Impact of Methods on Sensory Gating Indices

Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von David Schubring

an der

Universität Konstanz

Mathematisch-Naturwissenschaftliche Sektion

Fachbereich Psychologie

Konstanz, 2017
Tag der mündlichen Prüfung: 25.10.2017

1. Referentin: Prof. Dr. Brigitte Rockstroh
2. Referent: Prof. Dr. Christian Wienbruch
Summary

The present dissertation addresses sensory gating in schizophrenia and the impact different methodological approaches can have on the assessment of it and related connectivity measures. The auditory P50/M50 event-related brain potential/field is one of the most prominent parameters subject to sensory gating. A classical paradigm for its measurement consists of two identical click sounds (S1 & S2) presented 500 ms apart (paired-click protocol). The P50/M50 following the second click S2 is typically partially suppressed compared to the first, which is denominated as gating. Schizophrenia patients have less gating, which is interpreted as a lack of filtering of irrelevant information. The sensory gating deficit has been proposed as an endophenotype for schizophrenia and a link between deficient frontal control with sensory gating has been suggested. However, used methods and associated effect sizes vary considerably across studies using first-admission and/or using chronic patients. Both as a potential endophenotype and as a node in a potential brain network, the sensory gating deficit should be able to be reliably quantified.

Pilot study

The first step of this dissertation was to investigate the consistency of abnormal sensory gating in chronic schizophrenia patients (CHR) compared to healthy controls (HC) across methods in a pilot study. Magnetoencephalography (MEG) was measured in 58 CHR and 28 HC during a paired-click protocol. Sensory gating was evaluated in a large number of common preprocessing and quantification methods through all possible combinations. As preprocessing methods different strategies for noise correction, trial exclusion, artifact correction and filter
settings were compared. As quantification methods, different strategies for source projection, peak identification, peak scoring and calculating a sensory gating index were compared. The contrast between HC and CHR was used as a point of comparison. Depending on the combination of preprocessing methods, results showed either no HC vs CHR contrast at all, a tendency of a HC vs CHR contrast or significant HC vs CHR contrasts. Following the law of parsimony, only the minimum amount of preprocessing necessary to produce a HC vs CHR contrast was used in the later studies.

Study 1

Study 1 evaluated the impact of several quantification methods with an additional patient group of 35 first-admission (FA) schizophrenia patients. This was done to both estimate the consistency of abnormal sensory gating across stage of disorder and to have an patient group independent from the comparisons of the pilot study. Sensory gating was quantified on sensor and source levels as a ratio \( S2 / S1 \) and as a S1-minus-S2 difference, with M50 amplitude scored relative to baseline and relative to M100 and to M40. Independent of quantification method, patients showed less sensory gating than HC, with medium-to-large effect sizes, without differences between FA and CHR. Results indicate that the frequently reported sensory gating deficit in schizophrenia is robust to variations in quantification methods and stage of disorder.

Study 2

Study 2 evaluated connectivity dynamics within auditory – frontal cortex networks in a sensory gating context and related them to the robust sensory gating measures from study 1.
Background: Connectivity measures delineate the communication within and between neuronal cell assemblies that constitute large-scale default, functional, or salience networks in service of perceptual and cognitive function. Correspondingly, perceptual and cognitive dysfunction in schizophrenia has been related to dysfunctional communication within and between neuronal assemblies, hence, dysfunctional large-scale default and functional networks. Different connectivity measures verified dysfunctional connectivity, while dysfunctional communication by direction of information flow (top-down or bottom-up) remains to be specified. Here, we used Granger causality analysis of MEG data to delineate communication within a functional auditory-frontal network and test the hypothesis that the common auditory sensory gating deficit in schizophrenia patients is associated with altered communication within this network.

Method: Auditory sensory gating, indexed by MEG M50, was assessed in 72 healthy participants (H) and 56 schizophrenia patients (SZ) in a binaural auditory paired-stimulus task. Sources of M50 S1-S2 differences were localized in bilateral auditory (Heschl’s gyri) and bilateral frontal (mid-cingulate) regions. Information flow, indexed by 7-30 Hz oscillations, was analyzed within this network prior to and in response to the stimulus onset of paired clicks using Granger causality algorithms with either auditory or frontal sources as predictor of activity in the respective other source.

Results: H, but not SZ, exhibited pre-stimulus information flow (at 19 Hz) from right auditory to right frontal source, which varied with post-stimulus auditory gating. No significant frontal-to-auditory information flow and relationship with M50-gating was found in any group.

Conclusion: Results indicate bottom-up rather than top-down information flow mediating auditory sensory gating in H. Stimulus-unrelated (pre-stimulus)
information flow may reflect a preparatory process for the top-down modulation after stimulus onset. The lack of systematic pre-stimulus communication within an ‘auditory-frontal’ auditory-gating network might explain dysfunctional sensory gating in SZ.
Zusammenfassung


Pilotstudie

Der erste Schritt dieser Dissertation war die Beständigkeit von abnormem sensorischen Gating über verschiedene Methoden hinweg bei chronischen Schizophreniepatienten (CHR) verglichen mit gesunden Kontrollpersonen (HC) in einer Pilotstudie zu messen. Dabei wurde ein MEG von 58 CHR und 28 HC während einer Stimulation mit gepaarten Klicks aufgezeichnet. Sensorisches Gating

Studie 1

hin, dass die häufig berichteten Defizite im Sensorischen Gating robust sind gegen Variationen in Quantifizierungsmethoden und über Stadien der Krankheit hinweg.

Studie 2


Methode: Auditorisches Sensorisches Gating, indiziert von MEG M50, wurde in 72 gesunden Probanden (H) und 56 Schizophreniepatienten (SZ) in einer binauralen auditorischen Stimulation mit gepaarten Klicks gemessen. Die
Quellen der M50 S1-S2 Differenzen wurden in bilateralen auditorischen (Heeschl’sche Gyri) und bilateralen frontalen (mittleres Cingulum) Regionen lokализiert. Informationsfluss, indiziert von 7-30 Hz Oszillationen, wurde mithilfe von Granger-Kausalitäts Algorithmen innerhalb dieses Netzwerkes analysiert, sowohl vor als auch nach dem Stimulusbeginn von gepaarten Klicks. Dabei wurden entweder die auditorischen oder die frontalen Quellen als Prädiktor der Aktivität in den jeweils anderen Quellen verwendet.


Table of Contents

Summary ............................................................................................................................ I
Zusammenfassung ................................................................................................................ V
Table of Contents ............................................................................................................. IX
Related Publications .......................................................................................................... XI
List of Contributions ......................................................................................................... XIII
Danksagungen ...................................................................................................................... XV
Abbreviations ..................................................................................................................... XVII
1 Introduction ................................................................................................................... 1
  1.1 Overview .................................................................................................................... 1
  1.2 Sensory gating ........................................................................................................... 3
  1.3 Sensory gating and replication challenges ............................................................... 5
  1.4 Pilot Study: Comparison of Preprocessing and Quantification Paths ...................... 10
    1.4.1 Raw data ............................................................................................................. 10
    1.4.2 Noise correction ................................................................................................. 11
    1.4.3 Trial exclusion .................................................................................................... 12
    1.4.4 Artifact correction .............................................................................................. 14
    1.4.5 Filter .................................................................................................................. 16
    1.4.6 Sensor and source Space .................................................................................... 19
    1.4.7 Peak identification .............................................................................................. 22
    1.4.8 Peak scoring ....................................................................................................... 24
    1.4.9 Sensory gating index .......................................................................................... 25
    1.4.10 Summary .......................................................................................................... 25
  1.5 Granger causality ...................................................................................................... 28
  1.6 Outline and research aims of the present dissertation ............................................. 32
2 Study 1: Consistency of Abnormal Sensory Gating in First-Admission and Chronic Schizophrenia across Quantification Methods ...................................................... 33
  2.1 Introduction ................................................................................................................ 33
  2.1 Method ....................................................................................................................... 37
    2.2.1 Participants ......................................................................................................... 37
    2.2.2 Data collection .................................................................................................... 41
    2.2.3 Sensor-space analysis ....................................................................................... 42
    2.2.4 Source-space analysis ....................................................................................... 44
    2.2.5 Gating quantification ....................................................................................... 46
    2.2.6 Gating comparison ............................................................................................ 47
  2.3 Results ....................................................................................................................... 49
2.4 Discussion ................................................................. 53
2.5 Supplements ............................................................ 59

3 Study 2: Connectivity Dynamics within Auditory – Frontal Cortex Networks in Sensory Gating Context .................................................. 72
3.1 Introduction ........................................................................................................ 72
3.2 Method ............................................................................................................... 75
  3.2.1 Participants .......................................................... 75
  3.2.2 Data assessment .......................................................... 77
  3.2.3 Source reconstruction .................................................... 78
  3.2.4 Auditory sensory gating ..................................................... 79
  3.2.4 Directional connectivity ....................................................... 80
  3.2.4 Inferential statistics .......................................................... 81
3.3 Results .............................................................................................................. 82
3.4 Discussion ......................................................................................................... 87

4 General Discussion .............................................................................................. 93
4.1 Summary .......................................................................................................... 93
4.2 Methodological recommendations to study sensory gating in schizophrenia ........................................................ 95
4.3 Sensory gating as an endophenotype for schizophrenia.............................. 98
4.4 Limitations and implications for further research ........................................ 101
4.5 Conclusion ....................................................................................................... 104

List of Tables ......................................................................................................... 105
List of Figures ....................................................................................................... 106
References ........................................................................................................... 107
Related Publications

Parts of the data and results of this dissertation have been presented elsewhere.

Publications


Conference contributions


List of Contributions

The present thesis is part of my (DS) work as a doctoral student in bigger research projects funded by the German Research Foundation (Ro805/14-2) and the Illenauer Foundation (Christian Roller Award 2012), supervised by Prof. Dr. Brigitte Rockstroh (BR), Prof. Dr. Gregory A. Miller (GM) and Tzvetan Popov (TP). Almut Carolus (AC) and Petia Popova (PP), who worked also as doctoral students in the same research projects, contributed to parts of the study series. Additionally, Prof. Dr. Christian Wienbruch (CW) contributed by giving general scientific and methodological advice. In the following, contributions of BR, GM, TP, AC, PP, CW and DS are detailed based on the criteria suggested by the International Committee of Medical Journal Editors (ICMJE; http://www.icmje.org/).

BR, GM and TP developed the first study ideas (Chapter 2-3). AC, PP and DS conducted the studies (Chapter 2-3), including participants recruitment and data collection under supervision of BR and TP. MEG data analyses (Chapter 1.5, 2 & 3) were conducted by DS with important intellectual input from BR, GM and TP. For the pilot study (Chapter 1.5), CW also contributed with important methodological advice. DS prepared the manuscript draft for Study 1 (Chapter 2) with important intellectual input and assistance in revision from BR, TP and GM. DS prepared the manuscript draft for Study 2 (Chapter 3) with important intellectual input and assistance in revision from BR and TP. All authors approved the final manuscripts.

Some of the studies of the research projects were only briefly mentioned but not entirely included in this dissertation. The first of them is published as following: Carolus, A. M., Schubring, D., Popov, T. G., Popova, P., Miller, G. A., & Rockstroh, B. S. (2014). Functional cognitive and cortical abnormalities in chronic

For this study, BR, GM and TP developed the first study ideas. AC, PP and DS conducted the studies, including participant recruitment and data collection under supervision of BR and TP. Cognitive tests were analyzed by AC, MEG data was analyzed by AC and DS. AC prepared the first draft with important intellectual input and assistance in revision from BR, TP, GM and DS. DS also prepared the MEG figures of the draft. All authors approved the final manuscripts.


For this study, BR, GM and TP developed the first study ideas. AC, PP and DS conducted the studies, including participant recruitment and data collection under supervision of BR and TP. Data analyzes were conducted by TP. TP prepared the manuscript draft with important intellectual input from BR, TP, GM and DS. All authors approved the final manuscripts.
Danksagungen

Diese Arbeit wäre ohne die Hilfe und Unterstützung von zahlreichen Personen nicht möglich gewesen. Daher möchte ich mich bedanken:

Bei Brigitte Rockstroh für die Unterstützung und Betreuung in allen Phasen meiner Doktorarbeit.

Bei allen, die zum Gelingen der Studie und der Publikationen mit Rat und Tat beigetragen haben, insbesondere Almut Carolus, Tzvetan Popov, Greg Miller, Christian Wienbruch und Ursula Lommen.

Bei allen, die bei der Datenerhebung und -Auswertung geholfen haben, insbesondere Petia Popova, Johanna Kienle, Vanessa Hirt, Matthias Rack und Marion Reith.

Bei allen anderen meiner Arbeitsgruppe, die auf verschiedenste Weise beigetreten haben, insbesondere Thomas Kustermann, Britta Westner, Johanna Fiess, Vanessa Hirt und Dagmar Moret.


Bei allen Teilnehmerinnen und Teilnehmern der Studie für die Mitarbeit.

Für das Begutachten dieser Arbeit bei Christian Wienbruch und Harald Schupp.
Bei allen, die mich mit Diskussionen zur und Ablenkung von der Doktorarbeit unterstützt haben, insbesondere Johannes Doerflinger und Helge Giese.

Bei meinen Eltern Dorothea und Andreas Schubring für die Unterstützung auf meinem akademischen und privaten Weg.

Bei Laura König für Kommentare, Diskussionen und Korrekturen der Doktorarbeit und Unterstützung von Anfang bis Ende.
Abbreviations

a.u.  Arbitrary Units
ANOVA Analysis of variance
CA Component analysis
CHR Chronic schizophrenia patients
EEG Electroencephalography
FA First admission schizophrenia patients
FIR Finite Impulse Response
fT Femtotesla
g Hedge’s g effect size
GC Granger Causality
HC Healthy Controls
Hz Hertz
ICA Independent component analysis
IIR Infinite Impulse Response
ISI Inter-Stimulus-Interval
ITI Inter-Trial-Interval
M Mean
M100 Auditory MEG component 100ms after sound onset
M30 Auditory MEG component 30ms after sound onset
M40 Auditory MEG component 40ms after sound onset
M50 Auditory MEG component 50ms after sound onset
MEG Magnetoencephalography
ms Milliseconds
N100/N1 Negative auditory EEG component 100ms after sound onset
ns Not significant when p > .1
P50 Positive auditory EEG component 50ms after sound onset
PCA Principal component analysis
s Seconds
S1 First stimulus/click in the double click paradigm
S2 Second stimulus/click 500ms after S1 in the double click paradigm
<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>sMRI</td>
<td>Structural magnetic resonance images</td>
</tr>
<tr>
<td>SZ</td>
<td>Schizophrenia patients</td>
</tr>
<tr>
<td>T</td>
<td>Tesla</td>
</tr>
<tr>
<td>#</td>
<td>p &lt; .1</td>
</tr>
<tr>
<td>*</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>**</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>***</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Overview

Sensory gating is a widely studied phenomenon especially in schizophrenia patients since the first studies by Adler and colleagues (1982). It is commonly studied in a paired click paradigm with pairs of auditory click noises 500ms apart which are repeatedly presented. Typically, the neurophysiological response 50ms after the first click (EEG P50 or MEG M50) is stronger than after the second, which is interpreted as a way of filtering irrelevant information by the brain. This filtering is disturbed in schizophrenia patients and is thought to contribute to the flooding of the brain with an information overload (Hetrick, Erickson, & Smith, 2012; Venables, 1964). The sensory gating deficit has repeatedly been described as one of the most prominent, robust and reliable neuropsychological measures in schizophrenia in overview reviews and meta-analyses (e.g. Bramon, Rabe-Hesketh, Sham, Murray, & Frangou, 2004; Heinrichs, 2004; Rojas, 2014). As such, it has been related to a variety of schizophrenia characteristics, most notably the heredity: The deficit was found at early stages of illness, in subjects at risk and in family members (Ahveninen et al., 2006; Reig et al., 2011; Yee et al., 2010; Yee, Nuechterlein, Morris, & White, 1998) and has thus been suggested as an endophenotype of schizophrenia (e.g. Miller & Rockstroh, 2013).

Yet, other reviews of the literature have found considerable variability in effect sizes, including some null findings. Two different meta-analysis concluded that the variability of methods across studies and also the wide range of effect sizes between studies from different laboratories made the claim of sensory gating being an endophenotype for schizophrenia only viable for studies from certain
laboratories (de Wilde, Bour, Dingemans, Koelman, & Linszen, 2007; Patterson et al., 2008). Another early review of the literature even found a steadily declining amplitude of the studied event related potential P50 over the years, which might be explained by additional preprocessing methods (e.g. more advanced artifact removal or filtering), a bias in rating the potentials by hand or publication bias (Light & Braff, 1998).

Therefore, the first part of the present dissertation compared different methodological approaches to study the sensory gating phenomenon in schizophrenia patients set to evaluate the robustness of the sensory gating phenomenon across methods in a single, large dataset. This was done in two steps: First, a variety of common preprocessing methods were compared against each other to evaluate which show the a priori contrast of chronic schizophrenia patients versus healthy controls at all and to which extent. This was done as a pilot study, as it mainly examined the best way to get “clean data”, i.e. data that is artifact free or data in which artifacts are at least reduced as good as possible, for the specific test setting and to get a standard of comparison for the following approaches (benchmark). In a second step, the data was used as a means of comparing chronic and first admission schizophrenia patients in relation to healthy controls. This comparison evaluated the quantification methods, which should be independent of specific laboratory settings. In this step a variety of common sensory gating quantifications were compared against each other and the robustness of the sensory gating deficit was evaluated across patient groups via convergence (or non-convergence) of quantification methods (Study 1).

Once a robust measure has been established, the second part of the present dissertation examined the interplay of frontal and auditory networks which might contribute to sensory gating via frontal control. This was assessed similarly
to the method comparison of the first part: In a first step, the robustness of various candidate network measures was assessed, and, if successful, the influence of frontal control on sensory gating was analyzed together with a possible pathology of it in schizophrenia.

1.2 Sensory gating

Sensory gating is the filtering of both excess and trivial or redundant information, so that irrelevant stimuli are “gated out” of awareness to facilitate information processing of more relevant stimuli (Braff & Geyer, 1990). Accordingly, it has been studied either with excess information e.g. via habituation or prepulse inhibition (PPI) of a startle response or with redundant information e.g. via a double click paradigm.

The paradigms for PPI typically consist of a weak sensory event or prepulse presented repeatedly a short time (typically 30-500ms) before a strong, startle-inducing stimulus. After a few trials the startle response is typically smaller when preceded by a prepulse and that reduction is also less pronounced in schizophrenia patients (for a review see e.g. Braff, Geyer, & Swerdlow, 2001). However, a startle response is impractical to study in a MEG setting to analyze brain responses and connectivity, as muscle movements from startle responses are detrimental to MEG data quality.

More suitable for a MEG recording is the paired click paradigm, which targets the redundant information gating instead of the excess: Typically, two short, identical, white noise click-sounds are presented repeatedly (S1 & S2), separated by a fixed inter-stimulus interval (ISI) and a variable inter-trial interval (ITI). S1 is also referred to as the conditioning stimulus and S2 as the testing stimulus, as
S1 is thought to activate or condition the inhibition and S2 testing the inhibition’s strength (Adler et al., 1982; Freedman et al., 1987). The stimulus sequence used throughout this dissertation is depicted in Figure 1.1, with an ISI of 500ms and an ITI of 8+1 s.

![Stimulus sequence](image)

Figure 1.1: Stimulus sequence used in the present dissertation

Following an auditory event, the brain elicits a typical series of event related components, which are classified into early components, middle latency components and long latency components. The component most analyzed for sensory gating is the one occurring around 50ms after stimulus onset\(^1\), which is classified as P1 or P50 (or M50 with MEG recording) as the first long latency response. It is followed by the biggest, and therefore often easiest to identify, auditory component N1 (or M100) with opposite polarity (Picton, Hillyard, Krausz, & Galambos, 1974, see also Figure 1.1) and preceded by the M40 in the opposite and M30 in the same polarity.

Sensory gating is then reflected in the fact that the response to S2 is smaller than the response to S1. There are a variety of ways how this can be quantified, which is discussed later in the Pilot Study and Study 1.

---

\(^1\) Other components besides M50 are as well discussed to show sensory gating, see study 1 for a discussion on that matter.
1.3 Sensory gating and replication challenges

Sensory gating studies like neuroscience in general suffer from small sample sizes and insufficient statistical power\(^2\). For example, the mean power of all neuroscientific studies included in meta-analysis published in 2011 was 0.21 (Button et al., 2013). One of the meta-analyses evaluated by Button et al. even concerned sensory gating and had a mean statistical power of 0.10, ranging from

\(^2\) The statistical power is the probability of not committing a type II error (\(\beta\)) and is equal to \(1 - \beta\); or in other words the probability of correctly rejecting the null hypothesis when the null hypothesis is indeed false.
0.07 to 0.38 (Chang, Arfken, Sangal, & Boutros, 2011). However, statistical power does not only depend on sample size but also on the estimated effect size. If one is to assume the effect size of $d=0.19$ from the meta-analysis by Chang et al. (2011), the $N=121$ of the present dissertation would equate to a statistical power of 0.56, if one assumes the effect size of $d=1.28$ from the study by de Wilde et al. (2007), a $N=121$ would equate to a statistical power of $>0.99$ (Cohen, 1977). In any case, the number of subjects in the present dissertation is higher than in most other studies, which at least mitigates the problem of low power.

Another problem is the “researcher degrees of freedom”: There is a wide variety of ways sensory gating is being studied and even more ways to analyze neurophysiological data in general. As Simmons, Nelson, & Simonsohn (2011) pointed out, (undisclosed) flexibility in data analysis allows presenting almost anything as significant and it is often more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. Moreover, when anonymously surveying over 2000 psychologists, almost two third admitted to only selectively report measures for their studies (John, Loewenstein, & Prelec, 2012). This is probably at least one of the reasons why, depending on the metric, only 36% to 68% of 100 studies were successfully reproduced in a large reproducibility project (Open Science Collaboration, 2015). For the problem of flexibility of data collection and analysis, a full disclosure of all used methods and a comparison across these is presented. In short, data collection is restrained to one dataset and data analysis is explored in a wide variety of ways:

---

3 It examined the influence of the first and the second response to sensory gating indices and included 56 studies with a mean of 55 subjects per study (ranging from 20-309) and a Cohen’s $d=0.19$. This is discussed later in study 1.

4 And alpha error probability, which was held constant at 0.05 for two tailed tests in this example.
While the exact parameters of the double click paradigm (click intensity, click length, inter-trial interval, inter-stimulus interval, stimulus rise times) are also a source of variation, they were held constant across this dissertation to make a comparison within one dataset possible (Figure 1.1.1).

Another source of variation is whether the measurement is via EEG or MEG. To restrict analysis to one dataset, only MEG data are evaluated in this dissertation, as source-projected M50 (MEG) data distinguished patients and controls better than scalp EEG P50 (Edgar et al., 2003), with higher test-retest reliability (Lu et al., 2007) and better resolution for the P50/M50 generating Heschl's gyrus (Edgar et al., 2003).

When it comes to analyze a data set, decisions made during data analysis are dependent on one another which leads to a “garden of forking paths” (Gelman & Loken, 2014): The number of possible combinations of preprocessing steps and quantifying the results is almost endless, as each of the steps alters all that follow so each binary choice of options (e.g. whether to use a mean reference noise correction or an individual one) doubles the total number of possible end points. In the meta-analysis of de Wilde et al. (2007) and Patterson et al. (2008), varying reported methods across the literature included artifact rejection techniques, peak selection and especially filter settings, which ranged from 0.1 Hz high-pass to 30 Hz high-pass. Even if the mean effect sizes of sensory gating deficits in the different meta-analyses were high, e.g. 1.28 (SD=0.72) in the one by de Wilde et al. (2007), they varied considerably from d=0.26 (Ringel, Heidrich, Jacob, Pfuhlmann, Stoebert, & Fallgatter, 2004) to d=3.87 (Olincy et al., 2000) with a 95% confidence interval from -0.13 to 2.69. While de Wilde et al. (2007) found some of the variation in effect sizes explained by the differing methods like e.g. filter settings, detailed preprocessing methods such as exact filter parameters
were often missing completely or being reported only partially. Because of this considerable variation in effect sizes and the often incomplete reporting of preprocessing methods, a pilot study evaluated a number of common preprocessing and quantification methods for most of the possible combinations. This was done to get an estimation of the impact preprocessing choices can have on sensory gating results and to find the best way to both get “clean data” in the present data set while minimizing the distortion introduced by it.
Figure 1.3: Preliminary pipeline of tested methods. Only one path is displayed for the sake of clarity. HC: Healthy controls, CHR: chronic schizophrenia patients. Color indicates if the a priori group contrast was found, red: no contrast, light green: tendency of contrast with borderline significance, dark green: significant contrast. CA: component analysis, ICA: independent CA, PCA: principal CA, FIR: finite impulse response, IIR: infinite impulse response, \( \rightarrow \) forward filter, \(<\) backward filter, \(<>\) forward and backward filter.
1.4  **Pilot Study: Comparison of Preprocessing and Quantification Paths**

The pipeline leading from "raw" MEG data to a final sensory gating score was performed using a combination of three programs: The MATLAB-based open-source signal processing toolbox FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), complemented by in-house MATLAB functions, additional dipole fitting was accomplished via BESA Research 6.0 (MEGIS Software GmbH, Gräfelfing, Germany) and noise correction, trial exclusion and artifact correction was partly performed in PECAT (http://www.pecat.eu). The a priori hypothesis was that a contrast between chronic schizophrenia patients (CHR) and healthy controls (HC) exists and the methods were evaluated in the extent they were able to show that difference. If multiple methods led to similar results, the law of parsimony ("Ockham's razor") was used (Sarris & Reiß, 2005): From two otherwise identical alternatives, the simpler of the two should be used. Simpler in this case means especially less manipulation or interpretation of the original recording, not necessarily simpler in terms of easier to do. The pipeline was divided into the following steps (Figure 1.3):

**1.4.1  Raw data**

"Raw" data in this context means data from HC and CHR from study 1. To avoid a doubling of the sample description and data collection, the full experimental setup is described in chapter 2.2. Most importantly, only CHR and HC were selected to get a “benchmark” for the a priori CHR vs HC contrast which can be compared to the FA as an independent measure at a later step.
1.4.2 Noise correction

When recording MEG data from the brain, environmental noise is always recorded together with the signal of interest. While it is normally generated at larger distances the MEG, it is still many orders of magnitude larger than the signals from the brain (Vrba & Robinson, 2001). Therefore, noise correction is always necessary. It was performed via two strategies: Mean reference and individual reference.

1.4.2.1 Mean reference

During the MEG measurement, environmental noise was recorded by 11 reference magnetometers approximately 1 m from the head and subtracted from the remaining sensors. One path was to just use this automatic subtraction for further analyzes.

1.4.2.2 Individual reference

The automatic subtraction can be refined when looking at individual subjects’ noise levels and, if necessary, apply a second noise correction by hand as implemented in PECAT: In a first step, reference channels were individually filtered to contain only those frequencies which were still present in the data. In a second step, the proportion of correlation between each of the reference sensors and each of the 148 MEG sensors was determined as a weighing matrix. In the final step, each of the reference sensors was subtracted from the 148 MEG sensors relative to the weighing matrix.

1.4.2.3 Noise correction summary

When evaluating all further steps in the preprocessing path, no significant difference was found between the mean and the individual noise correction in
their ability to distinguish HC from CHR as both paths showed significant difference. Thus, following the law of parsimony, mean noise correction seemed to be enough to produce clean data.

### 1.4.3 Trial exclusion

Even in perfect recording conditions, some trials are contaminated with artifacts. For one, participants were recorded with eyes open to prevent them from falling asleep, so blinking was impossible to prevent. Also, other movements of the head can have a considerable impact on MEG signals. One way to deal with such artifacts is trial exclusion, for which four different strategies were compared: No exclusion, automatic exclusion, variance based exclusion and individual inspection.

#### 1.4.3.1 No exclusion

As a naïve base comparison, no trials were excluded.

#### 1.4.3.2 Automatic exclusion

Also, an automatic solution was probed as suggested in the FieldTrip Wiki (http://www.fieldtriptoolbox.org/tutorial/automatic_artifact_rejection). In a first step, MEG channels around the eyes were band-pass filtered from 1 to 15 Hz, z-transformed and then all trials over a certain threshold (4.4) were estimated to contain eye blinks and discarded. In a second step, all MEG channels were band-pass filtered from 110 to 140 Hz, z-transformed and all trials over a certain threshold (8) were considered to contain muscle artifacts and discarded. Remaining trials were unfiltered.
1.4.3.3 Variance based exclusion

A third approach was to individually plot the variance of each trial per participant and to discard outliers by visual examination. These trials were considered to contain movement artifacts and were removed at two times in the preprocessing path: Once, with a higher threshold before analyzing subcomponents via ICA or PCA and once with a lower threshold after removing PCA or ICA components. This was done to both avoid bigger artifacts to distort the component analyses and to correct trials with smaller artifacts via removal of PCA or ICA components without removing them completely. This individual removal of trials lowered the overall variance per trial by 87% (from $6.43 \times 10^{-25} \text{T}^2$ (SD = $4.80 \times 10^{-24} \text{T}^2$) before to $8.55 \times 10^{-26} \text{T}^2$ (SD = $2.30 \times 10^{-26} \text{T}^2$) after removal of artifact-contaminated trials).

1.4.3.4 Individual inspection

A fourth approach was to visually scan the whole data set per person for artifacts and remove them by hand. The guidelines to identify an artifact were based on typical templates of eye blinks and head movements, which were marked in the continuous signal and later removed.

1.4.3.5 Trial exclusion summary

When evaluating all further steps in the preprocessing path, all trial exclusion procedures showed a tendency of a HC vs CHR contrast, although the contrast was considerably more pronounced for both visual removal based on variance and visual removal based on the whole data set. Thus, following the law of parsimony, visual removal based on variance was judged to be the best compromise to get both clean data and not needlessly overanalyzing the data.
1.4.4 Artifact correction

Instead of removing trials with artifacts completely, they can also be corrected by identifying artifacts via component analyses (CA) and then removing just the components with artifacts and leaving the rest intact. This can lead to a better signal to noise ratio than trial exclusion, as more trials can be included in the final data file. For artifact correction, five different strategies and a combination of them were tested: No artifact correction, an independent component analysis (ICA), principal component analysis (PCA), “heart template CA” and “blink template CA”.

1.4.4.1 No artifact correction

First, as a naïve base comparison, no artifact correction method was employed. This led to more trials being rejected for paths containing the variance based exclusion, as this exclusion removes trials once before and once after artifact correction with different thresholds. As no artifact correction is used to lower the variance of contaminated trials and thus enabling them to be included, they are instead completely removed.

1.4.4.2 ICA

An ICA was tested as implemented and recommended by the FieldTrip wiki (www.fieldtriptoolbox.org/project/guidelines/paper/preprocessing), both with the infomax “runica” ICA algorithm (Bell & Sejnowski, 1995) and the “fastICA” algorithm (Hyvärinen & Oja, 2000). As the name suggestes, fastICA is computed considerably faster, while supposedly providing similar accuracy. Resulting component topographies and time courses were visually judged and removed, if they were considered to contain heart artifacts, eye blinks or muscle movements.
1.4.4.3 PCA

A principal component analysis PCA was tested as implemented in PECAT. The resulting components and topographies were judged with the same criteria as ICA components. The PCA analysis is described in greater detail by Reith (2014).

1.4.4.4 “Heart template” and “eye blink” CA

A component analysis based on template selection was probed, in which both templates of eye blinks and heart artifacts were visually selected for each participant and then removed, as implemented in PECAT.

1.4.4.4 Artifact correction summary

When not removing artifacts via component analyses, HC vs CHR showed the tendency of a difference. But the PCA analysis proved to be even worse than that: All combinations containing the PCA (PCA alone, PCA + ICA, PCA + template CA, PCA + template CA + ICA) produced null results without a difference between HC and CHR. This was considered to be due to a low resolution or separation of artifacts, in which components which contained e.g. (part of) the PQRST heart signal and were therefore removed also contained (part of) the M50 component. In contrast, ICA and the combination of ICA with template CA showed a clear HC vs CHR contrast in most of the following preprocessing steps. When comparing the infomax ICA algorithm with the fastICA algorithm, the infomax ICA algorithm produced stronger group differences, but when comparing ICA with and without the combination of template CA, no difference was seen. Thus, only the infomax ICA algorithm was judged to provide better results than no CA at all.
1.4.5 Filter

Event related components often have frequency characteristics distinct from artifacts. Therefore, filter are ubiquitous in EEG and MEG data to filter out noise and thus improving the signal to noise ratio (Widmann, Schröger, & Maess, 2015). Filter settings have varied considerably in the sensory gating literature (de Wilde et al., 2007; Patterson et al., 2008) and have generally been under-specified (Widmann et al., 2015). Present filters were selected to minimize filtering and the distortion it causes (for high-pass: low cutoff and low order; for low-pass: high cutoff) while removing slow drift and high-frequency noise. Several filter settings were compared:

First, several high-pass cutoff frequencies were evaluated based on those commonly appearing in the studies in the meta-analysis of de Wilde et al. (2007): 0.1 Hz, 1 Hz, 5 Hz and 10 Hz. Second, the FieldTrip default setting of using a infinite impulse response filter (IIR) proved to be inappropriate, as the infinitely long impulse response blurs S1 and S2 responses together to a certain extent. Instead, a high-pass finite impulse response (FIR) filter with a maximum filter order of 300 (thus spanning 442 ms on each side of the filtered point) was chosen to minimize blurring of S1 and S2 responses. The filter was applied in forward and reverse directions (and thus with symmetrical weights) both to allow a lower-filter order (since this doubles the effective filter order) and to avoid having a “biased” one-sided blurring or phase shift. Filtering only in one direction with otherwise identical filter characteristics would double the blur in this direction while eliminating it in the other, meaning that a forward filter would blur S2 responses more than S1 responses while a backward filter would blur the baseline. Another
tradeoff for filtering in both directions is that the baseline is distorted by the backward part of the filter, which was compensated in this study by a long 1000ms baseline (Figure 1.4).

![Figure 1.4: ERF of the sensors surrounding the M50 peak on the right hemisphere, averaged over all subjects with different filter settings. Band-pass-filtered from 1-55 Hz with FieldTrip default settings except highpass (hp) filter direction. S1: Onset of the first click, S2: Onset of the second click. M50 and M100: Auditory components.](image)

Third, considerable testing was performed to identify the lowest cutoff that successfully removed baseline drift while minimizing waveform distortion. The 1 Hz cutoff met that criterion for sensor data, although evaluation of its gain function, after application in the forward and reverse directions, indicated that the cutoff was actually 1.7 Hz. Source data needed a higher (5 Hz) cutoff to eliminate baseline drift. Such low-frequency noise is characteristically higher in MEG than in EEG, and S2 is more vulnerable than S1 because S2 is more distant from the
pre-S1 baseline when using the pre-S1 baseline for both responses. A possible confound of cutoff Hz with sensor vs. source analyses was unlikely, because the exploratory work at 5 Hz cutoff with sensor-space data produced results very similar to those at 1 Hz cutoff. Although Kanno, Nakasato, Murayama, & Yoshimoto (2000) noted that a high-pass filter of higher than 3 Hz might artificially increase M50 amplitude (while decreasing M100 amplitude), a different explanation might be possible: Königs & Gutschalk (2012) showed that the M50 time interval partly overlaps with the M100, meaning that the M100 might mask some of the M50 strength. Since the M100 has lower frequency components than the M50, a high-pass filter might not artificially increase the M50 but rather restore it to its original strength, unmasked by the filtered-out M100. This hypothesis remains to be substantiated by additional studies, especially simulations.

However, this was the first preprocessing step that produced different results depending on the further analyses: For sensor data, a 1 Hz highpass filter sufficed to produce a strong HC vs CHR contrast, while for source data a 5 Hz highpass filter was necessary. Also, less filtering was necessary to produce a HC vs CHR contrast when calculating the sensory gating index as a S1 – S2 difference, while a S1/S2 ratio was more baseline-dependent and needed more filtering to produce a HC vs CHR contrast (see below).

Finally, for the low-pass filter settings, differences were much less pronounced. The only relevant setting was the low-pass cutoff, which distorted results when it was lower than 50 Hertz. This was due to the M50, which was beginning to be filtered out at 50 Hertz and was completely removed by a filter cutoff of 30 Hertz. When the low-pass cutoff was above 50 Hertz, no difference was found for the automatic peak detection. However, the visual peak detection was less clear for signals without a low-pass filter, so a 80 Hertz low-pass filter was
1.4.6 Sensor and source Space

Another variation in the sensory gating literature is whether results derive “directly” from sensor recordings or are projected into source space with one of the various methods available. One of the main advantages of source reconstruction is generally to control for different head shapes and positioning within the sensor. Controlling for those two via source projection is considerably different between EEG and MEG, since source reconstruction is better and sensor positioning might be worse with MEG (the position of the sensors might be more controllable when putting an EEG net directly onto the head than when positioning the head inside a MEG helmet). Five different methods were probed here: Mean peak sensor, individual peak sensor, virtual sensors and virtual parcels derived from beamforming and dipole localization via multiple source probe scan (MSPS).

1.4.6.1 Sensor Space.

The simplest solution of just scoring the sensor which showed the strongest M50 when averaging over all subjects did not elicit a significant HC vs CHR contrast, probably because it did not control for individual head shape and position differences. Another possible solution to control for individual differences without source projection was to identify individual peak sensors per subject as follows. First, separately for each hemisphere and each component (M50, M100), a set of adjacent sensors closest to the peak pixel (averaged over all subjects) was identified. Then, for each subject, hemisphere and component, the sensor of
the set with the strongest component (M50, M100) was chosen for scoring. This was done both by visual inspection and by an automated script. For the visual inspection, the judgement was to include the sensor closest to the mean peak sensor with the strongest and (subjectively rated) clearest peak. For the automated scoring, the sensor with the strongest peak was chosen. Also, the automated script considered different set numbers, as 2 – 9 sensors closest to the mean peak were included. All in all, this individual scoring produced robust HC vs CHR contrasts for both visual inspection and automated rating for sensor sets of 3 – 9 sensors. Following the law of parsimony, automated scoring was chosen over individual rating, as subjective rating of a “clear peak” did not provide additional effect. Concerning the set of sensors searched for individual peaks, an exception to the law of parsimony was made: As brain asymmetries especially in the temporal lobe are found repeatedly in schizophrenia (for a review see Crow, 1990), a set of four sensors per hemisphere was chosen for study 1. Three sensors as the minimum amount required to elicit robust group differences and a fourth in shifted in the anterior direction, as the typical anterior shift of the right hemisphere (Yakovlevian torque) of the brain is less pronounced in schizophrenia patients (Bilder et al., 1993; Sommer, Aleman, Ramsey, Bouma, & Kahn, 2001), to not bias the sensor selection for one of the groups.

1.4.6.2 Source Space

Advantages and disadvantages of various source localizations have been discussed extensively elsewhere (e.g. Baillet, Mosher, & Leahy, 2001; Krim & Viberg, 1996; Michel et al., 2004). In short, two approaches were compared here, beamforming and dipole solutions, which have several major differences: While dipole solutions have an a priori assumption of an underlying dipolar model (a small number of dipoles in the brain can adequately model the whole surface...
recording), beamforming does not need an a priori model and does not try to explain the whole measured field. Instead each brain position (or voxel) is estimated as a contribution to the variance (not the strength) of the measured field while suppressing the contribution from all other sources. Moreover, dipole solutions are vulnerable when multiple sources are active at once and beamformer are vulnerable to activity from correlated sources. As both problems are potentially present in this dataset, adjustments were needed: For the dipole solution, this was adjusted by stepwise adding more sources if the initial set of sources did not provide a satisfactory solution. For the beamformer, this was adjusted by analyzing single trials instead of averaged activity, as the substantial trial-by-trial variability has been shown to mitigate problems from correlated sources (Dalal, Sekihara, & Nagarajan, 2006; Sekihara, Nagarajan, Poeppel, & Miyashita, 2001).

The beamforming procedure is described in greater detail in chapter 2.2.4. With this method, three approaches were probed: First, a virtual sensor was placed in the peak voxel with the greatest S1-S2 source strength difference averaged over all subjects. Second, a virtual sensor was placed in the peak voxel per group. Third, the source projection was averaged per brain region based on an automatic anatomical parcellation of brain areas according to the MNI atlas (Tzourio-Mazoyer et al., 2002) and time series of two “virtual parcels” were derived from averaging source-projected data of five adjacent 1 cm³ voxels (per MNI atlas: two for right Heschl’s gyrus, three for left) per hemisphere to establish a “virtual sensor” for each Heschl’s gyrus. For all of those beamformer results a HC vs CHR group difference was found, but it was significantly stronger for the averaged Heschl’s gyrus virtual parcel compared to the virtual sensor. Accordingly, the virtual parcel method was used in study 1.
The dipole fitting used multiple source probe scan (MSPS) as implemented in BESA 6.0. It is described in greater detail by Hirt (2014) and was similar to the approach used by Popov et al. (2011). In short, a least squares fitting algorithm was used to search within the head model for a location where the sources can explain a maximal amount of variance (Scherg & Picton, 1991). In this case, the sources were a pair of regional dipoles which were simultaneously fitted in the left and right hemisphere for a 20ms interval around the S1 M50 peak. This latency was also used to obtain the M50 source strength in response to S2. Similar approaches were used to identify M100 and M40 peaks for peak to peak scoring. As this fitting procedure required visual judgement of peaks and manual placement of dipole seeds, the first approach of one unblinded rater used in the study by Carolus et al. (2014) was later repeated using four blinded raters by Hirt (2014). While the rating of the one unblinded rater produced significant HC vs CHR contrasts, they could not be reproduced by the four blinded raters. This might be due to unclear signals or because rating the dipole fits could only insufficiently be trained. Because of this subjective influence and instability of results depending on the rater, only the beamformer source reconstruction and not the dipole solution was used later in study 1.

1.4.7 Peak identification

To identify the M40, M50 and M100 peaks, three solutions were probed: Mean peak latency, visual inspection and automatic detection.

1.4.7.1 Mean peak latency

The mean peak latency used the ERF from the average over all subjects to estimate the latency of the potentials and scored the potentials with this latency for every participant. No HC vs CHR contrast was found for this method, as the
inter-individual latency differences were probably too high to just consider the mean latency. It is important to note that mean peak latency was only tested with peak scoring relative to a 1000ms long pre-S1 baseline (see 1.4.8), as no interactions with different baselines were expected.

1.4.7.2 Visual inspection

Visual inspection means looking at the waveforms for each subject and scoring the peak latency for S1 and S2 by hand. The guideline was to identify the M100 as the strongest peak after click onset with roughly 100ms latency and then score the preceding peak as M50. This produced a HC vs CHR contrast both for sensor data and beamformer source data, but not for the dipole source data (see 1.4.6.2). It is important to note that visual inspection was only tested with peak scoring relative to a 1000ms long pre-S1 baseline (see 1.4.8), as no interactions with different baselines were expected.

1.4.7.3 Automatic detection

With automatic detection, the visual inspection procedure was modelled by a script: For each subject, the automatic algorithm determined the most negative and the most positive peak within 40-130 ms after S1 and S2 click onset for each hemisphere. As peak polarity does not necessarily have the same meaning in MEG as in EEG, the earlier of the two peaks was defined as M50 and the later as M100. This procedure accounted for potential individual differences in peak polarities and peak locations. For the M40, the algorithm picked M40 and M50 separately: when searching in one latency window for M40 and M50 simultaneously (e.g. searching within 35-60 ms after click onset) either M40 or M50 could not reliably be scored, depending on the searched latency-window. Therefore, 35-50 ms was searched for the most negative and most positive peak and the earlier of the two was defined as M40. As 35-50 ms might only contain the rising
flank of M50 and not necessarily capture the peak of M50, M50 scores were taken from the original scoring (searching 40-130 ms for the most positive and negative peak and defining the earlier of the two as M50). This ensures that the flank of the M100 is not mistaken as M40, as both share the same polarity (for details see chapter 2.2 and supplementary table 2.5A-C in chapter 2.5). It is important to note that automatic detection was not tested with dipole source data derived from BESA, as this algorithm was not easily applicable to the BESA GUI (see 1.4.6.2).

As both visual inspection and automatic detection elicited a significant group difference, the automatic detection was chosen for study 1 to minimize rater influence.

### 1.4.8 Peak scoring

Six different peak scoring methods were compared: M50 relative to a 1000ms pre-S1 baseline, M50 relative to a 200ms pre-S1 baseline, S1 M50 relative to a 200ms pre-S1 and S2 M50 relative to a 200ms pre-S2 baseline, M50 relative to M100, M50 relative to M40 and M100 relative to a 1000ms pre-S1 baseline.

All scorings of M50 relative to the different baselines produced significant HC vs CHR contrasts, although they were strongest for the 1000ms pre-S1 baseline. This was probably due to the specific 1Hz filter discussed above, which, while minimizing blurring of S1 and S2 together, distorted the first 200ms of pre-stimulus baseline by the backward part of the filter (see Figure 1.4).

For the peak-to-peak scorings, M50 relative to M100 produced robust group differences for all paths, but M50 to M40 only produced group differences for sensor data but not for source data. This was probably because the M40 is
easily obscured by the surrounding M30 and M50 components and generally much smaller than the M100.

For the M100 relative to baseline, no group differences were found.

As peak scorings are still a matter of discussion in the sensory gating literature and more generalizable to other settings than the previous preprocessing approaches, all peak scorings were used for study 1.

1.4.9 Sensory gating index

Three different sensory gating indices were compared, the most common ratio (S2/S1), the difference S1–S2 and the single peak comparison, where plain S1 and S2 values were compared between the groups (for a discussion of sensory gating scores see study 1).

Both S2/S1 ratio and S1–S2 difference produced robust group differences while plain S1 and S2 values did not. Still, as the choice of gating index is still discussed like the peak scoring in the sensory gating literature, all three were used for study 1.

1.4.10 Summary

All in all, several paths of combining preprocessing and quantification methods proved to be effective in (re-)producing the a priori HC vs CHR contrast. For the first preprocessing steps (noise correction, trial exclusion, artifact correction and filter) only the simplest solutions which proved effective were chosen for study 1, as these were considered to be the best compromise of getting both clean data and distorting the original recording the least amount possible. For the more advanced quantification methods (sensor/source space, peak identification, peak scoring and sensory gating index) most path were included in study 1
whether they produced a HC vs CHR contrast or not, as these were still subject of ongoing discussion in the sensory gating literature (Figure 1.5, for a discussion of quantification methods see study 1). Also, study 1 included not only healthy controls and chronic schizophrenia patients as in the pilot study but an additional patient group with first admission schizophrenia patients (FA). That way an independent evaluation of the selected methods could be done, as the sensitivity of the selected methods to early stages of the illness could be evaluated and the multiple comparison problem can be avoided for FA: As all selected methods are based on different analyses with different populations (HC vs CHR), the results for FA can serve as external validation.
Figure 1.5: Selected preprocessing and quantification methods for study 1. HC: Healthy controls, CHR: chronic schizophrenia patients, FA: first admission schizophrenia patients. ICA: independent CA, FIR: finite impulse response, <-> forward and backward filter.
1.5 **Granger causality**

Hemodynamic imaging studies found simultaneous activity of both frontal cortex and auditory cortex in a sensory gating context (Grunwald et al. 2003; Tregellas et al. 2007; Tu et al. 2013; Bak et al. 2013). While this suggests a frontal involvement in sensory gating, it does not necessarily involve a causal relationship between the two processes. To test whether the simultaneous activation is due to a connection of the two brain regions, the hypothesis of the simultaneous activation being a causal process can be tested using Granger Causality (GC). GC is a method to study an information flow between two or more signals and can also be used on MEG data (for a review see Friston, Moran, & Seth, 2013). Specifically, using GC, temporal dynamics between brain regions can be studied to infer temporal precedence of events.

Originally, GC was invented for econometric models (Granger, 1969) and awarded with a nobel prize. Clive W. J. Granger himself defined causality as follows:

“[…] We say that \( Y_t \) [time series one] is causing \( X_t \) [time series two] if we are better able to predict \( X_t \) using all available information than if the information apart from \( Y_t \) had been used.” (Granger, 1969, p. 428)

However, he later also criticized the usage of his work and definition in his nobel lecture (2003):

“At that time, I had little idea that so many people had very fixed ideas about causation, but they did agree that my definition was not true causation in their eyes, it was only Granger causation. I would
ask for a definition of true causation, but no one would reply. However, my definition was pragmatic and any applied researcher with two or more time series could apply it, so I got plenty of citations. Of course, many ridiculous papers appeared.”

So what GC and what are potential “ridiculous” pitfalls? Generally, GC is based on autoregression (AR), which is the regression of one variable on itself at a later time point. This is based on the idea that a signal has a partly predictable time course, which works best with stationary signals. In the simplest case of an univariate AR of one signal $X_t$ with an model AR$(k)$ of order $k$, time point $t$, weighting term $a_n$ and error term $e_{xt}$ the formula would look like this (Cohen, 2014):

$$X_t = \sum_{n=1}^{k} a_n X_{t-n} + e_{xt}$$

It is important to note that with a model order of $\geq 2$, $X_{t0}$ predicts also $X_{t2}$:

![Figure 1.6: Univariate autoregression.](image)

Another possibility would be signals with a predictable change, e.g. with a monotonous increase or decrease of signal strength. As these are implausible in the context of MEG signals, they will not be discussed here.
A simple example would be a sinus oscillation with some added random noise. One would be able to predict future time points based on previous time points, if the model order is high enough to include at least one full sinus oscillation. A more interesting and useful case is bivariate AR with two signals $X_t$ and $Y_t$ and four weighing factors $a_n$, $b_n$, $c_n$ and $d_n$ predicting itself and each other (Figure 1.7):

$$X_t = \sum_{n=1}^{k} a_n X_{t-n} + \sum_{n=1}^{k} b_n Y_{t-n} + e_{xyt}$$

$$Y_t = \sum_{n=1}^{k} c_n Y_{t-n} + \sum_{n=1}^{k} d_n X_{t-n} + e_{yxt}$$

Figure 1.7: Bivariate autoregression. $X \rightarrow Y$ predictions are highlighted in red, $Y \rightarrow X$ predictions in blue.
An example for this would be again two sinus oscillations with some added random noise, but this time the oscillations influence each other, that is they predict the time course of the other signal with a temporal lag of $t - n$. According to the definition of GC, one would be able to better predict future time points of one signal based on previous time points of both signals compared to only one signal. This means that the variance\(^6\) of the error term $e$ in the case of two signals ($e_{xy}$) used for prediction of $X_t$ is smaller than in the case of one signal used for prediction ($e_x$). Comparing both error terms is what constitutes the Granger Prediction, which is the natural logarithm $\ln$ of those error terms:

$$GrangerPrediction = \ln \left( \frac{\text{var}(e_x)}{\text{var}(e_{xy})} \right)$$

Several implications from this formula should be noted: For one, the logarithm of variances is distributed as a $\chi^2$ function which is useful for statistical evaluation (Seth, Chorley, & Barnett, 2013). Also, in the case of equal error terms for both models, the fraction would equate to one and the natural logarithm of one is zero. It is theoretically possible that the error term $e_{xy}$ is smaller than $e_x$, which would mean that the fraction is smaller than 1 and the Granger Prediction is negative. This, however, is practically implausible and could e.g. point to a violation of stationarity (M. X. Cohen, 2014). Normally, the Granger Prediction is expected to be always greater than zero, because generally the more predictors one includes, the better the fit. This is also a potential pitfall, as e.g. each region in the brain could be used to predict another region and the result will almost always be positive.

One solution to the “always positive” Granger Prediction is comparing two different Granger Predictions against each other. A Granger Prediction based on

\(^6\) The variance is a measure of the strength of the error term, as the mean of the error term should be zero.
the real data and can be compared to a Granger Prediction based on the same data with a reversed time axis (Haufe, Nikulin, Müller, & Nolte, 2013). Because Granger causality is sensitive to temporal order, reversing the time axis should also reverse the direction of ‘true’ information flow, while spurious causality due to, for instance, volume conduction or channel differences in signal-to-noise ratio (e.g., Nolte et al. 2008) should not be affected by a reversal of the time axis. Moreover, when taking the difference between two Granger Predictions of “true” and the “time-reversed” data, the values become approximately normally distributed and centered on zero. This allows for a better estimation of true information flow and statistical testing against zero (Cohen, 2014) and was applied accordingly in study 2.

1.6 Outline and research aims of the present dissertation

The first goal of this dissertation was to compare preprocessing and quantification methods in the extent they produce an a priori HC vs CHR contrast, which was done in the pilot study in chapter 1.4.

After establishing preprocessing methods that both show the a priori HC vs CHR contrast and only minimally change the original signal (Figure 1.5), the second goal of this dissertation was to compare quantification methods across three groups, HC, CHR and FA and to determine if they impact the group contrasts or are robust to variations.

Once a robust quantification has been established, study 2 of the present dissertation examined a possible causal relationship between auditory and frontal brain regions in a sensory gating context via a measure of granger causality which has been corrected for spurious causality (chapter 1.5).
2 Study 1: Consistency of Abnormal Sensory Gating in First-Admission and Chronic Schizophrenia across Quantification Methods

2.1 Introduction

Sensory gating refers to the phenomenon of a diminished response to the second of two identical stimuli (Adler et al., 1982). The typical procedure includes a series of click pairs, each of which prompts a positive deflection of the event-related brain potential (ERP) around 50 ms (P50 or the MEG counterpart M50). This modulation of P50/M50 amplitude is quantified by the P50 component of the ERP and dividing the second by the first response (S2/S1; see also de Wilde et al., 2007). The sensory gating phenomenon has been interpreted as reflecting an inhibitory mechanism that facilitates the filtering of irrelevant sensory information in service of “optimal information processing” (Wan et al., 2008, p. 91). Numerous studies have found that P50 and M50 distinguish schizophrenia patients (SZ) and healthy controls (see meta-analyses by Bramon et al., 2004; de Wilde et al., 2007; Heinrichs, 2004; Patterson et al., 2008). Patients typically show an abnormally high P50 or M50 ratio, interpreted as deficient sensory gating. A failure of this inhibitory or filtering mechanism is thought to contribute to perceptual and cognitive symptoms in psychosis such as sensory overload and cognitive fragmentation (Croft et al., 2001; Potter et al., 2006; Venables, 1964). Moreover, relationships to characteristic symptoms of SZ have been reported. Thoma et al. (2003, 2005) and Smith et al. (2010) reported specific relationships between M50 gating and symptoms in schizophrenia, and others did so for P50 gating and symptoms (Arnfred & Chen, 2004; Erwin, Turetsky, Mobeg, Gur, & Gur, 1998; Louchart-de
la Chapelle et al., 2005; Ringel et al., 2004; Yee, Nuechterlein, Morris, & White, 1998). In spectral analysis of a traditional EEG sensory gating paradigm, Keil and colleagues (2016) reported that SZ had less gamma-band power reduction from S1 to S2, which correlated with PANSS positive symptoms, and less alpha-band phase coherence reduction from S1 to S2, which correlated with PANSS negative symptoms. These two spectral measures of gating were in turn correlated. In sum, accumulating evidence supports sensory gating deficits, measured by P50/M50, as a phenomenon that consistently distinguishes schizophrenia patients and controls with medium to large effect sizes (Allen et al., 2009; Bramon et al., 2004) and with sufficient stability to be considered an endophenotype (for review see Thibaut et al., 2015; Miller & Rockstroh, 2013). Still, insufficient sensitivity and stability (Thibaut et al., 2015) and considerable variability in effect sizes (Light & Braff, 1998) have challenged the value of sensory gating as an important biomarker without combination with other electrophysiological indices (Thibaut et al., 2015).

A variety of methodological parameters have been discussed as sources of variation, one of the most controversial being the quantification of the gating ratio itself. Higher retest reliability was found for the S1 minus S2 difference than the S2/S1 ratio (Dalecki et al., 2011; Fuerst et al., 2007; Rentzsch et al., 2008). Test-rest reliability of the P50 ratio has sometimes approximated zero, whereas the M50 ratio fared well (Lu et al., 2007). There is also evidence of the N100 or M100 component showing the same reduction in schizophrenia patients (e.g., Smith et al., 2010), but a review found reduced N100/M100 only in about half of the relevant studies (Rosburg et al., 2008). The ratio quantification may be affected by group differences in P50, M50, N100, or M100 (Edgar et al., 2003; Jin & Potkin, 1996; Moran et al., 2012; Smith et al., 2010). Although some studies
report ratio findings to be carried mainly by S1 (Boutros et al., 2004; Brockhaus-Dumke et al., 2008; Smith et al., 2010; van Tricht et al., 2015), a review by Chang et al. (2011) found S2 to contribute more, and the S2/S1 ratio still more, than S1 to schizophrenia/control differences (see also Hamilton et al., under revision).

Sensory gating results may also vary with P50 amplitude scored relative to pre-stimulus baseline, to the immediately preceding N40 amplitude, or to the immediately subsequent N100 amplitude. When scoring relative to a pre-stimulus baseline, either the pre-S1 baseline or the pre-S2 baseline may be used for scoring the S2 response, with the result that either a potential drift is better accounted for (when scoring S2 50 relative to the pre-S2 baseline) or the baseline is not affected by later aspects of the S1 ERP (when scoring S2 M50 relative to the pre-S1 baseline). Another factor may be analog and digital filter settings, typically undercharacterized. Evaluating quantification methods based on source modeling, Huang and colleagues (2003) found that source-projected M50 data distinguished patients and controls better than scalp P50. Moreover, test-retest reliability was reported to be much better for source-projected M50 than for scalp P50 (Lu et al., 2007). Another advantage is that source projection controls for different head shapes, sensor orientation, and source orientation (Edgar et al., 2003). Source modeling is especially feasible when using MEG because it exploits the spatial-temporal sensitivity of MEG for tangentially oriented sources and avoids the problem of volume conduction. MEG is particularly well suited for characterizing generators located in or near Heschl's gyrus, as is often assumed for P50/M50.

Another possible source of variation among studies is stage of illness, which is also of substantive interest in determining how well the sensory gating deficit may serve as an endophenotype or developmental indicator of illness.
course (Miller & Rockstroh, 2013). Some studies found first-episode patients to show the gating deficit, and one has not.

In sum, inconsistencies in effect sizes make an evaluation of methods and results challenging when studies vary in how they elicit and quantify gating. The present study evaluated several variations in sensory gating scoring methods using a single, large data set, including some comparisons not provided in any single prior report. Comparing sensory gating indices in patient samples with ICD diagnoses of schizophrenia spectrum disorders who differed with respect to chronicity (first vs. multiple treatments) addressed the question of the extent to which inconsistencies in the literature might be due to differences in samples. The present study compared several methods of quantifying M50 gating:

- S2 divided by S1 ratio in sensor space, with M50 peak scored relative to pre-S1 baseline
- S2 divided by S1 ratio in source space, with M50 peak scored relative to pre-S1 baseline
- S2 divided by S1 ratio in sensor space, with M50 peak scored relative to M40 peak and relative to M100 peak
- S2 divided by S1 ratio in source space, with M50 peak scored relative to M40 peak and relative to M100 peak
- S1 minus S2 difference for M50 peak in sensor space, with M50 peak scored relative to pre-S1 baseline
- S1 minus S2 difference for M50 peak in source space, with M50 peak scored relative to pre-S1 baseline

Hypotheses for M50 were: (1) Patients show less gating (higher S2/S1 ratios and smaller S1 minus S2 differences) than controls, regardless of quantification method. (2) The effect size contrasting patients and controls varies as a
function of quantification method. (3) The effect size contrasting patients earlier and later in the course of illness varies as a function of quantification method. The impact of additional methodological aspects was explored by comparing baselines (baseline prior to S1 for scoring both S1 and S2 responses vs. baseline prior to each click), comparing peaks (M50 peak relative to M100 vs. M50 peak relative to M40), and evaluating M100 in parallel with M50.

2.1 Method

2.2.1 Participants

Patients meeting criteria for an ICD diagnosis of schizophrenia spectrum disorder (ICD-code F2x) were recruited from inpatient units at the local public psychiatry inpatient facility. These units specialize either in chronic patients with multiple prior inpatient treatments (CHR from hereon) or in patients who were admitted for inpatient treatment with an ICD-F2x diagnosis for the first time (FA from hereon). Diagnoses were given by experienced psychiatrists and psychologists based on ICD-10 criteria (Dilling et al., 2011) and a German standard guideline for ICD-diagnoses (AMDP-Skalen, Guy et al., 1982). Of 124 eligible patients, 31 patients were excluded if diagnoses did not match schizophrenia spectrum disorders, history of neurological conditions or disorders (including epilepsy or head trauma with a loss of consciousness), pre-admission neuroleptic treatment (for FA), or failure to complete the whole study. In the remaining sample of 93 patients meeting ICD-10 criteria for schizophrenia-spectrum disorders, 58 patients had been treated before (mean ± SD number hospital admissions 8.6 ± 7.2), therefore being labeled “chronic” (CHR). Patients (n = 35) who had been
admitted to the inpatient treatment for schizophrenia for the first time, were described as “first admissions” (FA) below. Some patients within this group had been admitted before the current admission for reasons other than schizophrenia diagnoses (e.g., unclear symptoms, depression or substance abuse), which resulted in a mean ± SD admission of 0.6 ± 1.2. FA had experienced initial symptoms on average for 8.7 weeks (range 1–52) before admission. In addition to the primary ICD-F2 diagnosis, 12 of the total 58 CHR and 12 of the total 34 FA met criteria for one or more comorbid ICD-10 diagnoses. Patient samples did not differ in ratio of comorbid vs. no comorbid diagnoses (see Table 2.1). Most of the 28 comorbid diagnoses in both samples involved substance dependence (F10.x, F11.x and F19.x). In the FA sample, one patient was diagnosed with comorbid social phobia and one with comorbid borderline personality disorder.

Within the patient sample, FA had a higher level of function (per DSM-IV Global Assessment of Functioning Scale) than CHR and, on average, received less neuroleptic medication (per CPZ), although overall symptom severity did not differ (see Table 2.1). Most patients received a combination of first- and second-generation neuroleptics; 3 CHR patients received additional antidepressants, 4 benzodiazepines, 7 anticholinergics, 3 mood stabilizers, and 2 analgesics. In the FA sample, 4 received additional antidepressants and 4 benzodiazepines. Patients, including those with comorbid SUD diagnosis, did not use any drugs in the week prior to MEG measurement, as verified by blood and breath tests. (Drug intake prior to admission was not controlled.) All participants refrained from smoking for at least one hour prior to measurement. Participants were recruited for a larger project on training-induced cortical reorganization in schizophrenia. Neurromagnetic indices and their association with cognitive tests for largely overlap-
ping samples were reported before (Carolus et al., 2014) and after targeted, neuroplasticity-oriented intervention (Popov et al., 2015). The present results used entirely different scoring of the MEG data assessed prior to training.

HC were included if they did not report any psychiatric or neurological disorders according to the Mini International Neuropsychiatric Interview including substance abuse (Ackenheil et al., 1999). This led to the exclusion of 6 from the originally screened 34 HC. SZ and HC were comparable with respect to gender distribution and age. As is typical with schizophrenia samples, CHR were older and FA younger than the single HC sample. HC had more years of education and smoked fewer cigarettes than SZ (see Table 2.1 and Supplementary Table 2.1A-C for adjustment of sensory gating group differences to demographic group differences).
### Table 2.1

**Demographic information on samples: healthy controls (HC), chronic schizophrenia patients (CHR) and firstly admitted schizophrenia patients (FA)**

<table>
<thead>
<tr>
<th></th>
<th>HC</th>
<th>CHR</th>
<th>FA</th>
<th>Group comparison</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td>14 ♂ / 14 ♀</td>
<td>39 ♂ / 19 ♀</td>
<td>26 ♂ / 9 ♀</td>
<td>$\chi^2=4.2$</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>29.3 (9.5)</td>
<td>36.8 (9.2)</td>
<td>22.4 (3.8)</td>
<td>F=35.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Education (years)</strong></td>
<td>15.0 (2.5)</td>
<td>14.5 (3.4)</td>
<td>12.3 (2.4)</td>
<td>F=7.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>LQ</strong></td>
<td>71.7 (61.5)</td>
<td>69.3 (53.4)</td>
<td>63.4 (62.5)</td>
<td>T&lt;1</td>
<td>ns</td>
</tr>
<tr>
<td><strong>Cig/day</strong></td>
<td>1.2 (3.1)</td>
<td>10.4 (13.0)</td>
<td>7.5 (8.6)</td>
<td>F=7.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>CPZ</strong></td>
<td>-</td>
<td>614.3 (409.3)</td>
<td>415.4 (358.6)</td>
<td>T=2.2</td>
<td>0.033</td>
</tr>
<tr>
<td><strong>PANSS (P/N/G)</strong></td>
<td>-</td>
<td>15.4 (5.2) / 16.0 (4.1) / 35.4 (7.7) / 12 / 46</td>
<td>41.3 (13.2) / 17.4 (7.1) / 30.0 (9.0) / 12 / 23</td>
<td>T&lt;1 / T&lt;1 / T&lt;1 / $\chi^2=2.1$</td>
<td>ns / ns / ns / ns</td>
</tr>
<tr>
<td><strong>Comorbid (yes/no)</strong></td>
<td>-</td>
<td>43.1 (13.2)</td>
<td>51.3 (13.2)</td>
<td>T=2.6</td>
<td>0.012</td>
</tr>
<tr>
<td><strong>GAF</strong></td>
<td>-</td>
<td>106 (16)</td>
<td>102 (13)</td>
<td>T=1.3</td>
<td>ns</td>
</tr>
<tr>
<td><strong>IQ</strong></td>
<td>-</td>
<td>8.6 (7.2)</td>
<td>0.6 (1.2)</td>
<td>T=4.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Number of admissions</strong></td>
<td>-</td>
<td>14 / 14</td>
<td>21 / 37</td>
<td>10 / 25</td>
<td>T=2.4</td>
</tr>
<tr>
<td><strong>MRI (yes/no)</strong></td>
<td>-</td>
<td>93.5 (2.8)</td>
<td>93.5 (3.2)</td>
<td>92.9 (3.0)</td>
<td>T=2.4</td>
</tr>
</tbody>
</table>

*Note. HC = healthy control group. CHR = chronic schizophrenia group. FA = first-admission schizophrenia group. Means and standard deviations (in parentheses); ns: not significant when p > .05). LQ: Lateralization quotient between 100 = 100% right-handed, -100 = 100% left-handed. MRI: Number of participants with an individual structural magnetic resonance image available. CPZ: chlorpromazine equivalent. PANSS: Positive and Negative Syndrome Scale (P: Positive, N: Negative, G: General; Kay et al., 1987). GAF: DSM-IV Global Assessment of Functioning scale. IQ: Premorbid IQ measured with a standard German test for premorbid intelligence, MWT-B (Lehrl, 2005). Group comparison: F-values from ANOVA comparing the three groups with F(2,118), t-values from independent-samples t-tests comparing the two patient groups with t(92). Number of admissions in FA included inpatient treatment for diagnoses other than schizophrenia.*
2.2.2 Data collection

Prior to MEG measurement, individual hearing levels were determined for each ear. The paired-click task comprised 100 pairs of 3 ms square-wave clicks (S1 & S2) presented with 500 ms onset-to-onset interstimulus intervals and a variable trial offset-to-onset interval of 7 to 9 s. Clicks were presented 60 dB above individual hearing level and delivered via 5 m non-ferromagnetic tubes. No performance task was involved, except that participants were asked to keep their eyes focused on a small fixation point throughout the procedure. Time of day varied randomly depending on participants’ treatment schedules, but most sessions took place in early afternoon (mean ± SD 13:46 ± 2 h 5 min). MEG was recorded in a shielded room with a 148-sensor MAGNES 2500 WH whole-head magnetometer (4D Neuroimaging, San Diego, USA) while subjects were in a supine position. Environmental noise was recorded by an additional 11 reference magnetometers approximately 1 m from the head and subtracted from the remaining sensors.

Data were continuously recorded with a sampling rate of 678.17 Hz and a bandpass filter of 0.1 to 200 Hz. The subject’s nasion, left and right ear canal, and head shape were digitized using a Polhemus 3Space FASTRACK. Epochs of 2 s before and 2 s after the first click of each trial were extracted. Heart and eye-blink artifacts were corrected by independent component analysis. Trials containing movement artifacts or SQUID jumps were rejected based on visual inspection: for each participant, trials with extreme variance were removed, judged as reflecting movement artifacts. This procedure identified a similar number of trials for the three groups (see Table 2.1) and a similar reduction in mean variance from $6.43 \times 10^{-25} \, T^2$ (SD = $4.80 \times 10^{-24} \, T^2$) before to $8.55 \times 10^{-26} \, T^2$ (SD = $2.30 \times 10^{-26} \, T^2$) after removal of artifact-contaminated trials. Offline treatment of
the MEG signals was accomplished using the MATLAB-based open-source signal processing toolbox FieldTrip (Oostenveld et al., 2011), complemented by in-house MATLAB functions.

### 2.2.3 Sensor-space analysis

Analyses reported here used a FIR high-pass filter (-6 dB half-amplitude at 1.7 Hz after application in both directions, 5 Hz transition band, single-pass order 300, thus spanning 442 ms on each side of the filtered point, maximum passband deviation of 0.5%, maximum stopband attenuation of 10.5 dB) and an IIR low-pass filter (-6 dB at 80 Hz after application in both directions, transition width 40 Hz, order 6, maximum passband deviation of 0% with a Chebyshev Type 2 response, maximum stopband attenuation of 192 dB). Both filters were Hamming-windowed and applied in forward and backward directions, thus being symmetrical and with zero-phase lag. A 1000 ms baseline preceding S1 onset was subtracted from each trial. (See Supplement chapter 2.5 p.67f and chapter 1.4.5 for discussion of extensive exploratory work undertaken to select the present digital filters and choice of baseline.) However, given that studies typically use shorter baselines (without discussion of the impact of filtering), all analyses were repeated using the more common 100 ms baseline. Moreover, both peaks (S1 and S2) were referred first to the same pre-S1 baseline. However, to be comparable to many studies of sensory gating measures, all analyses were repeated using separate pre-stimulus baselines for S1 and S2. Results are summarized in Supplementary Table 2.2, confirming identical results for 5 of the 6 quantifications. Because of this high concurrence and for the sake of clarity, results are reported here for the 1000 ms long pre-S1 baseline.
Figure 2.1: Topography 40-65 ms (M50, panel A) and 65-130 ms (M100, panel B) after S1 averaged across all participants. Sensors surrounding the averaged peak pixels used for automatic peakfinding are highlighted as larger dark circles.

M50 and M100 peaks were identified within 40-130 ms after S1 and S2 based on the topographic map for the overall sample average (Fehler! Verweisquelle konnte nicht gefunden werden.) as follows. First, separately for each hemisphere and each component (M50, M100), a set of three adjacent sensors closest to the peak pixel was identified. A fourth, adjacent sensor was added to the cluster that worked best for the goal of using a single set of sensors for all four hemisphere x component scores. S1 and S2 peaks were chosen for left and right sensor clusters using an automated script: for each subject, the automatic algorithm determined the most negative and the most positive peak within 40-130 ms after click onset for each sensor as illustrated in Figures 2.1 and 2.3. Then, of
the four sensors per hemisphere⁷, the sensor(s) with the largest positive and largest negative values were chosen for scoring. As peak polarity does not necessarily have the same meaning in MEG as in EEG, the earlier of the two peaks was defined as M50 and the later as M100. This procedure accounted for potential individual differences in peak polarities and peak locations (for details see Supplementary Table 2.5A-C).

2.2.4 Source-space analysis

Source reconstruction was based on individual structural magnetic resonance images (sMRI) that were obtained using a Philips Gyroscan ACS-T 1.5 T with field of view 256×256 and 200 sagittal slices. Via the methods of Popov et al. (2011), an automated source reconstruction procedure used individual digitized scalp shapes and individual sMRI co-registered to the MEG coordinate system via NUTMEG (Neurodynamic Utility Toolbox for Magnetoencephalography; Dalal et al., 2004) to construct a realistic, single-shell brain model (Nolte, 2003).

 footnote The present method of using a four-sensor array compensated sufficiently for individual head and location differences in the somewhat the same way as, but without the additional computational and assumptional burden of, source reconstruction. Although the exact number of sensors chosen surrounding the peak was arbitrary, additional exploratory analyses with 1-9 surrounding sensors on each hemisphere produced results similar to those for 3-9 sensors. This might be unique to MEG, since source reconstruction is better and sensor positioning might be worse (the position of the sensors might be more controllable when putting an EEG net directly onto the head than when positioning the head inside a MEG helmet).
For the 66 of the 121 subjects for whom the sMRI was not available, a MNI template brain (Montreal Neurological Institute (MNI), Montreal, Canada; http://www.bic.mni.mcgill.ca/brainweb) was transformed to best fit the subject’s digitized individual head shape (which was available from every subject) in place of the sMRI (see also Lecaignard et al., 2008; Keil et al., 2010). A voxel size of 10 mm³ was used for the leadfield computation portion of the source analysis.

A distributed source estimation was performed using a Linearly-Constrained Minimum Variance (LCMV) beamformer algorithm (Van Veen et al., 1997). The covariance of the sensor data used for source reconstruction was computed on the basis of the single trials. This method includes substantial trial-by-trial variability and has been shown to mitigate problems regarding reconstructing activity from correlated sources. The source estimation was registered to a common brain template, and automatic anatomical parcellation of brain areas was applied according to the MNI atlas (Tzourio-Mazoyer et al., 2002). Comparing averaged S1 (0 – 300 ms) and S2 (500 – 800 ms) responses confirmed both Heschl’s gyri as the atlas regions with the greatest reduction in source strength (Figure 2.2). Time series of primary auditory cortex activity were derived from source-projected data of five adjacent 1 cm³ voxels (per MNI atlas: two for right Heschl's gyrus, three for left) and averaged per hemisphere to establish a “virtual sensor” for each Heschl’s gyrus. The resulting source waveforms were processed with the same filter settings and peak-picking algorithm as sensor data, with two exceptions:

First, prior to M50-to-baseline scoring, waveforms were high-pass filtered at 5 Hz instead of 1 Hz because of drift that sometimes reduced the S2-evoked M50 peak amplitude to near zero and made ratio scores implausible. Filter char-
acteristics were those of the 1 Hz filter described earlier except for a 3 Hz transition band, maximum passband deviation from ideal filter of 0.4%, maximum stopband attenuation of 104 dB, and -6 dB half-amplitude at 5.9 Hz after application in both directions.

Second, M50 and M100 source-space scoring used the same script as did sensor-space scoring, except that no peak sensor was chosen, since only one virtual sensor was computed for each hemisphere’s source.

Figure 2.2: S1 minus S2 power in parcellated regions averaged across all participants (see main text). The red area highlights Heschl’s gyrus. Color shading indicates % power decrease of S2 relative to S1 with dark grey to red colors indicating areas with larger S1-S2 differences.

2.2.5 Gating quantification

Ratios and difference scores were computed for M50 and M100 from sensor- and source-space data. For ratio scores (S2 amplitude / S1 amplitude), am-
amplitude was determined (a) relative to the pre-S1 baseline (see p. 6 and Supplementary Table 2.2 on pre-S2 baseline scoring) and (b) relative to another component (M50 amplitude minus M40 or M100 amplitude). Difference scores were determined as S1 amplitude minus S2 amplitude. In analyses of sensor-space scores, the absolute values of the sensor scores were averaged across hemispheres before computing differences or ratios. This served to approximate the traditional scoring of scalp P50 at Cz relative to a laterally symmetrical reference, while controlling for the typically opposite polarities of the M50 at the surface of the left and right scalp. Source analyses were done individually for each hemisphere, because lateralization of the HC vs. CHR difference has been reported. For detailed per-hemisphere results see Supplementary Tables 2.7A-C. Group differences in scores of interest were evaluated using separate analyses of variance (ANOVA) with the between-subjects factor Group for all three groups. Two-tailed t-tests were used for comparing each pair of groups. Effect sizes were represented via Hedges’ g, for which the interpretive guidelines of can be used but which provides correction of Cohen’s d for small and uneven sample sizes.

2.2.6 Gating comparison

Finally, the hypotheses of varying magnitude of effect size as a function of quantification were tested by comparing the various quantifications within one analysis. For this purpose, the individual values for the six quantifications were ranked per quantification and then subjected to the Friedman test, a non-parametric statistical test similar to an ANOVA.
Study 1: Consistency of Abnormal Sensory Gating in First-Admission and Chronic Schizophrenia across Quantification Methods

Figure 2.3: Time course of activity for sensor and source analysis averaged across participants within group.

HC: healthy control group;
CHR: chronic schizophrenia group;
FA: first-admission schizophrenia group.

[T]: field strength in Tesla.
[a.u.]: source strength in arbitrary units.
[s]: time in seconds.
Light color regions indicate +/-1 standard error. Latency windows for scoring S1 and S2 M50 and M100 are highlighted in gray.
2.3 Results

Figure 2.3 illustrates resulting sensor-space waveforms for S1 and S2 responses averaged over individual peak sensors and averaged separately for the three groups. Table 2.2 and Figure 2.4 provide mean and standard deviations, effect sizes, and inferential statistics for the six quantification procedures: With one exception, M50 quantifications in sensor space and all left-hemisphere M50 quantifications in source space confirmed significant differences between HC and CHR or FA, respectively, but not between CHR and FA. (Source-space-based M50-ratio only approached significance, $p < .1$, for both HC vs. CHR and HC vs. FA.) These effects were associated with medium-to-large effect sizes. The Friedman test indicated no differences between the six quantifications ($\chi^2(5)=4.4$, $p=0.49$).

As groups differed in age and smoking habits, and as patient groups differed in medication (see Table 2.1), supplemental analyses of covariance with age (mean 30.9 years), smoked cigarettes per day (mean 7.4), and CPZ equivalent (mean 536 mg) as covariates evaluated potential confounds (per strategy discussed in Miller & Chapman, 2001, and Verona & Miller, 2015; for detailed results see Supplementary Table 2.1). With the one exception of a non-significant HC vs. CHR difference in M50 ratio (source estimate) when adjusting for age, effects were the same as reported in Table 2.2. Thus, confounding influences of age and medication status seem unlikely.

Group differences similar to those reported above for M50 quantifications (significant HC vs. SZ contrasts but no significant CHR vs. FA contrasts) were also observed for (a) single-trial peak S1 and S2 scoring, (b) other baselines, (c) M100 quantifications, (d) peak latency constraints of S2 to e.g. ± 10 ms around
the individual peak latency at S1, (e) scoring M50 relative to the immediately preceding M40 instead of the immediately subsequent M100 and (f) hemisphere specific results. Detailed results are provided in Supplementary Tables 2.2-7. Group differences as reported for M50 quantifications above were obtained in analyses (b) and (d), but no HC vs. CHR differences were found for analyses (a) and (c). For analysis (e), group differences were similar to those reported for the main M50 quantifications found for sensor data, though no HC vs. CHR differences were found for source data. For analysis (f), group differences were similar for sensor data of the left hemisphere (compared to averaged hemispheres), though sensor data of the right hemisphere generally showed weaker effect sizes and were only bordering significance for the S1-S2 M50 difference. For source data, only the left hemisphere showed significant group differences.
Study 1: Consistency of Abnormal Sensory Gating in First-Admission and Chronic Schizophrenia across Quantification Methods

Table 2.2

<table>
<thead>
<tr>
<th></th>
<th>M50-Baseline Ratio</th>
<th>M50-M100 Ratio</th>
<th>S1-S2 M50 Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensor</td>
<td>Source</td>
<td>Sensor</td>
</tr>
<tr>
<td>HC</td>
<td>0.48 (0.19)</td>
<td>0.54 (0.41)</td>
<td>0.49 (0.16)</td>
</tr>
<tr>
<td>CHR</td>
<td>0.65 (0.26)</td>
<td>0.68 (0.29)</td>
<td>0.61 (0.22)</td>
</tr>
<tr>
<td>FA</td>
<td>0.63 (0.28)</td>
<td>0.73 (0.37)</td>
<td>0.59 (0.18)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>4.46*</td>
<td>2.7#</td>
<td>3.67*</td>
</tr>
<tr>
<td>HC vs. CHR</td>
<td>0.70**</td>
<td>0.43#</td>
<td>0.60**</td>
</tr>
<tr>
<td>HC vs. FA</td>
<td>0.60*</td>
<td>0.50#</td>
<td>0.54*</td>
</tr>
<tr>
<td>FA vs. CHR</td>
<td>0.07ns</td>
<td>0.16ns</td>
<td>0.13ns</td>
</tr>
</tbody>
</table>

Note. HC = healthy control group. CHR = chronic schizophrenia group. FA = first-admission schizophrenia group. Sensor-space scores were averaged across hemispheres (see Methods). Top three rows: For both sensor and source data, scores reflect S2 divided by S1 amplitudes relative to 1000 ms pre-S1 baseline (M50-Baseline Ratio) and M50 peak relative to M100 (M50-M100 Ratio). S1-S2 M50 Difference: Scores reflect S1 minus S2 amplitude subtraction in fT for sensor data and in arbitrary units for source data. Scores are presented as means and standard deviations (in parentheses).

ANOVA: scores indicate F-values from three-group ANOVA with F(2,118). Two-group comparisons (HC vs. CHR, HC vs. FA, FA vs. CHR): scores indicate effect sizes in Hedge’s g, superscript significance from t-tests with t(85), t(62), and t(92), respectively.

Bottom four rows: Level of significance for group comparisons are indicated by superscript as ns: not significant when p > .1, # p < .1, * p < .05, ** p < .01.
Figure 2.4: Metrics combining S1 and S2 scores:

- **M50-Baseline Ratio**: S2/S1 ratio for M50 scored relative to pre-S1 baseline.
- **M50-M100 Ratio**: S2/S1 ratio for M50 scored relative to adjacent M100.
- **S1-S2 M50 Difference**: S1-S2 each scored relative to pre-S1 baseline.

Error bars indicate +/-1 standard error.

Numbers provide effect size expressed in Hedge’s g.

Level of significance for group comparisons are indicated by superscript as:
- **ns**: not significant when p > .1,
- # p < .1,
- * p < .05,
- ** p < .01.

S1-S2 difference measures are normalized via division by the mean of HC to allow a comparable scaling of source and sensor values.
2.4 Discussion

Inconsistent results as a function of methodological differences across studies have sometimes been cited as challenging the classification of the auditory sensory gating deficit as a core feature of schizophrenia psychopathology. The present report evaluated six common approaches within the same experimental protocol and for both chronic and first admission patients compared to healthy controls in order to determine, in the same data set, how substantial such discrepancies are.

Per hypothesis 1, a gating deficit in M50 in schizophrenia patients was observed consistently across gating quantification methods. The gating ratios varying between .59 and .73 for CHR and FA vs. and between .48 and .54 for HC are within the range reported in meta-analyses (de Wilde et al., 2007; Patterson et al., 2008), although effect sizes between .43 and .70 are less than those with Cohen’s d > 1 reported in other reviews (summarized in Thibaut et al., 2015). This is an appropriate foundation for the comparisons of methods that motivated the study, though it was expected given the ratios reported for the overlap in participants with Carolus et al. (2014). In the latter study, in which M50 was determined by dipole fitting by trained raters, ratios were 0.39 for HC, 0.50 for CHR, and 0.61 for FA. No HC vs. CHR differences were observed for any M100 or single-stimulus measure, evidence that the traditional P50 or M50 gating abnormality is not driven primarily by N100 or M100 or by only S1 or only S2 responses.

The second hypothesis, that quantification methods differ in the magnitude of effect sizes, was not confirmed. The 12 M50 gating effect sizes in Table 2.2 comparing HC to one of the SZ groups ranged from .53 to .71. (The respective effect sizes for the overlapping sample reported in Carolus et al., 2014, were .79
The 12 M50 gating quantifications were not markedly discrepant in effect size, and the Friedman test of heterogeneity did not find significant differences. The magnitude of effect sizes the HC vs. SZ group difference was largely consistent across quantification methods: for the same 12 M50 gating effect sizes in Table 2.2, HC and SZ were consistently distinguishable in M50 gating, with similar medium-to-large effect sizes across quantification methods. No HC vs. SZ differences were found for M100 or for S1 or S2 responses analyzed individually. Thus, regarding the primary motivation for the study, M50 gating and the SZ gating deficit were robust to differences in gating quantification method.

Hypothesis 3, that CHR/FA differences, if any, differ by quantification method, received no support, in that all six effect sizes in the bottom row of Table 2.2 were small (.07 to .27) and nonsignificant. This result is consistent with the characterization of the P50/M50 sensory gating deficit as a core feature of schizophrenia psychopathology evident early in the course of the disorder and not merely a consequence of chronic illness and/or multiple inpatient treatments.

Evaluation of multiple quantification methods in a given data set can compensate for some issues of data quality, as they have different strengths and weaknesses. For example, individual head shape and sensor-position differences can be compensated for by using individualized peak sensor selection or by analyzing in source space, and the impact of differences in overall sensor signal strength can be mitigated by using a ratio rather than a difference score or by using source analysis. Moreover, as the meta-analysis by de Wilde et al. (2007) emphasized, methodological differences between studies, which may contribute to inconsistent results, make it difficult to evaluate the stability of sensory gating effects across studies, such as to establish it as an endophenotype for schizophrenia. The present independence of effect sizes and statistical results from the
quantification methods evaluated here encourages the study of sensory gating deficits as a core feature of schizophrenia psychopathology and in particular as an endophenotype (Miller & Rockstroh, 2013; Thibaut et al., 2015).

Beyond the general consistency of the present MEG results with the gating literature, which is largely EEG-based, the present source analysis replicated the left-hemisphere dominance of the HC vs. CHR contrast reported in previous MEG gating studies. Specifically, only left-, not right-hemisphere quantifications showed a HC vs. CHR difference (left vs. right comparison not significant for M50 minus baseline ratio, \( t(120)<1, p>0.1 \), nor for M50 minus M100 ratio, \( t(120)<1, p>0.1 \), but highly significant for S1-S2 M50 Difference, \( t(120)=5.62, p<0.001 \)). Similar to the source analysis, sensor-space results confirmed stronger HC vs. CHR contrasts for left-hemisphere sensors.

Several limitations of the present study can be noted. There are significant advantages to the present comparison of gating quantification methods from a single large data set, thus avoiding cross-study differences in participants, recording, and preprocessing as a contributor to inconsistent results, but the generalizability of any one data set are necessarily limited. Systematic simulations to evaluate the effects of preprocessing options and their interactions with quantification method remain to be done. Still, present results suggest that that quantification method alone is not necessarily a barrier to successful use of P50/M50 in studies of normal function or psychopathology.

A second limitation is that the present comparisons were restricted to methods of quantification of sensory gating and did not directly address differences in how EEG and MEG signals appear at the scalp due to their somewhat different biophysical properties, which could be a source of inconsistent results. These differences were addressed in several indirect ways. First, for some
analyses, MEG sensor data were averaged across hemispheres to approximate midline EEG recordings. Second, we adjusted quantification methods: unlike in EEG recordings, in MEG recordings vertically oriented lateralized sources (such as Heschl’s gyrus for M50; though see Edgar et al., 2003, for evidence of cross-subject variability in orientation) typically produce magnetic fields, and thus MEG scalp components, with opposite polarities over homologous sites. (Conversely, EEG would do this with sources oriented left to right, and in such a case lateralized MEG recording would not be subject to such a polarity reversal.) This difference from typical EEG P50 studies was addressed here by identifying M50 and M100 as components having complementary polarities. Third, the possibly less consistent placement of the MEG helmet compared to a fixed EEG net was counteracted by choosing individual peak sensors or by source reconstruction. These strategies notwithstanding, EEG and MEG provide both redundant and complementary information, so a direct comparison of results would be limited by inherent differences, e.g. better sensitivity to radial (EEG) vs. tangential (MEG) sources. MEG provides advantages, such as better source reconstruction (Edgar et al., 2003), better test-retest reliability (Lu et al., 2007), and better differentiation of hemispheres, whereas EEG is cheaper to purchase and use and much more widely available. Even though the neural activity of interest in paired-click gating studies manifests in both electrical and magnetic fields, so that EEG and MEG measurements are equally of interest, most of the literature has used EEG. Thus, the present study would be fruitfully complemented by a similar EEG study.

Another possible source of differences across studies is stimulus properties. Characterization of the intensity of such a brief auditory stimulus in a way that would be replicable cross-lab is quite difficult. Stimulus rise times are often not reported and presumably not measured or controlled (they were not in the
present data set). Even measuring dB with such a brief stimulus is challenging. In any case, the generalizability of present results to stimulus properties other than those used in the present data set is unknown. However, the similarity of present sensor-space gating ratios to those in the literature provides some evidence that results apply to other contexts.

Age, medication, nicotine, and education may be considered potential confounds, both within a study that compares groups and across studies, especially because FA patients had been receiving neuroleptic medication for a shorter period than had CHR patients. However, taking these variables into account as covariates did not change the pattern of results (see Supplementary Table 2.1).

Two final limitations concern the time course of schizophrenia. The robustness of sensory gating deficits over the course of a disorder cannot be adequately evaluated by cross-sectional comparison of samples presumably at different stages of the disorder, with markedly different treatment histories. In addition, the inclusion of FA patients does preclude these patients having experienced psychotic symptoms before admission. That is, FA participants may not represent first-episode phenomena. Nevertheless, consistency of findings across two samples that differ in the amount and duration of treatment suggests that the gating deficit is not a result of long-term medication or other aspects of chronicity. Thus, present results are in line with evidence of the sensory gating deficit in first-episode patients (Yee et al., 2010) and support its interpretation as an index of treatment-independent psychopathology.

The comparison of several measures within the same data set suggests independence of auditory gating effects from the specific method of M50 quantification, supporting sensory gating as a robust assay of psychological and neural function. Moreover, the comparison of these measures in patient groups differing
in stage of illness supports treating abnormal gating as an index of fundamental psychopathology in schizophrenia and candidacy as an endophenotype. Although absence of statistical differences among quantification methods cannot constitute proof, present results, based on larger sample Ns than many past P50 and M50 studies, indicate that differences in quantification methods do not contribute critically to sensory gating differences reported in the literature.
### 2.5 Supplements

Supplementary Table 2.1A

*Metrics and group comparisons (as in Table 2.2) for quantifications adjusted for mean age = 30.9 years*

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Source</th>
<th>Sensor</th>
<th>Source</th>
<th>Sensor</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>0.480</td>
<td>0.541</td>
<td>0.498</td>
<td>0.527</td>
<td>76.9</td>
</tr>
<tr>
<td></td>
<td>(0.259)</td>
<td>(0.344)</td>
<td>(0.196)</td>
<td>(0.275)</td>
<td>(38.9)</td>
</tr>
<tr>
<td>CHR</td>
<td>0.631</td>
<td>0.660</td>
<td>0.602</td>
<td>0.657</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td>(0.282)</td>
<td>(0.388)</td>
<td>(0.221)</td>
<td>(0.312)</td>
<td>(43.5)</td>
</tr>
<tr>
<td>FA</td>
<td>0.648</td>
<td>0.759</td>
<td>0.606</td>
<td>0.724</td>
<td>52.4</td>
</tr>
<tr>
<td></td>
<td>(0.290)</td>
<td>(0.396)</td>
<td>(0.225)</td>
<td>(0.319)</td>
<td>(44.7)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>4.18*</td>
<td>2.93#</td>
<td>3.19*</td>
<td>3.96*</td>
<td>3.66*</td>
</tr>
<tr>
<td>HC vs CHR</td>
<td>0.54*</td>
<td>0.31_{ns}</td>
<td>0.48*</td>
<td>0.43#</td>
<td>0.50*</td>
</tr>
<tr>
<td>HC vs FA</td>
<td>0.60*</td>
<td>0.58*</td>
<td>0.50*</td>
<td>0.65*</td>
<td>0.57*</td>
</tr>
<tr>
<td>FA vs CHR</td>
<td>0.06_{ns}</td>
<td>0.25_{ns}</td>
<td>0.02_{ns}</td>
<td>0.21_{ns}</td>
<td>0.08_{ns}</td>
</tr>
</tbody>
</table>

*Note. HC = healthy control group. CHR = chronic schizophrenia group. FA = first-admission schizophrenia group. Top three rows: For both sensor and source data, scores reflect S2 divided by S1 amplitudes relative to 1000 ms pre-S1 baseline (Ratio M50-Baseline) and M50 peak relative to M100 (Ratio M50-M100). S1-S2 M50 Difference: Scores reflect S1 minus S2 amplitude subtraction in fT for sensor data and in arbitrary units for source data. Scores are presented as means and standard deviations (in parentheses).

ANOVA: scores indicate F-values from three-group ANOVA with F(2,118). Two-group comparisons (HC vs CHR, HC vs FA, FA vs CHR): scores indicate effect sizes in Hedge’s g, superscript significance from t-tests with t(85), t(62), and t(92), respectively.

Bottom four rows: Level of significance for group comparisons are indicated by superscript as ns: not significant when p > .1, # p < .1, * p < .05, ** p < .01.*
### Supplementary Table 2.1B

**Metrics and group comparisons (as in Table 2.2) for quantifications adjusted for mean cigarettes per day = 7.4**

<table>
<thead>
<tr>
<th>Sensor Source</th>
<th>Sensor</th>
<th>Source</th>
<th>Sensor</th>
<th>Source</th>
<th>Sensor</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HC</strong></td>
<td>0.476</td>
<td>(0.265)</td>
<td>0.501</td>
<td>(0.201)</td>
<td>0.514</td>
<td>(0.286)</td>
</tr>
<tr>
<td><strong>CHR</strong></td>
<td>0.646</td>
<td>(0.259)</td>
<td>0.611</td>
<td>(0.198)</td>
<td>0.671</td>
<td>(0.196)</td>
</tr>
<tr>
<td><strong>FA</strong></td>
<td>0.627</td>
<td>(0.254)</td>
<td>0.588</td>
<td>(0.195)</td>
<td>0.712</td>
<td>(0.278)</td>
</tr>
<tr>
<td><strong>ANOVA</strong></td>
<td>2.95*</td>
<td>2.05*ns</td>
<td>2.57*#</td>
<td>2.74*</td>
<td>2.70*</td>
<td>3.97*</td>
</tr>
</tbody>
</table>

*Note: as in Table 1A*

### Supplementary Table 2.1C

**Metrics and group comparisons (as in Table 2.2) for quantifications adjusted for mean CPZ = 536mg**

<table>
<thead>
<tr>
<th>Sensor Source</th>
<th>Sensor</th>
<th>Source</th>
<th>Sensor</th>
<th>Source</th>
<th>Sensor</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CHR</strong></td>
<td>0.634</td>
<td>(0.267)</td>
<td>0.604</td>
<td>(0.198)</td>
<td>0.669</td>
<td>(0.289)</td>
</tr>
<tr>
<td><strong>FA</strong></td>
<td>0.646</td>
<td>(0.272)</td>
<td>0.606</td>
<td>(0.201)</td>
<td>0.707</td>
<td>(0.290)</td>
</tr>
<tr>
<td><strong>t-test</strong></td>
<td>t&lt;1</td>
<td>t&lt;1</td>
<td>t&lt;1</td>
<td>t&lt;1</td>
<td>t&lt;1</td>
<td>t&lt;1</td>
</tr>
<tr>
<td><strong>FA vs CHR</strong></td>
<td>0.04*ns</td>
<td>0.11*ns</td>
<td>0.01*ns</td>
<td>0.13*ns</td>
<td>0.01*ns</td>
<td>0.20*ns</td>
</tr>
</tbody>
</table>

*Note: as in Table 1A except two-group comparison by t-test with t(92).*
Supplementary Table 2.2

*Metrics and group comparisons (as in Table 2.2) for quantifications with separate S1 and S2 baseline*

<table>
<thead>
<tr>
<th></th>
<th>M50-Baseline Ratio</th>
<th>M50-M100 Ratio</th>
<th>S1-S2 M50 Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Source</td>
<td>Sensor</td>
<td>Source</td>
</tr>
<tr>
<td>HC</td>
<td>0.47 (0.22)</td>
<td>0.57 (0.56)</td>
<td>0.49 (0.16)</td>
</tr>
<tr>
<td>CHR</td>
<td>0.79 (0.53)</td>
<td>0.67 (0.28)</td>
<td>0.61 (0.22)</td>
</tr>
<tr>
<td>FA</td>
<td>0.67 (0.36)</td>
<td>0.72 (0.36)</td>
<td>0.59 (0.18)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>5.22**</td>
<td>1.25ns</td>
<td>3.67*</td>
</tr>
<tr>
<td>HC vs CHR</td>
<td>0.70**</td>
<td>0.25ns</td>
<td>0.60**</td>
</tr>
<tr>
<td>HC vs FA</td>
<td>0.66*</td>
<td>0.33ns</td>
<td>0.54*</td>
</tr>
<tr>
<td>FA vs CHR</td>
<td>0.24ns</td>
<td>0.16ns</td>
<td>0.13ns</td>
</tr>
</tbody>
</table>

*Note.* HC = healthy control group. CHR = chronic schizophrenia group. FA = first-admission schizophrenia group. Top three rows: For both sensor and source data, scores reflect S2 divided by S1 amplitudes relative to 100 ms pre-stimulus baseline separate for S1 and S2 (Ratio M50-Baseline) and M50 peak relative to M100 (Ratio M50-M100). S1-S2 M50 Difference: Scores reflect S1 minus S2 amplitude subtraction in fT for sensor data and in arbitrary units for source data. Scores are presented as means and standard deviations (in parentheses).

ANOVA: scores indicate F-values from three-group ANOVA with F(2,118). Two-group comparisons (HC vs CHR, HC vs FA, FA vs CHR): scores indicate effect sizes in Hedge’s g, superscript significance from t-tests with t(85), t(62), and t(92), respectively.

Bottom four rows: Level of significance for group comparisons are indicated by superscript as ns: not significant when p > .1, # p < .1, * p < .05, ** p < .01.
Supplementary Table 2.3

*Metrics and group comparisons (as in Table 2.2) for S1- and S2-evoked response amplitudes*

<table>
<thead>
<tr>
<th></th>
<th>S1 M50</th>
<th></th>
<th>S1 M100</th>
<th></th>
<th>S2 M50</th>
<th></th>
<th>S2 M100</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensor</td>
<td>Source</td>
<td>Sensor</td>
<td>Source</td>
<td>Sensor</td>
<td>Source</td>
<td>Sensor</td>
<td>Source</td>
</tr>
<tr>
<td>HC</td>
<td>155</td>
<td>6.24</td>
<td>211</td>
<td>5.91</td>
<td>71</td>
<td>2.82</td>
<td>100</td>
<td>2.98</td>
</tr>
<tr>
<td></td>
<td>(54)</td>
<td>(3.58)</td>
<td>(72)</td>
<td>(3.74)</td>
<td>(27)</td>
<td>(1.81)</td>
<td>(37)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>CHR</td>
<td>134</td>
<td>5.73</td>
<td>185</td>
<td>5.51</td>
<td>82</td>
<td>3.61</td>
<td>104</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>(58)</td>
<td>(3.24)</td>
<td>(85)</td>
<td>(3.25)</td>
<td>(40)</td>
<td>(2.26)</td>
<td>(48)</td>
<td>(2.06)</td>
</tr>
<tr>
<td>FA</td>
<td>133</td>
<td>4.10</td>
<td>176</td>
<td>4.26</td>
<td>77</td>
<td>2.60</td>
<td>96</td>
<td>2.66</td>
</tr>
<tr>
<td></td>
<td>(51)</td>
<td>(2.79)</td>
<td>(77)</td>
<td>(2.96)</td>
<td>(29)</td>
<td>(1.50)</td>
<td>(37)</td>
<td>(1.68)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>1.5 ns</td>
<td>4.2*</td>
<td>2.3 ns</td>
<td>1.0 ns</td>
<td>3.4*</td>
<td>&lt;1 ns</td>
<td>1.7 ns</td>
<td></td>
</tr>
<tr>
<td>HC vs CHR</td>
<td>0.36 ns</td>
<td>0.15 ns</td>
<td>0.33 ns</td>
<td>0.12 ns</td>
<td>0.37#</td>
<td>0.06 ns</td>
<td>0.21#</td>
<td></td>
</tr>
<tr>
<td>HC vs FA</td>
<td>0.40 ns</td>
<td>0.67*</td>
<td>0.47#</td>
<td>0.49#</td>
<td>0.22#</td>
<td>0.14#</td>
<td>0.10 ns</td>
<td>0.17#</td>
</tr>
<tr>
<td>FA vs CHR</td>
<td>0.02 ns</td>
<td>0.53*</td>
<td>0.11 ns</td>
<td>0.39#</td>
<td>0.13#</td>
<td>0.50**</td>
<td>0.17#</td>
<td>0.39*</td>
</tr>
</tbody>
</table>

*Note.* HC = healthy control group. CHR = chronic schizophrenia group. FA = first-admission schizophrenia group. Top three rows: For both sensor and source data, scores reflect amplitudes relative to 1000 ms pre-S1 baseline in fT for sensor data and in arbitrary units for source data. Scores are presented as means and standard deviations (in parentheses).

ANOVA: scores indicate F-values from three-group ANOVA with F(2,118). Two-group comparisons (HC vs CHR, HC vs FA, FA vs CHR): scores indicate effect sizes in Hedge’s g, superscript significance from t-tests with t(85), t(62), and t(92), respectively.

Bottom four rows: Level of significance for group comparisons are indicated by superscript as ns: not significant when p > .1, # p < .1, * p < .05, ** p < .01.
Supplementary Table 2.4

Metrics and group comparisons (as in Table 2.2) for M100 scores instead of M50

<table>
<thead>
<tr>
<th>Sensor Source</th>
<th>M100-Baseline Ratio</th>
<th>M50-M100 Ratio</th>
<th>S1-S2 M100 Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>0.51 (0.23)</td>
<td>0.49 (0.16)</td>
<td>111 # (71)</td>
</tr>
<tr>
<td></td>
<td>0.58 (0.19)</td>
<td>0.53 (0.24)</td>
<td>3.6 (2.0)</td>
</tr>
<tr>
<td>CHR</td>
<td>0.61 (0.26)</td>
<td>0.61 (0.22)</td>
<td>81 3.1 (69)</td>
</tr>
<tr>
<td></td>
<td>0.64 (0.22)</td>
<td>0.67 (0.26)</td>
<td>3.1 (2.1)</td>
</tr>
<tr>
<td>FA</td>
<td>0.60 (0.30)</td>
<td>0.59 (0.18)</td>
<td>80 2.4 (67)</td>
</tr>
<tr>
<td></td>
<td>0.64 (0.20)</td>
<td>0.71 (0.30)</td>
<td>2.4 (1.9)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>1.38&lt;sup&gt;ns&lt;/sup&gt;</td>
<td>0.96&lt;sup&gt;ns&lt;/sup&gt;</td>
<td>3.67* 3.9*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.20&lt;sup&gt;ns&lt;/sup&gt;</td>
<td>2.39#</td>
</tr>
<tr>
<td>HC vs CHR</td>
<td>0.39# 0.29&lt;sup&gt;ns&lt;/sup&gt;</td>
<td>0.60** 0.55*</td>
<td>0.43# 0.24&lt;sup&gt;ns&lt;/sup&gt;</td>
</tr>
<tr>
<td>HC vs FA</td>
<td>0.33&lt;sup&gt;ns&lt;/sup&gt;</td>
<td>0.54*</td>
<td>0.46# 0.56*</td>
</tr>
<tr>
<td>FA vs CHR</td>
<td>0.03&lt;sup&gt;ns&lt;/sup&gt;</td>
<td>0.13&lt;sup&gt;ns&lt;/sup&gt;</td>
<td>0.02&lt;sup&gt;ns&lt;/sup&gt; 0.30&lt;sup&gt;ns&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Note. As in Supplementary Table 2.1A. M50-M100 ratio is unchanged from main text Table 2.2 and is repeated to facilitate comparison.
### Supplementary Table 2.5A

**Peak latencies (in ms)**

<table>
<thead>
<tr>
<th></th>
<th>S1 M50 Sensor</th>
<th>S1 M50 Source</th>
<th>S1 M100 Sensor</th>
<th>S1 M100 Source</th>
<th>S2 M50 Sensor</th>
<th>S2 M50 Source</th>
<th>S2 M100 Sensor</th>
<th>S2 M100 Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HC</strong></td>
<td>61 (11)</td>
<td>57 (13)</td>
<td>111 (11)</td>
<td>99 (15)</td>
<td>570 (21)</td>
<td>565 (22)</td>
<td>598 (22)</td>
<td>598 (21)</td>
</tr>
<tr>
<td><strong>CHR</strong></td>
<td>57 (11)</td>
<td>61 (17)</td>
<td>108 (13)</td>
<td>101 (17)</td>
<td>564 (21)</td>
<td>570 (23)</td>
<td>596 (21)</td>
<td>591 (26)</td>
</tr>
<tr>
<td><strong>FA</strong></td>
<td>59 (13)</td>
<td>61 (15)</td>
<td>107 (12)</td>
<td>96 (18)</td>
<td>577 (25)</td>
<td>569 (22)</td>
<td>587 (21)</td>
<td>597 (2)</td>
</tr>
</tbody>
</table>

ANOVA: 1.2<sup>ns</sup> 1<sup>ns</sup> 1.1<sup>ns</sup> 1.0<sup>ns</sup> 3.5<sup>*</sup> 3.7<sup>*</sup> 2.3<sup>ns</sup> 1<sup>ns</sup>

<table>
<thead>
<tr>
<th></th>
<th>HC vs CHR</th>
<th>HC vs FA</th>
<th>FA vs CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.37&lt;sup&gt;ns&lt;/sup&gt; 0.27&lt;sup&gt;ns&lt;/sup&gt; 0.26&lt;sup&gt;ns&lt;/sup&gt; 0.16&lt;sup&gt;ns&lt;/sup&gt; 0.30&lt;sup&gt;ns&lt;/sup&gt; 0.20&lt;sup&gt;ns&lt;/sup&gt; 0.08&lt;sup&gt;ns&lt;/sup&gt; 0.29&lt;sup&gt;ns&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.19&lt;sup&gt;ns&lt;/sup&gt; 0.32&lt;sup&gt;ns&lt;/sup&gt; 0.38&lt;sup&gt;ns&lt;/sup&gt; 0.14&lt;sup&gt;ns&lt;/sup&gt; 0.26&lt;sup&gt;ns&lt;/sup&gt; 0.20&lt;sup&gt;ns&lt;/sup&gt; 0.47&lt;sup&gt;#$&lt;/sup&gt; 0.07&lt;sup&gt;ns&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.15&lt;sup&gt;ns&lt;/sup&gt; 0.02&lt;sup&gt;ns&lt;/sup&gt; 0.09&lt;sup&gt;ns&lt;/sup&gt; 0.29&lt;sup&gt;ns&lt;/sup&gt; 0.55&lt;sup&gt;##&lt;/sup&gt; 0.01&lt;sup&gt;ns&lt;/sup&gt; 0.40&lt;sup&gt;ns&lt;/sup&gt; 0.23&lt;sup&gt;ns&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** HC = healthy control group. CHR = chronic schizophrenia group. FA = first-admission schizophrenia group. Top three rows: For both sensor and source data, scores reflect peak latencies in ms. Scores are presented as means and standard deviations (in parentheses).

ANOVA: scores indicate F-values from three-group ANOVA with F(2, 118). Two-group comparisons (HC vs CHR, HC vs FA, FA vs CHR): scores indicate effect sizes in Hedge’s g, superscript significance from t-tests with t(85), t(62), and t(92), respectively. Bottom four rows: Level of significance for group comparisons are indicated by superscript as ns: not significant when p > .1, # p < .1, * p < .05, ** p < .01.
No significant sensor quantifications combining S1 and S2 scores constraining S2 latency to within 10 ms of S1 latency were found. Because higher high-pass filtering leads to better peak detection, the same 5 Hz high-pass filter used for the source data was used for the sensor data in Tables 2.5B and 2.5C.
Supplementary Table 2.5C

Peak latencies (in ms) constraining S2 latency to within 10 ms of S1 latency

<table>
<thead>
<tr>
<th></th>
<th>S1 M50</th>
<th>S1 M100</th>
<th>S2 M50</th>
<th>S2 M100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensor</td>
<td>Source</td>
<td>Sensor</td>
<td>Source</td>
</tr>
<tr>
<td>HC</td>
<td>62</td>
<td>57</td>
<td>105</td>
<td>99</td>
</tr>
<tr>
<td>(12) (13)</td>
<td>(10) (15)</td>
<td>(13) (15)</td>
<td>(12) (15)</td>
<td></td>
</tr>
<tr>
<td>CHR</td>
<td>60</td>
<td>61</td>
<td>105</td>
<td>101</td>
</tr>
<tr>
<td>(12) (17)</td>
<td>(14) (17)</td>
<td>(12) (18)</td>
<td>(14) (19)</td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>62</td>
<td>61</td>
<td>101</td>
<td>96</td>
</tr>
<tr>
<td>(11) (15)</td>
<td>(13) (18)</td>
<td>(11) (15)</td>
<td>(14) (17)</td>
<td></td>
</tr>
</tbody>
</table>

ANOVA <1ns <1ns 1.2ns 1.0ns <1ns <1ns 1.1ns 2.2ns

<table>
<thead>
<tr>
<th></th>
<th>HC vs</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC vs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHR</td>
<td>0.16ns</td>
<td>0.27ns</td>
</tr>
<tr>
<td>HC vs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>0.03ns</td>
<td>0.32ns</td>
</tr>
<tr>
<td>FA vs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHR</td>
<td>0.13ns</td>
<td>0.02ns</td>
</tr>
</tbody>
</table>

Note. As in Supplementary Table 2.5A.

No significant sensor quantifications combining S1 and S2 scores constraining S2 latency to within 10 ms of S1 latency were found. Because higher high-pass filtering leads to better peak detection, the same 5 Hz high-pass filter used for the source data was used for the sensor data in Tables 2.5B and 2.5C.
Filter settings have varied considerably in the sensory gating literature (de Wilde et al., 2007; Patterson et al., 2008) and have generally been under-specified, as noted in the method section (see also Edgar et al., 2005). Present filters were selected to minimize filtering and the distortion it causes (for high-pass: low cutoff and low order; for low-pass: high cutoff) while removing slow drift and high-frequency noise. Extensive exploratory work was undertaken to select the present filters. First, several high-pass cutoff frequencies were evaluated based on those commonly appearing in the studies in the meta-analysis of de Wilde et al. (2007): 0.1 Hz, 1 Hz, 5 Hz, 10 Hz. Second, a high-pass FIR filter with a maximum filter order of 300 (thus spanning 442 ms on each side of the filtered point) was chosen to minimize blurring of S1 and S2 responses. The filter was applied in forward and reverse directions (and thus with symmetrical weights) both to allow a lower-filter order (since this doubles the effective filter order) and to avoid having a “biased” one-sided blurring or phase shift. Filtering only in one direction with otherwise identical filter characteristics would double the blur in this direction while eliminating it in the other, meaning that a forward filter would blur S2 responses more than S1 responses. Another tradeoff for filtering in both directions is that the baseline is distorted by the backward part of the filter, which was compensated for in this study by the long 1000 ms baseline. Third, considerable testing was done to identify the lowest cutoff that successfully removed baseline drift while minimizing waveform distortion. The 1 Hz cutoff met that criterion for sensor data, although evaluation of its gain function, after application in the forward and reverse directions, indicated that the cutoff was actually 1.7 Hz. Source data needed a higher (5 Hz) cutoff to eliminate baseline drift. Such low-frequency noise is characteristically higher in MEG than in EEG, and S2 is more vulnerable to it than is S1 when using the pre-S1 baseline for both responses, because S2
is more distant from the pre-S1 baseline. A possible confound of cutoff Hz with sensor vs. source analyses was unlikely, because the exploratory work at 5 Hz cutoff with sensor-space data produced results very similar to those at 1 Hz cutoff (see Table 2.2 and Supplementary Table 2.5). Although Kanno et al. (2000) noted that a high-pass filter of higher than 3 Hz might artificially increase M50 amplitude (while decreasing M100 amplitude), a different explanation might be possible: Königs and Gutschalk (2012) showed that the M50 time interval partly overlaps with the M100, meaning that M100 might mask some of M50 strength. Since M100 has lower frequency components than M50, a high-pass filter might not artificially increase the M50 but rather restore it to its original strength, unmasked by the filtered-out M100. This hypothesis remains to be substantiated by additional studies, especially simulations.


Supplementary Table 2.6

**Metrics for M40-M50 peak to peak scoring**

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Source</th>
<th>M40-M50 Ratio</th>
<th>S1 M40 latency Sensor</th>
<th>S1 M40 latency Source</th>
<th>S2 M40 latency Sensor</th>
<th>S2 M40 latency Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>0.52</td>
<td>0.52</td>
<td>38</td>
<td>39</td>
<td>542</td>
<td>540</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.82)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>CHR</td>
<td>0.68</td>
<td>0.68</td>
<td>38</td>
<td>38</td>
<td>541</td>
<td>539</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(2.43)</td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(5)</td>
</tr>
<tr>
<td>FA</td>
<td>0.66</td>
<td>0.32</td>
<td>38</td>
<td>38</td>
<td>540</td>
<td>540</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(1.25)</td>
<td>(2)</td>
<td>(2)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
</tbody>
</table>

ANOVA: 3.96* <1<sup>ns</sup> 1.06<sup>ns</sup> 1.33<sup>ns</sup> 1.73<sup>ns</sup> <1<sup>ns</sup>

HC vs CHR: 0.69** 0.08<sup>ns</sup> -0.17<sup>ns</sup> -0.37<sup>ns</sup> -0.32<sup>ns</sup> -0.17<sup>ns</sup>

HC vs FA: 0.51* -0.18<sup>ns</sup> 0.12<sup>ns</sup> -0.23<sup>ns</sup> -0.43# 0.02<sup>ns</sup>

FA vs CHR: -0.09<sup>ns</sup> -0.17<sup>ns</sup> 0.33<sup>ns</sup> 0.12<sup>ns</sup> -0.15<sup>ns</sup> 0.19<sup>ns</sup>

*Note. As in Supplementary Table 2.1A (left) and 2.5A (right).*

No significant quantifications for M40-M50 peak-to-peak scoring were found when searching in one latency window for M40 and M50 simultaneously (e.g. searching within 35-60 ms after click onset) either M40 or M50 could not reliably be scored, depending on the searched latency-window. Therefore, 35-50 ms was searched for the most negative and most positive peak and the earlier of the two was defined as M40. As 35-50 ms might only contain the rising flank of M50 and not necessarily capture the peak of M50, M50 scores were taken from the original scoring (searching 40-130 ms for the most positive and negative peak and defining the earlier of the two as M50). This ensures that the flank of the M100 is not mistaken as M40, as both share the same polarity.
Supplementary Table 2.7A

**Metrics and group comparisons for M50 quantifications per hemisphere**

<table>
<thead>
<tr>
<th>Sensor Source</th>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
<th>Both hemispheres</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td>0.48 (0.26)</td>
<td>0.54 (0.41)</td>
<td>0.51 (0.25)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td>0.63 (0.31)</td>
<td>0.67 (0.27)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td></td>
<td>0.48 (0.19)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td></td>
<td>0.56 (0.21)</td>
</tr>
<tr>
<td><strong>CHR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td>0.68 (0.29)</td>
<td>0.68 (0.29)</td>
<td>0.70 (0.49)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td>0.67 (0.27)</td>
<td>0.67 (0.26)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td></td>
<td>0.65 (0.23)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td></td>
<td>0.65 (0.23)</td>
</tr>
<tr>
<td><strong>FA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td>0.63 (0.26)</td>
<td>0.73 (0.37)</td>
<td>0.69 (0.45)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td>0.60 (0.25)</td>
<td>0.60 (0.28)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td></td>
<td>0.63 (0.21)</td>
</tr>
<tr>
<td>M50-Baseline Ratio</td>
<td></td>
<td></td>
<td>0.63 (0.21)</td>
</tr>
</tbody>
</table>

ANOVA: 5.03** 2.7# 2.05ns <1ns 4.46* 1.54ns

HC vs CHR: 0.71** 0.43# 0.45* 0.12ns 0.70** 0.40#

HC vs FA: 0.58* 0.50# 0.48* 0.11ns 0.60* 0.31ns

FA vs CHR: 0.16ns 0.16ns 0.02ns 0.24ns 0.07ns 0.10ns

**Note.** As in Supplementary Table 2.1A.

Supplementary Table 2.7B

**Metrics and group comparisons for M50 quantifications per hemisphere**

<table>
<thead>
<tr>
<th>Sensor Source</th>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
<th>Both hemispheres</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td>0.49 (0.24)</td>
<td>0.53 (0.24)</td>
<td>0.54 (0.19)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td>0.59 (0.20)</td>
<td>0.65 (0.25)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td></td>
<td>0.49 (0.16)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td></td>
<td>0.55 (0.15)</td>
</tr>
<tr>
<td><strong>CHR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td>0.62 (0.23)</td>
<td>0.67 (0.26)</td>
<td>0.65 (0.29)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td>0.65 (0.25)</td>
<td>0.65 (0.22)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td></td>
<td>0.61 (0.20)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td></td>
<td>0.61 (0.20)</td>
</tr>
<tr>
<td><strong>FA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td>0.60 (0.21)</td>
<td>0.71 (0.30)</td>
<td>0.61 (0.23)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td>0.63 (0.23)</td>
<td>0.63 (0.23)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td></td>
<td>0.59 (0.18)</td>
</tr>
<tr>
<td>M50-M100 Ratio</td>
<td></td>
<td></td>
<td>0.64 (0.20)</td>
</tr>
</tbody>
</table>

ANOVA: 3.15* 3.9* 1.85ns <1ns 3.67* 1.76ns

HC vs CHR: 0.54* 0.55* 0.42* 0.22ns 0.60** 0.32ns

HC vs FA: 0.49# 0.65* 0.35ns 0.14ns 0.54* 0.49*

FA vs CHR: 0.07ns 0.16ns 0.13ns 0.08ns 0.13ns 0.14ns

**Note.** As in Supplementary Table 2.1A.
**Supplementary Table 2.7C**

*Metrics and group comparisons for M50 quantifications per hemisphere*

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Source</th>
<th>S1-S2 M50 Difference</th>
<th>S1-S2 M50 Difference</th>
<th>S1-S2 M50 Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Left hemisphere</td>
<td>Right hemisphere</td>
<td>Both hemispheres</td>
</tr>
<tr>
<td>HC</td>
<td></td>
<td>81.1 (57.1)</td>
<td>86.7 (63.8)</td>
<td>83.9 (48.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.42 (2.79)</td>
<td>4.37 (3.64)</td>
<td>3.90 (2.50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44.6 (44.6)</td>
<td>59.8 (65.0)</td>
<td>52.2 (41.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.12 (2.25)</td>
<td>4.48 (3.97)</td>
<td>3.30 (2.52)</td>
</tr>
<tr>
<td>CHR</td>
<td></td>
<td>48.1 (40.8)</td>
<td>64.2 (82.0)</td>
<td>56.1 (46.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.50 (2.32)</td>
<td>3.83 (3.28)</td>
<td>2.66 (2.01)</td>
</tr>
<tr>
<td>FA</td>
<td></td>
<td>4.18**</td>
<td>1.43&lt;sub&gt;ns&lt;/sub&gt;</td>
<td>4.96**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.12**</td>
<td>&lt;1&lt;sub&gt;ns&lt;/sub&gt;</td>
<td>2.11&lt;sub&gt;ns&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

**ANOVA**

- HC vs CHR: 0.74**
- HC vs FA: 0.67*
- FA vs CHR: 0.08<sub>ns</sub>

*Note*. As in Supplementary Table 2.1A.
3 Study 2: Connectivity Dynamics within Auditory – Frontal Cortex Networks in Sensory Gating Context

3.1 Introduction

Connectivity dynamics in large-scale brain networks are addressed to decipher neuronal mechanisms serving psychological functions (Menon, 2011). Different connectivity measures delineate functional, task-related, default mode, or salience networks as neuronal substrate of perceptual, cognitive or affective functions (review Bressler & Menon, 2010; Menon, 2011; see also Braga et al., 2016). Hence, network connectivity measures served to understand perceptual, cognitive or affective dysfunction in psychiatric disorders (e.g., Woodcock et al., 2016; Ohtani et al., 2015; van Dellen et al., 2016; Repovs et al., 2011; Tu et al., 2012; LeRoux et al., 2014; Lavigne et al., 2015; Menon, 2011). Functional disconnection or “dysconnectivity” has been proposed to characterize schizophrenia (Ohtani et al., 2016; Repovs et al., 2011; van Dellen et al., 2016) with reduced connectivity between frontotemporal and frontoparietal regions as main sources (Menon, 2011, Ohtani et al., 2015; Woodcock, Wadehra, & Diwadkar, 2016; Lavigne et al., 2015; Leroux, Delcroux, & Dollfus, 2014; Tu et al., 2012, 2013; van Dellen et al., 2016; Repovs, Csernansky, & Barch, 2011). Moreover, perceptual and cognitive deficits in schizophrenia patients has been attributed to dysfunctional stimulus discrimination, which becomes manifest in robust measures like abnormal sensory gating (the smaller than normal difference between S1- and
S2-evoked P/M50 or larger-than-normal S2/S1 ratio of P/M50; overviews, for instance, Hanlon et al., 2005; Boutros et al., 2004; Heinrich, 2004; Wheeler et al., 2015; Patterson et al., 2008, Zhou et al., 2015; Li et al., 2016; Chen et al. 2013).

While this hypothesis is widely accepted and supported by abnormal activity in primary auditory cortex and abnormal activity in auditory-frontal networks associated with “gating”, the nature of network abnormality implementing normal and abnormal auditory gating, in particular the frontocortical influence on the auditory cortex processing of identical stimulus pairs remain unclear.

In particular, information flow within an assumed “auditory-frontal gating” network, has to be specified. Connectivity within and between auditory and frontal regions, defining a ‘gating network’, should elucidate the fronto-auditory interaction enabling gating. While hemodynamic imaging studies found simultaneous activity of both regions, frontal cortex and auditory cortex (Grunwald et al. 2003; Tregellas et al. 2007; Tu et al. 2013; Bak et al. 2013), electroencephalographic specification of temporal dynamics should inform the information flow between the components of the supposed network enabling gating.

Auditory sensory gating is commonly defined by the attenuation of electromagnetic brain responses to the second of two identical stimuli (S1, S2) presented in rapid succession (Rentzsch et al., 2008; Chang et al. 2011; Patterson et al., 2007). Scalp topography of the electroencephalographic (EEG) or magnetoencephalographic (MEG) auditory gating index, the P50 or M50 ratio, verified activity over the superior temporal gyrus (STG; Edgar et al., 2003). Source analyses indicated activity in a distributed network including auditory/temporal, frontal and parietal regions (Knott, Millar & Fisher, 2009; Korzyukov et al. 2007; Williams et al. 2011; Weiland et al., 2008). Structural and functional connectivity between these regions were verified in animal studies (Barbas 2004; Knight et al. 1999;
Romanski et al. 1999), and by human diffusion tensor imaging (DTI, Catani et al. 2002). Impaired top-down (frontal) influence on auditory gating was concluded from impaired auditory gating in patients with frontal-lobe lesions (Knight et al., 1989; Golubic et al., 2014). The interaction within this “gating network”, in particular the direction of information flow and the frontal (“top down”) influence on auditory gating processes remains to be verified. While Joos and colleagues (2014) proposed that stimuli in the paired-click task are processed simultaneously in auditory and frontal regions, lower (alpha) and higher (beta) frequency oscillations of frontal cell assemblies have been shown to exert a modulatory influence on sensory processes (e.g. Fries, 2015; Arnal et al. 2011; Foxe 2011). Sensory gating has been associated with different regions, including the dorsolateral prefrontal cortex (DLPFC), the hippocampus and the thalamus (Tregellas et al., 2007), but the anterior cingulate cortex (ACC) has been most prominently mentioned for its dysfunction in schizophrenia, including e.g. sensory input discrimination (for a review see Adam & David, 2007). As a wide variety of sources are being reported, the present analysis considered only the most often mentioned ACC as a theory-driven source. Additionally, it was evaluated if this matched the sources, which are actually manifest in the present task and MEG data as a data-driven approach.

Spontaneous, pre-stimulus (network) activity may reflect states that influence subsequent task-related activity (see Baenninger et al., 2016, for “temporally coherent networks” (TCN)). Thus, an “auditory-frontal gating” network may be evident prior to stimulus-onset. Therefore, and to provide stationary data, connectivity was determined during the period prior to the paired clicks and compared to post-stimulus connectivity.
Granger causality analysis of magnetoencephalographic source activity served to index information flow within the (hypothetical) “auditory-frontal gating” network, while the (hypothetical) frontal top-down control of auditory gating examined by correlation with the M50 gating difference score (S1-minus-S2). Although Granger analysis does not directly establish causal direction, this time-lagged correlational method suggests temporal precedence of events and is used to disambiguate causality: According to Granger (1969) time series A can be assumed to cause time series B if time series B is (significantly) better predicted when using information from time series A. Specific hypotheses were

(1) Frontal activity in alpha-beta frequency ranges predicts oscillatory source activity in the auditory cortex.

(2) This dynamic correlates with the auditory gating index.

(3) Reduced top-down control has been speculated for SZ, therefore results obtained from Granger causality analysis were compared between the healthy sample and a sample of schizophrenia patients (SZ) with the hypothesis that the frontal-to-auditory information flow was smaller in SZ compared to H.

(4) As a consequence, the impact on auditory gating was expected to be reduced in schizophrenia patients, manifest in lacking correlation.

3.2 Method

3.2.1 Participants

The present analyses include data from $n = 77$ healthy participants (H), and $n = 58$ schizophrenia patients (SZ), who were recruited within a study on cognitive training effects on sensory gating and test performance (results unrelated to those reported here have been reported in Popov et al. 2011; Carolus et al.)
2014; Popova et al., 2015). H were screened for current or lifetime psychiatric diagnoses with the M.I.N.I (International Neuropsychiatric) Interview (Ackenheil et al. 1999). SZ met ICD-10 (version 2011) criteria for schizophrenia spectrum disorders. Patients were recruited from inpatient units at the local center for psychiatry. For both groups, exclusion criteria included a history of neurological conditions or disorders, including epilepsy or head trauma with a loss of consciousness.

Table 3.1 summarizes demographic information of the sample. Groups differed in age (F(1,128) = 32.51, p < 0.001; SZ being older than H), and years of school education (F(1,127) = 5.53, p = 0.02; more in H than SZ), while gender distribution and handedness did not differ between groups.

All participants accomplished a paired-click task while the magnetoencephalogram (MEG) was monitored. Structural magnetic resonance imaging (sMRI) was measured from n = 42 H and n = 21 SZ in a separate session.

*Data assessment and analysis:* The paired-click task comprised 100 pairs of 3 ms square-wave clicks as S1 and S2 presented with 500 ms onset-to-onset interstimulus intervals and a variable offset to onset intertrial interval of 7 – 9 s. Clicks were presented 60 dB above the individual hearing level and delivered via 5 m non-ferromagnetic tubes. Participants were asked to keep their eyes focused on a small fixation point throughout the procedure.
Table 3.1

Demographic information for healthy participants \(n = 72\) and schizophrenia patients \(n = 58\) and statistical effects of group differences

<table>
<thead>
<tr>
<th></th>
<th>Healthy participants</th>
<th>Schizophrenia patients</th>
<th>Statistical Group differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (M±SD)</td>
<td>28.33±7.66</td>
<td>36.78±9.22</td>
<td>(F_{1,128} = 32.51^{**})</td>
</tr>
<tr>
<td>Gender (m/f)</td>
<td>44/28</td>
<td>39/19</td>
<td>(\text{Chi}^2 \ p &gt; 0.5)</td>
</tr>
<tr>
<td>school education (M±SD years)</td>
<td>15.75±2.64</td>
<td>14.49±3.44</td>
<td>(F_{1,128} = 5.53^{*})</td>
</tr>
<tr>
<td>LQ</td>
<td>72.81±56.59</td>
<td>69.29±53.38</td>
<td>(F &lt; 1)</td>
</tr>
</tbody>
</table>

Note. m: male; f: female, RH: right-handed per Laterality quotient (LQ) +80-100; LH: left-handed per laterality quotient -80-100; ambi: ambidexter per Laterality quotient -60 - +60. **: \(p < 0.001\); *: \(p < 0.05\).

3.2.2 Data assessment

Data assessment followed standard procedures (detailed in Carolus et al., 2014, Popov et al., 2011, 2015). Prior to MEG measurement, individual hearing levels were determined for each ear. The MEG was recorded using a 148-channel whole-head magnetometer (MAGNES 2500 WH, 4D Neuroimaging, San Diego, USA) while subjects were in a supine position. Data were continuously recorded with a sampling rate of 678.17 Hz and a bandpass filter of 0.1 to 200 Hz. The subject’s nasion, left and right ear canal, and head shape were digitized using a Polhemus 3Space FastTrack prior to each session.

MEG epochs of 2000 ms before and 2000 ms after each S1 onset were extracted from continuous recordings. Before correcting for heart and eye-blink artifacts via independent component analysis, trials with extreme variance of data points were removed as contaminated by movement artifacts or SQUID jumps.
Global noise was removed offline by subtracting external, non-biological noise recorded by eleven MEG reference channels. Offline treatment of the MEG signals was accomplished using the MATLAB-based open-source signal processing toolbox FieldTrip (Oostenveld et al. 2011), complemented by in-house MATLAB functions. Data were not subjected to a digital frequency filter at this stage to avoid distortion in the Granger analysis.

3.2.3 Source reconstruction

Source reconstruction was based either on individual structural magnetic resonance images that were obtained either using a Philips Gyroscan ACS-T 1.5 T (256×256 mm field of view, 1x1mm in-plane resolution, 200 sagittal slices and 1x1x1mm resulting voxel dimensions or on an affine transformation of an MNI-template brain (Montreal Neurological Institute (MNI), Montreal, Canada; http://www.bic.mni.mcgill.ca/brainweb) to the participant’s digitized individual head shape (see also Lecaignard et al. 2008; Keil et al. 2010). MEG source reconstruction employed an automated approach using individual head shapes and individual structural MRIs co-registered to the MEG coordinate system via NUT-MEG (Neurodynamic Utility Toolbox for Magnetoencephalography; Dalal et al. 2004). A realistic, single-shell brain model (Nolte, 2003) was constructed for each subject. Results for individual subjects were normalized to a common brain template for illustration (e.g., source grand averaging) and for statistical comparisons. A voxel size of 10 mm$^3$ was used for the leadfield computation.

A distributed source model was estimated using a Linearly-Constrained Minimum Variance (LCMV) beamformer algorithm (Van Veen et al. 1997). Automatic anatomical parcellation of brain areas was employed, one analysis focusing the analysis on two areas representing bilateral Heschl's gyri according to the
MNI atlas (Tzourio-Mazoyer et al. 2002). This a priori definition of region of interest (ROI) was confirmed by determining sources of evoked magnetic fields during 200 ms after S1 and 200 ms after S2-onset in both Heschl’s gyri and calculating the difference (S1 minus S2) source activity as index of M50 gating: The difference between S1 and S2 peaked at both Heschl’s gyri and a frontal source. This third source was located in the middle cingulate cortex (MNI atlas label “cingulum mid right” or CMR; from hereon named frontal source for the sake of brevity, see Figure 3.1 A). Source activity was determined for all voxels within each of the MNI-defined ROIs. This resulted in two voxels the right-hemispheric Heschl’s gyrus ROI (HGR), three voxel in the left Heschl’s gyrus ROI (HGL), and 15 voxel in the mid-cingulate ROI (CMR). (Because of hemispheric asymmetry of the primary auditory cortices, the right-hemispheric ROI comprised less voxel than the left-hemispheric ROI. The mean source strength over all voxels per ROI and event related fields (ERF) was used as measure of source activity within an ROI for the subsequent connectivity analyses.) For a virtual sensor in each of these three regions, magnetic field strength was calculated over a time window of 2000 ms before and 2000 ms after S1 (Figure 3.1 B).

3.2.4 Auditory sensory gating

The time series derived from the source activity of the temporal ROIs HGL and HGR were filtered with a 5 Hz high-pass filter (finite-impulse-response, order 300) and an 80 Hz low-pass filter (infinite-impulse-response, order 6). M50 and M100 peaks were automatically scored as the largest peak and the largest trough 40 – 130 ms after S1 and after S2. Because of arbitrary polarity in MEG source reconstruction M50 was defined as the first of two peaks if the second peak
showed reversed polarity. The difference (S1 minus S2) was identified as gating index.

3.2.4 Directional connectivity

Directional connectivity between auditory and frontal sources was analyzed with a nonparametric variant of bivariate Granger causality. The time-lagged correlation analysis suggests temporal precedence of events (Dhamala et al. 2008) and is used as index of the direction of information flow, measured as source activity in one ROI predicting activity in another ROI. In the frequency domain, Granger causality is based on the spectral transfer function and the noise covariance matrix, which can be obtained either from an autoregressive model or non-parametrically from a factorization of the spectral matrix. In the present analyses, nonparametric Granger causality was computed from Fourier spectra with a factorization based on the Wilson algorithm (Wilson et al. 2007), which avoids assumptions about specific autoregressive model order. Because Granger causality is sensitive to temporal order, reversing the time axis should also reverse the direction of ‘true’ information flow, while spurious causality due, for instance, to volume conduction or channel differences in signal-to-noise ratio (e.g., Nolte et al. 2008) should not be affected by a reversal of the time axis (see also Haufe et al. 2012). For each ROI and each subject, the forward spectrum during the 2000 ms prior to S1 was compared with the time reversed spectrum using two-tailed dependent-sample t-tests for successive 0.25 Hz frequency bins between 7 and 30 Hz. (Analysis of 0.25 Hz bins was enabled by additional 2000 ms mirror data padding within the 2000 ms window.) Figure 3.1 illustrates this comparison for the present results for source activity in Heschl’s gyrus (‘auditory’) predicting source activity in mid cingulate ROI (‘frontal’) and vice versa. The comparison of
Granger causality (GC) scores obtained from forward and reversed time series across frequency bins (abscissa) shows frequency ranges of overlap and of differences. According to Hauffe et al. (2013) frequencies of similar GC scores from forward and reverse time series indicate influence of spurious sources, and frequency ranges, in which GC scores differ, suggest ‘true’ information flow. Accordingly, granger causality spectra were transformed into difference scores (forward time series minus reversed time series) as estimate of true information flow. Finally, frontal to auditory information flow was combined with the auditory to frontal information flow as both represent the same connectivity and only differing in sign so that positive difference scores indicate frontal-to-auditory information flow and negative difference scores auditory-to-frontal information flow (Figure 3.1 C):

\[
\frac{(F \rightarrow A)^{t} - (F \rightarrow A)^{-t} - (A \rightarrow F)^{t} - (A \rightarrow F)^{-t}}{2}
\]


### 3.2.4 Inferential statistics

Frequencies bins of significant GC difference scores were determined with nonparametric statistical testing by calculating Monte-Carlo estimates of the significance probabilities from 1000 permutations of the distribution to correct for non-normality of the values as implemented in the FieldTrip Toolbox (Oostenveld et al., 2011). Hypothesis (1) predicting dominant frontal-to-auditory information flow was tested by comparing these values by two-tailed t-tests for dependent samples. The same analysis was completed for the stimulus-related window using a ‘frontal’ (cingulum mid right, CMR) and an ‘auditory’ (right Heschl’s gyrus, HGR) source.
Per hypothesis (2), frequencies (of the successive 0.25 Hz bins during the 2000-ms pre-S1 time window) with significant GC difference scores were correlated with the auditory gating index (S1-minus-S2 M50 ERF). Because of the non-normal distribution of scores two-tailed Spearman's rank correlation coefficients were determined. Hypothesis (3) on less information flow in SZ than H was tested with results for H. GC differences between H and SZ were examined by the same Monte-Carlo procedure like (1) and then tested with two-tailed t-tests for independent samples.

In a final step, a larger network was examined to explore the extent of the hypothetical “auditory gating network”. Granger causality analysis was repeated for bilateral cingular and auditory sources (CML, CMR, HGL, HGR), and the left and right anterior cingulum (per atlas CAR and CAL).

For methodological evaluation, a final analysis evaluated whether results depend on the restriction of the source fed into the granger analysis (maximum or adjacent regions).³

### 3.3 Results

Testing S1 source activity versus S2 source activity revealed significant differences in both heschl’s gyri for both H and SZ. For H a significant difference for the middle cinguli atlas regions were found with a peak in the right middle

---

³ When analyzing both hemispheres, the left and right parts of the cingulum can be represented either separately (as CMR/CML and CAR/CAL) or combined (as CM and CA), because they are bordering on each other and the resolution might not be accurate enough to separate them fully. While the separation of the cinguli gives a better estimation of hemispheric differences, a combined cingulum might have a better signal to noise ratio (provided there are only little hemispheric differences).
cingulum, for SZ these regions showed a peak as well but below significance threshold when cluster correcting for multiple comparisons. Compared to H, SZ showed similar differences but with lower overall significance and less dominance of the right hemisphere (Figure 3.1 A).

For the left heschl’s gyrus, the sensory gating index of the S1 M50 – S2 M50 difference for H ($M = 3.58 \times 10^{-9} \text{arb. unit}, SD = 2.86 \times 10^{-9} \text{arb. unit}$) and SZ ($M = 2.64 \times 10^{-9} \text{arb. unit}, SD = 2.18 \times 10^{-9} \text{arb. unit}$) differed significantly (t(132)=2.15, p=0.034). For the right heschl’s gyrus, the sensory gating index between H ($M = 5.70 \times 10^{-9} \text{arb. unit}, SD = 4.22 \times 10^{-9} \text{arb. unit}$) and SZ ($M = 5.69 \times 10^{-9} \text{arb. unit}, SD = 3.91 \times 10^{-9} \text{arb. unit}$) did not differ.

For H, granger causality measures between the strongest auditory (HGR) and frontal (CMR) source revealed a coherent characteristic with three peaks for 9 Hz (bottom-up), 15 Hz (top-down) and 19 Hz (bottom-up) and linear trends between them (Figure 3.1 C). Even though all three peaks are of equal strength, only the 19 Hz bottom-up peak was significantly different from zero when testing with nonparametric Monte Carlo estimates (from 19.25 – 19.5 Hz with p = 0.018).

For SZ, the granger causality measures fluctuated randomly without a significant difference from zero. The significant 19 Hz information flow of H differed significantly from SZ (p=0.044). It is unlikely that the absence of a significant information flow for SZ is just due to the marginally lower power of n=58 for SZ instead of n=77 for H because the sample-size independent effect size of this

---

9 The granger analysis corrected for this source strength difference by contrasting forward versus flipped time series, see method section.

10 Please note that the unit of the virtual sensors is in arbitrary units rather than tesla because the nonunitary gain of the beamformer was corrected by normalization which both provides unitary gain and removes units (Cheyne & Papanicolaou, 2015).
connection is much smaller as well: \( g=0.32 \) for H and \( g=-0.08 \) for SZ (in the opposite direction).

When correlating the sensory gating index of the M50 S1-S2 difference with auditory-to-frontal information flow, the significant 19 Hz bottom-up information flow of H showed a significant correlation with gating \( (r_s(75) = .318, p = 0.005, \text{Figure 3.1 D}) \). The (nonsignificant) 15 Hz top-down peak of H was also marginally correlating with the poststimulus sensory gating \( (r_s(75)=0.21, p=0.07) \), the (nonsignificant) 8 Hz peak showed no correlation.

As no information flow for SZ was found, no correlation with the sensory gating index was possible.
Table 3. 2

Extended network analysis

<table>
<thead>
<tr>
<th>Network</th>
<th>Time</th>
<th>Region</th>
<th>HC</th>
<th>SZ</th>
<th>HC vs CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMR – HGR</td>
<td>Prestim</td>
<td>CMR-HGR</td>
<td>↑ 19*</td>
<td>-</td>
<td>↑ 19*</td>
</tr>
<tr>
<td></td>
<td>Post-stim</td>
<td>CMR-HGR</td>
<td>↓ 29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CM – HGR/L</td>
<td>Prestim</td>
<td>CM-HGR</td>
<td>↑ 7</td>
<td>↓ 21-24 &amp; 27</td>
<td>↑ 7-10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CM-HGL</td>
<td>-</td>
<td>↓ 9</td>
<td>↓ 9</td>
</tr>
<tr>
<td>CMR/L – HGR/L</td>
<td>Prestim</td>
<td>CMR-HGR</td>
<td>↑ 19*</td>
<td>↑ 25</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CML-HGR</td>
<td>↑ 19*</td>
<td>↓ 10 &amp; 22</td>
<td>↓ 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CML-HGL</td>
<td>↑ 15*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CMR-HGL</td>
<td>↑ 15*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CAR/L – HGR/L</td>
<td>Prestim</td>
<td>CAR-HGR</td>
<td>↑ 12-15, 26-28</td>
<td>-</td>
<td>↑ 26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAL-HGR</td>
<td>↓ 15 &amp; 29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAL-HGL</td>
<td>↓ 21</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAR-HGL</td>
<td>↓ 8, ↑18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Post-stim</td>
<td>CAR-HGR</td>
<td>↑ 15</td>
<td>↑ 4-8</td>
<td>↑ 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAL-HGR</td>
<td>↓ 14-17</td>
<td>↑ 4-8</td>
<td>↓ 14-17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAL-HGL</td>
<td>-</td>
<td>↑ 14</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CAR-HGL</td>
<td>-</td>
<td>↑ 24</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note. Only when the frequency bins are wider than 2 Hz they are listed fully (even when extending outside the originally analyzed 7-30 Hz spectrum), otherwise only the mean frequency is listed. ↓ Top-down, ↑ bottom-up. HGR: Heschl’s gyrus right, HGL: Heschl’s gyrus left, CMR: mid cingulate right, CML: mid cingulate, CM: mid cingulate averaged over both hemispheres, CAR: cingulate anterior right, CAL: cingulate anterior left. Symbols: Green: HC > SZ. Red: SZ > HC. * correlates significantly with sensory gating within healthy participants.*
Figure 3.1: A: Statistical difference between S1 (0-200ms) and S2 (500-700ms) source strength. HGR: Heschl’s gyrus right, HGL: Heschl’s gyrus left, CMR: cingulum mid right. B: For the auditory gating task determined sources from A, the time course of activity in virtual sensors of H was plotted for three timewindows: “Pre”, the 2 second time window of prestimulus stationary activity (which was used as basis for connectivity measures), “Gating”, the 1 second poststimulus interval containing event related activity (including the S1 and S2 M50s which were used for the sensory gating difference) and “Post”, the 1 second time window of stationary activity after click onset from 1 – 2 seconds (which was used as basis for additional connectivity measures). S1 onset at 0s, S2 onset at 0.5s. C: Granger Causality values for CMR and HGR. Positive values indicate information flow from frontal to auditory, negative values from auditory to frontal. H: Healthy participants. SZ: Schizophrenia patients. Colored shaded areas signify standard errors. Grey shaded areas signify significant difference from zero for H. D: Correlation between prestimulus connectivity at 19 Hz and sensory gating for H.
These results were supplemented by the analyses involving additional network interactions (left-right HG, left-right CM, left-right CA, poststimulus time windows, see methods).

Generally, only one consistent pattern emerged for the various network-analyses which was the 4-node pre-stimulus CMR/L-HGR/L network showing bottom-up information flow for H in the lower beta-band which also correlated with sensory gating and was mirrored in the 2 node CMR-HGR network. Pre- and poststimulus networks showed little overlap and neither did HC and SZ (Table 3.2).

3.4 Discussion

The present study examined connectivity dynamics within auditory – frontal cortex networks in a sensory gating context for both healthy participants and schizophrenia patients.

Hypothesis (1), that frontal activity reflected in the alpha/beta frequency range influences oscillatory source activity in the auditory cortex preceding and during stimulus processing, was not confirmed. Instead, a connectivity in the opposite bottom-up direction was found in lower beta (19 Hz) frequencies. This index also correlated with the M50 sensory gating score, confirming hypothesis (2), that pre-stimulus connectivity predicted the auditory gating index.

Per Hypothesis (3) the auditory-to-frontal influence of healthy participants was absent in patients, who also did not show frontal top-down information flow. Likewise, schizophrenia patients showed no correlation between connectivity and sensory gating indices, supporting hypothesis (4).
While examining the strongest auditory and frontal nodes seemed to confirm almost all hypotheses at first, several major caveats have to be discussed: Firstly, the direction of information flow was contrary to the hypothesis, with pre-stimulus activity from the right heschl’s gyrus predicting the activity in the right mid cingulum instead of the other way around. One explanation for this could be, that this still reflects a frontal control network but in the preparatory stage: A bottom up information flow could be preparing the later top down control which is exerted only when the stimulus appears – but as this non-stationary poststimulus process cannot be analyzed using Granger causality (He & Maekawa, 2001), only the first part of this process is apparent.

Moreover, the specificity of the frequency band comprising only of two neighboring frequencies (19.25 & 19.5 Hz) might seem spurious. However, the frequency characteristic of the analyzed spectrum is not randomly fluctuating (like it does for the schizophrenia patients) but rather exhibiting a coherent characteristic with three peaks for 9, 15 and 19 Hz and linear trends between them (Figure 3.1 C). Thus, information flow may include neighboring frequencies but significant information flow is only seen at 19 Hz due to a poor signal to noise ratio. Moreover, Fries (2015) pointed out that multiple frequency rhythms can coexist and be nested within each other. Preceding the flank of the bottom-up 19 Hz peak is an equally strong 15 Hz top-down peak, which is not significant because of more variance. Yet, this top-down peak is also marginally correlating with the poststimulus sensory gating. One could speculate that there indeed is also a top-down influence nested within a bottom-up influence at different frequencies just border-line failing statistical levels because of insufficient signal to noise levels. Yet, it should also be acknowledged that present results only provide a weak hint at
such interactions and are by no means a definitive proof, especially because previous analyses with bidirectional connectivity in different frequency ranges found them to be further segregated (namely feedback in beta and feedforward in theta and gamma) (Bastos et al., 2015) and no study has found bottom-up beta connectivities before. Further analyses with greater N and a wider frequency range might strengthen (or refuse) these hints.

The predictive coding framework provides a promising explanation of the role of the frontal cortex in sensory gating: According to Knill and Pouget (2004), the “bayesian brain” tries to predict incoming information to facilitate processing it (Friston 2005, Hawkins and Blakeslee 2004, Mumford 1992, Rao and Ballard 1999). A model of expected sensory input is generated in higher cortical areas and modulates lower sensory areas (top-down). This model is updated by bottom-up information, e.g. by a mismatch of predicted and actual input (Rao & Ballard, 1999). Correspondingly, this is mostly studied via mismatch negativity in the auditory domain (e.g. Winkler and Czigler 2012), but the same principle leading to enhanced activity in mismatching tones can also be applied to suppressed activity in matching tones in this sensory gating paradigm.

Another caveat was that the sensory gating index did not differ between healthy participants and schizophrenia patients in the right heschl’s gyrus but only in the left. This hemisphere-specific gating deficit is known from other studies (J Christopher Edgar et al., 2008; Hirano et al., 2010; Thoma et al., 2003). However, it also makes a lack of information flow in schizophrenia patients difficult to relate to their illness when analyzing it only for right hemisphere, as cross-hemispheric connections are necessary to explain an involvement but are not analyzed here.

The simple network analysis focused on the pre-stimulus frontal and auditory regions with the biggest power difference between S1 and S2. However,
these regions might not necessarily be the ones with the most meaningful information flow for frontal control. When extending the network by also analyzing poststimulus time windows, the less powerful left hemisphere and the theoretically more meaningful frontal part of the cingulum, the pattern was less clear: Various information flows in both directions were found, neither converging between left and right auditory cortices nor between pre- and poststimulus time windows. The most promising (post-hoc) interpretation and entanglement of these complex network interactions might be external validation. As the study includes several independent analyses, they can be used as a cross-validation. In a first step, information flows in the healthy brain in candidate areas for sensory gating were identified. In a second step, only those information flows were considered which differed significantly between healthy controls and schizophrenia patients as only those may be meaningful for a disturbed network. In a third step, those differing information flows were related to the sensory gating index, which was recorded in an independent time window. In doing so, only one information flow fulfills all three criteria, which is the same lower beta, bottom-up information flow which was also found in the simple network analysis.

It is important to note, that while only testing granger causality in one group is subject to a multiple comparison problem as a wide array of frequencies, regions and time windows can potentially be tested, the external validation of the significant information flow is not. By reducing the tested frequencies only to those (eight) connectivities differing between healthy participants and schizophrenia patients the strength of the correlation (p=.005) remains significant when performing a bonferroni correction.

Results could disconfirm the commonly accepted model of top-down control - if the result could be substantiated and artificial reasons for lack of effect
can be ruled out! The assumption of top-down control is based on tasks other than paired-click. It is possible that this design prompts first of all bottom-up auditory processing, and that abnormal gating in SZ reflects abnormality on this fundamental level and not the commonly assumed hypofrontality or frontal disconnection. Only the comparison with other designs would rule out a task-specific result. Yet, connectivity was verified for the pre-stimulus interval, which should reflect a “default” rather than a “task-related” network activity. Why should frontal impact be evident constantly – unless the task-context primed the brain for preparatory frontal control in the anticipated task stimulus discrimination. Frontal information flow may be constrained to the challenge of frontal cognitive tasks.

Furthermore, comparing the simple and extended network analysis shows potential pitfalls of network analyses in general and especially Granger causality. As analyzing different networks in the context of sensory gating can be justified based on the literature or available data, interpreting only a single one can be misleading. Moreover, granger causality measures are only one of many different network analyses (Sakkalis, 2011), and other methods might also yield different results. Finally, the focus on alpha/beta frequencies was chosen because of a hypothesized top-down influence typically found in these frequencies (e.g. Arnal et al. 2011; Foxe 2011; Herzog et al. 2014; Fries, 2015; Skoblenick et al. 2016). As only a bottom-up information flow was found, which is typically expected in other (higher) frequencies (Buschman & Miller, 2007), different frequency windows might also be of interest.

Taken together this study provides a first glance at temporal dynamics of frontal involvement in auditory sensory gating, which appeared to be more complex than a simple top-down modulation, as only a bottom-up information flow
was found, discriminating both sensory gating efficiency in healthy participants and schizophrenia patients from healthy controls.
4 General Discussion

4.1 Summary

The present thesis examined various methods to study the sensory gating phenomenon in schizophrenia relative to patient groups in different stages of the disorder and to healthy controls. Compared to existing studies on sensory gating in schizophrenia, novel aspects of the present thesis were (a) to compare different methods in the same dataset instead of across studies and (b) to employ granger causality measures for a sensory gating paradigm in schizophrenia with MEG data.

Taken together results show that multiple preprocessing and quantification methods can yield vastly differing results and that the robustness of the results had to be validated. In a pilot study, robust preprocessing and quantification methods were validated by the a priori difference of healthy controls (HC) versus chronic schizophrenia patients (CHR). These methods were then used in study 1 to validate different quantification methods by comparing the sensory gating deficit in an independent sample of first admission schizophrenia patients (FA) and the same CHR and HC from the pilot study. Results revealed that all quantification methods which show a HC vs CHR contrast also show a HC vs FA contrast but do not show a CHR vs FA contrast. This convergence of methods between patient groups strengthens the robustness of the sensory gating deficit in schizophrenia independent of stage of the illness. Study 2 then used this robust sensory gating index to study the involvement of frontal regions in the auditory sensory gating process. When analyzing only the strongest, data driven frontal
source, an information flow opposite to the original hypothesis was found. Auditory to frontal connectivity preceding the sensory gating was found in healthy controls but not in schizophrenic patients. This connectivity also correlated significantly with sensory gating, meaning that more connectivity led to more filtering (Table 4.1).

Table 4.1

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Aims</th>
<th>Results</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>Compare preprocessing and quantification methods in the extent they produce an a priori HC vs CHR contrast</td>
<td>Considerable differences between the methods Only some methods produced a HC vs CHR contrast</td>
<td>Variable results make a validation of methods necessary. Rule of parsimony: Simplest methods which produced the HC vs CHR contrast used for further studies.</td>
</tr>
<tr>
<td>2</td>
<td>Compare quantification methods across three groups: HC, CHR and FA</td>
<td>All quantification methods which produce a HC vs CHR contrast also show a HC vs FA but no FA vs CHR contrast</td>
<td>Sensory gating is robust across quantification methods and stage of illness</td>
</tr>
<tr>
<td>3</td>
<td>Compare different frontal and auditory candidate networks and their causal relationship to sensory gating</td>
<td>An information flow from strongest auditory to strongest frontal source was found for HC but not for CHR This information flow correlates with sensory gating</td>
<td>Contrary to the initial hypothesis, a bottom up and not a top down information flow preceding the auditory stimulation predicts the extent of the following gating. This connectivity is disturbed in schizophrenia patients.</td>
</tr>
</tbody>
</table>

The results of the present dissertation may be discussed in the light of the following questions:

1. Which methods can be recommended to study sensory gating?
2. Can sensory gating be considered an endophenotype for schizophrenia?

3. How can sensory gating methods be used in other contexts?

4.2 Methodological recommendations to study sensory gating in schizophrenia

In the present pilot study and Study 1, different strategies for preprocessing and quantification methods were compared. First of all, just comparing all possible methods based on an a priori contrast of HC vs CHR and then picking these methods to evaluate said contrast is of course circular reasoning and not feasible by itself. It also shows that indeed any two group comparison could be presented as either significant or non-significant, if one would try all methods and report only the one with the desired result (c.f. Simmons et al., 2011). This highlights the importance of a full reporting and a thorough justification of used methods. In the case of the present dissertation, this was done via two strategies: Especially for the preprocessing methods, the law of parsimony was used (Sarris & Reiß, 2005): From two otherwise identical alternatives, the simpler of the two should be used. This was further substantiated by the fact that even the simplest of approaches – NO preprocessing at all – elicited at least a tendency of the a priori HC vs CHR contrast. So, preprocessing methods which reproduced this contrast in greater strength can be seen as carving out the pre-existing difference

---

11 There were even a few combinations of methods which showed a CHR vs FA contrast but were not included here because they did not show the a priori HC vs CHR contrast.

12 At least for the quantification methods which showed a HC vs CHR contrast at all, this was not the case e.g. with mean peak and mean latency scoring, which did not produce a HC vs CHR contrast regardless of the used preprocessing.
in greater detail. The other justification of methods concerned the FA, as the methods were judged based on a HC vs CHR contrast and then also tested on an independent third group.

Having said that, the following strategies for preprocessing and quantification methods can be recommended based on the aforementioned results: Classified as preprocessing methods different strategies of noise correction, trial exclusion, artifact rejection and filter settings were compared. Classified as quantification methods different strategies of sensor or source reconstruction, peak identification, peak scoring and sensory gating indices were compared. Generally, the rule of parsimony could be applied for all preprocessing methods that yielded the HC vs CHR (and accordingly also the HC vs FA) difference. As far as the present setting with a large number of participants in a sensory gating paradigm can be generalized, the following preprocessing methods can be recommended as the best compromise between removing unwanted noise and artifacts from the signal while retaining most of the original undistorted recording:

- Correcting environmental noise via mean reference
- excluding bad trials via a variance based criterion
- rejecting artifact contaminated components via ICA and
- remove baseline drifts with a low cutoff, forward and backward FIR high-pass filter with an order low enough to not blur S1 and S2 responses\(^\text{13}\)

Recommendations for filter settings, which are also a big source of variation in the sensory gating literature and are often underspecified (de Wilde et al., 2007; Patterson et al., 2008; Widmann et al., 2015), cannot be made independent

\(^{13}\) Exact filter recommendations depend on the setup, e.g. the length of the filter response depends both on the order of the filter and the sampling frequency.
of other method recommendations. The minimal “effective” filter settings in the present analysis consisted of a 1 Hz high-pass, order 300, forward and backward FIR filter while the low-pass filter settings did not matter as much as long as they stayed well above a 50 Hz cutoff. However, unlike the other tested preprocessing methods, the high-pass filter settings interacted with later quantification methods: Sensor and difference quantifications were less affected by a lower 1 Hz high-pass filter setting, while source and ratio quantifications needed a higher 5 Hz high-pass filter in some cases to produce the a priori HC vs CHR contrast. On the other hand, a high-pass filter of 5 Hz and higher affected the M100 more than the M50, because the M100 has a lower frequency characteristic than the M50. This in turn meant that such a filter is detrimental to a M50-M100 peak to peak scoring and a M50-M100 ratio.

For quantification methods, the rule of parsimony is applicable to source reconstruction and peak identification. The present solution of picking individual peak sensors and peak latencies proved to be comparable or even better than source reconstruction and manual identification of peaks. It has to be noted, though, that source reconstruction provides other advantages, such as better separation of signal sources which can also be of additional value, e.g. to estimate connectivity flows between brain regions as in study 2. However, the rule of parsimony is not strictly applicable to peak scoring and the sensory gating index. While additional constraints such as restricting S2 latency to within ±10ms of S1 can be ruled out as unnecessary sophistication, the choice of whether to score relative to different baselines or peak to peak and the choice of S1/S2 or S1-S2 is not different in complexity. But it can be judged in combination with the other aforementioned filter distortions: Scoring relative to a baseline is vulnerable to a distorted baseline e.g. by backward filtering. Also, scoring relative to shorter and
separate baselines for S1 and S2 (100ms before S1 and 100ms before S2) is even more vulnerable to a distortion by filtering, as this distortion is especially present right before click onset (Figure 1.4), so scoring S1 and S2 relative to the same 1000ms baseline preferable because it is more stable to filter distortions. Also, as a ratio score is more vulnerable to a baseline drift and thus needs more filtering, a difference measure is preferable in terms of simplicity. Finally, as the M50 and the M100 have different frequency characteristics (lowering the low-pass cutoff removes the M50 before the M100, increasing the high-pass cutoff removes the M100 before the M50), a S1-S2 M50 difference is not as easily distorted as a M50-M100 peak to peak scoring.

Taken together, the recommended quantifications based on this method comparison would be to remain in sensor space and pick automated individual peak sensor per hemisphere, identify peaks via automatic peak detection, peak scoring relative to a long pre-S1 baseline and a sensory gating index of a M50 S1-S2 difference. This combination of preprocessing and quantification methods was used in study 2 to relate the information flows to a sensory gating score. The only exception from the recommendations were the usage of source reconstruction instead of sensor scoring because the information flow involved picking different brain regions as network nodes which is only possible in source space.

4.3 Sensory gating as an endophenotype for schizophrenia

Sensory gating has been proposed as an endophenotype for schizophrenia (e.g. Miller & Rockstroh, 2013), meaning that it is a link between a genetic risk for and the manifestation of schizophrenia. To qualify as an endophenotype, it
has to satisfy the following criteria (Chan & Gottesman, 2008, p.94; Hasler, Drevets, Gould, Gottesman, & Manji, 2006, p.362f):

“1. An endophenotype is associated with illness in the population.

2. An endophenotype is heritable.

3. An endophenotype is state independent (manifests in an individual whether or not illness is active) but age normed and might need to be elicited by a challenge (e.g., glucose tolerance test in relatives of diabetics).

4. Within families, endophenotype and illness cosegregate.

5. An endophenotype identified in probands is found in their unaffected relatives at a higher rate than in the general population.

6. The endophenotype should be a trait that can be measured reliably, and ideally is more strongly associated with the disease of interest than with other psychiatric conditions (i.e., specificity).”

While most of the criteria are being met by sensory gating (e.g. Miller & Rockstroh, 2013; Thibaut et al., 2015), it is especially the reliable measurement of criterion 6 which is being challenged by the inconsistent results across studies. Particularly the meta-analyses from de Wilde et al. (2007) and Patterson et al. (2008) criticized that the wide range of employed methods, varying effect sizes and that the claim of a robust endophenotype can only be made by studies from certain laboratories.

Concerning the wide range of employed methods across studies, the present dissertation evaluated a wide range of methods within one dataset and found consistent results. Accordingly, as far as this dataset can be generalized: Yes, sensory gating can be measured reliably and sufficient for criterion 6. The convergence of methods can be seen as concurrent validity. Even though some
method combinations failed to even find a HC vs CHR contrast, almost all methods that did also found a HC vs FA difference, which could also be interpreted as a split-half reliability test: As no FA vs CHR difference was found for these methods, the both patient groups are similar in the extent of sensory gating deficit. Also, the first results of a frontal and auditory connectivity preceding and correlating with sensory gating could hint at a causal frontal involvement in the origin of sensory gating, but the evidence for that was weaker, as the methods did not converge but could only be externally validated.

However, if present results suggest similar effect sizes across methods, the variability across studies is probably due to other factors, which were not systematically varied but held constant in this study. One of these constant factors is the specific settings of the paradigm, e.g. the specific click characteristics or measurement with MEG instead of EEG. However, the present settings of the paradigm proved to be robust, which in turn still fulfills the criterion that the deficit has to be reliably measurable. As the pool of participants was also constant, another possible factor are the relationships to characteristic symptoms of SZ to sensory gating, which was found for M50 gating (Smith et al., 2010; Thoma et al., 2003, 2005), P50 gating (Arnfred & Chen, 2004; Erwin et al., 1998; Louchart-de la Chapelle et al., 2005; Ringel et al., 2004; Yee et al., 1998) and even spectral analyzes of EEG (Keil, Roa Romero, Balz, Henjes, & Senkowski, 2016). While analyzing different methods in the same dataset in this dissertation allows to better compare the methods, it did not allow to compare different datasets. So, the heterogeneity in schizophrenia samples could be one of the reasons of differing results.
Finally, study 2 tried to shed more light on a causal involvement of frontal regulation to sensory gating. Embedding sensory gating in a causal path to schizophrenia would also strengthen its role as an endophenotype, however, while the present results did find a frontal involvement in healthy controls which was disturbed in schizophrenic patients, it was in the opposite direction of the hypothesis. This hints at a more complex involvement than a simple top-down regulation and needs further research.

4.4 Limitations and implications for further research

Adding to the methodological recommendations (chapter 4.2) and the strengthening of sensory gating as an endophenotype for schizophrenia (chapter 4.3), several other implications for further research can be made based on present results.

For one, the methodological recommendations might be unique to the present experimental settings to some degree. Here, different strategies of analyzing the same dataset were explored, but there are also a variety of ways the datasets itself can vary. Some of the possible variations of experimental setup and subjects have been tested before, such as click intensity, sex, age and earphone type or speaker delivery, which all failed to show significant differences in the meta-analysis of Patterson et al. (2008). However, some variations of sound intensity did elicit significant changes in the P50 responses: Griffith et al. (1995) reported a HC vs CHR contrast for 30 and 50 dB sound intensity clicks, but not for 70 dB or louder and White & Yee (2006) did not find variations in the gating ratio, but in the P50 S1 amplitude depending on click intensity and background noise. As these variations of the paradigm were tested with the same method of
analyzing the dataset, interactions of paradigm with quantification methods remains to be verified.

Concerning optimal preprocessing methods, other variations are possible. A MEG which is e.g. situated in a more urban setting with more environmental noise sources might need a different noise correction than a MEG situated in a rural area. Also, the characteristics of “bad trials” might change when the participants are in a different setting, e.g. in a more or less agitated state or sitting instead of a supine position, and thus producing other artifact characteristics which might be better recognized by other trial and component rejection procedures. Concerning optimal preprocessing methods, the biggest difference is due to recording via MEG compared to EEG as, amongst other things, EEG is not reference free, more distorted by tissues of different conductivity and sensitive to radial sources. Accordingly, further research might substantiate the present methodological recommendations by further variations like using EEG instead of MEG.

Other research questions concern the relation of sensory gating to other indices, like cognitive measures or connectivity measures. As pointed out before, the variability of methods makes a thorough justification of them necessary and a possible solution for this would be convergence of methods or external validation. This is especially true when adding additional indices. As the paths tested in the pilot study alone add up to around 100,000 possible combinations, a further combination with e.g. 10 domains of the MATRICS Consortium Cognitive Battery (Nuechterlein & Green, 2006) would multiply that to a million possible combinations, so convergence of methods might be a strenuous task. Instead, one solution would be external validation by a replication of the exact settings where an
effect was found before and to use only them. This was done with the same dataset of this dissertation in the study of Carolus et al. (2014), as it used the same method as employed in an earlier study by Popov et al. (2011).

But not only exact replication of previously used methods can be used as external validation, it is also possible to use external validation within one dataset: Like the evaluation of the pilot study by an independent patient group later in study 1, a study relating sensory gating to other indices could analyze a similar split-half reliability test by evaluating the best methods on one half of the total test subject pool and testing them on the other.

Finally, another solution would be the aforementioned recommendation for best methods, which is to generally use the least amount of preprocessing necessary and, in the case of similar recording situations as in the present dissertation, to use the recommended methods specifically from 4.1.

In study 2, a combination of these was used. For one, the sensory gating index from the recommended pipeline of 4.1 was used as an external validation of the connectivity measures. Additionally, while not reported here to not multiply the results beyond readability, the connectivity measures were actually correlated with all gating indices used in study 1 from chapter 2.3 and the correlation held true.

To sum up, sensory gating can be used to relate to other indices, but they have to be carefully justified as bringing in external measures multiplies the amount of choices made exponentially and can easily lead to selective reporting.
4.5 Conclusion

Multiple methods yield different results – so the robustness of results is only valid if (a) methods converge to similar results or (b) methods can be validated by external justification. If they do not – no claim can be made, because each method is equally justifiable but leads to different results or, worse, if only one of the appropriate methods would be used, a false claim could be made. In the first study, the multiple comparison problem could be ameliorated by taking the HC vs CHR contrast as a benchmark and evaluate those with an independent FA patient group. By doing so, sensory gating deficits were found to be a robust measure in schizophrenia across patient groups and quantification methods. In the second study, a partial external justification could be found for a simple network model, but if the model was extended to other potential brain areas, no claim could be made. For the simple network model of information flows between the strongest auditory and frontal source, an information flow from auditory to frontal was found for healthy controls but not for schizophrenia patients. This information flow was correlated with the independent measure of sensory gating, which served as an external validation.
List of Tables

Table 2.1: *Demographic information on samples: healthy controls (HC), chronic schizophrenia patients (CHR) and firstly admitted schizophrenia patients (FA)* ................................................................. 40

Table 2.2: *Metrics and group comparisons for M50 quantifications* ............... 51

Supplementary Table 2.1: *Metrics and group comparisons for adjusted quantifications* ................................................................. 59
Supplementary Table 2.2: *Metrics and group comparisons for quantifications with separate S1 and S2 baseline* ......................................................... 61
Supplementary Table 2.3: *Metrics and group comparisons for S1- and S2-evoked response amplitudes* ................................................................. 62
Supplementary Table 2.4: *Metrics and group comparisons for M100 scores instead of M50* ................................................................. 63
Supplementary Table 2.5: *Metrics and group comparisons for quantifications with separate peak latencies* ......................................................... 64
Supplementary Table 2.6: *Metrics for M40-M50 peak to peak scoring* ............ 69
Supplementary Table 2.7A: *Metrics and group comparisons for M50 quantifications per hemisphere* ................................................................. 70

Table 3.1: *Demographic information for healthy participants and schizophrenia patients* ................................................................. 77

Table 3.2: *Extended network analysis* ................................................................. 84

Table 4.1: *Aims, results and conclusions of the present dissertation* ............... 94
List of Figures

Figure 1.1: Stimulus sequence used in the present dissertation ............................ 4
Figure 1.2: ERF of the sensors surrounding the M50 peak on the right
hémisphere, averaged over all subjects.................................................... 5
Figure 1.3: Preliminary pipeline of tested methods............................................. 9
Figure 1.4: ERF of the sensors surrounding the M50 peak on the right
hémisphere, averaged over all subjects with different filter settings...... 17
Figure 1.5: Selected preprocessing and quantification methods for study 1. ... 27
Figure 1.6: Univariate autoregression................................................................. 29
Figure 1.7: Bivariate autoregression................................................................. 30

Figure 2.1: Topography 40-65 ms (M50, panel A) and 65-130 ms (M100, panel
B) after S1 averaged across all participants .............................................. 43
Figure 2.2: S1 minus S2 power in parcellated regions averaged across all
participants.................................................................................................. 46
Figure 2.3: Time course of activity for sensor and source analysis averaged
across participants within group .................................................................. 48
Figure 2.4: Metrics combining S1 and S2 scores .............................................. 52

Figure 3.1: A: Statistical difference between S1 (0-200ms) and S2 (500-700ms)
source strength. B: Time course of activity in virtual sensors C: Granger
Causality values for CMR and HGR. D: Correlation between prestimulus
connectivity and sensory gating................................................................. 86
References


https://doi.org/10.1111/j.1469-8986.2010.01168.x

https://doi.org/10.1093/oxfordhb/9780199764228.013.6

https://doi.org/10.2307/2286629

https://doi.org/10.1037/0033-2909.112.1.155

https://doi.org/10.1007/s13398-014-0173-7.2

https://doi.org/10.1016/S0006-3223(01)01082-4

https://doi.org/10.1093/schbul/16.3.433


Dalal, S. S., Zumer, J. M., Agrawal, V., Hild, K. E., Sekihara, K., & Nagarajan,


https://doi.org/10.1111/j.1469-8986.2011.01262.x


https://doi.org/10.1111/j.1469-8986.2008.00682.x


Gelman, A., & Loken, E. (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Psychological Bulletin, 140*(5), 1272–1280. https://doi.org/10.1037/a0037714


Huang, M. X., Edgar, J. C., Thoma, R. J., Hanlon, F. M., Moses, S. N., Lee, R.


Louchart-de la Chapelle, S., Levillain, D., Ménard, J.-F., Van der Elst, A., Allio,


Nuechterlein, K. H., & Green, M. F. (2006). *MATRICS Consensus Cognitive*
References


Popov, T., Jordanov, T., Rockstroh, B., Elbert, T., Merzenich, M. M., & Miller, G.


