

**Neuronal Correlates of Comodulation Masking  
Release at the level of the Inferior Colliculus in  
the context of Spectro-temporal Receptive Fields  
and their corresponding Volterra Operators**

**Dissertation**

zur Erlangung des akademischen Grades

**doctor rerum naturalium (Dr.rer.nat.)**

genehmigt durch die Fakultät für Naturwissenschaften der  
Otto-von-Guericke-Universität Magdeburg



von: Dipl.-Phys. Jan-Philipp Diepenbrock  
geb. am 16.09.1982 in Bremen

Gutachter Prof. Dr. Frank W. Ohl  
Prof. Dr. J. Leo van Hemmen

eingereicht am 19.05.2016  
verteidigt am 15.02.2017

## Thanks

"Unless the Lord builds the house, the builders labor in vain." (Psalm 127, 1)

My supervisor **Prof. Frank W. Ohl** for providing me with the opportunity and support to write this thesis.

**Prof. J. Leo van Hemmen** for support in and introduction to the concept of differential sampling.

**Prof. Jesko L. Verhey** for support and help with comodulation masking release.

**Dr. Marcus Jescke** for teaching me single unit recordings and help with the evaluations.

**Dr. Arne F. Meyer** for support with the STRF evaluation and good collaboration.

# Summary

An important goal in auditory signal analysis is the prediction of unknown stimuli. If the mathematical structures that describe any response of a cell in the auditory pathway were known, this would be an important step not only for research but also for technical applications. The results could be implemented e.g., in any form of device for hearing impaired or for speech recognition algorithms. Some aspects of this vast topic are to be elucidated in this work, which is structured into four chapters. The first two chapters describe a correlate of a psychophysical phenomenon in the inferior colliculus, and point out the importance of the temporal and spectral structure of the used signals. It has been shown that these components do occur in natural signals and that they are important for speech recognition in noise. The second and the third chapter show how the responses presented in the first chapter might be predicted using a spectro-temporal receptive field (STRF). The third chapter gives an overview of how the STRF could be linked to a more general mathematical concept, the Volterra Series. A solid mathematical background of a model helps as it increases the possibility to interpret the results, see the limitations of the model and gain knowledge from the prediction failures. In the fourth chapter a method to infer the components of the Volterra Series is applied for the first time and it is shown how long term interactions in a neuron might be measured.

In the first chapter, one method in psychophysics is introduced that is used to gain insights into auditory detection of masked signals, comodulation masking release (CMR). More precisely, CMR describes the reduction in detection threshold if the masker contains coherent envelope fluctuations. This means that the masker has to fulfill certain temporal and spectral requirements. From the physiological perspective, another problem is finding a definition when a cell might have detected a signal. Neural correlates of CMR have been proposed at different stages of the auditory system. While later stages seem to suppress the response to the masker, earlier stages are more likely to enhance their response to the signal when the masker is comodulated. Using a flanking band masking paradigm, the chapter investigates how CMR is represented at the level of the inferior colliculus (IC) of the Mongolian gerbil. The result is that though elements from earlier stages (an increase in the signal response) are more dominant, the decrease in the masker response is an important feature.

In the second chapter, several aspects of CMR are further elucidated. Results from psychophysics that show how certain spectral properties influence CMR are also found in the physiological correlate of CMR. Further, individual properties of the units representing CMR

are investigated. It is shown how units with certain types of frequency response areas (FRA) are more likely to show CMR than others. The spectro-temporal receptive field (STRF) is introduced, which predicts a unit's response given the time-frequency representation of an incoming stimulus. It is shown what the requirements for the STRF would be in order to show the observed responses to CMR. Finally, the chapter shows how the response to the masker without the signal highly correlates with the unit's ability to show CMR. Since the analysis requires a temporal window of 50 ms, whereas STRFs for the IC usually look at the previous 30 ms, this effect demonstrates a shortcoming of the STRF in the effort to fully predict CMR.

Given the problems of the STRF, the third chapter shows an in-depth investigation of how an STRF might be turned into a Volterra operator. In order to do this, time-frequency representations are introduced as the STRF relies on them. The Volterra Series is introduced as a method that, like a function turns one variable into another, turns one function into another. The proposed method to turn the STRF into the mathematically more generally applicable and better defined and funded Volterra Series relies on the numeric representation of both of them but makes, other than that, as few assumptions as possible. Using some artificial data it is shown when the operations are equivalent and when the results of the STRF can only be approximated by the Volterra Series. The implications of these comparisons are a more thorough understanding of the STRF and the implicit assumptions made when using an STRF.

In the last chapter, measurement methods are applied for a first time that directly investigate the Volterra operator. The name of the method is differential sampling, and using a modified version of it long term interactions are demonstrated that cannot be found in the STRF. The chapter gives the experimental proof of concept of differential sampling as it was used to determine two Volterra operators and use them for an output prediction. It is shown that another benefit of the Volterra series is its interpretability in terms of differential equations and the resulting phase space. While from an experimental point of view the performance of Volterra operators still has to be tested (as it is shown here for the first time how to apply them in auditory neuroscience, no systematic investigation has been done yet), they offer the advantage of being clearly defined at a level, having a known convergence range and, by their more solid mathematical framework, offer more possibilities.



# Zusammenfassung

Einer der Schwerpunkte der auditorischen Forschung ist die Vorhersage des Antwortverhaltens einer Zelle im auditorischen System auf unbekannte Stimuli. Wenn man die dafür benötigten Funktionen kennte, so wäre dies zum einen ein bedeutender Schritt in der Grundlagenforschung, aber zum anderen auch für viele technische Anwendungen wichtig. So könnten z.B. Hörhilfen oder automatische Spracherkennungsalgorithmen verbessert werden, letztere da das Gehör bei vielen Störgeräuschen modernen Prozessoren überlegen ist. Einige Aspekte dieses großen Themenkomplexes werden in dieser Arbeit, die sich in vier Kapitel gliedert, betrachtet. Die ersten beiden Kapitel behandeln ein neuronales Korrelat eines psychophysikalischen Phänomens auf Ebene der Colliculi inferiores (IC). Hierbei sind vor allem die zeitliche und spektrale Struktur des Signals wichtig. Diese Effekte können auch bei natürlichen Signalen gefunden werden und können Einblicke in die Spracherkennung in Störgeräuschen bieten. Im zweiten und dritten Kapitel wird gezeigt, wie die im ersten Kapitel präsentierten Antworten anhand eines sogenannten Spekro-Temporalen rezeptiven Feldes (STRF) vorhergesagt werden könnten, also wie die Antworten mathematische reproduzierbar modelliert werden können. Das dritte Kapitel gibt dabei einen Überblick darüber, wie statt einer Operation mit dem STRF im Zeit-Frequenzraum eine äquivalente im Zeitraum anhand eines Volterra Operators durchgeführt werden kann. Dieses ist wichtig, da die Volterra Reihe ein allgemein Ansatz in der Mathematik ist, der eine parameterfreie Beschreibung der Funktion erlaubt. Eine genaue mathematische Beschreibung ist wiederum wichtig, um die Hintergründe und Limitationen eines Modells besser verstehen zu können und ist besser zu interpretieren wenn die Vorhersage inkorrekt ist. Im vierten Kapitel wird dann zum ersten Mal eine Methode angewandt, die direkt Komponenten des Volterra Operators misst und es wird gezeigt, wie das Zusammenwirken von Stimuli auf längeren Zeitskalen in einer Zelle untersucht werden kann.

Das erste Kapitel beginnt mit der Erklärung eines psychophysikalischen Phänomens, der komodulierten Maskierungsverringerng (CMR). Dabei geht es darum, wie sich ein Signal in einem Maskierer besser detektieren lässt, wenn dieser kohärente Schwingungen der Einhüllenden aufweist. Mit anderen Worten, der Maskierer muss bestimmte zeitliche und spektrale Anforderungen erfüllen. Dabei stellt sich von der physiologischen Seite zunächst die Aufgabe, eine sinnvolle Definition für das Entdecken eines Signals in einem Maskierer für eine Einzelzelle zu finden. Neuronale Korrelate von CMR wurden an verschiedenen Stationen des auditorischen Pfads gefunden, und oft auch unterschiedlich definiert. Generell lässt sich aber sagen, dass spätere Stationen eher die Antwort auf den Maskierer verringern, während frühere

Stationen ihre Antwort auf das Signal verstärken, wenn das Maskierungssignal kohärent moduliert ist. Anhand eines Flankenbandparadigmas wurde im ersten Kapitel untersucht, wie Einzelzellen im Colliculi inferiores der mongolischen Wüstenrennmaus auf ein Zielsignal antworten, wenn der Maskierer komoduliert ist. Es wird gezeigt, dass das Antwortverhalten Elemente aus früheren (Vergrößerung der Signalantwort) und späteren (Verringerung der Maskiererantwort) Ebenen vereint. Dabei ist die Vergrößerung der Signalantwort dominanter als die Verringerung der Maskiererantwort.

Im zweiten Kapitel werden verschiedene weitere Aspekte von CMR betrachtet. Es werden Ergebnisse aus der Psychophysik bestätigt, die die spektralen Eigenschaften von CMR betreffen. Dazu wurde für eine Gruppe von Zellen der spektrale Abstand der Flankenbänder systematisch variiert. Weiter wird gezeigt, dass insbesondere Zellen, deren Frequenzantworten für unterschiedliche Pegel (FRA) eher linear sind, eine höhere Wahrscheinlichkeit haben, ein Antwortverhalten, das CMR entspricht, zu zeigen. Die Linearität bezieht sich dabei darauf, dass die in der FRA gezeigte Antwort mit einem linearen Filter zwischen zwei stationären Nichtlinearitäten erklärt werden könnte. Typische nicht-lineare Zellen wären solche, die für eine bestimmte Frequenz bei einem Pegel eine Antwort zeigen, aber bei der Vergrößerung des Pegels dann nicht mehr reagieren. Es wird das spektro-temporal rezeptive Feld vorgestellt, welches die Antwort einer Zelle auf unbekannte Stimuli vorhersagen soll (wobei es eine linear antwortende Zelle beschreiben würde). Weiter wird gezeigt, dass die Fähigkeit einer Zelle, CMR zu zeigen, mit Eigenschaften ihrer Antwort auf einen bestimmten Maskierer korreliert. Da dieser Effekt auf einer Zeitskala von mindestens 50 ms stattfindet, dass STRF im IC aber auf einer Zeitskala von nur 30 ms bestimmt wird ist das STRF nicht geeignet alle Phänomene der CMR-Repräsentation im IC erklären zu können.

Die Struktur und Hintergründe des STRFs werden im dritten Kapitel untersucht. Dabei wird ein besonderer Schwerpunkt auf den Hintergrund des STRFs gelegt, nämlich die Zeit-Frequenz Darstellung eines Stimulus. Hierbei handelt es sich um eine zwei-dimensionale (Zeit und Frequenz) Darstellung eines eindimensionalen (rein zeitlichen) Signals, was zu einer Abhängigkeit der beiden Dimensionen führt, wodurch sie nicht mehr als unabhängig voneinander betrachtet werden können (anders als zum Beispiel zwei Ortsdimensionen, bei denen man sich entlang einer Achse bewegen kann während die Position auf der anderen Achse konstant bleibt). In dem Kapitel wird gezeigt, wie man mit wenigen, die Allgemeinheit wenig einschränkenden Annahmen, aus einem STRF ein Volterra Operator im numerischen Fall berechnen kann. Die Herleitung in der Numerik stellt keine große Einschränkung dar, da in der experimentellen Praxis alle Daten numerisch vorliegen. Es wird gezeigt, unter welchen Bedingungen identische Ergebnisse produziert werden können und wie das Ergebnis des STRFs, wenn diese Bedingungen nicht gegeben sind, angenähert werden kann. Anhand von Kapitel 3 ist es möglich, dass STRF und die bei der Nutzung implizit gemachten Annahmen besser zu verstehen und es in einem Zusammenhang mit der Volterra Reihe zu sehen.

Im vierten und letzten Kapitel wird dann zum ersten Mal experimentell eine Methode (Differential Sampling) angewandt, um Elemente der Volterra-Reihe direkt zu bestimmen. Weiter wird anhand von Doppelimpulsantworten gezeigt, dass im IC längere Zeitskalen als die des STRF gefunden werden können. Dabei bietet die Beschreibung einer Zelle durch

den Volterra Operator, dessen Fähigkeit zur Vorhersage anderer Stimuli noch nicht systematisch getestet wurde, Vorteile. Zum einen ist die Volterra Reihe eingebettet in die Theorie der Differentialgleichungen und ist damit mathematisch gut fundiert. Des Weiteren gibt es weniger Schwierigkeiten, den Pegel, der zur Messung genutzt wurde, zu interpretieren. Auch ist die Konvergenz besser definiert und der genauere mathematische Hintergrund bietet mehr Möglichkeiten.

# Contents

<b>1</b>	<b>An introduction to comodulation masking release at the level of the inferior colliculus</b>	<b>2</b>
1.1	Outline and key findings . . . . .	2
1.2	Abstract . . . . .	3
1.3	Introduction . . . . .	4
1.4	Materials and Methods . . . . .	5
1.5	Results . . . . .	9
1.5.1	Single unit responses . . . . .	9
1.5.2	Population analysis . . . . .	16
1.6	Discussion . . . . .	20
1.6.1	Comparison to previous results . . . . .	20
<b>2</b>	<b>Influence on comodulation masking release and possible predictors</b>	<b>22</b>
2.1	Outline and key findings . . . . .	22
2.2	Abstract . . . . .	23
2.3	Introduction . . . . .	24
2.4	Material and methods . . . . .	26
2.5	Results . . . . .	28
2.5.1	Narrow and V-shaped FRAs are most likely to show CMR . . . . .	28
2.5.2	Can the similarity index predict the units' ability to show CMR? . . . . .	30
2.5.3	The influence of the carrier and of the distance of the closest FB . . . . .	31
2.5.4	Spectro-temporal-receptive-fields of selected units . . . . .	33
2.6	Discussion . . . . .	36
2.6.1	The predictability via the FRA and the STRF . . . . .	36
2.6.2	Spectral distance . . . . .	38
2.6.3	High correlation between the first intervals of the codeviant (CD) response and Comodulation masking release (CMR) effects . . . . .	38
<b>3</b>	<b>Turning a spectro-temporal filter into a Volterra operator in the time domain</b>	<b>39</b>
3.1	Outline and key findings . . . . .	39
3.2	Abstract . . . . .	40
3.3	Introduction . . . . .	41
3.3.1	From the black box model to the spectro-temporal receptive field . . . . .	41

3.3.2	A brief introduction to systems identification . . . . .	42
3.3.3	Frequency and time-frequency representations . . . . .	43
3.3.4	The STRF . . . . .	45
3.4	Materials and Methods . . . . .	47
3.4.1	Time-Frequency analysis . . . . .	47
3.4.2	The STRF operation and the corresponding filter using square TFR forms . . . . .	47
3.4.3	Absolute value TFRs . . . . .	49
3.4.4	Parameters of the simulation . . . . .	51
3.5	Results . . . . .	52
3.5.1	Comparing the time-frequency representations . . . . .	52
3.5.2	An STRF based on a squared time-frequency representation can be exactly reproduced by a 2nd-order Volterra operator . . . . .	52
3.5.3	The dictionary of 2nd-order Volterra kernels . . . . .	52
3.5.4	The influence of the memory of the TFR filters on the kernel . . . . .	55
3.5.5	STRF based on the absolute value time-frequency representation, 1st- and 2nd-order approximations . . . . .	55
3.5.6	The approximated 2nd-order kernel is proportional to the exact kernel of the square form . . . . .	59
3.6	Discussion . . . . .	60
3.6.1	Comparison to the mapping of Wiener kernels to the STRF . . . . .	60
3.6.2	Specific results for absolute value and squared TFRs . . . . .	60
3.6.3	Benefits of converting an STRF into a Volterra series . . . . .	61
3.7	Appendix . . . . .	62
<b>4</b>	<b>Using double clicks to test spectro-temporal receptive fields and to estimate Volterra operators</b> . . . . .	<b>63</b>
4.1	Outline and key findings . . . . .	63
4.2	Abstract . . . . .	64
4.3	Introduction . . . . .	65
4.3.1	Example calculations to visualize Differential Sampling . . . . .	65
4.3.2	The STRF . . . . .	67
4.4	Materials and Methods . . . . .	68
4.4.1	Stimuli for the STRF and higher order experiment . . . . .	68
4.4.2	Evaluation . . . . .	69
4.5	Results . . . . .	71
4.5.1	Using the ergodic hypothesis to test for significance . . . . .	71
4.5.2	Examples of different units . . . . .	75
4.5.3	Second order Volterra estimation in the cochlear nucleus . . . . .	90
4.6	Discussion . . . . .	93
4.6.1	Technical aspects of the evaluation: delta function approximation . . . . .	93
4.6.2	The ergodic hypothesis and the comparison with the STRF . . . . .	93

# List of Figures

1.1	Schematic plot of the CMR paradigm . . . . .	8
1.2	Example unit with a suppressed CM masker response . . . . .	10
1.3	Example unit with a stronger decrease in the CM than in the RF masker . . . . .	12
1.4	Example unit that stopped to respond to the CM masker before it started to respond to the target signal . . . . .	14
1.5	Example unit not showing CMR due to FRA shape . . . . .	15
1.6	Example unit not showing CMR due to lacking masker suppression . . . . .	15
1.7	Population response to the CMR paradigm . . . . .	17
1.8	Relative changes in signal response increase and masker response decrease for CMR units . . . . .	19
1.9	Number of units with larger facilitation and suppression depending on the target signal SNR . . . . .	19
2.1	Schematic explanation of the Similarity Index . . . . .	27
2.2	Change in the CM masker response for phasic and tonic units . . . . .	30
2.3	SI between the RF and the other masker responses for different time intervals . . . . .	31
2.4	SI between different time intervals of the same maskers . . . . .	32
2.5	Results for a unit where the FB distance was varied . . . . .	34
2.6	STRFs of selected units . . . . .	35
2.7	Schematic drawing of possible forms of the STRF and their consequences for CMR behavior . . . . .	37
3.1	Schematic drawing of a time-frequency analysis operation . . . . .	48
3.2	Used frequency transformation matrices . . . . .	51
3.3	Time signals in normal and logarithmic form and possible TFRs . . . . .	53
3.4	A toy STRF and the exact corresponding Volterra kernel for square values TFRs . . . . .	54
3.5	The dictionary of 2nd order Volterra kernels resulting from STRF forms . . . . .	56
3.6	Gamma filter and implications for Volterra kernel shape . . . . .	57
3.7	Volterra kernels belonging to a toy STRF based on an absolute value TFR . . . . .	57
3.8	Results for the absolute value TFR . . . . .	58
3.9	The operations behind the STRF . . . . .	61

4.1	Example unit I: Interaction at 200 ms . . . . .	72
4.2	Example unit I: Interaction at 100 ms . . . . .	73
4.3	Example unit I: no Interaction at 50 ms . . . . .	74
4.4	Example unit II: FRA . . . . .	75
4.5	Example unit II: Interaction at 50 ms does not converge . . . . .	76
4.6	Example unit II: Interaction at 25 ms . . . . .	77
4.7	Example unit II: Interaction at 6 ms undetected . . . . .	78
4.8	Example unit II: Interaction at 13 ms . . . . .	79
4.9	Example unit II: STRFs of different TFRs and corresponding Volterra kernels	81
4.10	Example unit III: FRA that indicates non-linear behavior . . . . .	82
4.11	Example unit III: STRF . . . . .	82
4.12	Example unit III: Diagonal of higher order operators at 13 ms inter-click intevall . . . . .	83
4.13	Example unit III: Histogram of the variances . . . . .	83
4.14	Example Uni IV: FRA . . . . .	84
4.15	Example unit IV: Interaction at 100 ms not significant . . . . .	85
4.16	Example unit IV: Interaction at 25 ms not significant . . . . .	86
4.17	Example unit IV: Interaction at 6 ms . . . . .	87
4.18	Amplitude and significance of all units for different inter-click intervalls . . . .	89
4.19	CN: FRA, 2nd Volterra kernel measured with differential sampling and input output simulation, I . . . . .	91
4.20	CN: FRA, 2nd Volterra kernel measured with differential sampling and input output simulation, II . . . . .	92

## Chapter 1

# An introduction to comodulation masking release at the level of the inferior colliculus

### 1.1 Outline and key findings

- Comodulation masking release (CMR) is explained as an auditory effect to increase the detectability of a signal in a masker if the masker fulfills certain requirements
- To test for a physiological correlate of this effect, a signal was placed in three different maskers and the spike rate during the signal and during the masker was measured while the signal intensity was varied
- The key finding is that the signal representation of the preferred masker in the inferior colliculus contains a reduction in the masker response additionally to the facilitation in the signal response from earlier processing stages



## 1.2 Abstract

Auditory signals that contain coherent level fluctuations of a masker in different frequency regions enhance the detectability of an embedded sinusoidal target signal, an effect commonly known as Comodulation masking release (CMR). Neural correlates have been proposed at different stages of the auditory system. While later stages seem to suppress the response to the masker, earlier stages are more likely to enhance their response to the signal when the masker is comodulated. Using a flanking band masking paradigm, the present study investigates how CMR is represented at the level of the inferior colliculus of the Mongolian gerbil. The responses to a target signal at various sound pressure levels in three different masking conditions were compared. In one condition the masker was a 10-Hz amplitude modulated sinusoid centered at the signal frequency while in the other two conditions six off-frequency carriers (flanking bands) were added. From 81 units 26 showed a change that enhanced the detectability of the signal if the temporal modulation of the added flanking bands was identical to that of the masker at the signal frequency compared to the other two masking conditions. This study shows that the response characteristics of these neurons represent an intermediate stage between the representation in the cochlear nucleus and the auditory cortex. This means that the response is increased during the signal intervals but is also decreased for the following masker portions.

### 1.3 Introduction

One parameter that influences the detectability of an acoustic signal in a noisy environment is the modulation of the background noise. It has been shown that natural stimuli often show coherent envelope fluctuations (Nelken et al., 1999) across frequency. The auditory system can use these as a cue for signal detection. Psychophysical studies investigated the processing of such comodulated sounds by measuring the masking of narrowband targets, usually sinusoids, by comodulated maskers and comparing it to maskers with the same spectrum but different envelopes in on- and off-frequency regions, or maskers restricted to a narrow frequency range around the signal frequency. Masker comodulation often resulted in lower detection thresholds. This effect is referred to as comodulation masking release (CMR, Hall et al., 1984, for a review see Verhey et al. (2003)). It was investigated in humans but was also observed in other species, e.g., the mouse (*mus musculus*) (Klink et al., 2010) the European starling (*Sturnus vulgaris*) (Klump and Langemann, 1995) and the bottlenose dolphin (*Tursiops truncatus*) (Branstetter and Finneran, 2008).

Two different paradigms have been used to study neuronal correlates of CMR at different stages of the auditory pathway. One paradigm that was used for an effect called envelope locking suppression (ELS) was investigated with a band-widening type of CMR experiment (Las et al., 2005; Nelken et al., 1999). This means that signal detectability was measured with a single masker for various bandwidth (see Verhey et al., 2003). The other paradigm tested for CMR was a flanking-band type of experiment, where the masker consisted of narrowband masker components, one on-frequency masker and one or more off-frequency components (Grose and Hall, 1989; Moore et al., 1990; Dau et al., 2009; Verhey et al., 2013). Psychoacoustical simulations indicate different mechanisms underlying the masking release in the band-widening and flanking-band types of CMR experiment (Verhey et al., 1999; Piechowiak et al., 2007; Dau et al., 2013).

Despite these different mechanisms, the existing studies still suggest that some critical transformation of the signal representation takes place at the level of the inferior colliculus (IC). It occupies a strategic position in the ascending auditory system as almost all lower auditory nuclei project to its central nucleus (Oliver and Huerta, 1992; Casseday et al., 2002). Throughout the text, the term IC will refer to the central nucleus of the inferior colliculus. The IC receives direct, monaural pathways from the ventral and dorsal cochlear nucleus and indirect, binaural pathways from the lateral and medial superior olivary complex. Further, it is part of a multisynaptic pathway that includes a synapse in the lateral lemniscus (see Oliver and Huerta (1992)). Importantly, the projections from the dorsal cochlea nucleus are only excitatory (Semple and Aitkin, 1980; Davis, 2002; Malmierca et al., 2005). There, CMR was tested using a flanking band paradigm and it was found that for a subpopulation of cells, the response during the signal interval significantly increased in one of the maskers (Neuert et al., 2004). This was also shown in the dorsal cochlea nucleus (Pressnitzer et al., 2001). At the level of the IC, a band-widening experiment did not show strong effects (Las et al., 2005) that were present at subsequent processing stages, the medial geniculate body (MGB) and the auditory cortex. The MGB is the prime target of the IC and in the MGB it has been shown that a correlate of CMR does not increase the signal response but, instead decreases

the masker response.

Using an evaluation method modified from that in the cochlear nucleus (CN) to capture a wider range of effects and their influence on the representation, in this chapter it is investigated how the signal representation might change as expected based on the results from the studies of Pressnitzer et al. (2001); Neuert et al. (2004) and Las et al. (2005); Nelken et al. (1999).

## 1.4 Materials and Methods

Experiments were performed on 25 adult male ketamine/xylazine anesthetized Mongolian gerbils (*Meriones unguiculatus*), aged between 3-16 months with a body weight between 80-120 g. All experiments were conducted in accordance with the international National Institute of Health *Guidelines for Animals in Research* and with ethical standards for the care and use of animals in research defined by the German Law for the protection of experimental animals. All experiments were approved by an ethics committee of the State of Saxony-Anhalt, Germany.

*Surgical Procedure.* Anesthesia was a mixture of 45% ketamine (50mg/ml; Ratiopharm), 5% xylazine (Rompun, 2%; BayerVital) and 50% isotonic sodium chloride solution (154 mmol/L; Braun). Initially, a dose of 0.3 ml was given intraperitoneally. To maintain the anesthetized state during surgery and later during the course of the experiment, it was applied subcutaneously at a rate of 0.1 ml/ hr, or as necessary. The state of the anesthesia was ensured by monitoring the hindlimb withdrawal reflex and respiratory rate (Zandieh et al., 2003). Body temperature was kept at 37°C using a controlled heating blanket. A craniotomy was performed 2 mm caudal and 2 mm lateral of lambda using the major blood sinus as a landmark (Brückner and Rübsamen, 1995; Cant and Benson, 2005). A gold-plated pin (Amphenol) with good electric contact to the dura was implanted in the contralateral parietal bone and served as voltage reference. For stereotactic fixation, an aluminum bar was attached to the frontal bones with dental cement (Paladur; Heraeus Kulzer). Recordings were made with tungsten electrodes (2-4 M $\Omega$ , FHC) using a Plexon Multichannel Acquisition Processor or Omniplex System (Plexon Inc.). The electrode was lowered dorsoventrally until a multi-unit response to wideband noise or tone pips was observed. After this procedure, single units were identified via their signal-to-noise ratio (SNR) and post sorted using the software Offline Sorter (Plexon Inc). To ensure that the position within the IC was the central nucleus, the tonotopic gradient in the multi-unit response was measured.

*Stimulus generation and setup.* Acoustical stimuli were digitally synthesized, pseudo-randomized and controlled using Matlab (MathWorks) and amplified (depending on the setup used on the experimental day, either with an AMP 75 wideband power amplifier, Thomas Wulf, Frankfurt, Germany, or an Alesis RA 150 amplifier, inMusic Brands Inc) to a calibrated Canton Plus XS.2 speaker and presented in an acoustically and electrically shielded chamber. The distance between the speaker and the animal was either 1m (setup with the AMP 75) or 70 cm (setup with the RA 150). The highest presented sound level that was parametrically attenuated was 90 dB SPL.

*Recording paradigm.* Once a single unit with good SNR ratio was found, it was characterized on the basis of its frequency response area (FRA) using pure tones with a frequency in the range of 0.25 to 16 kHz and with a duration of 100 ms (including 5-ms raised-cosine-squared ramps at on and offset). The signals were separated by silent intervals of 500 ms. The signal frequencies were equally spaced on a logarithmic scale in 1/4 octave steps. The signal intensities were varied from a highly supra-threshold level down to threshold in steps of 10 dB. Each combination of level and frequency was repeated ten times.

After the FRA was measured, the unit was stimulated using the same type of stimuli as

described in Neuert et al. (2004) to investigate the unit response to CMR-stimuli. Three different masking conditions were used (see figure 1.1). For the reference (RF) condition, the masker consisted of a sinusoidally amplitude modulated AM pure tone at 100% modulation depth with a modulation frequency of 10 Hz. Given these parameters, a sufficiently large fraction of units in the IC should respond to the RF masker alone (Krishna and Semple, 2000; Joris et al., 2004), as it is a sinusoidal amplitude modulation as investigated in these studies. The carrier frequency was chosen in such a way that the unit showed a strong response in the FRA. As the carrier frequency was the same for the masker and for the signal, this masker is also called on-frequency masker. The masker duration was 500 ms, resulting in five full modulation cycles. The amplitude of the modulation was zero at the beginning of the masker, i.e., it was a cosine modulation with a phase of  $\pi$ . The level of the masker was chosen to be well above threshold, usually at least 30 dB. Thus, the masker can be described by the following equation

$$M_{RF}(t) = \frac{1}{a} (1 + \cos(2\pi \cdot 10 \cdot t + \pi)) \cdot \sin(2\pi \cdot F_c \cdot t) \quad (1.1)$$

with  $t$ , the time, the maximum amplitude  $2a$  and the carrier frequency  $F_c$ .

For the comodulated (CM) condition, the masker consisted of the on-frequency masker component and six further components at spectrally remote frequencies. The frequencies were chosen to be just outside the excitatory region of the unit according to the FRA and are likely to be inhibitory (Egorova and Ehret, 2008). These components are commonly referred to as flanking bands (FBs). The FBs were amplitude-modulated AM pure tones (of frequency  $FB_i$ ) with the same modulation frequency and starting phase of the modulation as the on-frequency masker. The resulting equation for the CM masker is

$$M_{CM}(t) = M_{RF}(t) + \frac{1}{a} \sum_{i=1}^6 (1 + \cos(2\pi \cdot 10 \cdot t + \pi)) \cdot \sin(2\pi \cdot FB_i \cdot t). \quad (1.2)$$

In the present study, the frequencies of the FBs were spaced at least 100 Hz apart from each other. Note that the level of each of the flanking band masker components was the same as for the RF masker. Thus the overall level of the CM masker was higher than that of the RF masker.

The third condition is the CD condition. The term codeviant refers to a masker with the same magnitude spectrum as the CM masker, but where the FBs had a modulation phase of 0 instead of  $\pi$ ,

$$M_{CD}(t) = M_{RF}(t) + \frac{1}{a} \sum_{i=1}^6 (1 + \cos(2\pi \cdot 10 \cdot t)) \cdot \sin(2\pi \cdot FB_i \cdot t). \quad (1.3)$$

The target signal was for all masking conditions (RF, CM and CD) a sequence of three consecutive 50-ms tone pips with the same frequency as the carrier frequency of the on-frequency masker,  $F_c$ . The tone pips were positioned into temporal minima of the masker: they lasted from 175 to 225, from 275 to 325 and from 375 to 425 ms. Each tone pip

was gated on and off using 25-ms cosine-squared ramps. The signal was played at five attenuations relative to the level of the (RF) on-frequency masker component: 0, -10, -20 and -30 dB. These relative signal levels will be referred to as signal-to-noise ratios (signal-to-noise ratio (SNR)) in the following. In addition, the masker was presented alone, referred to as masker-only (MO) condition in the following. All stimuli were presented at least 25 times. For units with a stable response, 50 repeats were recorded.

*Data analysis.* For the evaluation of each unit a delay time was determined that was optimal for the detection of the target signal. The target was presented in the minima of the masker, i.e., during the time intervals from 225 to 275, 325 to 375 and 425 to 475 ms after stimulus onset. The optimum delay time minimizes the response in the MO conditions during the time intervals where the target signal is added. The time was allowed to vary in a range from -10 to +30 ms. Negative delays were sometimes necessary since the next masker cycle started while the target tone pip was still present and some units showed a strong onset response to this next masker cycle. In the following analysis, the time scale was corrected for this delay.

The study used the difference between spikes in the signal interval and spikes in the masker interval as the criterion to differentiate between CMR and non-CMR representing units. For every SNR except the MO, it was tested whether (i) this difference was largest in the CM condition and (ii) there was an increase in the given signal response compared to the MO in the CM condition. A one-way ANOVA (MatLabR2006a) was performed to examine the statistical significance of the effect. The dependent variable was the masking condition (RF, CM and CD). If the ANOVA showed a significant effect of condition, *post-hoc* Scheffé's (multcompare; MatLabR2006a) tests were applied to test for significant differences between the masking conditions. If these tests showed significant differences in the aforementioned spike count between the CM and RF and the CM and CD condition ( $p < 0.05$ ) for at least two signal attenuations, the unit was defined as a CMR unit.

As a measure of the average response of the units, the responses to signal relative to the masker were summed up for all units. In addition, the response gain was calculated as  $(S_{\Delta} - S_M) / (S_{\Delta} + S_M)$ .  $S_M$  denotes the spike number in the MO condition and  $S_{\Delta}$  the spike rate at a given SNR. This results in four gain values, one for each SNR (0, -10, -20 and -30 dB) with the MO being the reference. The response gain varied between -1 and 1. This measure was used to compare the increase of spikes during the signal interval and the decrease of spikes during the masker interval. Furthermore, it served to indicate if the suppressive or the facilitative effects were dominant (at any given SNR) in the CM masker. The unit was defined as facilitative or suppressive at a certain SNR, depending on which value was larger in magnitude.

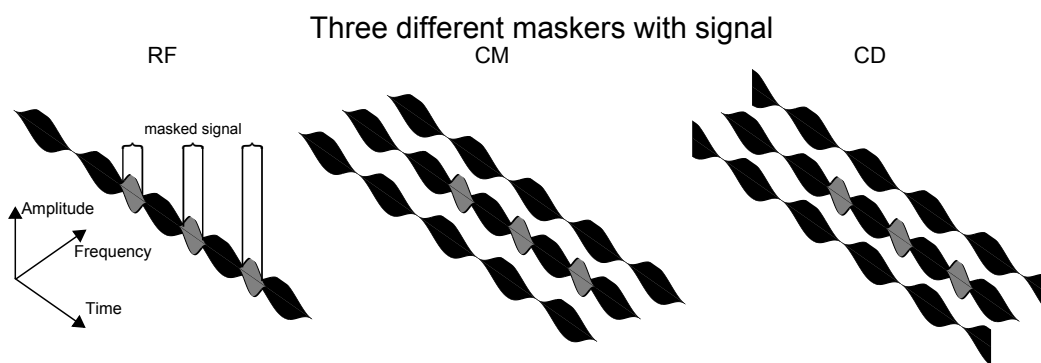


Figure 1.1: Schematic plot of 3D spectrograms of the paradigm used to investigate CMR in single units. Three masking conditions were used; the different maskers are plotted in black: reference (RF, left), comodulated (CM, middle), and codeviant (CD, right). The target signal (plotted in gray) was identical in all maskers. The SNR was set to 0 dB as can be seen by the amplitude of the signal, which is the same as that of the RF masker. In the middle and on the right only two (out of six) FBs are shown.

## 1.5 Results

### 1.5.1 Single unit responses

During the evaluation of the single unit responses, the FRAs will be described according to response and firing pattern. Many of the terms used for the units are self-explanatory, a more detailed explanation will follow in chapter 2.

Figure 1.2 shows the response of a unit with a "narrow" FRA and a "chopper"-like response pattern (see panel a). The FRA shows the firing patterns in response to a pure tone of the frequency and sound attenuation given by the respective coordinates on the abscissa and ordinate. The duration of each of the depicted firing patterns was 100 ms.

The signal frequency for this CMR experiment (and thus the center frequency of the on-frequency masker) was 4 kHz. The center frequencies of the FBs were 0.8, 1, 1.2 and 6.8, 7 and 7.2 kHz. The attenuation for each masker band (on-frequency masker and FBs) was 30 dB. The frequency level combination of the on-frequency band is indicated as an open circle, those of the FBs with filled circles.

Figure 1.2c shows in the bottom panels the response of the unit to the maskers without an added signal (MO condition). The comparison between the response to the on-frequency masker (RF condition, bottom left) and the masker in the CM condition (bottom middle), shows that the FBs were suppressing the response to the on-frequency masker. The response to the CD masker (bottom right) showed that the codeviant FBs enhanced the response to the Amplitude modulation envelope of the on-frequency masker component. The unit's envelope locking to the Amplitude modulation was increased in the CD masker and barely present in the RF masker.

For the CM condition, a clear signal response was already observed for the -30 dB SNR. This ratio increased when the signal level was further increased to -10 dB. There was a small decrease from the -10 dB to the 0 dB SNR (middle column panel c and top panel a). The unit's response to the signal increased in the RF condition, too, but to a lesser extent. In this condition, there was a minor reduction of masker response and there were prominent signal responses at the -20 and -10 dB SNR, but not as prominent as in the CM masker. The response was not affected by the presence of the signal for the CD masker. This unit did not show any suppressive effects in terms of a decrease of the masker response in the CM condition due to the addition of the signal, since the response to the CM masker alone was already completely suppressed. Panel b of figure 1.2 shows the summary response of the unit to the different masking conditions. Note that the spike rate in the signal-interval for the CM masker exceeded that for the RF masker only for the -10 and 0 dB SNRs. The spike rate during the signal-interval was lowest for the CD conditions for all SNRs. The masker portion of the response was hardly altered by the presence of the signal. This is very similar to the behavior of the CN units shown in Neuert et al. (2004). On basis of the criteria employed in the present study, this unit is classified as a CMR unit.

Figure 1.3 shows the response of another CMR unit with a "narrow" FRA (panel a). On the basis of the peri-stimulus time histogram (PSTH) it was classified as a "sustained" firing



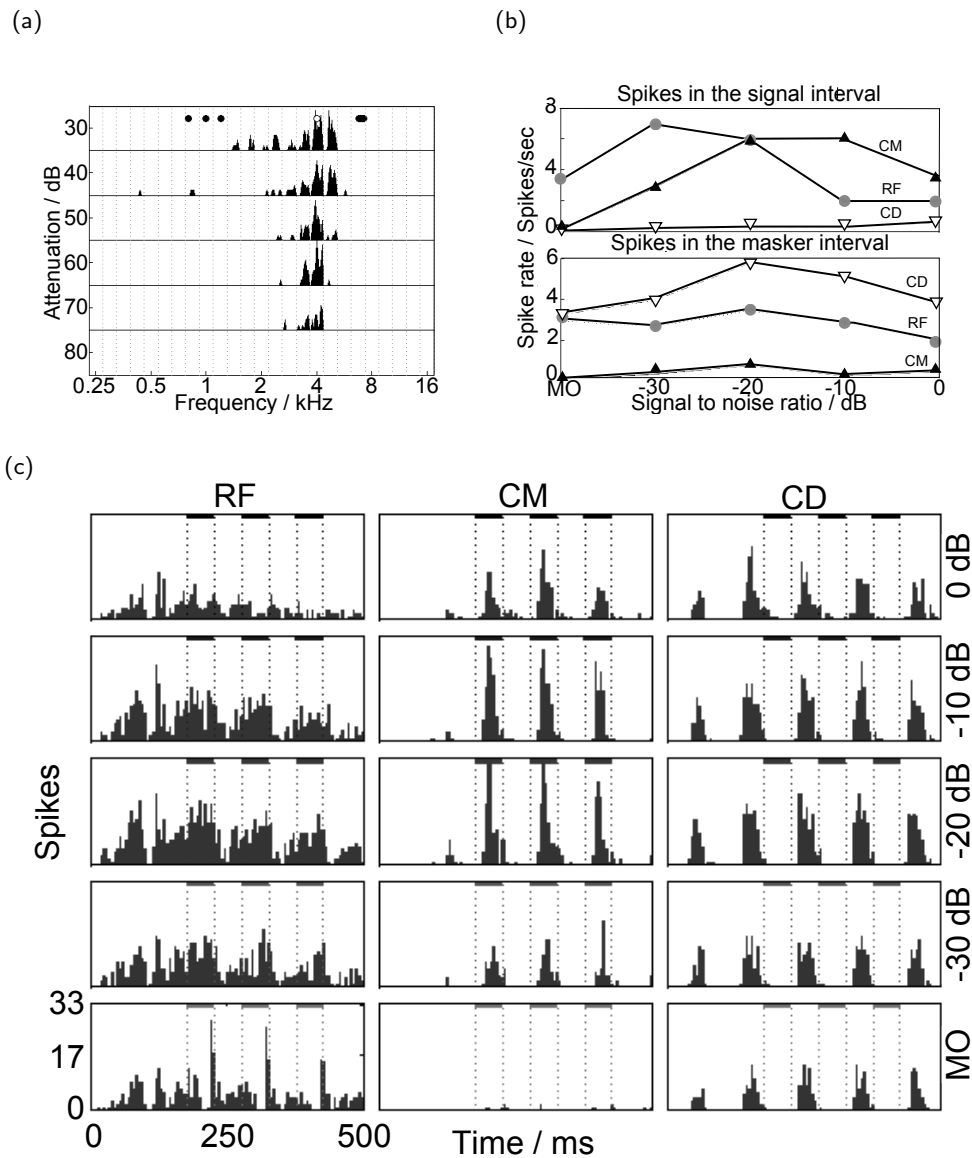


Figure 1.2: Results for a unit that did not respond to the CM masker but to the signal in the CM masker condition. (a) frequency response area (FRA), the white dot indicate frequency and level of the on-signal masker component, the black dots those of the flanking bands (FBs). For each frequency-attenuation combination the FRA shows the peri-stimulus time histogram (PSTH) for a 100-ms pure tone with this frequency and attenuation. (b) Spike rate in the signal interval and in the masker interval for the different masker conditions. (c) PSTHs (bin width 5 ms) for the three masking conditions (columns) and SNRs (rows). The bottom panels show the response to the masker-only (MO).

pattern. The signal frequency was 1 kHz. The FBs were positioned at 0.25, 0.35 and 1.7, 1.8, 1.9 and 2 kHz. The attenuation of each masker band was 40 dB.

A comparison of the response to the RF masker and the CM masker showed that the addition of FBs reduced the unit's response (see MO conditions, bottom panels of figure 1.3c). The effect was smaller than that for the unit in figure 1.2. The response to the CD masker was larger than that to the CM masker but smaller than that to the RF masker. For all maskers the response of the unit followed the envelope of the Amplitude modulation of the on-frequency masker.

A signal presented at a SNR of 30 dB below the masker level hardly changed the unit's response. For the CM and RF maskers, the response of the unit started to change at the -20 dB and further increased in the signal response at the -10 and 0 dB SNRs. For the CD and RF conditions, the addition of the signal hardly changed the response to the masker (panel b). In contrast, for the CM condition the addition of the signal increased the response in the signal interval and decreased the response to the masker in the masker interval. At the 0 dB SNR, the response for the CM condition showed a periodic change between a stronger signal response and a weaker masker response. Figure 1.3b shows that the response to the signal was stronger for the RF condition than for the CM condition at all signal levels. In terms of the gain functions, this unit showed more facilitation of the response to the signal than suppression of the response to the masker for all SNRs except for the -30 dB level. On basis of the signal response relative to the masker response the unit is classified as a CMR unit.

Figure 1.4 shows a unit that was classified as an "onset" unit with a "monotonic" V-shaped FRA (panel a). The on-frequency masker had a center frequency of 1 kHz. The FBs were spaced in 0.1 kHz intervals from 3.3 to 3.8 kHz. The attenuation of each masker band was 40 dB. The FBs were positioned spectrally above the signal due to the lack of non-excitatory frequencies below the signal frequency. The unit showed envelope locking for all masking conditions (compare bottom panels of figure 1.4c).

The MO responses to the different maskers shown in the bottom panel of figure 1.4c were similar. For all maskers, the response in the interval from 0-100 ms differed from the response in the interval from 100-200 ms. The difference was smallest for the PSTH of the RF masker: the response was slightly stronger in the first interval with a higher maximal rate and temporally more extended. For the CM masker, the difference was more pronounced. While there was a strong response to the first cycle of the masker modulation, it was less than half during the second one. A similar behavior was found for the CD masker, but the onset response was less pronounced than for the CM masker. The temporal extension of the firing pattern decreased for the CD masker in the second interval and thus reduced the spike count.

Figures 1.4c and 1.4b show that the response to the CM masker was largely suppressed when the signal was added to the masker at the -30 dB SNR. This is also observed for the RF condition at a SNR of -20 dB and for the CD condition at -20 to -10 dB SNR. For the CM condition, there was a weak response to the signal at the 0 dB SNR that was not present in the RF and CD condition (see also panel b). Note that the signal was always well above

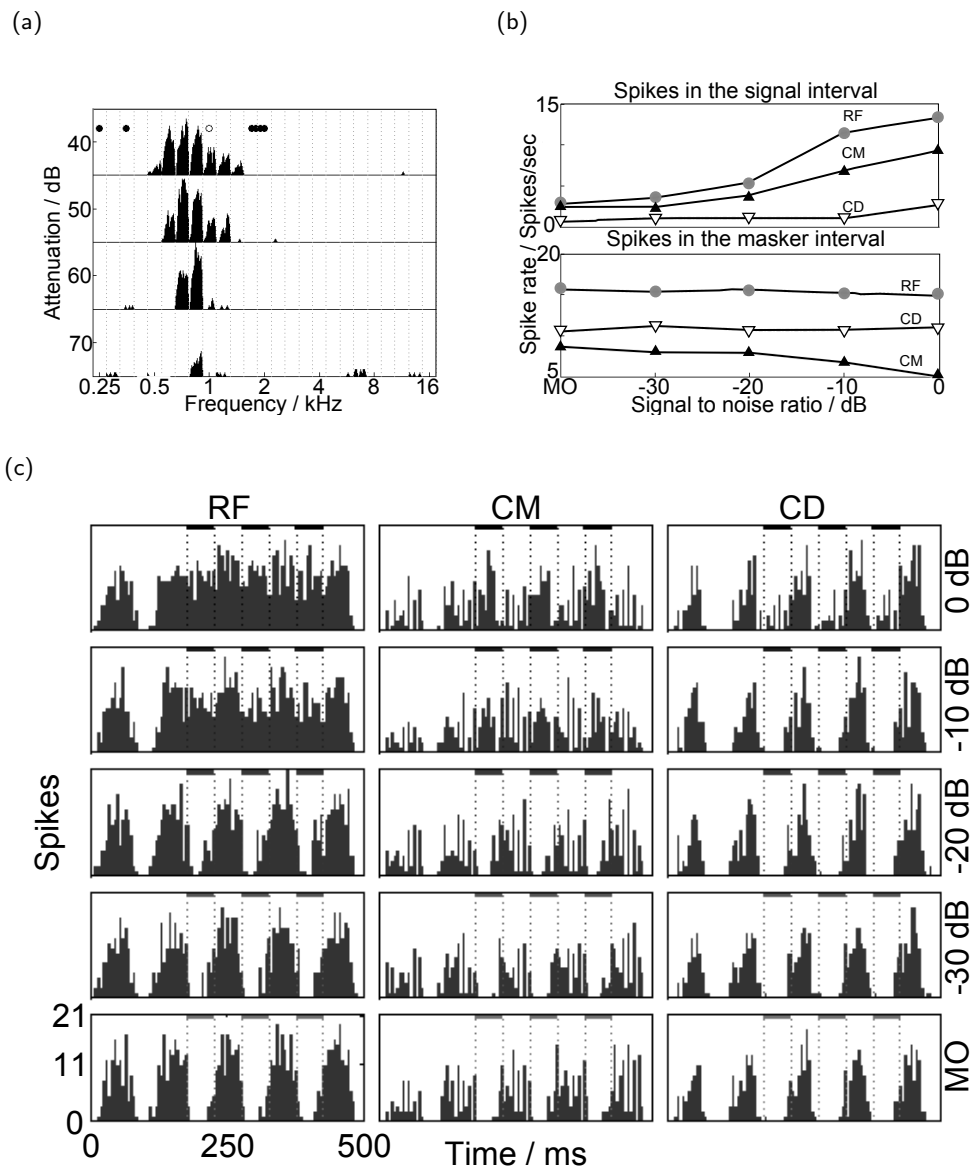


Figure 1.3: A unit showing a decrease in spike rate for high signal attenuations in the CM masker only, notation as in figure 1.2. (a) FRA, showing a "narrow" response area and sustained firing, (b) Spikes in the signal and in the masker for the different conditions. The increase in spike rate is highest for the RF masker at all signal-to-masker levels, but in the CM masker the response to the masker showed the strongest decline. (c) PSTH of CMR paradigm. Due to the decrease in the masker response in the CM masker, as seen in panel (b), the response to the signal became more prominent and the SNR in the response was increased.

threshold of the unit (see FRA in panel a) when it started to suppress the masker response. Overall, this unit was also classified as a CMR unit.

Whereas the previous three figures showed units classified as CMR units, figure 1.5 shows a unit that did not show CMR. The FRA is shown in panel a, the CMR response for selected SNRs in the CMR experiment are shown in panel b. The unit was classified as having a "tilt" FRA (figure 1.5a) and an "on-sustained"-like firing pattern. The signal frequency was 0.4 kHz and the FBs were spaced evenly in 100 Hz steps from 0.9 to 1.4 kHz. The attenuation of each masker band was 30 dB. There was a weak response to the RF masker in the MO condition and no response to neither the CD masker nor the CM masker (bottom panels figure 1.5b). The presence of the FBs seemed to suppress the unit's response entirely, regardless of their phase. The unit also responded to the signal in the RF masker at the 0 dB SNR, while there was no change in the firing patterns of the other two maskers. As the FRA was classified as "tilt", this unit showed strong non-linearities in its response. There were only few frequencies for which it responded reliably for a range of different attenuations.

Figure 1.6 shows another unit which did not fulfill the criteria of a CMR unit. The unit had a "narrow" shaped FRA and a "sustained" firing pattern. Its high spontaneous activity was suppressed at an attenuation of 30 dB for frequencies ranging from 0.6 to 2.8 kHz and around 4.8 kHz. The on-frequency masker had a center frequency of 4 kHz and a 30 dB attenuation. The FBs were positioned spectrally below at 1.5, 1.75 and 2.0 kHz and spectrally above at 4.75, 5 and 5.25 kHz, their attenuation was 30 dB.

Figure 1.6b shows similar responses to the CM masker and the RF masker. The addition of comodulated FBs did not reduce the response to the on-frequency masker, even though pure tones played at this level suppressed the spontaneous activity (see panel a).

The unit responded well to the signal in the CM condition, but this response was not significantly different from that in the RF condition. Especially, the contrast between the masker and the signal response was not enhanced by the addition of comodulated FBs. The response in the signal interval was suppressed in the CD condition, especially at -10 dB SNR.

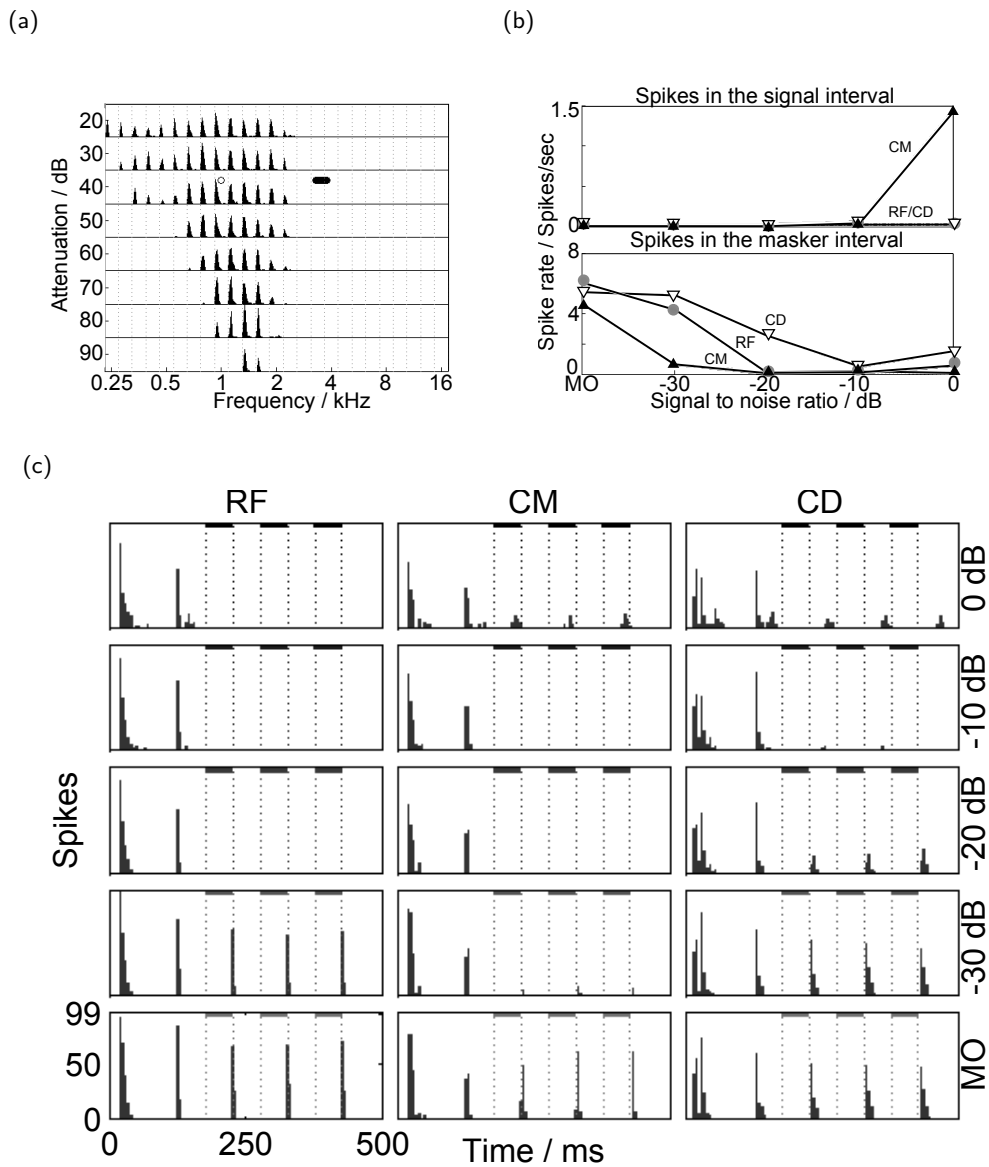


Figure 1.4: A unit showing suppression of the masker response at low signal attenuations, notation as in figure 1.2. (a) FRA (note the onset behavior of the unit), (b) Spike rates in the signal and masker responses. (c) PSTH of the CMR paradigm. Panel (b) and (c) show that the unit ceased to respond to the masker in the CM condition at -30 dB SNR, then at the RF and last in the CD condition. At 0 dB SNR there was a response to the signal in the CM condition.

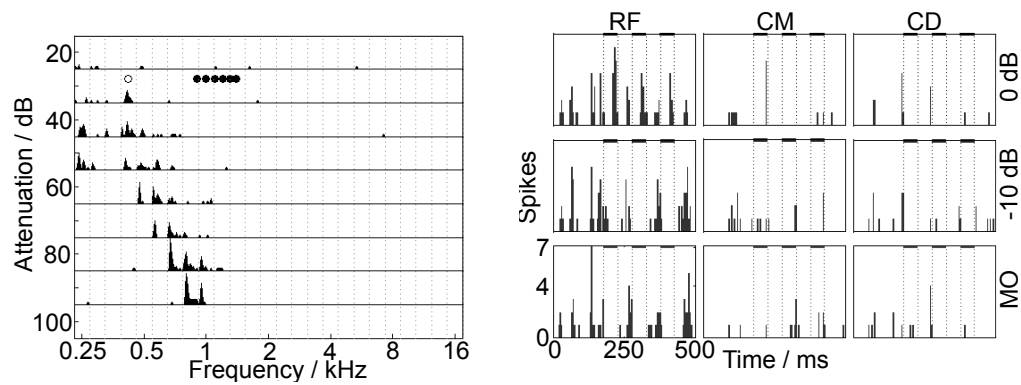


Figure 1.5: A unit with a "tilt" FRA (panel a) that did not show CMR. Panel (b) shows the response to the MO (bottom panels) and -10 and 0 dB SNR. Due to the form of the FRA this unit was unlikely to show CMR behavior in the first place.

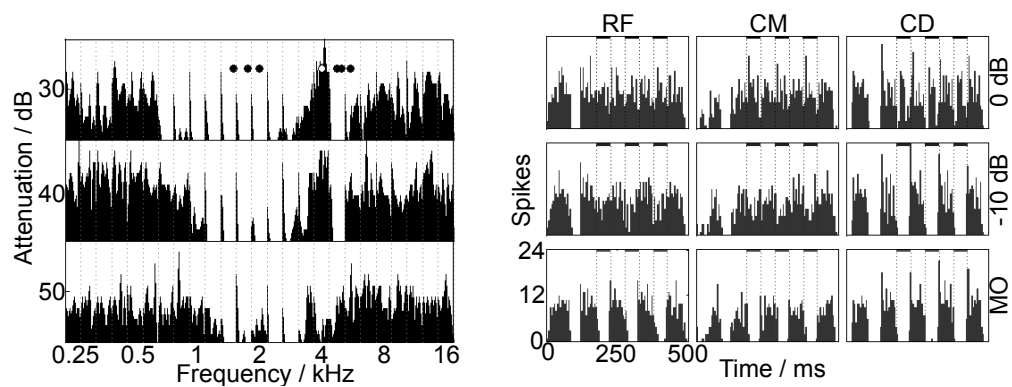


Figure 1.6: A unit that showed less suppression of the RF masker in the CM condition than expected from the FRA and that did not show CMR. The FRA (panel a) was classified as "narrow / close". Panel (b) shows the CMR response at the MO and the -10 and 0 dB SNRs. As the unit had a high spontaneous activity, suppressive areas can be seen in the FRA. This made the unit a likely candidate to show CMR, though it did not.

## 1.5.2 Population analysis

First, we investigated the relative role of CMR units in shaping the representation of the signal within the measured population. The subgroup of CMR units was analyzed more thoroughly in a next step; as units do suppress their masker response (as the units shown in figures 1.3 and 1.4), the influence of the masker response reduction compared to potential increase in signal response was investigated. This was done using the gain function as introduced above (see Materials & Methods section). The follow-up analysis on the CMR units was based on individual SNRs. Here it was investigated which mechanism (facilitation or suppression) was predominant in the units at a given SNR.

Figure 1.7a shows the individual results for the difference between the signal and the masker responses in the three different masking conditions. On the left the results for the RF condition are shown, in the middle the results for the CM condition and the results for the CD condition are shown on the right. Solid lines denote units that were classified as CMR units, dotted lines non-CMR units. In the CM condition there is a trend for larger responses as the signal-to-noise ratio is increased. Figure 1.7 shows the difference between the spikes in the signal and the spikes in the masker interval with respect to this difference for the masker-only (MO) SNR ratio as a function of target signal level. Panel (a) shows the average for all 81 units, with the filled upwards triangle denoting the CM condition, the circle the RF condition and the open downward pointing triangle the CD condition. For this sample, the result is the same for the RF and the CM condition in all but the 0 dB SNR ratio. At this highest SNR ratio the response in the CM masker is larger than in the other two conditions. The response in CD condition is lower for all SNR ratios. Panel (b) shows the results for the 55 non-CMR units. Here, the results are approximately the same for all masking conditions at the -30 and 0 dB SNR ratios. For the intermediate ratios, signal detectability in the RF condition is better than in the CM and CD conditions. Panel (c) shows the results for the 26 CMR units. Here, the response in the CM masker has the highest values for all except the -30 dB SNR ratio. At this SNR ratio the result is similar for the CM and the RF condition.

Figure 1.8 shows the relative change of spikes during the signal (panel a) and the masker (panel b) intervals for those units that were categorized as CMR units. The solid line indicates the average data; the symbols connected with thin lines show the data for the single units shown in figures 1.2-1.4. For the RF condition, the gain in the response increased monotonically from 0.06 to 0.47 when the SNRs increased from -30 to 0 dB. For the CM condition, it increased from 0.13 to 0.72 for the same signal-level range. The increase was more pronounced between the two lowest levels than at the higher levels. For the CD condition, the gain values were all below 0.1 for all signals, except at the 0 dB SNR where it rose to 0.48, i.e., the same value as found for the RF masker. At the -30 dB SNR, it was less than zero. While the increase in signal response was expected as the signal level was raised, the decrease in the response during the masker interval required an interaction between signal and masker. Figure 1.8b shows this decrease of the response in the masker interval when the signal is added. This was generally not as strong as the increase in the response during the signal interval for all masker types. For the RF condition, it started at -0.01 and decreased almost linearly to -0.21 at the highest signal level. For the CM condition,

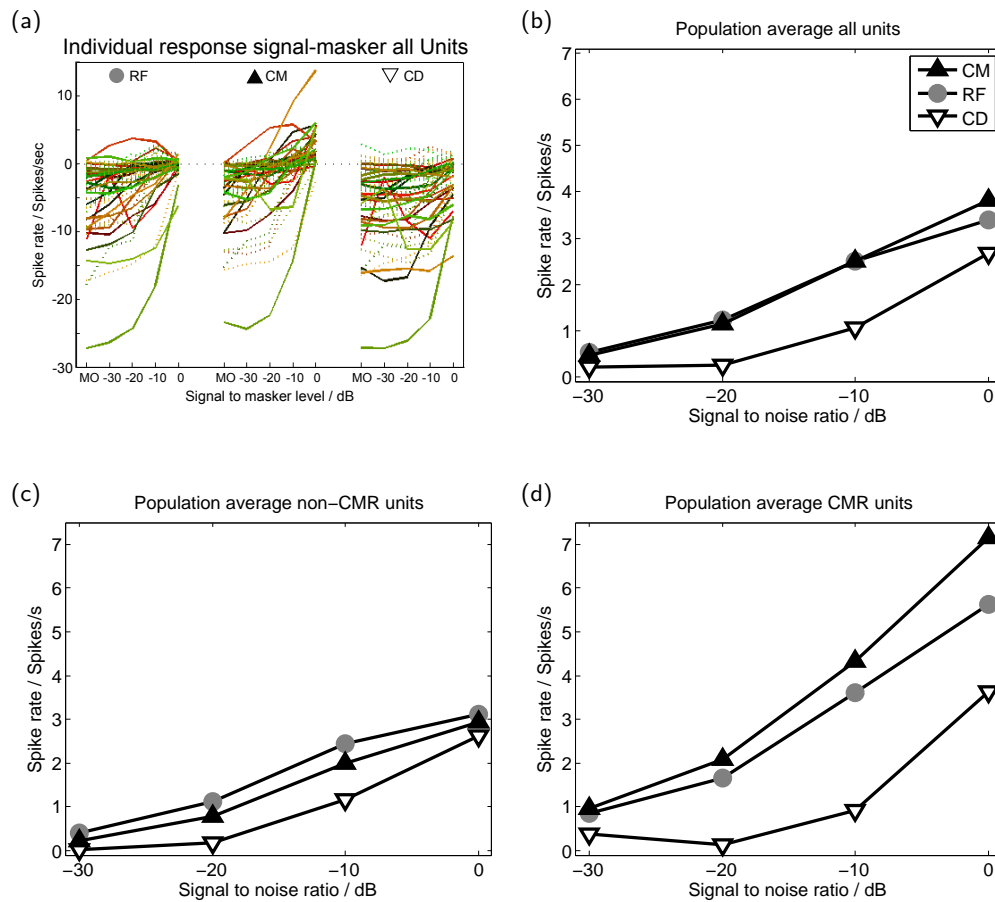


Figure 1.7: Population results: Panel (a): Individual differences between the spikes in the signal interval and the spikes in the masker interval of the 81 units. Solid lines refer to CMR units and dotted lines to non-CMR units. Other panels: Average for the differences between the spikes in the signal interval and the spikes in the masker interval with respect to the MO signal-to-noise ratio ratio. Panel (b) denotes the results for all units (81), panel (c) for the 55 non-CMR units and panel (d) for the 26 CMR units.



the decrease of the initial value was less linear but also monotonic from -0.05 to 0.30 as SNR increased. For the CD condition, the gain was nearly constant (approximately -0.05) for a large range of signal levels. Only at the highest signal level that it decreased to -0.17.

Single responses may differ from those average results in several aspects.

The unit of figure 1.2 (response indicated by squares in figure 1.8) is only showing signal facilitation. This is reflected in a large gain in the signal interval for the CM condition. The response gain for the masker interval in the CM condition was larger than 0.3 for all signal level. This was not shown since it is outside the range shown here. At the other end of the spectrum of possible effects is the unit shown in figure 1.4 (response indicated by diamonds). For the CM condition, this unit already shows a strong decrease of the response in the masker interval at -30 dB SNR. A decrease in the masker response was also found for the other two masker types. However, they were on average less pronounced than for the CM condition. The signal response for the CM condition increased with signal level but only at the SNRs higher than -20 dB. The absolute magnitude was smaller than the gain of masker response reduction at that level. Thus, at this level the unit was classified as a unit showing suppression. At the 0 dB SNR, the gain in the signal was one and larger in magnitude than the decrease (-0.78). At this SNR, the unit was defined as being facilitative. The unit shown in figure 1.3 (response indicated by circles) was close to average in terms of its suppression and facilitation of the masker and the signal, respectively. Even though the masker suppression just reaches -0.25 which is as strong as the facilitation at the -20 dB SNR this effect is important for the signal detection as it increases the differences in spike rate between signal and masker response. Note that the units shown in figures 1.3 and 1.4 would not have been defined as CMR units if they had been evaluated by the method in Neuert et al. (2004). This is due to the fact that these units did not meet the prerequisite that their strongest increase in spike rate during the signal was at two SNRs in the CM masker.

To elucidate the influence of the decrease of spike rate further, figure 1.9 shows the fractions of the dominating effect for the units, based on which of the gains for the CM condition was larger, the one to the signal or the one to the masker. For the -30 dB SNR, where the effects were still small (0.1 and -0.05 respectively) the population of 26 units is split into half. With increasing signal level, the amount of facilitative units raises from 65 (-20 dB) to 73 (-10 dB) and 85 (0dB) percent. This is in agreement with the results shown in figures 1.8a and 1.8b where the maximum relative change in signal response is 0.7 while it is -0.3 for the masker response.

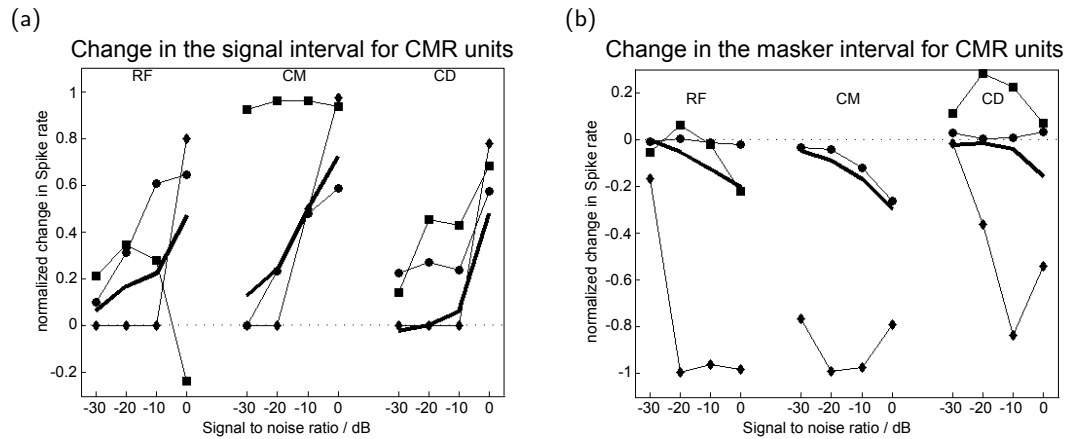


Figure 1.8: Relative changes of the spike rate in the signal interval (panel a) and masker interval (panel b) for the CMR units. Squares denote the unit shown in figure 1.2, circles that shown in figure 1.3 and diamonds that in figure 1.4. In figure b the display of the unit shown in figure 1.2 was omitted for the CM masker as all values were larger than 0.3.

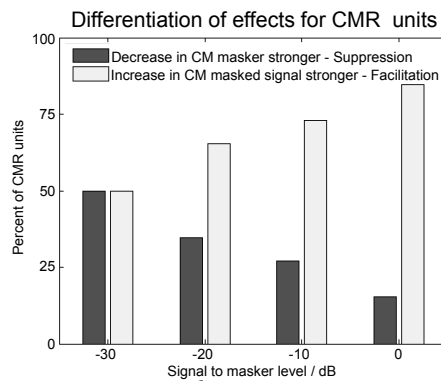


Figure 1.9: The amount of units for which the facilitation and suppression as shown in the CM condition in figure 1.8 is larger with respect to the SNR.

## 1.6 Discussion

The present study showed that a neural correlate of CMR can be found at the level of the IC as an increase of the signal response relative to the masker response when the masker was comodulated. The analysis used in the present study is a generalization of the one applied by Neuert et al. (2004) for the dorsal cochlear nucleus (DCN) units since it is also sensitive to a reduction of the response to the masker. This reduction in response to the CM masker was common to many CMR units although it was smaller than the increase in spike rate during the signal interval.

### 1.6.1 Comparison to previous results

At the level of the CN, it was argued that CMR originated from improved signal detection because the CM masker was largely suppressed by a wideband inhibitor. This inhibition seemed to be unaffected by the presence of the signal. An increase in signal level was reflected by an increase in the response to the signal. In both studies (Pressnitzer et al., 2001; Neuert et al., 2004), CMR units were units which showed a larger response to the signal for the CM masker than for the RF and the CD masker. These units are classified as being "facilitative" in the present study. Their proportion (at the 0 dB SNR) was 25% and thus comparable to the 25% reported in the CN. One may argue that the IC simply inherits the representation of comodulation from the CN since several studies indicated that DCN units project only excitatory to the IC (Malmierca et al., 2005; Davis, 2002; Semple and Aitkin, 1980). Davis (2002) claimed that especially closed units in the low frequency range resemble responses from the DCN.

The results shown in figure 1.8 indicate, however, that IC units do not simply reflect the behavior at lower levels. The CMR units of the present study usually show both, an increase in response to the signal as the signal level is increased as well as a decrease of the response to the masker. This seems to be inconsistent with the wideband inhibitor hypothesis at least if it is assumed to be nearly instantaneous. One may argue that such a decrease in response to the masker envelope cycle following a signal tone pip response originates from a certain fatigue of the unit due to firing to the preceding tone pip. Then, the reduction should be present in all masking conditions and should be stronger when the signal response is stronger. This is not observed in the data. For example, figure 1.3b shows a stronger increase in spike rate during the signal interval for the RF condition than for the CM condition, whereas the decrease in the response to the masker was stronger for the CM condition. In figure 1.4, the response to the masker was suppressed before any response to the signal was found.

It might be argued that the reduction of masker response in the presence of the signal could be explained on the basis of forward masking. According to Eggermont (2014), forward masking can be defined as a reduction of response to a signal caused by a preceding stimulus (see also Young and Sachs, 1973). For the present masker one would need to assume that the masker following a signal pip is masked. However, several findings argue against forward masking as the single source of the masker response reduction. First, the unit shown in figure 1.4 ceased to respond to the masker without showing a response to the signal. Second, the decrease in the spike rate for the CM masker found in the population analysis is stronger

than for the RF masker. This effect can only be explained on the basis of forward masking if the response to the signal is stronger in the CM than in the RF masker. Envelope locking suppression (ELS) as observed at the level of the auditory cortex (AC) requires a decrease in response of the masker envelope as the signal level is increased. This is also found for many of the present IC units showing CMR. However, a direct comparison is difficult since different maskers were used. The baseline condition in the studies on ELS was an unmodulated noise, whereas in the present study the maskers in all conditions were modulated. As a consequence, one may predict on the basis of the hypothesis of ELS a reduction of the response to the masker envelope in all conditions when the signal is present. This is consistent with the present study, where the increase in signal level decreased the masker response, i.e. showed a masker suppression (figure 1.8). Part of the definition of ELS was a reduction of masker response in the presence of pure tones at level that were below threshold in quiet for this signal. Such responses below the detection threshold of the unit were not investigated since they were only expected to be found at higher levels such as the MGB and the AC (Las et al., 2005). In our view, such a change in response pattern is not crucial for the neural correlate of the psychoacoustical effect of masking release due to comodulation. Psychoacoustical masked thresholds are usually well above thresholds in quiet. To account for the psychoacoustical effect, it is only required that the reduction is more pronounced when the masker is comodulated compared to the other conditions (RF and CD). A stronger suppression than facilitation is only observed for three CMR units at the 0 dB SNR while the vast majority showed stronger facilitation (23 of 26). The data indicate that masker suppression and signal facilitation are both observed as a neural correlate of masking release due to comodulation at this level of the auditory pathway (see figure 1.8).

## Chapter 2

# Influence on comodulation masking release and possible predictors

### 2.1 Outline and key findings

- Using the data and results from the previous chapter, the predictability of Comodulation masking release (CMR) is investigated
- It is shown how the frequency response area (FRA) and the Spectro-temporal receptive field (STRF) provide information for the CMR behavior of a unit
- The similarity of a response of a unit between the start and the ongoing signal correlates with its ability to show CMR
- Close spectral distance of the flanking bands to the masker are beneficial for CMR units

## 2.2 Abstract

About one third of all units in the inferior colliculus (IC) show Comodulation masking release (CMR) when tested with a flanking-band paradigm. These units represent CMR differently than those in the cochlear nucleus (CN), and the differing representation is consistent with the expected change along the auditory pathway. In this chapter, the predictability of the CMR response of a unit is investigated depending on other properties of a unit, the FRA, the STRF and the response to the masker only. It is shown that certain shapes of frequency response areas make the occurrence very unlikely, thus making other shapes a necessary condition. Similar observations are made for spectro-temporal receptive fields. By means of a Similarity Index (SI) that compares responses of a unit to different stimuli against each other, it is shown that while the masking of the on-frequency masker in the comodulated condition is of less importance, certain features in the codeviant masker response correlates well with CMR, making this response a possible predictor. As this effect takes place within the first 100 ms of signal presentation, it also shows the importance of the onset response of a unit. Another factor influencing CMR that is not directly related to the response of a unit to different stimuli or the masker itself is the distance between the closest flanking band and the signal frequency. As CMR effects correlate with this distance, a systematic investigation for 11 units was carried out showing that close flanking bands are beneficial for CMR, as suggested by psychophysical findings.

## 2.3 Introduction

Comodulation masking release (CMR), as it was introduced in the previous chapter, is the term for investigating the detectability of a signal depending on a masker structure. It is of importance in signal detection and it has been shown that the preferred maskers, comodulated ones, occur in natural sound environments (Nelken et al., 1999). In figure 1.7b it was shown that the effect, if evaluated as suggested, holds for the entire population of the 83 measured units. Comparing the reduction of the masker response with the increase of the signal response, the former was less prominent, nevertheless important (see, for example, the units shown in figures 1.3 and 1.4). After the detailed description of the effect, the question arises as to what other parameters of the unit might reveal about its ability to show CMR-like behavior. For this, a categorization into internal, external and mixed parameters or properties is suggested. The suggested internal parameters that are investigated are the frequency response area (FRA), estimated from pure tone stimulation, and the Spectro-temporal receptive field (STRF), calculated from frequency modulated tones. External and mixed parameters are not as easy to define; while the absolute distance of the closest flanking band clearly is an external parameter, the relative distance could be defined as either mixed or external. This is because the relative distance depends on the stimulation frequency, which might be chosen independently from the unit, but actually depends on the best frequency, and thus, on an intrinsic parameter.

The FRA is a standard method to characterize units along the auditory pathway (Ramachandran et al., 1999; Egorova et al., 2001; Young et al., 2005; Hernández et al., 2005, see figures 1.2-1.6 panel a), as they hardly need any computational power and depend only on a few, easily reproducible parameters. In lower areas than the inferior colliculus (IC), units can even be uniquely identified by their FRAs (see, for example, the scheme reviewed in Young et al., 2005), but the situation in the IC itself is more complex. In this chapter, a classification as suggested by Hernández et al. (2005) was used, though different ones have been suggested as schemes for the IC as well (see Egorova et al., 2001; Ramachandran et al., 1999). These classifications (as well as the one used here) only serve as a guideline, as Palmer et al. (2013) showed that FRAs in the IC are continuous, rather than of discrete class. Still, the categorization offers some criteria for preferred features of CMR units.

Another means to describe units in the auditory pathway is the STRF (Aertsen and Johannesma, 1981; Klein et al., 2000). Unlike the FRA, the STRF actually was intended to be a predictor for arbitrary stimuli (Eggermont et al., 1983; Klein et al., 2000) but it is more difficult to reproduce than the FRA. Further, it has more parameters and requires more computational power, as well as knowledge about the evaluation. For example, the stimuli, to obtain the STRF are ambiguous and for two different stimuli only two thirds of the units in the IC show identical results (Escabi and Schreiner, 2002). The STRF theory for the IC has been expended by Lesica and Grothe (2008), where a good prediction of the response of a unit was made using level dependent STRFs.

One advantage of the STRF, when compared with the FRA, is that it is able to show suppressive frequencies. The importance of these is investigated with a similarity index (SI, Las et al. (2005)). The Similarity Index (SI) is also used to investigate the onset effects of

the masker-only responses by comparing a response of a unit to the interval 0-100 ms of one masker with the interval 100-200 ms of the same masker.

Finally, the influence of the distance of the closest flanking band is investigated. For a flanking-band paradigm, the distance of the flanking band to the signal is of importance, according to Schooneveldt and Moore (1987); Piechowiak et al. (2007). Dau et al. (2013) showed how a non-linear model of the periphery could account for such effects. This is tested by systematically varying the distance of the flanking bands and testing for CMR effects.



## 2.4 Material and methods

The experimental design was as described in chapter 1.4 and the results (particularly for the masker-only (MO) responses and the CMR behavior of units) will be used. Additionally to the 81 units there, 11 further units were measured in two of the 25 animals in order to investigate the frequency dependence of CMR. For these units, the stimuli were repeated 20 times and the signal was played at four attenuations (0, -10, -20 dB and the MO condition) and the closest flanking band (FB) was placed 0.2, 0.5, 1, 2, 3 and, if the recording stability made it possible, 4 kHz above the signal frequency. The other FBs were placed in 0.1 kHz steps above the closest one. The silence interval between the stimuli was at least 500 ms.

Out of the pool of 81 units, for some with high stability, STRF's were recorded using superpositions of frequency modulated tones (FM Banks) at random starting times. A detailed description of the stimuli can be found in Meyer et al. (2014). Briefly, the FM Banks covered different frequency ranges randomly chosen from the range of 0.5 and 16 kHz. All frequency modulations were linear. The duration of one frequency modulated tone was 100 ms and they were grouped into continuous blocks of 10 sec. There were five different blocks and every block was repeated five times. The attenuation of the FM banks was the same as that of the on-frequency masker to ensure that both stimuli were played at a similar sound pressure level. For the analysis, the maximal informative dimension method (MI) was used (Sharpee et al., 2004). This method found the optimal filter for the given response that was elicited with the incorporated stimulus ensemble.

*Data Analysis.* Apart from an investigation into CMR behavior of a unit as described in chapter 1.4 which led to the two subgroups of CMR and non-CMR units, further distinctions were made. Units that responded with more than half the spikes in the first 100 ms in the combined three MO conditions were supposed to be unable to show CMR and were exempted in some of the later group comparisons where appropriate.

Units were classified according to the shape of their FRA and were divided into eight different categories. "V-monotonic" units had a V-shaped FRA and increased their firing rate at higher sound levels, whereas "V-non-monotonic" units responded with lower firing rates to higher sound levels. "Multi-peaked" units had several distinct regions in the FRA where they responded. Purely "inhibitory" units had a high firing rate which was reduced as a response to the pure tone stimulation. The categories are similar to those in Hernández et al. (2005). However, in contrast to Hernández et al. (2005) the present study did not differentiate between "high" and "low tilted" units. Two other categories in Hernández et al. (2005), "narrow" and "closed" units, were combined in the present study because recording stability did not always allow a complete characterization, i.e., a unit with a "narrow" FRA could have revealed a "closed" FRA, if higher stimulus levels had been tested. Furthermore, fully characterized narrow or closed units turned out to show similar response types in the CMR experiments. U-shaped units responded over a wide range of frequencies that was only slightly narrowed for low sound levels. Units not falling into one of these categories were classified as mosaic. For the final steps of data analysis, "mosaic", "tilt", "inhibitory" and "multi-peaked" units were combined in the category "other".

Additionally, the units were also classified according to their firing pattern at the best

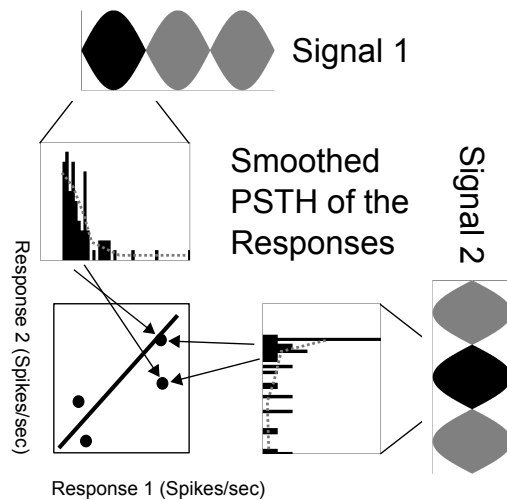


Figure 2.1: The Similarity Index (SI) compares two signals of equal duration against each other. High values mean that the second signal elicited a stronger response, while low values indicate a stronger response to the first signal. Mathematically it is the linear regression for the second response depending on the first.

frequency (BF) and best level as in Le Beau et al. (1996). The BF and best level are defined as the frequency and attenuation that elicited the maximum number of spikes during the 100 ms interval of the pure tone stimulation. Units were classified as (i) "onset" units when they only spiked at the beginning of the stimulus, (ii) "pauser" units when they showed an onset peak followed by a brief period without spikes before they started to gradually fire again, (iii) "on-sustained" when they continued to fire at a lower rate after a pronounced onset response, (iv) "sustained" units if they fired in response to the signal without pronounced peaks and (v) "chopper" units when they had several clear peaks in their response. Unlike Le Beau et al. (1996), no distinction was made between "chopper" and "onset-chopper" units because in the present study their firing patterns were too similar for a clear dissociation.

The data were analyzed with respect to the inhibition of the on-frequency masker when comodulated FBs were added (i.e. comodulated (CM) vs. RF) and the possibility of predicting CMR behavior by means of the response to the first cycles (see figure 2.1) for the different masking conditions. To this end the similarity index (SI, Las et al., 2005) was calculated that compared two peri-stimulus time histograms (PSTHs) (bin width 10 ms) of equal duration. Each PSTH was smoothed using a moving average technique that had an averaging window of 50 ms. The two smoothed responses of equal duration were plotted against each other and linearly interpolated (polyfit, Matlab R2006a). The slope of this linear interpolation was the SI (Las et al., 2005). An SI of one would indicate the same responses to different stimuli. Slopes of above or below 1 indicate a stronger or weaker response, respectively, for the condition plotted on the ordinate.

## 2.5 Results

### 2.5.1 Narrow and V-shaped FRAs are most likely to show CMR

Table 2.1 summarizes the responses of all recorded units. For each combination of type of FRA and type of PSTH, three numbers are shown: (i) Number of units showing CMR according to the criterion of the present study, (ii) number of units that did not respond with more than half the spikes in the first 100 ms of the MO conditions (see *Materials and Methods*) (iii) the total number of units. The failure to respond to the entire signal was distributed almost evenly across all response-area and firing-pattern types. Most of the recorded units were classified as "sustained". With 25 units they comprised almost a third of the total measured units. More than 40 percent of the units were classified as CMR units (11 out of 25). Fifty percent of all "pauser" units were classified as CMR units. About a third of all "onset" and "on-sustained" units were classified as CMR units. Only a small portion of the "chopper" fulfilled the criteria for CMR units. They were the second most common unit type. Units with a "narrow" or "closed" FRA are most likely to show CMR but only about ten percent of all units had such a FRA. More common were V-shaped FRA. Forty percent of all units with a monotonic "V shaped" FRA and 45 percent of those with a non-monotonic V-shaped FRA showed CMR. 29 of the 81 units did not fall into the categories and were thus classified as having "other" FRA. Of these units, five were "inhibitory", nine were "multipeaked" and seven were "tilt", the remaining ones had a "mosaic"-like FRA. Seventeen percent of them showed CMR. Units with an "other" type of FRA were more than a third of the entire measured population while "narrow / closed" and "U-shaped" units were the smallest fractions. None of the units with a U-shaped FRA showed CMR.

For this table, the FRAs were grouped according to their linearity. While "other" units can be characterized as being less linear than the other categories, they also show the least chance of performing CMR behavior. This does not seem to hold for "U-shaped" units though. Their responses are more linear according to this definition, but are less likely to be a CMR unit. Using a Fisher test, this difference in the amount of CMR units between the linear and non-linear FRA's is significant.

Comparing the different firing patterns further, figure 2.2 shows the middle panels from figure 1.8 (the CM responses) with the results of 1.8a at the top and the results of 1.8b at the bottom. The responses are differentiated with respect to their response types. Units characterized as "Onset", "On-sustained" and "Pauser" were more phasic in their behavior and thus grouped together. Their results are shown by the continuous line. The relative increase in spike rate for these units increased from 0.03, to 0.75 in the signal interval as it decreased from -0.05 to -0.47 in the masker interval. The second group of tonic units consisted of "sustained" and "chopper" ones. The increase in the signal interval for these units was from 0.21 to 0.69 while their decrease in the response to the masker interval was from -0.05 to -0.14. However, this latter decrease was only observed when the signal level was raised from -10 dB to 0 dB signal-to-noise ratio (SNR). In contrast, the response of the phasic units decreased continuously between the signal level of -30 dB and 0 dB. In the signal interval, the increase in spike rate was continuous for the tonic units, whereas it was almost constant for the phasic units.











	VM 	VNM 	N/C 	U 	Oth 	$\Sigma$
On 	2/3/4	0/1/1			2/3/6	4/7/11
Pa 	1/1/1	3/3/3			0/3/4	4/7/8
OS 	2/5/6	1/3/3	1/1/1	0/2/2	0/3/3	4/14/15
Su 	6/10/11	1/2/2	1/2/2	0/2/2	3/7/8	11/23/25
Ch 	0/4/5	1/3/4	2/3/3	0/2/2	0/7/8	3/19/22
$\Sigma$	11/23/27	6/12/13	4/6/6	0/6/6	5/23/29	26/70/81

Table 2.1: A table showing the number of CMR units (left), the number of units that were able to respond with less than half of the total spikes in the first 100 ms MO conditions (middle) and the total number of units (right). Abbreviations of the different firing types: Su - "sustained", OS - "on-sustained", On - "onset", Ch - "chopper", Pa - "pauser". Abbreviations of the different FRA types: VM - "V-shaped monotonic", VNM - "V-shaped non-monotonic", N/C - "narrow" or "closed", U - "U-shaped", Oth - "other" (including "inhibitory", "tilt", "multi-peaked", "mosaic").

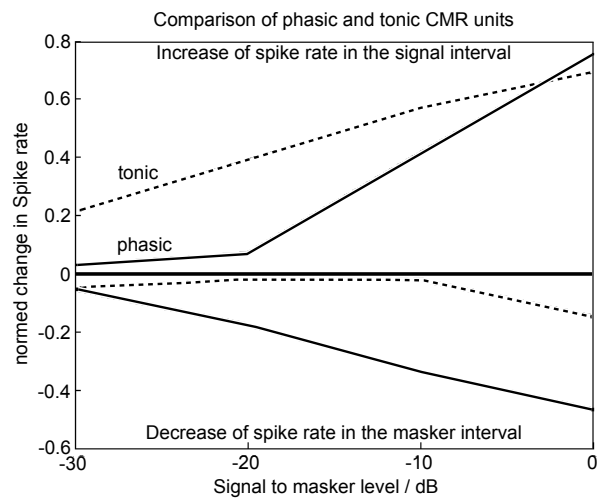


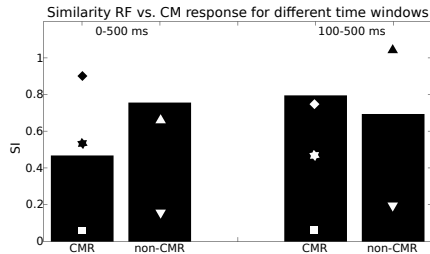
Figure 2.2: The results of the CM masker in figure 1.8a at the top, the normed increase of spike rate in the signal interval. The units are split up into two different subpopulations, one being more tonic units (dotted line) and one being more phasic units (continuous line). The bottom part of panel b shows the same, but for the normed decrease of spikes in the masker interval as in figure 1.8b.

## 2.5.2 Can the similarity index predict the units' ability to show CMR?

To answer this question the Similarity Index (SI) was calculated for the responses to the masker-only (MO) in two ways. In the first analysis the SI between different masking conditions was calculated. This was either done for the whole duration of the stimulus (0-500 ms) or for the four last cycles of the masker envelope (100-500 ms), i.e., excluding the onset response of the unit. The second analysis focused on changes in the firing pattern from the first cycle (0-100 ms) to the second cycle (100-200 ms) of the masker envelope within a masking condition. The SI was only calculated for the 70 units that showed a response to the amplitude modulation of the on-frequency masker (see Materials&Methods). If the SI is sensitive to the ability to show a CMR-like behavior then SI of the 26 CMR units should be significantly different from the SI for the remaining 44 units not showing CMR.

Figure 2.3 shows the SI between the reference (RF) and the CM (panel a) and RF and codeviant (CD) (panel b) maskers. Each panel shows four columns. The first two indicate the results for the whole duration of the stimulus, the other two when the analysis was restricted to the last 400 ms of the stimulus. The SI between the CM and the RF masker was always smaller than one indicating that for all units the response in the CM condition was weaker than that to the RF condition. For CMR units, the mean SI was 0.47 for the comparison of RF and CM condition when the whole duration of the stimulus was analyzed. It was considerably lower than that for the non-CMR units. This may indicate that the response to the on-frequency masker was suppressed by the presence of the FBs in the CM condition. However, the effect was not significant. Interestingly, similar SIs were obtained when the analysis was restricted to the last 400 ms of the stimulus which would not be expected on

(a)



(b)

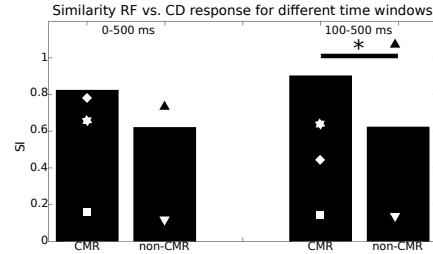


Figure 2.3: SI between different signals (RF and CM and RF and CD) for different time intervals (0-500 ms and 100-500 ms). Symbols as defined in figure 1.8. Downwards pointing triangle are magnitudes of the results for the unit shown in figure 1.5, upwards pointing triangles are the SI values from the unit shown in figure 1.6.

the basis across-frequency suppression. When SI was calculated for the comparison of RF and CD condition (panel b of figure 2.3) non-CMR units had a smaller SI than CMR units. The difference was significant when the interval 100-500 ms was analyzed.

Figure 2.4 shows the SI between the first (0-100 ms) and the second (100-200 ms) cycles of the on-frequency-masker envelope. For the RF and the CM masker, the units seem to be less responsive to the second cycle of masker envelope than to the first, regardless whether or not the unit showed a CMR-like behavior. This was different for the CD masker, in which units that showed CMR tended to respond with more spikes to the second cycle ( $SI > 1$ ). In contrast, non-CMR units had a SI of 0.5, i.e., similar to the SI for the other two masking condition. The difference for the CD masker for the CMR units and non-CMR units was statistically significant.

The comparison of the CMR units and the non-CMR units reveals that (i) the most informative measure of a unit's CMR response is the similarity between the initial and the following 100 ms of the response to the CD masker: if the similarity is high then the unit is likely to show CMR. (ii) A high similarity between the responses to the RF and the CD masker also is a good indicator while the suppression of the on-frequency masker in the CM condition is not a reliable criterion to differentiate between units showing CMR and those that do not show CMR but follow the envelope of the masker.

### 2.5.3 The influence of the carrier and of the distance of the closest FB

The average frequency of the target signal was 1.41 kHz (minimum 350 Hz, maximum 4 kHz) for all CMR units while that of the remaining units was 1.9 kHz (minimum 250 Hz, maximum 6.7 kHz). This difference was not significant. Further, the influence of the closest FB divided by the target signal was investigated. The mean for the CMR units was about

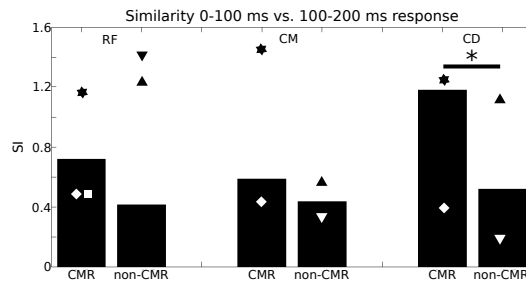


Figure 2.4: SI between different intervals (0-100 ms and 100-200 ms) of identical signals (RF, CM or CD masker). Symbols for the different units as in figure 2.3. For the unit in figure 1.5 calculating the SI was not possible in the CM condition as there were too few spikes.

1, meaning that the FB were one octave higher or lower than the target signal (minimum 0.3, maximum 3.8). For non-CMR units, this distance was 1.3 (minimum 0.2, maximum 4.7). Again, the difference was not significant. However, as Schooneveldt and Moore (1987) showed in their psychophysical study, there is an influence of the distance of the nearest FB on CMR. Thus, in a modified paradigm, eleven units were measured for various spectral distances of the FBs (see materials and methods). From these eleven units, only two showed CMR like behavior for any of the FB positions, i.e., a smaller portion than found in the overall population. This may be partly due to differences in the experimental paradigm between the 11 units and the other units. For the eleven units, 20 repetitions and 4 instead of 5 SNR ratios were used and all FBs were placed at higher frequencies. Only FBs spectrally above the signal frequency were used since (i) the distance was varied up to 4 kHz and (ii) using lower frequencies for lower distance would have complicated the experimental conditions.

To test the influence of the distance of the closest FB, eleven units were measured for various spectral distances of the FBs (see materials and methods). Figure 2.5 shows the response of one of the two CMR units. The nearest FB was positioned at 0.2, 0.5, 1, 2, or 4 kHz above the signal frequency (see panel a). This corresponds to a FB frequency relative to the signal frequency of 0.4, 1, 2, 4, 6, and 8. The unit showed a CMR-like behavior if the closest FBs were positioned 0.2 or 1 kHz above the signal frequency. In panel b, the increase in the response difference between the signal and the masker response is shown. Black triangles denote the CM, circles the RF and white triangles the CD masking condition. For the 0.2 kHz FB distance, the difference was best for all signal-to-noise ratio ratios in the CM condition. Interestingly, the response in the CD condition was better than that in the RF condition. For the 0.5 kHz FB distance, the response in the CM condition was not highest at 0 dB signal-to-noise ratio ratio, and, though it was higher than in the other masking conditions for the other two ratios, this effect was not significant and at this distance the unit did not show CMR-like behavior as defined here. At 1 kHz FB distance, the response in the CM condition clearly was higher than that in the RF and the CD conditions and the CMR-like behavior was most prominent. At 2 kHz, the response in the CM condition still was the best but failed to reach significance and at higher distances, the CMR-behavior vanished.

Panel c shows the SI from figure 2.3 in the lower two plots. While the SI between the RF and the CM masker confirm the trend observed for CMR units, namely lower values for CMR-like behavior, the situation is different for the SI between the RF and the CD masker. The number is close to the mean if CMR-like behavior occurs, but higher when the unit failed to show CMR. The top plot of figure 2.5c shows the significant result of figure 2.4. The trend seen in figure 2.4 is confirmed for 1 kHz FB distance, but not for 0.2 kHz distance.

#### 2.5.4 Spectro-temporal-receptive-fields of selected units

For some units, the STRFs for one level were recorded in addition to the FRA. The thorough investigation of STRFs with several attenuations as suggested by Lesica and Grothe (2008) was not possible due to stability limitations. In comparison to the FRA, the STRF has the advantage of not only indicating suppressive areas, but also the temporal dynamics.

Figure 2.7 shows what kind of STRF would be expected for the different effects encountered in the IC. An STRF can be used as a predictor by convolving the spectrogram of the stimulus in time with the STRF and integrating it over all frequencies (top right). The facilitation of the signal in the CMR paradigm of this study would not require temporal effects (figure 2.7, top), unlike the suppression of the masker response (middle). A suppressive field following the excitation would cause the unit to reduce its response in the masker intervals that follow a signal. The combination of both, particularly with a stronger suppression in the CM than in the RF masker, would require a temporal precise interplay of facilitation and suppression in order to be accounted for linearly (bottom). A non-linear interaction between the signal and the FB frequencies would require less temporal precision.

Figure 2.6 shows STRFs of the units shown in figures 1.2-1.5. Black dots denote the signal frequency, white dots the FBs. Figure 2.6a shows the STRF of the unit from figure 1.2, which has a response similar to the one suggested in the top part of figure 2.7. Large suppressive fields at frequencies of the FBs could explain the masking of the on-frequency masker in the CM condition. The response to the RF maskers is difficult to explain on the basis of the STRF, as it does not give a reason for the inability of the unit to follow the amplitude modulation more closely. Figure 2.6b shows the STRF for the unit shown in figure 1.3. This unit reduced its response to the masker in the CM condition. This would be resembled by the suggested situation of the bottom panel of figure 2.7. A shorter excitatory field compared to that in figure 2.6a could account for the correct representation of the AM envelope of the RF masker condition in the MO. Furthermore, the inhibitory side fields are not very prominent, which confirms the rather weakly masked on-frequency masker in the CM condition. The reduction of spike rate in the CM condition for increasing signal levels cannot be accounted for by the STRF. Figure 2.6c shows the STRF of the unit shown in figure 1.4, which also shows similar effects to those suggested in the bottom panel of figure 2.7. The reduction of the response in the masker interval for all masker conditions could be explained by the inhibitory field following the excitation at the signal frequency. The inversed fields at the FBs are in accordance with the increased signal response in the CM condition (see also figure 2.7). Figure 2.6d shows the STRF of the unit shown in figure 1.5. The excitatory field of the STRF is followed by a long suppressive one, and thus disabling any signal response. The neighboring suppressive fields for the FBs do not have an excitatory component following



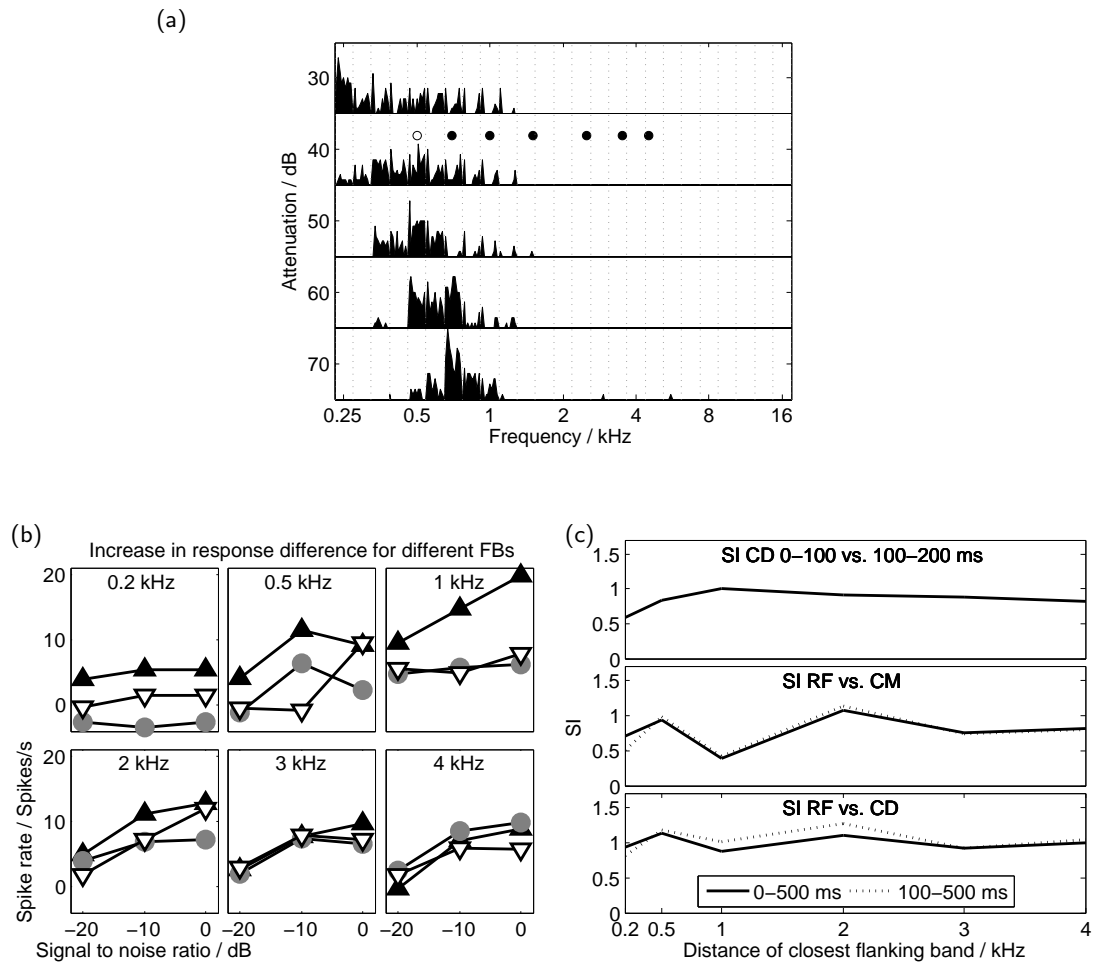


Figure 2.5: Unit for which the spectral position of the FBs was varied. (a) The FRA of the unit with a white dot at the stimulation frequency and black dots for the starting points of the closest FB. (b) Result for the individual FB distances (as in figure 1.7). (c) The similarity indexes: Top the CD 0-100 vs 100-200 ms intervals, middle the similarity between the RF and the CM condition (solid:0-500 ms, dotted: 100-500 ms) and bottom the SI between the RF and the CD condition, same time frames as for the middle panel.

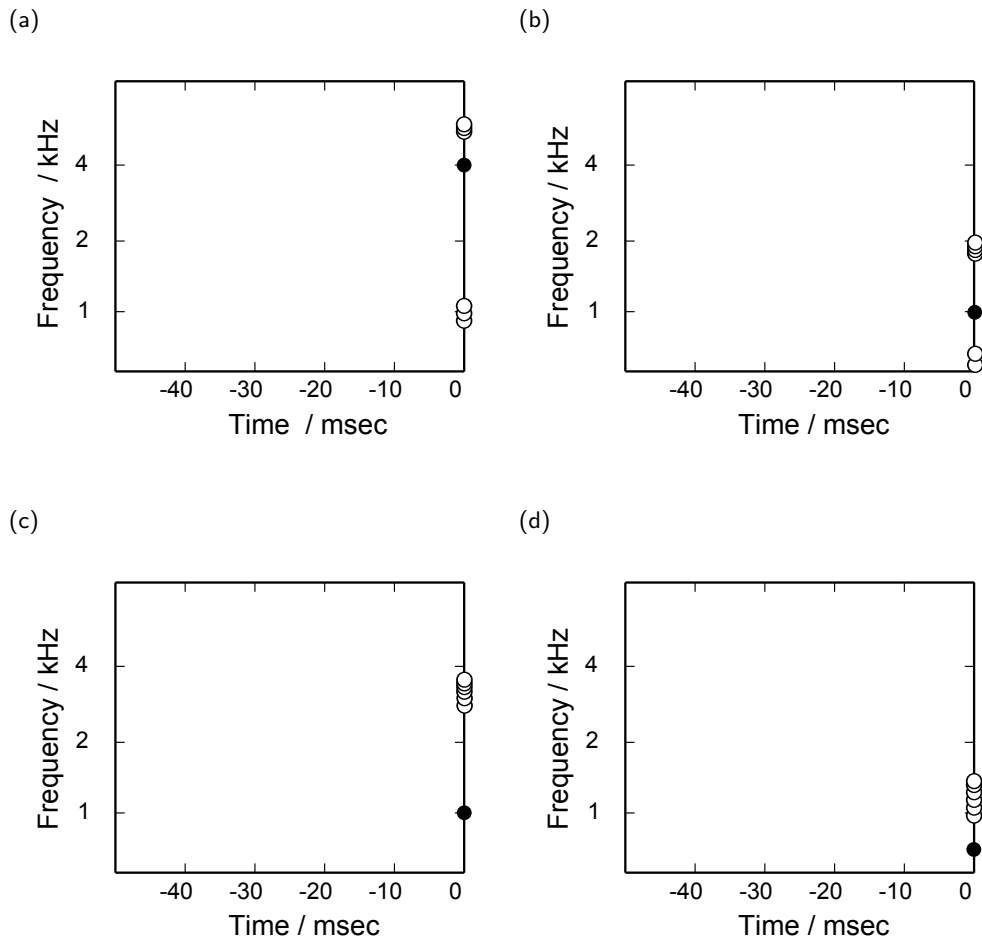


Figure 2.6: Examples of STRFs from different units. (a) shows the unit from figure 1.2, (b) the unit from figure 1.3, (c) the unit from figure 1.4 and (d) the unit from figure 1.5.

them that could cancel out this effect. Therefore, the STRF correctly indicates that the unit is not able to perform CMR. The STRFs cannot account for the differences in the responses in figure 2.4, as they generally do not account for onset responses. Thus, they would be unable to predict different responses to two identical consecutive intervals of a signal.

## 2.6 Discussion

In this chapter, it was shown that CMR could be explained by a linear STRF mechanism, even in the form of suppression and facilitation (see figure 2.7). Furthermore, units that did show CMR tended to have an FRA which either had a constant frequency response or increased its frequency range with increasing sound level, and thus had a more linear response. The response to the codeviant masker in the first 200 ms already contained enough information to give a good estimate as to whether a unit was capable of showing CMR. The expected results for frequency dependence could not be confirmed entirely.

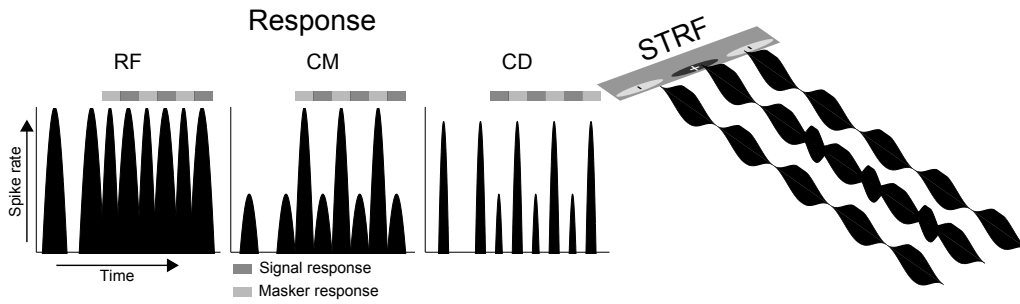
### 2.6.1 The predictability via the FRA and the STRF

In the last chapter, it was shown that the suppression of the comodulated masker response (compared to the other maskers) cannot be solely based on forward masking or onset effects. Thus, the hypothesis that the IC simply inherits the representation of CMR from the CN is rejected. However, the DCN only has excitatory projections to the IC (Malmierca et al., 2005; Davis, 2002; Semple and Aitkin, 1980) leaving the source of the suppressive effects open. Davis (2002) even claimed that especially closed units in the low frequency range resemble responses from the DCN. Figure 2.3 seems to further confirm this assumption, since it shows the importance of suppressing the on-frequency masker similar to the effect found at the level of the CN (the SI between RF and CM is lowest for CMR units). However, based on the results by Palmer et al. (2013), where it was shown that IC response form a continua and, hence, cannot solely represent lower stages, it seems clear that the idea of areas in the IC receiving input only from one other source does not hold.

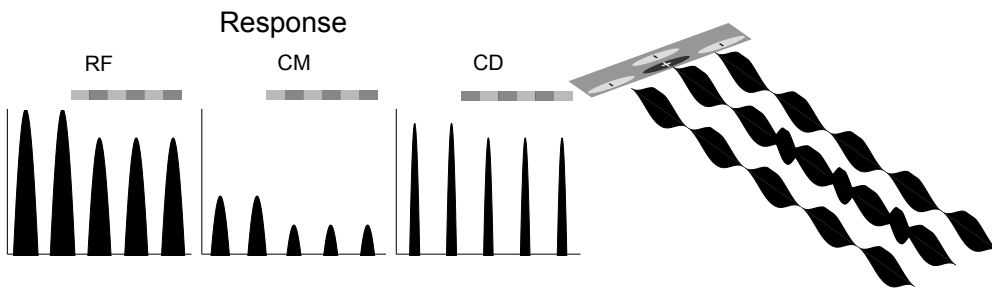
The possibility that CMR in the IC including facilitative and suppressive effects could be explained by an STRF (see figure 2.7 and, as an example, figure 2.6c) shows that this effect could be explained by a linear feed forward network because the STRF is a linear predictor. Table 2.1 confirms the notion that CMR is a more linear process if the responses of "V", "U" and "I"-shaped FRAs are considered more linear. An even better descriptor could be derived from STRFs determined at different sound levels (see Lesica and Grothe, 2008) but this would raise the question of how to define the appropriate sound level for a fully amplitude modulated tone as used in the present CMR paradigm.

The general importance of inhibition in the IC has been reviewed by Pollak et al. (2011), for units with non-V-shaped (including the "narrow" and "closed" ones) response areas by (LeBeau et al., 2001). Given that the DCN projections to the IC are only excitatory, these studies indicate that the inhibition of the masker signal has a different origin. The transformation of the stimuli response from the CN mainly showing signal facilitation for comodulated maskers to the IC which displays both signal facilitation, as well as masker suppression can be accounted for by assuming a simple linear parallel feed-forward network. A CMR unit from the CN projects has excitatory projections to two neurons, one directly in the IC and another one that projects to an interneuron which then inhibits the IC neuron. This network accounts for the reduction in the spike rate to the masker since it includes a temporal delay. Further, since a feed-forward network is essentially, linear it is plausible that many effects can be captured by an STRF.

Suppressive side fields in the STRF make the signal more prominent



A suppression following the excitation could decrease the masker response



Precise timing or non-linear interaction can account for the combination

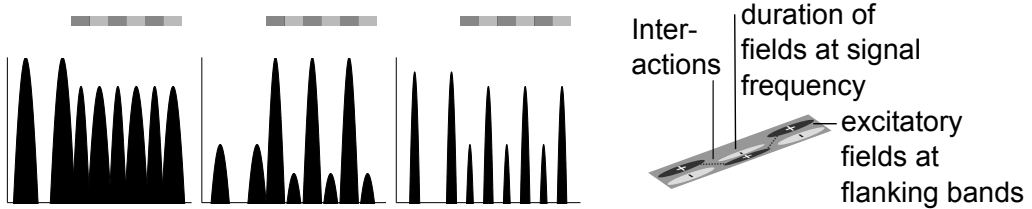


Figure 2.7: Schematic drawing of possible forms of the STRF and their consequences for CMR behavior.

## 2.6.2 Spectral distance

Figure 2.5b seems to confirm the psychophysical findings that flanking bands with little spectral distance to the signal are beneficial for CMR (Schooneveldt and Moore, 1987; Hall III et al., 1990; Piechowiak et al., 2007) which could be accounted for by the model from Dau et al. (2013). This notion might be biased, though, as the relative distances show no significant influence, a result that would have been expected. Another explanation for the significant correlation between absolute distance and CMR behavior could be due to the "different" units in table 2.1. Due to the shape of their FRAs, it was difficult to find a non-responsive region, and thus the flanking bands sometimes were placed in a larger spectral distance than for other units.

## 2.6.3 High correlation between the first intervals of the CD response and CMR effects

Figure 2.4 could be explained by the similarity between the structures of a signal in a CM masker and the starting interval of the CD masker. Keeping in mind that the CD masker consists of an on-frequency masker identically placed as the signal and codeviant flanking bands, the start of the CD masker corresponds to the response to one signal pip. The response to the second cycle of the AM, the interval from 100-200 ms, is the response to the same pip, but now it has been preceded by one cycle of flanking bands. This means that the response to the signal is better if the unit is first stimulated with 100 ms of flanking bands. The response to the signal in the CM condition has a similar context, the only difference is that the flanking bands preceding the signal now also contain the signal itself, the on-frequency masker. Note that the time of this effect is 50-100 ms and thus this is longer than the experimental results that the STRF 2.6 could detect.

## Chapter 3

# Turning a spectro-temporal filter into a Volterra operator in the time domain

### 3.1 Outline and key findings

- The assumptions underlying the Spectro-temporal receptive field (STRF) calculation are explained
- A general method to calculate Volterra kernels on the time domain that correspond to STRF operators is introduced
- The suggested method is applied to toy data that resembles true STRF forms

## 3.2 Abstract

The Spectro-temporal receptive field (STRF) is a common descriptor for single unit responses in the auditory pathway. One of its main purposes is the prediction of the unit's response to an arbitrary stimulus on the basis of its Time-frequency representation (TFR). The prediction works well if testing for stimuli similar to those that are used for the estimation of the STRF, but may fail to do so for other stimuli. When introduced, the STRF was derived from the 2nd-order Wiener operator, which is related to the Volterra series and is used to map one function to another one. The derivation is difficult, however, as the STRF is defined in the time-frequency space, which itself is not uniquely defined, and necessarily lacks some properties other two-dimensional spaces possess. In this chapter, a new numerical way of decomposing an STRF into a Volterra operator is shown. This is only possible when the precise form of the mapping from the time to the time-frequency space is known. The method can be generalized, though, and only requires the TFR to be causal and to depend on a linear transformation followed by a static non-linearity. A reduction of the sampling rate in the TFR is possible. Using toy data, it is shown that the STRF operation in the time-frequency space can be exactly replaced by a Volterra operator (or, in this case, equivalently, a Wiener operator) if the instantaneous non-linearity is of an integer power form is used, e.g., in the power spectrogram. For other non-linearities, a method of deriving a corresponding approximation is introduced, which could apply if working on an absolute value or logarithmic TFR. The implications of these comparisons are a more thorough understanding of the STRF and the implicit assumptions made when using an STRF.

### 3.3 Introduction

In the previous chapters, figures 1.1 and 2.7 were examples of a time-frequency representation (TFR) and a prediction made based on it, respectively. The general principle of a TFR is that it indicates the temporal and spectral content of an originally temporal signal simultaneously (Gröchenig, 2001). Predictions based on the TFR require a background in both predictor (or black box) models and time-frequency representations.

#### 3.3.1 From the black box model to the spectro-temporal receptive field

The behavior of a neuron in the auditory pathway in response to acoustic signals is an example for a black box model. The input, the acoustic signal, is known, and the output, the spike train or membrane potential, can be measured. If it was possible to obtain a universally valid transformation function, the output to any given input could be inferred. In this spirit Eggermont et al. (1983) (but see also Johannesma, 1980) suggested the "subject centered approach" for the auditory system and this is the general goal of system identification for the nervous and other systems (see, for example, Hung and Stark (1977)). System identification itself as a concept was most prominently introduced to neuroscience by Marmarelis and Marmarelis (1978). The basis for system identification are the Volterra and the Wiener series (Volterra, 1930; Wiener, 1966) that, given some requirements, can be transformed into one another. Normally, these series representations map one dimension onto another one, but they use integral kernels with more than one dimension.

For units in the auditory pathway, a descriptor that was intuitive and potentially able to predict the output of a unit to unknown stimuli was the spectro-temporal receptive field (STRF, Aertsen et al., 1980). The STRF is related to one component of the aforementioned Wiener series; namely, its 2nd-order operator. A certain type of TFR is required to derive the STRF from the 2nd-order Wiener kernel, either a Rihaczek distribution (Eggermont, 1993) or transformations from the Wigner class of TFRs (Klein et al., 2000). The inverse operation, turning 2nd-order Wiener kernels into STRFs, was performed by Lewis and van Dijk (2004), where the requirement was that the frequency domain be made up from a short-term Fourier transform.

The reason why only certain TFRs are suitable to relate the 2nd-order Wiener operator to the STRF is that TFRs are ambiguous (Gröchenig, 2001). While there is a large class of TFRs (called Cohen's class, Cohen, 1995; Gröchenig, 2001) that can be converted into another, it is not necessarily the case that one TFR leads to another one. The range of their application can be seen when considering that Gabor (1946) suggested using TFRs in auditory research that originated in quantum mechanics (Wigner, 1932). The problem in deriving an auditory filter from Wigner class or Rihaczek distributions is that neither of them are physiologically motivated. On the other hand, TFRs that originate from Gamma-tone filterbanks (Aertsen and Johannesma, 1980; Lopez-Poveda and Meddis, 2001; Hohmann, 2002) are physiologically motivated but, to the knowledge of the author, do not fall into any class. Thus, though commonly used, a direct connection from an STRF computed using a Gamma-tone filterbank to the mathematically founded Volterra series cannot be drawn.

This chapter will present a way of computing equivalent or approximated Volterra kernels



to STRFs given very few requirements: (i) The spectrogram is created from linear filters followed by a non-linear operation and (ii) it may have a lower sampling rate than the signal. The first requirement does not hold for the Wigner distribution, but it allows for more TFRs that are physiologically motivated. The example shown in this chapter does not stick to the Gamma-tone filterbank, either, but the frequencies are not linearly spaced and the method could be applied to the Gamma-tone filterbank as well.

Using a Volterra kernel instead of an STRF is beneficial in two aspects. First, it turns a parametric model into a non-parametric one (and thus gives a precise insight into the utilized parameters) and second, by this it makes it possible to compare different data sets with each other.

Before presenting the precise algorithm and its implications on toy data, an overview of systems identification and time-frequency analysis is to be given.

### 3.3.2 A brief introduction to systems identification

The Volterra and Wiener series are both examples of dynamic, non-linear, time-invariant functionals (the last attribute being possibly a problematic one in the auditory system) that are able to mathematically describe input-output systems. This means that a known output,  $x$ , depends on the input  $\varphi$  in an unknown manner. Here, we want to focus on a one-dimensional output (e.g., a spike train of a neuron or its membrane potential) and on a causal dependence, meaning that for  $\varphi = \varphi(\tau)$  and  $x = x(\varphi(\tau))$  it follows that  $x(t) \neq x(\varphi(\tau))$  for  $t < \tau$ . In the first step, the difference between the two series is to be disentangled; the main need for the Wiener series in this study stems from the fact that it, unlike the Volterra series, has traditionally been related to the STRF.

#### *The Volterra series*

The Taylor series can approximate one variable depending on another one to a desired degree of accuracy. As it defines the behavior of one variable with respect to another one, it is a function. If, however, one function is to depend (or be mapped) on another one, this is done via the Volterra series. Given the above defined input and output, this would mean a mapping of

$$x(t) = \int_{-\infty}^t d\tau k_1(t - \tau)\varphi(\tau) + \iint_{-\infty}^t d\tau_1 d\tau_2 k_2(t - \tau_1, t - \tau_2) x(\tau_1) x(\tau_2) + \dots \quad (3.1)$$

As was shown by van Hemmen et al. (2000), the Volterra series can be regarded as the solution to the differential equation around a stable fixed point

$$\frac{d\mathbf{x}}{dt} = \mathbf{x}' = \mathbf{f}(\mathbf{x}) + \varphi \quad (3.2)$$

where the variable  $\mathbf{x}$  was put into vector form. This vector form is crucial as any  $n$ -th order differential equation can be transferred into a first order differential equation with the

dimensionality  $n$  (Hirsch and Smale, 1974). The representation in (3.2) gives rise to the one in (3.1) and is more general. It can be shown that equation (3.1) results from (3.2) if it is  $q$ -analytic in a certain surrounding, and given (3.2), a representation in the form of (3.1) can be found, but not vice versa (van Hemmen et al., 2000).

Besides the analytic representation that has been used up to this point, numerical or discrete forms are to be introduced as well. This means that if  $y(t)$  is a continuous signal,  $y_n = y(n \cdot \Delta t)$  with  $n \in \mathbb{Z}$  and  $\Delta t$  being any discrete step in  $t$  is its discrete representation. The equation shown in (3.1) becomes, for example, (Dodd and Harrison, 2002)

$$x_t = \sum_{m=-M}^t k_{1,t-m} \varphi_m \Delta \tau + \sum_{m_1=-M}^t \sum_{m_2=-M}^t k_{2,t-m_1-m_2} \varphi_{m_1} \varphi_{m_2} (\Delta \tau)^2 + \dots \quad (3.3)$$

As was demonstrated in Barrett (1963), equation (3.1) results from (4.1) by taking  $\Delta \tau \rightarrow 0$  and using an infinite memory length,  $M \rightarrow \infty$ . The numerical form poses several problems in the notation, most notably that the integral kernels  $k$  now have an extra index that identifies their dimensionality. Furthermore, the vector  $\mathbf{x}$  now might refer to a sequence,  $\mathbf{x} = (x_1, \dots, x_n)^T$ , instead of a multicomponent signal such as  $\mathbf{x}(t) = (v_1(t), \dots, v_n(t))^T$ .

#### *The Wiener series*

The representation of the Wiener series would look like (3.1). To avoid confusion, the kernel variables  $k_n$  (Volterra) might be exchanged with  $g_n$  (Wiener). The Wiener series is best characterized in the same manner as it was motivated when it was introduced, via its kernel estimation. Finding the kernels  $k_n$  in 3.1 was solved by Wiener (1966) by using white noise for  $\varphi(t)$ . Through the properties of white noise, a cross-correlation of the output  $x(t)$  and the input  $\varphi(t)$  in  $n$  dimensions results in the kernel,  $g_n$ . Generally,  $g_n(t_1, \dots, t_n) \neq k_n(t_1, \dots, t_n)$  holds, and comparing the two kernels is difficult. The difference can be seen in equation (3.2). As every first order differential equation with a dimensionality of more than two, or a differential equation of at least 2nd order, can have phase dynamics, the stability in the phase space becomes important. Using white noise for as the input  $\varphi$  would mean that the position in the phase space might be shifted and hence it could lead to different solution. Hence, the Volterra and the Wiener series cannot be compared on this basis. In the appendix of this chapter, it is shown how to compare them if both series are truncated.

### 3.3.3 Frequency and time-frequency representations

Time signals can either be completely represented as a time signal,  $\varphi(t)$ , or as its Fourier transform,  $\tilde{\varphi}(\omega)$ , also known as its frequency representation.

*The definition and selected properties of the Fourier transform*

The definition of the Fourier transform is ambiguous in the coefficient chosen for the integral (Gröchenig, 2001). If  $\varphi(t)$  is any given function, then its Fourier transform is defined as

$$\tilde{\varphi}(\omega) = \frac{1}{a} \int_{-\infty}^{\infty} dt e^{-2\pi i \omega t} \varphi(t) \quad (3.4)$$

where  $a$  may be 1 or  $1/2\pi$  or  $1/\sqrt{2\pi}$ . Its choice determines the parameter  $\tilde{a}$  of the inverse transform,

$$\varphi(t) = \frac{1}{\tilde{a}} \int_{-\infty}^{\infty} d\omega e^{-2\pi i \omega t} \tilde{\varphi}(\omega). \quad (3.5)$$

Mathematical operators as the derivative of a function or the multiplication of two functions have analogues in the frequency domain. One property that is of importance for the analysis of linear systems is the convolution of functions in the frequency domain. If  $\varphi(t) = \int \alpha(\tau)\beta(t - \tau)d\tau$  then

$$\tilde{\varphi}(\omega) = \tilde{\alpha}(\omega)\tilde{\beta}(\omega). \quad (3.6)$$

On the other hand, if a signal is multiplied in the time domain, then this corresponds to a convolution in frequency.

*Two Dimensions from one - selected examples for TFRs*

The time signal  $\varphi(t)$  does not give an intuitive information about its spectral form  $\tilde{\varphi}(\omega)$  (at least for more complex signals), and vice versa. For example, a signal which is zero at all positions except for an interval  $[-t, t]$  where it consists of many different frequencies is difficult to estimate in the frequency domain. Time-frequency representations combine both properties and thus are necessarily two-dimensional. For all TFRs, two related principles hold: (i) As  $\varphi$  decays quickly if  $\tilde{\varphi}$  is smooth and vice versa, (ii)  $\varphi$  and  $\tilde{\varphi}$  cannot be simultaneously small. Selected examples of TFRs are:

- The short term Fourier transformation (STFT). As the name implies, it is a Fourier transform of the "recent past" of the signal. It contains  $w(t)$ , a window function that is to be chosen in a suitable way

$$\text{STFT } [\varphi](t, \omega) = \int_{-\infty}^{\infty} d\tau \varphi(\tau) \overline{w(\tau - t)} e^{-2\pi i \tau \omega}. \quad (3.7)$$

If  $w$  is a rectangular function, then the only free parameter left is its length or duration.

- The spectrogram, which is usually defined as being the absolute value of (3.7),

$$\text{SPEC} [\varphi] (t, \omega) = |\text{STFT} [\varphi] (t, \omega)| . \quad (3.8)$$

The power spectrogram denotes the square form of the spectrogram.

- Another quadratic representation is the Wigner distribution. It does not have free parameters like the STFT, and the TFR is

$$W [\varphi] (t, \omega) = \int_{-\infty}^{\infty} d\tau \varphi \left( t + \frac{\tau}{2} \right) \varphi \left( t - \frac{\tau}{2} \right) e^{-2\pi i \omega \tau} . \quad (3.9)$$

It is of importance as it shows a possible connection between STRFs and the Volterra series. A drawback of the Wigner distribution is, however, that it is not causal, and  $\varphi(t) = 0 \not\Rightarrow W [\varphi] (t, \omega) = 0$ .

- A fourth method motivated by auditory perception is the Gamma-tone Filterbank (Aertsen and Johannesma, 1980). It is usually characterized via its pulse response, and a possible time frequency representation (for a Gamma-tone Filter of the fourth order) could be

$$\text{GF} [\varphi] (t, \omega_c) = \int_{-\infty}^t d\tau h(\tau, \omega = \omega_c) \varphi(t - \tau) \quad (3.10)$$

with the kernel function (De Boer, 1975; Lopez-Poveda and Meddis, 2001)

$$h(t) = \begin{cases} kt^3 \exp(-2\pi Bt) \cos(2\pi \omega_c t + \theta) & (t \geq 0) \\ 0 & t < 0. \end{cases} \quad (3.11)$$

Here,  $B$  is the bandwidth,  $\omega_c$  is the center frequency,  $\theta$  is the phase and  $k$  the gain. Hohmann (2002) similarly gave a computationally efficient way of the gamma tone representation. Note that (3.10) is a special case of (3.7) if taking  $w(t) = kt^3 \exp(-2\pi Bt)$  and using a cosine instead of a Fourier transform.

From now on, TFRs of a signal will be denoted by  $s(t, \omega)$ , keeping in mind the different ways of creating them.

### 3.3.4 The STRF

For a two dimensional signal, the causal 1st-order Volterra operator would be (compare (3.1) with the the index on the integral kernel  $k$  omitted)

$$x(\sigma_1, \sigma_2) = \int_{-\infty}^{\sigma_1} \int_{-\infty}^{\sigma_2} d\tau_1 d\tau_2 k(\sigma_1 - \tau_1, \sigma_2 - \tau_2) \varphi(\tau_1, \tau_2) . \quad (3.12)$$

The STRF was derived from the 2nd-order non-linear operator, and thus has a different form (Aertsen et al., 1981)

$$x(t) = \iint d\tau d\omega \text{STRF}(\tau, \omega) s(t - \tau, \omega) \quad (3.13)$$

where  $\text{STRF}(t, \omega)$  denotes the STRF or integral kernel, the output  $x(t)$  is a spike rate and  $s(t, \omega)$  the time frequency representation of the input signal  $\varphi$ . Usually, the relation between the 2nd-order Wiener kernel and the STRF is illustrated by observing that (3.13) performs a convolution in the time domain and a multiplication in the frequency domain, which also corresponds to a convolution (see (3.6)). However, the reason for the relation with the 2nd-order Wiener kernel is the square representation of the signal in some TFRs.

## 3.4 Materials and Methods

### 3.4.1 Time-Frequency analysis

To create a squared TFR, a normed time signal was used, the factor 1.05 was added and a  $20 \cdot \log_{10}$  operation was performed to take into account the logarithmic processing of sound. Afterwards, the signal was sampled down and filtered with  $f$  different filters that had a memory of  $m$  (which resulted in a matrix  $\mathbf{A}$  of  $f$  rows or frequencies and  $m$  columns with the elements  $A_{fm}$ , see figure 3.1). To reduce the sampling rate,  $m$  positions of the input signal were filtered, and then, the next position was set to  $m - o$  where  $o < m$ . If  $o = m - 1$ , the sampling rate of the TFR was identical to that of the time signal. In the last step, the resulting matrix was squared element-wise and, if the absolute value TFR was used, the square root was taken. An example for a square TFR given a vector  $\mathbf{x} = x_1, x_2, \dots, x_7$  and  $f, m = 4$  and  $o = 1$  would be (with the filters grouped into a matrix for computational reasons)

$$\mathbf{S} = \begin{array}{c} \left( \sum_{m=1}^4 A_{1m} x_m \right) \left( \sum_{m=1}^4 \overline{A_{1m} x_m} \right) \\ \left( \sum_{m=1}^4 A_{2m} x_m \right) \left( \sum_{m=1}^4 \overline{A_{2m} x_m} \right) \\ \left( \sum_{m=1}^4 A_{3m} x_m \right) \left( \sum_{m=1}^4 \overline{A_{3m} x_m} \right) \\ \left( \sum_{m=1}^4 A_{4m} x_m \right) \left( \sum_{m=1}^4 \overline{A_{4m} x_m} \right) \end{array} \begin{array}{c} \left( \sum_{m=1}^4 A_{1m} x_{m+3} \right) \left( \sum_{m=1}^4 \overline{A_{1m} x_{m+3}} \right) \\ \left( \sum_{m=1}^4 A_{2m} x_{m+3} \right) \left( \sum_{m=1}^4 \overline{A_{2m} x_{m+3}} \right) \\ \left( \sum_{m=1}^4 A_{3m} x_{m+3} \right) \left( \sum_{m=1}^4 \overline{A_{3m} x_{m+3}} \right) \\ \left( \sum_{m=1}^4 A_{4m} x_{m+3} \right) \left( \sum_{m=1}^4 \overline{A_{4m} x_{m+3}} \right) \end{array} \cdot \quad (3.14)$$

For the absolute value TFR the a square root operation would be performed on every element of  $\mathbf{S}$  in (3.14).

### 3.4.2 The STRF operation and the corresponding filter using square TFR forms

The STRF was numerically implemented as shown in Machens et al. (2004). Let  $r_t$  be any result at the time  $t$  and let  $R_{f,t}$  be the STRF and  $S_{f,t}$  any time frequency representation at frequency  $f$  and time  $t$ , and let  $M$  be the memory length and  $F$  the number of frequencies of the STRF, then

$$r_t (\mathbf{A}^2) = r_0 + \sum_{f=1}^F \sum_{t_R=0}^{M-1} R_{f,M-t_R} S_{f,t-t_R} \cdot \quad (3.15)$$

Using what is known about the square-TFR from generalizing (3.14) and setting  $r_0 = 0$ , this results in

$$\sum_{f=1}^F \sum_{t_R=0}^{M-1} R_{f,M-t_R} \left( \sum_{n=1}^m A_{f,n} x_{t-n+1-t_R(m-o)} \right) \left( \sum_{n=1}^m \overline{A_{f,n} x_{t-n+1-t_R(m-o)}} \right) \cdot \quad (3.16)$$

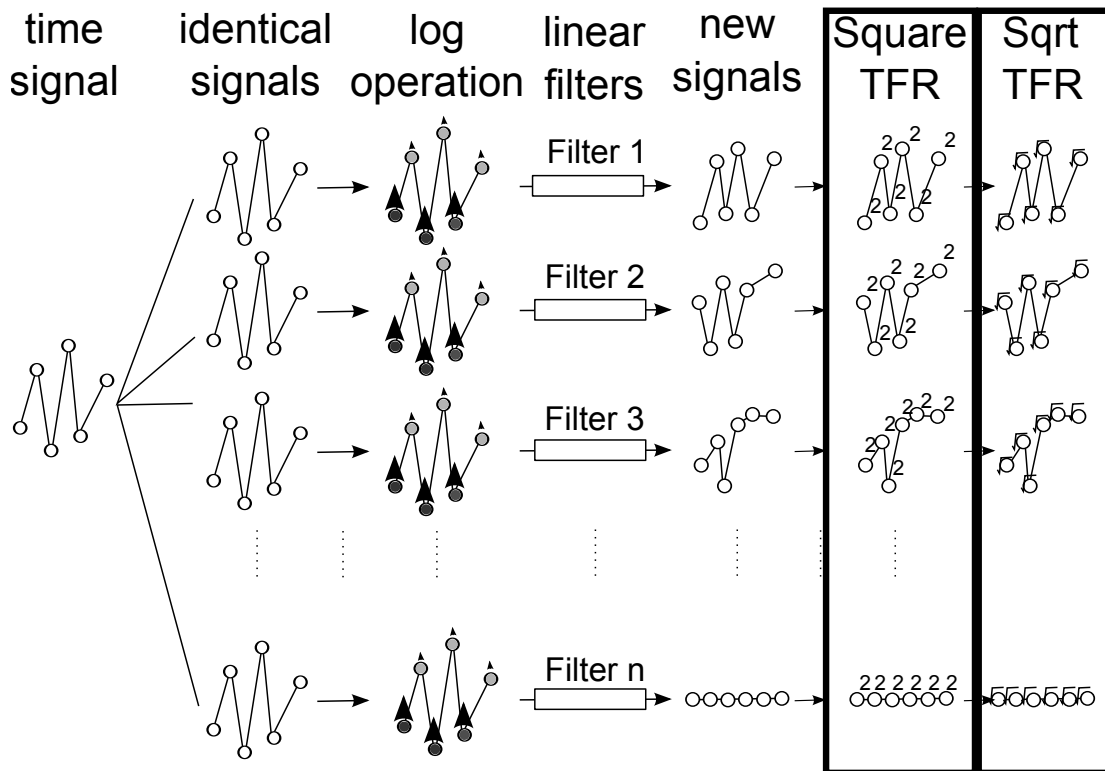


Figure 3.1: Schematic drawing of the TFR used in this chapter. The two frames on the right denote the two forms that were used. The absolute value TFR was created by taking the square root of the squared representation.

Creating an equivalent 2nd-order Volterra kernel means finding  $n$  and  $t_R$  for which

$$\tau_1 = t - n + 1 - t_R(m - o) \wedge \tau_2 = t - n + 1 - t_R(m - o) \quad (3.17)$$

holds, while  $t$  is defined to be  $m + z(m - o)$  for any integer  $z$  (which are the instants in time where the TFR is defined). There may be multiple combinations of  $t_R$  and  $n$  for which equation (3.17) holds, especially if  $\tau_1 = \tau_2$ . If this is true, then these combinations have to be summed up, and for any given  $\tau_1, \tau_2$ , the resulting kernel is

$$v_{\tau_1, \tau_2}(\mathbf{A}^2) = \sum_{t_R, n} \sum_{f=1}^F R_{f, M-t_R(n)} A_{f, n(t_R, \tau_1)} \overline{A_{f, n(t_R, \tau_2)}} \quad (3.18)$$

where  $k, n, k(n)$  and  $n(k)$  denote the pairs of values for which (3.17) holds. The first sum accounts for possible multiples of these combinations.

### 3.4.3 Absolute value TFRs

As the absolute values TFR could be calculated by taking the square root of the individual elements  $s_{ab}$  of  $\mathbf{S}$  in (3.14) and Volterra operators are defined as whole-number powers, the absolute valued TFR cannot be exactly reproduced by a Volterra operator. The approximation was done by calculating the  $n$ -dimensional Taylor series expansion for all elements of the past that contributed to the result. The result for the actual STRF operation turns from (3.15) (again, setting  $r_0 = 0$ ) to

$$r_t(|\mathbf{A}|) = \sum_{f=1}^F \sum_{t_r=0}^{M-1} R_{f, M-t_r} S_{f, t-t_r} = \sum_{f=1}^F \sum_{t_r=0}^{M-1} R_{f, M-t_r} \sqrt{\left( \sum_{n=1}^m A_{f, n} x_{t-n+1-t_r(m-o)} \right) \left( \sum_{n=1}^m \overline{A_{f, n} x_{t-n+1-t_r(m-o)}} \right)}. \quad (3.19)$$

The result of the corresponding Taylor series now depends on the derivatives of the  $x$  terms, with the zero order approximation being stationary (for now, even though dealing with numerical systems, the indices are to be written in brackets instead of subscripts again)

$$v_0(|\mathbf{A}|) = r_t(|\mathbf{A}|, (\mathbf{x} = \langle \mathbf{x} \rangle)) \quad (3.20)$$

with  $\langle \mathbf{x} \rangle$ , some characteristic vector of the time-input.

To obtain the 1st-order Volterra operator, it is approximated at every instant in the past that is in the combined memory of the STRF and the filter used at the corresponding frequency. This would imply that the approximated 1st-order Volterra operator  $v_1$  at the time point  $t_1$  in the past is



$$v_{\tau_1}(|\mathbf{A}|) = \left. \frac{\partial}{\partial x_{\tau_1}} r_t(|\mathbf{A}|) \right|_{(\mathbf{x}=\langle \mathbf{x} \rangle)} = \quad (3.21)$$

$$\sum_{t_R, n} \sum_{f=1}^F R_{f, M-t_R(n)} \frac{\overline{A_{f,n}(t_R, \tau_1)} (\sum_n A_{f,n} x_n) + A_{f,n}(t_R, \tau_1) (\sum_n \overline{A_{f,n} x_n})}{2\sqrt{(\sum_n A_{f,n} x_n) (\sum_n \overline{A_{f,n} x_n})}}.$$

Again, for the different  $x$  values  $\mathbf{x} = \langle \mathbf{x} \rangle$  holds. As, due to the overlap, some factor of  $x$  can occur at different frequencies, again a sum has to be taken.

The approximated 2nd-order Volterra operator is derived from the second derivative of the Taylor approximation of the result. Let

$$g = \sqrt{\left( \sum_n A_{f,n} x_n \right) \left( \sum_n \overline{A_{f,n} x_n} \right)} \quad (3.22)$$

then with  $r_t(|\mathbf{A}|)$  from (3.19)

$$v_{\tau_1, \tau_2}(|\mathbf{A}|) = \left. \frac{\partial^2}{\partial x_{\tau_1} \partial x_{\tau_2}} r_t(|\mathbf{A}|) \right|_{(\mathbf{x}=\langle \mathbf{x} \rangle)} = \quad (3.23)$$

$$\sum_{t_R, n} \sum_{f=1}^F R_{f, M-t_R(n)} \left( \frac{A_{f,n}(t_R, \tau_1) \overline{A_{f,n}(t_R, \tau_2)} + \overline{A_{f,n}(t_R, \tau_1)} A_{f,n}(t_R, \tau_2)}{2g} - \frac{1}{3g^3} \times \right.$$

$$\left( A_{f,n}(t_R, \tau_1) \sum_n \overline{A_{f,n} x_n} + \overline{A_{f,n}(t_R, \tau_1)} \sum_n A_{f,n} x_n \right) \times$$

$$\left. \left( A_{f,n}(t_R, \tau_2) \sum_n \overline{A_{f,n} x_n} + \overline{A_{f,n}(t_R, \tau_2)} \sum_n A_{f,n} x_n \right) \right).$$

### *The characteristic input vector*

The input signal has an influence on the resulting Volterra kernel since the Taylor series of the square-root function involves division by the original filter, which would be a product of the rows of the matrix. The sum of these, by definition, is zero, to prevent adding energy to the result when performing the filter operation.

A set of  $n$  vector was used, with  $n$  being the length of the frequency filters. These vectors consisted of random distributions of an equal amount of ones and zeros, and the corresponding kernels were determined  $n$  times. Eventually, the average of the kernels was taken as the resulting kernel.

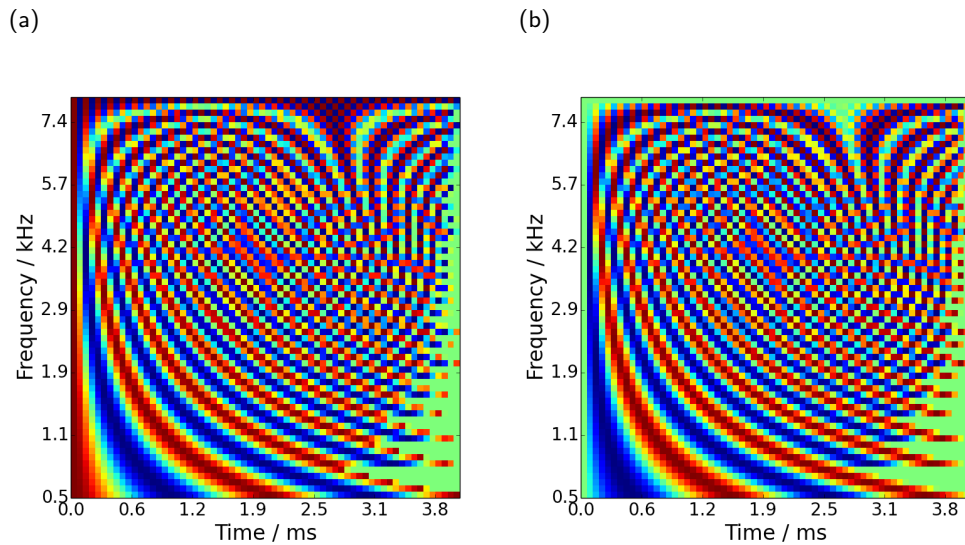


Figure 3.2: (a) Real part of the transformation matrix, (b) imaginary part. The minimum frequency was 500 Hz (the temporal resolution was 16 kHz), lower frequencies were not possible. The maximum frequency was the Nyquist frequency. As this is not a classical fast Fourier transformation matrix (which would start at 0 Hz), some of the row entries had to be truncated to keep the mean of zero.

#### 3.4.4 Parameters of the simulation

This study used a filter with a starting frequency of 0.5 kHz and a maximal frequency of 8 kHz. The frequencies were spaced logarithmically in order to have a higher frequency resolution in the low frequency part. The temporal resolution of the signal was 16 kHz, the overlap was 32 and the number of frequencies and memory length were both 64. The resulting sampling frequency of the TFR is 2 kHz. The used matrix (real and imaginary part) for  $\mathbf{A}$  is shown in figure 3.2.

## 3.5 Results

### 3.5.1 Comparing the time-frequency representations

Figure 3.3 shows two time signals that consisted of a series of pure tones (0.5, 1, 2, 4, and 8 kHz), one rising frequency modulated (FM) sweep, one falling frequency modulated sweep, two rising and falling sweeps, a raising and a falling frequency modulated sweep at the same time, and noise. Every signal fraction lasted 83 ms. Panel (a) shows the time signal in the range of  $[0, 2]$  and the corresponding TFRs, the squared one and the absolute value one. Panel (b) shows the same signal, but after the  $20 \cdot \log_{10}$  operation was performed on the original time signal. Again, the corresponding spectrograms are shown below. Regardless of the log operation, the signal to noise ratio is lower in the squared spectrogram. The absolute value spectrogram has artifacts at higher and lower frequencies when tested with pure tones, but, especially without the log operation, the frequency modulated sweeps are more prominent. Additionally, regardless of the log operation, the superposition of frequency modulated sweeps and the noise are more easily detected in the absolute value spectrograms. Artifacts at the first harmonics of the pure tones can be seen on the squared TFR when the log operation is performed on the time signal. These artifacts are more prominent and can be seen at different harmonics for the absolute value TFR. For the absolute value TFR, the first harmonic can be seen for the single and double sweeps as well, while in the square TFR neither harmonics nor the signal are easily detected. The noise is again better detectable for the absolute value TFR.

### 3.5.2 An STRF based on a squared time-frequency representation can be exactly reproduced by a 2nd-order Volterra operator

Figure 3.4 shows an artificial STRF with realistic parameters on the top left and the corresponding 2nd-order Volterra kernel on the top right. On the bottom the result to the stimuli shown in figure 3.3 is shown. The red line shows the result from the STRF on the square spectrogram domain and the black dotted line the result from the 2nd-order Volterra kernel on the time domain. The results are equal, as the STRF performs a linear operation on the TFR. The components of the input signal  $x$ , on the other hand, contributed multiples (squares) to the TFR, thus the result of the STRF depends on double multiples of the input signal. A 2nd-order Volterra operator performs the same action, and therefore, the results are equal. In the beginning of the signal, the 1 kHz pure tone hits the excitation and there is a weak rise in the output. Minor on- and offset effects can be seen. Afterwards, the 2 kHz pure tone hits the inhibitory field which results in a suppression. The higher frequencies do not elicit any output. The sweeps cause some major deflections, while the noise causes some minor ones in the output, as was to be expected by shape of the squared TFR.

### 3.5.3 The dictionary of 2nd-order Volterra kernels

Figure 3.5 shows toy STRFs and their corresponding 2nd-order Volterra kernels. Panels (a),(c),(e) show STRFs where only one entry at the beginning was set to one. The corre-

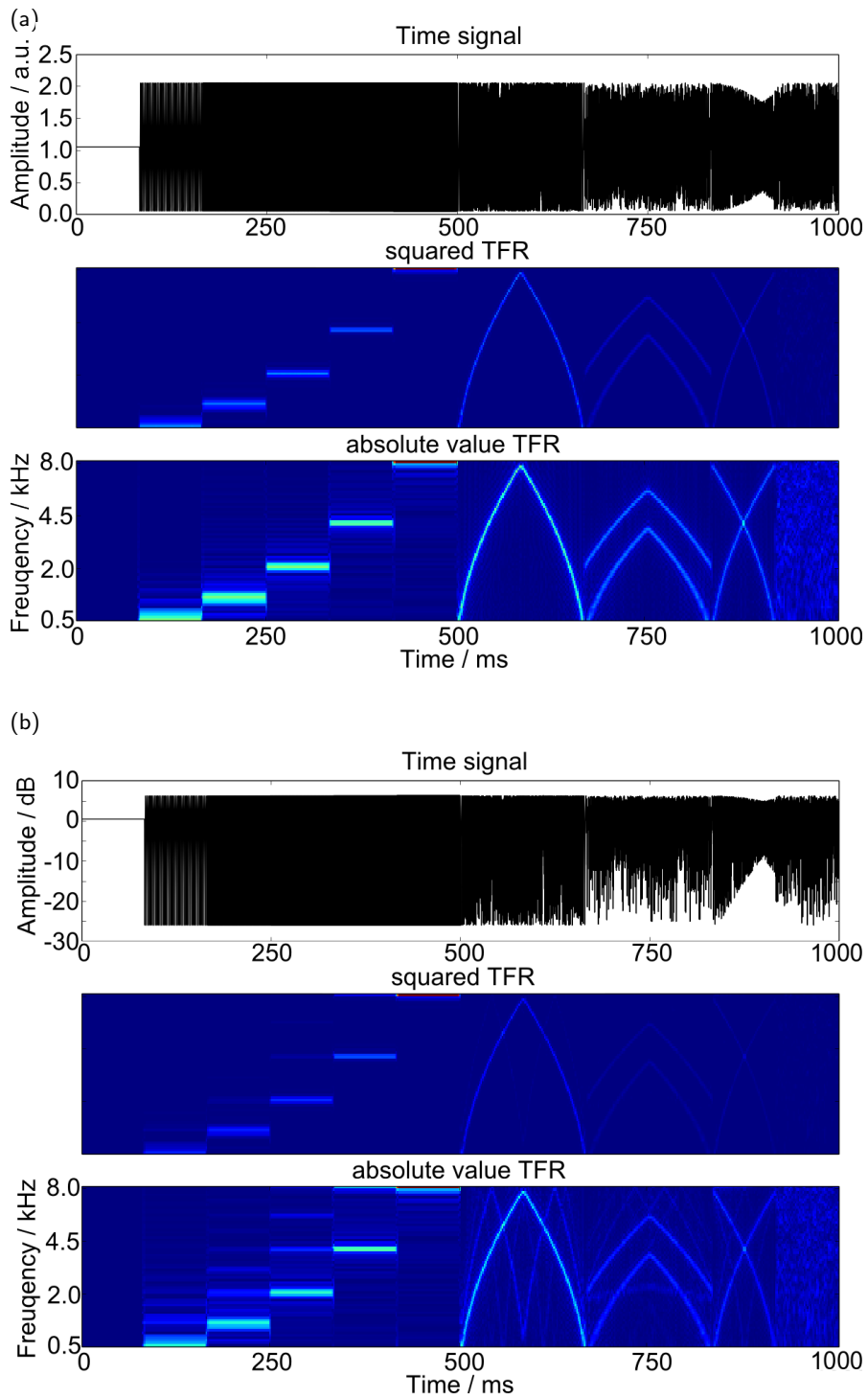


Figure 3.3: (a) A time signal, consisting of pure tones, frequency modulated sweeps, and noise, in two different TFRs. (b) The same signal, this time after the  $20 \cdot \log_{10}$  operation was performed on it.

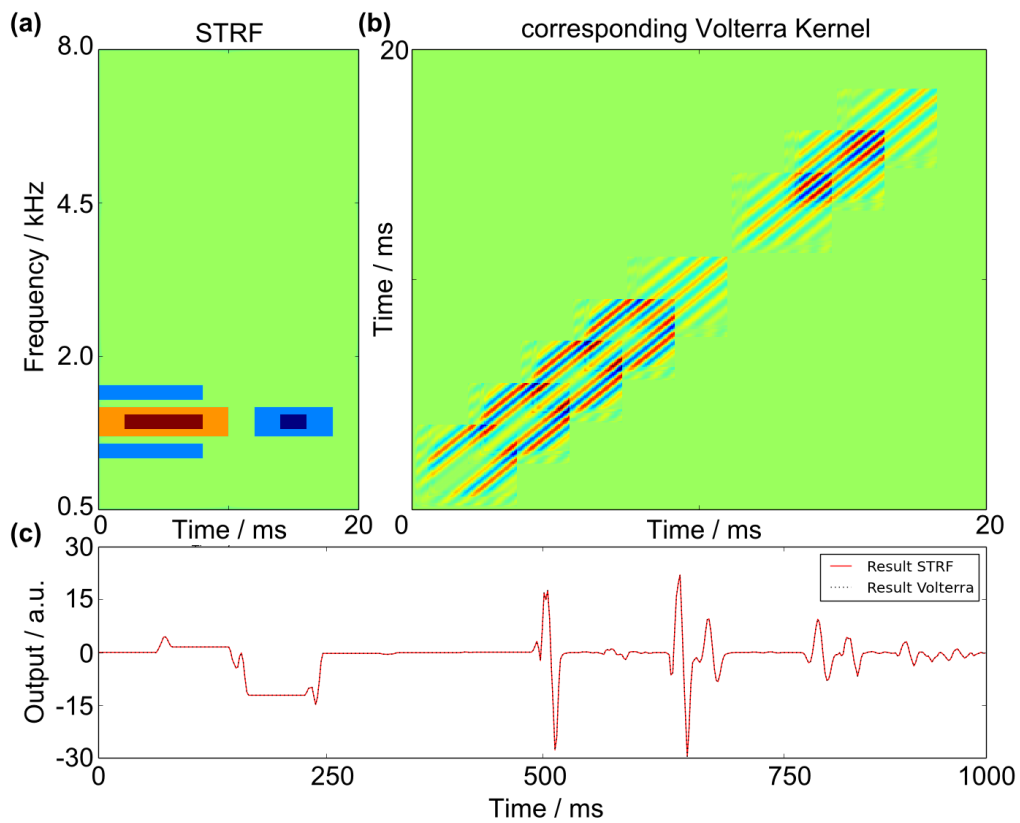


Figure 3.4: (a) A toy STRF with a realistic shape. (b) The equivalent 2nd-order Volterra kernel. (c) Red line: the result for the stimuli shown in figure 3.3 for the STRF operation on the squared TRF. Black dotted line: The results for the same stimuli for the 2nd-order Volterra operator performed on the time input.

sponding kernels start immediately, too, and show the structure of the filter collection at the according frequencies (compare figure 3.2). In panel (b), the influence of the time shift can be seen, which causes a mere delay in the kernel. Panel (c) shows the superposition of a prolonged excitation. The kernel is a sum of the individual elements, and thus, has higher amplitudes. What happens in the case of a large excitatory field is illustrated in panel (d). The elements of the Volterra kernel that have nearly zero time difference (those along the main diagonal) are summed up, as all filters have a peak around that time. The elements further away from the main diagonal are either smaller, or tend to zero, as they would for a sum of cosine functions with different frequency but constant phase. The influence of inhibitory fields is not shown, but corresponds to a switch in prefix: At the same position as an excitatory field, they would have the same structure in the kernel with an inverted sign.

#### 3.5.4 The influence of the memory of the TFR filters on the kernel

The corresponding kernels shown in figures 3.4 and 3.5 show that the duration of a kernel along its main diagonal is, as expected, equal to the duration of the STRF. Further, these figures show that only a limited amount in interactions of two points separated by  $\Delta\tau = |\tau_1 - \tau_2|$  is possible (see figure 3.6b). This maximal distance can be quantified; it is the maximum duration of the memory of a given frequency filter or the length of the filter matrix,  $m$  (compare equation (3.14)) times the temporal resolution of the original signal. In this study, this would correspond to  $64 \cdot 0.0625 \text{ ms} = 4 \text{ ms}$ . If taking a fourth-order Gamma-tone Filter with a center frequency of 1 kHz, cutoff frequencies of 0.5 and 8 kHz, respectively, and 3 filters per erb the maximum time constant for one filter is 20-25 ms, when the filter is at less than 1% of its highest peak (the form of the kernel or filter is shown in figure 3.6a).

#### 3.5.5 STRF based on the absolute value time-frequency representation, 1st- and 2nd-order approximations

Figure 3.7 shows the STRF (panel a) on an absolute value spectrogram as well as its corresponding first (panel b) and second (panel c) order Volterra kernels. The first kernel was calculated according to (3.21). The results from using the STRF on the input shown in figure 3.3 bottom are shown in figure 3.8 as a red line. The dotted lines indicate different approximations. The blue line denotes the result from the 1st-order Volterra kernel; the green line denotes that of the second Volterra kernel. The black is the sum of the two plus the offset term from (3.20), thus it corresponds to the exact 2nd-order Taylor approximation. The purple line is the result of the 2nd-order Volterra kernel as shown in figure 3.4 or, alternatively, the result from the STRF on the squared spectrogram. This is essentially a differently scaled version of the approximated one. While the influence of the linear kernel is small and sometimes of opposite sign as the actual result (see the peak down at 200 ms), a 2nd-order Volterra kernel is in close agreement with the result of an STRF on an absolute value TFR. The higher scaled kernel overestimates the response at pure tones and single sweeps, but performs better in noise and multiple sweeps; the lower scaled version does just the opposite.

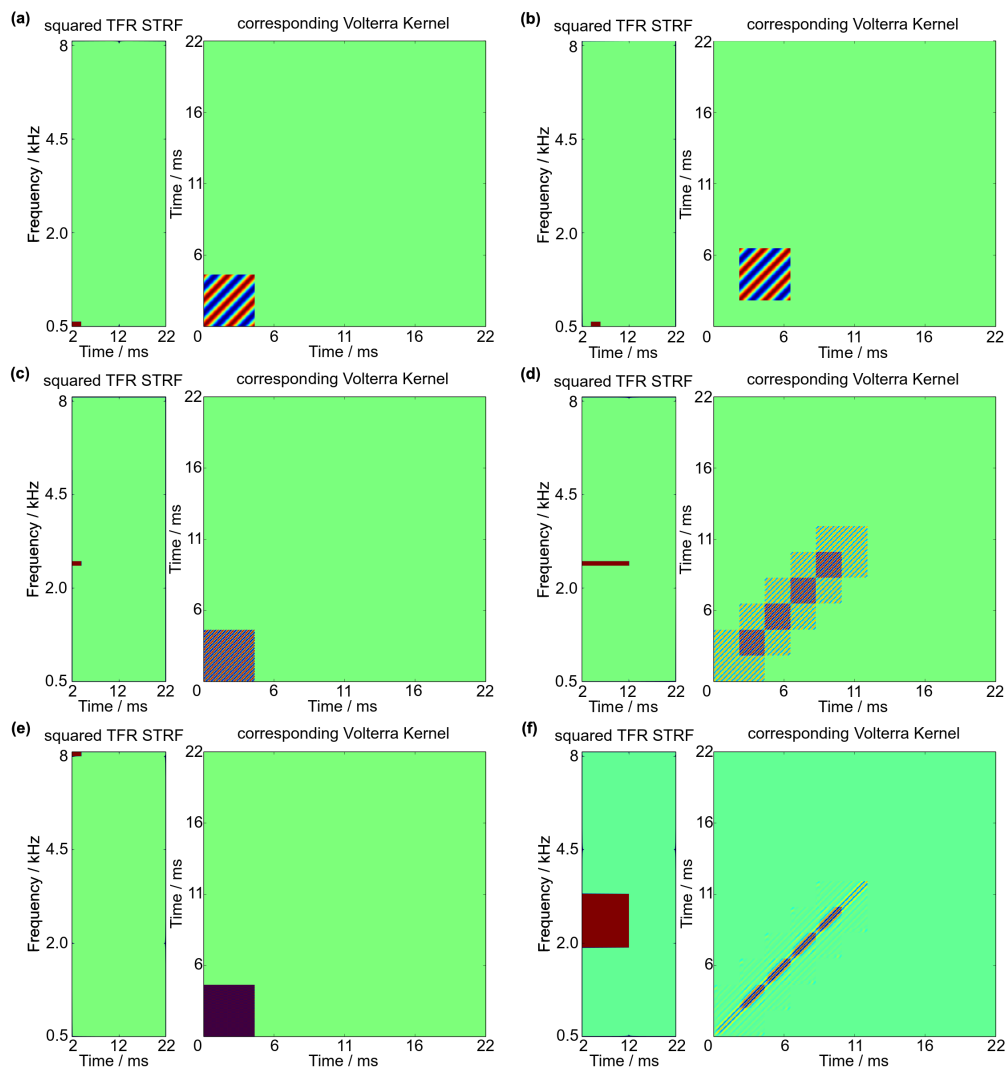


Figure 3.5: The resulting Volterra kernels are linear combinations with the STRF determining the weighting factors of the different filters. (a),(c),(e) STRFs that would be sensitive at one frequency and time frame. The corresponding kernels show the structure of the individual filters. (b) A shift in time of the STRF corresponds to an equal shift of the Volterra kernel. (d) If the field extends in time, so does the kernel. (f) A large excitatory field leads to a more complicated kernel, but this kernel remains a superposition of all individual elements in the STRF.

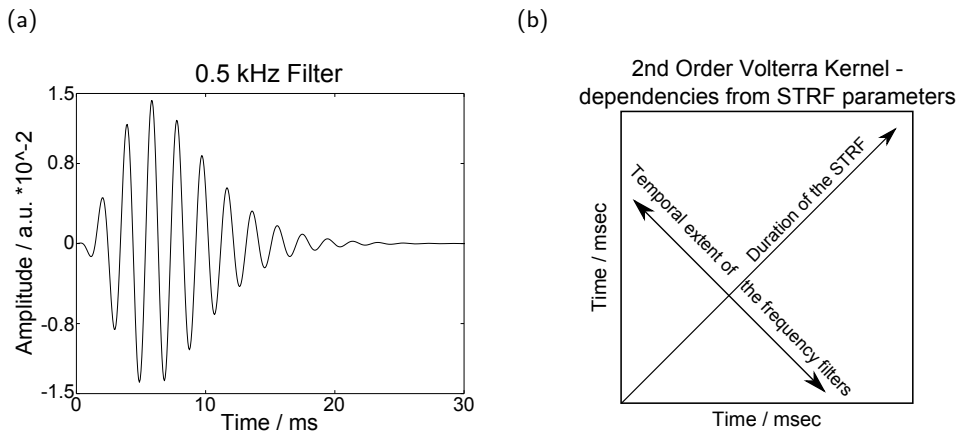


Figure 3.6: (a) Using the algorithm by Hohmann (2002), a Gamma tone filter was created for 0.5 kHz. (b) The scheme showing the dependencies of a 2nd-order Volterra operator on the parameters used to create an STRF.

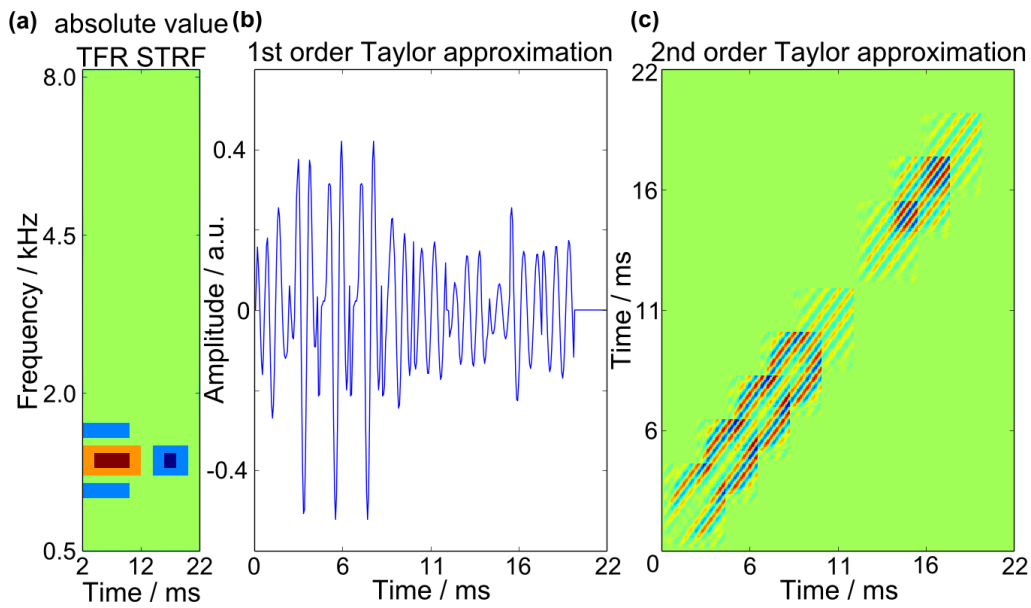


Figure 3.7: The same STRF as shown in figure 3.4 panel a, but now applied to an absolute value TFR (bottom of both panels in figure 3.3). Unlike the square TFR STRF, no exact Volterra kernel that could reproduce the output exists. (a) shows the 1st-order approximation based on equation (3.21), (b) the 2nd-order approximation according to equation (3.23).



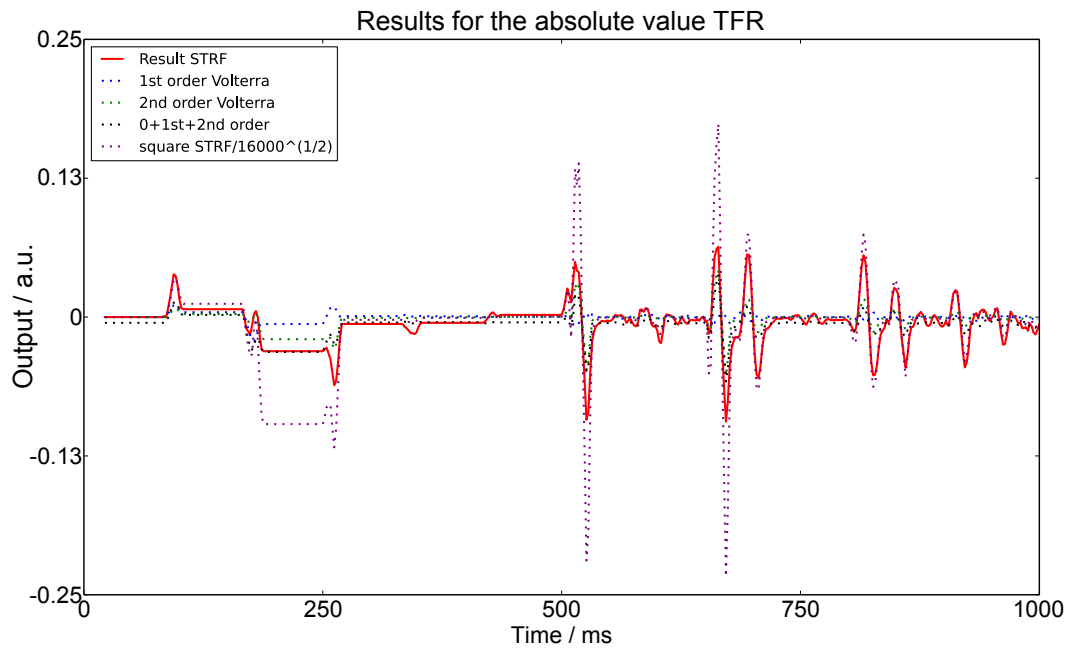


Figure 3.8: The response of the absolute value spectrogram (red line) of the STRF shown in figure 3.4 using the stimuli of the absolute value TFR, shown in figure 3.3b at the bottom. The dotted blue line is the result from the 1st-order Volterra kernel, the dotted green line the result from the 2nd-order Volterra kernel. The black dotted line is the 2nd-order Taylor approximation, the sum of the mean response of the STRF and the two kernels. The dotted purple line is the result from the Volterra kernel shown in figure 3.4, scaled down by the square root of the sampling frequency.

### 3.5.6 The approximated 2nd-order kernel is proportional to the exact kernel of the square form

The 2nd-order Volterra kernel shown in figure 3.7 is similar to that derived from the square TFR shown in figure 3.4. This can also be demonstrated by showing that equation (3.18) is proportional to (3.23). As a first observation, for (3.18)

$$v_{\tau_1, \tau_2}(\mathbf{A}^2) = \overline{v_{\tau_2, \tau_1}(\mathbf{A}^2)}. \quad (3.24)$$

The imaginary parts result from the filtering operation of the transformation into the time-frequency domain. Any results from a convolution with a 2nd-order Volterra operator for which (3.24) holds have an imaginary part of zero. Now, it has to be shown that

$$\text{Re}[v_{\tau_1, \tau_2}(\mathbf{A}^2)] \propto v_{\tau_1, \tau_2}(|\mathbf{A}|). \quad (3.25)$$

In order to show this without loss of generality, let  $F=1$  and  $\rho_i = \sum_n A_{f,n} x_n$  for a given vector  $\mathbf{x}_i$ . The receptive field  $R_{f, M-t_r(n)}$ , as well as the summation over multiple appearances of indices and the indices for  $A$  themselves are equal for both equations, and thus, do not have to be considered. The first term of (3.23),

$$\sum_{\rho_i} \frac{A_{f,n(t_R, \tau_1)} \overline{A_{f,n(t_R, \tau_2)}} + \overline{A_{f,n(t_R, \tau_1)}} A_{f,n(t_R, \tau_2)}}{2\sqrt{\rho_i \overline{\rho_i}}} \quad (3.26)$$

already satisfies (3.25) as  $\sum_{\rho_i} \frac{1}{\rho_i}$  is any number and  $A + \overline{A} = 2\text{Re}[A]$ . Thus, it has to be shown that

$$\frac{(A_{f,n(t_R, \tau_1)} \overline{\rho_i} + \overline{A_{f,n(t_R, \tau_1)}} \rho_i) \times (A_{f,n(t_R, \tau_2)} \overline{\rho_i} + \overline{A_{f,n(t_R, \tau_2)}} \rho_i)}{3(\sqrt{\rho_i \overline{\rho_i}})^3} \quad (3.27)$$

is proportional to (3.18). Still, factors of the form  $\rho_i^y$  with  $y \in \mathbb{Q}$  are ignored as they obey the proportionality. Thus, only two terms remain,

$$A_{f,n(t_R, \tau_1)} A_{f,n(t_R, \tau_2)} \overline{\rho_i}^2 \quad \text{and} \quad \overline{A_{f,n(t_R, \tau_1)}} \overline{A_{f,n(t_R, \tau_2)}} \rho_i^2. \quad (3.28)$$

It has to be shown that the sum over all used  $\mathbf{x}_i$  that are part of the  $\rho_i$  causes them to be averaged out. This is easy to see for the standard base,  $\mathbf{x}_1 = 1, 0, 0, \dots$ ,  $\mathbf{x}_2 = 0, 1, 0, \dots$ ,  $\dots$ ,  $\mathbf{x}_n = \dots, 1$ , however, it is not true for every base. Hence, the base has to be chosen in a suitable manner. Thus, the criteria for the base are that  $\rho_i = \sum_n A_{f,n} x_n \neq 0$  for any  $f$  and that the terms in (3.28) are zero when summed over all base entries.

The usage of a base instead of one vector in equations (3.21) and (3.23) ensures that the Taylor approximation of the Volterra operators holds for every input.

## 3.6 Discussion

In this study, it was shown how a system such as in figure 3.9 can be approximated by Volterra kernels in the time domain, similar to an idea by Palm (1979). The log operation, that may be part of the non-linearity, was performed before creating the TFR, and the filters to do so were less sophisticated than state of the art ones (for a complete discussion of the influence of the time-frequency representation see Gill et al. (2006)). Despite this, the general approach remains valid, and it does not rely on the Wigner or the Rihaczek class like other existing transformations and, unlike the derivation shown in Klein et al. (2000), it does not violate causality. Note that the non-linearity of the STRF results from the static non-linearity of the pre-processing and not from the integration of all frequencies. If the non-linearity is of the form of an integer power, the corresponding Volterra kernels turn out to be able to give the same result as the STRF.

### 3.6.1 Comparison to the mapping of Wiener kernels to the STRF

The difference of this study to Lewis and van Dijk (2004) is that the filters for the TFR are not necessarily part of a short-term Fourier transform, but the assumption of linear-nonlinear-linear model is similar. In the approach by Lewis and van Dijk (2004), the signals were filtered, squared and added, whereas in this study they were filtered again to simulate the STRF operation. As Lewis and van Dijk (2004) did not calculate Wiener kernels from STRFs but vice versa, there was no need for the following filter. The third important difference is the usage of Wiener instead of Volterra kernels. This might require some words about the general difference between Wiener and Volterra kernels. If a hypothetical system can be completely described by any finite order of a Volterra series, then this system can be described just as well as by a Wiener series. Equation (3.29) shows the means of converting one representation into the other. The benefit of the Volterra series is that it can be linked to a differential equation of the form (3.2) and occurs as a solution operator. Even if setting  $\varphi(t) = 0$ , this kind of equation does not necessarily have only one stable fixed point. If setting  $\varphi(t)$  equal to white noise, as would be required for a Wiener analysis, then this knowledge cannot be used any more. In terms of the example above, the Wiener and the Volterra series are only equal for a system described by equation (3.2), if and only if, it has one sole stable fixed point around the zero position. In this manner, the operators found for the STRF could be either Volterra or Wiener operators.

### 3.6.2 Specific results for absolute value and squared TFRs

Specifically, it was shown that square and absolute TFRs have, if no other non-linearities are involved, corresponding 2nd-order Volterra operators that have a limited range of non-linearity around the main diagonal. This means that the time signal can only interact with itself on a scale of several ms. Instead of investigating the effects of the operator, one could also investigate the differences in the two representations. Using any sort of square form is necessary as even a simple on and offset sensitive unit could not be modeled by a linear system. A linear system would always lead to some activity after the onset of a stimulus and

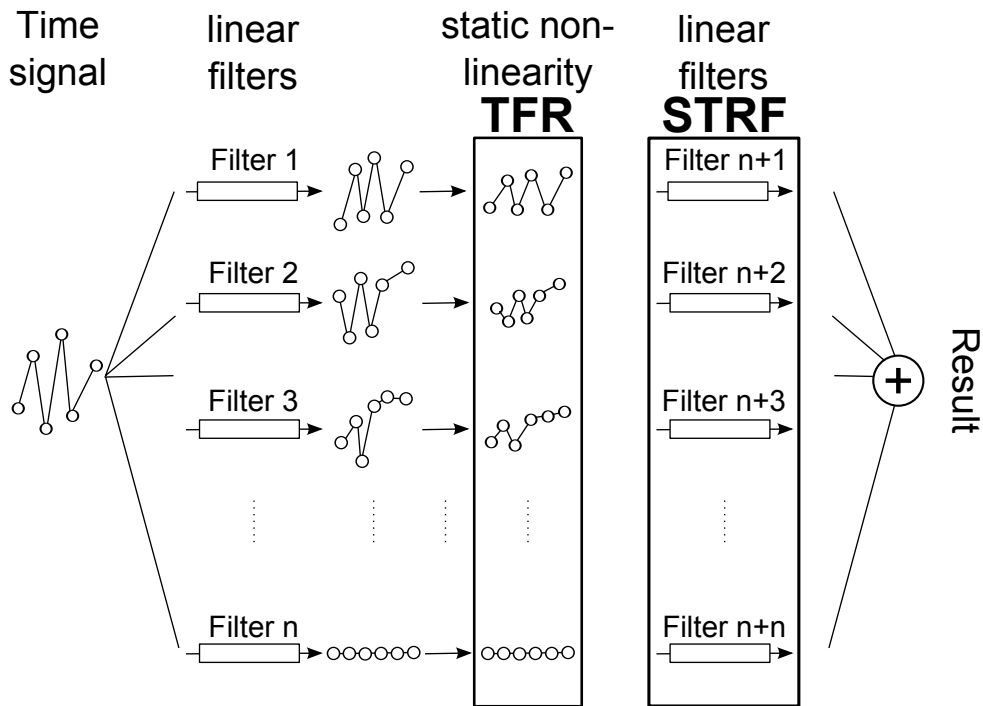


Figure 3.9: Schematic drawing of the assumptions of the proposed algorithm. The nonlinearities are entirely due to the TFR operation, while the influence of the STRF is linear.

after the offset, and an increase in activity at the beginning and a decrease at the end are not possible.

### 3.6.3 Benefits of converting an STRF into a Volterra series

Though sometimes the Volterra series is thought of as a parametric model (Wu et al., 2006) it still provides a better and deeper understanding of the parameters used than the STRF, for example. Using an STRF means choosing from a dictionary of possible 2nd-order Volterra kernels, determined by the form of the time-frequency representation. Though the trend is to increase the parameter space of the STRF (Lesica and Grothe, 2008; Ahrens et al., 2008; Rabinowitz et al., 2013), this algorithm could help to disentangle some of the methods. Further, it increases data comparability as, for example, the results by Pienkowski et al. (2009), using (3.29), could be directly compared to the recorded STRFs. Additionally, if the generalized linear-nonlinear model (Carruthers et al., 2013) was turned into a Volterra series by similar methods, even more complicated predictors such as the aforementioned one and the STRF could be directly compared.

### 3.7 Appendix

For a system with only one stable fixed point at the zero position, the Volterra operators could be recovered from the Wiener operators (and similarly vice versa) via

$$k_n(\tau_1, \dots, \tau_n) = \sum_{\nu=0}^{\infty} (-1)^\nu \frac{(n+2\nu)!}{n!\nu!} \left(\frac{P}{2}\right)^\nu \times \int d\sigma_1 \dots \int d\sigma_\nu g_{n+2\nu}(\tau_1, \dots, \tau_n, \sigma_1, \sigma_1, \dots, \sigma_\nu, \sigma_\nu) . \quad (3.29)$$

Note that the two operators of highest order are equal.

## Chapter 4

# Using double clicks to test spectro-temporal receptive fields and to estimate Volterra operators

### 4.1 Outline and key findings

- Differential Sampling, a method to calculate Volterra kernels directly, is introduced and explained
- Using Differential Sampling the existence of 2nd or higher order operators at time scales higher than that of the STRF is shown
- The difference between the two results is explained within the framework of ordinary differential equations
- Differential Sampling is used to investigate two units in the cochlear nucleus

## 4.2 Abstract

One goal in system identification is the correct prediction of the response of a system to a given stimulus. In auditory neuroscience, this led to the Spectro-temporal receptive field (STRF) that characterizes single units. The implicit assumptions made when using an STRF have been described in the previous chapter. These assumptions might be too restrictive, as many times the predictions of the STRF to unknown stimuli fail. A more general method to describe a single unit is the Volterra series, which, unlike the STRF is parameter-free. The integral kernels of the series can be calculated with a method called Differential Sampling that uses clicks at different sound levels. For 11 units, higher order operators were calculated for sample times at 6, 12, 25, 50, 100 and 200 ms. To investigate the STRFs more closely, they were transformed to an equivalent 2nd order Volterra operator. At all times significant responses in the kernel could be found for at least one unit. Significance was investigated by testing for ergodicity with respect to a self-leveling reference measurement. For two units in the cochlear nucleus, the 2nd order Volterra operator was calculated using Differential Sampling.

## 4.3 Introduction

The Volterra series was introduced in the previous chapter as a functional that describes the way an unknown system turns a known input into a measurable output. This can be imagined by using coefficients in a Taylor series to map one variable onto another, but using functions instead of variables. In this manner, the coefficients become integral kernels, with the disadvantage that higher kernels increase their number of free parameters by the power of themselves. This means, e.g. that while the  $x^2$  coefficient of a Taylor polynomial is of the same dimensionality as the one used for  $x$ , the 2nd-order integral kernel  $v_2(t_1, t_2)$  is a function on two dimensions while the first one  $v_1(t_1)$  is one-dimensional.

A new way of finding Volterra kernels was introduced by van Hemmen et al. (2000), where it was shown that they can be derived from a differential equation and that they can be directly accessed using Differential Sampling. This means that delta functions (or clicks) are taken as stimuli.

### 4.3.1 Example calculations to visualize Differential Sampling

The Volterra series basically states that any function  $x(t)$  that depends on another function  $\varphi(t)$  can be approximated by using integral kernel functions that are of the form  $k_1(\tau_1)$ ,  $k_2(\tau_1, \tau_2), \dots$ . Taking these functions in discrete steps  $\Delta\tau$  and looking at their numeric forms  $(x_t, \varphi_t, k_{n, \tau_1 \dots \tau_n})$ , this means (keeping the notation from chapter 3.3.2)

$$x_t = \sum_{m=-M}^t k_{1, t-m} \varphi_m \Delta\tau + \sum_{m_1=-M}^t \sum_{m_2=-M}^t k_{2, t-m_1 t-m_2} \varphi_{m_1} \varphi_{m_2} (\Delta\tau)^2 + \dots \quad (4.1)$$

where  $M$  denotes the memory length of the system. Staying within the numerical framework, the idea of Differential Sampling is to extract the functions  $k_{1, \tau_1}, k_{2, \tau_1 \tau_2}, \dots$ . This is done by choosing the  $\varphi_t$  function as clicks.

To illustrate this further, an actual calculation is done on formula (4.1). Let  $M = 3$ , and let the order of the system be 3, too. Hence,

$$\begin{aligned} x_3 = & \quad (4.2) \\ & k_{1, 3} \varphi_1 + k_{1, 2} \varphi_2 + k_{1, 1} \varphi_3 + \\ & k_{2, 33} \varphi_1^2 + k_{2, 22} \varphi_2^2 + k_{2, 11} \varphi_3^2 + 2k_{2, 31} \varphi_1 \varphi_3 + 2k_{2, 32} \varphi_1 \varphi_2 + 2k_{2, 21} \varphi_2 \varphi_3 + \\ & k_{3, 333} \varphi_1^3 + k_{3, 222} \varphi_2^3 + k_{3, 111} \varphi_3^3 + \\ & 3(k_{3, 332} \varphi_1^2 \varphi_2 + k_{3, 331} \varphi_1^2 \varphi_3 + k_{3, 322} \varphi_1 \varphi_2^2 + k_{3, 311} \varphi_1 \varphi_3^2 + \\ & k_{3, 221} \varphi_2^2 \varphi_3 + k_{3, 211} \varphi_2 \varphi_3^2) + 6k_{3, 321} \varphi_1 \varphi_2 \varphi_3. \end{aligned}$$

Multiple entries are due to the symmetry properties of the kernel.



*Investigating the first and second order kernel with Differential Sampling*

In order to determine  $k_1$  in equation (4.2), let

$$\gamma_+ = \begin{cases} \epsilon & \text{if } t = 1 \\ 0 & \text{else} \end{cases} \quad \text{and} \quad \gamma_- = \begin{cases} -\epsilon & \text{if } t = 1 \\ 0 & \text{else} \end{cases} . \quad (4.3)$$

If now the two  $\gamma$ -functions are taken as the input  $\varphi$  in equation (4.2), the results would be

$$x_3 [\gamma_+] = \epsilon k_{1,3} + \epsilon^2 k_{2,33} + \epsilon^3 k_{3,333} \quad (4.4)$$

and

$$x_3 [\gamma_-] = -\epsilon k_{1,3} + \epsilon^2 k_{2,33} - \epsilon^3 k_{3,333} . \quad (4.5)$$

Now,

$$x_3 [\gamma_+] - x_3 [\gamma_-] = 2\epsilon k_{1,3} + 2\epsilon^3 k_{3,333} \quad (4.6)$$

which means that, if  $\epsilon$  is chosen to be small enough, the subtraction of the two resulting terms gives the kernel. This of course holds for all time points after the impulses occurred.

The method can be generalized for higher order kernels, too. The second order kernel would be evaluated using four stimuli,

$$\begin{aligned} \gamma_{++} &= \begin{cases} \epsilon & \text{if } t = 0 \text{ or } t = 1 \\ 0 & \text{else} \end{cases} & \gamma_{--} &= \begin{cases} -\epsilon & \text{if } t = 0 \text{ or } t = 1 \\ 0 & \text{else} \end{cases} \\ \gamma_{+-} &= \begin{cases} \epsilon & \text{if } t = 0 \\ -\epsilon & \text{if } t = 1 \\ 0 & \text{else} \end{cases} & \gamma_{-+} &= \begin{cases} -\epsilon & \text{if } t = 0 \\ \epsilon & \text{if } t = 1 \\ 0 & \text{else.} \end{cases} \end{aligned} \quad (4.7)$$

Now the resulting 2nd-order kernel can be computed via

$$k_{2,32} = k_{2,23} = \frac{1}{\epsilon^2} (x_t [\gamma_{++}] + x_t [\gamma_{--}] - x_t [\gamma_{+-}] - x_t [\gamma_{-+}]) . \quad (4.8)$$

Generally, a fixed time interval  $\Delta t$  between two clicks played at all combinations shown in equation (4.9) can give the off-diagonal of the 2nd order Volterra operator along the diagonal  $\Delta t$ .

### Investigating higher order kernels using differently attenuated clicks

If taking a slightly different approach, the detection threshold of a system for a pulse is set to be at  $A$ , meaning that  $x(A + \epsilon) > 0$  and  $x(A - \epsilon) = 0$  in terms of the response. Equation (4.7) then becomes

$$\gamma_1 = \begin{cases} A & \text{if } t = 1 \text{ or } t = 2 \\ 0 & \text{else} \end{cases} \quad \gamma_2 = \begin{cases} A + \epsilon & \text{if } t = 1 \text{ or } t = 2 \\ 0 & \text{else} \end{cases}$$

$$\gamma_3 = \begin{cases} A + \epsilon & \text{if } t = 1 \\ A & \text{if } t = 2 \\ 0 & \text{else} \end{cases} \quad \gamma_4 = \begin{cases} A & \text{if } t = 1 \\ A + \epsilon & \text{if } t = 2 \\ 0 & \text{else.} \end{cases} \quad (4.9)$$

If doing the same computation as in (4.8) the result is

$$(x_t[\gamma_1] + x_t[\gamma_2] - x_t[\gamma_3] - x_t[\gamma_4]) = \quad (4.10)$$

$$2k_{2,12}\epsilon^2 + 3\epsilon^2(2Ak_{3,112} + 2Ak_{3,122} + \epsilon k_{3,112} + \epsilon k_{3,122}) \quad (4.11)$$

which means that this computation would indicate third (and higher) order kernels. Note that

$$x_t(\gamma_1) + x_t(\gamma_2) - x_t(\gamma_3) - x_t(\gamma_4) = 0 \text{ for } t < \delta t \quad (4.12)$$

holds regardless of the order of the system.

### 4.3.2 The STRF

On the other side of this rather unintuitive calculations is the STRF. Like the Volterra series, the STRF describes an unknown input-output system and performs a convolution in time. The difference is, that the series is truncated and that the input is given in form of a Time-frequency representation (TFR). Due to the additional parameter that is present in the input but lacking in the output, besides the convolution an integration about all frequencies is performed. The previous chapter showed how to convert an STRF into corresponding or equivalent Volterra kernels using general assumptions. The STRF does not describe a linear system, but rather one whose non-linearity is determined by a static one in the TFR.

Originally, STRFs were derived from 2nd-order Wiener kernels (Eggermont, 1993; Klein et al., 2000), and were evaluated in the same way; via the cross-correlation of (the TFR of) the input with the output. This is essentially the reverse correlation technique (De Boer and Kuyper, 1968). When dropping the assumption of a Wiener functional, optimizing the filter with respect to the input-output functions leads to estimators that depend less on the statistics of the input. This resulted in evaluation techniques such as the general linear

model (GLM, Nelder and Wedderburn (1972); Kouh and Sharpee (2009); Calabrese et al. (2011)), maximum informative dimension (MND, Sharpee et al. (2004) or the classification based receptive field (CbRF, Meyer et al. (2014)) that give similar results but need different amounts of data (Meyer et al., 2014). Especially the evaluation method by Meyer et al. (2014) makes use of the binary character of the spike train.

## 4.4 Materials and Methods

### *Experiments for the STRF and higher order kernel evaluation*

Experiments were performed on 11 adult male ketamine/xylazine anesthetized Mongolian gerbils (*Meriones unguiculatus*) aged 3-16 months with a body weight of 80-120 g. All experiments were conducted in accordance with the international National Institute of Health *Guidelines for Animals in Research* and with ethical standards for the care and use of animals in research defined by the German Law for the protection of experimental animals. All experiments were approved by an ethics committee of the State of Saxony-Anhalt, Germany. *Surgical procedure* and *Stimulus generation and setup* were as explained in chapter 1.4.

*Recording paradigm.* Once a single unit with good SNR ratio was found, it was characterized on the basis of its frequency response area (FRA) using pure tones with a frequency in the range from 0.25 to 16 kHz, and with a duration of 100 ms (including 5-ms raised-cosine-squared ramps at on- and offset). The signals were separated by silent intervals of 500 ms. The signal frequencies were equally spaced on a logarithmic scale in 1/4 octave steps. The signal intensities were varied from a highly supra-threshold level down to threshold in steps of 10 dB. Each combination of level and frequency was repeated ten times.

### *Experiment for the Differential Sampling approach*

For two units, recordings were made in the cochlear nucleus (CN) of the pigmented guinea pig. The detailed procedure can be found in Neuert et al. (2004). Briefly, the animals were anesthetized with urethane and Hypnorm was administered as supplemental analgesia. The animal was placed in a stereotaxic frame using hollow speculai in the ear canal. A small whole was made in the bulla and, using a silver-wire, the compound action potential (CAP) was measured. The whole was resealed with petroleum jelly. A craniotomy was performed, parts of the cerebellum were aspirated and a glass coated Tungsten microelectrode was placed in the cochlear nucleus. Electrodes were advanced using a hydraulic microdrive. Once a unit was found, it was characterized by measuring its response to varying frequencies and attenuations to determine the frequency response area (FRA). The stimuli were monitored using a condenser microphone. After measuring the FRA, double clicks combinations as in equation (4.7) were played. The time between the click combinations was twice the frequency that the unit was most sensitive to. Every click combination was repeated 100 times, the time between the first clicks of each stimulation was 200 ms. The clicks were played in a randomized manner.

#### 4.4.1 Stimuli for the STRF and higher order experiment

Additionally to the frequency response area (FRA), which is for descriptive purposes only, the units were stimulated with stimuli for the evaluation of an STRF, and with click stimuli to evaluate them via Differential Sampling.

*The STRF.* The stimuli used for the estimation of the STRF were as described in Meyer et al. (2014). It was an ensemble composed of frequency modulated FM tones. The mod-

ulation function was linear, meaning  $\sigma(t) = f_0 + \frac{f_1 - f_0}{t_1}$  with randomly drawn starting and ending frequencies and time,  $f_0$ ,  $f_1$  and  $t_1$ , respectively. The form of a sweep was then

$$s(t) = \cos(2\pi\sigma(t)t + \Phi) \quad (4.13)$$

with the phase  $\Phi$ . The entire stimulus  $\varphi$  consisted of a sum of random sweeps with an average density of 3 to 4 sweeps. The length of the sequence was 100 s and five different sequences were repeated five times each.

*The click stimuli.* The third set of stimuli consisted of single and double clicks at various attenuations. Single click stimuli were repeated at least 50 times and played at 3 or more different attenuations. To investigate higher order Volterra kernels according to (4.10), the double click stimuli were repeated at least 90 times with click intervals of 200, 100, 50, 25, 12.5 and 6.25 ms. (For one unit, the 200 ms interval was not used.) For every click interval, four different combinations of double clicks were played (compare equation (4.9)). Both sound levels were chosen in a way that a single click could elicit a response; the difference between the two levels was 5 or 10 dB.

#### 4.4.2 Evaluation

*STRF estimation.* The STRFs were evaluated using a generalized linear model (Nelder and Wedderburn, 1972; Kouh and Sharpee, 2009; Calabrese et al., 2011). This means, that the linear filter (the STRF operator) was followed by an instantaneous non-linearity that generated the spikes. In this study, only the STRF is of interest.

*Higher order kernel estimation.* The evaluation of the higher order Volterra kernels is straightforward (see equation (4.10)). These evaluations require a continuous output signal, which is not given by a measured spike train.

*Smoothing the peristimulus time histogram* To generate a smooth spike output function, the resulting spikes were converted into a density using the bayesian binning algorithm by Endres et al. (2008). Instead of creating a perstimulus time histogram (PSTH) from rigid bins and smoothing, this algorithm uses flexible bin sizes depending on the number of spikes. It has been shown to be an optimal estimator of the inferred PSTH Endres and Oram (2010)

*Testing the 2nd-order Volterra kernel for ergodicity* The data set was split into three subsets of equal size and for the subsets, equation (4.10) was used. In the next step, for each subset the difference between itself and the mean of the three sets was calculated. Using a fixed time window for all results of one unit to be defined later, the histogram of these differences was given a mean of zero and the variance was calculated for a collection of first clicks and for every following results individually. For the differences, an F-Test was performed and if the p-value was below 0.01, the results were deemed significant. This calculation made use of equation (4.12) that states that the response before the second click theoretically should be zero. The interval was estimated as follows. The starting point was taken from the responses to the first click. If the absolute value of the result of the addition and subtraction of the four stimuli was above 10% of the maximum of all the responses, then this was chosen as a starting point. The shortest estimate of all starting points was used;

sets where the interval between the two clicks was below an estimated starting time were not considered. In the same manner, the ending point was evaluated after the second click. One unit was excluded from further evaluation as the algorithm could not be implemented.

#### *Estimation of the 2nd order Volterra kernel*

In the experiment with clicks of opposite polarity, the evaluation was done by binning the spike responses into a PSTH of 0.1 ms binwidth. Afterwards, this PSTH was upsampled to 0.005 ms binwidth and smoothed using the autocorrelation of the single click response (see Schreiber et al. (2003)).

## 4.5 Results

### 4.5.1 Using the ergodic hypothesis to test for significance

Figures 4.1-4.3 show the probability density functions from all spikes (black) and three randomly selected disjunct subpopulations (red, blue, green) as well as the result from the Differential Sampling operation described in equation (4.10) and the entire spike count. All three results (starting at 200, 100 or 50 ms, depending on the interval between the clicks, denoted with black dots on the bottom left) were significant, even though the amplitude of these results was lower than that of the noise sample (denoted with white dots, if no white dots are indicated, then the click interval was too short to use). The results were significant because the individual results of the subpopulations converged well against their mean. Figure 4.1 shows the individual responses of the action potentials (right) and probability density functions (left) for a click interval of 200 ms. The results at the bottom left show that the results for the three different subgroups are similar, while on the right it can be seen that before the 2nd click, they do not necessarily even share the direction of magnitude. Figure 4.2 shows the same unit with a click interval of 100 ms, thus the diagonal of the 2nd-order Volterra kernel starting at 100 ms click difference. The amplitude of the signal was low and for the green subpopulation equal to zero, but it was significantly different compared to the noise reference of the unit. The overall amplitude of the noise (black line between the white dots) was larger than the amplitudes of the kernels in that as well as figure 4.1, but again rather noisy in its response. For figure 4.3, the resulting higher order estimate was equal to zero for the entire duration of the response window which was statistically significant, as it was zero in all three subpopulations. Thus, the statistical significance also indicated correctly zero responses.

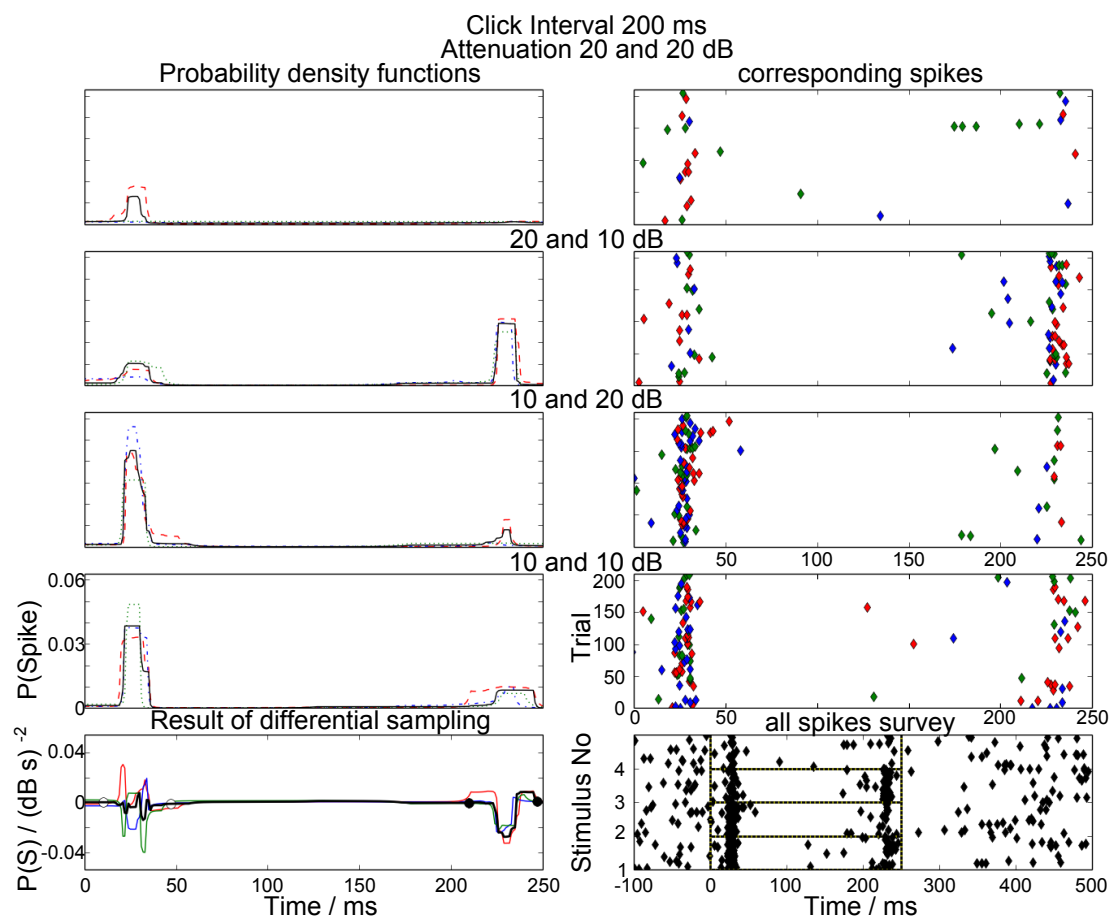


Figure 4.1: Top four rows: Results for the four differently attenuated click combinations with the different subgroups in the time interval 0–250 ms. Right: The scatter plots, with every subgroup being color-coded differently (red, blue, green). Left: The calculated probability density functions (pdf) for the scatter plots on the right, the different lines denoting the same subgroups and the black line the overall population. Bottom Left: The result of the double click paradigm, from the time point of the second click (200 ms) this corresponds to the higher order kernel. Again, the different results are color-coded, and black denotes the result from all spikes. The white dots in the plots show the time window in which the data for the noise reference was gathered. The black dots show the interval in which the diagonal of the kernel was evaluated. The mean of the different subgroups in the interval of the resulting kernel was significantly different from that of the noise reference. The bottom right shows the entire spike count for a longer time frame (-100 to 500 ms) for the different stimuli and trials; the black windows indicate the section that can be seen on the four plots above.



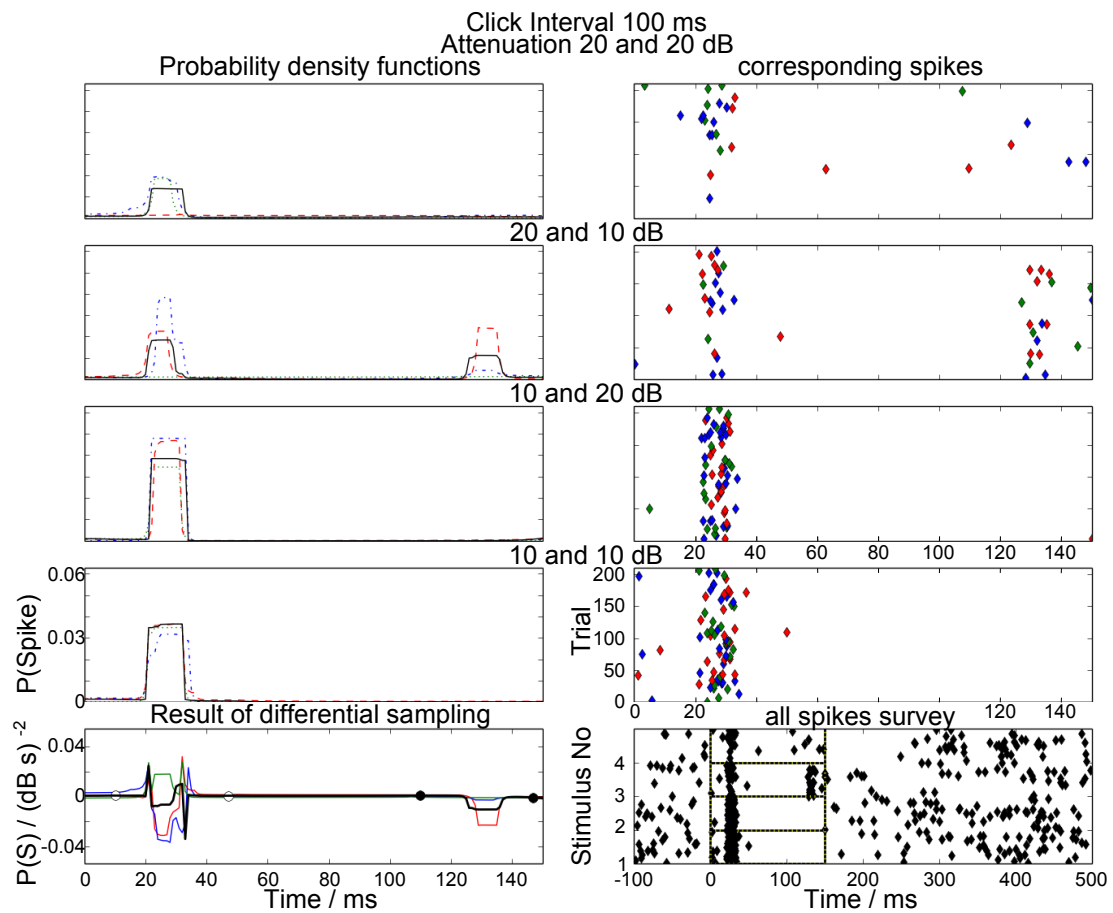


Figure 4.2: The same unit and notation as in figure 4.1, but for a click interval of 100 ms. Note that the amplitude of the result (bottom left) of the entire population is much larger than the result of the kernel. This noise reference in its overall amplitude is also larger than the result in figure 4.1, but again, the three subpopulations indicate similar results.

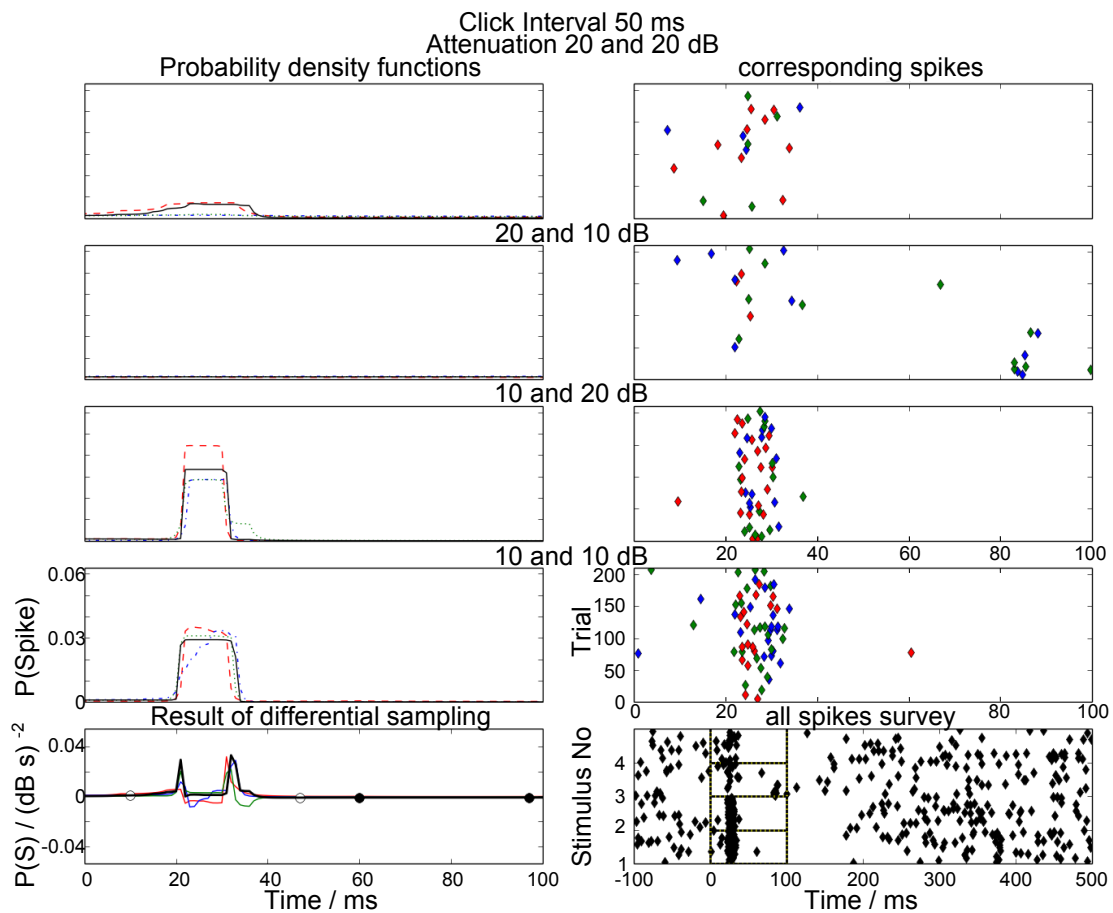


Figure 4.3: The same unit and notation as in figure 4.1 and 4.2 but for a click interval of 50 ms. This kernel reached statistical significance, as the individual responses did not differ from their mean. This is not in full agreement with the spike response, as it indicates a very weak suppression.

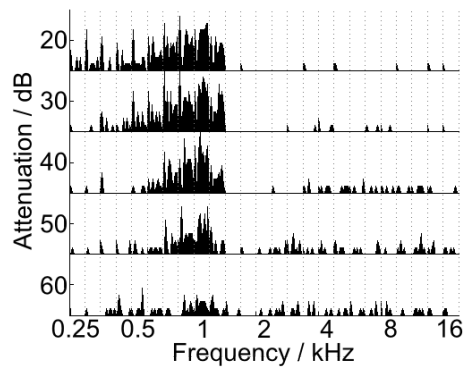


Figure 4.4: The frequency response area of the unit. Every square denotes the peri-stimulus time histogram of the indicated attenuation and frequency. This unit was excited at all attenuation for frequencies of 1.2 kHz and was suppressed at frequencies of 1.4 kHz. At lower attenuations (thus higher sound levels) the unit elicited a response at lower frequencies, too.

#### 4.5.2 Examples of different units

*An example unit with a high amplitude kernel at 25 ms and a potentially undetectable longer one at 6 ms*

Figures 4.4-4.9 show the pure tone response, STRFs with corresponding Volterra kernels and the responses of the double click paradigm of one single unit. Figure 4.4 shows the pure tone response of the unit, which showed ongoing activity for frequencies between 0.5 and 1.5 kHz for various sound levels. Above 1.5 kHz, the unit seemed to have suppressed its activity according to the pure tone response. Figures 4.5-4.7 show examples of the double click measurements for different click intervals (50, 25, 13 and 6 ms). Note that for this unit no recordings with a 200 ms inter-click interval were made.

Figure 4.5 shows the response to the required four different click combinations in the spike response and the probability density for the diagonal of the diagonals of the higher order kernels at 50 ms time difference as well as the result itself. It was not significantly different from noise, as the three subpopulations do not converge against their mean. For the time difference of 25 ms shown in figure 4.6, the resulting diagonal of the higher order kernels is significantly different from zero, which can also be seen in the spike response on the right side. Note that the response to the different individual click combinations did not converge against one mean (unlike their combination) but rather vary, about as much as those from figure 4.5. The deviations nullified each other in the total result which led to a prominent result.

Figure 4.8 shows the result for 12 ms interclick interval. Again, the result was significantly different from the noise reference, though the convergence of the subpopulations was not as strong as for the 25 ms click interval shown in figure 4.6. The spike response on the right confirms this result, as the second response subtracted from the first one shows a

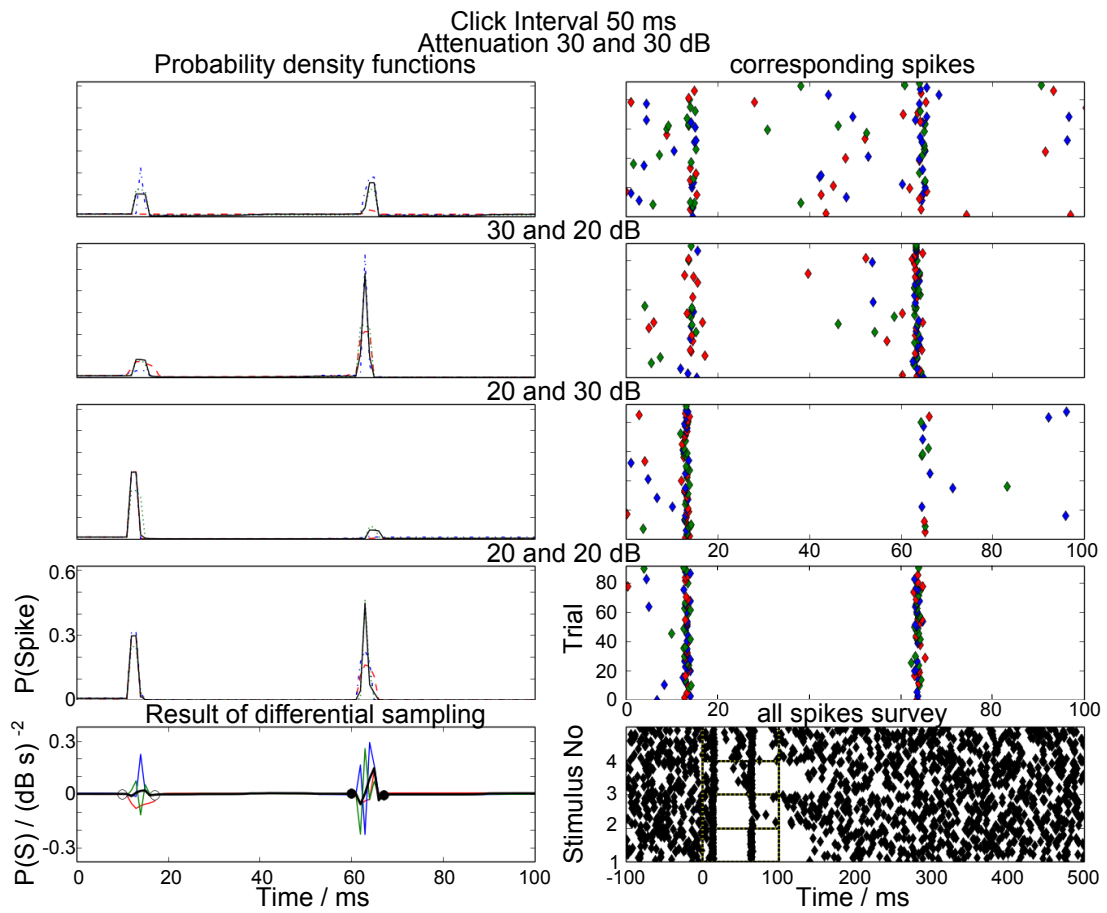


Figure 4.5: The same unit as in figure 4.4 and notation as in figure 4.1. The result of the Differential Sampling test for higher order kernels for this click interval (50 ms) was not significantly different from the noise reference of this unit.



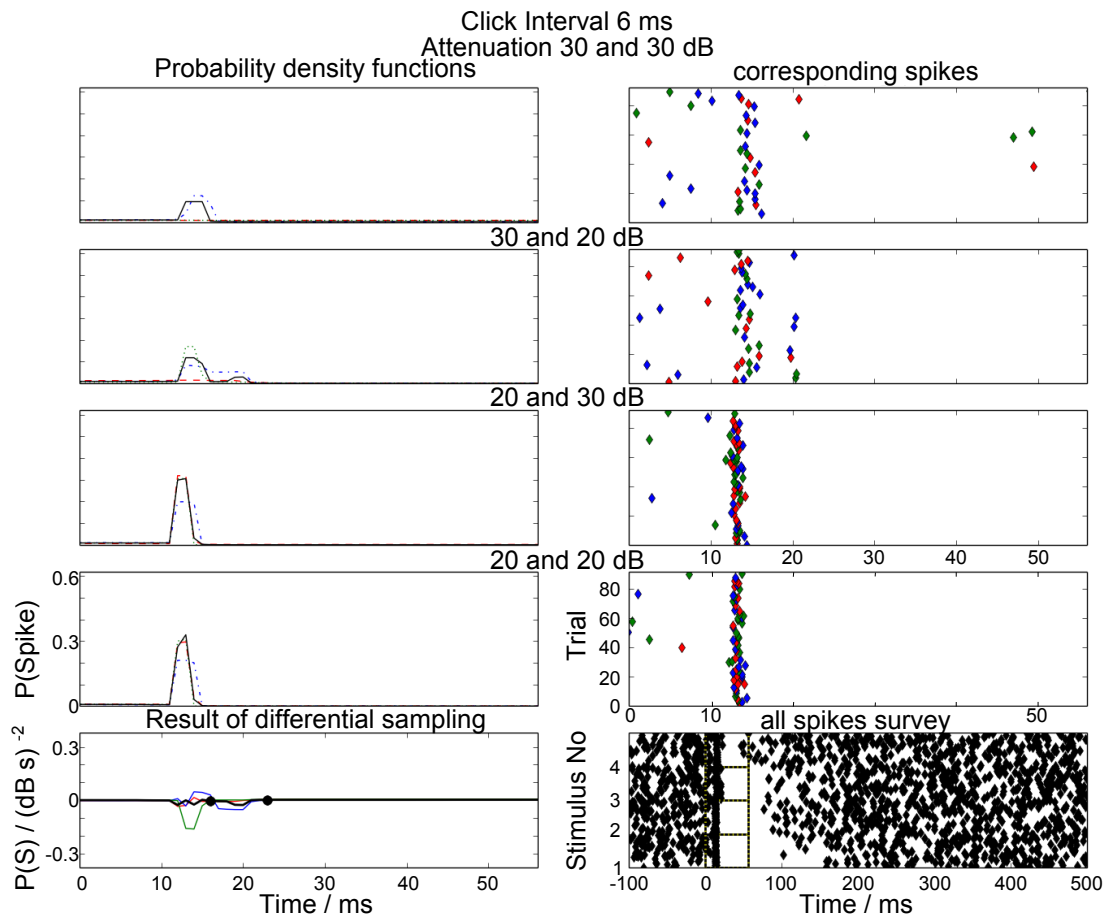


Figure 4.7: The same unit as in figures 4.4-4.9 and notation as in figure 4.1. For the click interval of 6.0 ms, the corresponding diagonals of the higher order operators are almost zero.

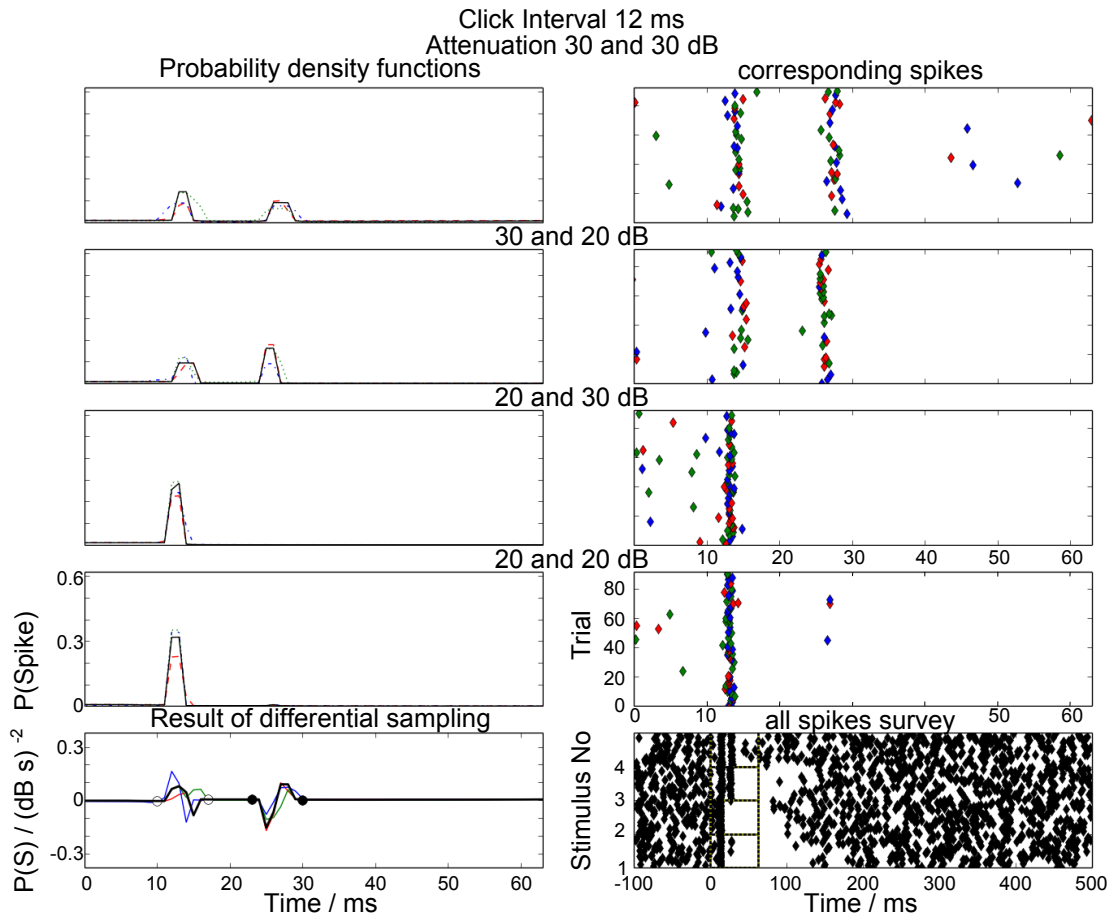


Figure 4.8: The same unit as in figures 4.4-4.9 and notation as in figure 4.1. For the click interval of 13.0 ms, the subpopulations were more divergent in respect to the mean than for 25 ms click interval (compare with figure 4.6, but still significantly different from the unit's noise reference).

strong difference, and the last responses are zero or very close to zero. Figure 4.7 shows the response for 6 ms click interval. It was statistically significant, but the amplitude is close to zero. In the scatter plot on the bottom right, different durations of the suppression of the spontaneous rate can be seen for the four different click-combinations on the time scale up to 180 ms, but these differences were too small to be significant.

Figure 4.9 shows the STRF of the squared TFR, the STRF of the absolute value time-frequency representation, the 2nd-order Volterra kernels corresponding both representations and the 1st-order Volterra kernel corresponding the absolute value STRF. It was shown in the previous chapter that the 2nd-order Volterra kernel approximation of the absolute value TFR STRF is a scaled version of the exact 2nd-order Volterra kernel of the square TFR STRF. The STRF of the square representation confirms the ongoing response seen in figure 4.4 and the frequency selectivity. Suppressive fields are to be seen at higher and lower

frequencies. The suppression is less prominent in the absolute value STRF, the excitation is slightly weaker but essentially the same. The red lines in the 2nd-order Volterra kernels denote the responses of the double clicks that gave results significantly different from the noise reference with amplitudes larger than 0.05 and their time frame. It can be seen that these higher order operators are much shorter, however with much longer interactions times. The resulting STRF showed interaction times around the main diagonal, on a time scale of less than 10 ms, and set in 13 ms later. The same holds true for the 1st-order kernel; the corresponding one is a band-pass filter with high magnitudes of the envelope between 20 and 37 ms.

#### *A unit changing its frequency range with sound level*

Figures 4.10-4.13 show a unit that has a pure tone response that could be defined as highly non-linear. For a linear system (as could be defined by the first operator only in equation 4.1), the increase of a response eliciting input would cause an elicited response as well. A reduced criterion would be to keep the response steady. For this unit, increasing a 1 kHz pure tone input causes it to stop its response, hence it is defined as less linear than the unit shown in figures 4.4-4.9. According to the FRA shown in figure 4.10, this unit was active around 1 kHz at low sound pressure levels and, upon increasing the sound level, the unit was then responsive at lower frequencies of less than 0.5 kHz. Figure 4.11 shows the STRF based on the squared time-frequency representation of the unit and the equivalent 2nd-order Volterra kernel recorded at 30 dB attenuation. Its excitatory and suppressive subfields were below 2 kHz and started immediately at the onset. In the Volterra kernel, the arrows denote the inter-click intervals at which Differential Sampling showed significant responses. The dotted lines denote time constants of more than 52 ms. The onset of the directly measured kernels was later than that one of the STRF corresponding kernel and, as for the other units shown, had a higher interaction time constant.

Figure 4.12 shows the result of the Differential Sampling operation for a click interval of 13 ms. This result was statistically significantly different from the noise reference, and, despite being noisy, can be seen to converge on several intervals of the time axis. Figure 4.13 shows the histogram of noise reference and the histogram of the response in figure 4.12. The overall histogram could indicate that the unit did not respond ergodic or that the response was too noisy for the amount of measurements taken; but the convergence in the kernel shown in 4.12 shows that ergodic responses were possible.



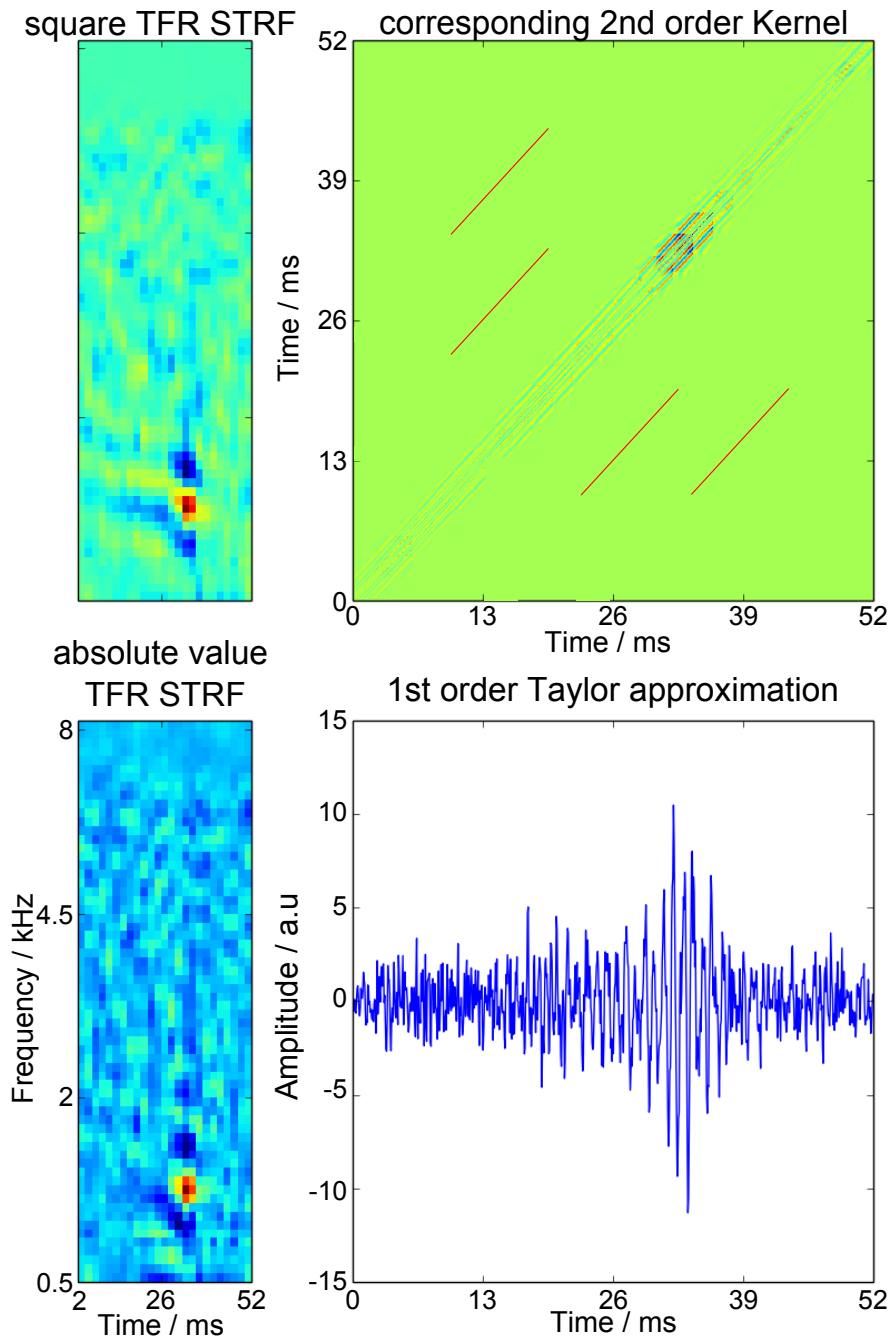


Figure 4.9: The STRFs of the unit shown in figures 4.4-4.7 recorded at 40 dB attenuation using two different representations and their corresponding Volterra kernels (the square STRF has the 2nd-order Volterra kernel as the exact analogue). The red lines indicate 2nd-order Volterra kernels found with double clicks that were significantly different from the noise reference and had an amplitude above 0.1.

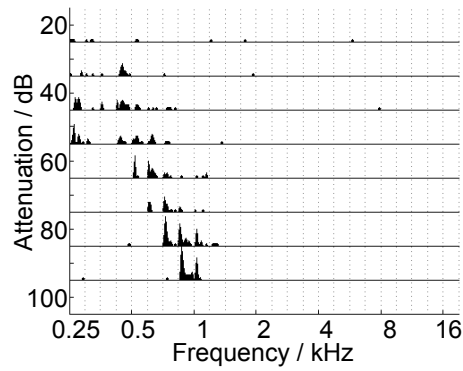


Figure 4.10: Frequency response area of a unit that changed its frequency selectivity for different attenuations.

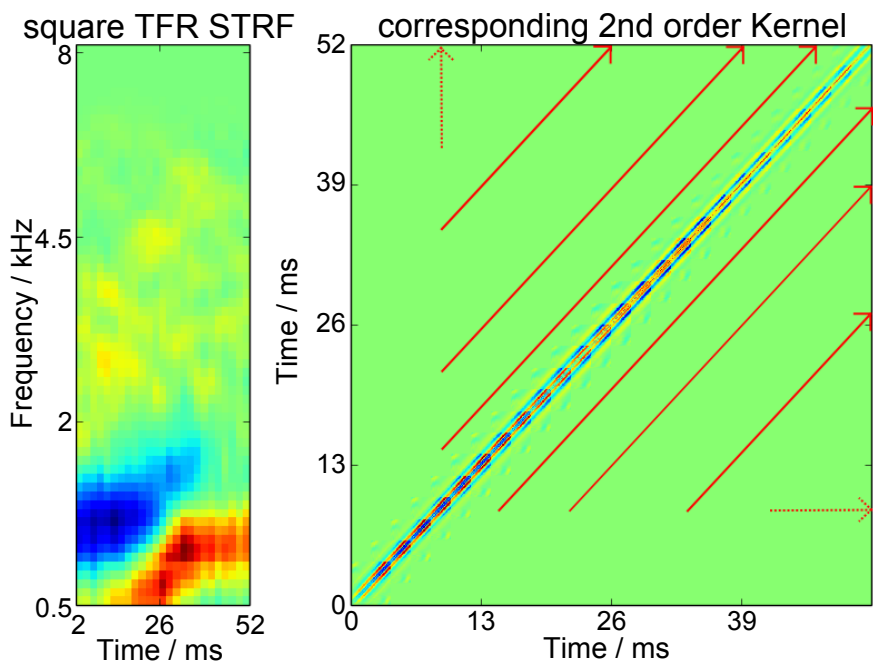


Figure 4.11: The STRF based on the squared TFR recorded at 30 dB attenuation of the unit shown in figure 4.10 and the equivalent 2nd-order Volterra kernel.

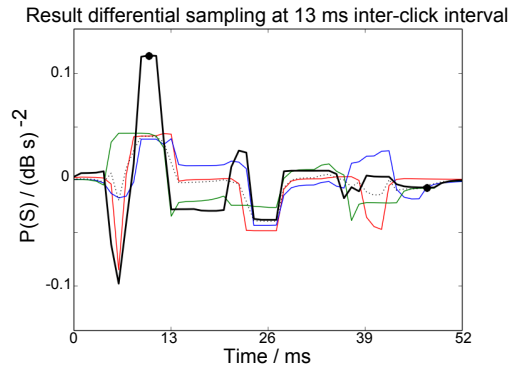


Figure 4.12: Resulting diagonal of higher order kernels for 13 ms click interval for the unit shown in figures 4.10-4.11.

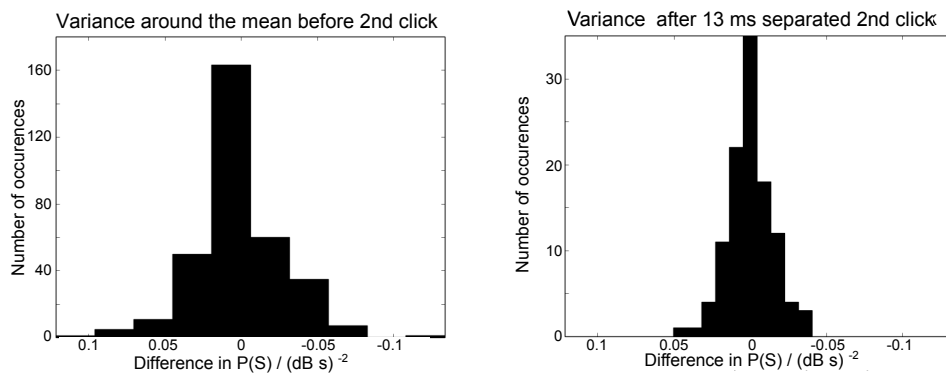


Figure 4.13: Histograms of the variance of the noise and the resulting 13 ms inter-click interval kernel for the unit shown in figures 4.10-4.12. The reference window for the noise and the higher order kernels was as described in Materials and Methods. For the noise reference, it included only results of the Differential Sampling operation before the 2nd click. The variance of the 13 ms inter-click interval response was significantly smaller than that of the noise reference according to the F-Test.

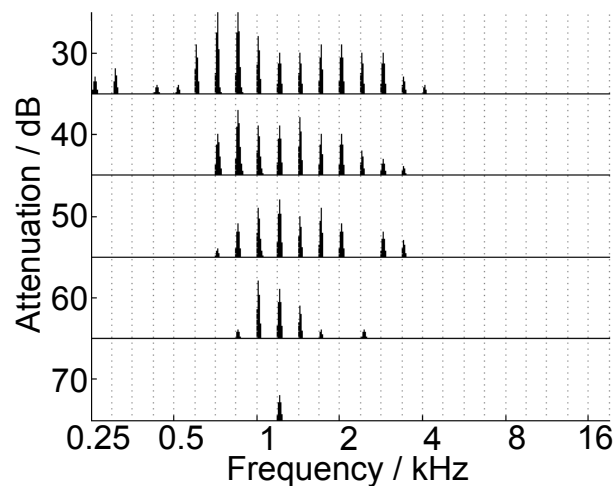


Figure 4.14: frequency response area of a unit that showed onset like behavior and an increase in frequency selectivity from 1.4 kHz at low sound levels to 0.25-4 kHz for high sound levels.

*A unit showing slow adaptation to clicks*

Figures 4.14-4.17 show the responses of a single unit to a pure tone stimulation and to different double clicks. The STRF response is not available as the unit was lost after the first repetition of STRF stimuli, and thus, the resulting STRF turned out to be too noisy. The diagonals of the higher order operators had high amplitudes, but only one of them (6 ms inter click interval, figure 4.17) turned out to be statistically significant from noise. In figure 4.15, in the scatter plots with the response where the first click was 60 dB attenuated, it can be seen that the unit responded with fewer spikes to the click during the course of stimulation. This indicates a long term adaptation to the sound level to which the evaluation method of using subsets is insensitive. The same observation can be made for the inter click interval of 25 ms shown in figure 4.16. Here, a shorter evaluation window for the result (denoted by the two black dots in the plot showing the result of the Differential Sampling operation) could have caused the result to be significant. Figure 4.17 shows the results of the paradigm for the 6 ms inter click interval.

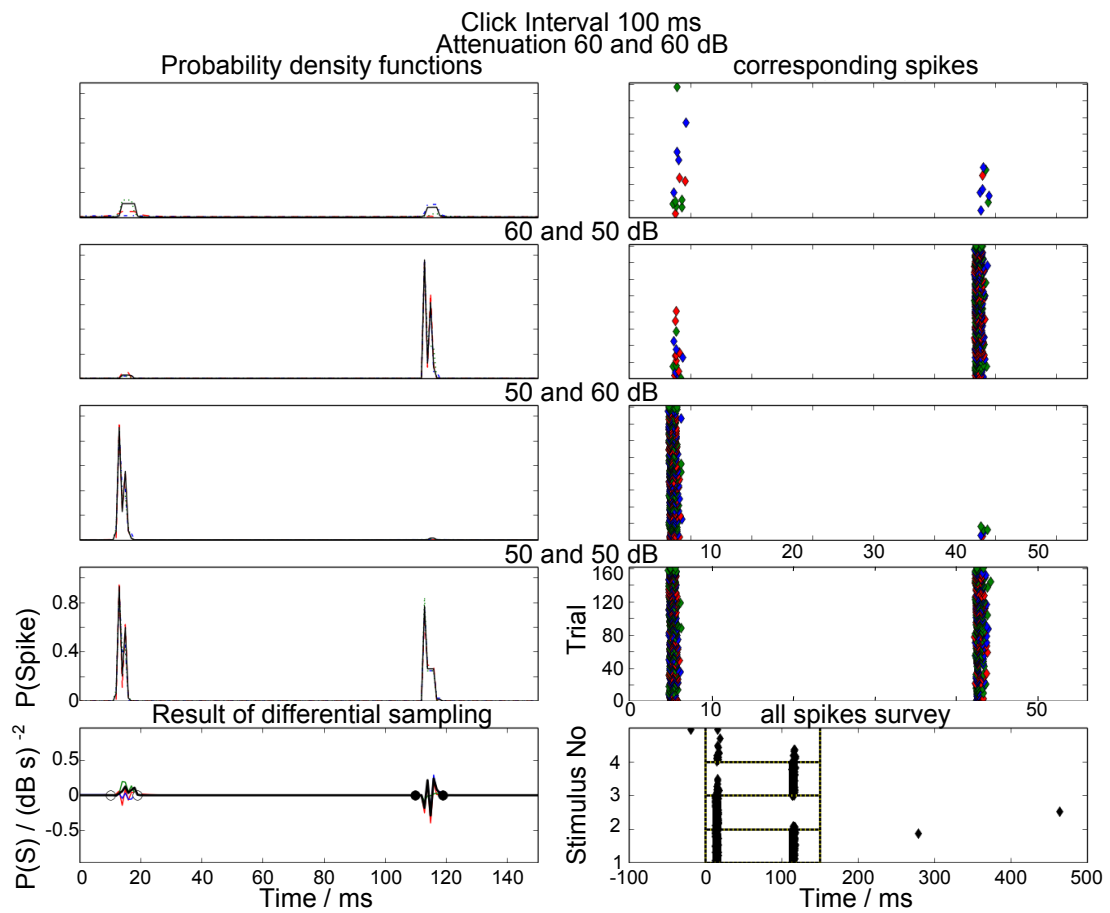


Figure 4.15: Responses to four different double clicks with 100 ms interval between the clicks and the results of the Differential Sampling operation for the unit shown in figure 4.14. The result (starting from 100 ms) was not significantly different from the noise reference.

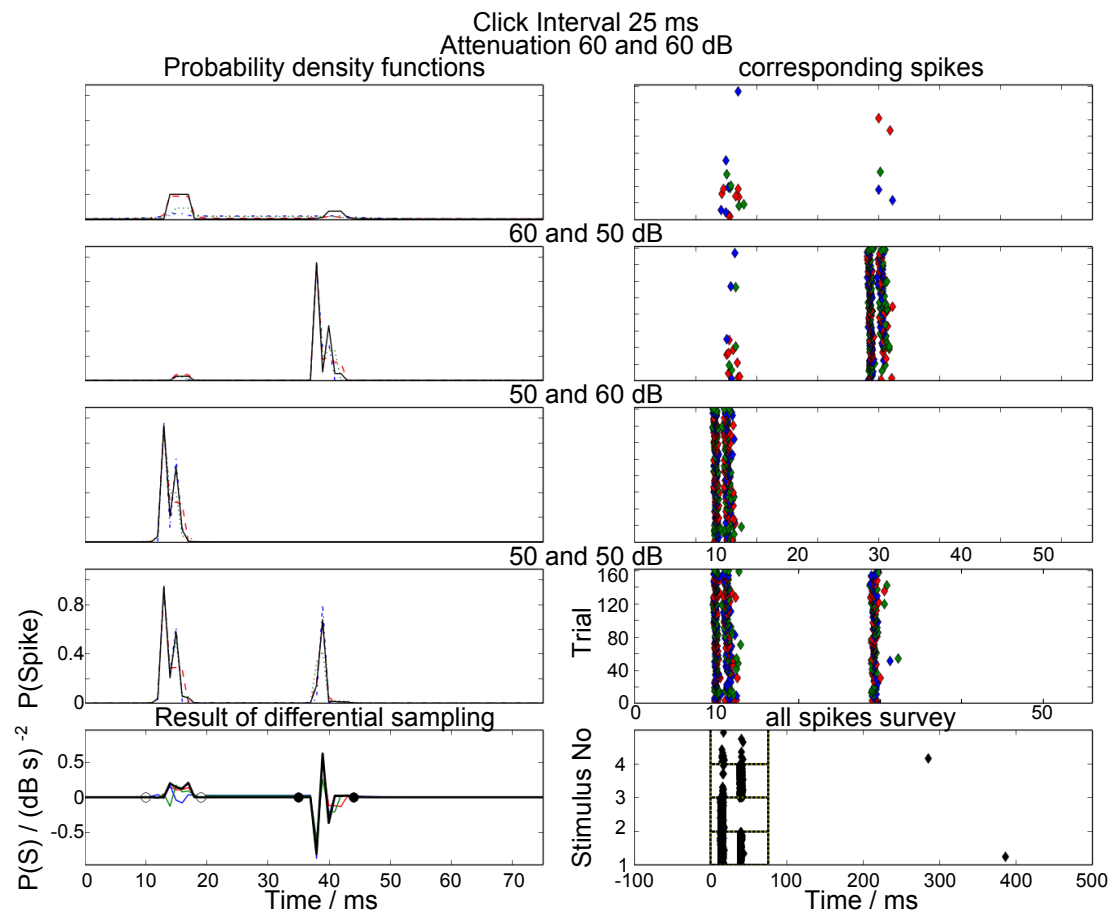


Figure 4.16: Responses and result of the Differential Sampling operation for the unit shown in figure 4.14 and a click interval of 25 ms. The result (starting from 25 ms) was not significantly different from the noise reference.

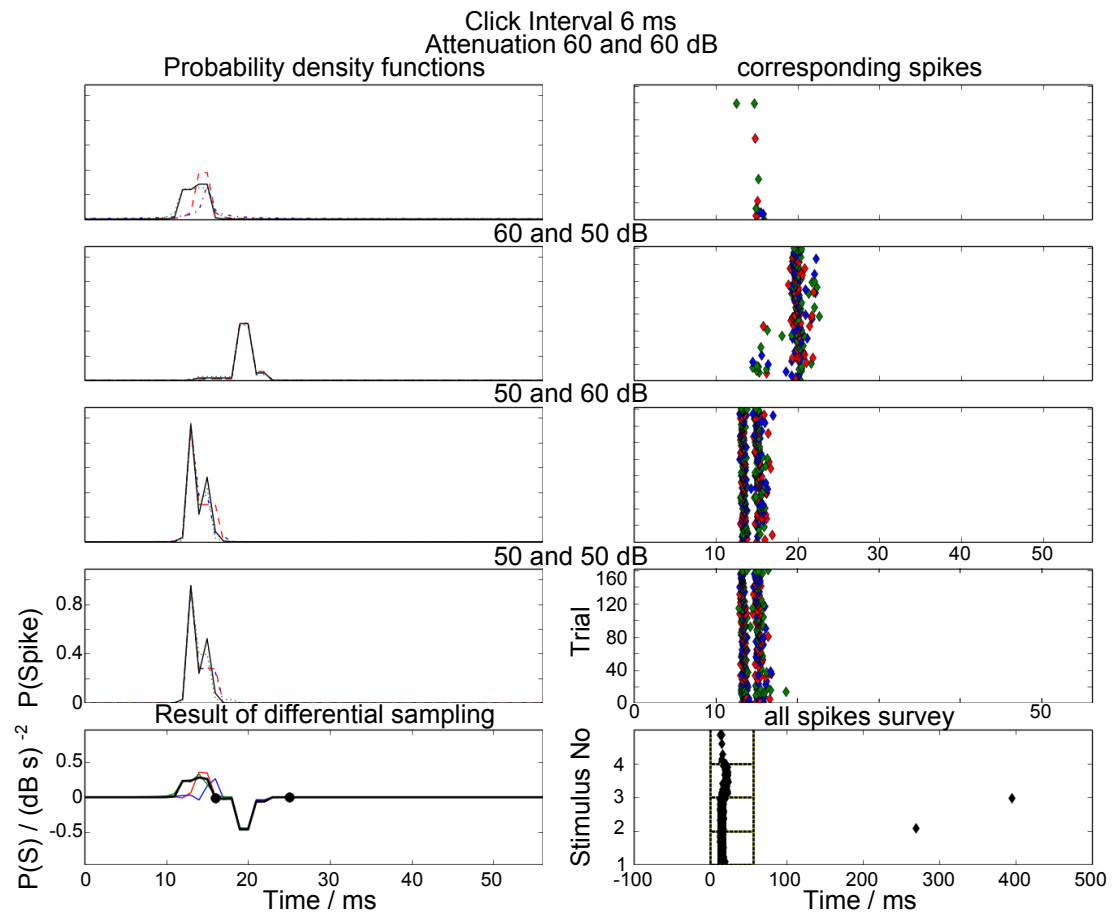


Figure 4.17: Responses and result of the Differential Sampling operation for the unit shown in figure 4.14 and a click interval of 6 ms. The result (starting from 6 ms) was significantly different from the noise reference.

Inter-click interval	200	100	50	25	12.5	6.25
Number of units with significant non-zero kernels	5	5	3	3	6	3
Av. duration of 2nd-order kernel/ ms	42	23	28	19	21	20
Shortest duration of 2nd-order kernel/ ms	13	7	13	7	7	9
Longest duration of 2nd-order kernel/ ms	98	37	37	37	37	37

Table 4.1: The number of units that had an amplitude of more than 0.05 and were statistically significant for the different inter-click intervals and the duration of the resulting kernels. The duration was estimated as described in Materials and Methods.

#### *Analysis of all units*

In total, 77 double click measurements were made in 13 units. Of these, 39 measurements were significant by means of the F-test, and had an amplitude larger than 0.05 (see also table 4.1 and figure 4.18). Two units showed significant results for all click intervals (one of them shown in figures 4.10-4.13) and one unit did not show any significant results at all. All other units showed significant results for one to three different click intervals (with the one shown in figures 4.4-4.7 not being tested for 0.2 ms inter-click interval).

Table 4.1 shows the temporal extent of the higher order operators estimated with the double click paradigm. This is an equivalent to the temporal extent of an STRF. As the onset of the kernel was usually after 10 ms (data not shown), the duration assumed in the STRF seems sufficient. The lowest temporal extent is about 10 ms, which corresponds to a frequency of 1 kHz.

Figure 4.18 shows the maximal amplitude of all units and indicates their statistical significance. The trend of weaker amplitudes for higher inter-click intervals confirms the approach. As several units were still responsive at the 200 ms inter-click interval, the total pause time between the clicks (500 ms) might have been too short for some units.



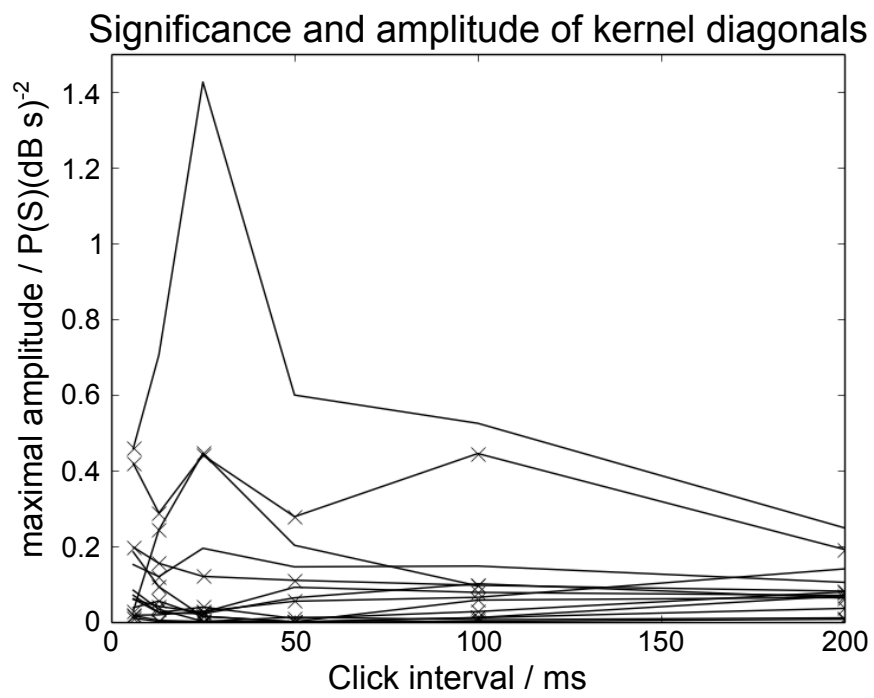


Figure 4.18: The highest amplitude of all units for the different inter-click intervals; crosses denote statistical significance. Some of the very high amplitudes were not considered significant due to the conservative time window (compare figure 4.16).

### 4.5.3 Second order Volterra estimation in the cochlear nucleus

Figures 4.19 and 4.20 show the FRA, 2nd order Volterra kernels and the simulated response of these kernels to various pure tones, respectively. The Volterra kernels were measured by using clicks of opposite polarity and applying equation (4.8) to the resulting spike train. The results were computed by convolving the 2nd order Volterra kernels with various pure tones numerically. The kernels correctly predict the frequency at which the unit is most responsive.

Most importantly, the temporal patterns on the Volterra kernels in figures 4.19 and 4.20 match the best frequencies of the units. While for the unit in figure 4.20 three fields of excitation and suppression are sufficient (counting only one side of the main diagonal, as the kernels are symmetric), the unit shown in figure 4.19 has 8. However, these 3 seem to suffice to make a good prediction of the output.

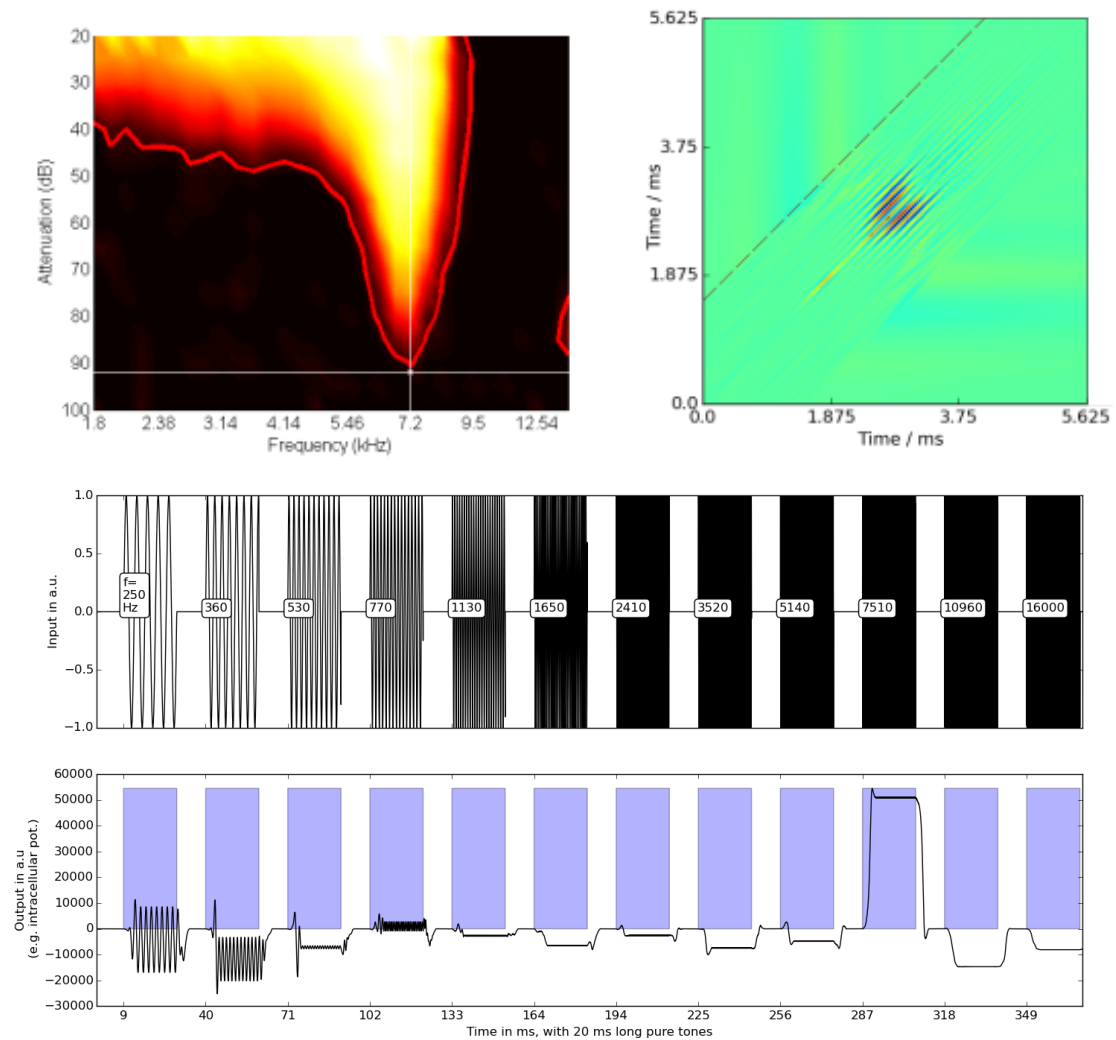


Figure 4.19: Top left: The response of the unit to various frequencies and attenuations. The unit is most responsive at 7.2 kHz. Top right: The 2nd order Volterra kernel of the unit, measured with combinations of double clicks. The diagonal line is at 1.5 ms, which was the longest time interval between two clicks. Bottom, upper part: Stimuli used as input for the convolution with the kernel at the top right. The frequencies are written in the boxes. The lower part shows the response of the convolution, the blue areas denote the stimulation times. At lower frequencies, phase locking can be observed. The 7510 Hz stimulus elicits a strong response, which is in good agreement with the data from the pure tone response.

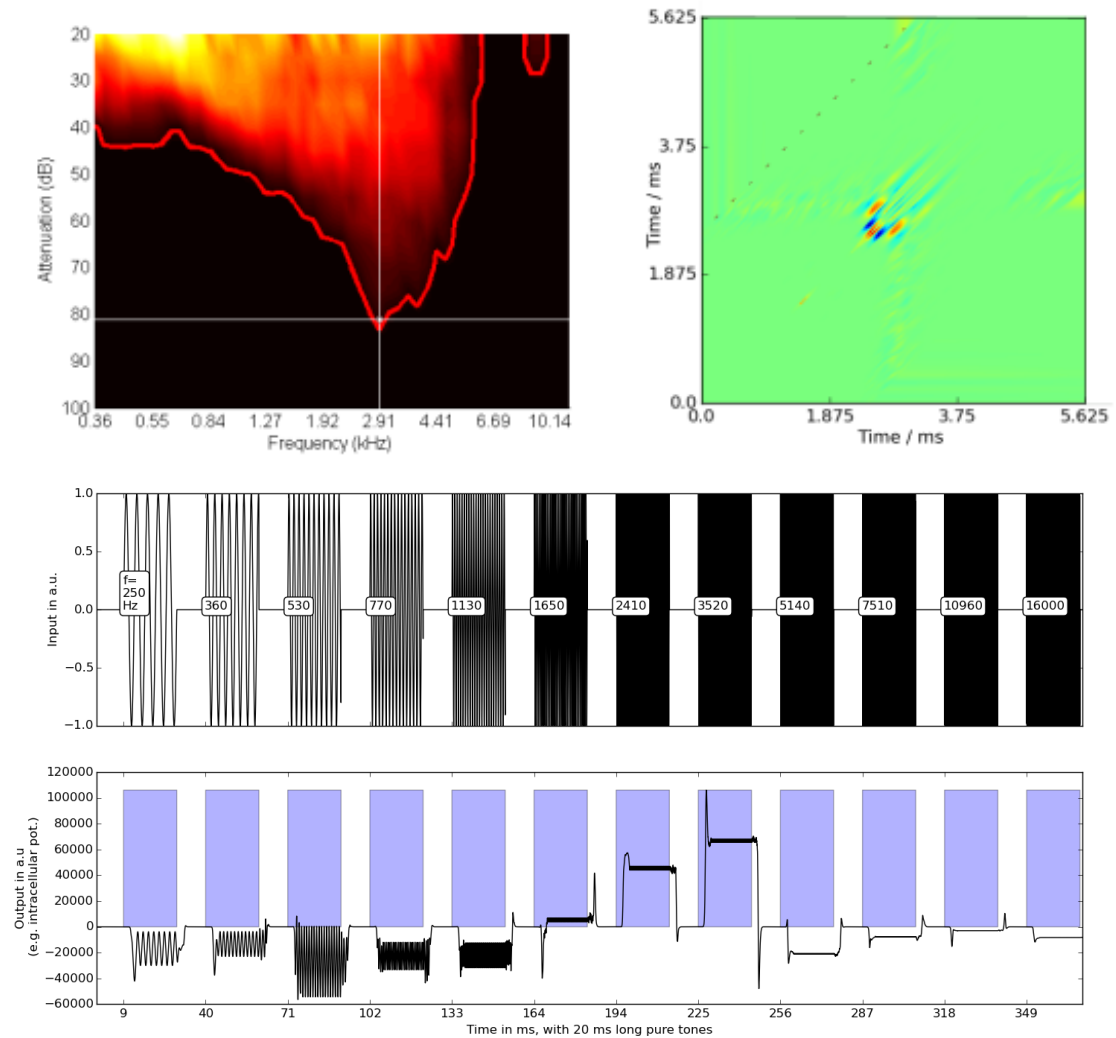


Figure 4.20: Same notation as in figure 4.19, but for a different unit. Here, the frequency where the unit responded best was 2.9 kHz. The diagonal on the kernel is at 2.5 ms, which again was the longest time interval between two clicks. The simulated pure tone response predicts good responses for 2410 and 3520 Hz, which again is in good agreement with the measured response.

## 4.6 Discussion

This study for the first time has implemented a method to measure 2nd-order Volterra kernels directly. Further, using a slightly different paradigm, results of higher order kernels were compared with common STRFs. It was shown that higher order operators exist on a much longer time-scale than usually assumed in the common STRF.

### 4.6.1 Technical aspects of the evaluation: delta function approximation

The use of approximated delta functions delivered by a speaker instead of real ones, which are difficult to produce, does not impose a large problem on the evaluation. As explained in van Hemmen et al. (2000), the recurrence relations ensure that if a function  $\varphi$  converges to a Dirac delta measure  $\delta$ , then

$$\varphi \rightarrow \delta \Leftrightarrow \int ds \varphi f \rightarrow \int ds \delta f \quad (4.14)$$

for real valued functions  $f$ . In the following, van Hemmen et al. (2000) showed that the Volterra kernels, identified with the approximate delta functions, also approximate the real kernel. This means that an approximated delta function as would be delivered by a speaker gives an approximated Volterra operator. The results shown in figures 4.19 and 4.20 show that the evaluated kernels really are capable of predicting responses.

### 4.6.2 The ergodic hypothesis and the comparison with the STRF

The main assumption (upon which the subsequent analysis relies) is that of an ergodic response. This means that over the period of repeated identical stimuli, the response of the unit remains similar with only minor changes.

#### *Experimental findings*

For the unit shown in figures 4.10-4.13 and 4.14-4.17 the data suggest that this might not be the case for different reasons. The unit shown in figures 4.10-4.13 shows a high variability in the sum of the first responses (figure 4.13 left), but on the basis of the stability of the kernel diagonal for the 13 ms inter click interval (figure 4.12), this was rejected. The unit shown in figures 4.14-4.17 shows slow adaptation to the stimulus for the more attenuated click (see the responses in figures 4.15 and 4.16). Adaptation to the stimulus level in the long term environment at the level of the inferior colliculus has been shown by Dean et al. (2005). It was described more thoroughly at the level of the auditory cortex by Ulanovsky et al. (2004), where longer time scales were found. The time scales at the level of the inferior colliculus were reported to be 160 ms (Dean et al., 2008), which is less than the 500 ms inter-stimulus interval used in this paradigm. Long-term adaptation, as described by Ulanovsky et al. (2004) could be observed for the 60 dB attenuated click in figures 4.15. Despite the adaptation, the response shown in 4.17 was stable, and thus this unit was not

rejected on the basis of the ergodic hypothesis either. Especially, the left-hand side of figure 4.17 shows that regardless of a weakened response to the stimulus during the course of the experiment, the quieter click was still able to reduce the response to the louder one.

#### *Comparison of the STRF and directly measured Volterra kernels*

Apart from the problems in finding Volterra operators, the advantages of the method presented are its mathematical clarity and its ability to investigate long term interactions. If it is assumed that the STRF gives a correct predictor for a given sound level, which is a reasonable assumption at the level of the inferior colliculus (Escabi and Schreiner, 2002; Escabi et al., 2003; Lesica and Grothe, 2008), then the difference between the two methods is as follows: Both operators are looking for a solution for the a dimension of

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}) \quad (4.15)$$

with the vector  $\mathbf{x}$  and the unknown non-linear function  $\mathbf{f}$ . The vector character might be due to unknown, or untestable other influences of the unit, or due to higher order derivatives of the spike rate. While Differential Sampling, due to the short inputs and pausing times, solves for (using  $n$  as the dimensionality of  $\mathbf{x}$  and  $\mathbf{f}$  and assuming that a higher temporal derivative is used)

$$x'_n = f_n(\mathbf{x}) + \sum \delta(t) \quad (4.16)$$

the STRF operator gives the solution for

$$x'_n = f_n(\mathbf{x}) + \cos(2\pi\sigma(t)t + \Phi) . \quad (4.17)$$

The data presented here first of all show how these operators might change, and that they are defined on different time scales. Both could describe the system in different steady states, but the state for the Volterra operator is much better defined than that for estimating the STRF. This can be seen by comparing the equations (4.16) and (4.17).

This knowledge could be used in higher, as well as lower processing stages. A first important step would be a generalization of equation (4.16) to parameterize the input conditions. This can be imagined similar to the approach taken in Kistler et al. (1997). There, the output was used to investigate different solution operators depending on the output, which is like including the initial conditions of equation (4.15). Further, the history of the input could be included as well, including the position in the phase space of (4.15). This is easier using Wiener kernel and is shown as early as the auditory nerve (Recio-Spinoso and van Dijk, 2006) that they change their form when investigated with higher sound pressure levels. The results by Escabi et al. (2003) and Dean et al. (2008) make it very likely that similar effects can be observed at the level of the inferior colliculus. Optimally, a similar method would be found for Volterra kernels that would allow a systematic investigation of their change.

The long interaction time observed in this study (up to 200 ms) can be seen in other click studies as well. Pienkowski et al. (2009) showed that using poisson distributed click trains, interaction times up to 500 ms can be reached in the auditory cortex. This approach again uses Wiener Operators, hence making a more detailed description of the phase of the unit impossible, but the time scale confirms the current findings.

Another example for an application would be the estimation of phase space and ongoing activity in higher areas. The study by Arieli et al. (1996) shows, cortical activity is largely driven by ongoing activity. Using local field potential information of the proximity, the initial conditions of (4.16) might be inferred and the predictability of the unit increased. The predictability would be increased even further, if the activity had several fix points in the phase space. Further, as Differential Sampling does not rely on spikes, but rather on a continuous output signal, intracellular recordings could offer new insights. This might also be possible with extracellular ones as the spike response is backwards correlated with the local field potential Okun et al. (2010) and thus could give insights on the response below the spiking threshold.

# Conclusion and outlook

In this thesis it was shown how a psychophysical phenomenon could be represented at the level of the midbrain (inferior colliculus) and why current predictor models would be unable to account for all the features shown. This is largely due to the linear behavior of the model, where here and in the following, any phenomenon that might be explained by performing one convolution will be called linear.

In the first chapter, the neuronal correlates of an effect that had been first discovered in psychophysics was investigated. This effect (called comodulation masking release, CMR) showed how the detection threshold for a signal could be decreased for certain maskers, meaning other simultaneous auditory signals. The effect was linked to a commonly used predictor for auditory neurons, the spectro-temporal receptive field (STRF). The STRF describes a mathematical operation based on the spectrogram or any other time-frequency representation of a stimulus. In chapter 3 it was shown how the STRF can be linked to the Volterra Series, which can be pictured as a Taylor Series with memory. The combined results of chapter 2 and chapter 3 confirmed some assumptions based on the STRF but raised objections to this sort of predictor as well. The distance and the position of the flanking bands are properties that might be correctly predicted by the STRF. On the other hand, the memory of the STRF is not sufficient to capture certain effects that were observed in the neuronal correlates of CMR. Another shortcoming of the STRF for CMR is its lack of definition regarding the stimulus strength; it is usually defined for one sound-pressure level, but as the CMR stimuli require amplitude modulation, problems arise. The nature of these problems and the mathematical background are further elucidated in the third chapter. In the fourth chapter, two new measurement methods are introduced. One shows how the masking of two double clicks can be used to show long term stimulus interactions. It is important to point out that these are defined for the system at rest, meaning in this particular case for the neuron that has not been stimulated for about half a second. The STRF, on the other hand, is supposed to predict the response for the stimulated (or excited) state of the system. Further, in a procedure called differential sampling double clicks of various polarities were used to estimate second order Volterra operators. This is the first application of a previously suggested measurement method.

The application of differential sampling could be used in earlier stages of the auditory system for a more strict and possibly robust mathematical analysis of its non-linearities. There, it might be possible to fully predict the behavior of a unit given higher order operators. In higher stages, especially the cortex, it could be used in intracellular recordings as it is more



robust and faster when used on a continuous output. Apart from reducing the distractions of the non-linearity of the output neuron, this would give the benefit of estimating the initial value problem. Further, as bifurcations might arise at this stage, different operators could be found, if the position in the phase space could be estimated by measuring other parameters like the local field potential. The experimental comparison of Volterra and Wiener operators lacks the theoretical background and thus might not be insightful. However, as differential sampling offers a new way of systematically investigating the input space, it would be interesting to develop new methods mapping high order operators into equivalent low order ones (in the spirit of finding eigen-vectors for matrices) and thus being able to possibly find higher order operators. These might then be able to predict responses to comodulation masking release or similar stimuli.



# Bibliography

- Aertsen, A. and Johannesma, P. I. M. (1980). Spectro-temporal receptive fields of auditory neurons in the grassfrog: I characterization of tonal and natural stimuli. *Biological Cybernetics*, 38(4):223–234.
- Aertsen, A., Olders, J., and Johannesma, P. I. M. (1981). Spectro-temporal receptive fields of auditory neurons in the grassfrog: Iii analysis of the stimulus-event relation for natural stimuli. *Biological Cybernetics*, 39(3):195–209.
- Aertsen, A. M. H. J. and Johannesma, P. I. M. (1981). The spectro-temporal receptive field. *Biological Cybernetics*, 42:133–143. 10.1007/BF00336731.
- Aertsen, A. M. H. J., Johannesma, P. I. M., and Hermes, D. J. (1980). Spectro-temporal receptive fields of auditory neurons in the grassfrog: Ii analysis of the stimulus-event relation for tonal stimuli. *Biological Cybernetics*, 38:235–248. 10.1007/BF00337016.
- Ahrens, M. B., Linden, J. F., and Sahani, M. (2008). Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *The Journal of Neuroscience*, 28(8):1929–1942.
- Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273(5283):1868–1871.
- Barrett, J. F. (1963). The use of functionals in the analysis of non-linear physical systems. *Journal of Electronics and Control*, 15(6):567–615.
- Branstetter, B. K. and Finneran, J. J. (2008). Comodulation masking release in bottlenose dolphins (*tursiops truncatus*). *The Journal of the Acoustical Society of America*, 124(1):625–633.
- Brückner, S. and Rübsamen, R. (1995). Binaural response characteristics in isofrequency sheets of the gerbil inferior colliculus. *Hearing research*, 86(1):1–14.
- Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., and Woolley, S. M. (2011). A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS One*, 6(1):e16104.

- Cant, N. B. and Benson, C. G. (2005). An atlas of the inferior colliculus of the gerbil in three dimensions. *Hearing research*, 206(1):12–27.
- Carruthers, I. M., Natan, R. G., and Geffen, M. N. (2013). Encoding of ultrasonic vocalizations in the auditory cortex. *Journal of neurophysiology*, 109(7):1912–1927.
- Casseday, J., Fremouw, T., and E., C. (2002). The inferior colliculus: a hub for the central auditory system. In Oertel D., Fay R.R., P. A., editor, *Integrative Functions in the Mammalian Auditory Pathway*, volume 15 of *Handbook of Auditory Research*, pages 238–318. Springer New York.
- Cohen, L. (1995). *Time-Frequency Analysis*. Prentice Hall.
- Dau, T., Ewert, S., and Oxenham, A. J. (2009). Auditory stream formation affects comodulation masking release retroactively). *The Journal of the Acoustical Society of America*, 125(4):2182–2188.
- Dau, T., Piechowiak, T., and Ewert, S. D. (2013). Modeling within-and across-channel processes in comodulation masking release. *The Journal of the Acoustical Society of America*, 133(1):350–364.
- Davis, K. A. (2002). Evidence of a functionally segregated pathway from dorsal cochlear nucleus to inferior colliculus. *Journal of neurophysiology*, 87(4):1824–1835.
- De Boer, E. (1975). Synthetic whole-nerve action potentials for the cat. *The Journal of the Acoustical Society of America*, 58(5):1030–1045.
- De Boer, E. and Kuyper, P. (1968). Triggered correlation. *Biomedical Engineering, IEEE Transactions on*, (3):169–179.
- Dean, I., Harper, N. S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature neuroscience*, 8(12):1684–1689.
- Dean, I., Robinson, B. L., Harper, N. S., and McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. *The Journal of Neuroscience*, 28(25):6430–6438.
- Dodd, T. and Harrison, R. (2002). A new solution to volterra series estimation. In *Proceedings of the 15th IFAC World Congress on Automatic Control, Barcelona, Spain*.
- Eggermont, J., Johannesma, P., and Aertsen, A. (1983). Reverse-correlation methods in auditory research. *Quarterly reviews of biophysics*, 16(03):341–414.
- Eggermont, J. J. (1993). Wiener and volterra analyses applied to the auditory system. *Hearing research*, 66(2):177–201.
- Eggermont, J. J. (2014). Animal models of auditory temporal processing. *International Journal of Psychophysiology*, (0):-.

- Egorova, M. and Ehret, G. (2008). Tonotopy and inhibition in the midbrain inferior colliculus shape spectral resolution of sounds in neural critical bands. *European Journal of Neuroscience*, 28(4):675–692.
- Egorova, M., Ehret, G., Vartanian, I., and Esser, K.-H. (2001). Frequency response areas of neurons in the mouse inferior colliculus. i. threshold and tuning characteristics. *Experimental Brain Research*, 140(2):145–161.
- Endres, D. and Oram, M. (2010). Feature extraction from spike trains with bayesian binning: latency is where the signal starts. *Journal of computational neuroscience*, 29(1-2):149–169.
- Endres, D., Oram, M., Schindelin, J., and Foldiak, P. (2008). Bayesian binning beats approximate alternatives: estimating peri-stimulus time histograms. In *Advances in neural information processing systems*, pages 393–400.
- Escabi, M. A., Miller, L. M., Read, H. L., and Schreiner, C. E. (2003). Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *The Journal of neuroscience*, 23(37):11489–11504.
- Escabi, M. A. and Schreiner, C. E. (2002). Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *The Journal of neuroscience*, 22(10):4114–4131.
- Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26):429–441.
- Gill, P., Zhang, J., Woolley, S. M., Fremouw, T., and Theunissen, F. E. (2006). Sound representation methods for spectro-temporal receptive field estimation. *Journal of computational neuroscience*, 21(1):5–20.
- Gröchenig, K. (2001). *Foundations of time-frequency analysis*. Springer.
- Grose, J. H. and Hall, J. W. (1989). Comodulation masking release using sam tonal complex maskers: Effects of modulation depth and signal position. *The Journal of the Acoustical Society of America*, 85(3):1276–1284.
- Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). Detection in noise by spectro-temporal pattern analysis. *The Journal of the Acoustical Society of America*, 76(1):50–56.
- Hall III, J. W., Grose, J. H., and Haggard, M. P. (1990). Effects of flanking band proximity, number, and modulation pattern on comodulation masking release. *The Journal of the Acoustical Society of America*, 87(1):269–283.
- Hernández, O., Espinosa, N., Pérez-González, D., and Malmierca, M. (2005). The inferior colliculus of the rat: a quantitative analysis of monaural frequency response areas. *Neuroscience*, 132(1):203–217.
- Hirsch, M. W. and Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*. Academic Press.

- Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. *Acta Acustica United with Acustica*, 88(3):433–442.
- Hung, G. and Stark, L. (1977). The kernel identification method (1910-1977) review of theory, calculation, application, and interpretation. *Mathematical Biosciences*, 37(3–4):135–190.
- Johannesma, P. (1980). Functional identification of auditory neurons based on stimulus-event correlation. *Psychophysical, physiological and behavioural studies in hearing*, pages 77–84.
- Joris, P., Schreiner, C., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, 84(2):541–577.
- Kistler, W., Gerstner, W., and Hemmen, J. (1997). Reduction of the hodgkin-huxley equations to a single-variable threshold model. *Neural Computation*, 9(5):1015–1045.
- Klein, D. J., Depireux, D. A., Simon, J. Z., and Shamma, S. A. (2000). Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *Journal of computational neuroscience*, 9(1):85–111.
- Klink, K. B., Dierker, H., Beutelmann, R., and Klump, G. M. (2010). Comodulation masking release determined in the mouse (*mus musculus*) using a flanking-band paradigm. *Journal of the Association for Research in Otolaryngology*, 11(1):79–88.
- Klump, G. and Langemann, U. (1995). Comodulation masking release in a songbird. *Hearing Research*, 87(1–2):157–164.
- Kouh, M. and Sharpee, T. O. (2009). Estimating linear-nonlinear models using renyi divergences. *Network*, 20(2):49–68.
- Krishna, B. S. and Semple, M. N. (2000). Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *Journal of Neurophysiology*, 84(1):255–273.
- Las, L., Stern, E. A., and Nelken, I. (2005). Representation of tone in fluctuating maskers in the ascending auditory system. *The Journal of Neuroscience*, 25(6):1503–1513.
- Le Beau, F. E., Rees, A., and Malmierca, M. S. (1996). Contribution of gaba- and glycine-mediated inhibition to the monaural temporal response properties of neurons in the inferior colliculus. *Journal of Neurophysiology*, 75(2):902–919.
- LeBeau, F. E. N., Malmierca, M. S., and Rees, A. (2001). Iontophoresis in vivo demonstrates a key role for gabaa and glycinergic inhibition in shaping frequency response areas in the inferior colliculus of guinea pig. *The Journal of Neuroscience*, 21(18):7303–7312.
- Lesica, N. A. and Grothe, B. (2008). Dynamic spectrotemporal feature selectivity in the auditory midbrain. *The Journal of Neuroscience*, 28(21):5412–5421.

- Lewis, E. R. and van Dijk, P. (2004). New variations on the derivation of spectro-temporal receptive fields for primary auditory afferent axons. *Hearing research*, 189(1):120–136.
- Lopez-Poveda, E. A. and Meddis, R. (2001). A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, 110(6):3107–3118.
- Machens, C. K., Wehr, M. S., and Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *The Journal of neuroscience*, 24(5):1089–1100.
- Malmierca, M., Saint Marie, R., Merchan, M., and Oliver, D. (2005). Laminar inputs from dorsal cochlear nucleus and ventral cochlear nucleus to the central nucleus of the inferior colliculus: Two patterns of convergence. *Neuroscience*, 136(3):883–894.
- Marmarelis, P. Z. and Marmarelis, V. Z. (1978). The white-noise method in system identification. In *Analysis of physiological systems*, pages 131–180. Springer.
- Meyer, A. F., Diepenbrock, J.-P., Happel, M. F. K., Ohl, F. W., and Anemüller, J. (2014). Discriminative learning of receptive fields from responses to non-gaussian stimulus ensembles. *PLoS ONE*, 9(4):e93062–.
- Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (1990). Across-channel masking and comodulation masking release. *The Journal of the Acoustical Society of America*, 87(4):1683–1694.
- Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, pages 370–384.
- Nelken, I., Rotman, Y., and Yosef, O. B. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*, 397(6715):154–157.
- Neuert, V., Verhey, J. L., and Winter, I. M. (2004). Responses of dorsal cochlear nucleus neurons to signals in the presence of modulated maskers. *The Journal of Neuroscience*, 24(25):5789–5797.
- Okun, M., Naim, A., and Lampl, I. (2010). The subthreshold relation between cortical local field potential and neuronal firing unveiled by intracellular recordings in awake rats. *The Journal of Neuroscience*, 30(12):4440–4448.
- Oliver, D. and Huerta, M. (1992). Inferior and superior colliculi. In Webster D.B, Popper A.N., F. R., editor, *The Mammalian Auditory System: Neuroanatomy*, pages 168–221. Springer New York.
- Palm, G. (1979). On representation and approximation of nonlinear systems part ii: Discrete time. *Biological Cybernetics*, 34:49–52.
- Palmer, A. R., Shackleton, T. M., Sumner, C. J., Zobay, O., and Rees, A. (2013). Classification of frequency response areas in the inferior colliculus reveals continua not discrete classes. *The Journal of physiology*, 591(16):4003–4025.

- Piechowiak, T., Ewert, S. D., and Dau, T. (2007). Modeling comodulation masking release using an equalization-cancellation mechanism. *The Journal of the Acoustical Society of America*, 121(4):2111–2126.
- Pienkowski, M., Shaw, G., and Eggermont, J. J. (2009). Wiener-volterra characterization of neurons in primary auditory cortex using poisson-distributed impulse train inputs. *Journal of neurophysiology*, 101(6):3031–3041.
- Pollak, G. D., Xie, R., Gittelman, J. X., Andoni, S., and Li, N. (2011). The dominance of inhibition in the inferior colliculus. *Hearing Research*, 274(1–2):27–39.
- Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (2001). Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. *The Journal of Neuroscience*, 21(16):6377–6386.
- Rabinowitz, N. C., Willmore, B. D. B., King, A. J., and Schnupp, J. W. H. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biol*, 11(11):e1001710–.
- Ramachandran, R., Davis, K. A., and May, B. J. (1999). Single-unit responses in the inferior colliculus of decerebrate cats i. classification based on frequency response maps. *Journal of Neurophysiology*, 82(1):152–163.
- Recio-Spinoso, A. and van Dijk, P. (2006). Analysis of responses to noise in the ventral cochlear nucleus using wiener kernels. *Hearing research*, 216:7–18.
- Schooneveldt, G. P. and Moore, B. C. J. (1987). Comodulation masking release (cmr): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band. *The Journal of the Acoustical Society of America*, 82(6):1944–1956.
- Schreiber, S., Fellous, J., Whitmer, D., Tiesinga, P., and Sejnowski, T. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52-54:925 – 931. Computational Neuroscience: Trends in Research 2003.
- Semple, M. and Aitkin, L. (1980). Physiology of pathway from dorsal cochlear nucleus to inferior colliculus revealed by electrical and auditory stimulation. *Experimental Brain Research*, 41(1):19–28.
- Sharpee, T., Rust, N. C., and Bialek, W. (2004). Analyzing neural responses to natural signals: maximally informative dimensions. *Neural computation*, 16(2):223–250.
- Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience*, 24(46):10440–10453.
- van Hemmen, J. L., Kistler, W. M., and Thomas, E. G. (2000). Calculation of volterra kernels for solutions of nonlinear differential equations. *SIAM Journal on Applied Mathematics*, pages 1–21.



- Verhey, J., Epp, B., Stasiak, A., and Winter, I. (2013). Can comodulation masking release occur when frequency changes could promote perceptual segregation of the on-frequency and flanking bands? In Moore, B. C. J., Patterson, R. D., Winter, I. M., Carlyon, R. P., and Gockel, H. E., editors, *Basic Aspects of Hearing*, volume 787 of *Advances in Experimental Medicine and Biology*, pages 475–482. Springer New York.
- Verhey, J. L., Dau, T., and Kollmeier, B. (1999). Within-channel cues in comodulation masking release (cmr): Experiments and model predictions using a modulation-filterbank model. *The Journal of the Acoustical Society of America*, 106(5):2733–2745.
- Verhey, J. L., Pressnitzer, D., and Winter, I. M. (2003). The psychophysics and physiology of comodulation masking release. *Experimental Brain Research*, 153(4):405–417–.
- Volterra, V. (1930). *Theory of functionals and of integral and integro-differential equations*. Blackie, London. reprinted by Dover 2005.
- Wiener, N. (1966). Nonlinear problems in random theory. *Nonlinear Problems in Random Theory*, by Norbert Wiener, pp. 142. ISBN 0-262-73012-X. Cambridge, Massachusetts, USA: The MIT Press, August 1966.(Paper), 1:–.
- Wigner, E. (1932). On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40(5):749.
- Wu, M. C.-K., David, S. V., and Gallant, J. L. (2006). Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.*, 29:477–505.
- Young, E. and Sachs, M. B. (1973). Recovery from sound exposure in auditory nerve fibers. *The Journal of the Acoustical Society of America*, 54(6):1535–1543.
- Young, E. D., Yu, J. J., and Reiss, L. A. (2005). Non-linearities and the representation of auditory spectra. *International review of neurobiology*, 70:136.
- Zandieh, S., Hopf, R., Redl, H., and Schlag, M. (2003). The effect of ketamine/xylazine anesthesia on sensory and motor evoked potentials in the rat. *Spinal Cord*, 41(1):16–22.

Jan-Philipp Diepenbrock  
Birkenweg 21 A  
27777 Ganderkesee

## **Erklärung**

Hiermit erkläre ich, dass ich die von mir eingereichte Dissertation zu dem Thema

Comodulation Masking Release at the level of the Inferior Colliculus in the context of  
Spectro-Temporal Receptive Fields and their corresponding Volterra Operators

selbstständig verfasst, nicht schon als Dissertation verwendet habe und die benutzten Hilfsmittel und Quellen vollständig angegeben wurden.

Weiterhin erkläre ich, dass ich weder diese noch eine andere Arbeit zur Erlangung des akademischen Grades doctor rerum naturalium (Dr. rer. nat.) an einer anderen Einrichtung eingereicht habe.

Ganderkesee,