche Maxima in der Autokorrelationsfunktion zeigt (Abb. 5.1), die eine Verschiebung der wahrgenommenen Tonhöhe bewirkt, wurde diese Bedingung verworfen.

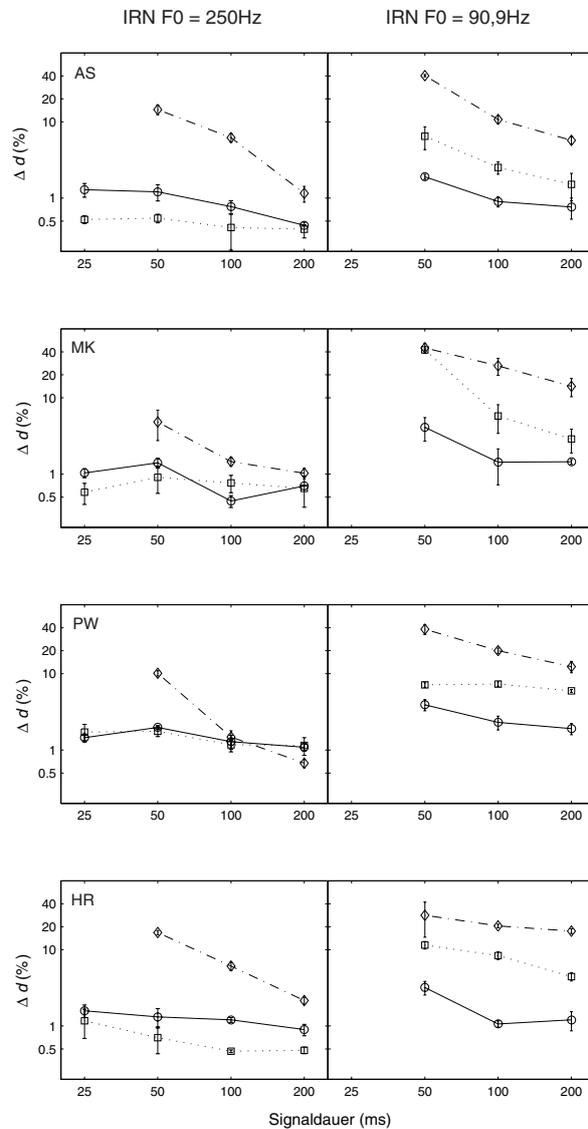
Die auf einem IBM-kompatiblen PC generierten Signale wurden auf einen Tucker-Davis-Technologies System3 RP2 *Real-Time Processor* übergeben. Die analoge Ausgabe der Stimuli erfolgte über den integrierten D/A-Wandler und einen Tucker-Davis-Technologies System3 HB 7 Kopfhörerverstärker, dessen Endabschwächung auf -18 dB eingestellt war. Der mittlere Ausgabepegel lag bei 64 dB SPL  $\pm$ 6 dB. Die Endausgabe erfolgte diotisch über AKG K240DF Kopfhörer. Im Hintergrund wurde dauerhaft ein 10 kHz breites rosa Rauschen geboten, um hörbare Verzerrungsprodukte unterhalb des Passbandes zu maskieren. Es wurde von einem Brüel&Kjaer 1405 Rauschgenerator erzeugt und mit einem Pegel von 16 dB SPL bei 1 kHz ausgegeben.

### Paradigma

Die F0 DLs wurden durch ein adaptives 2AFC Paradigma ermittelt. Der adaptive Parameter war  $\Delta F_0$ . Dieser wurde dann in das entsprechende  $d(\text{ms})$  umgerechnet. Zu Beginn einer Messung lag der Unterschied der beiden IRN F0s bei 40 Hz. Dieser Anfangswert hat sich während der Trainingsphase der Probanden als günstig erwiesen. Das  $\Delta F_0$  wurde nach einer *two-down one-up* Regel, welche die 71% korrekt Marke entlang einer psychometrischen Funktion ermittelt (Levitt, 1971), um den Faktor 1,41 (entspricht ca.  $\sqrt{2}$ ) verändert. Dies ermöglichte eine rasche Annäherung an die Schwelle. Es wurden 16 Umkehrpunkte aufgenommen. Das geometrische Mittel der letzten 12 Umkehrpunkte ergab die Schwelle. In einen Durchlauf wurden für eine ausgewählte F0 und einem ausgewählten Filter (LOW, MID oder HIGH) die F0 DLs für die verschiedenen Signaldauern in randomisierter Reihenfolge erhoben. Pro Sitzung wurden mehrere Durchläufe absolviert. Jeweils mit verschiedenen Kombinationen von F0 und Bandpassfilter. Die Dauer einer Sitzung betrug im Schnitt ca. 90 Minuten. Die Schwelle der einzelnen Probanden für jede Signalkonfiguration errechnete sich aus dem Mittelwert von drei aufeinanderfolgenden Messungen. Die Auswahl wurde über ein graphisches Interface auf einem Touchscreen getroffen. Nach jeder Entscheidung wurde ein visuelles Feedback gegeben.

### 5.2.2 Ergebnisse

Die Ergebnisse von Teil I sind in Abb. 5.2 zu sehen. Da die Ergebnisse ziemlich heterogen ausgefallen sind, wurden keine Mittelwerte zwischen den Probanden ermittelt, sondern die Mittelwerte der Einzelnen getrennt voneinander dargestellt. Die Quadrate zeigen die Resultate des LOW-Filterbereiches, die Kreise die des MID-Filters, die Rauten die des HIGH-Filters. Bei einer IRN F0 von 250 Hz ist zwischen dem LOW und dem MID Bandpassbereich nur ein geringer bis gar kein Unterschied in den F0 DLs festzustellen. In beiden Fällen ist der F0 DL nur in gerigem Maße von der Signaldauer abhängig. Die Schwellen für den HIGH-Filterbereich heben sich deutlicher ab. Während sie für eine Signaldauer von 200 ms noch relativ nah an den Werten für MID und LOW sein können (MK und PW), steigen die Schwellen mit abnehmender Signaldauer bei allen Probanden an. War die Signaldauer nur 25 ms, konnte kein Proband das Intervall mit der höheren Tonhöhe eindeutig identifizieren.



**Abbildung 5.2:** F0 DLs für IRN mit F0=250Hz oder F0=90,9Hz für verschiedene Bandpassbereiche. Die IRNs wurden jeweils mit dem Abschwächungsfaktor eins berechnet. Die Quadrate repräsentieren den LOW-Filter, die Kreise den MID-Filter und die Rauten den HIGH-Filter. Für eine IRN F0 von 90,9Hz wurden nur drei Signaldauern gemessen. Bei einer IRN F0 von 250Hz und einer Signaldauer von 25 ms im Filter HIGH, war keiner der Probanden in der Lage eindeutig das Intervall mit der höheren Tonhöhe zu detektieren.

Ist die IRN  $F_0=90,9$  Hz erhöhen sich die F0 DLs sowohl mit ansteigender Mittenfrequenz bzw. unterer Eckfrequenz des Passbands, als auch mit abnehmender Signaldauer.

### 5.2.3 Diskussion

Die F0 DLs, die für eine IRN  $F_0$  von 250 Hz erzielt wurden, sind im Einklang mit den Ergebnissen von Plack und Carlyon (1995). In beiden Arbeiten sind die F0 DLs für spektral aufgelöste Stimuli (hier LOW und MID) relativ unabhängig von der Signaldauer. Für spektral nicht aufgelöste Stimuli dagegen (hier HIGH) steigen die F0 DLs ab einer Signaldauer von  $<100$  ms stärker an (v.a. Probanden MK und PW in Abb. 5.2). Bei den F0 DLs in dieser Arbeit ergeben sich z.T. auch schon deutliche Unterschiede bei einer Signaldauer von 100 ms (Probanden AS und HR in Abb. 5.2). Bei Plack und Carlyon (1995) liegen die F0 DLs für spektral aufgelöste HCs zwischen 0,5 und 1 %. Auch damit sind die aktuellen Ergebnisse im Einklang. Ist die IRN  $F_0=250$  Hz, liegen die F0 DLs für spektral aufgelöstes IRN ebenfalls in diesem Bereich.

Die Ergebnisse für IRN mit einer  $F_0$  von 90,9 Hz stellen ein anderes Bild dar. Insgesamt sind die Schwellen etwas höher als bei einer  $F_0$  von 250 Hz. Bei Plack und Carlyon (1995), die HCs mit den  $F_0$ s 62,5 und 250 Hz verwendeten, waren zumindest im spektral aufgelösten Bereich die F0 DLs für beide HCs gleich. Dass dies bei IRN nicht der Fall ist, könnte damit begründet sein, dass die Klangfarbe von IRN rauschhafter wird, je niedriger die IRN  $F_0$  ist. Aus den Ergebnissen der Periodizitätsdetektion (Abb. 2.3) geht hervor, dass IRN mit  $F_0=90,9$  Hz eine etwas schwächer ausgeprägte Tonhöhe besitzt als IRN mit  $F_0=250$  Hz. Dort musste IRN von GN unterschieden werden. Die Schwelle für  $F_0=250$  Hz lag bei einem  $g$  von -44 dB. Die interpolierte Schwelle für  $F_0=90,9$  Hz wäre bei ca. -39,5 dB. Geht man davon aus, dass die PS (*pitch strength*) die F0 DLs bestimmt, wäre es demzufolge eine logische Konsequenz, dass die F0 DLs für die längere Verzögerung höher ausfallen.

Des Weiteren steigen die F0 DLs für eine Signaldauer von 50 ms deutlich an, selbst bei spektral aufgelösten IRNs. Wenn man die ACF von IRN mit  $F_0=90,9$  Hz mit einer Stimulusdauer von 50 ms berechnet, stellt man fest, dass auch hier noch Artefakte zu sehen sind (keine Abbildung). Diese sind zwar deutlich weniger prominent und nicht auf Anhieb hörbar wie bei einer Signaldauer von 25 ms, dennoch könnten sie die F0 DLs beeinflusst haben. Bei nicht aufgelösten IRNs sind die F0 DLs insgesamt höher, gleich den Ergebnissen für eine IRN  $F_0$  von 250 Hz. Dagegen ist ein stärkerer Einfluss der Signaldauer auf die F0 DLs im Vergleich zu spektral aufgelösten Signalen in diesem Fall nicht zu beobachten. Warum sich IRN mit  $F_0=250$  und IRN mit  $F_0=90,9$  Hz in dieser Hinsicht unterscheiden, ist unklar. Möglicherweise haben auch hier die Artefakte in der ACF einen Einfluss. Außerdem ist zu beobachten, dass hier die F0 DLs mit der unteren Eckfrequenz (LCF) des Passbandes steigen (Vergleich: kein Unterschied zwischen LOW und MID bei IRN  $F_0=250$  Hz). Ist die  $F_0=90,9$  Hz, werden die Stimuli sowohl in MID als auch in HIGH spektral nicht aufgelöst. Das ist konsistent zu Ergebnissen von Houtsma und Smurzynski (1990). Mit der LCF des Passbandes erhöhen sich die Ordnungszahlen der, im Signal enthaltenen, HC-Komponenten - oder analog in diesem Experiment die Ordnungszahlen der *ripple*. Ist die LCF = 1375 Hz, ist Ordnungszahl des niedrigsten vorkommenden *ripples* gleich 16. Bei

einer LCF von 3400 Hz wäre 38 die niedrigste Ordnungszahl der im Signal vorhandenen *ripple*<sup>2</sup>. Houtsma und Smurzynski (1990) haben für HCs gezeigt, dass sich die DLs mit der Ordnungszahl der niedrigsten Harmonischen erhöhen.

### 5.3 Ermittlung des Abschwächungsfaktors

Die Ausprägtheit der Tonhöhe (PS) lässt sich bei IRN sehr leicht stufenlos variieren, indem man den Abschwächungsfaktor  $g$  ändert. In der Schleife des Verzögerns und Addierens zur Berechnung eines IRN Stimulus, wird die verzögerte Kopie des Rauschens zuerst mit  $g$  multipliziert bevor diese zum Eingangsruschen zurückaddiert wird (siehe Abb. 1.1). Üblicherweise nimmt  $g$  einen Wert zwischen 0 und 1 an<sup>3</sup>. Je geringer der Wert für  $g$  ist, desto schwächer ist die PS (Yost, 1996b). In Teil II wurde der Wert für  $g$  ermittelt, mit dem ein IRN, das spektral aufgelöst wird die gleiche PS besitzt wie ein IRN mit identischem  $d$ , das spektral nicht aufgelöst wird. Dieser Wert wurde für jede Versuchsperson nach individuellem, subjektiven Empfinden ermittelt.

#### 5.3.1 Stimuli und Paradigma

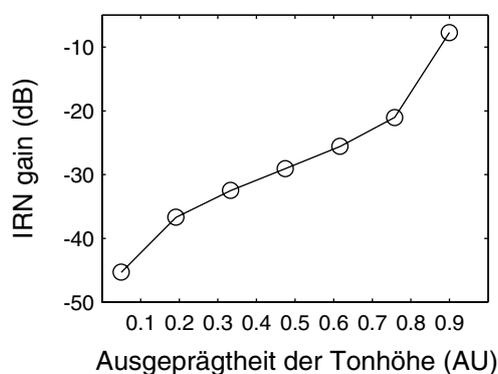
Da die Versuchsanordnung in allen drei Teilen des Experiments identisch ist, wird nicht mehr darauf eingegangen. Ebenso waren die Ausgabepegel der Signale und des rosa Rauschens ebenfalls bei allen drei Teilen gleich.

Hier wurden IRNs mit gleichen F0s, aber verschiedenen Bandpassbereichen miteinander verglichen. Die Passbänder wurden derart gewählt, dass die Signale einmal spektral aufgelöst wurden und einmal nicht. Das bedeutet konkret, dass für IRN F0=250 Hz die Filterbedingungen MID (aufgelöst) und HIGH (nicht aufgelöst) miteinander verglichen wurden und bei einer IRN F0 von 90,9 Hz die Bedingungen LOW und MID. Um ein besseres Gefühl für die PS zu erlangen, wurden in einem Intervall nicht einzelne Töne, sondern Melodien, bestehend aus fünf Tönen, vorgespielt. Die Einzeltöne wurden mit einer Bandbreite von einer halben Oktave um den Grundton zentriert, entweder um 250 oder um 90,9 Hz. Diese fünf „Noten“ wurden in zufälliger Weise aneinandergereiht, um eine Melodie zu bilden. In beiden Intervallen wurde dieselbe Melodie präsentiert. Nach jeder Entscheidung wurde eine neue Melodie berechnet. Die spektral nicht aufgelösten IRNs wurden mit  $g=1$  und 32 Iterationen berechnet, die spektral aufgelösten IRNs mit variablem  $g$  und ebenfalls 32 Iterationen. Entlang einer linearen Achse von null bis eins, die in frei gewählten Einheiten (*arbitrary units*, AU) die Ausprägtheit der Tonhöhe definiert, wurden sieben Punkte

---

<sup>2</sup>Man errechnet die Ordnungszahl der niedrigsten Komponente, indem man die UCF durch die F0 teilt:  $1375/90,9=15,125$ . Das bedeutet, dass die 15. Komponente gerade nicht in das Passband fällt. Somit ist die 16. die Niedrigste, die im Stimulus enthalten ist. Für LCF=3400 Hz:  $3400/90,9=37,4$ , daher ist die 38. Komponente die niedrigste im HIGH Filter.

<sup>3</sup>Der Abschwächungsfaktor  $g$  kann auch negative Werte annehmen, was zu Veränderungen in der Tonhöhenempfindung führt. Darauf wird in dieser Arbeit nicht eingegangen. Informationen dazu können Yost (1996a) entnommen werden.



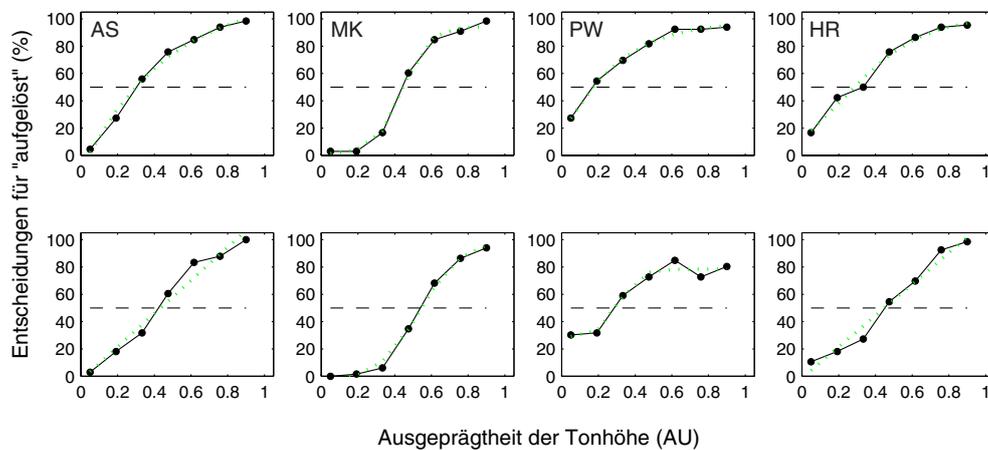
**Abbildung 5.3:** Diese Graphik zeigt die Werte der Umrechnung von der Ausgeprägtheit der Tonhöhe entlang einer Achse frei gewählter Einheiten von 0 bis 1 (*arbitrary units*; AU) in IRN  $gain$ (dB). Diese Werte gelten nur für IRN, das mit 32 Iterationen berechnet wurde.

ausgewählt, die äquidistant entlang dieser Achse sind. Dieser Wert wurde in  $g$  (dB) umgerechnet (Abb. 5.3). Diese Umrechnung ist nur gültig für IRN mit 32 Iterationen. Innerhalb eines Intervalls blieb  $g$  gleich.

Mittels eines 2AFC Paradigmas wurde eine psychometrische Funktion gemessen. In jedem Durchgang wurde in einem der beiden Intervalle eine Melodie bestehend aus spektral *nicht* aufgelösten IRNs präsentiert, die alle mit  $g=1$  berechnet wurden. Im anderen Intervall wurde eine Melodie präsentiert die aus spektral aufgelösten IRNs mit  $g<1$  bestand. Für jeden der sieben getesteten  $g$ -Werte wurden 66 Einzelentscheidungen aufgenommen. Diese wurde in drei Sitzungen pro Proband aufgeteilt. Jede Sitzung bestand aus 22 Präsentationen pro  $g$ , welche in randomisierter Reihenfolge dargeboten wurden. Daraus ergab sich eine Anzahl von insgesamt 154 Durchgängen pro Sitzung. Die Aufgabe bestand darin, auszuwählen, welcher Melodie leichter zu folgen war. Ein Feedback wurde nicht gegeben, da die Antwort weder „richtig“ noch „falsch“ war, sondern nach dem subjektiven Empfinden gefragt wurde. Es wurde ausgewertet wie oft (in Prozent) sich der Proband für das Intervall mit dem spektral aufgelöstem IRN entschieden hat. Mit Hilfe einer fit-Funktion wurde der 50% Wert entlang der psychometrischen Funktion jedes Probanden ermittelt.

### 5.3.2 Ergebnisse und Diskussion

In Abb. 5.4 werden die psychometrischen Funktionen aller Versuchspersonen für beide gemessenen  $ds$  dargestellt. In der oberen Reihe sind dabei die Daten für IRN  $F_0=250$  Hz abgebildet und in der unteren Reihe die Daten für IRN  $F_0=90,9$  Hz. In dieser Abbildung ist der Wert für die PS (in AU) gegen die Anzahl der Entscheidungen für das Intervall mit



**Abbildung 5.4:** Psychometrische Funktionen zur Ermittlung des  $g$ -Wertes, bei dem die Probanden spektral aufgelöstem IRN die gleiche Ausprägtheit an Tonhöhe zugeschrieben haben wie einem spektral nicht aufgelöstem IRN mit identischem  $d$ . Es wurde die Ausprägtheit der Tonhöhe entlang einer normierten Skala von 0 bis 1, gegen den Prozentsatz an Entscheidungen aufgetragen, die auf das Intervall mit dem spektral aufgelöstem IRN gefallen sind. Die obere Reihe zeigt die Ergebnisse der vier Probanden für IRN  $F_0=250$  Hz. Die untere Reihe zeigt die entsprechenden Ergebnisse für IRN  $F_0=90,9$  Hz. Die Fitfunktion ist grau dargestellt, die gestrichelte Linie deutet die 50% Marke an.

**Tabelle 5.1:** Die  $g$ -Werte in dB, die zur Berechnung der IRN-Stimuli in Teil III herangezogen wurden.

Proband	$g$ -Werte in dB	
	IRN F0=250 Hz	IRN F0=90,9 Hz
AS	-33,5	-30
MK	-36,7	-27,4
PW	-37,6	-33,4
HR	-33,8	-29,7

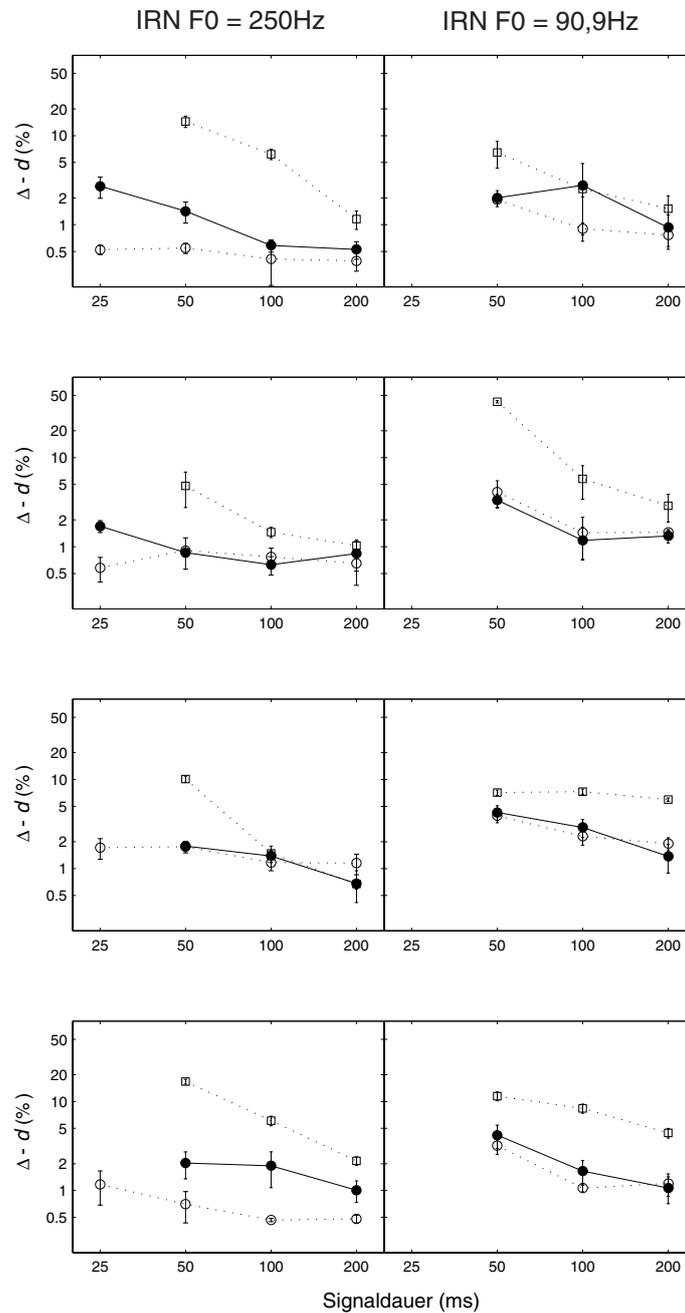
dem spektral aufgelöstem IRN aufgetragen. Je geringer die PS desto weniger oft haben sich die Probanden für das Intervall mit der spektral aufgelösten Melodie entschieden. Der Schnittpunkt der Fitfunktion mit der 50 % Linie entspricht dem Punkt, an welchem der Proband dem spektral aufgelösten und dem spektral nicht aufgelösten IRN die gleiche Ausprägtheit der Tonhöhe zuschreiben. Aus diesem Wert wurde dann  $g$  in dB errechnet. Die Werte für  $g$  in dB, mit dem die IRN Stimuli Teil III berechnet wurden, sind für jeden Probanden und jedes  $d$  in Tabelle 5.1 angegeben.

## 5.4 Unterscheidungsschwellen von IRN-Stimuli mit verändertem gain

Im dritten und letzten Teil dieser experimentellen Serie wurden erneut F0 DLs für IRN mit denselben F0s wie in Teil I gemessen, mit folgender Veränderung: alle IRNs waren spektral aufgelöst, allerdings mit einem  $g$  berechnet, welches den Stimuli die gleiche Ausprägtheit an Tonhöhe verleiht wie den entsprechenden spektral nicht aufgelösten IRNs. Im Vergleich mit den Schwellen des ersten Teils soll ersichtlich werden, welchen Einfluss die Ausprägtheit der Tonhöhe auf die F0 DLs hat. Im Vordergrund des Interesses steht v.a. die Frage: Was beeinflusst die F0 DLs stärker - ob die Stimuli spektral aufgelöst sind oder nicht, oder wie stark oder schwach ausgeprägt die Tonhöhe ist? Deshalb war es ausreichend, in diesem Teil die F0 DLs jeweils nur für spektral aufgelöste IRNs zu messen.

### 5.4.1 Stimuli und Paradigma

Die IRNs wurden nach derselben Methode berechnet wie in Teil I. Das geometrische Mittel der F0 beider zu vergleichender IRNs betrug auch hier entweder 250 oder 90,9 Hz. Die Anzahl der Iterationen lag bei 32. Für  $g$  wurden die Werte eingesetzt, die im vorangegangenen Experiment ermittelt wurden. Das bedeutet, dass für jeden Probanden die IRN-Stimuli mit einem anderen  $g$  berechnet wurden. Tabelle 5.1 gibt die individuellen Werte an. Die Filterbedingungen beschränkten sich auch LOW für IRN mit einer F0 von 90,9 Hz und MID für eine IRN F0 von 250 Hz. Somit sind die Signale stets spektral aufgelöst.



**Abbildung 5.5:** F0 DLs für IRN F0=250Hz in der linken Spalte und IRN F0=90,9Hz in der rechten Spalte. Die Stimuli wurden jeweils mit einem  $g < 1$  berechnet. Die ausgefüllten Kreise, verbunden durch die durchgezogenen Linie stellen die hier gemessenen F0 DLs dar. Die Signale waren hier stets spektral aufgelöst. Die gestrichelten Linien zeigen die Vergleichswerte aus Teil I (Abb. 5.4) an. Die offenen Kreise repräsentieren die Werte für dasselbe Passband wie die geschlossenen Kreise (MID für IRN F0=250 Hz und LOW für IRN F0=90,9 Hz); beide spektral aufgelöst). Die offenen Quadrate repräsentieren die Werte für den spektral nicht aufgelösten Bandpassbereich (HIGH für IRN F0=250 Hz und MID für IRN F0=90,9 Hz).

### 5.4.2 Ergebnisse

Wie in Teil I, fallen auch hier die Ergebnisse recht heterogen aus. Da es, wie gesagt, Ziel dieser Messung war, die Ergebnisse mit denen der Ersten zu vergleichen, werden in Abb. 5.5 die hier gewonnenen F0 DLs zusammen mit denen aus Teil I zusammen dargestellt. Die geschlossenen Kreise stellen die Ergebnisse dieser Messung dar: Passband MID für IRN F0=250 Hz und LOW für IRN F0=90,9 Hz. Der  $g$  der IRN Stimuli war in diesen Messungen stets kleiner als eins (Werte in dB für jeden Probanden können Tabelle 5.1 entnommen werden). Die offenen Symbole zeigen die Ergebnisse aus der ersten Messung in denen  $g$  gleich eins war. Die offenen Kreise repräsentieren dabei wieder Ergebnisse für spektral aufgelöste Stimuli, die offenen Quadrate die Ergebnisse für spektral nicht aufgelöste Stimuli (Passband HIGH respektive MID). Für IRN F0=90,9 Hz ist zwischen den F0 DLs gleicher Filterbedingung (LOW) kein Unterschied zu erkennen (Ausnahme Hörer AS). Für IRN F0=250 Hz ist der Trend sehr ähnlich, wenn auch nicht mit gleicher Deutlichkeit. Hörer PW und HR reicht z.B. eine Signaldauer von 25 ms nicht aus, um die beiden IRNs eindeutig zu unterscheiden. Dies war auch in der spektral nicht aufgelösten Situation aus Teil I der Fall (offene Quadrate). Bei HR liegen darüberhinaus die F0 DLs gemessen für IRNs mit  $g < 1$  genau zwischen den F0 DLs der spektral aufgelösten und nicht aufgelösten Situation aus Teil I. Betrachtet man AS und MK, stiegen die F0 DLs zumindest für sehr kurze Signaldauern (50 und 25 ms bzw. nur 25 ms) an, was in der ersten Messung auch nicht der Fall war. Insgesamt sind die F0 DLs der spektral aufgelösten IRNs mit  $g < 1$ , den F0 DLs von spektral aufgelösten IRNs mit  $g = 1$  wesentlich ähnlicher als denen spektral *nicht* aufgelöster IRNs mit  $g = 1$ .

### 5.4.3 Diskussion

Plack und Carlyon (1995) hatten in ihrer Arbeit F0 DLs für HCs mit den Grundfrequenzen 62,5 und 250 Hz gemessen und gezeigt, dass der Einfluss der Signaldauer auf die F0 Diskriminierung für spektral aufgelöste HCs geringer war als für spektral nicht aufgelöste. Die Kernfrage des vorliegenden Experiments lag darin, zu ergründen, ob sich diese Diskrepanz darauf zurückführen lässt, dass HCs mit spektral nicht aufgelösten Harmonischen eine schwächer ausgeprägte Tonhöhe besitzen. Wäre das tatsächlich der Fall, würde man erwarten, dass bei gleich ausgeprägter Tonhöhenempfindung keine Unterschiede bei der F0 Diskriminierung mehr auftreten. Im Hinblick auf die F0 DLs von IRN Stimuli würde das bedeuten, dass die Schwellen gemessen für ein spektral nicht aufgelöstes IRN und die Schwellen gemessen für ein spektral aufgelöstes IRN, berechnet mit einem  $g < 1$ , welches dem Stimulus eine ungefähr gleich ausgeprägte Tonhöhenempfindung verleiht, annähernd gleich sein müssten. In Abb. 5.5 müssten die geschlossenen Kreise (aufgelöstes IRN mit  $g < 1$ ) und die offenen Quadrate (nicht aufgelöstes IRN mit  $g = 1$ ) sehr dicht beisammen liegen. Das ist eindeutig nicht der Fall. Vielmehr ist der Unterschied zu den F0 DLs von spektral aufgelösten IRNs mit  $g = 1$  sehr viel geringer. Möglicherweise müsste der IRN Abschwächungsfaktor noch weiter reduziert werden, um einen größeren Effekt zu erzielen. Betrachtet man aber die psychometrische Funktion aus Teil 2 (Abb. 5.4) sieht man, dass

diese monoton steigend ist, also mit stärkerer Ausprägtheit der Tonhöhe die Anzahl der Entscheidungen für die spektral aufgelöste Melodie ("Welcher Melodie kann leichter gefolgt werden?") sich stetig erhöht. Um einer Melodie gut folgen zu können, d.h. sie mitsingen/-summen zu können, muss man in der Lage sein, die einzelnen Noten gut voneinander abgrenzen zu können. Das geht umso einfacher, je deutlicher bzw. eindeutiger einer Note eine bestimmte Tonhöhe zugewiesen werden kann. Insofern ist auch davon auszugehen, dass der Schnittpunkt mit der 50 % Marke ein sinnvoller Wert für die Berechnung der IRN Stimuli in Teil III war. Die vorliegenden Ergebnisse legen aber den Schluss nahe, dass die Ausprägtheit der Tonhöhe nur geringen oder auch gar keinen Einfluss auf gerade wahrnehmbare Tonhöhenunterschiede von IRN hat. Das ist allerdings schwer vorstellbar, da generell eine eindeutigere Identifizierung von Objekten die Diskriminierung gegenüber ähnlichen Objekten erleichtert. Vielleicht ist aber auch für IRN der Zusammenhang zwischen PS und F0 DL nicht linear. Es wäre möglich, dass sich der Zusammenhang asymptotisch verhält und dass ab einem gewissen Punkt die Schwelle abrupt ansteigt. Wegen der ohnehin rauschhaften Klangfarbe von IRN, stellt sich das Gehör wohl darauf ein und unterdrückt den Rauschanteil. Die Unterscheidungsschwellen wären sicherlich höher müsste man ein IRN mit einem HC vergleichen. In der Arbeit von Carlyon und Shackleton (1994) erzielten die Probanden bessere Ergebnisse, wenn sie HCs mit gleicher PS verglichen. Aufgrund der Ergebnisse wäre es interessant, F0 DLs in Abhängigkeit von  $g$  und somit der PS zu bestimmen.

Insgesamt muss festgestellt werden, dass die Unterschiede der F0 DLs nicht durch die unterschiedliche PS bedingt sind. Auch in Teil III, den F0 DLs für IRN mit  $g < 1$ , scheinen die Ergebnisse primär durch Auflösbarkeit der spektralen Komponenten bestimmt zu sein. Eine Ergänzung um F0 DLs von spektral nicht aufgelösten IRNs mit  $g < 1$  könnte darüber weitere Informationen bringen.

Eine Reihe von psychoakustischen Messungen haben Unterschiede in der Wahrnehmung, je nach Auflösbarkeit/Nicht-Auflösbarkeit der spektralen Komponenten, gezeigt (z.B. Carlyon und Shackleton (1994); Houtsma und Smurzynski (1990); Plack und Carlyon (1995); Shackleton und Carlyon (1994)). Insofern ist die Argumentation, es gäbe zwei unterschiedliche Mechanismen zur Extraktion der F0 nicht abwegig. Auch die vorliegenden Ergebnisse scheinen die gleiche Tendenz zu haben.

Houtsma und Smurzynski (1990) haben aber in ihrer Diskussion bereits darauf hingewiesen, dass auch die Möglichkeit besteht, dass es sich um nur einen neuronalen Mechanismus handelt der, je nach Input aus der Peripherie unterschiedlich reagiert. Die Autoren beschreiben, wie ein Modell von Sruлович und Goldstein (1983) in der Lage ist die korrekte Tonhöhe für einen Harmonischen Komplex vorherzusagen, unabhängig davon ob dessen Komponenten spektral aufgelöst werden oder nicht. Darüberhinaus spiegelt sich in der Simulation auch die schwächere Ausprägung der Tonhöhe wider.

Meddis und O'Mard (1997) haben ihr Tonhöhenmodell mit dem Paradigma von Carlyon und Shackleton (1994) getestet. Es wurden HCs mit den gleichen Eckfrequenzen wie hier die Bedingungen LOW und HIGH, bandpassgefiltert. Die F0s waren 100 und 102 Hz. Die SACGs (*summary autocorrelogramms*, siehe Abb. 2.10) der LOW-Stimuli wiesen deutliche Unterschiede zwischen 100 und 102 Hz auf, die der HIGH-Stimuli dagegen nicht. Um den Befund zu untermauern, haben die Autoren zusätzliche Passbänder getestet, einmal MID

(entsprechend dem hier beschriebenen) und MID/HIGH (2500-3500 Hz). Die Euklidische Distanz der SACGs der beiden Komplexe nahm mit steigendem Passband stetig ab. Insofern hat dieses Modell gezeigt, dass man die Diskrepanzen, die es in der Wahrnehmung zwischen spektral aufgelösten und nicht aufgelösten Stimuli gibt, mit einem einzigen Mechanismus zur Extraktion der F0 nachbilden kann. Diese Simulation mit IRNs mit entsprechendem  $\Delta F_0$  und verschiedenen Abschwächungsfaktoren durchzuführen, könnte vielleicht helfen, die Ursache zu klären, warum sich die F0 DLs durch Verringern der Ausprägtheit der Tonhöhe nicht erhöht haben.



# Kapitel 6

## Periodizitätsdetektion: Ein Mechanismus oder zwei?

In dieser Arbeit wurden Experimente dargestellt, die sich zum einen mit der Wahrnehmung von Amplitudenmodulation befassen, zum anderen mit der Wahrnehmung von Tonhöhe. Bis auf das Experiment aus Kapitel 4, haben alle Versuche eines gemeinsam: In allen Versuchen ging es letztendlich um die Erfassung von Periodizität. Amplitudenmodulierte Stimuli haben, in der Regel, eine periodische Hüllkurve bzw. Einhüllende. Bei den tonalen IRN-Stimuli ist der Träger periodisch. Kapitel 2 wirft bereits die Frage auf: Gibt es möglicherweise einen einzigen Mechanismus, der Periodizität sowohl in der Einhüllenden als auch im Träger eines Signals erfassen kann?

Die Ergebnisse haben zumindest gezeigt, dass die Wahrnehmung dieser beiden Formen von Periodizität nicht unabhängig voneinander ist. Gäbe es zwei völlig unabhängige Mechanismen, würde man keine Interaktion erwarten. Es hat sich aber auch eine deutliche Asymmetrie, im Ausmaß wie sich Träger- und Einhüllendenperiodizität gegenseitig in ihrer Wahrnehmung beeinflussen, heraus gestellt. Die Bildung von Tonhöhe sowie deren Ausprägtheit spielen dabei eine entscheidende Rolle. Die Bildung von Tonhöhe ist ein aktiver Prozess. Abgesehen von Sinusreintönen, die in natürlicher Umgebung nur äußerst selten vorkommen, muss von einem komplexen Signal erst einmal die Grundfrequenz  $F_0$  extrahiert werden, welche die Tonhöhe bestimmt. Die Autokorrelationsfunktion ist ein mathematisches Werkzeug, um die Periodizität eines Signals zu erfassen. Elektrophysiologische Messungen am Hörnerv der Katze (Cariani und Delgutte, 1996a,b; Horst et al., 1986) haben gezeigt, dass eine Autokorrelation des zeitlichen Aktivitätsmusters im Hörnerv eine  $F_0$  errechnet, die zutreffend für eine Vielzahl an Teststimuli ist. Da der Hörnerv diese wichtigen Informationen enthält, kann man davon ausgehen, dass diese auch genutzt werden.

## 6.1 Ein physiologisch basiertes Tonhöhenmodell

Der *Nucleus Cochlearis*, CN, ist der einzige Kern der aufsteigenden Hörbahn, der direkten Input vom Hörnerv bekommt. Die Chop-S-Neurone dieses Kerns haben sich als geeignet erwiesen, die Periodizität im Antwortmuster des Hörnervs zu erfassen (Wiegrebe und Winter, 2001; Winter et al., 2001). Wiegrebe und Meddis (2004) haben ein Computer-Modell entwickelt, das diese Zellen zur Tonhöhenbestimmung mit einbezieht. Das Modell beinhaltet eine detaillierte Implementation der auditorischen Peripherie. Die Aktivität des Hörnervs wird durch die Chop-S-Zellen evaluiert. In ihrem Antwortmuster zeigen sie eine bevorzugte intrinsische Oszillation (*chopping rate*, CR), die unabhängig von der jeweiligen Anregungsfrequenz (*characteristic frequency*, CF) ist. Die Autoren verwenden ein Computer-Modell der Chop-S Neurone, das von Hewitt et al. (1992) entwickelt wurde. Das Modell von Wiegrebe und Meddis (2004) fasst die Zellen zu Chop-S-Populationen zusammen. Dabei bilden Neurone, gleicher CR aber unterschiedlicher CF, eine Population. Die entscheidende Eigenschaft dieser Neurone ist die Tatsache, dass sie ihre Aktivität mit der Periodizität bzw. der F0 ihres Inputs synchronisieren, sofern diese im Bereich ihrer CR liegt. Außerdem findet diese Synchronisation nicht nur bei den Zellen statt, die spektral auf die F0 abgestimmt sind, sondern auch bei denjenigen, die spektral auf ein ganzzahliges Vielfaches der F0 abgestimmt sind. Diese selektive Synchronisation wird in „*CR-locked rate*“, die im Folgenden als Synchronisationsrate bezeichnet wird, gemessen. Die Synchronisationsrate ist das ausschlaggebende Maß dafür, welche Tonhöhe und welche Ausgeprägtheit der Tonhöhe (PS) das Modell einem Stimulus zuordnet. Dieses Modell ist in der Lage, diverse Phänomene der Tonhöhenwahrnehmung zu simulieren.

### 6.1.1 Die Ausgeprägtheit von Tonhöhe in simulierten Chop-S Neuronen

**Auswirkungen auf die Unterscheidungsschwellen für spektral aufgelöstes und spektral nicht aufgelöstes IRN:**

Houtsma und Smurzynski (1990) haben Unterscheidungsschwellen (*difference limen*, DL) für Harmonische Komplexe (HC) mit elf Komponenten gemessen. Die Ordnungszahl der ersten physikalisch vorhandenen Harmonischen wurde dabei ständig erhöht. Harmonische mit einer Ordnungszahl größer elf werden von der Peripherie nicht mehr spektral aufgelöst und bewirken eine deutlich schwächer ausgeprägte Tonhöhenempfindung als Komplexe mit Harmonischen niedriger Ordnungszahlen, die spektral aufgelöst werden können. Da die Tonhöhe eines HC durch die Grundkomponente (F0) bestimmt wird, beziehen sich auch die Unterscheidungsschwellen darauf; gemessen werden also F0 DLs. Die Ergebnisse von Houtsma und Smurzynski (1990) haben gezeigt, dass die F0 DLs für die HCs mit niedrigen Harmonischen kleiner waren als für HCs mit hohen Harmonischen. Diese Tatsache, dass spektral aufgelöste Komplexe eine deutlich stärker ausgeprägte Tonhöhe produzieren als spektral nicht aufgelöste, werten die Autoren als Ursache für die unterschiedlichen Schwellen. Die Chop-S-Neurone im Modell von Wiegrebe und Meddis (2004) zeigen für einen HC

mit den Komponenten h1-h11 eine Synchronisationsrate (SR), die knapp drei mal so hoch ist, wie für einen HC mit den Komponenten h16-h26. Somit sind die Simulationsergebnisse mit denen von Houtsma und Smurzynski (1990) im Einklang.

Plack und Carlyon (1995) haben F0 DLs von HCs in Abhängigkeit von der Signaldauer gemessen. Auch sie unterschieden dabei nach spektral aufgelösten und nicht aufgelösten HCs. Auch hier waren die Schwellen höher wenn die Komponenten eine hohe Ordnungszahl besaßen, also spektral nicht aufgelöst werden konnten. Darüberhinaus hat sich in diesem Fall auch die Signaldauer stärker auf die Unterscheidungsschwellen ausgewirkt, als bei den spektral aufgelösten Komplexen. Die Autoren dieser Arbeit werteten diesen Befund als Hinweis darauf, dass zwei unterschiedliche Mechanismen der F0-Extraktion existieren: Einer für spektral aufgelöste und ein anderer für spektral nicht aufgelöste Komplexe. Das Ziel der Versuche aus Kapitel 5 war es, zu klären, ob nicht eher die Ausprägtheit der Tonhöhe (*pitch strength*, PS) die Ursache für die Ergebnisse von Plack und Carlyon (1995) sein könnte. So wurden F0 DLs von IRN in Abhängigkeit der Signaldauer gemessen, da IRN die Möglichkeit bietet, die PS selektiv und stufenlos zu variieren. Im Einklang mit den anderen beiden Arbeiten waren auch in diesem Experiment, bei größtmöglicher PS, die F0 DLs für spektral aufgelöstes IRN geringer, als diejenigen für spektral nicht aufgelöste. In einer zweiten Messung wurde die PS der spektral aufgelösten IRN-Stimuli auf ein Maß reduziert, das der PS spektral nicht aufgelöstem IRN entsprach. Unter der Annahme, dass die PS Grund für die Divergenz der Schwellen ist, wäre die Erwartung folgende gewesen: Ist die Tonhöhe ähnlich schwach ausgeprägt wie bei einem spektral nicht aufgelösten IRN, sind auch die F0 DLs ähnlich hoch. Die Ergebnisse haben aber gezeigt, dass dies nicht der Fall ist. Stattdessen haben sich die Schwellen gegenüber der maximalen PS kaum verändert. Kann nun ein solches Ergebnis mit den simulierten Chop-S Neuronen rekonstruiert werden?

Wenn ein HC aus Harmonischen zusammengesetzt ist, deren Ordnungszahlen höher sind als fünfzehn, ist die Bandbreite der angeregten auditorischen Filter derart hoch, dass am Filterausgang die Periodizität lediglich in der Einhüllenden konserviert ist. In diesem Fall ist die SR der simulierten Chop-S Neurone deutlich verringert (siehe oben). Die Ordnungszahlen der Frequenzmaxima (*ripple*) der spektral nicht aufgelösten IRN-Stimuli waren ebenfalls höher als fünfzehn.

Die PS von IRN kann auf zwei verschiedene Arten verändert werden. Eine Möglichkeit besteht darin, die Anzahl der Iterationen zu verändern. Je höher die Anzahl der Iterationen, desto stärker ausgeprägt ist die Tonhöhe. Proportional zur Anzahl der Iterationen, ändert sich auch die SR der simulierten Chop-S Neurone. Mathematisch gesehen, wird die PS von IRN durch den Korrelationsgrad des ersten Maximums in der Autokorrelationsfunktion (AC1) bestimmt (siehe Abschnitt 1.1.1). Dieser konvergiert mit steigender Iterationszahl gegen eins. Die zweite Möglichkeit, die PS von IRN zu verändern, ist mittels des Abschwächungsfaktors. Reduziert man den Abschwächungsfaktor von eins gen Null, reduziert sich proportional auch die PS. Da sich eine Veränderung des Abschwächungsfaktors gleichermaßen auf die AC1 auswirkt wie die Anzahl der Iteration, ist davon auszugehen, dass auch die SR der simulierten Chop-S Neurone sich proportional zum Abschwächungsfaktor verhält.

Die PS der IRN-Stimuli mit verringertem Abschwächungsfaktor lag in diesem Experiment ungefähr zwischen 0,3 und 0,5 AU (*arbitrary units*). Das Modell errechnet für diese

Werte eine SR von ca 13-16. Für einen HC mit den Komponenten h16-h26 errechnet das Modell eine SR von ca 10. Ist die Periodizität nur in der Einhüllenden vorhanden, ist das SR also geringer als für IRN, dessen Periodizität zwar im Träger vorhanden (da spektral aufgelöst), aber abgeschwächt ist. Insofern wäre die Vorhersage des Modells qualitativ richtig, denn es würden geringere F0 DLs für IRN mit verringertem Abschwächungsfaktor vorraussagen.

Jedoch implizieren die Ergebnisse, dass kein linearer Zusammenhang zwischen der PS und den F0 DLs besteht. Die Diskussion in Kapitel 5 weist bereits darauf hin, dass sich der Zusammenhang asymptotisch verhalten könnte. In weiterführenden psychoakustischen Messungen wäre es sinnvoll, F0 DLs für IRN-Stimuli in Abhängigkeit der PS zu messen, um deren Zusammenhang zu klären. Darüberhinaus wäre es sinnvoll, mit dem Modell zu versuchen, die gemessenen F0 DLs zu rekonstruieren. Dazu müsste man es noch mit einem Entscheidungsmodul ergänzen. Denkbar wäre auch hier ein optimaler Detektor (Dau et al., 1996). Das Konzept der Kreuzkorrelation mit einer internen Repräsentation des überschwelligen Signals hat sich auch in den Simulationen zur Interaktion von Träger- und Einhüllendenperiodizität (Kapitel 2) als vorteilhaft erwiesen.

### **Auswirkungen für die Interaktion von Träger- und Einhüllendenperiodizität:**

In Kapitel 2 wurde gezeigt, dass Trägerperiodizität die Detektionsschwellen für Amplitudenmodulation (AM) im Tonhöhenwahrnehmungsbereich erhöht. Im umgekehrten Fall hat eine zusätzliche AM die Detektionsschwellen für Trägerperiodizität in diesem Bereich nicht beeinflusst. Da die Tonhöhe das wesentliche Entscheidungskriterium für die Probanden darstellte, ist es plausibel, dass die relativ stark ausgeprägte Tonhöhe von IRN die relativ schwach ausgeprägte Tonhöhe von SAMN effektiver maskiert hat. Eine Modellsimulation für die Tonhöhe von sinusförmig-amplitudenmoduliertem Rauschen (SAMN) sagt die richtige Tonhöhe voraus, die SR ist aber auch hier nur schwach ausgeprägt, genauso, wie auch die Ausgeprägtheit der wahrgenommenen Tonhöhe bei SAMN nur schwach ist. Da bei SAMN die Periodizität in der Einhüllenden liegt, ist dies ein weiterer Hinweis darauf, dass Einhüllendenperiodizität generell eine niedrigere SR zur Folge hat als Trägerperiodizität. Das wiederum bekräftigt die Annahme, dass Tonhöhe und deren Ausgeprägtheit, das wesentliche Entscheidungskriterium war. Gleichzeitig bietet dies eine Erklärung für die Asymmetrie der Interaktion.

### **6.1.2 Periodizitäten unterhalb der Tonhöhengrenze**

Für IRN F0s bzw. Modulationsfrequenzen unterhalb der Tonhöhengrenze (*lower limit of pitch*: Krumbholz et al. (2000); Pressnitzer et al. (2001)) wurde die Detektion von Trägerperiodizität durch zusätzliche AM ebenso beeinträchtigt wie die Detektion von AM durch Trägerperiodizität. In diesem niedrigen Frequenzbereich wird anstelle einer Tonhöhe ein repetitives Element wahrgenommen. Dabei kann leicht verwechselt werden, ob dieses vom IRN oder der Modulation stammt. Je niedriger die Frequenz, desto genauer kann man den Schwankungen im Signal folgen. In diesem niedrigen Frequenzbereich scheint es eher so zu

sein, dass eine Art zeitliches Fenster das Signal über dessen Verlauf abtastet und die Beobachtungen in einer Art kurzzeitigen Gedächtnisspeicher sammelt, wie es z.B. das „multiple looks“ Modell von Viemeister und Wakefield (1991) vorschlägt. Maßgeblich sind daher die Veränderungen des Signals anstelle einer stationären Empfindung.

In Kapitel 3 wurden die Detektionsschwellen einer Signalmodulation in Gegenwart einer maskierenden AM gemessen. Die Modulationsfrequenzen lagen unterhalb der Tonhöhen- grenze. Die Maskierer besaßen alle die gleiche Periodizität und teilweise das gleiche Modula- tionsspektrum. Der zeitliche Verlauf der Wellenform war bei allen Maskierern verschieden. Der wesentliche Unterschied in der Wellenform war die Länge der negativen Halbwelle, also die Breite des Tals, in dem der Träger abgeschwächt ist. Verändert man bei einer Rechteckfunktion das Tastverhältnis, verändert man somit das Verhältnis von Berg zu Tal.

Die Ergebnisse haben gezeigt, dass bei der Detektion einer Signalmodulation in Ge- genwart einer maskierenden AM eine Kombination aus lokaler Modulationstiefe, also dem Grad der Abschwächung, und Dauer der anhaltenden Modulationstiefe, entsprechend der Breite der Täler, ausschlaggebend ist. Bei einer Stimuluskonfiguration betrug die Frequenz der Signalmodulation 128 Hz und produzierte somit eine wahrnehmbare Tonhöhe. Zwar war auch in dieser Situation eine Kombination aus lokaler Modulationstiefe und deren anhalten- der Dauer ausschlaggebend für die Detektionsschwellen, dennoch sprechen die Ergebnisse für eine Trennung von Signal und Maskierer in unterschiedliche auditorische Objekte.

Es ist also wahrscheinlich, dass es zwei Mechanismen der Periodizitätskodierung gibt, jedoch erfolgt die Teilung eher anhand der Periodendauer, die dem Reziproken der Frequenz entspricht, als anhand dessen, ob die Periodizität in der Einhüllenden oder im Träger vorhanden ist.

### 6.1.3 Ein Tonhöhenmechanismus

Für den Bereich der Tonhöhenwahrnehmung kann ein Mechanismus zur Periodizitätsko- dierung durchaus ausreichend sein. Das hat die Diskussion in Anbetracht des Modells von Wiegrebe und Meddis (2004) gezeigt. Dieses Modell ist in der Lage, die Frequenz der Tonhöhe sowie deren Ausprägtheit qualitativ korrekt zu erfassen. Das gilt für Tonhö- he, bedingt durch Einhüllendenperiodizität wie bei SAMN, als auch für Tonhöhe, bedingt durch Trägerperiodizität, wie bei IRN. Ebenso werden die Unterschiede zwischen spektral aufgelösten und spektral nicht aufgelösten komplexen Tönen qualitativ korrekt simuliert. Die Autoren selbst weisen auf noch bestehende Schwachstellen hin: Das ist vor allem die Tatsache, dass kaum Neurone gefunden werden, deren CR niedriger als 100 Hz oder grö- ßer 500 Hz ist. Der Tonhöhenwahrnehmungsbereich für komplexe Töne liegt aber zwischen 33 und 1400 Hz. Die Chop-S Zellen können allerdings auch die Subharmonischen von kom- plexen Tönen mit höherer F0 erfassen, indem sie mit einer Periodendauer synchronisieren, die einem ganzzahligen Vielfachen der Stimulusperiode entspricht. Oktavenverwechslungen kommen bei der Beurteilung von Tonhöhe sehr häufig vor. Es wird in der Arbeit diskutiert, dass dieses Modellverhalten eventuell das neuronale Korrelat für diese Verwechslungen ist. Darüber hinaus bezieht das Modell keinerlei binaurale Tonhöhenphänomene mit ein. Man

muss aber bedenken, dass man sich im CN auf einer sehr niedrigen Stufe der aufsteigenden Hörbahn befindet.

## 6.2 Kodierung der Einhüllenden

Unabhängig von der Tonhöhenwahrnehmung gibt es Neurone die speziell die Einhüllende eines Stimulus kodieren. Das haben verschiedene elektrophysiologische Arbeiten gezeigt (z.B. Langner (1992), Joris und Yin (1995) oder Grothe et al. (2001)). In der Arbeit von Grothe et al. (2001) werden Modulations-sensitive Neurone mit phasischem Antwortmuster gezeigt, die die Fähigkeit zum „phase-locking“ besitzen. Das bedeutet, dass sie ihr Antwortmuster mit der Phase des Modulators synchronisieren können. Da der Bereich der Modulationsfrequenzen, der von diesen Neuronen erfasst wird, im Bereich zwischen 50 und 400 Hz liegt, ist es wenig wahrscheinlich, dass sie in der Lage sind, AM-Frequenzen, die tiefer als 30 Hz sind, zu kodieren. Allerdings wäre es durchaus möglich, dass mit Hilfe dieser Art von Neuronen Rauigkeit erfasst werden kann. Die Detektionsschwellen für Sparse Noise (SN) haben einen engen Zusammenhang mit dem dekadischen Logarithmus des vierten Moments der Stimuli gezeigt. Dieses Maß gibt Auskunft über die physikalische Rauigkeit eines Stimulus, die sich wiederum auf das Modulationsspektrum auswirkt. Die maximale Rauigkeit im psychoakustischen Sinne wird mit Modulationsfrequenzen um die 70 Hz erreicht (Zwicker und Fastl, 1999). Wobei dieser neue Stimulus, SN (Hübner und Wiegrebe, 2003), möglicherweise durch weiterführende Verhaltenstests neue Aspekte zur psychoakustischen Rauigkeit beitragen kann. Zunächst aber könnte man die Rauigkeit von SN mit einem psychoakustischen Modell berechnen lassen. Die Ergebnisse könnten gezieltere Hinweise auf weiterführende Versuche geben.

Auf jeden Fall liegen die Modulationsfrequenzen, die die Modulations-sensitiven Neurone in der Arbeit von Grothe et al. (2001) erfassen, in einem für Rauigkeit relevanten Bereich. Insofern ist es plausibel, diese Neurone als mögliches neuronales Substrat für dessen Kodierung in Betracht zu ziehen. Bislang wurden diese phasisch antwortenden Zellen nur mit Stimuli gereizt, deren Einhüllende periodisch war. Die Einhüllende von SN dagegen ist aperiodisch. Es wäre sicher interessant, mittels elektrophysiologischer Methoden die Reaktion von Modulations-sensitiven Neuronen auf Stimuli wie SN mit einem breitbandigen Modulationsspektrum zu untersuchen.

# Literaturverzeichnis

- Aures, W. (1985). Ein Berechnungsverfahren der Rauigkeit. *Acustica*, 58:268–281.
- Bilsen, F. A. und Ritsma, R. J. (1969). Repetition pitch and its implication for hearing theory. *Acoustica*, 22:63–68.
- Burns, E. und Viemeister, N. (1976). Nonspectral pitch. *J. Acoust. Soc. Am.*, 60:863–868.
- Burns, E. M. (1981). Played-again sam: Further observation on the pitch of amplitude-modulated noise. *J. Acoust. Soc. Am.*, 70:1655–1660.
- Buus, S. und Florentine, M. (1985). *Time Resolution in Auditory Systems*, chapter Gap detection in normal and impaired listeners: The effect of level and frequency, pages 159–179. Springer Verlag, New York.
- Cariani, P. A. und Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.*, 76:1698–1716.
- Cariani, P. A. und Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift and pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J. Neurophysiol.*, 76:1717–1734.
- Carlyon, R. P. und Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *J. Acoust. Soc. Am.*, 95(6):3541–3554.
- Dau, T., Kollmeier, B., und Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.*, 102(5):2892–2905.
- Dau, T., Kollmeier, B., und Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration. *J. Acoust. Soc. Am.*, 102(5):2906–2919.
- Dau, T., Püschel, D., und Kohlrausch, A. (1996). A quantitative model of the „effective“ signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.*, 99(6):3615–3622.

- Eddins, D. A. (1993). Amplitude-modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region. *J. Acoust. Soc. Am.*, 93:470–479.
- Eddins, D. A. (1999). Amplitude-modulation detection at low- and high-audio frequencies. *J. Acoust. Soc. Am.*, 105:829–837.
- Eddins, D. A., Hall, J. W., und Grose, J. H. (1992). Detection of temporal gaps as a function of frequency region and absolute noise bandwidth. *J. Acoust. Soc. Am.*, 91:1069–1077.
- Ewert, S. (2002). *Auditory spectro-temporal processing in the envelope-frequency domain: Experiments and models*. Dissertation, Carl von Ossietzky Universität, Oldenburg. Online-Version im Internet unter: <http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2003/eweaud02/pdf/eweaud02.pdf>.
- Ewert, S. und Dau, T. (2000). Characterizing frequency selectivity for envelope fluctuations. *J. Acoust. Soc. Am.*, 108(5):1181–1196.
- Ewert, S., Verhey, J., und Dau, T. (2002). Spectro-temporal processing in the envelope-frequency domain. *J. Acoust. Soc. Am.*, 112:2921–2931.
- Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys.*, 12:47–65.
- Formby, C. (1985). Differential sensitivity to tonal frequency and to rate of amplitude modulation of broadband noise by normally hearing listeners. *J. Acoust. Soc. Am.*, 78:70–77.
- Forrest, T. G. und Green, D. M. (1987). Detection of partially filled gaps in noise and the temporal modulation transfer function. *J. Acoust. Soc. Am.*, 82(6):1933–1943.
- Frisina, R. D., Smith, R. L., und Chamberlain, S. C. (1990a). Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. *Hearing Research*, 44:99–122.
- Frisina, R. D., Smith, R. L., und Chamberlain, S. C. (1990b). Encoding of amplitude modulation in the gerbil cochlear nucleus: II. Possible neural mechanisms. *Hearing Research*, 44:123–144.
- Grothe, B., Covey, E., und Casseday, J. H. (2001). Medial superior olive of the big brown bat: neural responses to pure tones, amplitude modulations, and pulse trains. *J. Neurophysiol.*, 86(5):2219–2230.
- Grunwald, J.-E., Schörnich, S., und Wiegrebe, L. (2004). Classification of natural textures in echolocation. *Proc. Nat. Acad. Sci.*, in press.
- Hartmann, W. M. und Pumplin, J. (1988). Noise power fluctuations and the masking of sine signals. *J. Acoust. Soc. Am.*, 83:2277–2289.

- Hübner, M. und Wiegrebe, L. (2003). The effect of temporal structure on rustling-sound detection in the gleaning bat, *Megaderma lyra*. *J. Comp. Physiol. A*, 189:337–346.
- Helmholtz, H. L. F. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. F. Vieweg, Braunschweig.
- Hewitt, M. J., Meddis, R., und Shackleton, T. M. (1992). A computer model of a cochlear-nucleus stellate cell: Responses to amplitude-modulated and pure-tone stimuli. *J. Acoust. Soc. Am.*, 91(4):2096–2109.
- Horst, J. W., Javel, E., und Farley, G. R. (1986). Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms. *J. Acoust. Soc. Am.*, 79:398–416.
- Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *J. Acoust. Soc. Am.*, 85:1676–1680.
- Houtsma, A. J. M. und Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.*, 87(1):304–31.
- Joris, P. X. und Yin, T. C. (1995). Envelope coding in the lateral superior olive of cat. I. Sensitivity to interaural time differences. *J. Neurophysiol.*, 73:1043–1062.
- Krumbholz, K., Patterson, R. D., und Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *J. Acoust. Soc. Am.*, 108(3):1170–1180.
- Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, 60:115–142.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 49:467–477.
- Licklider, J. C. R. (1951). A duplex theory of pitch and modulation perception. *Experientia*, 7:128–133.
- Meddis, R. und Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I:Pitch identification. *J. Acoust. Soc. Am.*, 89(6):2866–2882.
- Meddis, R. und Hewitt, M. J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II:Phase sensitivity. *J. Acoust. Soc. Am.*, 89(6):2883–2894.
- Meddis, R. und O'Mard, L. (1997). A unitary model of pitch perception. *J. Acoust. Soc. Am.*, 102(3):1811–1820.
- Münkner, S. (1993). *Modellentwicklung und Messungen zur Wahrnehmung nicht stationärer Signale*. Dissertation, Georg-August Universität, Göttingen.

- Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing*. Academic Press.
- Neuweiler, G. (1993). *Biologie der Fledermäuse*. Georg Thieme Verlag, Stuttgart.
- Patterson, R. D. und Moore, B. C. J. (1986). *Frequency Selectivity in Hearing*, chapter Auditory filters and excitation patterns as representations of frequency resolution. Academic Press, London.
- Patterson, R. D., Robinson, K. D., Holdsworth, J., McKeown, D., Zhang, C., und Allershand, M. (1992). *Auditory Physiology and Perception*, chapter Complex sounds and auditory images. Pergamon, Oxford.
- Plack, C. J. und Carlyon, R. P. (1995). Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *J. Acoust. Soc. Am.*, 98(3):1355–1364.
- Pressnitzer, D., Patterson, R., und Krumbholz, K. (2001). The lower limit of melodic pitch. *J. Acoust. Soc. Am.*, 109:2074–2084.
- Rhode, W. S., Oertel, D., und Smith, P. H. (1983). Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus. *J. Comp. Neurol.*, 213:448–463.
- Schulze, H. und Langner, G. (1997). Periodicity coding in the primary auditory cortex of the Mongolian Gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm. *J. Comp. Physiol. A*, 181:651–663.
- Shackleton, T. M. und Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.*, 95(6):3529–3540.
- Shailer, M. J. und Moore, B. C. J. (1983). Gap detection as a function of frequency, bandwidth and level. *J. Acoust. Soc. Am.*, 74:467–473.
- Shailer, M. J. und Moore, B. C. J. (1985). Detection of temporal gaps in band-limited noise: effects of variations in bandwidth and signal-to-masker ratio. *J. Acoust. Soc. Am.*, 77:635–639.
- Smith, Z. M., Delgutte, B., und Oxenham, A. J. (2002). Chimeric sounds reveal dichotomies in auditory perception. *Nature*, 416:87–90.
- Snell, K. B., Ison, J. R., und Frisina, D. R. (1994). The effects of signal frequency and absolute bandwidth on gap detection in noise. *J. Acoust. Soc. Am.*, 96(3):1458–1464.
- Srulovicz, P. und Goldstein, J. L. (1983). A central spectrum model: A synthesis of auditory nerve timing and place cues in monaural communication of frequency spectrum. *J. Acoust. Soc. Am.*, 73:1266–1276.

- Terhardt, E. (1998). *Akustische Kommunikation*. Springer Verlag, Berlin, Heidelberg.
- Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *J. Acoust. Soc. Am.*, 66:1364–1380.
- Viemeister, N. F., Stellmack, M. A., und Byrne, A. J. (2004). *Auditory signal processing: Physiology, psychoacoustics and models*, chapter The role of temporal structure in envelope processing. in press.
- Viemeister, N. F. und Wakefield, G. H. (1991). Temporal integration and multiple looks. *J. Acoust. Soc. Am.*, 90(2 Pt.1):858–865.
- Wiegrefe, L. (2001). Searching for the time constant of neural pitch extraction. *J. Acoust. Soc. Am.*, 109:1082–1091.
- Wiegrefe, L. und Meddis, R. (2004). The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus. *J. Acoust. Soc. Am.*, 115(3):1207–1218.
- Wiegrefe, L. und Patterson, R. (1999). The role of envelope modulation in spectrally unresolved iterated rippled noise. *Hearing Research*, 132:94–108.
- Wiegrefe, L. und Winter, I. M. (2001). Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus. *J. Neurophysiol.*, 85:1206–1219.
- Winter, I. M., Wiegrefe, L., und Patterson, R. D. (2001). The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. *J. Physiol (Lond)*, 537(2):553–566.
- Yin, T. C. und Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. I. Neural mechanisms. *J. Neurophysiol.*, 65:465–488.
- Yost, W. (1996a). Pitch of iterated rippled noise. *J. Acoust. Soc. Am.*, 100(1):511–518.
- Yost, W. (1996b). Pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.*, 100(5):3329–3335.
- Zwicker, E. und Fastl, H. (1999). *Psychoacoustics - Facts and Models*. Springer Verlag, second updated edition.



# Lebenslauf

## Persönliche Daten

Alexandra Stein  
Guldeinstr. 31  
80339 München

Tel.: (089) 54 84 78 11

e-Mail: stein@zi.biologie.uni-muenchen.de

Geb. am 18.06.1970 in Wichita Falls, Texas, USA

Mutter eines Sohnes

## Schulbildung

1976	Victor Valley Christian School in Victorville, Ca., USA
1977–1980	Finkenburg Grundschule in Wittmund
1980–1983	Our Lady of the Desert School in Apple Valley, Ca., USA
1983–1985	Alexander-von-Humboldt Gesamtschule in Wittmund
1985–1991	Apian-Gymnasium in Ingolstadt
01.07.1991	allg. Hochschulreife

## Studium

WS 1991/92–WS 1999/2000 Studium der Biologie an der LMU München

- Hauptfach: Neurobiologie
- Nebenfächer: Zoologie, Systematische Botanik, Ökologie

April–Juli 1997 Forschungsaufenthalt am HYDRA-Institut für Meereswissenschaften, Fetovaia, Elba, Italien

08/1997–11/1998 Babypause

06/1999–03/2000 Diplomarbeit am Institut für Systematische Botanik der LMU München

- Betreuer: Prof. Dr. Jürke Grau
- Thema: Die Makrophytenverteilung auf den Coralligène-Blöcken an der Punta di Fetovaia, Elba, Italia

01.03.2000 Erlangung des Diploms

08/2000–04/2004 Promotion bei Prof. Dr. Dr. h.c. G. Neuweiler.

- Thema: Funktionelle Mechanismen der auditorischen Kodierung für Träger- und Einhüllendenperiodizität beim Menschen

August 2001 Forschungsaufenthalt am CNBH, Department of Physiology, University of Cambridge, U.K.

# Publikationen

- Stein, A., Ewert, S. und Wiegrebe, L. (2003). Interaction between pitch and modulation perception in broadband stimuli: a test for the autocorrelation and the modulation filterbank model. *J. Acoust. Soc. Am.*, in review.
- Wiegrebe, L., Stein, A. und Meddis, R. (2004). Coding of pitch and amplitude modulation in the auditory brainstem: one common mechanism? in *Proceedings of the International Symposium in Hearing*, edited by D. Pressnitzer, A. de Cheveigne, S. MacAdams and L. Collet.



# Danksagung

Zuerst möchte ich mich bei **Herrn Prof. Dr. G. Neuweiler** für die Überlassung des Themas bedanken. Vielmehr noch möchte ich diese Gelegenheit nutzen, mich für die gute Lehre während meiner Studienzzeit zu bedanken. Er hat mich als Studentin dazu ermutigt, nicht alles was man hörte als Gegeben zu betrachten sondern eigene Gedanken zu entwickeln, und, viel wichtiger, diese Gedanken auch zu äußern, denn man fühlte sich immer ernst genommen.

Bei **Herrn Dr. habil. Lutz Wiegrebe** möchte ich mich dafür bedanken, dass er mich in seine Arbeitsgruppe aufgenommen hat und für die Betreuung meiner Arbeit, für die vielen Hilfestellungen und Diskussionen. Nicht zu vergessen diverse Raclett-, Fondue- und Videoabende, die er für „seine Jungs und Mädels“ immer wieder veranstaltet hat.

Um bei der Arbeitsgruppe zu bleiben: **Petra Weissenbacher** danke ich für ihr stets offenes Ohr und die Abwechslung an tristen Tagen. Bei **Jan Grunwald** muss ich mich für eine ganze Menge bedanken. Für die unglaubliche Hilfsbereitschaft, die er immer und in allen Belangen an den Tag legt. Vor allem für das unermüdliche Beantworten vieler sich wiederholender dummer Fragen und die nette Gesellschaft, die vielen Gespräche über Fachliches und gar nicht Fachliches. Aber an Hilfsbereitschaft mangelte es in unserer Arbeitsgruppe sowieso nicht. **Mathias Hübner, U.D.O. Gröger, Maike Schuchmann** und **Sven Schörnich** haben sich ganz freiwillig (oder??) und höchst uneigennützig, da unbezahlt, als Versuchspersonen zur Verfügung gestellt, wie auch **Tina Stein, Axel Schulke** und **Franky Borina**. Ganz lieben Dank dafür! Dies gilt natürlich auch allen anderen „Hörern“ die ich nicht namentlich aufführe.

**Markus Drexl** war so lieb mich von seinen frisch erworbenen LaTeX Fähigkeiten profitieren zu lassen (Sowie auch der Jan, der mir nie böse war, dass er immer seine Bücher von meinem Schreibtisch wieder einsammeln musste).

Für eine tolle Atmosphäre mit viel Spaß sorgten abwechselnd alle am Lehrstuhl: Andi, Ariane, Britta, Claudia, Dieter, Elisabeth, Frank, Jan, Jenny, Lutz, Maike, Manfred, Markus, Mathias, Petra, Rudi (der Berg/das Balan ruft!), Sven, Silvana, Tom, Udo, Uwe, Vroni. Wirklich eine ausnehmend lustige und angenehme Arbeitsgruppe!

**Stephan Ewert** danke ich für die tolle Zusammenarbeit und den äußerst hilfreichen Diskussionen und Kommentaren zum Manuskript. Auch er hatte immer ein offenes Ohr für meine Fragen, und derer gab es viele, vor allem in physikalischen Belangen. Vielen Dank auch für die vielen aufmunternden mails, wenn sich mal wieder Verzweiflung breit gemacht hatte.

Stets mit Rat und Tat zur Seite stand mir **Anneliese Heigl**. Ihre Hilfe war wirklich von unschätzbarem Wert, vor allem die viele und liebe Betreuung von Yannick. Was wäre die Welt ohne Großeltern - oder Eltern - **Gerd** und **Waltraud Stein**, meine Eltern, haben mich (und damit auch Yannick), wie immer eigentlich, mit all ihren Mitteln und Möglichkeiten gerne unterstützt. Ich hoffe sie wissen auch ohne viele Worte, dass ich ihnen unendlich dankbar bin. An dieser Stelle muss ich auch meine Schwester **Tina** nochmal erwähnen. Denn meine ganze Familie sehr dazu beigetragen, dass ich immer wieder einen freien Kopf für meine Arbeit haben konnte.

**Christiane Schroth** und **Albrecht Kühn** möchte ich für ihre wertvollen Tipps rund um die Regeln der deutschen Schriftsprache danken. All meinen Freunden danke ich für ihr Interesse an meiner Arbeit und für ihren Glauben an mich und meine Fähigkeiten als, wie Dani so schön formulierte: Doktorandinmutterköchinholundbringdienstkrankenschwestersofortrichtigeentscheidungentreffenmüsserin..... Das hat mir sehr geholfen in den „harten Zeiten“ durchzuhalten und nicht den Mut zu verlieren. Die schönen Stunden der Abwechslung haben ebenfalls ihren Beitrag dazu geleistet.

# Ehrenwörtliche Erklärung

Die Arbeit wurde von mir selbstständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt.