MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES LEIPZIG

Sonja Lattner: Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformation. Leipzig: Max Planck Institute of Cognitive Neuroscience 2002 (MPI Series in Cognitive Neuroscience; 29)

# Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformation

Der Humanwissenschaftlichen Fakultät
der Universität Potsdam
im Mai 2002 eingereichte

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

DOCTOR PHILOSOPHIAE
(Dr. phil.)

von Sonja Lattner, M.A.

Die Annahme der Dissertation haben empfohlen:

1. Prof. Dr. Angela Friederici
   Universität Potsdam
2. Prof. Dr. Douglas Saddy
   Universität Potsdam

Danksagung

J. Kreiman schrieb 1997 in ihrem Artikel Theory and Practice in Voice Perception Research:“ Given its interdisciplinary nature, the study of vocal quality has no clear academic home. Students at the beginning of their studies may find it difficult to acquire the necessary interdisciplinary expertise [...]“[1]
Ich hatte das Glück, im Rahmen meiner Doktorarbeit am Max Planck Institut für Neuropsychologische Forschung interdisziplinäre Expertise nicht nur im Hinblick auf die Stimmverarbeitung, sondern vor allem auch in Bezug auf die Methoden, mit denen diese in der vorliegenden Arbeit untersucht wurde, erwerben zu können.
Für die großartigen Arbeitsmöglichkeiten bedanke ich mich vor allem bei Frau Prof. Angela Friederici. Außerdem danke ich Herrn Alter, dem Leiter der Nachwuchsgruppe “Neurokognition der Prosodie“, in deren Rahmen diese Arbeit entstand.
Für ihre Unterstützung und zahlreiche Diskussionen danke ich besonders Martin Meyer und Thomas Gunter, die mir allzeit zur Seite standen. Burkhard Maeß und Claudia Hruska danke ich für ihre Einführung in die Arbeit mit dem MEG und dem EEG. Für den Ideenaustausch und die Unterstützung bei der Datenauswertung danke ich Burkhard Maess, Thomas Knoesche, Michael Schauer und Yunhua Wang, die am Gelingen der MEG-Studien wesentlich beteiligt waren. Sylvia Stasch, Kristin Schloßhauer und Yvonne Wolff danke ich für ihre Hilfe bei der Datenerhebung. Für wertvolle Hinweise und Anregungen, sowie für die Beantwortung meiner oft zahlreichen praktischen und theoretischen Fragen danke ich Claudia Hruska, Monica de Filippis, Claudia Friedrich, Annett Schirmer, Christian Fiebach, Stefanie Kruck, und Volker Bosch. Andrea Gast-Sandmann danke ich für die Mitgestaltung der Abbildungen in der vorliegenden Arbeit und in den damit verbundenen Publikationen. Frau Lewin gilt schließlich mein besonderer Dank für ihren unermüdlichen Einsatz bei der Beschaffung der manchmal sehr alten und

[1] J. Kreiman: *Theory and Practice in Voice Perception Research.* In K.Johnson & J.W. Mullenix : Talker Variability in Speech ProcessingSSan Diego 1997, S. 101/2

oft sehr neuen Publikationen zum Thema.
Für den Rückhalt und Ausgleich, die ich bei ihnen immer gefunden habe, danke ich meiner Familie, Stefan, Kristin und meinen Freunden.

# Inhaltsverzeichnis

**Inhaltsverzeichnis** i

**Einleitung** 1

**I Einführung in die Methoden** 5

**1 Grundannahmen neurophysiologischer Methoden** 7

**2 Elektroenzephalographie** 9

**3 Magnetenzephalographie** 13

**4 Neurophysiologische Komponenten** 19

4.1 Kennzeichen neurophysiologischer Komponenten . . . . . 19

4.2 N100 . . . . . 21

4.3 Mismatch Negativierung . . . . . 23

4.4 N2b und P300 . . . . . 26

4.5 N400 . . . . . 29

**II Verarbeitung von Stimminformation** 33

**5 Produktion, Akustik und Wahrnehmung** 35

**6 Sprechernormalisierung** 41

**7 Neuronale Grundlagen** **49**

**III Empirischer Teil** **57**

**8 EXPERIMENTALSERIE I: Prototypikalitätseffekte bei der Stimmwahrnehmung** **59**

8.1 Experiment Ia . . . . . 59
8.1.1 Fragestellung . . . . . 59
8.1.2 Methode . . . . . 61
8.1.2.1 Stimuli . . . . . 61
8.1.2.2 Probanden . . . . . 65
8.1.2.3 Versuchsablauf . . . . . 66
8.1.2.4 Datenerhebung . . . . . 66
8.1.2.5 Quellanalyse . . . . . 67
8.1.3 Ergebnisse . . . . . 68
8.1.4 Diskussion . . . . . 70
8.2 EXPERIMENT Ib . . . . . 73
8.2.1 Fragestellung . . . . . 73
8.2.2 Methode . . . . . 73
8.2.2.1 Stimuli . . . . . 73
8.2.2.2 Probanden . . . . . 74
8.2.3 Ergebnisse . . . . . 74
8.2.4 Diskussion . . . . . 76
8.3 Allgemeine Diskussion der Experimente Ia und Ib . . . . . 78

**9 EXPERIMENT II: Stimminformation und phonetische Verarbeitung** **83**

9.1 Experiment II . . . . . 83
9.1.1 Fragestellung . . . . . 83
9.1.2 Methode . . . . . 85
9.1.2.1 Stimuli . . . . . 85
9.1.2.2 Probanden . . . . . 86

9.1.2.3 Versuchsablauf . . . . . 86
9.1.2.4 Datenerhebung . . . . . 86
9.1.2.5 Quellanalyse . . . . . 86
9.1.3 Ergebnisse . . . . . 89
9.1.4 Diskussion . . . . . 92

**10 EXPERIMENTALSERIE III: Stimminformation und semantische Verarbeitung 97**

10.1 Allgemeine Fragestellung . . . . . 97
10.2 Versuchsmaterial . . . . . 99
10.3 Experiment III a: Reaktionszeitstudie - Sprecherdiskrimination und semantische Verarbeitung . . . . . 102
10.3.1 Fragestellung . . . . . 102
10.3.2 Methode . . . . . 104
10.3.2.1 Probanden . . . . . 104
10.3.2.2 Versuchsablauf . . . . . 104
10.3.3 Ergebnisse . . . . . 105
10.3.4 Diskussion . . . . . 110
10.4 Experiment IIIb: EKP-Studie - Sprecherdiskrimination . . . . . 115
10.4.1 Fragestellung . . . . . 115
10.4.2 Methode . . . . . 117
10.4.2.1 Probanden . . . . . 117
10.4.2.2 Stimuli . . . . . 117
10.4.2.3 Versuchsanordnung . . . . . 118
10.4.2.4 EEG-Aufnahme . . . . . 118
10.4.2.5 Versuchsablauf . . . . . 119
10.4.3 Ergebnisse . . . . . 120
10.4.4 Diskussion . . . . . 127
10.5 Experiment IIIc: EKP-Studie - Passives Zuhören . . . . . 132
10.5.1 Fragestellung . . . . . 132
10.5.2 Methode . . . . . 132

10.5.2.1 Probanden . . . . . 132
10.5.2.2 Stimuli . . . . . 133
10.5.2.3 Versuchsanordnung . . . . . 133
10.5.2.4 EEG-Aufnahme . . . . . 133
10.5.2.5 Versuchsablauf . . . . . 133
10.5.3 EKP-Ergebnisse . . . . . 133
10.5.4 Diskussion . . . . . 138
10.6 Allgemeine Diskussion der Experimente IIIa, IIIb und IIIc . . . . . 140

**IV Zusammenfassung und abschließende Diskussion 145**

**11 Abschließende Diskussion 147**
11.1 Theoretischer Ausgangpunkt . . . . . 147
11.2 Methodischer Hintergrund . . . . . 149
11.3 Zusammenfassende Diskussion der Ergebnisse . . . . . 150
11.4 Resumée und Ausblick . . . . . 159

**V APPENDIX 161**

**A Experimentalserie I 163**

**B Experiment II 167**

**C Experimentalserie III 169**

**Literaturverzeichnis 179**

# Einleitung und Überblick

*Even the most sophisticated state-of-the-art speech-recognition systems cannot compare to the speed and efficiency of the human listener.*

- David B. Pisoni -

*Stimmen. Ich denke sie dringen tiefer in uns ein als andere Dinge.*

- George Eliot -

Bei der lautsprachlichen Kommunikation dekodiert der Hörer die vom Sprecher intendierte Nachricht aus dem akustischen Sprachsignal. Dabei werden die relevanten akustischen Muster des Sprachschalls mentalen Repräsentationen der Sprachlaute (Phonemen) zugeordnet und der akustische Strom in Wörter, Phrasen und Sätze segmentiert. Über diese linguistische Information hinaus wird vom Hörer aber auch para- und extralinguistische Information verarbeitet. Paralinguistische Information ist nonverbal und vom Sprecher meist intendiert. Sie kennzeichnet einen Satz beispielsweise als vertrauliche Mitteilung oder ironisch gemeinte Äußerung. Extra-linguistische Information umfaßt all jede Parameter, die in der Regel nicht der Kontrolle durch den Sprecher unterliegen, die aber Aussagen über sein Alter, seine Geschlechtszugehörigkeit usw. erlauben. Die neurowissenschaftliche Forschung hat sich vornehmlich mit denjenigen Aspekten der Sprache befaßt, die verbal kodiert sind und mit der linguistischen Struktur und Bedeutung zusammenhängen, d.h. mit Semantik, Syntax, Morphologie und Phonologie. In neuere Zeit, auch aufgrund verbesserter Möglichkeiten der auditorischen

Stimuluspräsentation, wurde zudem die kognitive Verarbeitung suprasegmentaler Merkmale und paralinguistischer Information untersucht (siehe z.B. Steinhauer, Alter & Friederici, 1999; Pihan, Ackermann & Altenmüller, 1997; Buchanan, Lutz, Mirsazade, Specht, Shah, Zilles et al., 2000; Kotz, Meyer, Alter, Besson, von Cramon & Friederici, in press; Schirmer, Kotz & Friederici, in press). Den extralinguistischen Parametern wurde bisher hingegen wenig Beachtung geschenkt. Die vorliegende Arbeit wendet sich diesem Aspekt zu und befaßt sich insbesondere mit den neuronalen Korrelaten, die mit der Wahrnehmung von Stimminformation einhergehen. Als Untersuchungsmethoden wurden mit der Elektroenzephalographie und der Magnetenzephalographie zwei Verfahren gewählt, die eine nicht-invasive Beobachtung auditorischer Verarbeitungsprozesse erlauben. Die Methoden werden im ersten Teil der Arbeit beschrieben. Dort erfolgt ebenfalls eine Darstellung derjenigen elektrophysiologischen Parameter (Komponenten), die für das weitere Verständnis der Arbeit notwendig sind. Es schließt sich eine Einführung an den inhaltlichen Gegenstand der Arbeit an: Zunächst werden die akustischen Parameter der menschlichen Stimme und die daraus resultierenden Eigenarten für ihre Perzeption beschrieben. Danach wird ein Überblick über die bisherigen Erkenntnisse aus der experimentellen und klinischen Erforschung der Stimmwahrnehmung gegeben, wobei vor allem drei Aspekte deutlich werden:

1. Die menschliche Stimme zeichnet sich im Vergleich zu Tönen durch besondere akustische Parameter aus, die in einer spezifischen Verarbeitung reflektiert sind.

2. Die Verarbeitung von phonetisch-sprachlicher Information und Sprechstimme scheinen nicht unabhängig voneinander zu verlaufen.

3. Bei der Verarbeitung von Stimminformation handelt es sich um einen sehr schnellen Prozeß, der wenig Verarbeitungskosten verursacht und deshalb nur in subtilen experimentellen Anordnungen beobachtbar gemacht werden kann. Es ist jedoch unklar, inwieweit beobachtete Effekte in Verhaltensexperimenten von der jeweiligen Experimentalaufgabe abhängig sind.

Diese drei Faktoren der Stimmwahrnehmung werden in der vorliegende Arbeit untersucht, wobei es vornehmlich um ihre neuronalen Korrelate in der frühen auditorischen Verarbeitungsphase gehen soll. Jedem der drei Aspekte ist dabei ein Abschnitt des empirischen Teils der Arbeit gewidmet.

Die erste Experimentalserie umfaßt zwei MEG-Experimente. In diesen wird mit Hilfe eines etablierten Experimentalparadigma (Oddball-Design) gezeigt, daß die Einzigartigkeit menschlicher Stimmen bereits in ihrer frühen auditorischen Verarbeitung genutzt wird, bzw. daß Abweichungen in der Prototypikalität bereits in diesem frühen Zeitbereich detektiert werden. Die Ergebnisse lassen weiterhin vermuten, daß die Wahrnehmung der Stimminformation nicht durch die unabhängige Verarbeitung einzelner akustischer Parameter geschieht, sondern daß die Gesamtkonfiguration, bzw. der holistische Gesamteindruck von Bedeutung ist.

Der zweite Abschnitt des empirischen Teils besteht aus einer MEG Studie, die das Verhältnis von Stimminformation und phonologischen Parametern in der präattentiven Sprachverarbeitung untersucht. Es wird neurophysiologische Evidenz für die Annahme präsentiert, daß die Verarbeitung beider Informationsarten parallel geschieht und daß es sich dabei um nicht-unabhängige, interagierende Prozesse handelt.

Im dritten Abschnitt wird mit einem Reaktionszeitexperiment und zwei EKP-Studien der Einfluß der Stimminformation auf das semantische Satzverstehen untersucht. Es wird gezeigt, wie eine Variation der Experimentalaufgabe die neurophysiologischen Korrelate der sprecherspezifischen und der semantischen Verarbeitung moduliert. Die Daten deuten darauf hin, daß das Perzeptionssystem auf die Dekodierung der intendierten Nachricht ausgelegt ist und Sprecherunterschiede automatisch ausgeglichen werden, so daß semantische Prozesse davon nicht unbedingt betroffen sind.

Im letzten Teil der Arbeit werden die Daten der Experimente zusammengefaßt. Der Beitrag der vorliegenden Arbeit zum Verständnis der Verarbeitung lautsprachlicher Signale wird diskutiert und ein Ausblick auf weitere Fragestellungen gegeben.

# Teil I

# Einführung in die Methoden

# Kapitel 1

# Grundannahmen neurophysiologischer Methoden

Die Signalübermittlung innerhalb und zwischen Nervenzellen (Neuronen) verläuft auf elektrochemischen Wege, durch den intrazellulären Austausch von Neurotransmittern, sowie durch die Bewegung von Ionen innerhalb und außerhalb der Zellen. Da Ionen elektrisch geladene Teilchen sind, ist ihre Bewegung als Stromfluß aufzufassen. Man beschreibt diese Signalübermittlung auch als *neuronales Feuern* bzw. als *neuronale Aktivierung*, während die Nervenzellen selbst als *Quellen* der Aktivierung bezeichnet werden. Wenn auch das explizite Zusammenspiel von Zellaktivierung und dem korrelierten sensorischen, motorischen bzw. kognitiven Resultat nicht vollständig geklärt ist, so wird doch angenommen, daß jede Aktivierung einem Zweck dient, bzw. einen Verarbeitungsschritt widerspiegelt (Papanicolaou, 1998). Da zahlreiche basale und kognitive Verarbeitungsvorgänge gleichzeitig ablaufen, bedarf es besonderer Methoden und experimenteller Versuchsanordnung, um Aktivierungen gezielt isolieren und untersuchen zu können. Dazu stehen verschiedene Technologien zur Verfügung, z.B. die räumlich hoch-

auflösenden, bildgebenden Verfahren wie die Positronen-Emissions-Tomographie (PET) und die funktionale Magnetresonanztomographie (fMRT). Diese Methoden machen sich den Zusammenhang von neuronaler Aktivierung und damit verbundenen Stoffwechselvorgängen im Gehirn zunutze. Allerdings sind sie somit an den Blutkreislauf gebunden, der mit deutlicher Zeitverzögerung von zwei bis sechs Sekunden auf eine neuronale Aktivierung reagiert. Die Anwendung der räumlich hochauflösenden Verfahren bedeutet also in der Regel einen Verlust der zeitlichen Auflösung. Die vorliegende Arbeit fokussiert hingegen auf zeitliche Verläufe früher kognitiv-perzeptueller Verarbeitungsschritte. Deshalb finden hier mit der Elektroenzephalographie und der Magnetenzephalographie zwei neurophysiologische Untersuchungsmethoden Verwendung, die Lage sind, neuronale Aktivierung im Millisekundenbereich abzubilden. Sie werden im folgenden Abschnitt genauer beschrieben.

# Kapitel 2

# Elektroenzephalographie

Die Aktivität von Hirnneuronen bei der Verarbeitung sensorischer, motorischer oder psychischer Ereignisse ist unter bestimmten Bedingungen im Elektroenzephalogramm (EEG) meßbar. Eine Voraussetzung für die Aktivitätsregistrierung ist dabei ein zeitlich synchrones Feuern einer hinreichend großen Anzahl parallel ausgerichteter Nervenzellen. Da eine parallele Zellanordnung innerhalb des Neokortex nur für die Pyramidenzellen gegeben ist, sind diese als Hauptgeneratoren des EEG-Signals anzusehen.[1] Ebenfalls parallel ausgerichtete thalamische Zellpopulationen sind aufgrund ihres relativ großen Abstands zur Kopfoberfläche hingegen vernachlässigbar (Elbert, 1993). Die Summe der synchronen Einzelzellaktivierungen wird als ein einziger *Dipol* interpretiert, d.h. als ein räumlich eng umschriebener, zielgerichteter elektrischer Strom zwischen einem positiven und einem negativen Pol. Ein solcher Dipol evoziert extrazelluläre Ausgleichs- oder Volumenströme, die dem Weg des geringsten Widerstandes folgend, über die inaktive weiße Substanz, Hirnhäute und Knochen einen Ladungsausgleich anstreben. Sie sind die Grundlage der elektroenzephalographischen Messung. Das EEG-Signal selbst wird dann als Differenz der resultierenden elektrischen Spannung an Meß- und Referenzelektroden registriert. Die Elektroden sind dabei in der Regel an der Kopfoberfläche angebracht z.B. mittels einer dehnbaren Elektro-

[1] Für eine detaillierte Beschreibung siehe z.B. Birbaumer & Schmidt (1990); Rösler (1982); Papanicolaou (1998)

denhaube.[2] Die Positionierung der Elektroden erfolgt nach einer international anerkannten Konvention, dem sogenannten 10-20er System (siehe Abbildung 2.1).

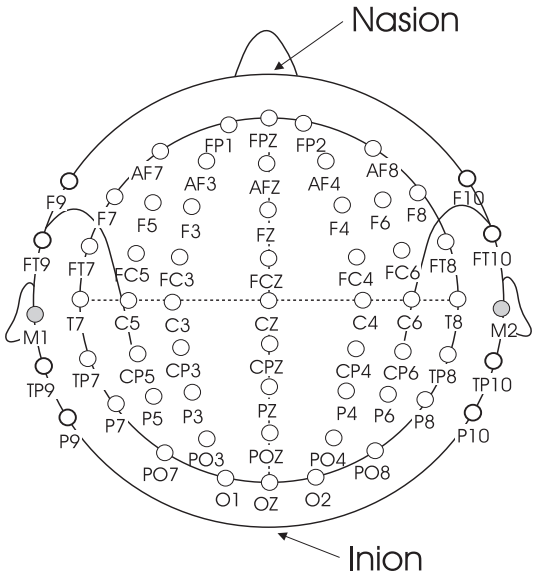



Abbildung 2.1: Das 10-20er System, die internationale Konvention der Elektrodenanordnung. Die ursprüngliche Version der Anordnung stammt von Jasper (1958). Der Bedarf an weiteren Elektroden führte zu der hier abgebildeten Erweiterung durch die Amerikanischen Gesellschaft für Elektroenzephalographie (Sharborough, Chatrian, Lesser, Lüders, Nuwer & Picton, 1991). Die Referenzpunkte der Elektrodenanordnung sind auf anatomischer Grundlage definiert: Der äußere Elektrodenkranz führt durch Nasion und Inion. Folgt man der Mittellinie von Nasion zu Inion, so liegt der zweite Ring nach 10 % dieser Distanz , die frontozentrale Elektrode (Fz) nach weiteren 20%. Die Mittelelektrode (CZ) liegt an der Schnittstelle zwischen der Inion/Nasion-Verbindung und der orthogonal dazu verlaufenden Verbindungslinie der präaurikulären Punkte. Die Nomenklatur der Elektroden orientiert sich an den darunterliegenden Hirnregionen. Dabei sind folgende Indizes gebräuchlich: *a= anterior; c = central; f = frontal; m = mastoid; p = parietal; o = occiptial; t = temporal; z = auf der zentralen Mittellinie.* Elektroden über der rechten Hemisphäre tragen einen geradzahligen, Elektroden über der linken Hemisphäre einen ungeraden Index. Grau markiert sind hier die Referenzelektroden an den Mastoiden.

[2] In der klinischen Anwendung setzt man präoperativ bisweilen auch interkraniale Meßelektroden ein, die in tiefere Hirnschichten vordringen, sowie Netzelektroden, die direkt an der Kortexoberfläche appliziert werden.

Wichtig für eine spätere Interpretation der Messungen sind die Referenzelektroden, die in der Regel entweder an den Mastoiden oder an der Nase angebracht sind. Das gemessene Spannungsmuster wird dann als zeitlich fortlaufendes EEG-Signal abgebildet. Zunächst einmal besteht es aus einer stark schwankenden neuronale Grundaktivität von etwa 50 - 150 $\mu$V, die primär mit einfachen Bewußtseins- und Aufmerksamkeitsprozessen in Verbindung gebracht wird.
Die Wahrnehmung und kognitive Verarbeitung eines Reizes führt zu einer charakteristischen Verlagerung dieser Grundaktivität. Es resultiert ein sogenanntes ereigniskorreliertes Potential (EKP).[3] In der Regel sind EKPs wesentlich schwächer als die Grundaktivität. Da sie im Gegensatz zu dieser jedoch ein reizabhängiges, regelmäßiges Muster aufweisen, gelingt es durch Mittelung der Hirnreaktion auf identische oder ähnliche Reize, das Signal-Rausch-Verhältnis, d.h. das Verhältnis von EKP zur Grundaktivität zu verbessern. Während nämlich die Mittelung für das zufällige Aktivationsmuster der Grundaktivität eine Abschwächung bedeutet, bewirkt sie gleichzeitig eine Verstärkung und damit Isolierung der regelmäßigen ereigniskorrelierten Potentiale (siehe Abbildung 2.2). Das EKP-Signal selbst besteht aus spezifischen negativen oder positiven Potentialverlagerungen, die aufgrund invarianter stimulusabhängiger Merkmale in *Komponenten* klassifiziert werden (siehe Kapitel 4.1). In Bezug auf die Lokalisierung der neuronalen Aktivierungsquellen, die für die jeweilige Komponente verantwortlich sind, kann die EKP-Messung durch die Magnetenzephalographie ergänzt werden.

[3] engl. *event-related potential* (ERP). In der basalen auditorischen Verarbeitung ist bisweilen auch von akustisch evozierten Potentialen (AEP) die Rede.

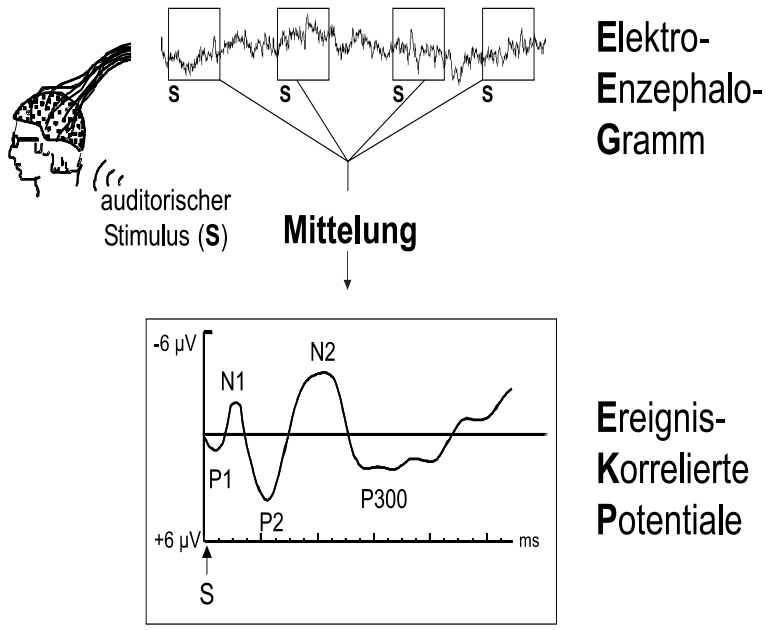



Abbildung 2.2: Das EEG-Signal reflektiert die Hirnaktivität. Durch die Mittelung bestimmter Signalepochen, die zum Ereignis S (auditorische Stimuluspräsentation) zeitlich korreliert sind, werden die regelmäßgen EKPs isoliert. Die zufällige Hintergrundaktivität wird herausgemittelt. Das EKP besteht aus negativen und positiven Potentialen, die aufgrund charakteristischer Eigenschaften als Komponenten klassifiziert werden (z.B. N1, P2, N2, P300).

# Kapitel 3

# Magnetenzephalographie

Jeder Stromfluß, also auch die elektrische Aktivität der Nervenzellen, erzeugt ein Magnetfeld. Bei einer synchronen Aktivierung hinreichend vieler Neurone paralleler Anordnung ist der resultierende magnetische Fluß im Magnetenzephalographen meßbar. Für kognitionspsychologische Studien sind dabei vor allem stimulusabhängige *ereigniskorrelierte Felder (EF)* von Bedeutung. Diese Felder haben eine Stärke von 50 - 500 Femtotesla (fT), sind also im Vergleich zum Erdmagnetfeld (70 $\mu$T) sehr schwach (Sams & Hari, 1991). Für die Messung ereigniskorrelierter Felder bedarf es daher neben einer entsprechend abgeschirmten Umgebung einer speziellen Technologie, die in Abbildung 3.1 dargestellt ist. Resultat der Messung ist ein Magnetenzephalogramm (MEG), d.h. für jeden Meßkanal eine Darstellung des Magnetfeldstärkeverlaufs in der Zeit. Durch physikalisch-mathematische Modellierung wird vom ereigniskorrelierten Magnetfeld auf die möglichen Quellen der neuronalen Aktivität geschlossen. Von großer Bedeutung ist dabei das verwendete Volumenleitermodell des Kopfes/Gehirns. Ein hinreichendes Ergebnis kann bereits mit einer Simulierung des Kopfes als konzentrische Kugelschalen erzielt werden. Jede der Schalen stellt dabei eine der anatomischen Schichten Haut, Schädel, Liquor, Kortex mit ihrer spezifischen Leitfähigkeit dar.

Oft findet auch die Randelemente-Methode (engl. *Boundary Element Modell, BEM*)

Verwendung (Hämäläinen & Sarvas, 1989). Bei einem BEM-Modell wird die Oberfläche einer Gewebeschicht, an der sich die Leitfähigkeit ändert, als ein Netz aus Dreiecken dargestellt (siehe Abbildung 3.2). Dieses Modell ist anatomisch präziser als das Kugelschalenmodell und kann ggf. für jeden Probanden/Patienten individuell erstellt werden z.B. anhand von anatomischen Daten einer kernspintomographischen Messung.

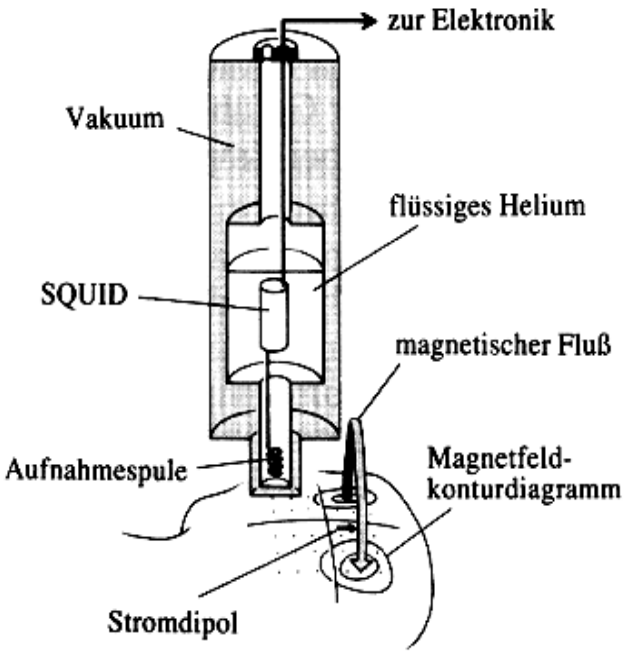



Abbildung 3.1: Die Abbildung zeigt den schematischen Aufbau eines Einkanal-Magnetenzephalographen. Eine Aufnahmespule, das sog. Magnetometer, erfaßt die Hirnaktivität: Dort wird eine Spannung induziert, sobald sich das durch die Spule hindurchgehende Magnetfeld ändert. Um Störungen durch thermale Aktivität zu vermeiden, werden zur Verstärkung des MEG-Signals supraleitende SQUIDS (= Supercoducting Quantum Intereference Devices) eingesetzt, die sich durch eine hohe Sensitivität, aber auch durch einge geringe Anfälligkeit gegenüber thermaler Aktivität auszeichnen. Das Funktionsprinzip basiert auf der Supraleitfähigkeit eines Niobium-Drahtes, der in einer Trommel mit flüssigem Helium auf der konstanten Temperatur von -270° C gehalten wird. Das in der vorliegenden Arbeit verwendete Ganzkopf-MEG enthält 148 Kanäle (Magnetometer), sowie 11 Referenzkanäle, mit deren Hilfe der Einfluß von äußeren Störfeldern verringert wird (Bildquelle: Springer & Deutsch, 1995, S. 116).

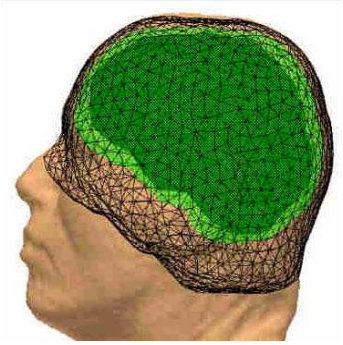


Abbildung 3.2: BEM-Modell des Kopfes: Die Gewebeschichten (Haut, Schädel, Liquor und Kortex) sind als Dreieckselemente unterschiedlicher Leitfähigkeit abgebildet. (Bildquelle: http://www.neuro.com/neuroscan/curdetail.htm)

Die exakteste, dreidimensionale Modellierung aller Gewebsschichten kann mit Hilfe der finiten Elementen Methode (FEM) erzielt werden (für eine ausführlichere Beschreibung siehe z.B. Anwander, Kuhn, Reitzinger & Wolters, 2001). FEM-Modelle haben den Vorteil, daß auch Leitfähigkeitsunterschiede innerhalb einer Schicht, z.B. aufgrund von Frakturen der Schädeldecke oder Hirntumoren, modelliert werden können. Ist eine hohe räumliche Auflösung erwünscht, so sind Finite-Elemente-Modelle den Boundary-Element-Modellen überlegen. Für die einfache experimentelle Anwendung sind sie jedoch aus Gründen der übermäßig hohen Rechenzeit noch sehr unökonomisch und Verbesserungen der mathematischen Lösungsalgorithmen müssen abgewartet werden.

Für die Rekonstruktion der Aktivierung stehen mehrere Möglichkeiten zur Verfügung. Meist wird eine Modellierung einzelner Aktivierungsquellen als äquivalente Stromdipole (engl. *equivalent current dipole, ECD*) durchgeführt, die man über ihre Stärke, ihre Ausrichtung und ihre Lokalisation in Bezug zu festgelegten anatomischen Referenzpunkten definiert (siehe z.B. Knösche, 1997; Scherg & von Cramon, 1985). Da magnetische Felder im Gegensatz zu elektrischer Aktivierung durch die Schädelknochen und das Gewebe kaum abgedämpft oder verzerrt werden, sind die Lokalisationen der Quellen mathematisch präzise rekonstruierbar (Papanicolaou, 1998). Der Hauptunterschied zwischen EEG und

MEG liegt daher in dem Zweck zu dem sie eingesetzt werden: Nur selten werden bei der Elektroenzephalographie dipolare Quellen lokalisiert, vielmehr interessiert hier der zeitliche Aktivierungsverlauf, während sich die topographische Beschreibung auf grobe Klassifikationen wie "parietal" oder "rechts anterior" beschränkt. Hingegen wird beim MEG in der Regel eine größere Anzahl von eng beieinander liegenden Abtastkanäle verwendet, um eine gute räumliche Erfassung der Aktivierung zu erzielen. Zudem heben sich die magnetischen Felder der Volumenströme aufgrund ihrer irregulären Orientierung auf, während die Aktivierung des primären Dipol durch das MEG-Signal abgebildet wird. Es ist jedoch zu beachten, daß bei der Magnetenzephalographie nur solche Felder gemessen werden, deren Quelle tangential zur Kopfoberfläche ausgerichtet ist, während beim EEG die Volumenströme aller Quellen zum Signal beitragen. Eine interessante Fragestellung ist daher, wie verschiedene neuropsychologische Methoden (z.B. MEG, EEG und fMRI) auf überzeugende Weise miteinander verknüpft werden können, um ein vollständiges Bild bestimmter neuronaler Aktivierungen zu erlangen.

Unabhängig von der Tatsache, ob die Modellierung der neuronalen Aktivität auf EEG oder MEG-Daten beruht, stellt sich jedoch immer das sogenannte **Inversionsproblem**: Sind Lokalisation, Stärke und Anzahl von neuronalen Quellen bekannt, so kann eindeutig auf das resultierende Korrelat (EKP oder Magnetfeld) geschlossen werden. Dies bezeichnet man als die Lösung des Forward-Problems. In der experimentellen Messung soll nun jedoch vom beobachteten Korrelat auf die zugehörige Quellkonfiguration geschlossen werden. Dies geschieht durch die Lösung des Forward-Problems für eine hypothetische Quelle. Das so berechnete Ergebnis wird mit dem tatsächlich gemessenen Korrelat verglichen und der Restfehler bestimmt. Die hypothetische Quellenkonfiguration wird dann verändert und die Simulation so lange wiederholt, bis die Lösung optimiert, d.h. der Restfehler sehr gering ist. Es gibt jedoch prinzipiell für jede inverse Lokalisierung unendlich viele Lösungen, denn unendlich viele Konfigurationen (d.h. unendlich viele Kombinationen aus Anzahl, Stärke und Ort einer Quelle) erzeugen identische Meßergebnisse. Die Lösung des Inversionsproblems liegt daher in Vorannahmen oder Vorwissen (z.B. über physiologische Plausibilität), die in Form von Restrik-

tionen in die Modellierung einfließen müssen. Liegen hinsichtlich der Anzahl von Aktivierungsquellen keine konkreten Hypothesen vor, oder ist das beobachtete Magnetfeld nicht plausibel mit einem Dipolmodell zu erklären, so kann alternativ eine Stromdichte-Berechnung durchgeführt werden. Die in der vorliegenden Arbeit verwendeten Minimum-Norm-Schätzungen (Hämälainen & Ilmoniemi, 1994; Wang, Williamson & Kaufman, 1992) modellieren die Quellaktivierung z.B. als Stromdichte auf der Kortexoberfläche, wobei die Gesamtsumme der Aktivität so gering wie möglich gehalten wird. Bei dieser Vorgehensweise ist die Aktivierung zwar auf die Kortexoberfläche eines standardisierten Gehirns beschränkt, aber Gruppenmittelwerte werden so in der Regel reliabler beschrieben, als bei einer Dipolmodellierung (Hari, Levänen & Raji, 2000).

Wie oben bereits erwähnt, werden sowohl der Elektro- als auch der Magnetenzephalographie die gemessenen Hirnpotentiale als *Komponenten*, d.h. als reliable Korrelate der perzeptuell-kognitiven Verarbeitung interpretiert. Eine Definition der Komponente, sowie eine Übersicht über diejenigen Komponenten, die für das Verständnis der vorliegenden Arbeit von Bedeutung sind, wird im folgenden Abschnitt gegeben.

# Kapitel 4

# Neurophysiologische Komponenten

## 4.1 Kennzeichen neurophysiologischer Komponenten

Stehen elektrische Hirnpotentiale in einem kausalen Zusammenhang zur Verarbeitung eines Ereignisses, so sollten sie unter gleichen Bedingungen replizierbar sein. Ein in dieser Hinsicht reliables Hirnpotential wird als *Komponente* bezeichnet. Im Allgemeinen werden zur Definition einer Komponente folgende Faktoren herangezogen:

1. Polarität des Potentials im Vergleich zu einer Baseline, d.h. im Vergleich zu einem Signalabschnitt, in dem es zwischen den Experimentalbedingungen keinen Unterschied gibt (z.B. der Zeitbereich unmittelbar vor der Stimuluspräsentation).

2. Latenz zwischen dem definierten Zeitpunkt $t_0$ (z.B. Stimulusbeginn) und dem Zeitpunkt des Maximum $t_{max}$ der Komponente.

3. Verteilung der Aktivität auf der Kopfoberfläche, die durch den Vergleich der Signale verschiedener Elektroden ermittelt wird.

4. Funktionale Abhängigkeit von kognitiven Prozessen wie z.B. der Aufmerksamkeit oder der Reizmodalität (z.B. auditorische vs. visuelle Reizpräsentation).

Aufgrund der großen Übereinstimmungen interpretiert man diese nicht als eigenständige Komponenten, sondern als spezifischen Variationen einer zugrundeliegenden Komponente.

Die Nomenklatur von Komponenten orientiert sich an diesen Charakteristika. So wird die Polarität durch die Buchstaben N für Negativität und P für Positvität verwendet. Der Zahlenindex beschreibt entweder die Latenz oder die Ordnungszahl der Komponente, und Varianten einer Komponenten aufgrund funktionaler Abhängigkeiten werden mit lateinischen Buchstaben gekennzeichnet (Beispiele: N100, P3a).

Um eine einfache Klassifikation zu schaffen, wurden Komponenten unterteilt in *exogene* Komponenten, die eine Verarbeitung der basalen physikalischen Parameter des Stimulus widerspiegeln, und *endogene* Komponenten, die eine spätere Stufe der Reizverarbeitung reflektieren und von den sogenannten "Top-Down"-Faktoren wie Aufmerksamkeit oder Aufgabenrelevanz beeinflußt werden. In der strikten Form ist diese Unterteilung als zu starke Vereinfachung abzulehnen (Coles, 1995), denn auch die meisten exogenen Komponenten lassen sich durch kognitive Vorgaben wie Aufmerksamkeit variieren, und viele der endogenen Komponenten stehen unter dem Einfluß von Stimulusattributen wie z.B. der Präsentationsmodalität. Es wurde daher vorgeschlagen, die strikte Dichotomie eher durch ein Kontinuum zu ersetzen, auf dem frühe Komponenten (innerhalb der ersten 100 ms) als eher exogen und spätere Komponenten als eher endogen einzuordnen wären (siehe z.B. Coles, 1995).

Im folgenden Abschnitt werden nun diejenigen Komponenten vorgestellt, die im Kontext dieser Arbeit von Bedeutung sind. Die Übersicht beschränkt sich daher vornehmlich auf die auditorische Modalität.

## 4.2 N100

Die N100 Komponente beschreibt die frühe kortikale Verarbeitung perzeptuellen Inputs. Es wird angenommen, daß sie aus mehreren Subkomponenten besteht (Woods, Alain, Covarrubias & Zaidel, 1995; Alain, Woods & Covarrubias, 1997), die auf unterschiedliche neuronale Generatoren zurückzuführen sind. Die magnetischen Korrelate dieser Subkomponenten wurden meist im oder nahe des primären auditorischen Kortex, bzw. Heschl'scher Gyrus lokalisiert (Elberling, Bak, Kofoed, Lebech & Saermark, 1982; Hari, Aittoniemi, Jarvinen, Katila & Varpula, 1980) .

Das magnetische Äquivalent der N100, die sog. M100 oder N100m, wird in der Regel als einzelner Dipol dargestellt (Roberts, Ferrari, Stufflebeam & Poeppel, 2000). Da die N100 frühe Verarbeitungsschritte widerspiegelt, sind neben der Lokalisation vor allem Latenzunterschiede in Reaktion auf verschiedene Stimuli von Bedeutung (eine ausführliche Übersicht dazu findet sich bei Roberts et al. 2000). Beispielsweise zeigten Untersuchungen eine Korrespondenz zwischen der Frequenz eines Sinustons, bzw. seiner Intensität und der Latenz der N100m (Roberts & Poeppel, 1996; Stufflebeam, Poeppel, Rowley & Roberts, 1998). Bei der Präsentation leiserer Töne war eine längere Gipfellatenz der N100m zu beobachten als bei lauten. Für alle Intensitätslevel war jedoch zu beobachten, daß hochfrequente Töne eine kürzere Latenz evozierten als niederfrequente. Wie Roberts & Poeppel (1996) jedoch zeigen, ist die minimale Latenz bei einer Frequenz von 1000 Hz erreicht. Die Autoren folgern daraus, daß sich die Latenz der N100m für Sinustöne aus einem Grundbetrag von etwa 100 ms und einem weiteren Betrag, der etwa drei Periodendauern der jeweiligen Stimulusfrequenz beträgt (30 ms bei 100 Hz, 3 ms bei 1000 Hz) zusammensetzt. Bei der Untersuchung komplexerer Klänge, bei denen neben der Grundfrequenz (Pitch) auch das Frequenzspektrum (Timbre) verarbeitet wird, ist der Stand der Forschung bzw. die Interpretation der Ergebnisse uneinheitlicher. Pantev und Kollegen sehen in der N100 ein Korrelat der wahrgenommenen Tonhöhe, unabhängig von der spektralen Zusammensetzung (Pantev, Elbert, Ross, Carsten & Terhart, 1996). Hingegen postulie-

ren Langner und Mitarbeiter eine periodotopische Organisation des auditorischen Kortex, der eine Auflösung aller Frequenzen durch unterschiedliche Neuronenpopulationen erlaubt (Langner, Sams, Heil & Schulze, 1997). Jedoch sind die methodologischen Unterschiede zwischen diesen Studien zu groß, als daß eine Entscheidung zugunsten der einen oder anderen getroffen werden könnte (Roberts et al., 2000). Grundsätzlich bedeutet diese Diskussion aber auch, daß Unklarheit darüber besteht, wie die N100 aus psychologischer Sicht zu interpretieren ist. Einerseits könnte sie ein bloßes Abbild der physikalischen Stimulusparameter sein, d.h. lediglich die Umkodierung akustischer Energie in korikal-neuronale Aktivität. Alternativ könnte es sich bei der N100 auch um die frühe Phase eines mentalen Repräsentationsaufbaus handeln. Sie würde dann also eher an das Perzept als an die physikalischen Eigenschaften des Inputs gebunden sein. Einen Hinweis zugunsten der letzteren Interpretation liefert die Studie von Roberts, Ferrari & Poeppel (1998). Hier wurden komplexe Klänge erstellt, die aus einer hohen (1000 Hz) und einer niedrigeren (100 Hz) Frequenz bestanden. Der relative Anteil der hohen und der niedrigeren Frequenz wurde variiert und die Töne den Probanden in randomisierter Reihenfolge dargeboten. Die Experimentalaufgabe bestand in der Entscheidung, ob der gehörte Ton eher hoch oder eher tief war. Einer kontinuierlichen Manipulation des Stimulus entsprach ein kategoriales Antwortverhalten, also eine steile Kategorisierungskurve mit einem kleinen Unsicherheitsintervall. Interessanterweise verhielt sich die Latenz der N100m ebenfalls kategorial. Sie korrelierte also eher mit der Repräsentation als mit dem akustischen Spektrum der Stimuli. Für Sprachlaute wurde ein solches Verhalten der N100m bisher noch nicht gezeigt. Zwar gibt es einige Studien die Latenzunterschiede der N100m auf verschiedene Vokalqualitäten finden (Diesch et al., 1996; Poeppel & Roberts, 1996; Poeppel et al., 1997) jedoch können diese ebenfalls mit einer rein akustischen Analyse der Formantstruktur erklärt werden, denn größere Latenzen traten immer dann auf, wenn der erste Formant niedrig war. Die Frage, inwieweit die N100 bereits höhere perzeptuelle Verarbeitungsschritte widerspiegelt bedarf also weiterer Forschung (Roberts et al., 2000).

## 4.3 Mismatch Negativierung

Die Mismatch Negativierung (MMN) kann formal als eine neuronale Reaktion auf eine Abweichung (Deviant) in einer Serie von konstanten Reizen (Standards) definiert werden. Bei den Standardreizen muß es sich nicht um physikalisch identische Stimuli handeln. Es können auch verschiedene Reize als Standards eingesetzt werden, die jedoch im Hinblick auf ein bestimmtes Merkmal konstant sind. Devianten stellen dann in Bezug auf dieses Merkmal eine Abweichung dar (siehe z.B. Aulanko, Hari, Lousna, Näätänen & Sams 1993; Paavilainen, Simola, Jaramillo, Näätänen & Winkler 2001a). Ein Deviant kann erster Ordnung sein, d.h. beispielsweise bei der Präsentation von Tönen in basalen Parametern wie der Tonhöhe oder Dauer abweichen, oder höherer Ordnung, wie eine Abweichung hinsichtlich des phonologischen Status (siehe Näätänen & Escera, 2000). Die MMN wird vornehmlich der auditorischen Modalität zugeordnet. Zwar sind ähnliche Komponenten auch für die visuelle und taktile Wahrnehmung beschrieben worden (Alho, Woods, Algazi & Näätänen, 1992; Czigler & Csibra, 1990; Kekoni, Hämäläinen, McCloud, Reinikainen & Näätänen, 1996), doch sind diese nicht wie die auditorische MMN aufmerksamkeitsunabhängig (s.u.), weisen außerdem eine andere Topographie auf und sind daher auf andere neuronale Generatoren zurückzuführen (Schröger, 1998). Die Mismatch Negativierung läßt sich als Differenzkurve zwischen dem EKP der Deviantverarbeitung und dem EKP auf Standardreize darstellen. Die Gipfellatenz dieser Differenzkurve liegt üblicherweise 100-300 ms nach Beginn des Deviants. Die MMN wird in der Regel gegen die Nasenelektrode abgeleitet und weist dann eine frontozentrale Topographie auf. Wird als Referenz eine Nasenelektrode verwendet, so kehren sich die Potentiale bei Rereferenzierung gegen Elektroden unterhalb der Sylvischen Fissur (z.B. den Mastoidelektroden) um. Dieses Inversionskriterium gilt jedoch nur für Devianten erster Ordnung und ist als Identitätskriterium für die MMN nicht unumstritten (Gomes, Bernstein, Ritter, Jr. Vaughan & Miller, 1997).

Es gibt für die MMN zwei funktionale Erklärungsansätze: Die Refraktärhypothese (Schröger, 1998) beruht auf der Tatsache, daß ein identischer repetitiver Reiz (der Standardstimulus) zu einer Abschwächung der neuronalen Reaktion auf diesen Reiz führt, da sich die Neurone in einem Refraktärzustand befinden (*neuronale Adaptation*). Auf die abweichenden Eigenschaften des Deviants reagieren nun Neurone, die sich nicht in einem Refraktärzustand befinden, und es kommt somit zu einer stärkeren Aktivität. Während diese Hypothese für Devianten erster Ordnung plausibel erscheint, ist sie jedoch für Abweichungen einer höheren Dimension nicht unbedingt einsichtig. Ebensowenig läßt sich damit die Mismatch Negativierung in Reaktion auf Reizauslassungen (Yabe, Tervaniemi, Reinikainen & Näätänen, 1997), Intensitätsabschwächung (Näätänen, 1992) oder Variationen des Interstimulusintervalles (Czigler, Csibra & Csontos, 1992) erklären. Alternativ wird die MMN daher als Ergebnis eines Vergleichsprozesses interpretiert, der einen Unterschied zwischen der neuronalen Repräsentation des aktuellen Inputs und der sensorischen Gedächtnisspur der Standardstimulation registriert. In einer aktuellen Studie zeigen Jacobsen & Schröger (2001), daß sich die beiden Erklärungshypothesen für Abweichungen der Frequenz nicht ausschließen müssen. Die Autoren präsentierten Sinustöne entweder als Devianten bei einem konstanten Standardton (Oddball-Block) oder mit gleicher Wahrscheinlichkeit in einer Folge unterschiedlicher Töne, die sich nicht unmittelbar wiederholen (Kontroll-Block). In der Oddball-Bedingung findet sich wie erwartet eine deutliche MMN. Für die Kontroll-Bedingung, in der die Refraktärhypothese ausgeschlossen werden muß, findet sich ebenfalls eine Negativierung, die jedoch deutlich schwächer ist als für den Oddball-Block. Die Autoren schließen daraus, daß für Frequenzabweichungen der Anteil gedächtnisbasierter Aktivität nur gering ist, während der Hauptteil der MMN mit der Refraktärhypothese erklärt werden muß.

Gegenüber Manipulationen der Aufmerksamkeit erwies sich die Mismatch Negativierung als außerordentlich resistent. So berichtet Schröger (1998) in einer Übersicht einige Studien, die eine MMN bei schlafenden Säuglingen und Koma-Patienten fanden. Experimentelle Settings, bei denen die Aufmerksamkeit nicht auf den Deviant gelenkt ist, liefern aber nicht nur eine stabile Mismatch Negati-

vierung, sondern haben zudem den Vorteil, daß die MMN nicht von Aufmerksamkeitskomponenten wie der P300 überlagert ist. Lediglich eine Primäraufgabe, mit der eine starke Bindung der Aufmerksamkeit einhergeht, wie z.B. eine dichotische Höraufgabe, führte zu einer Abschwächung der MMN (Woldorff, Hackley & Hillyard, 1991). Sehr subtile Deviationen, wie eine Verkleinerung des Interstimulusintervalls um lediglich 100 ms evozierten hingegen nur dann eine MMN, wenn der Aufmerksamkeitsfokus auf der Detektion dieser Abweichung lag, nicht aber, wenn die Probanden während der Stimulation ein Buch lasen (Schröger & Wolff, 1996). Insgesamt gilt die MMN als eine sehr robuste Komponente, die aufgrund ihrer Automatizität breite Anwendung in der klinischen Forschung und Diagnose findet (Näätänen & Escera, 2000). Für die experimentelle Forschung ist die Mismatch Negativierung jedoch nicht zuletzt auch deswegen interessant, weil sie Aussagen über mentale Langzeitrepräsentationen bestimmter akustischer Reize erlaubt. So weist die Amplitude der MMN im Gegensatz zur eher exogenen M100 eine Abhängigkeit von der phonologischen Repräsentationen der Sprachlaute im Gehirn auf; sie ist im Vergleich zu prototypischen Phonemdevianten geringer, wenn der Deviant ein nicht-prototypisches, z.B. fremdsprachliches Phonem umfaßt (Näätänen, Lehtokoski, Lennes, Cheour, Houtilainen, Ilvonen et al., 1997; Winkler, Kujala, Tiitinen, Sivonen, Alku, . et al., 1999; Aaltonen, Niemi, Nyrke & Tuhkanen, 1987; Phillips, Pellathy, Marantz, Yellin, Wexler, Poeppel et al., 2000; Ikeda, Hayashi, Hashimoto, Otomo & Kanno, 2002).

Aufgrund ihrer Gipfellatenz von 100-300 ms zählt die MMN zur Gruppe der N200 Komponenten . Sie wird auch als N2a bezeichnet und so von der aufgabenabhängigen N2b unterschieden. Das magnetische Korrelat der Mismatch Negativierung, das Mittels der Magnetenzephalographie gemessen werden kann, wird als MMNm oder als Mismatch Field (MMF) bezeichnet. Die neuronalen Generatoren der Mismatch Negativierung sind mit Hilfe der Elektroenzephalographie (Scherg, Vajsar & Picton, 1989; Giard, Perrin, Pernier & Boiuchet, 1990; Giard, Lavikainen, Reinikainen, Perrin, Bertrand, Pernier et al., 1995), der Magnetenzephalographie (Hari, Hämäläinen, Ilmionemi, Kaukoranta, Alho, Näätänen et al., 1984; Scherg & Berg, 1991), der Positronen-Emissions Tomographie (Tervaniemi, Medvedev, Al-

ho, Pakhomov, Roudas, van Zuijen et al., 2000) und der funktionellen Kernspintomographie (Opitz, Mecklinger, von Cramon & Kruggel, 1999; Celsis, Boulanouar, Ranjeva, Berry, Nespoulous & Chollet, 1999) untersucht worden. Diese Lokaliationsstudien deuten darauf hin, daß ihre Hauptgeneratoren im Temporallappen nahe des auditorischen Kortex liegen. Zusätzlich zu der temporalen Hauptaktivierung wird eine frontale Subkomponente der MMN berichtet (Giard, Perrin, Pernier & Boiuchet, 1990; Opitz, Mecklinger, von Cramon & Kruggel, 1999; Rinne, Alho, Ilmoniemi, Virtanen & Näätänen, 2000). Da diese frontale Aktivierung zeitlich auf die Hauptkomponente folgt, wird vermutet, daß sie bereits zu einer Kette von Prozessen gehört, die zu einer unwillkürlichen Aufmerksamkeitsverlagerung (*attentional switch*) führen. Da diese frontale Subkomponente bei magnetenzephalographischen Messungen nicht zu beobachten war, wird vermutet, daß es sich dabei um gyrale Aktivierungen handelt, deren radiales Magnetfeld im MEG nicht detektierbar ist (Rinne et al., 2000).

## 4.4 N2b und P300

Die Mismatch Negativierung (auch N2a) zeichnet sich durch ein hohes Maß an Automatizität aus und tritt relativ unabhängig von der Experimentalaufgabe auf. Muß der Proband jedoch den Devianten aktiv vom Kontextreiz unterscheiden, so folgt der Mismatch Negativierung, bzw. überlagert diese ein weiteres negatives Potential, die N2b. Entscheidendes Kriterium für das Auftreten einer N2b ist, daß die präsentierte Abweichung des perzeptuellen Inputs direkte Aufgabenrelevanz besitzt (Ritter, Paavilainen, Lavikainen, Reinikainen, Alho, Sams et al., 1992; Novak, Ritter & Vaughan, 1992). Die N2b ist modalitätsunspezifisch und wird funktional als Korrelat einer bewußten Stimulusdiskrimination (Halgren, Marinkovich & Chauvel, 1998; Kekoni, 1998), bzw. eines Orientierungsprozesses hin zum relevanten Reiz (Opitz, 1999; Pritchard, Shappell & Brandt, 1991) interpretiert. In EKP-Studien zeigt die N2b häufig eine maximale Amplitude an zentralen Elektroden, eine Polaritätsinversion wie bei der MMN tritt aber nicht auf. StDer neuronale

Hauptgenerator der N2b wird aufgrund von Studien mit interkraniellen Elektroden im rhinalen Kortex vermutet (Halgren, Marinkovich & Chauvel, 1998).

Im Zusammenhang sowohl mit der Mismatch Negativierung als auch mit der N2b wird häufig eine positive EKP-Komponente beobachtet, die etwa 300-350 ms nach dem devianten Ereignis ihr Maximum erreicht und folglich als P300 bezeichnet wird (Sutton, Braren, Zubin & John, 1965). Verschiedentlich ist auch von einer *late positive component* (*LPC*) odereinem *late positive complex* (ebenfalls *LPC*) die Rede. N200 und P300 Komponenten können zwar in bestimmten experimentellen Situation unabhängig voneinander evoziert werden (Sams et al., 1985; Ritter et al., 1992), in der Regel treten sie jedoch zusammen auf, so daß von einem "N2/P3 Muster" die Rede ist, bzw. von einem "N2/P3/SW Komplex", wenn sich eine breite Negativierung (slow wave) anschließt (z.B. Halgren et al., 1998). Ähnlich wie bei den N200 Komponenten richtet sich die Klassifikation von funktionalen Varianten der P300 nach Automatizität und Aufgabenrelevanz. Die P3a (Squires, Squires & Hillyard, 1975) tritt dann auf, wenn die präsentierte Abweichung keine Aufgabenrelevanz besitzt, wobei ein subjekiv selteneres Ereignis, ebenso wie eine größerer Unterschied zwischen Deviant und Kontext zu einer größeren P3a-Amplitude führen (siehe Pritchard, 1991). Die sogenannte Novel P3 tritt bei völlig neuartigen, nicht repetitiven Devianten auf. Obwohl es Hinweise darauf gibt, daß es sich bei der Novel P3 um eine komplexere Komponente handelt als bei der P3a (siehe Opitz, 1999; Comchero & Polich, 1998, 1999), werden die Begriffe bisweilen synonym gebraucht (z.B. Friedman et al., 2001; Knight, 1996), da beide eine ähnliche topographische und zeitliche Struktur aufweisen. Sowohl die Novel P3 als auch die P3a auf typische (d.h. nicht neuartige) Stimuli sind durch eine relativ kurze Latenz, eine große Amplitude und ein anteriores Maximum charakterisiert, das auf neuronale Generatoren im Frontallappen zurückgeführt wird. Funktional werden die P3a und die Novel P3 als Korrelate einer unwillkürlichen Aufmerksamkeitsverlagerung und Orientierungsreaktion interpretiert. Ist die Detektion der Abweichung im Oddball Experiment aufgabenrelevant, so zeigt das EKP im Zeitbereich um 300 ms eine weitere postitive Komponente, die auf der

Kopfoberfläche parietal verteilt ist,[1] und eine etwas längere Gipfellatenz als die P3a aufweist (Squires, Squires & Hillyard, 1975; Courchesne, Courchesne & Hillyard, 1978). Dieses Potential wurde zuerst von Sutton et al. (1965) beschrieben und wird als P3b bezeichnet. Es wird angenommen, daß die P3b Pozesse der aufgabenabhängigen Stimulusevaluation reflektiert, wobei eine schwierigere Experimentalaufgabe zu einer geringeren P3b-Amplitude führt. Es wurde daher vorgeschlagen, daß die P3b Amplitude den Grad der Konfidenz des Probanden bei der Entscheidung reflektiert (Hillyard et al., 1971). Ein interessantes Paradigma zur Untersuchung von P3a und P3b ist das *3-Stimulus-Design*. Hier werden neben einem Kontextreiz (Standard) zwei infrequente, deviante Reize präsentiert, von denen einer mit einer Experimentalaufgabe verbunden ist (Target), der andere nicht (NonTarget). Ist die Standard/Target Diskrimination leicht, so geht damit eine größere Amplitude der Target-P300 als der NonTarget-P300 einher. Ist die Diskrimination jedoch schwierig, so ist die NonTarget-P300 besonders an den frontalen Elektroden größer, während die Target-P300 eine geringere Amplitude und eine deutliche parietale Topographie aufweist (Comchero & Polich, 1998, 1999). Ist der NonTarget-Reiz jedoch dem Standard sehr ähnlich, so ist die NonTarget-P300 geringer, die Target Verarbeitung verläuft jedoch unabhängig davon (Comchero & Polich, 1999). Es folgt daraus, daß die beiden Komponenten relativ unabhängig voneinander sind und vermutlich auf unterschiedliche neuronale Generatoren zurückzuführen sind (Comchero & Polich, 1999; Halgren et al., 1998), und daß die Schwierigkeit der Primäraufgabe, die über die Verteilung von Aufmerksamkeitsressourcen entscheidet, einen Einfluß auf die Latenz und Amplitude beider Komponenten hat. Die P300-Komponenten sind modalitätsunspezifisch und werden mit kognitiven Prozessen wie der bewußten Stimulusevaluation bzw. eine Erneuerung der mentalen Umweltrepräsentation (*context updating*) in Verbindung gebracht (Näätänen & Picton, 1986; Donchin & Coles, 1988; Don-

[1] Die topographischen Angaben gelten für die Verwendung von Skalpelektroden. Halgren, Marinkovich & Chauvel (1998) berichten in ihrem Übersichtsartikel präoperative Studien mit Tiefenelektroden, die für die N2/P3 Komplexe weitverteilte Netzwerke von Generatoren beobachteten. Für die P3a wurden dort Generatoren im inferioren Parietallappen, im Gyrus cinguli und im dorsolateralen präfrontalen Kortex berichtet, für die P3b im Hippocampus, im superioren temporalen Sulcus, im lateralen orbitofrontalen Kortex, sowie im Sulcus interparietalis.

chin, Spencer & Dien, 1997; siehe aber auch Johnson, 1993 und Verleger, 1988).

## 4.5 N400

Als Beginn der elektrophysiologischen Untersuchung von Sprachverstehensprozessen, können die ersten Studien zur N400-Komponente von Kutas und Hillyard (Kutas & Hillyard, 1980a,b) angesehen werden. Diese Studien befaßten sich nicht mehr mit der basalen Verarbeitung physikalisch devianter Reize, sondern untersuchten die Verarbeitung semantischer Verletzungen. Bei dem ursprünglich verwendeten Paradigma wurden Sätze Wort für Wort auf einem Monitor präsentiert, während die ereigniskorrelierten Potentiale des lesenden Probanden abgeleitet wurden. Bei einem Teil der Sätze beinhaltete das satzfinale Wort eine semantische Verletzung, z.B. *"He spread the warm bread with socks."* statt *"He spread the warm bread with butter."*. Etwa 400 ms nach Beginn dieser Verletzung war im Vergleich zu den korrekten Sätzen eine ausgeprägte Negativierung zu beobachten.
In Folgeexperimenten stellte sich heraus, dass dieser N400-Effekt Schwierigkeiten bei der semantischen Integration eines Wortes in den Kontext abbildet. Die Amplitude der N400 variiert mit dem Grad der Verletzung,[2] ist aber auch für semantisch korrekte, doch assoziativ unerwartete Wörter beobachtet worden. So berichten z.B. Fischler, Bloom, Childers, Arroyo & Perry (1984) einen N400-Effekt für das unerwartete satzfinale Nomen in semantisch korrekten Sätzen wie "A robin is not a tool".
Formal ist der N400-Effekt also definiert als eine stärkere Negativierung im EKP auf ein Wort, das den semantischen Erwartungen widerspricht, im Vergleich zu einem Wort, daß an derselben Stelle im Satz die semantischen Erwartungen

[2] Kutas & Hillyard (1984) beobachteten eine geringere N400-Amplitude, wenn das kontextuell inkongruente Wort mit dem erwarteten Wort semantisch verwandt war, als wenn es semantisch unrelatiert war. Im Beispielsatz *"The pizza was too hot to eat/drink/cry"* wurde für jedes der beiden semantisch inkongruentn Wörter (*drink/cry*) eine deutliche N400 ausgelöst. Da *drink* jedoch semantisch assoziiert ist mit dem erwarteten Infinitiv *eat*, war in diesem Falle die N400 schwächer als bei dem unrelatierten Verb *cry*.

erfüllt.[3] Da nicht nur ein Satz einen kontextuellen Rahmen stellen kann, wird auch bei semantisch unrelatierten Wortpaaren ein N400-Effekt beobachtet (Holcomb & Neville, 1990) ebenso wie bei nichtsprachlichem Stimulusmaterial wie Bildern (**?**McPherson & Holcomb, 1999) und Gesichtern (Bobes, Valdes-Sosa & Olivares, 1994).

Desweiteren zeigt sich für die N400 eine Abhängigkeit von der relativen Häufigkeit eines Wortes in einer Sprache (sie ist bei niederfrequenten Wörtern grundsätzlich größer). Eine Kontrolle der Wortfrequenz und kontextuellen Plausibilität ist im Experiment daher unerläßlich. Als generelles Maß für die Auftretenswahrscheinlichkeit eines Wortes in einem Kontext wird die sogenannte CLOZE-Wahrscheinlichkeit (Taylor, 1953) ermittelt. Hierbei werden Satzrahmen vorgegeben und Testpersonen sollen das fehlende Wort ergänzen (z.B. *"Die Feuerwehr hat den heißen Brand ...*). Anschließend wird die CLOZE-Wahrscheinlichkeit berechnet als relative Häufigkeit eines Wortes (z.B. *"gelöscht"*) für den jeweiligen gegebenen Satzrahmen. Eine Wiederholung identischer Sätze führt zu einer Verringerung der N400, was insbesondere für niederfrequente Wörter gilt (Rugg, 1990; Besson, Kutas & Van Petten, 1992). Bei der Wiederholung eines Wortes in anderem Satzkontext wird dieser Reduktionseffekt allerdings nicht beobachtet (Besson et al., 1992).

Die N400 wurde sowohl für die visuelle Domäne (Lesen) als auch bei auditorischer Präsentation berichtet. Während sie in der visuellen Modalität eine etwas rechtslateralisierte zentro-parietale Verteilung aufweist, zeichnet sich die N400 in der auditorischen Domäne durch eine bilaterale oder auch linkslaterale Topographie aus, sowie durch eine etwa 100 ms kürzere Latenz. Die auditorisch evozierte N400 ist außerdem zeitlich ausgedehnter, was im allgemeinen darauf zurückgeführt wird, daß der akustische Input ebenfalls ein zeitlich ausgedehntes Signal ist. Desweiteren könnte von Bedeutung sein, daß die Dauer und die Zeitpunkte der jeweiligen Worterkennung (*recognition points*) für die auditorischen Sti-

[3] Wie Van Petten & Kutas (1991) zeigten, tritt eine N400-artige Negativierung prinzipiell für jedes Wort im Satz auf. Sie ist am Satzanfang größer und nimmt mit zunehmender kontextueller Information wird sie geringer, da die semantische Integration leichter wird. Der N400-Effekt meint hingegen eine *stärkere* Negativierung auf einen unerwarteten Begriff im Vergleich zum EKP auf ein erwartetes Wort.

muli zeitlich recht verschieden sind und es deshalb bei der EKP-Mittelung zu einer Verschmierung (*Jitter*) der Komponente kommt. Unterstützung erfuhren diese Annahmen durch eine Studie von Mäkelä und Kollegen, die für semantisch inkongruente satzfinale Wörter eine breite N400 Komponente fanden, wenn diese eine lange und variable Dauer besaßen, hingegen eine scharf umrissene N400, wenn diese Wörter eine kurze und relativ invariable Dauer besaßen (Mäkelä, Mäkinen, Nikkilä, Ilmoniemi & Tiitinen, 2001). Außerdem wurden die neuronalen Generatoren der auditorischen N400m mittels einer Dipolmodellierung beschrieben. Sie lagen in beiden Hemisphären nahe des auditorischen Kortex. Hemisphäreneffekte der Dipolstärken, Latenzen oder Lokalisationen waren nicht zu beobachten. In einer vergleichbaren MEG Studie wurde hingegen eine stärkere Aktivierung der linken Hemisphäre beobachtet, die dem N400 Effekt in der rechten Hemisphäre zeitlich vorausging (Friederici, Wang & Maess, 2001). Dieses Aktivierungsmuster wurde auf das Zusammenspiel von lexikalischer Verarbeitung im linken medialen temporalen Gyrus und den etwas späteren, bilateralen Assoziations- bzw. Integrationsprozessen zurückgeführt. Außerdem wurden in dieser Studie frontale Aktivierungsfoki gefunden, die evtl. Prozesse des episodischen Gedächtnisses widerspiegeln könnten. Diese Daten sind im Einklang mit der gängigen Annahme, daß an der Generierung der N400 ein Netzwerk mehrerer Prozesse und neuronaler Strukturen beteiligt ist (Kutas & Federmeier, 2000).[4]

Die hier vorgestellten Studien haben gezeigt, daß die Registrierung und Analyse ereigniskorrelierter Hirnpotentiale die Untersuchung unterschiedlicher Stufen der auditorischen Sprachverarbeitung erlauben. Die dargestellten Methoden wurden in der vorliegenden Arbeit genutzt, um die auditorische Perzeption von Stimminformation zu untersuchen. Um die theoretische Motivation der Arbeit darzulegen, wird im folgenden Teil eine Einführung in ihre inhaltliche Fragestellung gegeben.

[4]Wie bereits die EKP-Studien, so weisen auch die Lokalisationsstudien auf eine Abhängigkeit der N400 Komponente von der Präsentationsmodalität hin. Allerdings sind die Ergebnisse hier eher uneinheitlich; zu den neuronalen Generatoren der N400 beim Lesen von Sätzen siehe Helenius, R, Service & Connolly (1998); Nobre & McCarthy (1994, 1995); Halgren, Baudena, Heit, Clarke & Marinkovich (1994)

# Teil II

# Verarbeitung von Stimminformation

# Kapitel 5

# Produktion, Akustik und Wahrnehmung lautsprachlicher Signale

Das wichtigste Medium der menschlichen Kommunikation ist die lautsprachliche Äußerung, d.h. die regelgeleitete Produktion eines akustischen Signals durch den Sprecher, das durch den Hörer dekodiert wird. Die phonetische Produktion eines Sprachsignals läßt sich mit Hilfe des Quelle-Filter-Modells (Fant, 1960) beschreiben: Durch periodische Schwingungen der Stimmlippen im Kehlkopf (*Quelle*) entsteht ein Signal, das durch den Vokaltrakt, also von den supraglottalen Strukturen des Rachen- , Mund- und Nasenraumes "gefiltert" wird: Bedingt durch Reflexion und Absorption der Schallwellen werden bei der Vokalisation bestimmte Frequenzanteile des Quellsignals verstärkt, andere gedämpft (siehe Abbildung 5.1).

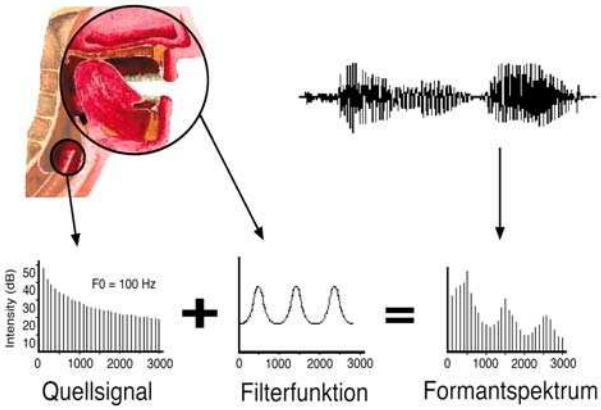



Abbildung 5.1: Das Quelle-Filter-Modell (schematische Darstellung). Durch die Stimmlippenschwingung im Kehlkopf entsteht das Quellsignal, das eine bestimmte Grundfrequenz (F0) besitzt und durch die Resonanzeigenschaften des Vokaltrakts verändert wird. Diese Veränderungen lassen sich durch eine entsprechende mathematische Filterfunktion beschreiben. In Kombination mit den akustischen Eigenschaften der Signalabstrahlung an den Lippen, die eine zusätzlichen Dämpfung höherer Frequenzen bedeuten (im Bild nicht dargestellt), resultiert ein charakteristisches Frequenzspektrum, dessen Gipfel als Formanten (F1,F2...) bezeichnet werden. Bei der Produktion verschiedener stimmhafter Laute ändert sich aufgrund der artikulatorischen Bewegungen von Lippen und Zunge die Filterfunktion und damit auch das Formantspektrum. Bei stimmlosen Konsonanten (wie /s/ oder /k/) wird das Quellsignal ersetzt durch eine Friktion oder die Bildung und Lösung eines artikulatorischen Verschlusses. Die daraus resultierende hochfrequente Rauschspektren werden ähnlich wie das Stimmsignal von den übrigen Teilen des Vokaltrakts "gefiltert".

Diese Resonatorfunktion des Vokaltrakts ist abhängig von seiner jeweiligen geometrische Struktur, a) von den dynamischen Veränderungen während der Artikulation und b) von den anatomisch-physiologischen Eigenschaften des Sprechers (wie z.B. den Proportionen des Rachenraumes).

Bei der Sprachwahrnehmung wird nun das akustische Signal in Einzellaute, bzw. in phonologisch distinktive, funktionale Segmente (*Phoneme*) dekodiert. Dabei ist allerdings noch weitgehend unklar, wie diese Rücktransformation der Schallwellen in die bedeutungstragenden Einheiten geschieht. Die Annahme *akustischer Invarianzen* besagt, daß jedes Phonem sich als eine charakteristische Kombination akustischer *Cues*[1] beschreiben läßt, die bei jedem Auftreten des Phonems

[1] Es ist hier zu unterscheiden zwischen den phonetischen *Merkmalen*, die im Englischen als *features*

vorhanden sein müssen und so seine Identifikation durch den Hörer ermöglichen. Desweiteren sollten die Sprachlaute bei der Produktion zeitlich aufeinander folgen und als eine solche zeitliche Abfolge auch wieder zu dekodieren sein (*Linearitätsbedingung*). Mit der Entwicklung akustischer Analyseverfahren wurde jedoch bald deutlich, daß diese Annahmen für die Lautsprache, insbesondere für die kontinuierliche Spontansprache nicht zutreffen:

> *It has become apparent to speech researchers over the last 40 years that the invariance and linearity conditions are almost never met in the actual speech signal.* [S. 3]
> *The assumption that a small set of acoustic features or attributes in the acoustic signal provide cues to linguistic contrast was the motivation for the search for invariant cues which was central to speech perception research from the mid-50s up until the present time. The more the signal was explored for invariant acoustic cues the more the problems of variability and non-linearity became evident.* [S. 6]
>
> Wright, Frisch und Pisoni 1997

Die Probleme von akustischer Varianz und Nicht-Linearität sind eng miteinander verbunden, denn eine Ursache der akustischen Varianz liegt im Einfluß der lautlichen Umgebung auf die artikulatorische und damit auch akustische Ausprägung eines Sprachsegments. Dieses Phänomen wird als Koartikulation beschrieben. Andere Varianzfaktoren sind beispielsweise die Silbenposition des jeweiligen Lautes, oder die Akzentverteilung des Wortes. Zu diesen strukturell bedingten Einflüssen kommt die Tatsache, daß selbst zwei phonologisch gleiche Äußerungen eines Sprechers artikulatorisch und akustisch niemals völlig identisch sind. Insbesondere Änderungen des Sprechtempos oder der Artikulationspräzision erhöhen diese *Intra-Sprechervariabilität*. Noch stärker sind die akustischen Unterschiede allerdings zwischen den Sprachsignalen verschiedener Sprecher (*Inter-Sprechervariabilität*). Diese lassen sich als Unterschiede a) des Quellsignals oder b) der Filterfunktion beschreiben.

Das wichtigste Merkmal des Quellsignals ist die Grundfrequenz (F0), welche die Geschwindigkeit der Stimmlippenschwingung beschreibt und sich perzeptuell in

bezeichnet werden (wie z.B. [+stimmhaft] oder [+labial]) und ihrer akustischen Substanz, den *Cues* wie z.B. eine bestimmte Voice-Onset Time oder eine charakteristische Formanttransition, die eine Perzeption der distinktiven Merkmale erst ermöglichen.

der Tonhöhe (engl. *pitch*) niederschlägt.[2] Beispielwerte für die mittlere Grundfrequenz männlicher und weiblicher Sprecher sind in Tabelle 5.1 aufgeführt.

| **Sprecher** | **F0** |
|---|---|
| männlich | 100-150 Hz |
| weiblich | 190-250 Hz |

Tabelle 5.1: Typische Werte für die Grundfrequenz weiblicher und männlicher Sprecher nach Neppert & Pétursson (1992). Es ist bei diesen Werten zu beachten, daß sie aus Untersuchungen einer Vielzahl von untrainierten Sprechern und Sprecherinnen und mit sehr unterschiedlichen Äußerungstypen. Gerade bei professionellen Sprecherinnen werden häufig deutlich tiefere Mittelwerte gefunden (siehe Eckert & Laver, 1994; Slembek, 1998).

Ein weiterer Parameter des Quellsignals ist die Stimmqualität. Aus akustischen Untersuchungen ist bekannt, daß sich männliche und weibliche Stimmen hinsichtlich der glottalen Schwingungsfunktion unterscheiden, was z.B. zu einem unterschiedlichen Grad von Behauchtheit führt (z.B. Mendoza et al., 1996). In Wahrnehmungsexperimenten hat sich dieser Parameter allerdings als unbedeutende für die Sprecheridentifikation herausgestellt (Lavner et al., 2000; Kreiman, 1997; Mullenix et al., 1995).

Unterschiede in den Filterparametern sind vor allem auf anatomische Unterschiede zwischen den Vokaltrakten zurückzuführen. Wie zwei verschiedene Instrumente (z.B. Klarinette und Geige) einen Ton gleicher Tonhöhe, aber mit jeweils charakteristischem Timbre produzieren können, so sind auch menschliche Stimmen aufgrund ihrer spektralen Timbre-Zusammensetzung unterscheidbar. Diese spezifischen Resonanzeigenschaften ergeben sich aus den anatomischen Besonderheiten des jeweiligen Vokaltrakts. Z.B. zeichnen sich die Vokaltrakte männlicher Sprecher im Vergleich zu denen weiblicher Sprecherinnen durch einen

[2] Auch die F0 Kontur bei der Satzäußerung ist ein Ausdrucksmittel, das von verschiedenen Sprechern unterschiedlich eingesetzt wird und daher zur Sprecheridentifikation genutzt werden kann (Abberton & Fourcin, 1978). Es sei jedoch darauf hingewiesen, daß in der vorliegenden Arbeit eher diejenigen Sprechermerkmale von Bedeutung sind, die unmittelbar mit Beginn des Sprecherwechsels wahrgenommen werden, statt der F0 Kontur also eher das F0 Register (pitch level), das eine hohe von einer tiefen Stimme unterscheidet.

überproportional größeren Rachenraum aus. Dieser anatomische Unterschied bewirkt bei der Lautproduktion spektrale Unterschiede im Bereich des dritten und vierten Formanten (siehe Abbildung 5.2). D.h. die Kombination aus Quellsignal und Vokaltraktfilter bewirkt das charakteristischen Formantspektrum bzw. das Timbre der individuellen Stimme.

Allerdings sind die individuell verschiedene Realisierungen einer Äußerung nicht nur auf physiologische Beschränkungen zurückzuführen. So wäre es beispielsweise rein physiologisch durchaus möglich, daß - zumindest alle Vertreter desselben Geschlechts - auf demselben Grundfrequenzniveau sprechen. Vielmehr spielt hier der “Sprechhabitus“, d.h. die individuellen Sprechgewohnheiten des Einzelnen eine Rolle, die eine Unterscheidung verschiedener Stimmen und ggf. das Wiedererkennen bekannter Sprecher erlauben (Laver & Mackenzie-Beck, 2001).

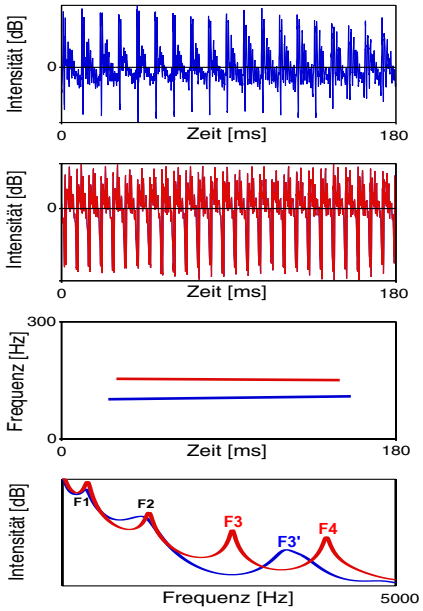



Abbildung 5.2: Beispiel für akustische Unterschiede zwischen einer männlichen und einer weiblichen Stimme: Der Vokal /a/ wurde hier produziert von einer weiblichen und einer männlichen Stimme: **1)** das Oszillogramm der männlichen Stimme; **2)** das Oszillogramm der weiblichen Stimme; **3)** die Grundfrequenzverläufe, bestimmt durch das Quellsignal [rot - weibliche Stimme; blau - männliche Stimme] **4)** Formantspektren bestimmt durch die Filterwirkung des Vokaltrakts [rot - Spektrum der weiblichen Stimme; blau - Spektrum der männlichen Stimme])

# Kapitel 6

# Sprechernormalisierung und Verarbeitung von Sprecherinformation

Die sprecherspezifischen Anteile im Sprachsignal tragen einen Informationsgehalt, der eine Einschätzung (Kategorisierung) des Sprechers, und gegebenenfalls sogar seine Identifikation erlaubt. So umfassen diese Sprechercharakteristika temporäre Merkmale wie den Gesundheitszustand, dauerhaft erworbene Eigenschaften wie den dialektalen Hintergrund und schließlich physiologische Parameter wie das ungefähre Alter und das Geschlecht des Sprechers (vgl. Brown, 1980; Nolan, 1983; Hecker, 1971).

Zahlreiche Studien versuchten mit einer multidimensionalen Analyse die entscheidenden akustischen Parameter für die Sprecheridentifikation herauszufinden (für eine Übersicht siehe Kreiman, 1997). Dabei stellten sich sowohl die Tonhöhe (Grundfrequenz) als auch das Timbre (Formantstruktur) als entscheidende Faktoren heraus. Zwischen einzelnen Studien ergaben sich jedoch Unterschiede in der Gewichtung der beiden Signalparameter. In einer kürzlich publizierten Experimentalserie untersuchten Lavner, Gath & Rosenhouse (2000) ausführlich die

relative Bedeutung verschiedener Signalparameter für die Identifikation bekannter Stimmen. Als Stimulusmaterial verwendeten sie den Vokal /a/ geäußert von mehreren Sprechern. Diese Sprachsignale wurden gezielt mit Hilfe verschiedener Filtertechniken, die beispielsweise eine isolierte Verschiebung einzelner Formanten ermöglichen, manipuliert und der resultierende Effekt auf die Sprechererkennung untersucht. Allgemein erwies sich die spektrale Zusammensetzung des Signals, insbesondere der 3. und 4. Formant von Bedeutung. Van Dommelen (1990) zeigte hingegen in seinen Experimenten zur Stimmerkennung, daß das Frequenzspektrum zwar eines der stärksten Wiedererkennungsmerkmale ist, für Personen mit auffallend hohen oder auffallend tiefen Stimmen erwies sich jedoch die Grundfrequenz als besonders bedeutsam. Was distinktive akustische Sprechermerkmale betrifft, zeigen sich Unterschiede also nicht nur zwischen den Experimenten, sondern auch innerhalb einzelner Studien in Abhängigkeit von den präsentierten Stimmen. Ein extremer Fall sind in diesem Zusammenhang die Stimuli, die Remez, Fellowes & Rubin (1997) in ihren Experimenten verwendeten. Dort wurde das gesamte Stimmsignal durch einen Sinuston ersetzt. Somit waren zwar die artikulatorisch-phonetischen Merkmale erhalten, das Timbre jedoch drastisch reduziert. Die Probanden waren nun trotz dieser starken Manipulation in der Lage, die Sprecher zu unterscheiden. Daß sie diese Aufgabe bewältigen konnten, muß aber im Gegensatz zu der von Remez formulierten Schlußfolgerung nicht unbedingt bedeuten, daß diese Analyse der rein phonetischen Merkmale auch bei der Erkennung natürlicher Stimmen eine besondere Rolle spielt.

Es ist oben im Zusammenhang mit der Grundfrequenz beschrieben worden, daß sich das Wiedererkennungssystem auf besonders saliente akustische Merkmale, bzw. auf die Abweichungen von der prototypischen Stimme verläßt. Wie andere Studien zeigen, werden solche untypischen Stimmen auch nach einem Zeitraum von mehreren Wochen besser wiedererkannt. Dagegen werden "durchschnittlich" klingende Stimmen häufig fälschlicherweise als *bekannt* gewertet (Kreiman, 1997). Stimmenvarianz erleichtert also die Sprecherunterscheidung. Hingegen ist anzunehmen, daß sie die phonetische Mustererkennung bei der Sprachverarbeitung erschwert, da die Sprechercharakteristika zu unterschiedlichen Ausprägungen

der akustischen Cues führen. Doch auch bei physikalisch sehr unterschiedlichen Sprachsignalen gelingt es den Hörern in der Regel mühelos und scheinbar automatisch die enkodierte Nachricht zu verstehen. Wie das menschliche Perzeptionssystem diese Leistung vollbringt, ist Gegenstand der Untersuchungen zur *Sprechernormalisierung*.

In behavioralen Studien gelang es, den subtilen Einfluß der Stimminformation auf die Sprachverarbeitung nachzuweisen. Ladefoged & Broadbent (1957) berichten, daß die Identifikation einzelner synthetisierter Vokale abhängig von der Formantstruktur der Vokale in einem vorangestellten Trägersatz ist. Die Autoren vermuteten daher, daß sich das Wahrnehmungssystem sich an den jeweiligen Sprecher adaptiert. Besteht ein erhöhter Adaptationsbedarf, so geht dieser mit einem Anstieg der Verarbeitungskosten einher. Mullenix, Pisoni & Martin (1989) zeigten beispielsweise, daß bei gleichzeitiger Präsentation eines Hintergrundrauschens die Wortidentifikation besonders dann erschwert war, wenn die Wörter von wechselnden Sprecher geäußert wurden. Auch bei einer Nachsprechaufgabe erfolgte die Wortidentifikation immer dann langsamer und fehlerhafter, wenn der Sprecher zwischen den einzelnen Wörtern wechselte, als wenn die Äußerungen eines Sprechers geblockt dargeboten wurden. In dieselbe Richtung weisen klinische Befunde von Ziegler, Jochmann & Zierdt (1999). In einer Silbendiskriminationsaufgabe erzielten Aphasiker[1] wesentlich schlechtere Resultate, wenn die Silben von unterschiedlichen Sprechern stammten, als wenn sie von demselben Sprecher geäußert wurden (die Probanden machten mehr Fehler, die mittlere Reaktionszeit war signifikant länger). Die gesunde Kontrollgruppe machte zwar nur wenig Fehler, zeigte aber ebenfalls eine signifikante Verzögerung der Antwortreaktion bei wechselndem Sprecher. Als Erklärung für solche Ergebnisse, die auf einen zusätzlichen Verarbeitungsaufwand beim Sprecherwechsel hindeuten, wurde ein zusätzlicher, möglicherweise zeitlich vorausgehender Norma-

[1] Unter einer Aphasie versteht man eine Störung der Sprachverarbeitung und/oder Sprachproduktion, die auf hirnphysiologische Ursachen zurückzuführen ist. Je nach dem Ort und dem Ausmaß der Läsion können unterschiedliche Funktionen betroffen sein. In der erwähnten Studie von Ziegler und Kollegen handelte sich um eine sehr heterogene Gruppe von Aphasikern unterschiedlicher Syndrome. Allen gemeinsam war jedoch Störungen der phonologischen Verarbeitung.

lisierungsprozeß postuliert. Die *abstraktionistischen Normalisierungstheorien*[2] betrachten sprecherspezifische Information als unerwünschtes "Rauschen", das die Extraktion der sprachlichen Nachricht aus dem Signal erschwert (Ladefoged & Broadbent, 1957; Sussman, 1986; Bladon, Henton & Pickering, 1984; Syrdal & Gopal, 1986). Sprachperzeption beinhaltet nach ihrer Ansicht eine Normalisierung im Sinne einer Transformation des Signalinhalts auf eine abstrakte linguistische Ebene, auf der die sprechereigentümlichen Merkmale nicht mehr vorhanden sind.[3] In diesem "Abstreifen" der sprecherspezifischen Information, sah man eine Erklärung dafür, wie ein kontinuierliches und variables Sprachsignal in diskrete Einheiten dekodiert werden kann, und somit auf der kognitiven Ebene eine perzeptuelle, phonetische Konstanz entsteht (Wright et al., 1997). Probleme bereiten dem strikt abstraktionistischen Ansatz aber jene Befunde, die auf eine akustisch-physikalische Repräsentation des Sprachsignals hinweisen. Daß sich das kurzzeitige Behalten von Wortlisten unter Bearbeitung einer zusätzlichen Gedächtnisaufgabe als besonders schwierig erwies, wenn die Wörter von verschiedenen Sprechern stammten, deutet bereits darauf hin, daß die Sprecherinformation nicht einfach abgestreift wird (Martin, Mullenix & Summers, 1989). Auf eine längerfristige Speicherung der akustischen Stimmparameter weisen eine ganze Reihe von Studien impliziter als auch expliziter Gedächtnisfunktionen hin (Church & Schacter, 1994; Schacter & Church, 1992; Palmeri, Goldinger & Pisoni, 1993; Goldinger, Pisoni & Logan, 1991; Goldinger, 1997; Goldinger, Kleider & Shelley, 1999). Im Lichte dieser Untersuchungen erscheint eine Transformation der Signalinformation auf eine abstrakte phonetische Ebene unter Vernachlässigung aller sprecherrelevanten Merkmale als unwahrscheinlich. Es scheint vielmehr ein

[2]Sie werden auch als symbolistischer oder analytischer Ansatz bezeichnet

[3]Verschiedene Vertreter dieses theoretischen Ansatzes nahmen als Zeitpunkt der Transformation unterschiedliche Stufen im Prozeß der Sprachwahrnehmung an und hielten unterschiedliche Anteile im Sprachsignal für wichtig. Die als extrinsischer Ansatz bekannt gewordene Richtung geht davon aus, daß für die Normalisierung ein hinreichend langes Sprachsignal vorhanden sein muß. Dagegen nehmen die Vertreter der intrinsischen Normalisierung an, daß jedes Sprachsignal selbst-normalisierend wirkt, daß es also einen Perzeptionsalgorithmus gibt, der das Verstehen ermöglicht, ohne daß er sich zuvor an den Sprecher anpassen muß. Die intrinsische Richtung unterteilt sich zudem in diejenigen die statische Anteile im Sprachsignal für entscheidend halten (d.h. vor allem die Vokale und das F0 Level) und diejenigen, die gerade dynamische Anteile wie Vokaltransitionen und Diphtonge als bedeutsam ansehen (siehe Haley & Ohde, 1996; Nygaard, Sommers & Pisoni, 1994; Green, Tomiak & Kuhl, 1997)

wechselseitiger Einfluß der Verarbeitungsprozesse von Sprecherinformation und Sprache vorzuliegen. Oben wurde bereits erwähnt, daß ein Sprecherwechsel die Wortverarbeitung unter ungünstigen Bedingungen zusätzlich erschwert. Eine bessere Wortverarbeitung wurde hingegen für bekannte, d.h. in einer Trainingsphase kennengelernte Stimmen beobachtetet (Nygaard et al., 1994). Im Gegenzug findet eine erfolgreichere Stimmerkennung dann statt, wenn die Äußerungen in der Muttersprache des Hörers präsentiert wurden, als wenn der Hörer die präsentierte Sprache nicht beherrschte (Goggin, Thompson, Sturbe & Simental, 1991).

Zeigen diese Daten bereits, daß zwischen der Sprachverarbeitung, insbesondere der phonetischen Analyse, und der Stimmenperzeption ein enger Zusammenhang besteht, so lassen sie jedoch die Frage offen, zu welchem Zeitpunkt im Sprachverarbeitungprozeß die Stimmverarbeitung stattfindet, d.h. ob sie sich *vor* der phonetischen Verarbeitung oder *parallel* zu dieser vollzieht. Zu diesem Aspekt wurden Reaktionszeitexperimente durchgeführt, bei denen beide Parameter direkt als abhängige bzw. unabhängige Variable verwendet wurden. Das klassische Experimentaldesign dieser Studien ist das Speeded-Classification Paradigma von Garner (1974). Hierbei variieren die Stimuli in zwei möglichen Dimensionen, wobei die Versuchspersonen eine Stimulusdimension zu beurteilen, und die andere zu ignorieren haben. Sie bekommen dabei jeweils ein Stimulus-Paar präsentiert und sollen entscheiden, ob die Reize identisch waren oder nicht (*same-different judgement*). Für nicht-unabhängige Prozesse werden dabei folgende Annahmen getroffen: Variiert die nicht-beachtete (nb) Dimension willkürlich, so ergibt sich im Vergleich zu den Reaktionszeiten der Kontrollbedingung (d.h. bei konstanter nb-Dimension) eine Verzögerung für die zu treffende Entscheidung. Variieren die Dimensionen aber in festen Paarungen so ergibt sich ein Redundanzgewinn, d.h. kürzere Reaktionszeiten mit einer geringeren Varianz. Variiert die nb-Dimension willkürlich, so tritt eine Interferenz auf, die einen Anstieg der Reaktionszeiten bewirkt. Für unabhängige Prozesse sind diese Effekte nicht zu erwarten. Diese Speeded-Classification Paradigma wurde auf die Verarbeitung von sprachlicher bzw. sprecherspezifischer Information zu übertragen. Bei einer

Beurteilung von Wortpaaren (selbes Wort/anderes Wort; selber Sprecher/anderer Sprecher) war ein robuster Redundanzeffekt, so wie eine Inteferenz in den Reaktionszeiten zu beobachten (Wood, 1974; Mullenix & Pisoni, 1990). Dabei war stets ein stärkerer Einfluß der Stimminformation auf die Wortverarbeitung beobachten als umgekehrt, was zu der Schlußfolgerung führte, daß es sich zwar um parallele und interagierende Prozesse handelt, daß die Sprecherverarbeitung aber der phonetischen Dekodierung zeitlich vorausgeht: *"In layman's terms, phonetic processes "wait" for at least partial voice processing to occur before they proceed."* (Mullenix, 1997, S. 71). Diese Schlußfolgerung ist jedoch nicht zwingend, denn ein Problem stellt für diese Experimente die Vergleichbarkeit der Bedingungen dar. So ist denkbar, daß die phonetische Unterscheidung aufgrund der akustischen Salienz grundsätzlich schwieriger ist, als die Stimmenunterscheidung. Ein Weg zur Überprüfung dieser Salienz-Hypothese liegt in der Betrachtung der Kontrollbedingungen, d.h. der Detektion einer eindimensionalen Abweichung von Wort oder Sprecher (vgl. Mullenix & Pisoni, 1990). Sind die Dimensionen grundsätzlich gleich salient und gleich schwierig zu diskriminieren, so sollte sich zwischen diesen Bedingungen kein signifikanter Reaktionszeitunterschied zeigen.

Bei Wood (1974) liegt jedoch in der Kontroll-Bedingung ein signifikanter Unterschied vor: Bei der Wahrnehmung von CV-Silbenpaaren wurde ein Wechsel der Grundfrequenz signifikant schneller detektiert, als ein Wechsel der konsonantischen Artikulationsstelle. In der Studie von Mullenix & Pisoni (1990) ist dieser Effekt in der Kontroll-Bedingung zwar nicht signifikant, aber es zeigt sich ein Trend in Abhängigkeit der verwendeten Experimentalreize: Bei einem 2 Sprecher $\times$ 2 Wörter-Design, wurde der Sprecherwechsel etwa 30 ms früher detektiert als der Wortwechsel, der in einem Stimmhaftigkeitskontrast bestand (*bad* vs. *pad*). Das Experiment wurde wiederholt, wobei 16 Sprechern und 16 Wörtern präsentiert wurden Die Wortunterscheidungen in dieser zweiten Bedingung umfaßte neben der Stimmhaftigkeit nun auch z.T. Vokalkontraste (z.B. *back* vs. *bit*). Hier betrug die Differenz der Reaktionsgeschwindigkeiten für die Wortbedingung nur noch 9 ms. Ein Sprecherwechsel hingegen war immer ein Wechsel zwischen männlicher und weiblicher Stimme. Würden auch Sprecher gleichen Geschlechts mit-

einander verglichen, sowie phonetische Kontraste einer größeren Salienz (z.B. verschiedene Vokale und Plosivlaute vs. Frikative am Wortbeginn) so steht zu vermuten, daß die Asymmetrie, die eine frühere Verarbeitung der Stimminformation und eine davon abhängige Wortverarbeitung suggeriert, nicht mehr auftritt.

Zusammenfassend kennzeichnete Mullenix (1997) die Verarbeitung von Stimminformation als obligatorischen, da sie mit Verarbeitungskosten einhergeht, jedoch als nicht-automatischen Prozeß. Bei der Verarbeitung von Sprechstimme und phonetischen Parametern handelt es sich offenbar um schnelle Prozesse der basalen Sprachwahrnehmung, die möglicherweise miteinander interagieren. Behaviorale Experimente haben hier interessante Daten hervorgebracht. Gerade in Bezug auf schnelle und unbewußte Wahrnehmungsprozesse kann es zudem aufschlußreich sein, neuronale Korrelate der Perzeption zu betrachten, d.h. zum einen mittels neurophysiologischer Verfahren Prozesse der Stimmverarbeitung zu untersuchen, und zum anderen klinische Daten heranzuziehen, die zeigen, daß durch einen Schädigung umschriebener Hirnareale Komponenten im auditorischen Verarbeitungsprozeß isoliert werden können.

# Kapitel 7

# Die neuronalen Grundlagen der Stimmverarbeitung

Schallreize, die isoliert dem linken oder dem rechten Ohr dargeboten werden, werden über die Hörbahn vor allem in den kontralateralen auditorischen Kortex geleitet. Da man von funktionalen Unterschieden der beiden Hemisphären ausgeht, sollte sich bei gleichzeitiger Präsentation zweier Reize (z.B. zweier unterschiedlicher Silben) derjenige durchsetzen, der von der funktional dominanten Hemisphäre verarbeitet wird. In den ersten Untersuchungen dieser Art wurden verschiedene sprachliche Stimuli gleichzeitig über Kopfhörer präsentiert und sollten vom Hörer identifiziert werden. Es zeigte sich ein Vorteil für diejenigen Reize, die auf dem rechten Ohr dargeboten wurden (*right ear advantage, REA*), was auf eine funktionale Überlegenheit der linken Hemisphäre für die Verarbeitung von Sprachreizen deutet (Kimura, 1967).[1]

Solche dichotischen Hörexperimente lieferten im Hinblick auf die Lateralisation der Stimmverarbeitung jedoch sehr unterschiedliche Ergebnisse. Für die Verarbeitung von Stimmen berichten beispielsweise Doehring & Batholomeus (1971)

[1] Der REA wurde jedoch kritisiert, da unklar ist, ob er tatsächlich die Dominanz der linken Hemisphäre für sprachliche Reize widerspiegelt, oder vielmehr eine ganze Reihe vor allem aufmerksamkeitsrelatierter Prozesse Für einen direkten Vergleich des REA und der aufgabenunabhängigen Mismatch Negativierung bei der Silbenwahrnehmung Shtyrov, Kujala, Lyytinen, Kujala, Ilmoniemi & Näätänen (siehe 2000a).

einen marginalen Vorteil für die linke Hemisphäre, Riley & Sackheim (1982) einen Vorteil der rechten Hemisphäre, Tartter (1984) keinen Hemisphärenunterschied und Landis, Buttet, Assal & Graves (1982) einen Vorteil für die linke Hemisphäre bei weiblichen Stimmen und einen Vorteil für die rechte Hemisphäre bei männlichen Stimmen. Kreiman & Van Lancker (1988) fanden in ihrem dichtotischen Hörexeriment für die Verarbeitung bekannter Stimmen (von Schauspielern und Politikern) einen klaren Vorteil der rechten Hemisphäre. Sie erklären die uneinheitlichen Ergebnisse in den anderen Studien damit, daß dort unbekannte Stimmen verwendet wurden, in deren Verarbeitung beide Hemisphären involviert seien. Diese Hypothese stützen sie auf ihre Untersuchungen von Patienten mit fokalen Hirnläsionen. So wurde eine Phonagnosie, d.h. ein Defizit bei der Identifikation bekannter Stimmen, nach rechtsparietaler Läsion beobachtet,[2] während die Unterscheidung unbekannter Stimmen nach unilateraler Schädigung des linken oder rechten Temporallappens beeinträchtigt war (Van Lancker & Kreiman, 1987; Van Lancker et al., 1988, 1989). Die Autoren vermuten, daß die Wahrnehmung bekannter und unbekannter Stimmen möglicherweise in qualitativ unterschiedlicher Weise vonstatten geht. Während bekannte Stimmen eher "holistisch" mit bestehenden Musterrepräsentationen (*Templates*) verglichen werden, vollziehe sich die Diskrimination unbekannter Stimmen als eine Extraktion und ein Vergleich von einzelnen Stimmerkmalen.[3] Unterstützt werden diese Annahmen durch die Patientenstudie von Ziegler & Kraus (1999). Sie berichten eine eingeschränkte Performanz bei der Diskrimination verschiedener Sprecher bei einer Schädigung entweder der rechten oder der linken Hemisphäre. Zusätzlich wurde dort die Wortverarbeitung der Patienten getestet. Bei linkshemisphärischer Läsion war neben der Einschränkung der Sprecherunterscheidung eine Beeinträchti-

[2]Diese Beobachtung wird durch experimentelle Ergebnisse unterstützt. In einer PET-Studie, in der gesunde Probanden Stimmen aus der vorangegangenen Trainingsphase wiedererkennen sollten wurde neben einer Aktivität im superioren frontalen Gyrus, den temporalen Polen und dem Gyrus parahippocampalis, auch Aktivierungen im okzipito-parietalen Übergangsbereich und vor allem im inferioren parietalen Gyrus der rechten Hemisphäre gefunden (Imaizumi, Mori, Kiritani, Kawashima & Sugiura, 1997).

[3]Diese eher prozeß-basierte Unterscheidung der Hemisphären ist auch heute noch weitgehend akzeptiert. Es wird allerdings versucht, sie als allgemeine, d.h. modalitätsunabhängige Verarbeitungsstrategien nachzuweisen, sie auf eine neurophysiologische Basis zu stellen und insbesondere zu untersuchen, wie beide als komplementäres System wirken und ihre Prozesse integrieren (Banich & Heller, 1998).

gung der Wortdiskrimination zu beobachten, sowie eine wechselseitige Interferenz: Defizite bei der Wortdiskrimination waren größer, wenn die Wörter von unterschiedlichen Sprechern stammten. Umgekehrt war die Sprecherunterscheidung erschwert, wenn die Sprecher unterschiedliche Wörter äußerten. Klinische Daten weisen also darauf hin, daß die phonetische Sprachverarbeitung und die Wahrnehmung von Sprecherinformation zumindest teilweise dieselben neuronalen Netzwerke rekrutieren.

Die Annahme einer bilateralen Verarbeitung unbekannter Stimmen, erfährt weitere Unterstützung durch die funktionell bildgebenden Verfahren. In einer Studie von Belin und Kollegen wurden verschiedene Stimmsignale (Einzelworte, Sätze, Lachen, Seufzen usw.) und als Kontrollstimuli "Geräusche" (non-vocal sounds; zum Beispiel Naturgeräusche, Tierrufe, mechanische Geräusche) präsentiert. Die Hirnreaktionen wurden mit Hilfe der funktionalen Magnetresonanztomographie (fMRT) erfaßt. Es trat bei der Verarbeitung der Stimmsignale eine bilaterale Aktivierung entlang des superioren temporalen Sulcus (STS) auf, wobei die stimmsensitiven Areale stärker in der rechten Hemisphäre aktiviert wurden und zwar vor allem in drei Regionen: a) nahe des temporalen Pols, b) nahe des Heschl'schen Gyrus und c) im posterioren Anteil des STS. Diese Aktivierungen erwiesen sich als unabhängig von der sprachlichen Struktur der präsentierten Signale. In einem Folgeexperiment blieb das Aktivierungsmuster gleichermaßen erhalten, wenn nur sprachliche oder nur nichtsprachliche Stimmsignale gegen Kontrollstimuli kontrastiert wurden. Daher bezeichnen die Autoren diese Areale entlang des STS als "voice selective" (Belin, Zatorre, Lafalle, Ahad & Pike, 2000; Belin, Zatorre & Ahad, 2002). Einen expliziten Vergleich der Aktivierungen bei sprachlichen Stimuli und nichtsprachlichen Stimmsignalen zeigten Binder und Kollegen (Binder et al., 2000) in einem fMRT-Experiment. Sie verwendeten in ihrer Studie frequenzmodulierte Töne im Vergleich zu drei verschiedenen Arten von Sprachreizen: a) Wörter, b) phonologisch legale Pseudowörter und c) rückwärts abgespielte Einzelworte (*reversed speech*). Die direkten Vergleiche der neuronalen Reaktion auf die einzelnen Stimmsignale mit der Reaktion auf die Töne zeigte in allen drei Fällen eine stärkere Aktivierung für das Stimmsignal bilateral entlang des STS

und des mittleren lateralen STG (superiorer temporaler Gyrus). Die Aktivierungsmuster der drei einzelnen Kontraste unterschieden sich jedoch nicht voneinander. Daraus folgern Binder und Kollegen, daß die bilaterale Aktivierung vor allem auf akustische Merkmale des Inputs zurückzuführen ist, während feinere Unterteilungen aufgrund ihres Experimentaldesigns verlorengingen.[4]

Sowohl Belin et al. (2000, 2002) als auch Binder et al. (2000) heben also die Rolle des STS/STG bei der Verarbeitung von Stimminformation hervor. Sie verweisen dabei auch auf entsprechende Einzelzellableitungen im Tierexperiment. So wurden im auditorischen Kortex von Rhesusaffen z.B. Neurone nachgewiesen, die auf artspezifischer Rufe ansprechen (Tian et al., 2001; Eggermont, 2001, für eine Übersicht ähnlicher Studien). Aufgrund derartiger Befunde werden zwei Verarbeitungspfade von den primären Wahrnehmungsarealen zum präfrontalen Kortex postuliert. Einer der Pfade sei für diese eher "inhaltliche" Information des akustischen Signals zuständig ("WHAT-Pfad"), der andere für die räumliche Lokalisation der Lautquelle ("WHERE-Pfad"). Dieses Konzept scheint nicht zuletzt deswegen plausibel, weil auch für die visuelle Wahrnehmung bei Menschen und Primaten ein dorsaler WHERE-Pfad für die räumliche Zuordnung und ein ventraler WHAT-Pfad für die Objekterkennung beschrieben wurde (Mishkin, 1979; Ungerleider & Mishkin, 1982; Wilson et al., 1993).[5] Unter Verwendung bildgebender Verfahren (PET, fMRT) wurden verschiedene Vorschläge für derartige Verarbeitungsströme bei der auditorischen Wahrnehmung gemacht (Binder et al., 1996, 2000; Belin & Zatorre, 2000; Scott et al., 2000; Wise et al., 2001). Berichten diese Studien auch häufig eine bilaterale Verarbeitung des Lautsprachsignals, ist allerdings nicht ausgeschlossen, daß die beiden Hemisphären einen unterschiedlichen Bei-

[4]Widersprüche zu anderen Studien, wie Démonet et al. (1992) und Binder et al. (1997), die Aktivierungsunterschiede für sprachliche und nichtsprachliche Experimentalreize berichten, werden als aufgabenbedingt interpretiert. In diesen Untersuchungen sollte eine explizite Beurteilung jedes Stimulus gegeben werden, während in der geschilderten Studie die Probanden nur auf den Blockstart und das Ende zu reagieren hatten. Eine von Binder und Kollegen durchgeführte Reanalyse der Daten mit einem laxeren Schwellenwert der voxelbasierten Signifikanz und ohne Begrenzung der Clustergröße lieferte dann auch Ergebnisse, die mit den anderen Studien weitgehend übereinstimmten (siehe Binder et al., 2000).

[5]Eine Erweiterung um einen "HOW"-Pfad, der insbesondere bei Bewegungs-/Handlungsbeobachtungen eine Rolle spielt und hier die Position des Zuschauers als Bezugssystem mit einbezieht, erfuhr dieser Ansatz durch Studien von Milner & Goodale (1995); Creem & Proffit (2001). Zur auditorischen Verarbeitung siehe auch Belin & Zatorre (2000).

trag dazu leisten (Hickok & Poeppel, 2000). Die frequenzabhängige Verarbeitung auditorischen Inputs wurde von Ivry und Robertson in ihrer DFF Theorie (*double filtering by frequency*) zusammengefaßt (Ivry & Robertson, 1998). Nach Ansicht der Autoren ist allein die spektrale Analyse Grundlage der Verarbeitung. Nachdem eine Art "Frequenzzentrum" bestimmt ist, das für die Analyse wichtig ist, werden die höheren Frequenzen vornehmlich von der linken, die tieferen von der rechten Hemisphäre verarbeitet. Andere Wissenschaftler machen hingegen eher die Verarbeitung zeitlicher Strukturen für die Hemisphärenasymmetrien verantwortlich (Belin, Zilbovicius, Crozier, Thivard & Fontaine, 1998; Tallal, Miller & Fitch, 1993). In ihren eigenen und in den von ihnen zitierten Studien zeigte die linke Hemisphäre einen deutlichen Vorteil bei der Verarbeitung schneller spektraler Änderungen, wie z.B. Formanttransitionen, die rechte Hemisphäre hingegen für langsame, wie z.B. der Sprechmelodie. Zatorre hingegen schreibt der linken Hemisphäre eine bessere Auflösung zeitlicher Muster zu, was zu einer schlechteren Frequenzauflösung führe, während die rechte Hemisphäre besser im Bereich der spektralen Analyse arbeite bzw. für die langsameren Veränderungen der Grundfrequenzkontur zuständig sei (Zatorre, 1997). Weitere theoretische Ansätze, die ähnlich argumentieren, sind in Hickok & Poeppel (2000) aufgeführt. Zur Absicherung dieser Konzepte ist jedoch noch weitere Forschung notwendig, die idealerweise auch durch Verfahren höherer zeitlicher Auflösung ergänzt werden sollte. Ein Beispiel für die frühe auditorische Verarbeitung lautsprachlicher Signale stellen die Untersuchungen von Hickok und Poeppel dar, die aufgrund von MEG-Studien einen bilateralen Verarbeitungspfad im posterioren Temporallappen für die frühe lautbasierte Repäsentation des akustischen Inputs postulieren (Hickok & Poeppel, 2000). Dieser interagiert nach Ansicht der Autoren mit inferior frontalen und inferior parietalen Hirnregionen, die für die Artikulationsmotorik zuständig sind bzw. eine Verbindung zwischen Sprachproduktion und Sprachwahrnehmung darstellen, und mit einem auditorisch-konzeptuellen Pfad, der im okzipito-temporo-parietalen Übergangsbereich angesiedelt wird (Poeppel & Rowley, 1996; Hickok & Poeppel, 2000).

Die Stimmenverarbeitung spielt in diesem Modell bisher keine Rolle und tatsächlich

befassen sich nur sehr wenige EKP- und MEG-Studien mit der Verarbeitung von Stimminformation. Der Unterschied zwischen der Verarbeitung bekannter und unbekannter Stimmen im EKP wurde beispielsweise erst kürzlich untersucht (Schweinberger, 2001). Beim aktiven Wiedererkennen der Stimmen nach einer Primingphase ergab sich für unbekannten Stimmen eine anhaltende zentro-parietale Negativierung (450 bis mindestens 800 ms nach Stimulusbeginn), für die bekannten Stimmen (von Schauspielern, Politikern usw.) hingegen eine stärkere temporale Negativierung zu beobachten, die mit einer aufgabenbedingten Gedächtnisaktivierung in Verbindung gebracht wurde. [6] Außerdem zeigten die unbekannten, nicht aber die bekannten Stimmen einen Priming-Effekt, d.h. eine anhaltende Negativierung ab 800 ms für die ungeprimten Experimentalreize. Senkfor & Van Petten (1998) fanden hingegen bei einer Worterkennungsaufgabe keinen solchen Priming-Effekt, der auf eine automatische Aktivierung der Stimminformation im episodischen Gedächtnis hinweist. EKPs auf Wörter, die dort in der Testsphase mit derselben Stimme wie in der vorangegangenen Trainingsphase präsentiert wurden, unterschieden sich nicht von den EKPs auf Wörter die mit einer anderen Stimme dargeboten wurden. Nur wenn die Aufgabe darin bestand, zu überprüfen, ob auch die Sprechstimme dieselbe wie in der Testphase war, trat für die bekannten Experimentalreize eine frontale Positivierung ab 700 ms nach Stimulusbeginn auf. Diese Studie zeigt, wie eine Manipulation der Experimentalaufgabe die EKP-Korrelate bei der Stimmwahrnehmung beeinflußt, denn ein elektrophysiologisches Korrelat war nur dann zu beobachten, wenn die Stimminformation im Fokus der Experimentalaufgabe stand.

Mit der aufgabenunabhängigen, präattentiven Verarbeitung eines Sprecherwechsels befaßte sich die Untersuchung von Titova & Näätänen (2001), die auf einem Oddball-Paradigma basierte. In dieser Studie wurden Äußerungen des Vokals

[6] Siehe hierzu auch die fMRT-Studie von Stevens & Schwartzreich (2000). Dort wurde bei einem Vergleich von Stimm-, bzw. Wortinformation im Arbeitsgedächtnis eine Aktivierung des STG gefunden, die bei Wortinformation eher linkslateralisiert und bei Stimminformation eher rechtlateralisiert war. Zudem war bei dem Vergleich von Stimmen und von Tönen eine Aktivierung im inferioren frontalen Gyrus beobachtet, die ein typisches Korrelat derartiger Verarbeitungsprozesse im Arbeitsgedächtnis darstellt (siehe z.B. Owen et al., 1999) Daß dies bei einem Vergleich der Wortinformation nicht auftrat, deutet auf eine qualitativ andere, vermutlich abstrakte Verarbeitung linguistischer Information im Arbeitsgedächtnis hin.

/e/ von verschiedenen Sprechern präsentiert. Als Standardstimulus diente dabei eine männliche Stimme und als Devianten drei weibliche Stimmen. Die Autoren zeigten, daß ein Sprecherwechsel eine präattentive Mismatch Negativierung evoziert. Keinen Unterschied in den ereigniskorrelierten N1-P2 Mustern fand sich dagegen in der Studie von Tiitinen et al. (1999). Dort wurde das periodische Quellsignal des Stimulus /a/ grundsätzlich beibehalten, aber die Regelmäßigkeit der periodischen Schwingung variiert, so daß der Eindruck einer "weichen" bzw. einer "gepreßten" Stimme entstand. Eine solche Variation des Stimmreizes ist offenbar zu schwach, um ein meßbares elektrophysiologisches Korrelat zu evozieren. Alku und Kollegen (2001) ersetzten in ihren Stimuli hingegen das gesamte Glottissignal durch Rauschen unter Beibehaltung der Filterparameter und fanden im Vergleich zu nicht-manipulierten Vokalen eine Amplitudenreduktion der N100m Komponente. Sie weisen auf die Bedeutung der Stimminformation bei der Verarbeitung lautsprachlicher Signale hin und schließen aus ihren Ergebnissen, *"that the characterization of vowels* [...] *ignoring the role of the glottal excitation is an oversimplication understanding speech perception."* (Alku et al. 2001, S.28). Nicht nur für die Sprechererkennung, sondern auch für das effektive Verarbeiten von Lautsprache scheint die Stimminformation also eine bedeutende Rolle zu spielen.

Die in diesem Abschnitt gegebene Übersicht zeigt, daß sich die menschliche Stimme durch einzigartige Parameter auszeichnet, die sie von anderen akustischen Signalen unterscheidet und daß diese Besonderheit in ihrer neuronalen Verarbeitung reflektiert ist. Dieser Sensitivität des auditorischen Verarbeitungssystems für die menschliche Stimme ist der erste Abschnitt des folgenden, empirischen Teils der Arbeit gewidmet. Die nachfolgenden zwei Experimentalserien untersuchen dann die Bedeutung der extralinguistischen Stimminformation für das Verstehen des sprachlichen Inhalts: Experiment II befaßt sich mit dem Zusammenhang von Stimmverarbeitung und Sprachlautverstehen. Experiment III überprüft, ob die Stimminformation auf die semantische Verarbeitung einen Einfluß hat.

# Teil III

# Empirischer Teil

# Kapitel 8

# EXPERIMENTALSERIE I: Prototypikalitätseffekte bei der Stimmwahrnehmung

Die ersten beiden MEG-Experimente dieser Arbeit hatten zum Ziel, die präattentive Reaktion auf verschiedene Stimmreize zu untersuchen. Mit einem Oddball-Paradigma wurde insbesondere die Rolle der Prototypikalität bei der Stimmwahrnehmung untersucht.

## 8.1 Experiment Ia

### 8.1.1 Fragestellung

Wie im vorhergehenden Kapitel beschrieben, deuten Studien darauf hin, daß das menschliche Wahrnehmungssystem die einzigartigen Frequenzmuster der menschlichen Stimme im Vergleich zu Tönen oder Rauschen auf besondere Weise verarbeitet (Belin et al., 2000; Binder et al., 2000; Alku et al., 2001). Des-

weiteren wurde elektrophysiologische Evidenz dafür präsentiert, daß ein Sprecherwechsel präattentiv detektiert werden kann: Titova & Näätänen (2001) berichten eine Mismatch Negativierung für den Wechsel einer männlichen Stimme (Standardstimulus) zu verschiedenen weiblichen (Devianten). Ein solcher Sprecherwechsel beinhaltete in diesem Falle einen deutlichen Anstieg in der mittleren Grundfrequenz, von mindestens 80 Hz. Die Amplitude der MMN korrelierte dabei mit der perzeptuellen und akustischen Ähnlichkeit (Grundfrequenz und erste zwei Formanten) zwischen Standard und jeweiligem Devianten. Eine zunehmende Amplitude mit zunehmender akustisch-perzeptueller Devianz wird jedoch auch auch für Frequenzabweichungen von Sinustönen berichtet (z.B. Sams et al., 1985). So ist unklar, ob die von Titova und Näätänen beobachteten Prozesse eine Spezifität der Stimmwahrnehmung widerspiegeln, oder basale, unspezifische Vorgänge der auditorischen Verarbeitung.

Ebenfalls mittels des Oddball-Paradigmas konnte gezeigt werden, daß Phoneme als Musterrepräsentation bzw. prototypische Gedächtnisspur zentral repräsentiert sind (Näätänen et al., 1997). So wurde berichtet, daß fremdsprachliche Phoneme (Näätänen et al., 1997; Winkler et al., 1999) oder weniger prototypische Exemplare eines Phonems (Aaltonen et al., 1987) eine Mismatch Negativierung hervorrufen, die schwächer ist als die MMN auf prototypische Sprachlaute[1]. Ebenso riefen identische akustische Abweichungen im Oddballparadigma unterschiedliche Mismatch Reaktionen hervor, je nachdem, ob sie perzeptuell zu einem Wechsel der Phonemkategorie führten, oder nicht (Phillips et al., 2000; Kasai et al., 2001). In diesen Fällen ist der Mismatch-Detektor also unabhängig von der akustischen Stimulusbeschaffenheit, aber abhängig von der zentralen Repräsentation der Sprachlaute im Gehirn.

Das folgende Experiment sollte untersuchen, ob eine solche Repräsentation auch bei der Verarbeitung von Stimminformation eine Rolle spielt, d.h. ob auch bei der präattentiven Wahrnehmung von Stimmen Prozesse auftreten, die für die Stimmwahrnehmung *spezifisch* sind. Hierzu wurden Stimmen unterschiedlicher

[1] d.h. Phoneme, die in der Muttersprache vorkommen oder die in einer Sprache vorkommen, die vom Hörer beherrscht wird.

*Prototypikalität* präsentiert. Es wird davon ausgegangen, daß das auditorische System aufgrund seiner Hörerfahrung eine Art Musterrepräsentation (*Template*) typischer männlicher und weiblicher Stimmen besitzt. Unbekannte Stimmen werden dann anhand dieser Stimmtemplates als mehr oder weniger prototypisch eingestuft (vgl. Kreiman, 1997). Um nun die Prototypikalität und die physikalischen Ähnlichkeit zwischen Standard und Deviant experimentell zu dissoziieren, wurde ein Oddball Paradigma mit folgenden Stimuli verwendet: Als Standardreiz wurde eine männlichen Stimme eingesetzt, als Devianten drei weibliche Stimmen, die unabhängig in den beschriebenen Dimensionen variierten: a) im Grad der Prototypikalität (natürliche, F0-manipulierte und synthetisierte Stimme) und b) im Grad der physikalischen Ähnlichkeit zum Standard (dieselbe oder eine andere Grundfrequenz). Um die auditorische Verarbeitungsprozesse an die normale Sprachverarbeitung anzugleichen, wurden Einzelwörter als Stimulusmaterial eingesetzt. Diese wurden einer akustischen Analyse, sowie einem Akzeptabilitäts- und Ähnlichkeitsrating unterzogen.

### 8.1.2 Methode

#### 8.1.2.1 Stimuli

*Akustische Parameter der Stimuli*

Alle Experimentalreize waren Äußerungen des Nomens “Dach“. Die natürlichsprachlichen Stimuli wurden aus einem Datenpool von Einzelwortaufnahmen aufgrund ihrer ähnlichen prosodischen Kontur und ihrer hohen Übereinstimmung der segmentalen Dauer ausgewählt. Der Experimentalreiz M stammte von einem männlichen Sprecher und hatte eine mittlere Grundfrequenz von 80 Hz. Da er bei diesem Experiment die Funktion des Standardstimulus hatte, wird er hier als **M_std** bezeichnet. Der Deviant **W** stammte von einer Sprecherin und hatte eine mittlere Grundfrequenz von 167 Hz. Der zweite Deviant **G** entsprach Stimulus W, jedoch war mit Hilfe des PRAAT Spracheditors[2] die Grundfrequenz auf

[2] http://fonsg3.uva.nl.praat

80 Hz gesenkt und damit dem Standardstimulus angeglichen worden. Diese Manipulation des Quellsignals ist möglich, ohne daß sich große Veränderungen in den Filterparametern, d.h. in der Formantstruktur ergeben. Der dritte Deviant **S** war ein vollkommen synthetisch hergestelltes Sprachsignal. Zur Erstellung dieses Experimentalreizes wurde die MBROLA Diphonsynthese[3] verwendet. Als Vorbild der Synthese wurde Stimulus W verwendet, die mittlere Grundfrequenz lag daher ebenfalls bei 167 Hz. Eine genauere Analyse der Stimuli zeigte aber für Stimulus S eine ungewöhnliche Verteilung der Spektralenergie, die in Abbildung 8.1 näher beschrieben ist. Die Intensitäten der Stimuli wurden mit Hilfe des Cool-Edit Spracheditors normalisiert, die Dauer betrug 397 ms für den Standard und 378 $\pm$ 6 ms für die Devianten.

Wie Tabelle 8.1 zeigt, sind sich die Formanten für die drei weiblichen Stimmen relativ ähnlich, während die männliche Stimme eindeutig andere Formantwerte aufweist. Eine Übersicht der Stimuluseigenschaften ist in Tabelle 8.2 gegeben.

| Stimulus | Sprecher | F0 [Hz] | F1 [Hz] | F2 [Hz] | F3 [Hz] |
|---|---|---|---|---|---|
| M (*"Dach"*) | männl. | 80 | 806 | 1653 | 4616 |
| W (*"Dach"*) | weibl. | 167 | 736 | 1361 | 2897 |
| G (*"Dach"*) | F0 man. | 80 | 718 | 1328 | 2777 |
| S (*"Dach"*) | synth. | 167 | 856 | 1308 | 2824 |

Tabelle 8.1: Übersicht der Stimuluseigenschaften II: mittlere Grudfrequenz, und Formant 1 - Formant 3.

*Stimulus-Rating*

Zwei Grupppen phonetisch naiver Hörer, die nicht an den späteren MEG-Experimenten teilnahmen, wurden gebeten, die Stimmen zu beurteilen. In einer Familiarisierungsphase wurden die Stimuli je zweimal vorgespielt mit einem Interstimulusintervall von 3000 ms. Bei der dritten Präsentation sollte Hörergruppe 1 (N=15) die Natürlichkeit der Sprachsignale auf eine Fünf-Punkte-Skala beurteilt

[3] http://tcts.fpms.ac.be/synthesis; deutsche Datenbank, weibliche Stimmme

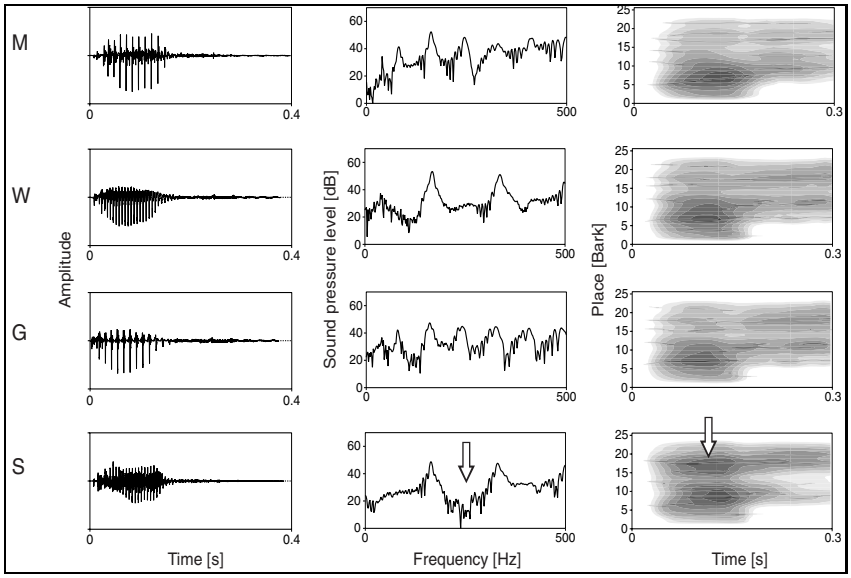



Abbildung 8.1: Experiment I: Physikalische Eigenschaften der Stimuli. Oszillogramm (links), Spektrum (mitte) und Cochleagramm (rechts) der vier Experimentalreize. Die Stimuli besaßen alle in etwa dieselbe Dauer, Bedingung M und G zeichnen sich durch eine niedrigere Grundfrequenz aus als die anderen beiden Experimentalreize und sind sich daher physikalisch ähnlich. Die Pfeile indizieren die spektralen Besonderheiten des synthetisierten Stimulus. Zum einen zeigt das Spektrum ein Frequenzdefizit im Spektralbereich um 300 Hz; zum anderen wird im Cochleagram, das die Frequenz-Ort-Abbildung des akustischen Inputs in der Cochlea darstellt, ein zusätzliches Energieband um 16-18 Bark sichtbar.

werden (1 = sehr natürlich, 5 = sehr künstlich). Da der Begriff der "Natürlichkeit" jedoch nur vage definiert ist und eine unbestimmte Eigenschaft sprachlicher Signale darstellt (Sonntag, 1999; Nusbaum, Francis & Henly, 1995), wurden die Hörer gebeten, im vierten Präsentationsdurchgang den "Wohlklang" der Stimmen auf einer Fünf-Punkte-Skala zu beurteilen (1 = sehr angenehm, 5 = sehr unangenehm).

Die Ergebnisse des Ratings sind in Abbildung 8.2 dargestellt. Das Natürlichkeitsrating ergab einen signifikanten Unterschied in der Beurteilung der menschlichen Stimmen (M, W), die als sehr natürlich eingeschätzt wurden ($<$ 2 Punkte) und der Stimmen, die computermanipulierten bzw. synthetisierten Sprachsignale. Der Unterschied zwischen diesen beiden Clustern war signifikant (t-test:

| Stimulus | Ähnlichkeit | Prototypikalität |
|---|---|---|
| M | (+) | + |
| W | - | + |
| G | + | - |
| S | - | - |

Tabelle 8.2: Stimuluseigenschaften: a) physikalische Ähnlichkeit zum Stimulus M (Standardreiz Experiment Ia). Sowie die Prototypikalität/für eine menschliche Stimme.

$t_{(14)} = 9.37, p < 0.0001$). Es zeigten sich jedoch innerhalb der Gruppen keine signifikanten Unterschiede, weder für M und W ($t_{(14)} = 1.29, p = 0.22$) noch für G und S (t $<$ 1). In der Beurteilung des Wohlklangs wurden die Stimuli M und W als angenehm empfunden ($<$ 2 Punkte; keine signifikanten Unterschiede). Für den grundfrequenzmanipulierten Reiz G resultierte ein mittlerer Wert von 3.9 Punkten, der sich signifikant von den natürlichen Stimmen unterschied ($t_{(14)} = 9.32, p <$ 0.0001). Das synthetische Sprachsignal S wurde als unangenehmstes Sprachsignal beurteilt (4.5 Punkte) und unterschied sich damit signifikant von Stimulus G ($t_{(14)} = 1.29, p < 0.05$). Es läßt sich also festhalten, daß im Rating die nicht-menschlichen Sprachsignale G und S deutlich von den menschlichen Stimmen M und W unterschieden werden. Dabei zeigt sich eine Divergenz in der Beurteilung von Natürlichkeit und Wohlklang: Die Beurteilungen der beiden nicht-natürlichen Experimentalreize G und S unterscheiden sich in Bezug auf die "Natürlichkeit" der Sprachdaten nicht signifikant. Aber hinsichtlich deren Wohlklang trat ein signifikanter Unterschied auf. Hier kommt dem Stimulus G eine Mittelposition zwischen natürlichsprachlichen Äußerungen und der als sehr unnatürlich bewerteten Computerstimme (S) zu.

Hörergruppe 2 (N=15) sollte die Ähnlichkeit der drei weiblichen Stimmen (W, G, S) zur männlichen Stimme M auf einer Fünf-Punkte-Skala beurteilen (1 = sehr ähnlich; 5 = sehr unähnlich). Das grundfrequenzmanipulierte Signal erhielt dabei den niedrigsten Wert (wurde also als am ähnlichsten eingeschätzt), während das synthetische Signal als sehr unähnlich beurteilt wurde. Dieser Unterschied

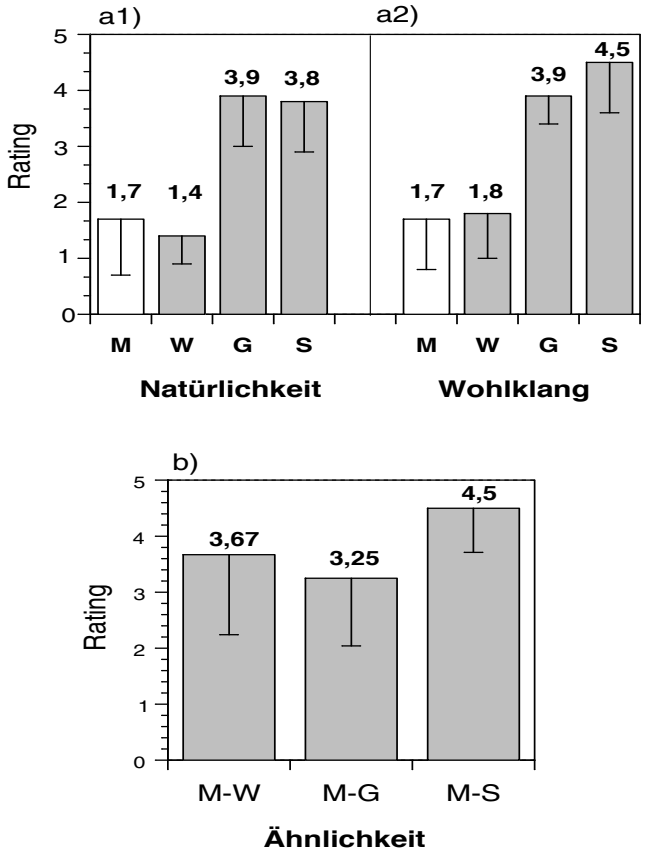



Abbildung 8.2: Stimulusbeurteilung: Rating a1) der Natürlichkeit, a2) des Wohlklangs und b) der Ähnlichkeit zu der männlichen Stimme (M) auf einer Fünf-Punkte-Skala (für die Absolutwerte siehe Appendix A.1).

war signifikant ($F_{(1,14)} = 3.36, p < 0.01$). Die natürliche weibliche Stimme nahm eine Mittelposition ein und unterschied sich von den anderen Bedingungen nicht signifikant (t $<$ 1).

#### 8.1.2.2 Probanden

An der MEG-Studie (Experiment Ia) nahmen 10 Probanden (6 männlich) im Alter von 21 bis 30 Jahren (Durchschnitt: 24 Jahre) teil. Alle waren deutsche Muttersprachler und nach dem Edinburgh Inventar (Oldfield, 1971) als Rechtshänder

klassifiziert. Keiner der Teilnehmer berichtete audiologische Beeinträchtigungen. Die Probanden wurden für ihre Teilnahme am Experiment bezahlt.

#### 8.1.2.3 Versuchsablauf

Der Proband saß in einem Meßsessel, der eine weitgehende Fixierung des Kopfes erlaubt, und wurde gebeten, sich möglichst wenig zu bewegen oder zu blinzeln. In etwa 1,20 m Abstand von den Augen befand sich ein Projektionsschirm (32 cm $\times$ 22 cm). Auf diesem war während der Stimulation ein Film ohne Ton zu sehen. Die Probanden waren instruiert, sich auf den Film zu konzentrieren und den auditorischen Input nicht zu beachten. Vor dem Beginn des Experiments wurde die Hörschwelle des Probanden ermittelt[4] und die Stimuli 40 dB über diesem Wert präsentiert (d.h. bei etwa 60 dBSPL) unter Verwendung der Präsentationssoftware ERTS (Exerimental Runtime System, Version 3.25, Berisoft Corporation 1995) und Ohrschläuchen aus Kunststoff zur Schallübertragung. Die Stimuli wurden pseudorandomisiert im Oddball Paradigma präsentiert, wobei vor jedem Deviant mindestens drei Standards auftraten. Der Onset-to-Onset Intervall betrug 900 ms. Insgesamt wurde jeder Deviant 500 mal präsentiert, und besaß eine relative Häufigkeit von etwa 0.05 (die Anzahl der Standardreize betrug 9214).

#### 8.1.2.4 Datenerhebung

Die Hirnaktivität der Probanden wurde mit einem 148-Kanal Ganzkopf-Magnetometer (MAGNES WHS 2500, 4D-Neuroimaging) aufgezeichnet, wobei die Daten eines Probanden in zwei Sitzungen zu jeweils vier Blöcken an zwei unterschiedlichen Tagen erhoben wurden. Vor der Messung wurde mit Hilfe des Polhemus FastTrack Systems die Position des Sensors in Referenz zum festgelegten Nasion-Ohr Koordinatensystem gemessen. Vor und nach jedem Block wurde

[4] Hierzu wurde das System SET ASG verwendet. Ein gesprochener Satz wurde mehrfach präsentiert. Einmal wurde die und die Lautstärke dabei von 0 dB sukzessive erhöht, und der Proband sollte angeben, ab wann er ein Signal hört. Im zweiten Durchgang wurde der Satz ab etwa 80 dB schrittweise abgedämpft und der Proband sollte angeben, ab wann er nichts mehr hört. Diese Prozedur wurde für jedes Ohr separat durchgeführt und der Mittelwert aus den insgesamt vier Einzeldaten als Hörschwelle gewertet.

diese Position überprüft um Bewegungsartefakte von der weiteren Datenanalyse auschließen zu können. 148 MEG-Kanäle, 11 Referenzkanäle, sowie ein Elektrookulogramm wurden aufgezeichnet. Die Abtastrate betrug 254 Hz. Das MEG-Signal wurde online bandpaßgefiltert (0.1-100 Hz), zur Baselinekorrektur wurde offline eine weitere Filterung von 1.5-20 Hz durchgeführt. Es erfolgte eine Mittelung über die jeweils 700 ms langen Signalepochen (von -200 bis 500 ms relativ zum Stimulusbeginn). Vor der Mittelung wurden die Epochen durch ein automatisiertes Verfahren auf Einflüsse der Augenbewegung und andere Artefakte hin überprüft. Sie wurden von der weiteren Analyse ausgeschlossen, wenn sie Abtastpunkte enthielten, zu deren Zeitpunkt entweder die Standardabweichung des EOG in einem 200 ms Zeitfenster einen Wert von 30 Mikrovolt überstieg, oder die die Standardabweichung eines MEG Kanals innerhalb eines 3 Sekunden Intervalls einen Wert von 1100 fT überstieg bzw. wenn die Änderung innerhalb der Epoche mehr als 3000 fT betrug. Ausserdem wurden MEG-Kanäle von der weiteren Analyse ausgeschlossen, wenn der Median der Korrelationskoeffizienten für die validen Epochen benachbarter Kanäle unter .75 lag. In die Analyse der Standardbedingung gingen nur diejenigen Epochen ein, die direkt vor einer Deviant-Epoche lagen.

#### 8.1.2.5 Quellanalyse

Die Quellanalyse wurde mit Hilfe der Auswertungssoftware CURRY 4.0 (Philips GmbH), durchgeführt. Ein standardisiertes BEM Modell (Ferguson, Zhang & Stroink, 1994), das auf den Talairach-skalierten anatomischen Daten von 50 Personen basierte, wurde als Volumenleiter verwendet. Das MEG Signal wurde separat für jeden Block gemittelt. Nach der Umrechnung auf eine Standard-Sensorposition des jeweiligen Probanden wurden die Werte normalisiert und das Mittel aller Blöcke gebildet. Die Stromdichten wurden unter Verwendung des Minimum-Norm-Least-Square Algorithmus (Wang, Williamson & Kaufman, 1992; Hämälainen & Ilmoniemi, 1994) berechnet. Die Quellstärken wurden als Summe der individuellen Stromdichten über der jeweiligen Hemisphäre ermittelt.

### 8.1.3 Ergebnisse

Die ereigniskorrlierten magnetischen Felder zeigen für die devianten Bedingungen zwischen 100 ms und 300 ms eine deutliche MMNm (siehe Abbildung 8.3). Für zwei Kanäle im Zentrum des evozierten Magnetfeldes der linken und rechten Hemisphäre wurden statistische Analysen (gepaarter t-Test) durchgeführt, die zeigen, daß zwischen Standard und jeweiligem Devianten ein signifikanter Unterschied im Zeitfenster von 100-200 ms vorlag (Kanal 88: Deviant W $t_{(9)} = 2.23, p = 0.052$; Deviant G $t_{(9)} = 3.96, p < 0.01$; Deviant S $t_{(9)} = 5.73, p < 0.001$; Kanal 78 : Deviant W $t_{(9)} = 4.09, p < 0.01$; Deviant G $t_{(9)} = 10.76, p < 0.0001$; Deviant S $t_{(9)} = 9.7, p < 0.0001$ ).

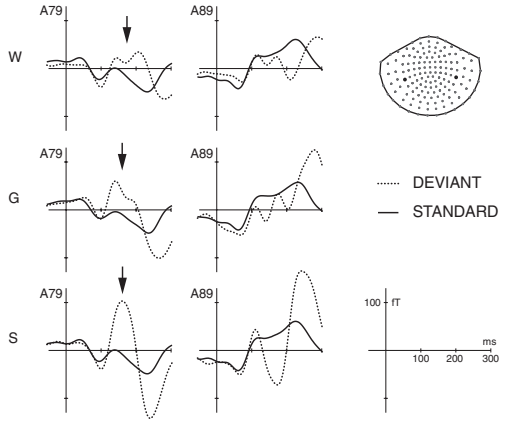



Abbildung 8.3: Gemittelte ereigniskorrlierten Magnetfeldstärken über dem linken (Kanal 78) und rechten (Kanal 88) auditorischen Kortex. Im Vergleich zum Standardstimulus M_std zeigen die Reaktionen auf die Deviants eine erhöhte Mismatch-Aktivierung zwischen 100 und 300 ms nach Stimulusbeginn (für den äquivalenten Kanal der anderen Hemisphäre ergibt sich stets eine Umkehr der Magnetfeldrichtung).

Die Aktivierungsverteilungen der einzelnen Probanden wiesen nur vereinzelt dipolare Muster auf. Daher wurde statt einer Dipolmodellierung die Berechnung der Stromdichten als Maß der Aktivierung durchgeführt. Abbildung 8.4 zeigt die mittleren Stromstärken der linken und rechten Hemisphäre, für den Bereich 100-300 ms. Für die drei Devianten zeigt sich in beiden Hemisphären ein Anstieg der mittleren Stromdichte W $<$ G $<$ S.

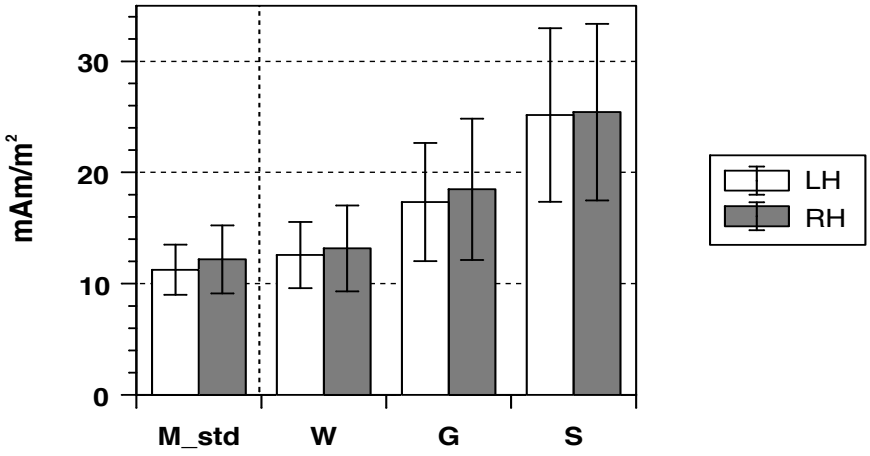



Abbildung 8.4: Mittlere Stromdichten von 100-300 ms nach Stimulusbeginn für die vier Experimentalbedingungen, aufgeteilt nach Hemisphären (li=links; re=rechts).

Die zweifaktorielle Varianzanlyse BEDINGUNG $\times$ HEMISPHÄRE zeigte einen signifikanten Haupteffekt BEDINGUNG ($F_{(3,27)} = 20.06, p < 0.0001$), jedoch keinen Haupteffekt HEMISPHÄRE und keine Interaktion der Faktoren. In Anlehnung an die Rating-Studie wurde die statistische Signifikanz der Unterschiede in Bedingung W und G, sowie G und S überprüft. Der t-Test zeigt jeweils einen signifikante Differenz (W vs. G: $t_{(9)} = 4.15, p < 0.01$ G vs. S: $t_{(9)} = 3.06, p < 0.05$).

Um den Zeitverlauf der Aktivierung zu untersuchen, wurden die Stromdichtewerte jedes Probanden für jeweils 50 ms breite Zeitfenster von 100-300 ms gemittelt. Diese Mittelwerte wurden einer Varianzanlyse der Faktoren DEVIANT (W, G, S), ZEITFENSTER (100-150 ms, 150-200 ms, 200-250 ms, 250-300 ms) und HEMISPHÄRE (links, rechts) unterzogen. Es zeigte sich wiederum kein HEMISPHÄREN-Haupteffekt und ebenfalls keine Interaktion des Faktors HEMISPHÄRE mit einem der anderen Faktoren. Es lag hingegen ein Haupteffekt DEVIANT ($F_{(2,18)} = 19.21, p < 0.0001$) vor, sowie eine signifikante Interaktion DEVIANT $\times$ ZEITFENSTER ($F_{(6,54)} = 6.68, p < 0.0001$). Die Auflösung der Interaktion zeigte für jedes Zeitfenster einen signifikanten Haupteffekt DEVIANT, der jedoch vor allem im letzten Zeitfenster schwächer war. Die Auflösung der Haupteffekte zeigte, daß die Unterschiede zwischen den Bedingungen signifikant waren (siehe Tabel-

le 8.3; für den zeitlichen Verlauf siehe auch Appendix A.1 und A.2).

| Zeitfenster | Bedingungen | *df* | *F* | *p* |
|---|---|---|---|---|
| **100-150 ms** | W G S | 2, 18 | 16.80 | 0.0001 |
| | W G | 9 | 0.17 | 0.69 |
| | G S | 9 | 6.11 | 0.05 |
| **150-200 ms** | W G S | 2, 18 | 17.10 | 0.0001 |
| | W G | 9 | 21.06 | 0.01 |
| | G S | 9 | 3.89 | 0.1 |
| **200-250 ms** | W G S | 2, 18 | 18.55 | 0.0001 |
| | W G | 9 | 15.17 | 0.01 |
| | G S | 9 | 23.16 | 0.001 |
| **250-300 ms** | W G S | 2, 18 | 9.41 | 0.01 |
| | W G | 9 | 13.69 | 0.01 |
| | G S | 9 | 8.4 | 0.05 |

Tabelle 8.3: Signifikanz der DEVIANT-Effekte in den einzelnen Zeitfenstern.

Die räumliche Verteilung der Aktivierung ist in Abbildung 8.5 dargestellt. Sie zeigt ein Hauptaktivierungsgebiet im medialen und superioren Temporallappen.

### 8.1.4 Diskussion

Die Ergebnisse der Studie zeigen, daß das auditorische System präattentiv auf Stimmenunterschiede reagiert. Wie erwartet, wurden durch die devianten Reize Mismatch Negativierungen ausgelöst. Die Hauptquelle dieser Aktivierungen liegt im oder nahe des auditorischen Kortex (Heschl'scher Gyrus), was in anderen MMN-Studien, die eine Stromdichte-Modellierung durchführten, ebenfalls beobachtet wurde (Giard et al., 1990; Rinne et al., 2000). Der Vergleich der Deviant-Bedinungen zeigte einen Anstieg der korrelierten Stromdichten, der mit der Einschätzung des Wohlklangs korrespondierte. Je unangenehmer das Signal im Vortest beurteilt wurde, desto höher war die ereigniskorrelierte Stromdichte.

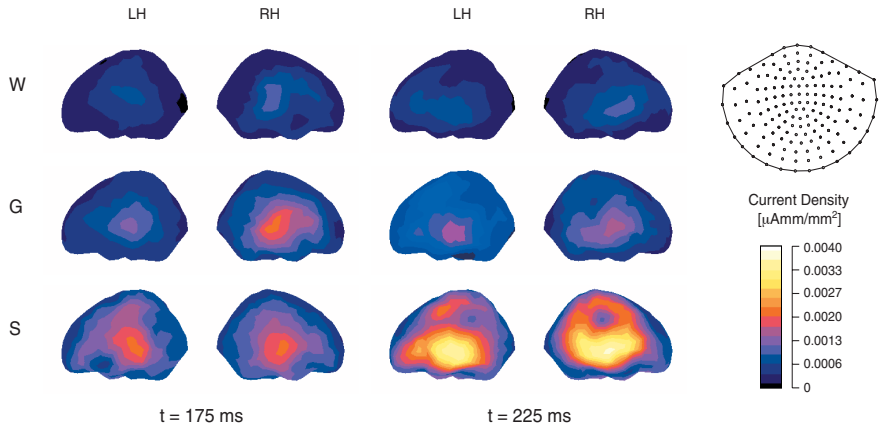



Abbildung 8.5: Verteilung der ereigniskorrelierten Aktivierungen auf die drei Devianten W, P und S: Stromdichten der linken (LH) und rechten (RH) Hemisphäre zu den Zeitpunkten t= 125 ms und t = 175 ms.

Im Ähnlichkeitsrating wurde hingegen nicht signifikant zwischen den Stimmen W und G unterschieden. Beide wurden als dem Standardstimulus gleich ähnlich eingeschätzt. Einige Hörer berichteten spontan, daß die Beurteilung der Stimmenähnlichkeit sehr schwer falle, was möglicherweise auf eine Überschneidung der Parameter Prototypikalität und physikalische Ähnlichkeit zurückzuführen ist. Das synthetische Signal, das akustisch und in Bezug auf die Prototypikalität die größte Abweichung darstellte, wurde eindeutig als dem Stimulus M unähnlich klassifiziert. Die explizite Einschätzung der "Natürlichkeit" zeigte ebenfalls keine graduelle Abstufung. Hier kategorisierten die Hörer a) in Stimmsignale, die von menschlichen Sprechern stammten und dem erwarteten Klang einer weiblichen und männlichen Stimme weitgehend entsprachen und b) in Stimmsignale die von der Norm abwichen. War jedoch der Wohlklang zu beurteilen, so wurde außerdem noch zwischen den beiden nicht-natürlichen Sprachsignalen unterschieden, wo bei das grundfrequenzmanipulierte Signal akzeptabler eingeschätzt wurde als die synthetische Stimme. Die gemessene Stromdichte scheint ein physiologisches Korrelat dieser feineren Differenzierung zu sein, denn auch hier war eine vergleichbare Abstufung zu beobachten. Interessanterweise ist die neurophysiologische Reaktion relativ unabhängig von der physikalischen Ähnlichkeit der Devi-

ants zum Standard. Würde diese Eigenschaft nämlich eine Rolle spielen, so hätte der grundfrequenzveränderte Stimulus (G), der sich ja nur im Stimmtimbre (Filterparameter), nicht aber in der Tonhöhe (Quellsignal) vom Standard unterschied, eine geringere Aktivierung aufweisen müssen, als Deviant W, der in beiden Parametern abwich. Der synthetisierte Stimulus S, der ein nicht-prototypisches Timbre für eine menschliche Stimme besaß, evozierte die stärkste neurophysiologische Reaktion.

Die vorliegenden Daten deuten darauf hin, daß ungewöhnliche Stimminformation auch präattentiv auf besondere Weise verarbeitet wird. Damit besteht aber auch die Möglichkeit, daß Quelle- und Filterparameter bei diesen abweichenden Stimmen anders verarbeitet werden als bei natürlichen Lautsprachsignalen. Bevor die Implikationen der vorgestellten Daten eingehend diskutiert werden, war in Experiment Ib daher die Frage zu untersuchen, wie sich die Stimmverarbeitung bei natürlichen Sprachsignalen gleichen Wohlklangs vollzieht.

## 8.2 EXPERIMENT Ib

### 8.2.1 Fragestellung

Das folgende Oddball-Experiment untersuchte die Frage, wie das auditorische System auf zwei natürliche Sprachreize reagiert, die vom Standardreiz in unterschiedlichen Parametern abweichen. Es wurden dieselben Experimentalreize verwendet wie in der vorhergehenden Studie, jedoch wurde nun das grundfrequenzveränderte Sprachsignal als Standardstimulus eingesetzt (G_std), und die Stimuli W, M und S als Devianten. Es wurde angenommen, daß das künstliche Sprachsignal S wiederum die stärkste Stromdichte evozieren würde. Für die Verarbeitung der beiden natürlichen Stimuli W und M ergaben sich zwei Hypothesen:
**A) Natürlichkeitshypothese**: Die beiden Stimuli unterscheiden sich nicht in ihrer Natürlichkeit und deswegen sollte auch ihre Verarbeitung zu keinen signifikanten Unterschieden hinsichtlich der Stromdichte führen.
**B) Parameterhypothese**: Die ereigniskorrlierten Stromstärken bei der Verarbeitung der Stimuli W und M sollten unterscheiden sich. Sollte sich dabei eine höhere Stromdichte für die weibliche Stimme zeigen, so würde das bedeuten, daß die Grundfrequenz hierfür eine entscheidende Rolle spielt, weil sich die weibliche Stimme in diesem Merkmal stärker vom Standard unterscheidet als die männliche. Sollte eine höhere Stromdichte für den Stimulus M auftreten, so wäre daraus zu schließen, daß die spektrale Zusammensetzung (d.h. das Stimmtimbre) eine größere Bedeutung trägt.

### 8.2.2 Methode

#### 8.2.2.1 Stimuli

*Akustische Parameter der Stimuli*

Alle Experimentalreize entsprachen denen aus Experiment I. Allerdings wurde nun der grundfrequenzveränderte Stimulus als Standard-Reiz verwendet. Deshalb trägt er hier die Bezeichnung G_std, während das Sprachsignal des männlichen Sprechers nunmehr als M bezeichnet wird.

*Stimulusrating*

Es wurde ein Ähnlichkeitsrating durchgeführt, an dem dieselben 15 Versuchspersonen teilnahmen, die auch schon in der Ähnlichkeitseinschätzung von Experiment Ia teilgenommen hatten. Sie konnten nun die Sprachsignale wieder alle beliebig oft anhören und sollten die Ähnlichkeit der Stimuli M und W zum grundfrequenzveränderten Signal G beurteilen. Die männliche Stimme (M) erzielte dabei einen Wert von 3.2, Stimulus W einen Wert von 3.7. Es ergaben sich jedoch keine signifikanten Unterschiede zwischen den Bedingungen ($t < 1$).

#### 8.2.2.2 Probanden

Es nahmen 10 Probanden (4 männlich) im Alter von 22 bis 27 Jahren (Durchschnitt: 25 Jahre) an dem Experiment teil. Alle waren Muttersprachler des Deutschen und nach dem Edinburgh Inventar (Oldfield, 1971) als Rechtshänder klassifiziert. Keiner der Teilnehmer berichtete audiologische Beeiträchtigungen. Die Probanden wurden für ihre Teilnahme am Experiment bezahlt.
Versuchsdurchführung, Datenerhebung und Quellanalyse entsprachen denen von Experiment Ia.

### 8.2.3 Ergebnisse

Die ereigniskorrelierten magnetischen Felder zeigen in den devianten Bedingungen wiederum zwischen 100 ms und 300 ms eine deutliche Mismatch Negativierung (siehe Abbildung 8.6). Für die ausgewählten Kanäle 78 und 88 zeigten sich für die gepaarten t-Tests über die Magnetfeldstärken wiederum signifikante Unterschiede zwischen Standardstimulus und jeweiligem Devianten (Kanal 78: Deviant M $t_{(9)} = 2.16, p = 0.059$; Deviant W $t_{(9)} = 5.31, p < 0.001$; Deviant S $t_{(9)} = 8.52, p < 0.0001$; Kanal 88: Deviant M $t_{(9)} = 2.57, p = 0.05$; Deviant W $t_{(9)} = 4.84, p < 0.001$; Deviant S $t_{(9)} = 5.14, p < 0.001$).[5]

[5] Die Effektstärken deuten an, daß Stimulus M eine schwächere Reaktion auslöste, als Deviant W. Es ist jedoch zu beachten, daß es sich hier nur um das Signal zweier Kanäle handelt. Eine andere Ausrichtung oder Lokalisation des Stromdipols könnte dafür verantwortlich sein. Aussagekräftiger ist

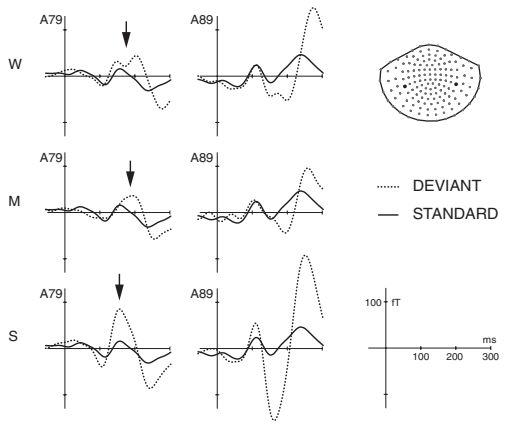



Abbildung 8.6: Gemittelte ereigniskorrlierten Magnetfeldstärken über dem linken (Kanal 78) und rechten (Kanal 88) auditorischen Kortex. Im Vergleich zum Standardstimulus G_std zeigen die Reaktionen auf die Deviants eine erhöhte Mismatch-Aktivierung zwischen 100 und 300 ms nach Stimulusbeginn.

Abbildung 8.7 zeigt die mittleren Stromstärken der linken und rechten Hemisphäre, für den Bereich 100-300 ms. Die ereigniskorrelierten Stromdichten zeigen wie erwartet eine schwache Reaktion auf G_std, und eine die höchste Stromdichte für das synthetisierte Sprachsignal S. Dabei weisen alle Bedingungen eine stärkere Beteiligung der linken Hemisphäre auf.

Die zweifaktorielle Varianzanlyse BEDINGUNG $\times$ HEMISPHÄRE zeigte einen signifikanten Haupteffekt BEDINGUNG ($F_{(3,27)} = 21.61, p < 0.0001, \varepsilon = 0.6871$), sowie einen Haupteffekt HEMISPHÄRE ($F_{(3,27)} = 6.33, p < 0.05$), jedoch keine signifikante Interaktion der Faktoren. Entscheidend ist jedoch der Vergleich zwischen den Devianten M und W. Diese beiden Bedingungen gingen daher in die ausführliche Analyse der Faktoren DEVIANT (M, W), ZEITFENSTER (100-150 ms, 150-200 ms, 200-250 ms, 250-300 ms) und HEMISPHÄRE (links, rechts) ein. Dabei zeigte sich lediglich der schon erwähnte HEMISPHÄREN-Unterschied ($F_{(1,9)} = 6.58, p < 0.05$), ansonsten aber keinerlei signifikante Haupteffekte oder Interaktionen.

Die räumliche Verteilung der Aktivierung ist schließlich in Abbildung 8.8 darge-

in jedem Falle die Modellierung der Stromdichten, die das Signal aller Kanäle berücksichtigt und in der Tat keinen Unterschied in den Reaktionen auf die Devianten W und M aufzeigt.

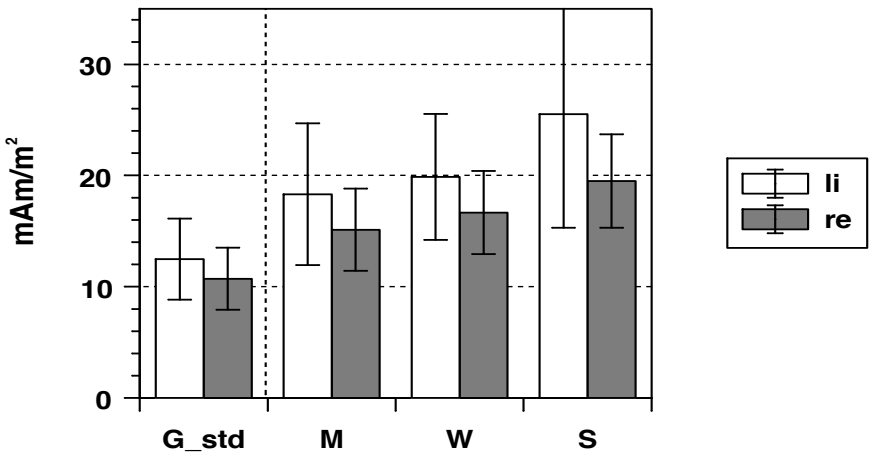



Abbildung 8.7: Mittlere Stromdichten von 100-300 ms nach Stimulusbeginn für die vier Experimentalbedingungen, aufgeteilt nach Hemisphären (li=links; re=rechts).

stellt. Sie zeigt ein Hauptaktivierungsgebiet im medialen und superioren Temporallappen.

### 8.2.4 Diskussion

Alle devianten Stimuli evozierten eine Mismatch Negativierung. Diese MMNm zeigte in allen Bedingungen eine stärkere Beteiligung der linken Hemisphäre. Die Ursache, für die eher bilaterale Verteilung der Aktivierung in Experiment Ia kann in der Tatsache liegen, daß die Probanden dort einen niedrigeren durchschnittlichen Händigkeitsquotienten (vgl. Oldfield, 1971) besaßen (Experiment Ia: 84.7, Experiment Ib: 94.5), da die Händigkeit einen Einfluß auf die Lateralisation der Sprachverarbeitung hat (Knecht et al., 2000; Lurito & Dzemidzic, 2001; Poeppel et al., 1997).[6] Trotz dieser Unterschiede in den ereigniskorrelierten Stromdichten stimmten die beiden Experimente hinsichtlich des *Effektes* des Sprecherwechsels überein, der Effekt war in beiden Fällen bilateral verteilt.

Repliziert wurde weiterhin der Akzeptanz-Effekt der vorhergehenden Studie: Wie-

[6] Unterstützung erfährt diese Annahme weiterhin durch die Tatsache, daß im nachfolgenden Experiment II, bei dem der mittlere Händigkeitsquotient der Probanden ebenfalls sehr hoch (bei 97) lag, auch eine Linkslateralisation zu beobachten war.

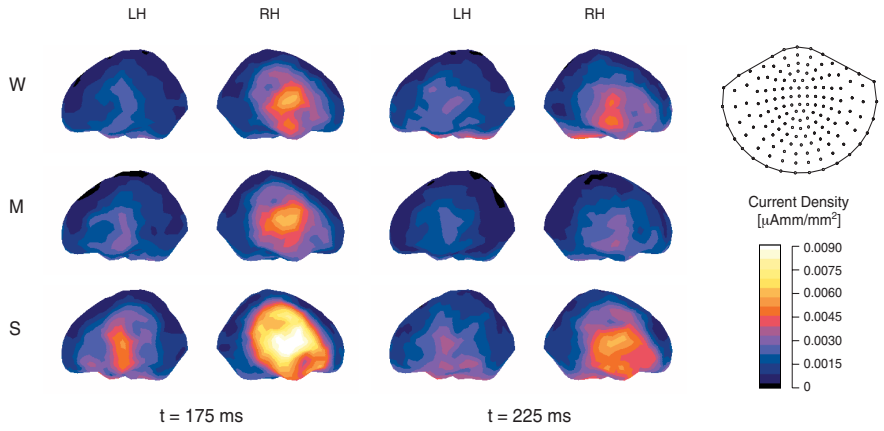



Abbildung 8.8: Verteilung der ereigniskorrelierten Aktivierungen auf die drei Devianten W, P und S: Stromdichten der linken (LH) und rechten (RH) Hermisphäre zu den Zeitpunkten t= 125 ms und t = 175 ms.

derum evozierte das künstlich synthetisierte Sprachsignal die höchste Stromdichte. Das Hauptinteresse des Experiments lag jedoch auf den beiden natürlichen Sprachsignalen W und M. Weder die Einschätzung im Ähnlichkeitsrating, noch die mittleren ereigniskorrelierten Stromdichten im Zeitraum von 100-300 ms unterschieden sich signifikant voneinander. Dieser Befund unterstützt die Annahme, daß die Reaktionen tatsächlich jeweils in Bezug auf ein prototypisches Template für eine natürliche Männer/Frauenstimme erfolgen und daher auch für die physikalisch sehr verschiedenen Stimuli in gleichem Maße ausfallen. Alternativ wäre denkbar, daß die Parameterhypothese zutrifft, und den Filter- bzw. Quellparametern dieselbe Gewichtung zukommt. Das hieße, die Stimmen würden parameterweise verglichen und dem Unterschied der Grundfrequenz (G_std vs. W) käme dieselbe Bedeutung zu, wie dem Unterschied der Filterparameter (G_std vs. M). Im Zusammenhang mit Experiment Ia ist das, wenngleich nicht auszuschließen, aber wenig wahrscheinlich. Dort wurde die ungewöhnliche Filter/Quelle-Konfiguration des grundfrequenzveränderten Stimulus nämlich *nicht* parameterweise mit den physikalischen Eigenschaften des Standardstimulus verglichen, sondern Deviant G evozierte trotz der akustischen Ähnlichkeit eine signifikant höhere Stromdichte als die weibliche Stimme W. Die Daten des vorliegenden Ex-

perimentes werden daher im Sinne Natürlichkeitshypothese interpretiert.

## 8.3 Allgemeine Diskussion der Experimente Ia und Ib

In Tierexperimenten wurde eine Selektivität für die auditorische Verarbeitung artspezifischer Vokalisationen nachgewiesen (Tian et al., 2001; Eggermont, 2001). Bildgebende Studien beim Menschen weisen auf eine "Selektivität" für die menschliche Stimme hin (Belin et al., 2000; Binder et al., 1997). Die vorliegenden Experimente, insbesondere die starken Reaktionen auf ungewöhnliche Stimmsignale, bestätigen diese Befunde und zeigen darüber hinaus, daß die spezifischen Prozesse der Stimmwahrnehmung schon in der präattentiven auditorischen Verarbeitung zu finden sind. In eine ähnliche Richtung deutet die Studie von Levy, Granot & Bentin (2001). Dort wurde eine P3-Komponente für die Wahrnehmung der menschliche Singstimme im Vergleich zu verschiedenen Instrumenten gefunden[7]. Diese neuronale Reaktion wurde als ein Indikator dafür interpretiert, daß die menschliche (Sing-)Stimme im Gegensatz zum Klang verschiedener Instrumente eine besondere Bedeutung (*ecological validity*) besitzt und auf spezifische Weise verarbeitet wird. Liegt bei diesen Experimenten jedoch ein Wechsel der konzeptuellen Kategorie ("Musikinstrumente") vor, so zeigen die hier vorgestellten Experimente, daß das Verarbeitungssystem auch auf Abweichungen der Stimmnatürlichkeit von Sprachsignalen reagiert.

Dabei deuten die ereigniskorrelierten Mismatch-Reaktionen insbesondere auf den grundfrequenzveränderten Devianten darauf hin, daß bei einem Wechsel der Stimme kein Vergleich einzelner Signalparameter (F0, Formanten) von Standard und Deviant durchgeführt wird (siehe Diskussion von Experiment Ia). Vielmehr ist die *Gesamtkonfiguration* aus Tonhöhe und Timbre der jeweiligen Stimme für die Perzeption entscheidend. Sprachsignale, deren Parameterkonfiguration von der Zusammensetzung natürlicher Stimmsignale abwich, evozierten grundsätzlich eine höhere Stromdichte als natürliche Sprachsignale (Experiment Ia). Expe-

[7] Die visuelle Inspektion der abgebildeten ereigniskorrelierten Potentiale zeigen im frühen Zeitbereich eine Negativierung, die durchaus als Mismatch Negativierung interpretiert werden kann. Diese wird in der Publikation jedoch nicht diskutiert.

riment Ib bestätigte diesen Befund und zeigte darüber hinaus, daß Stimmsignale gleicher Natürlichkeit keine Unterschiede in der Stromdichte evozieren, wenn sie nur in der Grundfrequenz oder nur in der Formantstruktur abweichen. Wiederum scheint also die Gesamtkonfiguration aus Tonhöhe und spektralen Eigenschaften entscheidender für die neuronale Reaktion. Die Art der Verarbeitung von Stimminformation kann demnach als *konfigural* oder *holistisch* bezeichnet werden.

In einige Publikationen wird die Ansicht vertreten, daß die MMN auf komplexe Töne grundsätzlich eine gestaltweise Verabeitung von Merkmalskombinationen reflektiert (Paavilainen et al., 2001b; Takegata et al., 1999; Gomes et al., 1997). Dieser Ansatz ist jedoch nicht unumstritten (vgl. Deacon et al., 1998). Unabhängig davon, ob komplexe Signale *generell* in holistischer Weise verarbeitet werden oder nicht, scheint aber zumindest die spezifische Wahrnehmung der Stimminformation solchen Verarbeitungsmechanismen zu unterliegen.

Wie bereits beschrieben (siehe S. 50) nehmen Van Lancker & Kreiman (1987) eine holistische Verarbeitung nur für die Wahrnehmung bekannter Stimmen an. Für die Verarbeitung unbekannter Stimmen postulieren sie jedoch eine parameterweise Analyse der Stimmen. Wie oben dargestellt beruhen diese Annahmen vor allem auf Erkenntnissen über die allgemeinen Verarbeitungsweisen der Hemisphären und sind nicht direkt experimentell belegt worden. Die Ursache dafür, daß die hier vorgestellten Daten in eine andere Richtung weisen, könnte zum einen an der iterativen Präsentation der Stimmen liegen, denn diese sind dadurch nicht mehr völlig unbekannt, auch wenn sie nicht mit besonderen Erinnerungen oder dem Wissen um die Identität des Sprechers verbunden sind. Darüber hinaus gibt es aber Beobachtungen, die dafür sprechen, daß es sich nicht um ein bloßes Artefakt des Versuchsdesigns handelt. Die Berücksichtigung der Gesamtkonfiguration einzelner Parameter spielt nämlich auch in der visuellen Perzeption eine bedeutsame Rolle.

Es finden sich Hinweise, daß auch die Wahrnehmung von Gesichtern in holistischer Weise stattfindet (vgl. Hancock, Bruce & Burton, 2000). In der Studie von Young, Hellawell & D.C. (1987) wurden z.B. Gesichter präsentiert, die aus zwei Fotographien unterschiedlicher Personen zusammengesetzt waren. Han-

delte es sich bei dem Ergebnis der Synthese um ein mögliches menschliches Gesicht, so war es relativ schwer, die Identitäten der einzelnen Gesichtsteile zu bestimmen. Ergab die Synthese aber kein mögliches Gesicht (*misaligment*), so war diese Zuordnung deutlich einfacher. Auch hier kann Ähnlichkeit der einzelner Parameter (Mund, Nase, Augen) mit der jeweiligen Originalfotographie nicht ohne weiteres abgeglichen werden, sondern unterliegt dem Einfluß der Gesamtkonfiguration. Ein anderer bekannter Effekt ist die Thatcher-Illusion (Thompson, 1980) verdeutlicht, die zeigt, daß eine Inversion verschiedener Gesichtsteile langsamer detektiert und nicht in vollem Maße erfaßbar ist, wenn das entsprechende Portrait um 180° gedreht ist. Bei normaler Ausrichtung hingegen ist eine effektive konfigurale Verarbeitung möglich (Lewis & Johnston, 1997). Da die Gesichter- und Stimmverarbeitung eine Reihe von Parallelen aufweisen,[8] ist die Annahme ähnlicher, konfigural-holistischer Verarbeitungsweisen plausibel. Hinzukommt, daß beide Fähigkeiten auf einem hochvariablen Input beruhen. Dennoch gelingt es in beiden Dimensionen in der Regel mühelos, tausende von Konfigurationen derselben Grundparameter zu diskriminieren, so daß Mitmenschen unterschieden und wiedererkannt werden können. Eine holistische Verarbeitungsweise scheint ein effektiver Weg, diese Leistung zu vollbringen.

Als Haupterkenntnisse aus den ersten beiden Experimenten der vorliegenden Arbeit läßt sich festhalten, daß es Hinweise auf eine konfigurale Verarbeitung der Stimminformation gibt und daß nicht-protoypische (computermanipulierte/synthetische) Stimmsignale im Oddball-Paradima eine Mismatch Reaktion evozieren, die sich signifikant von der neuronalen Antwort auf durchschnittliche menschliche Stimmen unterscheidet. Die Daten sprechen damit eher für eine Abhängigkeit von permanenten prototypischen Musterrepräsentationen als von kurzzeitigen, experimentell bedingten Gedächtnisspuren einzelner Experimentalreize.

[8] Das Gesicht und die Stimme sind beide wichtige Merkmale in der Personenwahrnehmung. Sie kodieren die gleiche Art von Information (Alter, Geschlecht etc.). Studien zeigen, daß die Fähigkeiten zur Gesichter- und Stimmenerkennung sich auch auffällig parallel entwickeln (Mann, Diamond & Carey, 1979). In beiden Fällen scheinen dem Wiedererkennen bekannter Personen und der Unterscheidung unbekannter Personen unterschiedliche Verarbeitungsstrategien zugrundezuliegen (Hancock et al., 2000) und schließlich werden sowohl Stimmen, als auch Gesichter, die von der Norm abweichen, besser erinnert (siehe Kreiman, 1997; bzw. Hancock et al., 2000). Für Übereinstimmungen in den neuronalen Computationsmechanismen der Parameterkodierung siehe Shamma (2001).

Das synthetische Sprachsignal evozierte in beiden Experimenten die höchste Stromdichte. Aufgrund der Tatsache, daß hier ein passives Oddball-Paradigma verwendet wurde, ist es nicht möglich zu sagen, ob mit der eingeschränkten Natürlichkeit der Stimminformation auch die phonetische Dekodierung erschwert ist, und diese Erschwernis im Anstieg der Mismatchreaktion mit abgebildet wird. Es sei aber erwähnt, daß psycholinguistische Experimente zeigen, daß trotz einer guten Verständlichkeit von synthetischer Sprache, ihre Unnatürlichkeit, zu höheren Verarbeitungskosten führt (Nix, Mehta, Dye & Cutler, 1993; Duffy & Pisoni, 1992; Mackie, Dermody & R., 1987).[9] Die Abhängigkeit vom phonologischen Status, d.h. von der Prototypikalität der Sprachlaute und die damit verbundene relative Unabhängigkeit von der physikalischen Stimulusbeschaffenheit ist oben bereits verwiesen worden (siehe S. 25). In Studien, die sich diese Fragestellung widmeten, waren für nicht-prototypische Laute stets schwächere Mismatch Reaktionen gefunden worden als für prototypische Devianten. Eine einzige Studie berichtet hingegen einen umgekehrten Effekt: In der Studie von Sandridge & Boothroyd (1996) wurden im Oddball-Experiment verschiedene CV-Silben präsentiert, darunter auch eine Silbe, die einen Pseudovokal enthielt. Die Amplitude der ereigniskorrelierten MMN war für den Pseudovokal signifikant größer als für die natürlichen Sprachlaute. Dieser Pseudovokal bestand aus einer iterativen Aneinanderreihung einer einzelnen Periodenschwingung aus einer Aufnahme des Vokals /I/. Eine solche Manipulation führt jedoch nicht nur zu einem ungewöhnlichen Formantspektrum, sondern auch zu einem sehr unnatulichen Stimmsignal, das kaum noch als menschliche Stimme identifizierbar ist. Die Daten der hier vorgestellten Experimente Ia und Ib legen nahe, daß die größere Amplitude eher auf diese Unnatürlichkeit im Stimmsignal zurückzuführen ist als auf die spektrale Phonemqualität. Nach dieser Interpretation führt eine Prototypikalitätsverletztung sprachlicher und nichtsprachlicher Information zu qualitativ verschiedenen

[9] Aufgrund der relativ hohen Kontrollierbarkeit akustischer Parameter ist die Verwendung synthetisierte Sprache als experimentelles Stimulusmaterial beliebt. Da aber also sowohl auf Verhaltensebene als auch in neurophysiologischen Studien Unterschiede in der Verarbeitung natürlicher und künstlicher Sprache gezeigt wurden, sind die Vor- und Nachteile der experimentellen Verwendung synthetisierter Sprachsignale im Einzelfall abzuwägen und ist aus methodologischen Gründen zumindest auf eine unkontrollierte Mischung natürlicher und synthetischer Sprache zu verzichten.

Mismatch Negativierungen. Ob und wie die Detektion einer phonetischen Abweichung und die Perzeption eines Sprecherwechsels *interagieren*, ist mittels neurophysiologischen Verfahren jeoch bisher nicht untersucht worden. Dieser Fragestellung ist das folgende Experiment gewidmet.

# Kapitel 9

# EXPERIMENT II: Stimminformation und phonetische Verarbeitung

## 9.1 Experiment II

### 9.1.1 Fragestellung

Die kognitive Verarbeitung von Lautsprache beinhaltet unter anderem die Dekodierung der phonetischen Cues, sowie die Extraktion der Sprecherinformation aus dem Signal. Während in den vorangegangenen Experimenten allein die Perzeption der Stimminformation untersucht wurde, soll es nun um ihre Interaktion mit der phonetischen Verarbeitung[1] gehen.

Im Kapitel 6 der vorliegenden Arbeit sind bereits klinische und behaviorale Daten vorgestellt worden, die darauf hindeuten, daß Wort- und Stimmverarbeitung

[1] Tatsächlich werden im folgenden Experiment verschiedene Wörter präsentiert. Da aber keine Aussagen über etwaige lexikalisch-semantische Verarbeitungsprozesse getroffen werden, und diese zu einem so frühen Zeitpunkt nach Stimulusbeginn im passiven Oddball-Paradigma auch nicht zu erwarten sind, wird im folgenden allein auf die phonetische Verarbeitung der Experimentalreize Bezug genommen.

sich gegenseitig beeinflussen. Im folgenden Experiment wurde der Frage nachgegangen, ob sich Hinweise auf eine solche Interaktion bereits in der präattentiven auditorischen Verarbeitung wiederfinden, oder ob die Effekte eher der zeitlich späteren, höheren Verarbeitung zuzuordnen sind. Um aufgabenspezifische Einflüsse zu vermeiden, wurde wie in den vorhergehenden Studien ein passives Oddball-Paradigma gewählt, wobei die devianten Reize hinsichtlich der phonetischen oder der sprecherspezifischen Information abwichen. Wie oben beschrieben, waren Mismatch-Experimente bisher entweder der phonetischen Devianz (siehe Näätänen, 2001, für eine Übersicht) oder der Abweichung der Stimminformation (Titova & Näätänen, 2001) gewidmet, wobei für jede der Devianzdimensionen eine MMN beobachtet wurde. Unbekannt ist jedoch das *Zusammenspiel* phonetischer und stimmrelevanter Information in der präattentiven Phase der neuronalen Sprachverarbeitung. In der vorliegenden MEG-Studie werden daher die Mismatch-Reaktionen auf die Verletzungen beider Dimensionen miteinander verglichen. Es wurde erwartet, daß sich die Mismatch Reaktionen je nach Dimension der Abweichung unterscheiden, d.h. daß also die Latenz, Stärke und Lokalisation des äquivalenten Stromdipols reflektieren sollte, ob es sich um einen Wechsel der Sprechstimme oder des Wortes handelt. Es stellt sich beispielsweise die Frage, ob sich schon in der präattentiven Verarbeitung eine Spezialisierung der Hemisphären zeigt. In den oben geschilderten MEG-Experimenten wurde eine bilaterale Verteilung der Mismatch Reaktion für die Stimmabweichung beobachtet. Für die Verarbeitung der phonetischen Abweichung wird hingegen aufgrund anderer Studien (z.B. Näätänen et al., 1997; Rinne et al., 2000; Kasai et al., 2001; Näätänen, 2001, für eine Übersicht) eine Linkslateralisation der magnetischen Felder erwartet. Sollte sich tatsächlich ein Hemisphärenunterschied zeigen, so wäre im nächsten Schritt die Frage der Hemisphärenbeteiligung bei gleichzeitiger Abweichung in beiden Dimensionen interessant. Um dies zu untersuchen, wurde als dritter Deviant ein Stimulus präsentiert, der sich vom Standardreiz sowohl hinsichtlich der Stimminformation als auch hinsichtlich des geäußerten Wortes unterschied, also eine zweidimensionale Abweichung beinhaltete. Diese Experimentalbedingung ist auch deswegen von besonderem Interesse, weil sie mögli-

cherweise eine Aussage darüber erlaubt, ob es sich bei der präattentiven Verarbeitung von phonetischer Struktur und Stimme um serielle oder parallele Verarbeitungsprozesse handelt. Eng damit in Verbindung steht die Frage, ob diese Prozesse unabhängig voneinander sind oder interagieren. Sollte bereits für die frühe, präattentive Wahrnehmung eine Interaktion paralleler Verarbeitungsprozesse vorliegen, ist für die zweidimensionale Devianz ein nicht-additives/subadditives Aktivierungsmuster zu erwarten, d.h. eine MMNm, die sich nicht durch eine einfache Addition der zwei Einzelabweichungen erklären läßt. Auf der anderen Seite sollte es zu einem additiven Verhalten kommen, wenn die akustischen Parameter, die zur Detektion eines Sprecherwechsels bzw. einer phonetischen Abweichung führen, unabhängig voneinander verarbeitet werden.
Um das zeitliche und räumliche Auflösungsvermögen des MEG zu nutzen, wurde eine Dipolmodellierung der neuronalen Aktivität durchgeführt.

### 9.1.2 Methode

#### 9.1.2.1 Stimuli

Die natürlichsprachlichen Stimuli wurden aus einer Menge von Einzelwortaufnahmen aufgrund ihrer ähnlichen prosodischen Kontur und ihrer hohen Übereinstimmung der segmentalen Dauern ausgewählt. Der Standardstimulus (Std) bestand aus einer Äußerung des Wortes "Test" eines männlichen Sprechers. Die drei Devianten bestanden aus
**Ph**: dem phonetischen Devianten, einer Äußerung des Wortes "Dach" desselben (Standard-)Sprechers;
**Sp**: dem Sprecher-Devianten, einer Äußerung des Wortes "Test" einer Sprecherin;
**PhSp**: dem zweidimensionalen Devianten, der eine kombinierten Abweichung der phonetischen Struktur und der Sprecherinformation umfaßte, nämlich die Äußerung des Wortes "Dach" einer Sprecherin.
Die Dauerstrukturen der Stimuli sind in Appendix B.1 beschrieben, ihre spektralen Eigenschaften in Appendix B.2. Die Lautstärken wurden normalisiert und die

Stimuli 40 dB über der vorher ermittelten Hörschwelle (s. Experiment Ia) präsentiert.

#### 9.1.2.2 Probanden

Es nahmen 20 Probanden an der Studie teil. Aufgrund der unten beschriebenen Kriterien gingen in die Auswertung 11 Versuchsteilnehmer (9 männlich) im Alter von 18 bis 26 Jahren (Durchschnitt: 23 Jahre) ein. Alle waren Muttersprachler des Deutschen und nach dem Edinburgh Inventar (Oldfield, 1971) als Rechtshänder klassifiziert. Keiner der Teilnehmer berichtete audiologische Beeiträchtigungen. Die Probanden wurden für ihre Teilnahme am Experiment bezahlt.

#### 9.1.2.3 Versuchsablauf

Die Versuchsdurchführung entsprach den vorhergehenden Experimenten. Auch hier waren die Probanden instruiert, sich auf einen Film (ohne Ton) zu konzentrieren. Jeder Deviant wurde pseudorandomisiert präsentiert (Oddball Paradigma) und hatte insgesamt eine relative Häufigkeit von 0.05.

#### 9.1.2.4 Datenerhebung

Siehe EXPERIMENT Ia.

#### 9.1.2.5 Quellanalyse

Da dieses Experiment feine Unterschiede in der neurophysiologischen Reaktion untersuchen sollte, spielte die räumlich/zeitliche Auflösung der Aktivierungsmuster eine besondere Rolle. Es wurde daher eine Dipolmodellierung unter Verwendung eines standardisiertes BEM-Modells durchgeführt. Dieses wurde dann nicht nur im Hinblick auf die Kopfposition normiert, sondern auch auf die individuelle Kopfgröße und -form skaliert. Die Aktivierung der ersten 300 ms nach dem Stimulus-Onset wurden als Mismatch-relatierte Aktivität akzeptiert. Zunächst wurden diejenigen Probanden identifiziert, die das typische dipolare Muster der N1m

Komponente aufwiesen (vgl. Nakasato et al., 1995). Ein Beispiel ist in Abbildung 9.1 dargestellt. Ein Proband wurde wegen technischer Probleme von der Ana-

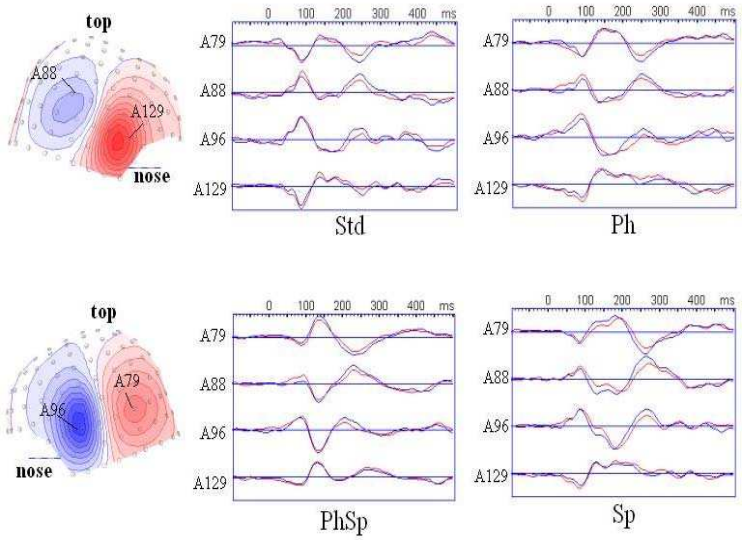



Abbildung 9.1: Ereigniskorrelierte Magnetfelder: Links die typischen dipolaren Magnetfeldmuster (oben: Ansicht der rechten Hemisphäre, unten: Ansicht der linken Hemisphäre); rechts die Magentfeldstärken im Verlauf der ersten 500 ms nach Stimulusbeginn und der Baseline (-200 - 0 ms) für zwei ausgewählte MEG-Kanäle im Zentrum des Magnetfeldes. Dargestellt ist die Stärke des Magnetfeldes in der jeweiligen Bedingung (grand average) für die erste und zweite Sitzung. Der Verlauf der Magnetfeldstärken der beiden Sitzungen unterscheidet sich nicht signifikant und demonstrieren somit die Reliabilität der Messungen.

lyse ausgeschlossen, drei Probanden wiesen in einer oder mehreren Bedingungen kein entsprechendes dipolares Feld in der linken Hemisphäre auf, einer in der rechten Hemisphäre. Bei drei weiteren Probanden zeigte sich in keiner Hemisphäre ein dipolares Muster. Eine häufige Ursache für das Fehlen der dipolaren Aktivierung ist eine stark ventrale Lokalisation des Dipols (bei einem großen Kopfumfang), die bewirkt, daß das Magnetfeld von den fixen MEG Kanälen nicht abgefaßt werden kann.[2] Die Daten einer weiteren Versuchsperson wurden aus-

[2]Die Anzahl der ausgeschlossenen Probanden ist relativ hoch. In einer methodologisch vergleichbaren Studie von Nakasato et al. (1995) traten bei 4 von 20 Versuchspersonen keine klar dipolaren Muster auf. In diesen Fällen wurden nur die Daten der entsprechenden Hemisphäre und Bedingung zurückgewiesen. Da es im vorliegenden Experiment aber um den direkten Vergleich der Bedingungen

geschlossen, da sie dem 4 $\sigma$-Kriterium des statistischen Ausreißer-Test[3] nicht standhielten. Die Daten der verbleibenden 11 Probanden gingen in die statistische Analyse (ANOVA) ein. Es ist wahrscheinlich, daß die N1m-Komponente der Devianten, eine Kombination aus einer N100m Komponente und einer sie überlagernden Mismatch Negativierung ist. Wegen der physikalischen Unterschiede zwischen Standard und Devianten und den möglischerweise daraus resultierenden Unterschieden bereits in den Primärkomponenten (N100m) wurde allerdings von der oft üblichen Differenzbildung (Deviant - Standard) zur Extraktion der MMNm abgesehen. Um eine neutralere gemeinsame Bezeichnung für Standard und Devianten zu verwenden, wird dieses ereigniskorrelierten Magnetfeld, das durch einen negativen Stromdipol erklärt werden kann, im Folgenden als N1m bezeichnet. Zunächst wurden die Latenzen der N1m pro Versuchsperson, Bedingung und Hemisphäre ermittelt. Dann wurde für jeden Probanden ein 20 ms breites Zeitfenster um das entsprechende Maximum ausgewählt und ein Paar fixer Dipole angepaßt (*fixed dipole fitting*, siehe Mosher & Leahy 1998). Die jeweiligen Dipol-Positionen wurden auf das standardisierte Hirnmodell transformiert. Zur Berechnung der Dipolstärke wurden die Positionen konstant gehalten und die Aktivierungsverläufe der N1m für jeden Probanden innerhalb der ersten 300 ms nach Stimulusbeginn berechnet. Es ist hierbei davon auszugehen, daß sich die Lokalisation im Zeitrahmen der N100 und P200 weniger als 1 cm ändert (Rif, Hari, Hämäläinen & Sams, 1991). Selbst bei einer Bewegung des Dipols (bzw. einer sukzessiven Aktivierung unterschiedlicher Neuronenpopulationen) wären die resultierenden Aktivierungstopographien so ähnlich, daß die Annahme einer einzelnen Quelle zur Erklärung der Aktivierung in diesem Zeitbereich eine adäquate Näherung darstellt (vgl. Hari et al., 2000). Um die Dipolstärken zu berechnen, wurden die resultierenden Aktivierungskurven auf der Zeitachse so angepaßt, das die N1-Maxima übereinstimmten (*time alignment*). Durch diese Anpassung wurde der Jitter aufgrund unterschiedlichen Latenzen zwischen den Probanden ausgeglichen. Zehn Abtastwerte (insgesamt 40 ms) um die Maxima der N1m gin-

ging, wurde ggf. alle Daten einer Versuchperson zurückgewiesen und somit ein strenges within-Design eingehalten (die verschiedenen Bedingungen umfassen Daten derselben Versuchspersonen).

[3](vgl. Sachs, 1997, S.364 ff.)

gen in die statistische Analyse der mittleren Dipolstärken ein.

### 9.1.3 Ergebnisse

Tabelle 9.1 faßt die Ergebnisse der statistischen Analyse von Dipollokalisation, Dipolstärke und Aktivierungslatenzen zusammen.

| *Faktor* | Dipollokalisation | | | Dipolstärke | Gipfellatenz |
|---|---|---|---|---|---|
| | x | y | z | | |
| *Bedingung* | * | - | - | ** | * |
| *Hemisphäre* | ** | * | - | ** | - |
| *Bedingung* × *Hemisphäre* | - | - | - | - | - |
| Std vs. Ph | * | | | ** | - |
| Std vs. Sp | ** | | | ** | ** |
| Std vs. PhSp | ** | | | ** | * |

Tabelle 9.1: Statistische Analyse der Dipollokalisationen (x: posterior-anterior; y: lateral - medial; z: dorsal - ventral), Dipolstärken und Latenzen. (**: $0.001<p<0.01$; *: $0.01<p<0.05$)

*Dipol-Lokalisationen*

Die Dipollokalisation wurde statistisch als separate Analysen (ANOVAS) der drei Raumkoordinaten überprüft (x: posterior-anterior; y: lateral - medial; z: dorsal - ventral). Dazu wurden jeweils eine zweifaktorielle Varianzanalyse der Faktoren BEDINGUNG (Std, Ph, Sp, PhSp) und HEMISPHÄRE (recht, links) durchgeführt. Für die x-Achse zeigten sich dabei sowohl ein Haupteffekt BEDINGUNG ($F_{(3,30)} = 4.34, p < 0.05$) als auch HEMISPHÄRE ($F_{(1,10)} = 15.29, p < 0.01$) aber keine signifikante Interaktion zwischen beiden Faktoren. Für die y-Achse war lediglich ein tendenzieller HEMISPHÄREN-Unterschied zu beobachten ($F_{(1,10)} = 3.57, p = 0.08$), für die z-Achse ein starker Trend für den Faktor BEDINGUNG ($F_{(1,10)} = 2.78, p = 0.06$). Die HEMISPHÄREN-Effekte sind darauf zurückzuführen, daß die Dipole in der rechten Hemisphäre im Vergleich zu denen der linken Hemisphäre anterior und eher lateral lokalisiert sind. Der Haupteffekt BEDINGUNG

für die x-Koordinate wurde zunächst aufgelöst in Einzelvergleiche jedes Devianten mit dem Standard (einseitige, gepaarte t-Tests). Es ergab sich hier für alle Devianten signifikante Unterschiede im Vergleich zur Standardbedingung oder zumindest einen starken Trend (Ph: $t_{(10)} = 2.13, p = 0.0585$; Sp: $t_{(10)} = 3.25, p < 0.01$ PhSp: $t_{(10)} = 3.04$ , $p < 0.05$). Die Devianten wurden separat auf Lokalisationsdifferenzen überprüft, jedoch lieferte keine der dreifach gestuften einfaktorielle ANOVAS für die einzelnen Raumkoordinaten hier signifikanten Unterschiede. Es lagen also zwischen den Devianten und dem Standard Lokalisierungsunterschiede vor, jedoch unterschieden sich die drei Deviant-Bedingungen untereinander nicht (siehe Abbildung 9.2).

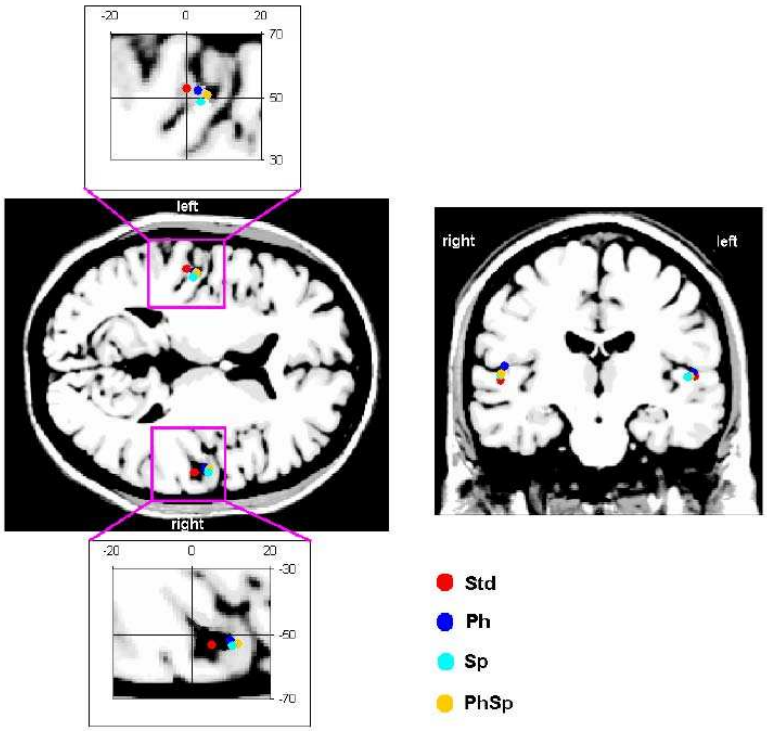
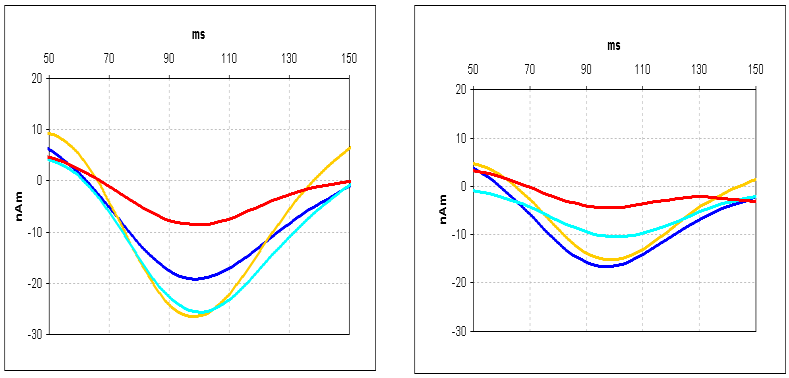



Abbildung 9.2: Ergebnisse der Dipollokalisation (Mittel aller Probanden) abgebildet auf einem Standard-Gehirn (Montreal Brain, McGill); axiale und koronale Schichtführung.

*Dipolstärken*

Die Stärken der Dipole sind in Abbildung 9.3 dargestellt. Ihre statistische Analyse (ANOVA der Faktoren BEDINGUNG $\times$ HEMISPHÄRE) zeigt einen signifikanten Haupteffekt HEMISPHÄRE ($F_{(1,10)} = 9.82, p < 0.05$), der eine stärker Aktivierung der linken Hemisphäre abbildet, sowie einen Haupteffekt BEDINGUNG ($F_{(3,30)} = 9.19, p < 0.001$), aber keine Interaktion der Faktoren. Der BEDINGUNGS-Effekt wurde aufgelöst in die paarweisen Vergleiche (einseitiger, gepaarter t-Test) zwischen jeweiligem Deviant und Standard; sie zeigten in allen drei Fällen signifikante Unterschiede auf (Ph: $t_{(10)} = 3.32, p < 0.01$; Sp: $t_{(10)} = 5.00$ , $p < 0.001$ PhSp: $t_{(10)} = 8.70$ , $p < 0.0001$). Die drei Deviant-Bedingungen unterschieden sich dabei nicht signifikant voneinander. Die Interaktion HEMISPHÄRE $\times$ DEVIANT lag im Trendbereich ($F_{(3,30)} = 2.76, p = 0.0877$), was auf eine tendenziell symmetrischere (bilaterale) Dipol-Aktivierung für Bedingung PhSp zurückzuführen ist.




linke Hemisphäre rechte Hemisphäre

Abbildung 9.3: Verlauf der Dipolstärken für die N1m in der linken und rechten Hemisphäre (Std: rot; Ph: dunkelblau; Sp: hellblau; PhSp: gelb).

*Latenzen*

Die Latenzen der N1m sind in Abbildung 9.4 dargestellt.

Hier ergab die ANOVA der Faktoren BEDINGUNG $\times$ HEMISPHÄRE einen signifikanten Haupteffekt BEDINGUNG ($F_{(3,30)} = 3.81, p < 0.02$), aber keinen si-

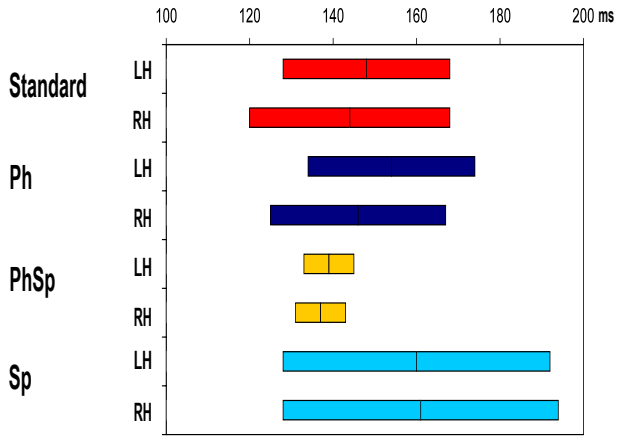



Abbildung 9.4: Latenzen der N1m in der linken und rechten Hemisphäre (Std: rot; Ph: dunkelblau; Sp: hellblau; PhSp: gelb)

gnifikanten Effekt HEMISPHÄRE und keine signifikante Interaktion der Faktoren. Die Einzelvergleiche (paarweise t-Tests) zwischen jeweiligem Devianten und der Standardbedingung zeigte, daß sich die Latenzen für Bedingung Ph, sowie für Bedingung PhSp nicht vom Standard unterschieden. Für Bedingung Sp (Sprecherabweichung) waren die Latenzen signifikant länger als für den Standardstimulus ($t_{(10)} = 3.49, p < 0.01$). Für die zweidimensionale Verletzung (PhSp) waren die deutlich kürzer ($t_{(10)} = 1.67, p < 0.01$) und besaßen außerdem eine deutlich geringere Varianz. Hier ergab nun auch die ANOVA für die drei Devianten einen signifikanten Haupteffekt DEVIANT ($F_{(1,10)} = 4.61, p < 0.05$). Wie Tabelle 9.2 zeigt, liegt dort kein Unterschied zeischen Sp und Ph vor. Der Deviant-Effekt ist vor allem auf Bedingung PhSp zurückzuführen, die eine kürzere Latenz und geringere Varianz aufweist (siehe Abbildung 9.4).

### 9.1.4 Diskussion

Wie schon im vorhergehenden Experiment, ist auch hier für alle Bedingungen (inklusive des Standards) eine Linkslateralisation der präattentiven auditorischen Verarbeitungskomponenten zu beobachten. Da sich zwischen Standard und De-

| *Bedingungen* | dF | t | p |
|---|---|---|---|
| Ph vs. Sp | 10 | 1.6 | n.s. |
| Ph vs. PhSp | 10 | 2.24 | $<0.05$ |
| Sp vs. PhSp | 10 | 2.35 | $<0.05$ |

Tabelle 9.2: Statistische Analyse der N1m-Latenzen (einseitiger t-Test, aufgrund der Mehrfach-Testung Bonferoni-korrigiert.)

viant keine signifikanten Unterschiede in dem frühen Zeitbereich ergeben, ist davon auszugehen, daß der Devianzeffekt bilateral verteilt ist. Für Phonemdevianz bei Einzelvokalen und Silben ist zwar auch häufig eine Linkslateralisation des Mismatch-Effekts beschrieben worden (Näätänen et al., 1997; Rinne, Alho, Alku, Holi, Sinkkonen, Virtanen et al., 1999; Kasai, Yamad, Kamio, Nakagome, Iwami, Fukuda et al., 2001, Shtyrov et al., 2000a, Shtyrov et al., 2000b, Alho, M., A., Huotilainen, Virtanen, Aulanko et al., 1998a), einige Studien berichten aber ebenfalls einen bilateralen MMN-Effekt bei Vokalen (Jaramillo, Ilvonen, Kujala, Alku, Tervaniemi & Alho, 2001) oder Silben (Mathiak, Hertrich, Lutzenberger & Ackermann, 1999, 2000).

Alle die Dipole der Devianzbedingungen lagen im Vergleich zum Standard anterior. Diese Überlagerung der N100m mit einer anterioren MMNm wird in der Literatur berichtet (Csepe et al., 1992) und auch in der EKP-Literatur erreicht die MMN-Differenzkurve eine maximale Amplitude an den frontalen bzw. fronto-zentralen Elektroden (Rinne et al., 2000). Die vielfach berichtete zusätzliche rechts-frontale Komponente der MMN, die mit einer unwillkürlichen Aufmerksamkeitsverlagerung in Verbindung gebracht wird (Näätänen et al., 2001; Escera, 2000), wurde hier allerdings nicht detektiert. Diese Subkomponente der MMN wurde in EKP-Experimenten untersucht (Rinne et al., 2000; Giard et al., 1990), konnten im MEG aber nicht beobachtet werden (Rinne et al., 2000). Es ist daher sehr wahrscheinlich, daß sie entweder in tieferen Hirnschichten generiert wird, oder radial ausgerichtet und daher für das MEG nicht erfaßbar ist.

Die Dipole der rechten Hemisphäre lagen im Vergleich zu denen der linken He-

misphäre anterior und lateral (Rinne et al., 2000). Dieser Befund wird ebenfalls in anderen Studien berichtet (Reite et al., 1995; Koyama et al., 2000; Nakasato et al., 1995). Vermutlich ist er auf allgemeine anatomische Unterschiede der Temporallappen beider Hemisphären zurückzuführen, wie sie in MRT-Studien aufgezeigt werden (Penhune, Zatorre, MacDonald & Evans, 1996; Good, Johnsrude, Ashburner, Henson, Friston & Frackowiak, 2001). Es waren keine Lokalisationsunterschiede zwischen den Deviant-Bedingungen zu beobachten. Dies kann nun zum einen bedeuten, daß für die frühe Wahrnehmung der Stimuli identische Neuronenpopulationen zuständig sind. Lokalisationsunterschiede werden für die MMNm in Reaktion auf unterschiedliche CV Silben von Diesch & Luce (1997) berichtet, waren in einer anderen Untersuchung (Alho et al., 1998a) aber nicht zu beobachten. Es ist jedoch möglich, daß die Dipole im hier vorgestellten Experiment so nahe beieinander lagen, daß Unterschiede durch den Konfidenzintervall von 0.95 nicht abgedeckt wurden. Bei der hier gegebenen Varianz und den methodisch bedingten Modellierungsbeschränkungen (Hari et al., 2000) können Unterschiede von weniger als 10 mm nicht statistisch erfaßt werden. Die in der Studie von Diesch & Luce (1997) untersuchten konsonantische Kontraste evozierten tatsächlich Lokalisationsunterschiede von weniger als 10 mm. Es ist aus dem hier vorliegenden Null-Effekt also nicht unbedingt zu schließen, daß für die Verarbeitung von Stimm- und Sprachlautinformation bzw. für die Detektion der jeweiligen Abweichung identische Neuronenpopulationen verantwortlich sind. Sollten jedoch unterschiedliche Hirnareale an diesen Prozessen beteiligt sein, beträgt ihr räumlicher Abstand weniger als 1cm und ist mit der hier verwendeten Methode nicht detektierbar bzw. statistisch erfaßbar. Interessant ist der Befund, daß sich die Dipolstärken für die Devianten untereinander nicht unterscheiden. Das bedeutet, daß sich die Reaktion auf die zweifache Abweichung nicht durch eine Addition der eindimensionalen Devianzen erklären läßt. Mismatch-Studien, die nichtsprachliche Experimentalreize verwenden, liefern diesbezüglich ein eher heterogenes Bild. Ein additives Amplitudenverhalten, das auf eine unabhängige Merkmalsverarbeitung schließen läßt, berichten Schröger (1995) für Frequenz und Präsentationsrichtung und Levänen et al. (1993) sowie Wolff & Schöger

(2001) für Tonhöhe und Dauer. Für die Parameter Frequenz und Dauer finden Paavilainen et al. (2001b) ein additives Verhalten der MMN-Amplitude, Wolff & Schöger (2001) hingegen eine Subadditivität. Eine MMN-Amplitude, die hingegen deutlich geringer ist als die Summe der Einzelbedingungen wird für dreidimensionale Abweichungen gefunden (Paavilainen et al., 2001b für Frequenz, Intensität und SOA[4]; Wolff & Schöger, 2001 für Frequenz, Dauer und Intensität). In der vorliegenden Studie ist die Dipolstärke der Bedingung PhSp durch die Summe der Dipolstärken von Bedingung Ph und Sp nicht zu erklären. Dieser Befund spricht bereits gegen eine unabhängige präattentive Verarbeitung der akustischen Parameter einer phonetischen und einer sprecherbezogenen Abweichung. Unterstützende Evidenz kommt zudem von der Latenzanalyse. Auffällig sind insbesondere die kurzen Mittelwerte, sowie die geringere Varianz der N100/MMN-Latenzen für die zweidimensionale Devianz. Offenbar bedeutet die gleichzeitige Abweichung von Wort und Stimme für den Mismatch-Detektor eine Redundanz, die zu einer schnelleren Reaktion Abweichung führt.

Dieser Effekt ist durchaus mit dem redundanzbedingten Gewinn in den entsprechenden Reaktionszeitstudien vergleichbar. Im Speeded-Classification Paradigma (Garner, 1974), das auch eine (bewußte) Abweichungsdetektion ist, kommt es bei einer festen Kombination von Sprechstimme und phonetischem Inhalt ebenfalls zu kürzeren Reaktionszeiten und einer Reduktion der Varianz (Mullenix & Pisoni, 1990). Aus diesen Befunden wurde geschlossen, daß die Verarbeitung phonetischer und sprecherbezogener Information integrativ, d.h. nicht-unabhängig verläuft. Dies wird durch die vorliegende MEG-Studie für die präattentive Wahrnehmung bestätigt.

Eine Asymmetrie, die auf eine schnellere Informationsverarbeitung für die Stimmeninformation spricht, wurde in den Reaktionszeitexperimenten (Mullenix & Pisoni, 1990; Wood, 1974), nicht aber in den elektrophsiologischen Daten gefunden. Ein Grund für diese Diskrepanz könnte in der unterschiedlichen Salienz des jeweiligen Kontrastes liegen: Der Stimmenwechsel könnte im Reaktionzeit-

[4]SOA: Stimulus Onset Asynchrony bezeichnet den Zeitintervall zwischen dem jeweiligen Beginn zweier aufeinanderfolgender Experimentalreize (vgl. Boff et al., 1986)

experiment nur deswegen schneller verarbeitet werden und die Wortverarbeitung scheinbar abhängig davon sein, weil er im Vergleich zum phonetischen Kontrast den akustisch stärkeren Reiz darstellt (siehe Seite 46). Die Daten der vorliegenden MEG-Studie stimmen mit dieser Hypothese überein. Hier umfaßte die Reizdevianz neben einem Unterschied in der Stimmhaftigkeit des wortinitialen Konsonanten auch ein Vokalkontrast und einen Unterschied in der Artikulationsstelle des nachfolgenden Frikativs (*Test* vs. *Dach*). Bei einer akustisch so salienten Abweichung sollte man ebenfalls keinen Vorteil für die Verarbeitung der Stimminformation gegenüber diesem phonetischen Kontrast erwarten. Die elektrophysiologischen Korrelate lassen in der Tat keine generell frühere Verarbeitung für die Stimminformation vermuten. Vielmehr weist Experiment II auf eine integrative Verarbeitung von Stimminformation und phonetischen Cues hin.

Die bisherigen Experimente I und II zeigen, daß die Detektion eines Sprecherwechsels auf einer frühen Ebene des Sprachwahrnehmungsprozesses stattfindet. Grundlage dieser Experimente war das passives Oddball-Paradigma. Dieses ist hervorragend geeignet, um die präattentive auditorischen Verarbeitung zu untersuchen und wurde hier eingesetzt, um Aussagen über die Charakteristika der Stimmverarbeitung, sowie ihrer Interaktion mit der *strukturellen*, d.h. phonetischen Verarbeitung zu treffen. Im Vergleich zur natürlichen Sprachverarbeitung stellt die iterative Einzelwortpräsentation des Oddball-Paradigmas allerdings eine starke Abstraktion dar. Daher wurde in den nachfolgenden Experimenten, die der *inhaltlichen* Verarbeitung von Lautsprache gewindmet waren, ein Paradigma verwendet, das sich stärker an den alltäglichen Prozessen der Sprachwahrnehmung orientiert. Indem der Einfluß der Stimminformation auf das semantische Satzverstehen untersucht wurde, sollte im dritten und letzten Abschnitt des empirischen Teils der Frage nachgegangen weden, mit welchen Kosten die Verarbeitung der Sprecherinformation verbunden ist.

# Kapitel 10

# EXPERIMENTALSERIE III: Stimminformation und semantische Verarbeitung

## 10.1 Allgemeine Fragestellung

Die folgende Experimentalserie untersuchte das Zusammenspiel von Stimminformation und semantischer Bedeutung beim Sprachverstehen. Ist ein wechselseitiger Einfluß von Stimmenverarbeitung und phonetischer Dekodierung wahrscheinlich, da letztere direkt von der Sprechervariabilität betroffen ist, so handelt es sich beim Verstehen des semantischen Inhalts jedoch um einen Prozeß, der der höheren und zeitlich späteren Verarbeitungsprozessen zugeordnet wird. In den Experimenten IIIa-c sollte nun getestet werden, ob der Einfluß abweichender Sprecherinformation das inhaltliche Verstehen einer Äußerung erschwert, oder ob die Normalisierungsprozesse automatisiert sind, so daß auf Ebene der semantischen Verarbeitung keine Effekte zu beobachten sind.

Prinzipiell wäre es möglich, diese Annahmen auf der Ebene von Einzelwörtern (d.h. z.B. der Präsentation von Wortpaaren) zu testen. In der folgenden Experi-

mentalserie wurde der Zusammenhang zwischen semantischem Verstehens und der Verarbeitung der Sprecherinformation aus zwei Gründen auf Satzebene untersucht, wobei erste dieser Gründe methodischer, der zweite inhaltlicher Art ist: Man benötigt für eine semantisch relatierte Experimentalaufgabe hinreichend viele verschiedene Stimuli, um Wiederholungseffekte zu vermeiden. Um ein günstiges Signal-Rausch-Verhältnis zu erhalten, verlangen neurophysiologischen Untersuchungsmethoden zudem eine ausreichend hohe Anzahl von Versuchsdurchgängen. Es ist aber aufgrund der intra-individuellen Varianz schwierig, eine genügend hohe Anzahl an unterschiedlichen Einzelwortaufnahmen eines Sprechers zu erstellen, die tatsächlich auch als Äußerungen dieses einen Sprechers identifizierbar sind, da die Präsentationszeit bei Einzelwörtern sehr kurz ist, der Hörer eine Sprecherrepräsentation also anhand weniger Informationen aufbauen muß.[1]
Der zweite Grund für die Verwendung von Sätzen als Experimentalreize liegt darin, daß mit den folgenden Experimenten nicht nur präattentive auditorische Prozesse untersucht werden sollten, sondern die bewußte semantische Verarbeitung. Die Stimuli sollten daher an die natürliche Sprachsituation angepaßt werden, weshalb hier nun keine Wortpaare, sondern Sätze präsentiert wurden.
Es wurde von folgenden Hypothesen ausgegangen. Durch die auditorische Präsentation von Teilsätzen, wie z.B. *"Der Vegetarier hat das frische Obst..."* werden beim Hörer aufgrund seiner Kommunikationserfahrung implizit mindestens folgende zwei Annahmen getroffen. Erstens wird angenommen, daß der Satz sinnvoll, d.h. semantisch kongruent beendet wird. Zweitens wird erwartet, daß derjenige Sprecher, der den Satzanfang geäußert hat, auch den Satz vervollständigt. Verletzungen dieser Annahmen würden also eine semantische Inkongruenz oder einen Mismatch der Sprecherrepräsentation und ggf. beides zugleich bedeuten.
Die Verarbeitung dieser qualitativ unterschiedlichen Verletzungen sollte nun anhand einer Reaktionszeitstudie und zweier EKP-Experimente untersucht werden.

[1] Ein Ausweg wäre natürlich die Beschränkung auf den den grundlegenden Unterschied männlicher/weiblicher Stimmen. In der vorliegenden Studie sollte aber auch zwischen Sprechern gleichen und unterschiedlichen Geschlechts unterschieden werden, was das Problem der Identifikation/Diskrimination natürlich verschärft. Alternativ könnte die Präsentation sehr außergewöhnlicher Stimmen eine solche Aufgabe erleichtern. Wegen der damit verbundenen Besonderheiten, wie sie z.B. in Experiment Ia dargestellt wurden, wurde davon jedoch abgesehen.

Da in diesen drei Experimenten dasselbe Satzmaterial verwendet wurde und da bei der Erstellung der Stimuli einige Besonderheiten zu berücksichtigen sind, erfolgt nun zunächst eine Beschreibung des Versuchsmaterials.

## 10.2 Versuchsmaterial

Die Experimentalreize bestanden aus einzelnen Sätzen, die semantisch kongruent oder inkongruent waren. Die semantisch kongruenten Sätze besaßen die Struktur *"Der Vegetarier hat das frische Obst gegessen."; "Die Hausfrau hat das weiße Hemd gebügelt."* Die semantisch inkongruenten Versionen der Sätze wurden durch Austausch von jeweils zwei der Partizipien am Satzende erstellt (*"Der Vegetarier hat das frische Obst gebügelt."; "Die Hausfrau hat das weiße Hemd gegessen."*). Wie in einem Vortest mit 35 Testpersonen ermittelt wurde, hatten die semantisch korrekten Sätze eine CLOZE-Wahrscheinlichkeit [2] von 0.68, die semantisch inkorrekten eine CLOZE-Wahrscheinlichkeit von 0. Neben der semantischen Kongruenz wurde als zweite unabhängige Variable die Sprecherkontinuität in drei Stufen manipuliert: In Bedingung **Std** äußert ein einziger (Standard-)Sprecher den kompletten Satz. In Bedingung **M** produziert dieser Standardsprecher den Satz bis zum vorletzten Wort, das Partizip stammt von einem anderen männlichen Sprecher. In Bedingung **W** produziert der Standardsprecher den Satz bis zum vorletzten Wort, das Partizip wurde von einer weiblichen Stimme gesprochen. Eine Übersicht aller Bedingungen findet sich in Tabelle 10.1 (die Sätze sind außerdem in Appendix C aufgeführt).

[2] (siehe Seite 4.5 dieser Arbeit, sowie Taylor, 1953)

| | | SEMANTIK | |
|---|---|---|---|
| | | Match | Mismatch |
| SPRECHER | Std | Der Vegetarier hat das frische Obst gegessen. | Der Vegetarier hat das frische Obst gebügelt. |
| | M | Der Vegetarier hat das frische Obst gegessen. | Der Vegetarier hat das frische Obst gebügelt. |
| | W | Der Vegetarier hat das frische Obst gegessen. | Der Vegetarier hat das frische Obst gebügelt. |

Tabelle 10.1: Übersicht der sechs Experimentalbedingungen. Der Sprecherwechsel ist durch farbige Schrift markiert (schwarz: Standardsprecher; blau: männliche Stimme; rot: weibliche Stimme)

Abbildung 10.1 illustriert den Grundfrequenzverlauf anhand eines Beispielsatzes in den Bedingungen Std, M und W. Es ist ersichtlich, dass die F0-Werte der Bedingungen Std und M, d.h. der beiden Männerstimmen, sich wenig unterscheiden, während die weibliche Stimme eine deutlichere Abweichung darstellt. Ebenso ist das Stimmtimbrenbei Sprechern desselben Geschlechts in der Regel ähnlicher.

Die Verwendung kontinuierlicher Sprache in EKP-Studien geht mit besonderen

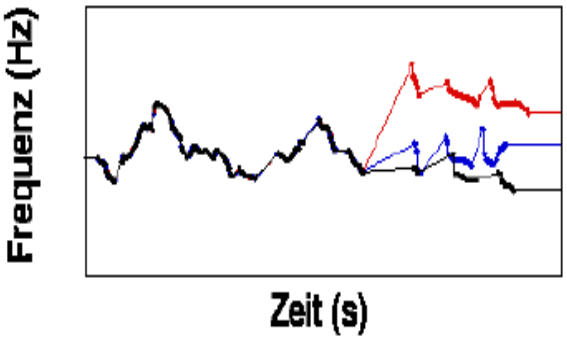



Abbildung 10.1: Grundfrequenzverläufe aller drei Sprecherbedingungen des Satzes *"Die Buben haben den kalten Saft getrunken."* *(schwarz: Standardsprecher; blau: männliche Stimme; rot: weibliche Stimme)*

Anforderungen einher, insbesondere im Hinblick auf die Koartikulation und den prosodischen Verlauf der Sätze (O'Halloran, Isenhart, Sandmann & Larkey, 1988; Holcomb & Neville, 1991; Hagoort & Brown, 2000). Daher wurde bei der Erstel-

lung des Versuchsmaterials auf die Einhaltung bestimmter Kriterien geachtet:

1.) Prosodie und Koartikulation: Die Experimentalstimuli wurden von den Sprechern als vollständige Sätze gelesen.[3]

2.) Artikulatorische Plausibilität: Die Sätze waren derart konstruiert, daß das vorletzte Wort im Satz auf einen alveolaren oder dentalen Plosiv endete (in wenigen Fällen auf einen Frikativ). Durch die Artikulationsbewegung der Zunge von diesem Laut zum nachfolgenden initialen /g/ des Partizips entstand eine artikulatorisch bedingte Pause. Diese wurde als Schnittpunkt bei der Erstellung der Bedingungen M und W genutzt.

3.) Die Lautstärke der Stimuli wurde normalisiert.

4.) Kontrolle der Sprechgeschwindigkeit: Die akustische Analyse ergab folgende mittlere Dauer der semantisch korrekten Sätze: Std 2600 ms, M 2650 ms, F 2650 ms; und der semantisch inkorrekten Sätze: Std 2590 ms, M 2620 ms, W 2620 ms. Sind die Mittelwerte der Standardbedingung auch etwas niedriger, so ergab die statistische Analyse (t-Test) doch keinen signifikanten Unterschied zwischen den Bedingungen. Zusätzlich wurde die Dauer der ersten Silbe der satzfinalen Partizipien (/gə/) gemessen. Sie betrug im Durchschnitt 74 ms und unterschied sich nicht signifikant zwischen den Bedingungen (für die Silben der kongruenten (k) und inkongruenten (i) Sätze: Std 74ms (k)/69ms (i); M 81ms (k)/74 ms (i); W 78 ms (k)/73 ms (i); die mittlere Differenz zwischen den Sprecherbedingungen lag bei weniger als 7 ms).

[3] Damit kann zwar prinzipiell nicht ausgeschlossen werden, daß schon während des Satzproduktion Hinweise auf die semantische Kongruenz des finalen Wortes gegeben werden, jedoch liefert die auditive Beurteilung keinen Anhaltspunkt dafür. Da zudem die Trägersätze (d.h. der jeweilige Satz bis zum vorletzten Wort) für die drei Sprecherbedingungen jeweils identisch war, ist zumindest zwischen den Sprecherbedingungen kein Unterschied zu erwarten.

## 10.3 Experiment III a: Reaktionszeitstudie - Sprecherdiskrimination und semantische Verarbeitung

### 10.3.1 Fragestellung

Zunächst sollte auf Verhaltensebene untersucht werden, wie ein unerwarteter Sprecherwechsel in Kombination mit einer semantischen Verletzung im Satzkontext verarbeitet wird. Dabei ging es vornehmlich um zwei Fragen:
1. Zeigt sich in den Verhaltensdaten eine Interaktion der Verabeitung semantischer und sprecherbezogener Information?
2. Spiegeln die Verhaltensdaten bei der Sprecherunterscheidung und/oder bei der semantischen Beurteilung den größeren physikalischen Unterschied zwischen der männlichen und der weiblichen Stimme (Std, W) im Vergleich zu den beiden männlichen Stimmen (Std, M) wider?
Zur Beantwortung dieser Fragen wurden zwei Versuchspersonengruppen gebildet, die entweder eine Sprecherunterscheidung oder eine semantische Beurteilung durchzuführen hatten.

Für die Sprecherdiskrimination wurden in den Verhaltensdaten deutliche Unterschiede in Abhängigkeit von der Aufgabenschwierigkeit erwartet: Für Bedingung W sollte sich eine geringe Fehlerzahl und eine niedrige Reaktionszeit ergeben, da die Frauenstimme im Hinblick auf ihre physikalischen und perzeptuellen Parameter leicht von der Männerstimme (Std) zu unterscheiden ist. Die Diskrimination der beiden Männerstimmen (M und Std) sollte hingegen schwieriger sein, und daher mit einer größeren Fehlerzahl und längeren Reaktionszeiten einhergehen. Interessant ist die Frage nach der wechselseitigen Interferenz der Semantik und Sprecherunterscheidung. Hier ergeben sich zwei Hypothesen. Zum einen werden im vorhandenen Material die akustischen Hinweise auf einen Sprecherwechsel im Signal bereits mit der ersten Partizipsilbe geliefert, d.h. etwa 50-60 ms bevor die lexikalisch-semantische Verarbeitung beginnen kann, man würde daher eine Detektion des Sprecherwechsels erwarten, die unabhängig von der semantischen

Kongruenz des jeweiligen Satzes verläuft. Andererseits ist davon auszugehen, daß die Sprechernormalisierung ein unbewußter Prozeß ist. Eine aufgabenbedingte Bewußtmachung, bedeutet möglicherweise einen zeitliche Ausdehnung der Verarbeitungsprozesse, die dann parallel zum inhaltlich-semantischen Satzverstehen stattfinden. Sollte sich ein Einfluß der Semantik zeigen, so würde dies bedeuten, daß die bewußte Sprecherdiskrimination nicht so stark vom Signal abstrahiert, daß die Bedeutung des Satzes vernachlässigt wird.

Für die semantische Beurteilung werden hinsichtlich der Fehlerraten keine Unterschiede zwischen den Sprecherbedingungen erwartet, da diese alle dieselben Sätze umfassen. Im Hinblick auf die Reaktionszeiten sind Vorteile für semantisch kongruente Sätze zu erwarten, deren letztes Wort durch den vorangegangenen Satzkontext geprimt ist, längere Reaktionszeiten hingegen für die semantisch inkongruenten Sätze (siehe z.B. Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980; Holcomb & Neville, 1990). Ein solcher Trend wurde hier zumindest auch für die Standardbedingung (kein Sprecherwechsel) angenommen. Es stellt sich aber die Frage, wie sich ein Wechsel der Stimme auf die semantische Beurteilung der Sätze auswirkt. Da davon auszugehen ist, daß die Verarbeitung/Normalisierung von Stimminformation einer semantischen Integration zeitlich vorgelagert, aber auch ein sehr schneller Prozeß ist, sollte man für die Bedingungen M und W gleiche oder etwas längere Reaktionszeiten als für den Standardsprecher erwarten. Liegt jedoch ähnlich wie für die phonetische Dekodierung eine Interaktion mit der Stimmenverarbeitung vor, so sind kürzere Reaktionszeiten für die Detektion der semantischen Inkongruenz zu erwarten, wenn diese mit einem zusätzlichen Sprecherwechsel verbunden ist; im Gegenzug wären aufgrund eines Antwortkonflikts längere Reaktionszeiten auf die semantisch kongruenten Sätze der Bedingungen M und W zu erwarten ("Nein"-Antwort auf semantischen Match).

### 10.3.2 Methode

#### 10.3.2.1 Probanden

32 Versuchspersonen (16 männlich) im Alter von 19 - 30 Jahren (Durchschnitt: 23 Jahre) nahmen am Experiment teil. Alle Probanden waren deutsche Muttersprachler und nach dem Edinburgh Inventory (Oldfield, 1971) als Rechtshänder klassifiziert. Sie wurden für ihre Teilnahme am Experiment bezahlt. Die Probanden wurden nach dem Zufallsprinzip in zwei Gruppen geteilt, wobei jede der Gruppen acht weibliche und acht männliche Probanden umfaßte. Wie unten beschrieben, unterschieden sich die Probandengruppen hinsichtlich der experimentellen Aufgabe (Gruppe1: Sprecherunterscheidung; Gruppe2: semantische Beurteilung).

#### 10.3.2.2 Versuchsablauf

Die Probanden saßen in einer schallarmen Kabine. Vor Ihnen befand sich ein Monitor in etwa einem Meter Abstand, links und rechts davon Lautsprecher vom Typ JBL XL 300. Die Stimuli wurden vom PC über einen Vorverstärker (Digilink III., Conrad Audio) und über die Lautsprecher dargeboten. Die Experimentalaufgabe der jeweiligen Versuchspersonengruppe war während des gesamten Experiments in Form einer kurzen Frage auf dem Bildschirm zu sehen.
**Gruppe1:** “Derselbe Sprecher?“
**Gruppe2:** “Sinnvoll?“
Die Probanden hielten eine Tastatur mit zwei Drucktasten in den Händen. Der linken Taste war jeweils die Ja-Antwort zugeordnet, der rechten Taste die Nein-Antwort. Die Probanden waren instruiert, bei jedem Versuchsdurchgang möglichst schnell per Tastendruck zu antworten. Die Reaktionszeitmessung startete jeweils mit dem Beginn des letzten Wortes. Um eine Ausbalancierung der Ja- und Nein-Antworten zu erreichen, erfolgte die Verteilung der Bedingungen, wie sie in Tabelle 10.2 dargestellt ist. Die Blockreihen wurde über die Versuchsteilnehmer hinweg rotiert. Die Reihenfolge der Bedingungen innerhalb der Blöcke wurde randomisiert.

| | | SEMANTIK | | Block |
|---|---|---|---|---|
| | | kongruent | inkongruent | |
| **SPRECHER** | Std1 | N=30 | N=30 | Block 1 |
| | M | N=30 | N=30 | |
| **SPRECHER** | Std2 | N=30 | N=30 | Block 2 |
| | W | N=30 | N=30 | |

Tabelle 10.2: Experiment IIIa: Stimulusverteilung in den beiden Experimentalblöcken. Die Tabelle gilt für beide Gruppen/Aufgabenstellungen. Std1 und Std2 bezeichnet dieselben Sätze (ausbalanciert über die Versuchspersonen) die entweder mit Bedingung M (Block1) oder Bedingung W (Block2) präsentiert wurden.

### 10.3.3 Ergebnisse

Es gab bei der Messung der Reaktionszeiten eine formaler Antwortintervall von zehn Sekunden nach Satzbeginn, innerhalb dessen eine Antwort erfolgen sollte. Die tatsächlichen Antwortzeiten waren jedoch deutlich kürzer als dieser Maximalintervall. Es wurde post hoc eine statistische Ausreißerkorrektur durchgeführt: Nur diejenigen Reaktionszeitwerte gingen in die Analyse ein, die nicht mehr als die zweifache Standardabweichung vom Mittelwert der Reaktionszeiten (pro Bedingung und Proband) entfernt lagen. Appendix C.1 zeigt die Mittelwerte und Standardabweichungen der Antwortquoten und Reaktionszeiten beider Gruppen. Die Ergebnisse für der Gruppen und Blöcke wurden getrennt analysiert. Da davon auszugehen ist, daß der Standardsprecher deutlich leichter von der weiblichen Stimme zu diskriminieren ist als von der männlichen, werden diese Reaktionen getrennt nach Blöcken ausgewertet (daher die Bezeichnungen Std1 und Std2; siehe Tabelle 10.2).

**Sprecherunterscheidung (Gruppe1)**

**Antwortquote**

Die Antwortquote wurde als relative Häufigkeit der richtigen Antworten (% korrekt) festgelegt. Appendix C.1 zeigt die relativen Häufigkeiten der korrekten Antworten. Es war eine hohe Antwortquote für die Bedingungen Std1, W und Std2 zu beobachten. Mehr Fehler wurden in Bedingung M gemacht, und hier insbesondere bei den kongruenten Sätzen. Auch die Standardabweichung ist in dieser Bedingung deutlich erhöht. Die zweifache Varianzanalyse (ANOVA) mit den Faktoren SEMANTIK (kongruent vs inkongruent) $\times$ SPRECHER (Std1, M, Std2, W) zeigte, daß diese Besonderheit der kongruenten Sätze der Sprecherbedingung M statistisch signifikant waren: Es trat ein Haupteffekt SPRECHER auf ($F_{(3,45)} = 3.82, p < 0.05$), kein signifikanter Haupteffekt SEMANTIK, jedoch eine signifikante Interaktion der Faktoren ($F_{(3,45)} = 3.33$ , $p < 0.05$).

**Reaktionszeiten**

Die Reaktionszeiten bei der Sprecherunterscheidung sind in Abbildung 10.2 dargestellt (für die genauen Werte siehe Appendix C.1). Die Daten zeigen eine deutlich längere Antwortzeit für die Unterscheidung der beiden männlichen Sprecher als für die Unterscheidung zwischen Standardsprecher und weiblicher Stimme. Außerdem traten in den Bedingungen Std1, W und Std2 längere Reaktionszeiten für die inkongruenten Sätze auf. Bei Bedingung M war ein umgekehrter Effekt zu beobachten. Für die korrekten Sätze der Bedingung M war auch die Standardabweichung deutlich erhöht. Eine Überprüfung der individuellen Probandendaten schließt einen Speed-Accuracy-Tradeoff aber aus, d.h. kürzere Reaktionszeiten gingen nicht mit einer Zunahme an Fehlern einher.

In Tabelle 10.3 sind die Ergebnisse der statistischen Analyse aufgeführt. Die zweifaktorielle Varianzanalyse SPRECHER $\times$ SEMANTIK zeigte für die Unterscheidung der beiden männlichen Stimmen einen signifikanten Haupteffekt SPRECHER und eine signifikante Interaktion beider Faktoren, die auf den gegensätzlichen SEMANTIK Effekt in den beiden Bedingungen zurückzuführen ist. Für

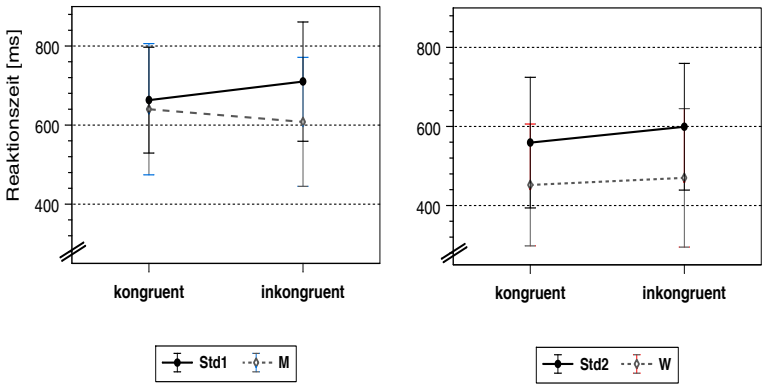



Abbildung 10.2: *Sprecherunterscheidung:* Reaktionszeiten in den beiden Experimentalblöcken für die Sprecherunterscheidung bei semantischer Kongruenz und Inkongruenz

die Unterscheidung von Standardsprecher und weiblicher Stimme beschreibt der SPRECHER Haupteffekt, daß die Reaktion auf die weibliche Stimme signifikant schneller erfolgte. Außerdem waren in beiden die Reaktionszeiten auf die kongruenten Sätze signifikant kürzer (Haupteffekt SEMANTIK). Da die Reaktion auf die inkongruenten Sätze in Bedingung W etwas schneller erfolgte, war ein Trend zu einer Interaktion der Faktoren zu beobachten. In den bisherigen Analysen wurden immer “Ja“- und “Nein“-Antworten verglichen. Ein Vergleich der beiden “Nein“-Antwortbedingungen W und M zeigt, daß wie erwartet, die weibliche Stimme signifikant schneller detektiert wurde, als der Wechsel zwischen den ähnlicheren männlichen Stimmen. Der umgekehrte SEMANTIK-Effekt in Bedingung M führte auch hier zu einer signifikanten Interaktion der Faktoren SPRECHER und SEMANTIK.

Da die Interaktionen in allen Bedingungen entweder signifikant sind, oder zumindest ein starker Trend besteht, wurden die SEMANTIK-Effekte separat für die einzelnen Sprecherbedingungen untersucht. Wie Tabelle 10.4 zeigt, liegt ein deutlicher SEMANTIK-Effekt in beiden Standardsprecher-Bedingungen (Std1, Std2) vor, in Bedingung W ein starker Trend. Der Unterschied zwischen semantisch kongruenten und inkongruenten Sätzen ist in Bedingung M ebenfalls signifikant, allerdings ist dort die Antwortzeit für die korrekten Sätze signifikant länger. Der

| Bedingung | *FAKTOREN* | *df* | *F* | *p* | |
|---|---|---|---|---|---|
| | SEMANTIK | 1,15 | 1.53 | n.s. | |
| Std1, M | SPRECHER | 1,15 | 10.81 | $< 0.001$ | *** |
| | SEMANTIK $\times$ SPRECHER | 1,15 | 14.75 | $< 0.01$ | ** |
| | SEMANTIK | 1,15 | 12.36 | $<0.01$ | ** |
| Std2, W | SPRECHER | 1,15 | 49.75 | $< 0.0001$ | *** |
| | SEMANTIK $\times$ SPRECHER | 1,15 | 3.71 | 0.0731 | |
| | SEMANTIK | 1,15 | $<1$ | n.s. | |
| M, W | SPRECHER | 1,15 | 17.7 | $< 0.001$ | *** |
| | SEMANTIK $\times$ SPRECHER | 1,15 | 13.47 | 0.01 | ** |

Tabelle 10.3: *Sprecherunterscheidung:* Analyse der Reaktionszeiten

| *K vs. I* | *df* | *F* | *p* | |
|---|---|---|---|---|
| Std1 | 1,15 | 17.82 | $< 0.001$ | *** |
| M | 1,15 | 6.31 | $< 0.05$ | * ! |
| Std2 | 1,15 | 12.67 | $< 0.01$ | ** |
| W | 1,15 | 4.30 | 0.056 | |

Tabelle 10.4: *Sprecherunterscheidung:* Unterschiede zwischen semantischer Kongruenz und Inkongruenz in den einzelnen Sprecherbedingungen.(Beachte: In Bedingung M ist die Reaktionszeit bei semantischer Kongruenz länger!)

SEMANTIK-Effekt, d.h. die Reaktionszeitdifferenz (inkongruente Sätze - kongruente Sätze) war in den Fällen, in denen ein Sprecherwechsel stattfand, stets schwächer als in den zugehörigen Standardsprecherbedingungen. Die Bedingungen Std1 und Std2 unterschieden sich hierbei nicht signifikant ($F_{(1,15)} < 1$). Zur Illustration der SEMANTIK-Effekte siehe Abbildung 10.3.

Die beschriebenen Daten zeigen deutliche Unterschiede zwischen den Sprecherbedingungen, sowie einen Einfluß der semantischen Kongruenz. Bevor die Effekte jedoch im einzelnen diskutiert werden, sollen zunächst die Ergebnisse für die zweite Probandengruppe dargestellt werden.

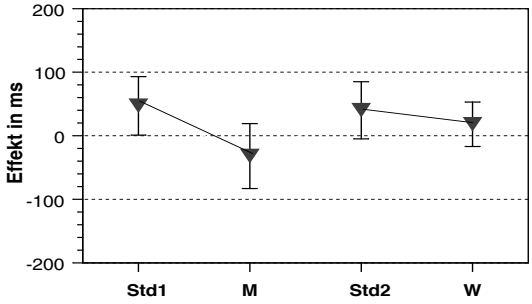



Abbildung 10.3: *Sprecherunterscheidung:* Reaktionszeitdifferenzen für semantische Inkongruenz-semantische Kongruenz aller drei Sprecherbedingungen

**Semantische Beurteilung (Gruppe2)**

**Antwortquote**

Wie Appendix C.1 zeigt, lagen die Antwortquoten hier alle über 90 %. Die Varianzanalyse (ANOVA) mit den Faktoren SPRECHER (Std1, M, Std2, W) $\times$ SEMANTIK zeigt für die Semantik keine signifikante Haupteffekte oder Interaktionen. Es trat allerdings ein Haupteffekt SPRECHER auf ($F_{(3,45)} = 4.77$ , $p < 0.01$), der auf die Bedingung Std2 zurückzuführen ist. Wurden in den anderen Bedingungen durchschnittlich etwa zwei von 30 Antworten falsch beantwortet, so wurden in Std2 durchschnittlich drei Versuchsdurchgänge falsch beantwortet.

**Reaktionszeiten**

Die Reaktionszeiten für die semantische Beurteilung sind in in Abbildung 10.4 dargestellt (für die Absolutwerte siehe Appendix C.1 auf Seite 170).

Es traten hier in allen Sprecherbedingungen längere Reaktionszeiten für die semantisch inkorrekten Sätze auf (für die Absolutwerte siehe Appendix C.1). Allerdings war dieser Semantik-Effekt in den Bedingungen unterschiedlich stark. Wie die Tabelle 10.5 zeigt, ergab die ANOVA mit den Faktoren SPRECHER $\times$ SEMANTIK für die Unterscheidung der beiden Sprecher (Std1, M) keinen Haupteffekt SPRECHER, jedoch sowohl einen Haupteffekt SEMANTIK als auch eine In-

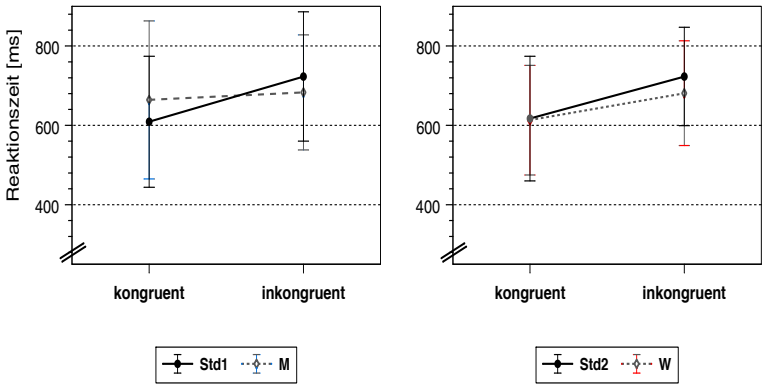



Abbildung 10.4: *Semantische Beurteilung:* Reaktionszeiten für die Sprecherbedingungen bei semantischer Kongruenz und Inkongruenz

teraktion SPRECHER × SEMANTIK. Diese ist auf den schwächeren SEMANTIK-Effekt für Bedingung M zurückzuführen, der sowohl auf die schnellere Antwort bei den inkongruenten Sätzen, als auch auf die verzögerte Reaktionszeit bei kongruenten Sätzen zurückzuführen ist. Der entsprechende statistische Vergleich der Bedingungen Std2 und W zeigte ebenfalls einen SEMANTIK-Haupteffekt und eine signifikante Interaktion der Faktoren. Diese beruht hier nun auf der im Vergleich zum Std kürzeren Reaktionszeit bei den inkongruenten Sätzen der Bedingung W.

Die Analyse der Reaktionszeiten für inkongruente und kongruente Sätze unterscheiden sich signifikant für alle Bedingungen außer M (Tabelle 10.6). Wie die Abbildung 10.5 zeigt ist der SEMANTIK-Effekt in Standardbedingungen besonders deutlich.

### 10.3.4 Diskussion

Zwei Probandengruppen hatten unterschiedliche Reaktionszeitaufgaben mit identischem Stimulusmaterial durchzuführen. Probandengruppe 1 sollte einen Sprecherwechsel auf dem satzfinalen Wort detektieren. Wie erwartet zeigen die Verhaltensdaten, daß die Unterscheidung von männlicher und weiblicher Stimme leichter ist, als zwischen den beiden männlichen Stimmen. Die deutlich höher-

| Bedingung | *FAKTOREN* | *df* | *F* | *p* | |
|---|---|---|---|---|---|
| | SEMANTIK | 1,15 | 30.04 | $< 0.0001$ | *** |
| Std1, M | SPRECHER | 1,15 | $<1$ | n.s. | |
| | SEMANTIK $\times$ SPRECHER | 1,15 | 13.27 | $< 0.01$ | ** |
| | SEMANTIK | 1,15 | 44.13 | $<0.0001$ | *** |
| Std2, W | SPRECHER | 1,15 | 2.48 | n.s. | |
| | SEMANTIK $\times$ SPRECHER | 1,15 | 7.45 | $<0.05$ | * |
| | SEMANTIK | 1,15 | 1.05 | n.s. | |
| M, W | SPRECHER | 1,15 | 10.57 | $< 0.01$ | ** |
| | SEMANTIK $\times$ SPRECHER | 1,15 | 5.40 | $<0.05$ | * |

Tabelle 10.5: *Semantische Beurteilung:* Analyse der Reaktionszeiten

| *K vs. I* | *df* | *F* | *p* | |
|---|---|---|---|---|
| Std1 | 1,15 | 58.28 | $< 0.0001$ | *** |
| M | 1,15 | $< 1$ | n.s. | |
| Std2 | 1,15 | 44.39 | $< 0.0001$ | *** |
| W | 1,15 | 25.99 | $< 0.0001$ | *** |

Tabelle 10.6: *Semantische Beurteilung:* Unterschiede zwischen semantischer Kongruenz und Inkongruenz in den einzelnen Sprecherbedingungen.(Beachte: In Bedingung M ist die Reaktionszeit bei semantischer Kongruenz länger!)

en Grundfrequenz und das geschlechtsspezifische Timbre der weiblichen Stimme führt nicht nur zu einer höheren Erkennungsrate, sondern auch zu deutlich niedrigeren Reaktionszeiten bei der Diskrimination. Ebenfalls war ein Einfluß der semantischen Information auf die Sprecherunterscheidung zu beobachten. Wie aus zahlreichen Studien bekannt (siehe Holcomb & Neville, 1990), werden die Entscheidungen für die semantisch kongruente Sätze signifikant schneller getroffen, als für semantisch inkongruente. Obwohl nun für die erste Experimentalgruppe die semantische Kongruenz nicht im Fokus der Aufgabenstellung stand, waren die Reaktionszeiten kürzer, wenn die Sätze semantisch kongruent waren,

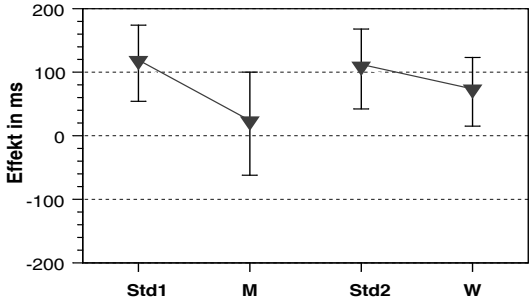



Abbildung 10.5: *Semantische Beurteilung:* Reaktionszeitdifferenzen für semantische Inkongruenz-Kongruenz aller drei Sprecherbedingungen

als wenn sie eine semantische Inkongruenz aufwiesen. Dies galt in zwei Fällen:
1. Wenn der Sprecher konstant blieb (Std-Bedingung), wurden der stärkste Semantik-Einfluß d.h. die größten Unterschiede zwischen semantisch kongruenten und inkongruenten Sätzen beobachtet. Offenbar waren hier die Bedingungen für eine automatische semantische Verarbeitung am günstigsten. Zu den Effekten mag auch beigetragen haben, daß in dieser Bedingung auf alle Sätze mit "Ja" (gleicher Sprecher) zu antworten war. Diese Antwortvorgabe könnte im Falle der semantischen Inkongruenz zu einer zusätzlichen Verzögerung geführt haben.
2. Im Falle des Wechsels von einer männlichen zu einer weiblichen Stimme trat ebenfalls ein Semantik-Effekt auf. Dieser war jedoch schwächer, was dadurch zu erklären ist, daß die Doppelverletzung (semantische Inkongruenz und Sprecherwechsel) die zu fällende Nein-Entscheidung beschleunigt. Ein Antwortkonflikt führte möglicherweise zu einer Verzögerung der Reaktion bei den kongruenten Sätzen.
Für Bedingung M war der Semantik-Effekt schwächer als für Bedingung W. Dieser Unterschied ist aber weniger auf einen Unterschied für inkongruenten Sätze zurückzuführen, als vielmehr auf eine Reaktionszeitverzögerung für die kongruenten Sätze der Bedingung M. Als Erklärung wäre denkbar, daß hier der Antwortkonflikt stärker wirkt, weil die Sprecherunterscheidung schwierig ist, daher länger dauert und die späteren Semantikprozesse stärker auf die Reaktion einwirken als in Bedingung W. Es ist jedoch auch möglich, daß nicht der Antwortkonflikt,

sondern die physikalische Ähnlichkeit die beobachteten Ergebnisse dieser Bedingung erklärt. Die Antwortrate deutet bereits darauf hin, daß für die Sätze dieser Bedingung eher die Tendenz besteht, sie als "selber Sprecher" zu werten. Die Reduktion des Semantik-Effekts könnte darauf zurückzuführen sein, daß die semantisch wohlgeformten Sätze der Bedingung M aufgrund einer größeren physikalischen Ähnlichkeit zum Standardsprecher tatsächlich schwieriger zu unterscheiden sind als die inkongruenten Sätze. Es erscheint nicht unplausibel, daß eine Unsicherheit in der Beurteilung zu einer Verzögerung der Reaktionszeiten führte und sich daher ein reduzierter Semantik-Effekt ergibt. Als experimentell überprüfbare Hypothese folgt aus dieser Annahme, daß eine niedrigere Antwortquote auch dann auftreten sollte, wenn die Sprecherbewertung ohne Zeitdruck erfolgt. Diese Hypothese wird in Experiment IIIb untersucht.

Die zweite Probandengruppe sollte dieselben Sätze auf ihre semantische Kongruenz hin beurteilen. Die Antwortquote zeigt hier einen Deckeneffekt. Damit ist die Reaktionszeit die eigentlich interessante abhängige Variable. Im Vergleich zur den Ergebnissen für die Sprecherunterscheidung liegen die Mittelwerte der unterschiedlichen Bedingungen nahe beieinander; d.h. der Einfluß des Sprecherwechsels ist hier deutlich geringer. Betrachtet man allerdings den reinen Semantik-Effekt (d.h. die Differenz der Reaktionszeiten für die inkongruenten im Vergleich zu den kongruenten Sätzen), so zeichnet sich ein Unterschied zwischen den Sprecherbedingungen ab.

(1) In der Bedingung ohne Sprecherwechsel (Std) ist der stärkste Semantik-Effekt vorhanden, der plausible Satzkontext beschleunigt die Verarbeitung des letzten Wortes.

(2) Bei dem Wechsel zu einer weiblichen Stimme ist ein ähnlicher Trend zu beobachten. Für die kongruenten Sätze der Bedingung W ist zwar keinerlei Auswirkung des Sprecherwechsels zu erkennen, es ergibt sich jedoch eine beschleunigte Ablehnung semantisch inkongruenter Sätze.

(3) Auch in Bedingung M werden die inkongruenten Sätze schneller abgelehnt als in der Standardbedingung.

Die Daten zeigen, daß die Beurteilung der semantischen Kongruenz im Reakti-

onszeitexperiment nicht unabhängig von der Stimminformation durchgeführt wird; inkongruente Sätze die einen Sprecherwechsel beinhalten, werden schneller abgelehnt als bei konstantem Sprecher. Im Falle der kongruenten Sätze ergibt sich kein signifikanter Unterschied zwischen den Sprechern. Der Mittelwert in Bedingung M liegt jedoch tendenziell höher als in den anderen Bedingungen. Da die Daten beider Probandengruppen/Experimentalaufgaben für die kongruenten Sätze der Bedingung M denselben Trend aufweisen, liegt die Vermutung nahe, daß seine Ursache im Stimulusmaterial zu suchen ist. Doch wie oben beschrieben, lieferte die Analyse von Stimuluslänge und Sprechgeschwindigkeit keine Erklärung für die beobachteten Reaktionszeiten. Tatsächlich ergab eine Post-hoc-Betrachtung der Ergebnisse, daß die Ursache möglicherweise eher auf Seiten der Probanden als auf Seiten der Stimuli zu suchen ist, denn nur etwa die Hälfte der Probanden wiesen keinen oder einen umgekehrten Semantik-Effekt für die Bedingung M auf. Alter, Geschlecht und Reading span (Danemann & Carpenter, 1980) korrelierten mit der Stärke des Effektes aber nicht. Auch ein Speed-Accuracy Trade-Off konnte ausgeschlossen werden. Die Daten werden daher auf den Antwortkonflikt ("ja" bei schwierig zu detektierendem Sprecherwechsel) in Bedingung M zurückgeführt, auf den einige Probanden offenbar besonders sensitiv reagieren.

Zusammenfassend ist zu sagen, daß die Verhaltensdaten auf einen wechselseitigen Einfluß von semantischer und sprecherbezogener Signalinformation deuten. Vor allem die Beschleunigung bei der Ablehnung inkongruenter Sätze, die einen Sprecherwechsel beinhalten, erwies sich als robuster Effekt. Es sollte weiter untersucht werden, inwieweit diese Beobachtungen spezifisch für die Anforderungen im Reaktionszeitexperiment sind. Mit Hilfe ereigniskorrelierter Hirnpotentiale ist es möglich, die neuronale Verarbeitung des Gehirns in direkterer Weise zu beobachten und so den angenommenen Antwortkonflikt zu vermeiden bzw. zu kontrollieren. Die nachfolgenden EKP-Untersuchungen sollen daher nähere Hinweise auf den zeitlichen Verlauf der beobachteten Effekte liefern, und aufgabenspezifische Einflüsse von der perzeptuell-kognitiven Verarbeitung dissoziieren.

## 10.4 Experiment IIIb: EKP-Studie - Sprecherdiskrimination

### 10.4.1 Fragestellung

Die folgende EKP-Untersuchung war eng an die vorhergehenden Reaktionszeitexperimente angelehnt und verwendete dieselben sechs Experimentalbedingungen. Um EKP- und Reaktionszeitdaten vergleichen zu können, bestand die Experimentalaufgabe wiederum in der Unterscheidung der Sprechstimmen, allerdings konnte die Entscheidungsaufgabe ohne zeitliche Beschränkung durchgeführt werden.

Eine EKP-Studie, die sich mit der semantischen Verarbeitung und - allerdings nur als Kontrollbedingung - mit den Auswirkungen eines Sprecherwechsels befaßt hat, ist die Untersuchung von McCallum, Farmer & Pocock (1984). In dieser Studie wurden Sätze präsentiert, bei denen das letzte Wort entweder eine semantische Inkongruenz aufwies oder einen Sprecherwechsel beinhaltete (während der Trägersatz von einer männlichen Stimme geäußert wurde, stammte das letzte Wort dann von einer Sprecherin). Die Aufgabe der Probanden bestand im passiven Zuhören, wobei nach jeweils 40 Sätzen eine Distraktoraufgabe durchzuführen war. Für die semantische Inkongruenz fanden McCallum und Mitarbeiter wie zu erwarten eine N400-Komponente, für den Sprecherwechsel eine breite Positivierung, die als eine Überlagerungen mehrerer Komponenten interpretiert wurde, als P300, slow-wave (von den Autoren als "P600" bezeichnet) und satzfinaler Positivität ("sentence wrap-up"). Für beide Inkongruenzen trat zudem eine schwache N200 Komponente auf, die in allen Sätzen von einer N100 überlagert und nur schwer von dieser zu dissoziieren war. Um ihre Experiment mit den visuellen Studien zur N400 (z.B. Kutas & Hillyard, 1980a,b) vergleichen zu können, präsentierten McCallum und Kollegen die Experimentalsätze sehr langsam gesprochen. Die ausgeprägte N100 ist auf die Pausen, die so zwischen den Wörtern auftraten und auf den abprupten Worteinsatz zurückzuführen. Um eine weniger artifizielle Präsentation zu gewährleisten, wurden die Sätze im vor-

liegenden Experiment mit einer natürlichen Sprechgeschwindigkeit dargeboten. Für solche kontinuierlichen Äußerungen ist mit einer reduzierten N100 zu rechnen (Friederici et al., 1993; Hagoort & Brown, 2000; O'Halloran et al., 1988), so daß die N200-Komponente besser untersucht werden kann. Die im Experiment von McCallum et al. (1984) beobachtete N200 für die semantische Verletzung und den Sprecherwechsel ist möglicherweise auf zwei qualitativ verschiedene Komponenten zurückzuführen.

Für den Sprecherwechsel ist denkbar, daß es sich dabei um eine Mismatch Negativierung handelt. Dies würde bedeuten, daß eine Sprecherevozierte MMN nicht auf die iterative Präsentation eines Standardreizes im Oddball-Paradigma angewiesen ist. Vielmehr wäre davon auszugehen, daß der Hörer auch bei der auditorischen Wahrnehmung einer variablen Äußerung eine mentale Repräsentation der Sprechstimme aufbaut. Die Verletzung dieser Repräsentation durch einen plötzliche Sprecherwechsel würde sich dann in der MMN manifestieren.

Für die semantische Verletzung könnte die frühe Negativierung hingegen als N250 Komponente interpretiert werden. Hagoort & Brown (2000) und vor allem van den Brink, Brown & Hagoort (2001) haben gezeigt, daß eine solche Komponente immer dann eine größere Amplitude aufweist, wenn aufgrund der kontextuellen Erwartung eine erschwerte lexikalische Selektion stattfindet.[4]

Wie oben erwähnt, war im folgenden Experiment eine Sprecherdiskriminationsaufgabe durchzuführen, bzw. die Probanden sollen nach jedem Satz entscheiden, ob während der Präsentation der Sprecher gewechselt hat oder nicht. Die Fokussierung auf den Sprecherwechsel sollte zu einem N200/P300-Komplex im Falle eines solchen Wechsels führen. Dieses bipolare Aktivierungsmuster sollte sich dabei aus aufgabenunabhängigen Komponenten (MMN, P3a) bestehen, sowie aus Komponenten, welche die kognitiven Anforderungen der Experimentalaufgabe reflektieren (N2b, P3b).

[4]Wird der Satz mit dem kontextuell erwarteten Wort beendet, tritt eine solche N250 ebenfalls auf, jedoch mit geringerer Amplitude. Aus der Tatsache, daß sie jedoch auch für die semantisch kongruente Bedingung zu finden ist, schließen van den Brink et al. (2001), daß es sich bei der N250 um einen Marker für lexikalische Selektion handeln muß und nicht, wie von Conolly & Phillips (1994) und Conolly et al. (1995) vorgeschlagen, um das Korrelat eines phonologischen Mismatch zwischen präsentiertem und erwartetem Wort.

Aufgrund der Reaktionszeitergebnisse (Experiment IIIa) ist zu erwarten, daß dabei nicht vom semantischen Inhalt abstrahiert wird. Das heißt, daß auch bei einer Sprecherdiskriminationsaufgabe eine N250 und eine eine N400 Komponente auftreten sollten. Die im vorliegende Experiment ebenfalls präsentierte doppelte Inkongruenz, d.h. die Kombination aus semantischer Verletzung und Sprecherwechsel, erlaubt nun die Untersuchung von möglichen Interaktionen zwischen der Verarbeitung von Stimminformation und dem semantischen Satzverstehen. Sollte es zu einer solchen Interaktion in den neurophysiolgischen Korrelaten kommen, so gibt es zwei Zeitfenster, in denen das geschehen kann: Im frühen Zeitfenster um 200-300 ms, in dem der Sprechereffekt (MMN/N2b) und die lexikalische Selektion (N250) angenommen werden, oder im späteren Zeitfenster (um 400-600 ms), in dem N400-Komponente zu erwarten ist.

### 10.4.2 Methode

#### 10.4.2.1 Probanden

27 Versuchspersonen (13 männlich) im Alter von 18 - 31 Jahren (Durchschnitt 23 Jahre) nahmen an dieser EKP-Studie teil. Alle Probanden waren deutsche Muttersprachler und nach dem Edinburgh Inventory Oldfield (1971) als Rechtshänder klassifiziert. Die Probanden wurden für ihre Teilnahme am Experiment bezahlt und hatten nicht an der oben beschriebenen Reaktionszeitstudie teilgenommen.

#### 10.4.2.2 Stimuli

Die Stimuli umfaßten dieselben Bedingungen wie in dem vorhergehenden Reaktionszeitexperiment. Allerdings wurden jetzt zur einer Verbesserung des Signal-Rausch-Verhältnisses mehr Stimuli verwendet. Eine Übersicht der Stimulushäufigkeiten findet sich in Tabelle 10.7. Zur Ausbalancierung der Ja/Nein-Antworten wurden ausserdem Filler-Sätze konstruiert. Diese besaßen eine andere syntaktische Struktur als die oben beschriebenen Trial-Sätze; der Sprecherwechsel, sowie die semantische Anomalie traten an beliebiger Stelle im Satz auf (z.B.: *"Berlin ist die neue Hauptstadt/Tischdecke Deutschlands."*)

Insgesamt ergab sich damit eine Stimulus-Zusammenstellung wie in Tabelle 10.7 beschrieben (auf die Frage nach dem Sprecherwechsel war insgesamt 288 mal mit “Ja“ und 288 mal mit “Nein“ zu antworten; die Hälfte der Sätze jeder Bedingung war semantisch inkongruent).

| | | Trialsätze | | Fillersätze | |
|---|---|---|---|---|---|
| | | **SEMANTIK** | | **SEMANTIK** | |
| | | Kongruenz | Inkongruenz | Kongruenz | Inkongruenz |
| **SPRECHER** | Std | N=48 | N=48 | N=96 | N=96 |
| | M | N=48 | N=48 | N=24 | N=24 |
| | W | N=48 | N=48 | N=24 | N=24 |

Tabelle 10.7: Stimulusverteilung des Experiments IIIb

#### 10.4.2.3 Versuchsanordnung

Die Versuchsanordnung entsprach dem vorhergehenden Versuch. Die Präsentation der Stimuli erfolgte mittels der Software ERTS (Experimental Runtime System, Version 3.25, Berisoft Corporation 1995). Visuelle Hinweise wurden auf einem Computermonitor präsentiert, auditive Reize über einen Vorverstärker (Digilink III., Conrad Audio) und Lautsprecher (JBL XL 300). Die Sätze wurden in normaler Zimmerlautstärke (etwa 60 dB) präsentiert.

#### 10.4.2.4 EEG-Aufnahme

Das EEG wurde mit 25 AG/AGCL Elektroden abgeleitet, die integrierter Bestandteil einer elastischen Haube waren (Electrocap Inc.). Die gewählten Elektrodenpositionen sind in Abbildung 10.6 aufgeführt.

Die Hirnpotentiale wurden zunächst monopolar gegen die Referenzelektrode der linken Mastoidposition (A1) abgeleitet. Später wurde “off-line“ gegen die ver-

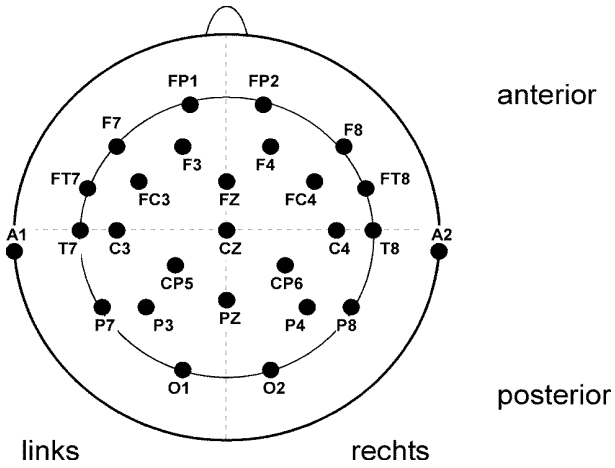



Abbildung 10.6: Experiment IIIb: Elektrodenpositionen

bundenen Mastoide (A1, A2) re-referenziert. Ein bipolares Electrookulogramm (EOGV, EOGH) registrierte Augenbewegung und Lidschlag, so daß diese Artefakte ebenso wie Drifts einzelner Elektroden von der weiteren Analyse ausgeschlossen werden konnten. Die Impedanzen aller Elektroden wurden unter 5 kΩ gehalten. Das EEG Signal wurde von einem PC bei einer Abtastrate von 250 Hz und einer Quantisierung von 22 Bit aufgezeichnet, vor der Auswertung wurde das EEG-Signal zudem mit 45 Hz tiefpaßgefiltert. Die ereigniskorrelierten Potentiale wurden nur für die korrekt beantworteten Versuchsdurchgänge ausgewertet. Trials mit falscher oder keiner Antwort, sowie EKPs, die von Artefakten z.B. von Augenbewegungen überlagert waren, wurden durch eine visuelle Inspektion der Rohdaten ausgeschlossen. Die EKPs wurden für 900 ms nach Beginn der Abweichung und mit einer Baseline von 200 ms vor Beginn des letzten Wortes ermittelt. Bei statistischen Analysen mit mehr als einem Freiheitsgrad im Zähler wurde eine Greenhouse-Geisser-Korrektur durchgeführt.

#### 10.4.2.5 Versuchsablauf

Etwa zehn Minuten vor dem Beginn des Experiments erhielt der Proband eine schriftliche und mündliche Einweisung in die Experimentalaufgabe. Dabei wurde

er instruiert, nach jedem gehörten Satz zu entscheiden, ob der Sprecher während des Satzes gewechselt hat, oder nicht. Der Versuchsteilnehmer wurde daraufhin in einer kurzen interaktiven Einführung von jeweils zwei Sätzen mit den drei Stimmen vertraut gemacht und führte danach einen Übungsblock von insgesamt zehn Sätzen durch. Eine einzelne Aufgabendurchführung sah folgendermaßen aus: 1000 ms vor dem Satzbeginn erschien ein Fixationskreuz auf dem Monitor, das während der Satzpräsentation und 900 ms danach zu sehen war. Danach erschien bis zur Beantwortung und maximal für 6000 ms die Frage auf dem Monitor "Hat der Sprecher gewechselt?", auf die der Proband sollte per Knopfdruck mit "Ja" (linker Daumen) oder "Nein" (rechter Daumen) antworten sollte. Es folgte die explizite Auffoderung "Jetzt blinzeln!". Nach 3000 ms erschien das Fixationskreuz und der nächste Satz wurde präsentiert. Die Probanden wurden gebeten, während der Satzpräsentation möglichst nicht zu blinzeln.
Im Gegensatz zur Reaktionszeitstudie bestand hier nun für die Durchführung der Entscheidungsaufgabe kein Zeitdruck. Vielmehr waren die Probanden instruiert, erst bei Erscheinen der Frage auf dem Monitor, d.h. deutlich nach der Satzpräsentation zu antworten.

### 10.4.3 Ergebnisse

**Verhaltensdaten** Tabelle 10.8 zeigt die Antwortquoten (gemessen als relative Häufigkeiten der korrekten Antworten). Demnach gelang die Sprecherunterscheidung für die Bedingungen Std und W. Die Antwortquoten unterschieden sich nicht für semantisch korrekte und inkorrekte Sätze. Für die Bedingung M hingegen war eine niedrigere Antwortquote zu beobachten als in den Bedingungen Std und W. Es lag hier besonders bei den semantisch korrekten Sätzen eine Tendenz vor, diese Bedingung als "selber Sprecher" zu werten. Eine Varianzanalyse wurde für die Faktoren SPRECHER $\times$ SEMANTIK durchgeführt. Wie Tabelle 10.9 zeigt, waren beide Haupteffekte signifikant und es ergab sich eine signifikante Interaktion beider Faktoren. Die Antwortquoten der kongruenten und der inkongruenten Sätze unterschieden sich nur für Sprecherbedingung M signifikant

($F_{(1,26)} = 34.04, p < 0.0001$).

| *SEMANTIK* | *SPRECHER* | **% Korrekt** |
|---|---|---|
| kongruent | **Std** | 97.45 [4.41] |
| | **M** | 84.34 [10.96] |
| | **W** | 98.23 [3.69] |
| inkongruent | **Std** | 97.14 [2.90] |
| | **M** | 92.67 [7.25] |
| | **W** | 98.15 [3.34] |

Tabelle 10.8: *aktive Sprecherunterscheidung:* Antwortquoten und Standardabweichungen für die verschiedenen Sprecherbedingungen

**EKP-Ergebnisse** Es werden zwei Vergleiche diskutiert. Zum einen werden die Korrelate des Sprecherwechsels betrachtet, indem die EKPs der Bedingungen Std, M und W verglichen werden. Zum anderen werden die Korrelate der Integration von semantisch kongruenten und inkongruenten Sätzen in den drei Sprecherbedingungen betrachtet.

*Frühe Negativierung*

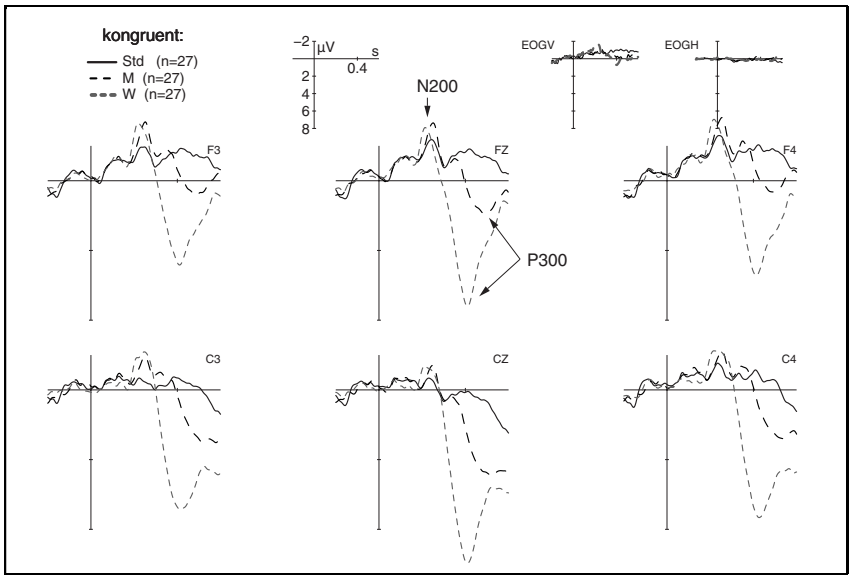
Vergleicht man die drei Sprecherbedingungen Std, M und W, so zeigt sich sowohl in den semantischen kongruenten, als auch in den inkongruenten Sätzen (Abbildung 10.7) für die Bedingung M und W eine frühe Negativierung (N200), die besonders an den anterioren Elektroden ausgeprägt ist. Die Gipfellatenz an FZ beträgt bei den semantisch kongurenten Sätzen 216 ms für Bedingung W und

| *FAKTOREN* | *df* | *F* | *p* |
|---|---|---|---|
| SPRECHER | 2,52 | 24.61 | $< 0.0001$ |
| SEMANTIK | 1,26 | 21.19 | $< 0.0001$ |
| SPRECHER $\times$ SEMANTIK | 2,52 | 27.84 | 0.0001 |

Tabelle 10.9: *Aktive Sprecherunterscheidung:* Antwortquoten bei den verschiedenen Sprecherbedingungen

248 ms für Bedingung M; bei semantischer Inkongruenz 212 ms für Bedingung W und 265 ms für Bedingung M. Für die Bedingung Std ist diese Komponente deutlich schwächer und hat eine Latenz von 244 ms (semantisch kongruente Sätze) bzw. 256 ms (inkongruente Sätze). Eine globale Varianzanalyse der mittleren Amplituden wurde durchgeführt für den Zeitbereich von 180 bis 280 ms mit den Faktoren SPRECHER (Std vs. M vs. W), SEMANTIK (kongruent vs. inkongruent) HEMISPHÄRE (links vs. rechts) und REGION (anterior vs. zentral vs. posterior). Der Faktor REGION umfaßte dabei die Elektroden F3, F7, F4, F8 (anterior), T7, C3, T8, C4 (zentral), sowie P3, P7 P4, P8 (posterior). Der Faktor HEMISPHÄRE umfaßte die Elektroden F3, F7, C3, T7, P3, P7 (links) und F4, F8, C4, T8, P4, P8 (rechts). Der Zeitbereich wurde mit 180-280 ms so gewählt, daß er den frühen Effekt in den relevanten Bedingungen möglichst symmetrisch abdeckt.

Es ergab sich kein signifikanter Haupteffekt SEMANTIK und auch keine signifikante Interaktion SEMANTIK $\times$ SPRECHER. Eine Interaktion war für die Faktoren SEMANTIK $\times$ REGION zu beobachten ($F_{(2,52)} = 6.56, p < 0.01, \varepsilon = 0.6592$), eine separate Analyse der drei Regionen zeigte dort allerdings keinen bedeutsamen SEMANTIK-Effekt. Für die anterioren Elektroden zeigte sich aber ein Trend zu einer stärkeren Negativität bei semantischer Inkongruenz ($F_{(1,26)} = 3.43, p = 0.0752$). Es ergab sich ein signifikanter Haupteffekt SPRECHER ($F_{(2,52)} = 5.53, p < 0.01, \varepsilon = 0.9421$), sowie eine signifikante Interaktion SPRECHER $\times$ REGION ($F_{(4,104)} = 4.13, p < 0.05$). Eine separate Analyse der Elektrodenregionen zeigte einen signifikanten SPRECHER-Effekt an anterioren und zentralen Elektroden



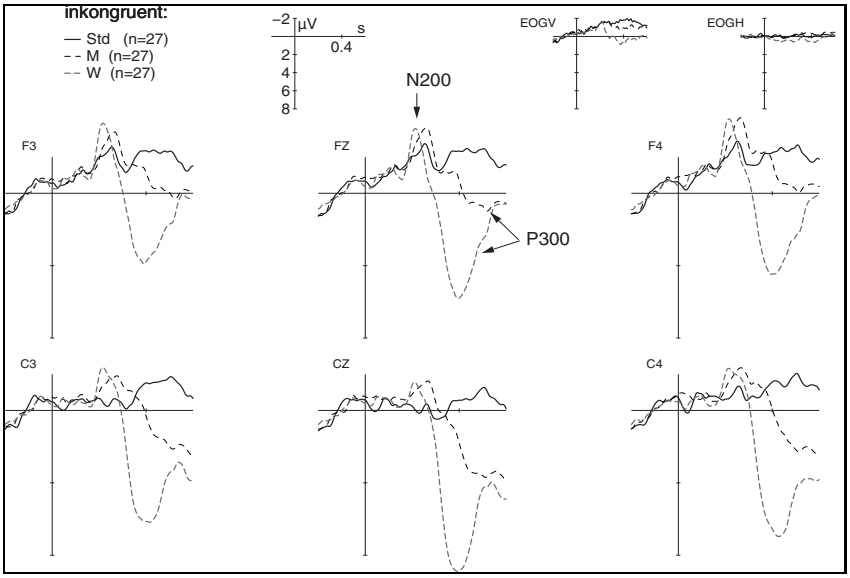




Abbildung 10.7: *Aktive Sprecherunterscheidung* EKPs an ausgewählten Elektroden infolge semantisch kongruenter (oben) und inkongruenter (unten) Sätze der drei verschiedenen Sprecherbedingungen Std, M, und W

auf (anterior: $F_{(2,52)} = 4.00, p < 0.05$; zentral: $F_{(2,52)} = 10.14, p < 0.001$). Signifikante Interaktionen mit dem Faktor HEMISPHÄRE traten nicht auf.

Der Haupteffekt SPRECHER wurde nun für die anterioren und zentralen Elektroden aufgelöst. Da sich ein signifikanter Unterschied sowohl zwischen Std und M (anterior: $F_{(1,26)} = 7.46, p < 0.05$; zentral: $F_{(1,26)} = 9.65, p < 0.01$) als auch zwischen Std und W ( anterior: $F_{(1,26)} = 3.5, p = 0.0725$ zentral: $F_{(1,26)} = 16.14, p < 0.001$) ergab, wurden die EKPs dieser beiden Bedingungen an den anterioren und zentralen Elektroden direkt miteinander verglichen. Dazu gingen die beiden Differenzen M-Std und W-Std in die Varianzanalyse ein. Die ANOVA der Faktoren SPRECHER (W, M) $\times$ REGION (anterior, zentral) zeigte hier allerdings keinerlei Haupteffekte oder Interaktionen.

*Positivierung*

Die frühe Negativierung wird in Bedingung W gefolgt von einer starken Positiverung (P300) mit einem zentralen Schädelverteilung und einer Gipfellatenz von 396 ms an CZ. Diese Positivierung ist deutlich schwächer für die Bedingung M und nicht vorhanden in Bedingung Std. Sie wird als P300 Komponente interpretiert. Da sich die Zeitfenster der späteren Negativierung (N400) und der sprecherbedingten P300 überlagern, wurde letztere getrennt für die semantisch kongruenten und inkongruenten Sätze berechnet. Die globale Analyse (ANOVA) der Faktoren SPRECHER (Std, M, W), HEMISPHÄRE (links,rechts) und REGION (anterior, zentral, posterior)[5] zeigte für semantisch kongruente und inkongruente Sätze einen Haupteffekt SPRECHER (kongruent:$F_{(2,52)} = 89.87, p < 0.0001, \varepsilon = 0.9237$ inkongruent: $F_{(2,52)} = 90.23, p < 0.0001, \varepsilon = 0.9529$), sowie eine SPRECHER $\times$ REGION Interaktion (kongruent: $F_{(4,104)} = 3.63, p < 0.05, \varepsilon = 0.4986$ inkongruent: $F_{(4,104)} = 4.46, p < 0.05, \varepsilon = 0.4214$). Eine Interaktion SPRECHER $\times$ HEMISPHÄRE trat nur für die semantisch inkorrekten Sätze auf ($F_{(2,52)} = 7.64, p < 0.01, \varepsilon = 0.9940$).

Der Haupteffekt SPRECHER wurde aufgelöst zu einer getrennten Analyse für die beiden Abweichbedingungen, d.h. es wurden separate Analysen für die Be-

[5] die Elektrodenzuordnung war identisch mit derjenigen der N200-Analyse

dingungen M vs. Std und W vs.Std durchgeführt. Für die Bedingungen M vs. Std zeigte sich ein signifikanter SPRECHER-Haupteffekt (kongruent: $F_{(1,26)} = 8.26, p < 0.01$; inkongruent: $F_{(1,26)} = 12.50, p < 0.01$), der nicht mit den Faktoren HEMISPHÄRE (F<1), aber mit dem Faktor REGION interagierte (kongruente: $F_{(2,52)} = 8.26, p < 0.01, \varepsilon = 0.5841$; inkongruente: $F_{(2,52)} = 8.05, p < 0.01, \varepsilon = 0.4495$). Für die inkongruenten Sätze war außerdem die dreifache Interaktion SPRECHER $\times$ REGION $\times$ HEMISPHÄRE signifikant ($F_{(2,52)} = 4.49, p < 0.05, \varepsilon = 0.8081$). Die Verteilungseffekte beschreiben hierbei einen größeren P300 Effekt für die zentralen Elektroden, der im Fall der inkongruenten Sätze ausserdem an den rechtshemisphärischen Elektroden stärker war.
Für die Bedingungen W und Std war der SPRECHER-HAUPTEFFEKT ebenfalls signifikant (kongruent:$F_{(1,26)} = 130.83, p < 0.0001$; inkongruent: $F_{(1,26)} = 140.22, p < 0.0001$). Desweiteren war eine Interaktion mit dem Faktor REGION zu beobachten (kongruent: $F_{(2,52)} = 5.17, p < 0.05, \varepsilon = 0.5500$; inkongruent: $F_{(2,52)} = 5.28, p < 0.05, \varepsilon = 0.5684$), der auf eine zentrale Verteilung des P300-Effekts in Reaktion auf die weibliche Stimme (Bedingung W) hindeutet.

*Späte Negativierung* Im direkten Vergleich der kongruenten und inkongruenten Sätze tritt etwas später als die Positivierung, aber noch von ihr überlagert, eine deutliche Negativierung (N400) hervor, die in einer separaten Varianzanalyse untersucht wurde. Die Abbildungen 10.8-10.10 zeigen für alle drei Sprecherbedingungen (Std, M, und W) jeweils die EKPs auf die semantisch korrekten vs inkorrekten Bedingungen. Für die statistischen Analyse wurden 9 Elektroden (FZ, CZ, PZ, F3, F4, C3, C4, P3, P4) im Zeitfenster 470-720 ms betrachtet. Dieser SEMANTIK-Effekt war für alle drei Sprecherbedingungen signifikant (Std: $F_{(1,26)} = 20.65, p < 0.0001$; M: $F_{(1,26)} = 8.05, p < 0.0087$; W: $F_{(1,26)} = 6.24, p < 0.05$). In allen Bedingungen war diese Negativierung vor allem an den zentralen und posterioren Elektroden ausgeprägt. Eine Interaktion ELEKTRODE X SEMANTIK wurde jedoch nur für Bedingung Std signifikant ($F_{(8,208)} = 9.08, p < 0.0005, \varepsilon = 0.2467$ ), wobei sich ein Maximum für den SEMANTIK-Effekt an Elektrode PZ fand. Für Bedingung M lag die Interaktion im Trendbereich $F_{(8,208)} = 3.00, p = 0.55, \varepsilon = 0.2680$),

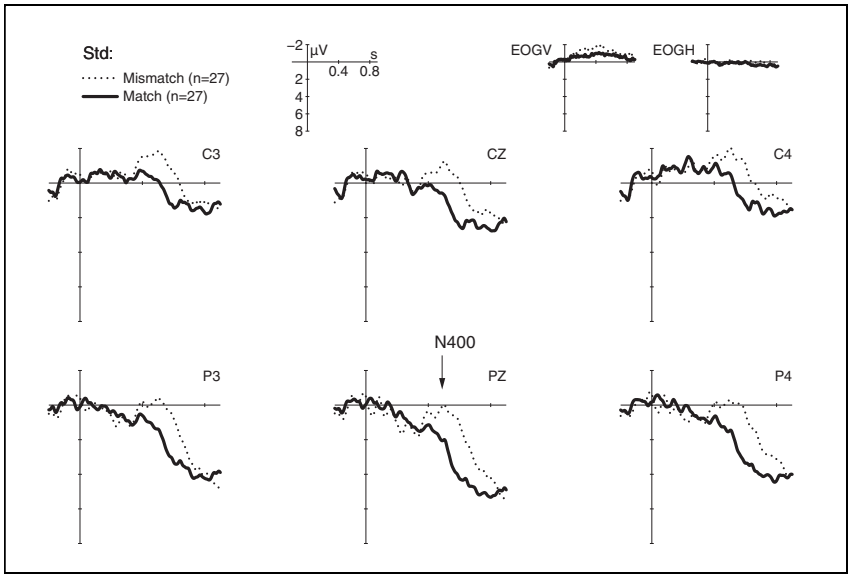



Abbildung 10.8: *Aktive Sprecherunterscheidung* EKPs infolge semantisch kongruenter und inkongruenter Sätze der Sprecherbedingung Std

eine maximale Differenz aus inkongruenten und kongruenten Sätzen ergab sich hier ebenfalls für die Elektrode PZ. Für den Vergleich zwischen den Sprecherbedingungen wurden die jeweiligen Differenzwerte aus den EKPs auf semantisch kongruente und semantisch inkongruente Sätze herangezogen. Sie sind als EKP-Differenzkurven in Abbildung 10.11 dargestellt. Die frontalen Elektroden, auf denen der Effekt nur schwach zu sehen war, wurden von der statistische Analyse ausgeschlossen. Es ergab sich weder für die durchschnittlichen noch für die maximalen Amplituden ein signifikanter Haupteffekt SPRECHER und ebenfalls keine Interaktion SPRECHER $\times$ ELEKTRODE. Da sich durch die Differenzbildung das Signal-Rausch-Verhältnis zusätzlich verschlechterte, wurde von einer Analyse der Gipfellatenzen abgesehen. Die visuelle Inspektion zeigt aber zumindest, daß das bestdefinierte Potenzial für die Standardbedingung erfolgte, während die Varianz (Jitter) für die Bedingungen M und W höher erscheint.

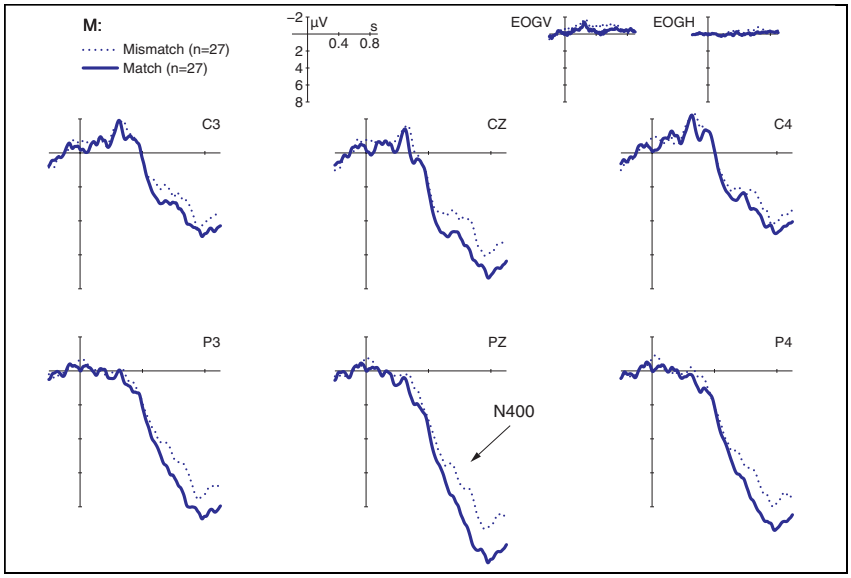



Abbildung 10.9: *Aktive Sprecherunterscheidung* EKPs infolge semantisch kongruenter und inkongruenter Sätze der Sprecherbedingung M

### 10.4.4 Diskussion

Die Performanzdaten der Sprecherdiskrimination bestätigen im wesentlichen die Daten aus der vorhergehenden Reaktionszeitstudie. Auch hier erwies sich die Unterscheidung der zwei Männerstimmen als deutlich schwieriger als die Detektion der weiblichen Stimme. Auch ohne Zeitdruck auf die Entscheidung, waren besondere Schwierigkeiten für die semantisch kongruenten Sätze der Bedingung M zu beobachten. Wie bereits oben vermutet, scheint die Ähnlichkeit zwischen den beiden männlichen Stimmen in dieser Bedingung besonders groß zu sein. Es ist daher wahrscheinlich, daß die längeren Reaktionszeiten für diese Sätze in Experiment IIIa auf die Unsicherheit in der Diskrimination zurückzuführen sind. Die vermeintliche Reduktion des Semantik-Effekts beruht demnach nicht auf einem Einfluß der Semantik, sondern auf die grundsätzlich größeren Diskriminationsschwierigkeiten.

Im vorliegenden Experiment wurden ereigniskorrelierte Potentiale für das jeweils letzte Wort im Satz gemessen. Im Vergleich zu McCallum et al. (1984) war in der

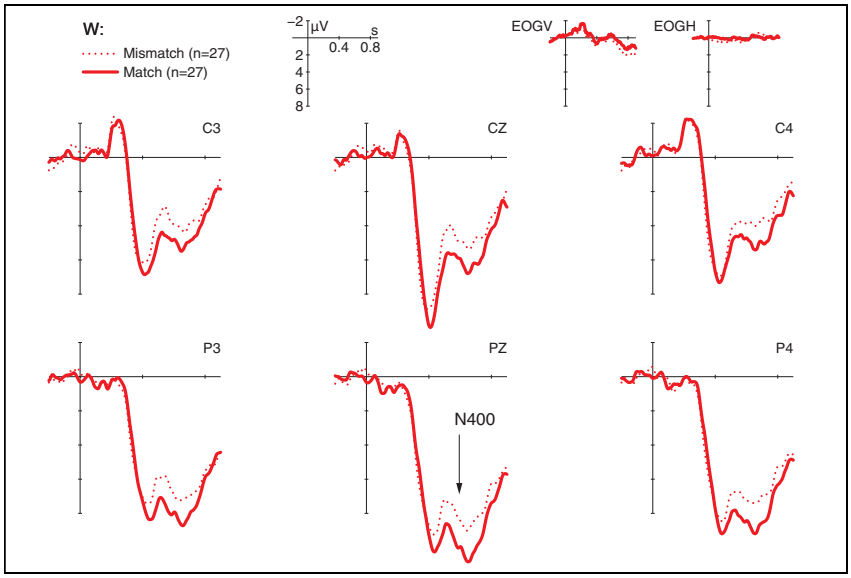



Abbildung 10.10: *Aktive Sprecherunterscheidung* EKPs infolge semantisch kongruenter und inkongruenter Sätze der Sprecherbedingung W

vorliegenden Studie wie erwartet für die Präsentation kontinuerlicher Lautsprache eine Reduktion der N100 Komponente zu beobachten (vgl. Friederici et al., 1993; Hagoort & Brown, 2000; O'Halloran et al., 1988).

Alle Bedingungen, die einen Sprecherwechsel beinhalteten, die also eine Verletzung der Kontinuitätsannahme darstellten, evozierten eine Negativierung um 200 ms (N200) nach Beginn der aufgabenrelevanten Abweichung (Sprecherwechsel). Eine ähnliche Negativierung zeigte sich auf die Verletzung einer Kontinuitätsannahme in der Studie von Hahne, Schröger & Friederici (2002). Dort wurde bei dem letzten Wort eines Satzes die Richtung geändert, aus der das Sprachsignal präsentiert wurde. Dieser Wechsel der Hörrichtung evozierte eine Negativierung, die von den Autoren als MMN interpretiert wurde. In der vorliegenden Studie zeigte sich also für den Sprecherwechsel ebenfalls eine Negativierung um 200 ms, die sich für die Bedingungen M (männliche Stimme) und W (weibliche Stimme) nicht unterschied. Für die aufmerksame Detektion einer Abweichung kann erwartet werden, daß der aufgabenrelevante Sprecherwechsel neben der Mismatch

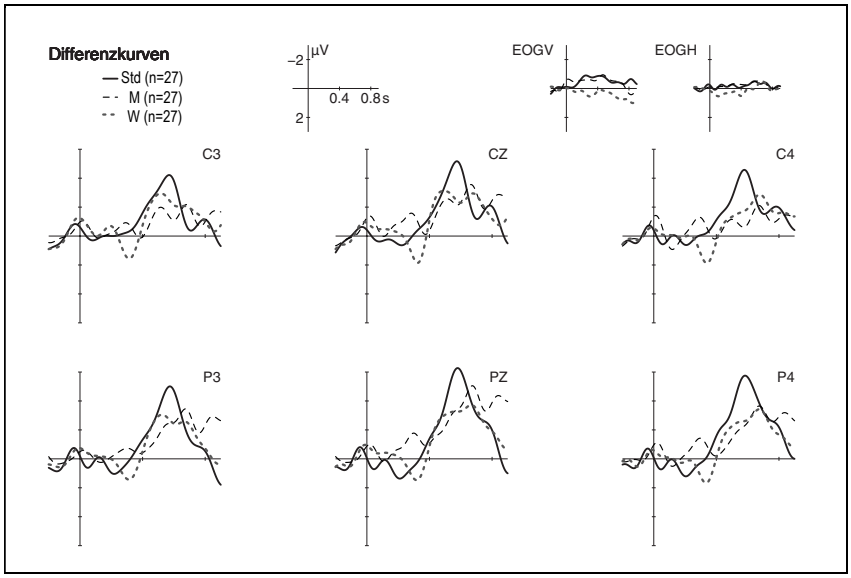



Abbildung 10.11: *Aktive Sprecherunterschiedung* Differenzkurven der EKPs infolge der jeweils semantisch kongruenten und inkongruenten Sätze der drei Sprecherbedingungen (für diese Abbildung mit 7 Hz tiefpaßgefiltert). Die bestausgeprägte N400 ist für die Standardbedingung zu beobachten. Die statistische Analyse weist allerdings keine signifikanten Unterschiede zwischen den Bedingungen auf.

Negativierung eine N2b Komponente evoziert (siehe Pritchard et al., 1991). Diese Komponenten weisen zwar in der Regel ein zentrales, bzw. frontozentrales Maximum auf, aber frontale N2b Komponenten sind für die auditorische Wahrnehmung auch berichtet worden (ebenda). Die beobachteten Potentiale werden daher als eine mögliche Überlagerung von MMN und aufgabenabhängiger N2b interpretiert. Wenn auch die Anteile aus MMN und N2b an dieser frühen Negativierung für Bedingungen M und W unterschiedlich sein können, so war jedoch keine Differenz in der resultierenden Amplitude dieser Negativierung (N200) zu beobachten. Die späteren Latenzen für Bedingung M deuten jedoch darauf hin, daß dort die Gipfellatenz eher der N2b-Komponenete zuzuordnen sind. Die weibliche Stimme evoziert hingegen eine starke und frühe MMN, da sie sich von Standardsprecher deutlicher unterscheidet, als die deviante Männerstimme.

Auf die frühe Negativierung folgte eine deutliche positive Komponente, die als

P300 interpretiert wird. Diese P300 ist ebenso in eine stimulusabhängige P3a- und eine aufgabenbedingte P3b-Teilkomponente zu unterteilen. Die P3a-Komponente ist abhängig von den physikalischen Parametern des Sprecherwechsels und daher für Bedingung M geringer als für Bedingung W. Außerdem reflektiert sie die subjektive Häufigkeit eines Ereignisses (siehe Pritchard, 1991). Kategorisieren die Hörer in männliche (Std, M) vs. weibliche (W) Stimmen, so kann die stärkere P3a-Komponente auch damit erklärt werden, daß die weibliche Stimme, die seltenere Kategorie darstellt.

Für die P3b ist eine Abhängigkeit der Amplitude von der Schwierigkeit der durchzuführenden Reizunterscheidung berichtet worden (siehe Kapitel 4.4). Die P3b-Komponente tritt daher für beide Devianzbedingungen auf, ist aber für Bedingung M, die eine schwierigere Entscheidung verlangt, geringer.

Ein Effekt der semantischen Kongruenz zeigte sich in dem frühen Zeitbereich um 200-300 ms nur tendenziell. Andere Studien, bei denen die Experimentalaufgabe auf die semantische Integration fokussiert, berichten hingegen einen derartigen N250 Effekt (Hagoort & Brown, 2000; van den Brink et al., 2001). Der Grund für das Ausbleiben einer solchen Komponente in der vorliegenden Studie könnte in der Aufgabenstellung liegen, die nicht auf die semantische Verarbeitung abzielt. Durch eine Änderung der Experimentalaufgabe wurde diese Erklärungshypothese im anschließenden Experiment überprüft.

Als Reaktion auf die semantische Inkongruenz war in allen Sprecherbedingungen eine N400 Komponente zu beobachten. Wie sich bereits im Reaktionszeit-Experiment IIIa abzeichnete, wird die semantische Inkongruenz also auch dann verarbeitet, wenn die Experimentalaufgabe nur auf die Sprecherunterscheidung abzielt. Andere Studien, die auf der Wortebene durchgeführt wurden, berichten hingegen bei ähnlichen Aufgaben, die sich auf die physikalischen Eigenschaften der Stimuli beziehen, ein Ausbleiben der N400. Chwilla, Brown & Hagoort (1995) fanden beispielsweise bei visueller Wortpräsentation eine N400 nur für die lexikalische Entscheidung, nicht aber, wenn die Probanden eine Beurteilung der Wortlänge abzugeben hatten. Bei auditorischer Wortpräsentation trat die N400 nur dann auf, wenn die Probanden instruiert waren, sich die Wörter einzuprägen,

nicht aber, wenn sie lediglich Pseudowörter zählen sollten (Bentin et al., 1993). Diese Ergebnisse zeigen, daß die N400 nur bei der "tiefen", d.h. semantischen Verarbeitung evoziert wird. Für die hier präsentierten Ergebnisse bedeutet dies, daß die Sprecherinformation nicht so stark vom Sprachsignal abstrahiert wird, daß es nicht mehr zu einer semantischen Verarbeitung käme. Vielmehr erfolgt die semantische Verarbeitung bei gegebenem Satzkontext und natürlicher Sprechrate auch dann, wenn sie nicht im Fokus der Experimentalaufgabe steht.

Als das wichtigste Ergebnis der Studie bleibt jedoch festzuhalten, daß die ereigniskorrelierten Hirnpotentiale keine Hinweise auf eine Interaktion zwischen Sprecherbedingung und semantischer Kongruenz lieferten. Vielmehr scheint der Sprecherwechsel früh detektiert zu werden, aber keinen Einfluß auf die semantische Integration bzw. die N400 Komponente zu haben.

Die hier durchgeführte EKP-Studie zeigte allerdings eine Reihe aufgabenabhängiger Effekte: Die frühe Negativierung (N200) ist möglicherweise eine Überlagerung von N2b und MMN. Die P300-Komponente hat einen eher stimulusabhängigen (P3a) und einen aufgabenbedingten (P3b) Anteil. Auch das Ausbleiben eines frühen Effekts der semantischen Inkongruenz (N250) kann auf die Experimentalaufgabe zurückzuführen sein. Um diese aufgabenrelatierten Einflüsse zu kontrollieren, wurde in einer weiteren Studie das identische Satzmaterial präsentiert, ohne daß die Probanden eine spezifische Aufgabe zu bearbeiten hatten.

## 10.5 Experiment IIIc: EKP-Studie - Passives Zuhören

### 10.5.1 Fragestellung

In diesem Experiment sollten die Auswirkungen eines Sprecherwechsels auf das Satzverstehen untersucht werden, ohne daß die neurophysiologische Reaktion durch eine Experimentalaufgabe beeinflußt ist. Dazu wurde das vorangegangene Experiment wiederholt, wobei die Probanden nun die Sätze lediglich anhören sollten.
Es wurden für die elektrischen Hirnpotentiale, die mit dem satzfinalen Partizip zeitlich korreliert sind, folgende Vorhersagen getroffen:

1. Wegfall des aufgabenbedingten N2b/P3b-Komplexes bei einem Sprecherwechsel (M, W)
2. größere Mismatch Negativierung für Bedingung W als für Bedingung M aufgrund der größeren physikalischen und perzeptuellen Distanz zum Standardsprecher
3. größere P3a für Bedingung W als für Bedingung M aus demselben Grund
4. größere N250 bei semantischer Inkongruenz aufgrund einer erschwerten lexikalischen Selektion
5. größere N400 bei semantischer Inkongruenz aufgrund einer erschwerten semantischen Integration

### 10.5.2 Methode

#### 10.5.2.1 Probanden

Von gemessenen 29 Versuchspersonen gingen 25 (13 männlich) im Alter von 20-30 Jahren (Durchschnitt 24 Jahre) in die Auswertung ein. Fünf Probanden wurden aufgrund eines zu schlechten Signal-Rausch-Verhältnisses und zu starker Elektrodendrifts von der Analyse ausgeschlossen. Alle Probanden waren deutsche Muttersprachler und nach dem Edinburgh Inventory (Oldfield, 1971) als

Rechtshänder klassifiziert. Die Probanden wurden für ihre Teilnahme am Experiment bezahlt und hatten nicht an der oben beschreibenen Reaktionszeit- oder EKP-Studie teilgenommen.

#### 10.5.2.2 Stimuli

Das Stimulusmaterial entsprachen dem von Experiment IIb (siehe auch Tabelle 10.7).

#### 10.5.2.3 Versuchsanordnung

Die Versuchsanordnung und die zur Präsentation verwendete Hard- und Software entsprachen Experiment IIIb.

#### 10.5.2.4 EEG-Aufnahme

Die Erfassung des EKP-Signals erfolgte in gleicher Weise wie in Experiment IIIb.

#### 10.5.2.5 Versuchsablauf

Etwa zehn Minuten vor dem Beginn des Experiments erhielt der Proband eine schriftliche und mündliche Einweisung, durch die er instruiert wurde, während der Satzpräsentation nicht zu blinzeln. Vor dem jeweiligen Satzbeginn erschien für 1000 ms ein Fixationskreuz auf dem Bildschirm, das während der Satzpräsentation und bis 900 ms danach dort zu sehen war. Daraufhin erschien die explizite Aufforderung, kurz zu blinzeln. Nach 3000 ms erschien das Fixationskreuz des nächsten Versuchsdurchgangs.

### 10.5.3 EKP-Ergebnisse

Es waren für den Vergleich der Sprecher wiederum eine frühe Negativierung (MMN) und eine darauf folgende Positivierung (P300) zu beobachten. Für den

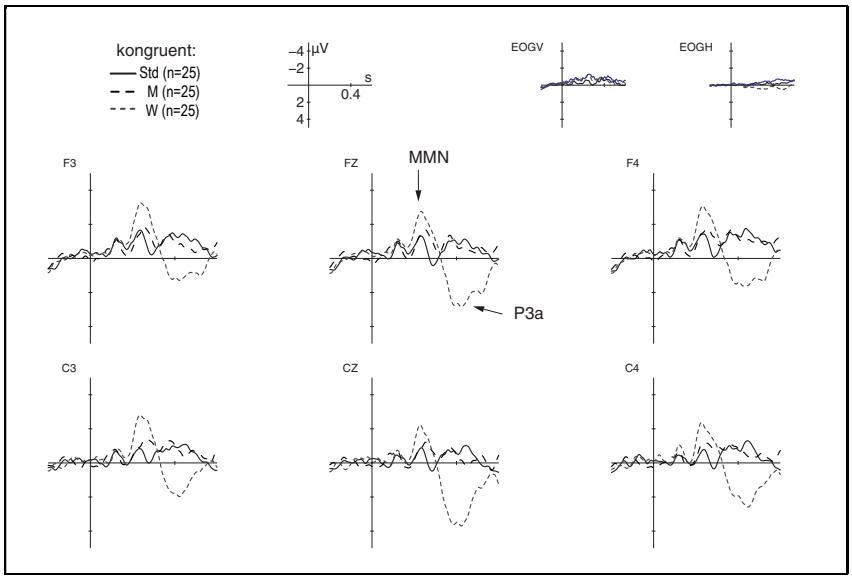
Vergleich der semantisch inkongruenten und kongruenten Sätze eine frühe Negativierung (N250) und eine späte Negativierung (N400).

*Frühe Negativierung (200-300 ms)*

Bei einem Vergleich der Sprecherbedingungen fällt insbesondere für die Sätze der Bedingung W im Vergleich zur Standard-Bedingung eine deutliche frühe Negativierung an den anterioren Elektroden auf, die als MMN interpretiert wird (siehe Abbildung 10.12). Es wurde für den Zeitbereich 200-300 ms eine globale ANOVA der mittleren Amplituden für die Faktoren SPRECHER (Std, M, W), SEMANTIK (kongruent, inkongruent), HEMISPHÄRE (links, rechts) und REGION (anterior, zentral, posterior) durchgeführt.[6] Es zeigte sich ein signifikanter Haupteffekt SPRECHER ($F_{(2,48)} = 15.55, p < 0.0001$). Dieser konnte auf einen signifikanten Unterschied zwischen den Bedingungen Std und M ($F_{(1,24)} = 6{,}24, p < 0.05$) , sowie Std und W ($F_{(1,24)} = 20.82, p < 0.0001$) zurückgeführt werden. Die Mismatch-Differenzen M-Std und W-Std wurden wiederum einer separaten Analyse unterzogen und wiesen dort ebenfalls signifikante Unterschiede auf ($F_{(1,24)} = 15.4, p < 0.0001$), die auf eine deutlich größere Amplitude der MMN für Bedingung W zurückzuführen war. Die globale ANOVA zeigte desweiteren eine REGION $\times$ SPRECHER Interaktion $F_{(1,24)} = 3.45, p < 0.05$). In allen drei Regionen ergab sich ein signifikanter SPRECHER-Haupteffekt (anterior: $F_{(2,48)} = 14.25, p < 0.0001$; zentral: $F_{(1,24)} = 14.41, p < 0.0001$; posterior: $F_{(1,24)} = 8.65, p < 0.001$). Die Auflösung dieser Haupteffekte zeigt aber nur für die Bedingungen Std und W einen signifikanten Unterschied in allen Regionen (anterior: $F_{(1,24)} = 18.58, p < 0.001$ ; zentral: $F_{(1,24)} = 18.66, p < 0.001$ ; posterior: $F_{(1,24)} = 13.63, p < 0.01$). Die Bedingungen Std und M unterschieden sich hingegen auf dem 5%-Niveau nur an anterioren ($F_{(1,24)} = 7.21, p < 0.05$) und zentralen ($F_{(1,24)} = 5.15, p < 0.05$) Elektroden, nicht aber in der posterioren Region ($F_{(1,24)} = 3.48, p = 0.0743$).

Die Omnibus-Analyse zeigte hingegen weder signifikante Interaktion der Faktoren SPRECHER $\times$ SEMANTIK, noch einen signifikanten SEMANTIK-Haupteffekt. Al-

[6] Für eine Zuordnung der Elektroden zu den Faktoren HEMISPHÄRE und REGION, siehe Experiment IIIb.



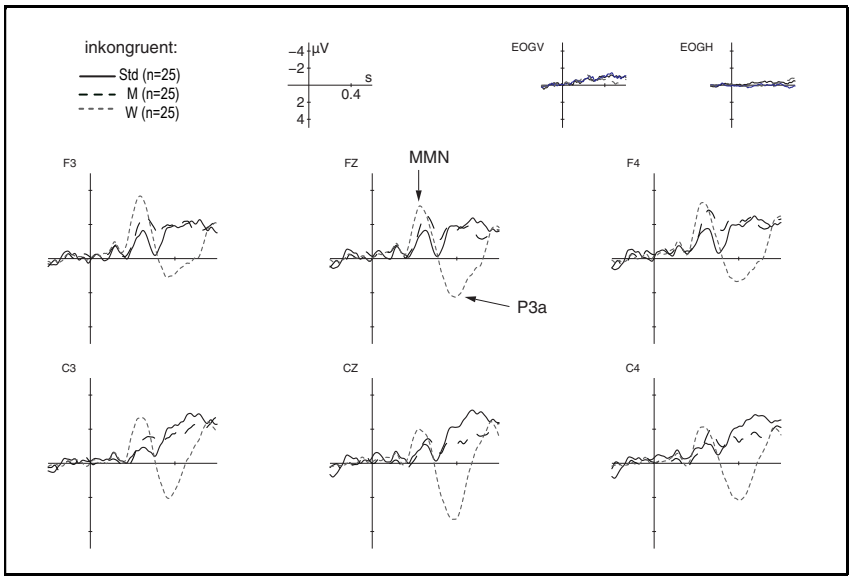




Abbildung 10.12: *Passives Hören* EKPs infolge semantisch kongruenter (oben) und inkongruenter (unten) Sätze der drei verschiedenen Sprecherbedingungen Std, M, und W

lerdings war eine signifikante Interaktion SEMANTIK $\times$ HEMISPHÄRE zu beobachten, die auf einen SEMANTIK-Effekt (N250) in der rechten Hemisphäre ($F_{(1,24)} = 6.92, p < 0.05$), nicht aber in der linken (F<1) zurückgeführt werden konnte.

*Positivierung (350-500 ms)*

Auf die MMN folgte eine Positivierung, die als P300 interpretiert wird. Wie die MMN ist auch sie für Bedingung W am stärksten ausgeprägt. Da sie im Falle der inkongruenten Sätze von einer späteren Negativierung (N400) überlagert war, wurde die statistische Analyse der P300 für kongruente und inkongruente Sätze getrennt vorgenommen. Für das Zeitfenster 350-500 ms wurde eine ANOVA der Faktoren SPRECHER (Std, M, W), REGION (anterior, zentral, posterior) und HEMISPHÄRE (links, rechts) durchgeführt.

Sowohl für die kongruenten als auch für die inkongruenten Sätze waren ein SPRECHER-Haupteffekt (kongruent: $F_{(2,48)} = 21.40, p < 0.0001, \varepsilon = 0.8953$; inkongruent: $F_{(2,48)} = 30.85, p < 0.0001, \varepsilon = 0.7992$), eine Interaktion SPRECHER $\times$ HEMISPHÄRE (kongruent: $F_{(2,48)} = 11.43, p < 0.001, \varepsilon = 0.8918$; inkongruent: $F_{(2,48)} = 4.40, p < 0.05, \varepsilon = 0.0200$), sowie eine Interaktion SPRECHER $\times$ REGION $\times$ HEMISPHÄRE (kongruent: $F_{(2,48)} = 3.29, p < 0.05, \varepsilon = 0.6264$; inkongruent: $F_{(2,48)} = 3.06, p < 0.05, \varepsilon = 0.5843$) zu beobachten. Eine SPRECHER $\times$ REGION Interaktion trat nur für die inkongruenten Sätze auf ($F_{(4,96)} = 4.23, p < 0.05, \varepsilon = 0.5424$).

Der SPRECHER-Haupteffekt wurde aufgelöst und zeigte einen signifikanten Unterschied zwischen den Bedingungen Std und W (kongruent: $F_{(1,24)} = 33.83, p < 0.0001$; inkongruent: $F_{(1,24)} = 32.57, p < 0.0001$) nicht jedoch für Std und M (kongruent: F<1; inkongruent: $F_{(1,24)} = 1.89, p = 0.1816$).

Der Faktor SPRECHER wurde nun aufgeteilt und eine separate Analyse der Bedingungen Std vs. M und Std vs. W. Für die Bedingungen Std vs. M zeigte sich für die kongruenten Bedingungen kein SPRECHER-Haupteffekt und auch keine Interaktionen mit den Faktoren HEMISPHÄRE und REGION (alle F <1), für die inkongruenten einzig eine SPRECHER $\times$ REGION Interaktion $F_{(1,24)} = 6.25, p < 0.05, \varepsilon = 0.6041$), die eine stärkere Positivierung an zentralen Elektroden als an

anterioren und posterioren Elektroden in Bedingung M beschreibt. Die ANOVA der Bedingungen Std vs. W zeigte hingegen einen SPRECHER-Haupteffekt (kongruent: $F_{(1,24)} = 33.83, p < 0.0001$; inkongruent: $F_{(1,24)} = 39.83, p < 0.0001$), keine Interaktion mit dem Faktor REGION (F<1) jedoch eine signifikante Interaktion mit dem Faktor HEMISPHÄRE (kongruent: $F_{(1,24)} = 20.23, p < 0.0001$; inkongruent: $F_{(1,24)} = 8.03, p < 0.01$). Dieser letztere Effekt beschreibt eine größere P300 Amplitude in Reaktion auf die weibliche Stimme (Bedingung W) an den Elektroden über der rechten Hemisphäre.

*Späte Negativierung (440-690 ms)*

Ein N400-Effekt wurde für den späteren Zeitbereich erwartet. Beim direkten Vergleich der EKPs auf semantisch kongruente und semantisch inkongruente Sätze zeigte sich dort für alle drei Sprecherbedingungen eine zentro-parietale Negativierung (siehe Abbildungen 10.13 bis 10.15). Der Zeitbereich 440-690 ms deckt diese Negativierung in allen Sprecherbedingungen ab und wurde daher für die statistische Analyse der mittleren Amplituden and den Elektroden C3, CZ, C4, P3, PZ, P4 ausgewählt. Die einfaktoriellen ANOVAS zeigen dann auch jeweils einen signifikanten Haupteffekt SEMANTIK (Std: $F_{(1,24)} = 33.37, p < 0.0001$; M: $F_{(1,24)} = 36.79, p < 0.0001$; W: $F_{(1,24)} = 20.341, p < 0.0001$). Eine statistische Analyse der Differenzkurven (inkongruent-kongruent; siehe Abbildung 10.16) zeigte einen ELEKTRODEN-Haupteffekt ($F_{(5,120)} = 3.45, p < 0.01, \varepsilon = 0.6260$). Dieser war auf eine größere Amplitude an den Elektroden der Mittellinie (CZ, PZ) zurückzuführen, interagierte aber nicht mit den drei Sprecherbedingungen. Zwischen den Sprecherbedingungen zeigten sich signifikante Unterschiede weder in der mittleren oder maximalen Amplitude, noch in den Gipfellatenzen (Std: 570ms; M: 582ms; W: 571ms).

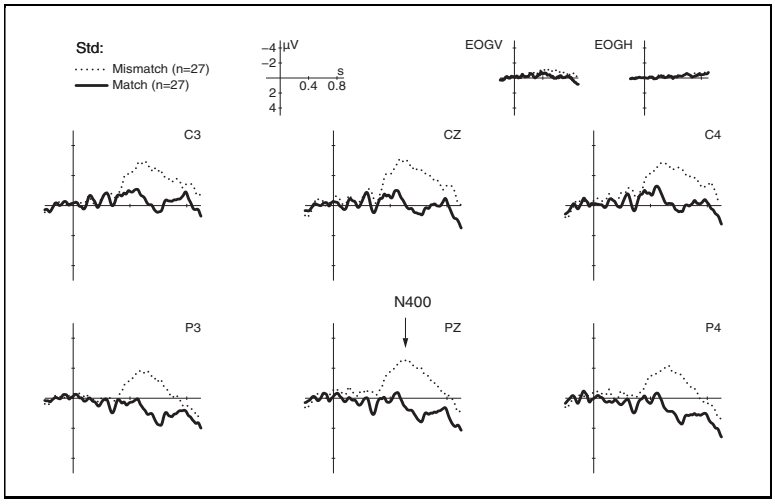



Abbildung 10.13: *Passives Hören* EKPs infolge semantisch kongruenter und inkongruenter Sätze der Sprecherbedingung Std

### 10.5.4 Diskussion

Die Hypothesen für Experiment IIIc werden weitgehend bestätigt. Wie erwartet, tritt ohne Experimentalaufgabe keine aufgabenrelatierter N2b/P3b Komplex auf, während eine Mismatch Negativierung jedoch auch beim passiven Zuhören zu beobachten ist. Im Einklang mit den Hypothesen ist die MMN im Falle der weiblichen Stimme deutlich größer als in den Vergleichsbedingungen, da hier ein salienterer physikalischer und perzeptueller Unterschied vorliegt (vgl. Titova & Näätänen, 2001). Dieser ist auch in der größeren P3a-Komponente für Bedingung W reflektiert. Daß dieser Effekt so deutlich ist, bzw. für den Wechsel vom Standardsprecher zur anderen männlichen Stimme keine nachweisbare P300 auftritt, kann zudem auf die Abhängigkeit der P3a von der subjektiven Häufigkeit eines Ereignisses zurückgeführt werden (vgl. Diskussion des Experiments IIIb).

Ohne den Aufgabenfokus auf der Stimmenverarbeitung scheint der semantische Inhalt für den Hörer an Bedeutung zu gewinnen. Bereits im frühen Zeitfenster ist eine deutliche Negativierung auf semantische Inkongruenz zu beobachten. Wie

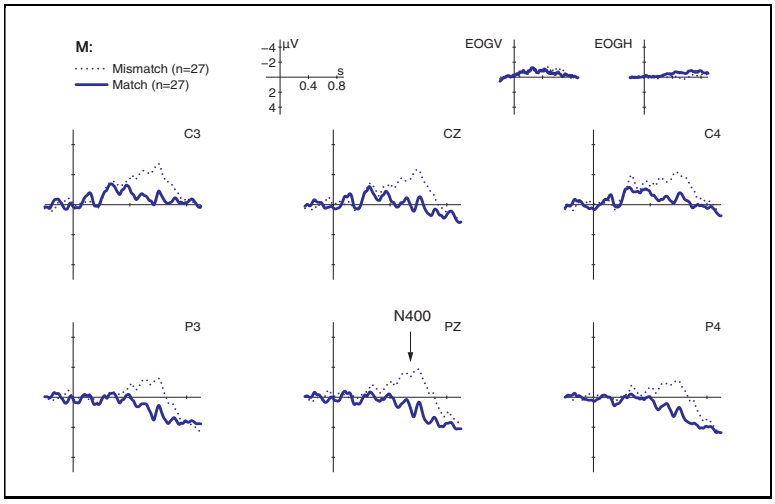



Abbildung 10.14: *Passives Hören* EKPs infolge semantisch kongruenter und inkongruenter Sätze der Sprecherbedingung M

schon im vorangegangene Experiment beschrieben, wird ein solcher früher Effekt der Semantik mit lexikalischen Prozessen in Verbindung gebracht (siehe Hagoort & Brown, 2000) und weist für die Standardbedingung auch die klassische parietale Topographie auf. Weshalb sie im Gegensatz zu den anderen Studien aber eine stärkere Beteiligung der rechten Hemisphäre zeigt, ist unklar. Eine Interaktion mit den Sprecherbedingungen lag nicht vor, diese Negativierung scheint sich also tatsächlich allein auf die sprachliche Verarbeitung zu beziehen.

Wie erwartet, ist in allen Fällen eine deutliche N400 zu beobachten. Es ergibt sich keine Interaktion der N400 mit dem Sprecherwechsel. Da keine Aufmerksamkeitsressourcen auf die bewußte Sprecherunterscheidung verwendet werden mußten, wirken sich die Normalisierungsprozesse offenbar auf die spätere Ebene der semantischen Satzverarbeitung nicht mehr aus.

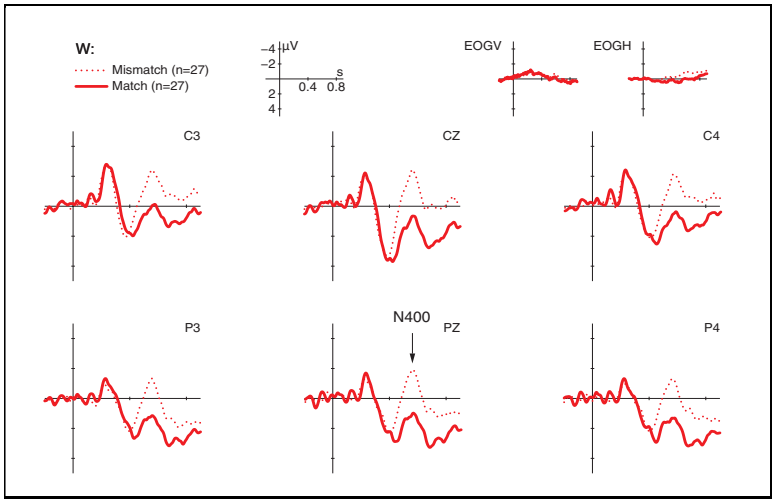



Abbildung 10.15: *Passives Hören* EKPs infolge semantisch kongruenter und inkongruenter Sätze der Sprecherbedingung W

## 10.6 Allgemeine Diskussion der Experimente IIIa, IIIb und IIIc

In einem Reaktionszeitexperiment und zwei EKP-Studien wurde das Zusammenspiel von Sprecherinformation und semantischem Inhalt untersucht. Dazu wurde getestet, wie sich ein unerwarteter Sprecherwechsel im Satz unter verschiedenen Instruktionen auf das Verhalten und die neurophysiologischen Korrelate der Verarbeitung auswirkt.

Zunächst einmal zeigen die vorliegenden Daten, daß es möglich ist, diese schnellen Prozesse mit relativ natürlichem Sprachmaterial, d.h. mit kontiuierlichen Satzäußerungen zu untersuchen: Auch ohne die Verwendung identischer Sprachsignale (vgl. Experimente I und II) wird offenbar eine zentrale Repräsentation des (Standard-)Sprechers aufgebaut, die bei einem Stimmenwechsel verletzt wird. Die neuronalen Korrelate dieser Verletzung sind dann als EKP-Komponenten (N200/P300) meßbar.

Sowohl auf der Verhaltensebene als auch in der elektrophysiologischen Hirnantwort wurden für die semantische Verarbeitung die erwarteten Effekte gefunden:

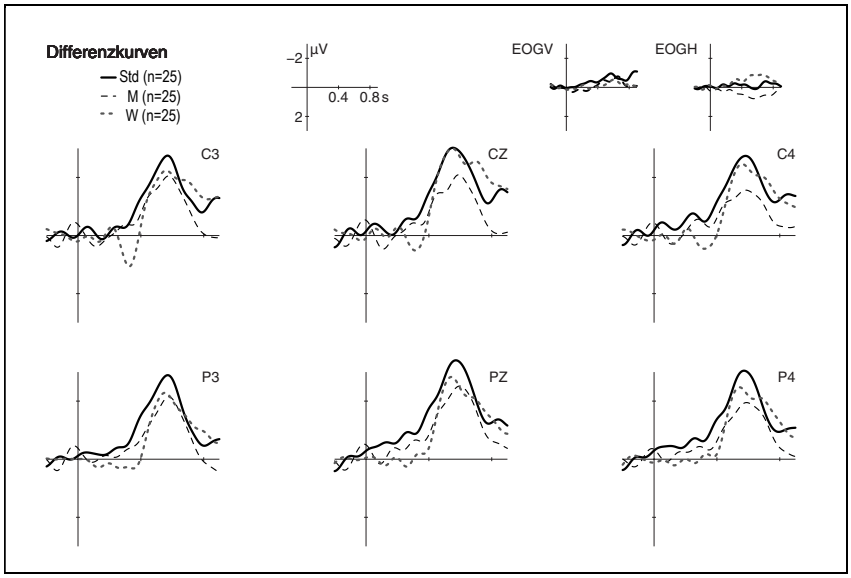



Abbildung 10.16: *Passives Hören* Differenzkurven der EKPs infolge der jeweils semantisch kongruenten und inkongruenten Sätze der drei Sprecherbedingungen (für diese Abbildung mit 7 Hz tiefpaßgefiltert. Es liegen keine statistisch signifikanten Unterschiede zwischen den Sprecherbedingungen vor.)

Die kürzeren Reaktionszeiten für die kongruenten Sätze sind auf ein Priming des kontextuell angemessenen Wortes zurückzuführen; die N400-Komponente in den EKP-Experimenten deutet auf eine Schwierigkeit der semantischen Integration kontextuell inkongruenter Wörter. Die inhaltsorientierte Verarbeitung scheint hierbei der Defaultmodus der Sprachwahrnehmung zu sein, denn die Semantik-Effekte treten auch dann auf, wenn die Primäraufgabe nur eine Sprecherunterscheidung verlangt. Dieses Primat der semantischen Verarbeitung wird vermutlich noch dadurch verstärkt, daß der Sprecherwechsel innerhalb eines Satzes dargeboten wurde. So wäre zu testen, ob sich die Effekte beispielsweise auch bei der Darbietung von Wortpaaren zeigen (auf die damit verbundenen Probleme wurde aber bereits im Abschnitt "Allgemeinen Fragestellung" verwiesen).

Zur Untersuchung einer Interaktion semantischer und stimmrelatierter Verarbeitungsprozesse ist, wie in der vorangegangen MEG-Studie, vor allem die Doppelverletzung, d.h. die Kombination aus semantischer Inkongruenz und Spre-

cherwechsel, aufschlußreich. Sollte im Reaktionszeitexperiment eine semantische Beurteilung der Sätze durchgeführt werden, so zeigte sich auf Verhaltensebene eine beschleunigte Ablehnung inkongruenter Sätze, wenn diese auch einen Sprecherwechsel beinhalteten. Bei einer Sprecherdiskriminationsaufgabe fand eine schnellere Detektion des Sprecherwechsels bei semantischer Inkongruenz statt. Der scheinbare Unterschied zwischen den Sprecherbedingungen, d.h. der schwächere Einfluß der Semantik aufgrund längerer Reaktionszeiten für kongruente Sätze der Bedingung M, konnte durch die Performanzdaten des Experiments IIIb erklärt werden. Auch wenn die Probanden nicht unter Zeitdruck standen, wurden gerade in dieser Bedingung signifikant mehr Fehler gemacht als in den anderen. Die Ähnlichkeit zwischen den Männerstimmen gerade bei den Experimentalreizen dieser Bedingung hat zumindest für einen bedeutenden Anteil der Probanden im Reaktionszeitexperiment zu längeren Reaktionzeiten geführt. Grundsätzlich deuten die Reaktionszeitergebnisse zunächst darauf hin, daß beide Informationsdimensionen interagieren.

Im EKP-Experiment IIIb, bei dem ebenfalls die Detektion eines Sprecherwechsels durchzuführen war, konnte dieser Befund auf neurophysiologischer Ebene jedoch nicht bestätigen. Es war keine Interaktion von Sprechstimme und semantischer Kongruenz zu beobachten. Möglicherweise war der Einfluß des Sprecherwechsels auf die N400-Komponente, aufgrund des ungünstigen Signal-Rausch-Verhältnisses statistisch nicht faßbar. Sollte ein derartiger Einfluß vorliegen, so muß es sich um einen aufgabenbedingten Effekt handeln, wie das darauffolgende EKP-Experiment verdeutlicht. In dieser Untersuchung (Experiment IIIc) war mit der Satzpräsentation keine explizite Exprimentalaufgabe verbunden (passives Zuhören). Ohne Experimentalaufgabe waren keine aufgabenabhängigen EKP-Komponenten (N2b/P3b) zu beobachten. Die Mismatch Negativierung für den Sprecherwechsel erwies sich nun als Korrelat der physikalischen/perzeptuellen Ähnlichkeit d.h. sie war für den Wechsel zur weiblichen Stimme größer, als für den Wechsel zwischen den männlichen Sprechern. Eine Interaktion war weder in diesem frühen Zeitbereich um 250 ms noch für die N400-Komponenten der verschiedenen Sprecherbedingungen zu beobachten. Dieses Resultat läßt vermu-

ten, daß es sich auch bei den im Reaktionszeitexperiment beobachteten Effekten um Einflüsse zu handelt, die vor allem mit den Entscheidungsprozessen der Experimentalaufgabe (Ja/Nein-Antwort) zusammenhängen.

Die EKP-Experimente zeigen, daß sich bei der Sprachwahrnehmung die Detektion des Sprecherwechsels in einer Mismatch-Negativierung manifestiert, während die semantische Integration in der deutlich späteren N400-Komponente reflektiert ist. Weder im frühen Zeitfenster der MMN noch im späteren Zeitfenster der N400 sind wechselseitige Einflüsse von Sprecherinformation und semantischer Kongruenz zu beobachten. Eine Interaktion, wie sie z.B. in Experiment II für das Auftreten eines Sprecherwechsels und die phonetische Verarbeitung berichtet wurden, waren hier nicht zu beobachten. Damit weisen die Ergebnisse darauf hin, daß es sich bei der Verarbeitung von Sprecherinformation und der Bewertung semantischen Inhalts um zwei unabhängige Prozesse handelt. In der visuellen Domäne wurden ähnliche Befunde für einen Wechsel der Schriftgröße beim Lesen von Sätzen berichtet (Kutas & Hillyard, 1980a). Auch dort wurden eine N400 Komponente für die semantische Inkongruenz und eine P300 Komponente für den physikalischen Mismatch beobachtet, die nicht miteinander interagierten. Obwohl es sich bei der Verarbeitung von Stimminformation um eine evolutionsbiologisch bedeutsamere Funktion handelt als die Buchstabengröße beim Lesen, so scheint die semantisch-inhaltliche Verarbeitung jedoch in beiden Fällen unabhängig von Verletzungen der Sprecher-/Schriftkontinuität zu verlaufen. Das bedeutet auch, daß sich zumindest bei prototypischen Stimmen eine Normalisierung bzw. eine effektive Extraktion invarianter Parameter ohne Kosten für das Sprachverstehen vollzieht.

# Teil IV

# Zusammenfassung und abschließende Diskussion

# Kapitel 11

# Abschließende Diskussion

## 11.1 Theoretischer Ausgangpunkt

Die gesprochene Sprache stellt die wichtigste Form der menschlichen Kommunikation dar. Ihre Erforschung beschränkt sich allerdings hauptsächlich auf ihren funktionellen Charakter, d.h. auf die Aspekte, die dem strukturellen und inhaltlichen Verständnis des Gesagten dienen. Die Bedeutung der ebenfalls im Sprachsignal enthaltenenen Stimminformation ist, obgleich ihre Verarbeitung eine grundlegende und auch phylogenetisch bedeutsame kognitive Fähigkeit darstellt, in den Neurowissenschaften wenig untersucht worden. Dafür mag es vor allem zwei Gründe geben. Zum einen besteht in der alltäglichen Kommunikationssituation zwischen Sprecher und Hörer ein Sichtkontakt. Die Einschätzung der Sprechers erfolgt also nicht allein über das auditorische System, sondern auch über die visuelle Modalität. (Besondere Bedeutung gewinnt die Stimminformation für den Hörerallerdings, wenn die normalen Kommunikationsbedingungen eingeschränkt sind, z.B. beim Radiohören, bei Hören von Hörspielen, u.U. beim Telefonieren und immer dann, wenn ein Sichtkontakt zwischen den Kommunikationspartnern nicht möglich ist.)
Eine zweite wichtige Ursache für die Vernachlässigung der Sprecherinformation ist in ihrer mangelhaften theoretisch-formalen Beschreibung zu suchen. Sprache

läßt sich als ein strukturiertes System diskreter linguistischer Einheiten beschreiben, die in regelhafter Weise zu sprachlichen Äußerungen kombiniert werden. Eine vergleichbare Beschreibung liegt für die Stimminformation nicht vor. Zwar gibt es Erweiterungen des Internationalen Phonetischen Alphabets und auch andere Klassifizierungssysteme, mit denen eine normierte Beschreibung von Stimmen geregelt werden soll, doch beruhen diese Klassifikationen meist auf eher impressionistischen Begriffen ("helle" vs. "dumpfe" Stimme) und sind daher bei ihrer Anwendung durch verschiedene Hörer wenig reliabel (Laver, 1980).

Erst in letzter Zeit wurden auch neurophysiologische Untersuchungen zur Wahrnehmung der Stimminformation durchgeführt. Diese Studien, die sich beispielsweise der funktionellen Magnetresonanztomographie bedienten (Belin et al., 2000, 2002; Binder et al., 2000), deuten darauf hin, daß die menschliche Stimme akustische Eigenarten besitzt, die in einer spezifischen neuronalen Verarbeitung reflektiert sind, wobei insbesondere die Bedeutung des STG/STS[1] herausgestellt wurde. Teilstrukturen dieser Hirnregionen, die in den Untersuchungen eine starke Aktivierung aufwiesen, wurde daher als *stimmselektive Areale* bezeichnet worden (Belin et al., 2000).[2] Unterstützende Befunde liefern klinische Studien, die zeigen, daß die Diskrimination verschiedener Stimmen bei Läsionen dieser Hirnregionen gestört ist (Van Lancker & Kreiman, 1987; Van Lancker et al., 1988, 1989; Ziegler & Kraus, 1999). Als einen zweiten wichtigen Befund berichten klinische Untersuchungen (Ziegler & Kraus, 1999), wie auch Verhaltensexperimente (z.B. Mullenix et al., 1989; Mullenix & Pisoni, 1990; Nygaard et al., 1994; Goggin et al., 1991) eine Interaktion der Stimm- und Sprachverarbeitung. Es ist jedoch unklar, welcher kognitiven Verarbeitungsebene diese Interaktion zugeordnet werden muß.

Die vorliegende Arbeit befaßte sich mit diesen Problemen und verfolgte drei Fragestellungen, denen jeweils ein Abschnitts des empirischen Teils gewidmet war. In der Experimentalserie I wurde unter Verwendung der Magnetenzephalographie die *Spezifität* der präattentiven Stimmwahrnehmung untersucht. Experiment II be-

[1] STG: superiorer temporaler Gyrus; STS: superiorer temporaler Sulcus

[2] Es stellt sich allerdings die Frage, ob diese Hirnregionen tatsächlich selektiv auf menschliche Stimmen reagieren oder auch auf andere Signale einer bestimmten akustischen Struktur; d.h. es bleibt abzuwarten, ob eine derartige Aktivierung funktional oder formal zu erklären ist, wenn sich diese Ansätze überhaupt für die Verarbeitung von Stimminformation trennen lassen.

faßte sich mit der Interaktion von Sprechstimme und phonetischer Information. Im dritten Abschnitt (Experiment IIIa-c) wurde das Zusammenspiel von Stimmwahrnehmung und inhaltlich-semantischer Verarbeitung untersucht.

## 11.2 Methodischer Hintergrund

Als Untersuchungsmethoden wurden mit der Elektroenzephalographie (EEG) und der Magnetenzephalographie (MEG) zwei Verfahren gewählt, die eine nicht-invasive Beobachtung auditorischer Verarbeitungsprozesse, im Millisekundenbereich erlauben.

In den MEG-Studien der vorliegendenden Arbeit (Experiment I und II) wurde zur Untersuchung der Stimmperzeption ein Oddball-Paradigma verwendet. Bei diesem Paradigma wird durch die iterative Präsentation eines Standardstimulus eine Gedächtnisspur im sensorischen Gedächtnis aufgebaut; tritt eine devianter Stimulus auf, so evoziert dieser ein charakteristisches ereigniskorreliertes Hirnpotential, die Mismatch Negativierung (MMN), bzw. im MEG deren magnetisches Korrelat, die MMNm (Näätänen, Gaillard & Mäntysalo, 1978). Da sich die Mismatch Negativierung als sehr robust gegenüber verschiedenen Aufmerksamkeitsmanipulationen erwiesen hat, wird sie als ein Korrelat der *präattentiven*, d.h. vorbewußten auditorischen Wahrnehmung interpretiert (Näätänen, 1992). Die Experimentalserie III, die mit Hilfe ereigniskorrelierter Hirnpotentiale (EKP) Einflüsse der Stimme auf die Satzverarbeitung untersuchte, ergänzt das Oddball-Paradigma, indem sie zeigt, daß die in den MEG-Studien beobachteten Effekte nicht von der iterativen Präsentation der Standardreize und Devianten abhängig sind.

## 11.3 Zusammenfassende Diskussion der Ergebnisse

**Experimentalserie I: Prototypikalitätseffekte bei der Stimmwahrnehmung**

Experiment Ia untersuchte die präattentive Reaktion auf Stimmen unterschiedlicher Prototypikalität. Ziel war es, herauszustellen, inwieweit eine Mismatch-Reaktion auf einen Sprecherwechsel, wie sie von Titova & Näätänen (2001) beobachtet wurde, tatsächlich eine *stimmspezifische* Verarbeitung widerspiegelt, d.h. inwieweit diese auf Langzeitrepräsentationen prototypischer Stimmen beruht, oder auf unspezifische auditorische Verarbeitungmechanismen zurückgeführt werden muß. Dazu wurden im Oddball-Experiment neben einer männlichen Stimme als Standardstimulus (M_Std) drei Devianten präsentiert: eine natürliche weibliche Stimme (W), eine grundfrequenzmanipulierte, ungewöhnlich tiefe weibliche Stimme (G) und ein unnatürliches, synthetisiertes Sprachsignal (S). Die Stimuli bestanden dabei jeweils aus demselben Einzelwort (*Dach*). Die magnetischen Korrelate der neurophysiologischen Stimulusverarbeitung wurden mittels eines Magnetenzephalographen aufgezeichnet und durch eine Stromdichtemodellierung analysiert. Alle Devianten evozierten wie erwartet eine Mismatch Reaktion. Es war ein Anstieg der Stromdichte zu beobachten, der sich umgekehrt zur Prototypikalität der Stimmen verhielt (die natürliche weibliche Stimme evozierte die schwächste, die synthetische Stimme die stärkste Mismatch-Reaktion). Dieser Anstieg war unabhängig von der physikalischen Ähnlichkeit: die tiefe Frauenstimme evozierte eine höhere Mismatch-Reaktion als die natürliche weibliche Stimme, obwohl sie der Männerstimme (M_Std) aufgrund derselben Grundfrequenz akustisch ähnlicher war. Experiment Ib bestätigte die Beobachtungen der ersten Studie. In diesem Experiment wurde die untypische, tiefe Frauenstimme als Standardstimulus eingesetzt, so daß ein direkter Vergleich der Reaktionen auf die natürliche männliche Stimme (im Vergleich zum Standard eine Abweichung im Formantspektrum), und der weiblichen Stimme (Abweichung in der Grundfrequenz) möglich war. Die ereigniskorrelierten Stromdichten der Mismatch-Reaktion unterschieden sich nicht signifikant zwischen diesen Devianzbedingungen. Aus diesen Ergebnissen wurde gefolgert, daß für die Stimmwahrnehmung offenbar

keine Hierarchie der untersuchten akustischen Parameter (Formanten, F0) existiert, sondern daß ihre Konfiguration, d.h. der holistische Gesamteindruck die Mismatch-Reaktion bestimmt. Ob diese Annahme für die auditorische Verarbeitung komplexe Signale im allgemeinen gilt (siehe z.B. Paavilainen et al., 2001b; Deacon et al., 1998, für eine Gegenposition), muß zukünftige Forschung erweisen. Hinweise auf eine solche holistische Verarbeitung finden sich allerdings auch in der Forschung zur Gesichtswahrnehmung (Lewis & Johnston, 1997; Hancock, Bruce & Burton, 2000). Diese Beobachtung und die Vielzahl von Parallelen zwischen Gesichts- und Stimmwahrnehmung (siehe z.B. Mann, Diamond & Carey, 1979) geben Anlaß zu der Vermutung, daß es sich bei der konfiguralen Verarbeitung um einen domänenübergreifenden Mechanismus handelt, der immer dann eine besondere Rolle spielt, wenn eine Vielzahl von Kombinationen ähnlicher Basisparameter unterschieden werden muß.

Eine Abhängigkeit von Langzeitrepräsentationen, wie sie die Experimente Ia und Ib für die Stimmwahrnehmung nahelegen, wurde in Mismatch-Experimenten auch für Phonemdevianten berichtet. Wird im Oddball-Design jedoch ein für den Hörer nicht-prototypischer Sprachlaut als Deviant präsentiert, d.h. ein Sprachlaut, der in den vom Hörer beherrschten Sprachen nicht auftritt und von ihm auch behavioral nicht vom Standard unterschieden werden kann, so ist die Amplitude der ereigniskorrelierten Mismatch Negativierung *geringer* als die der MMN auf prototypische Phonemdevianten (Näätänen, Lehtokoski, Lennes, Cheour, Houtilainen, Ilvonen et al., 1997). Dieser Effekt tritt unabhängig vom Ausmaß der akustischen Devianz auf. Wie von (Näätänen et al., 1997) vorgeschlagen, liegt der Grund für die stärkere Aktivierung bei prototypischen Phonemdevianten darin, daß ein phonetischer Kategoriewechsel zu zusätzlicher Mismatch-Aktivierung führt. Ein solcher Kategoriewechsel wird aber nur bei prototypischen Devianten wahrgenommen, für nicht-prototypische Phoneme fällt die Mismatch-Reaktion daher geringer aus. Da hingegen für nicht-prototypische Stimmen ein Anstieg in der Mismatch-Reaktion beobachtet wurde (Experiment Ia und Ib), ist davon auszugehen, daß sich die Mismatch-Prozesse auf nicht-prototypische Exemplare für die beiden Dimensionen (Stimmen, Sprachlaute) qualitativ unterscheiden. Der Frage, ob ein

solcher Unterschied auch in der Verarbeitung natürlichsprachlicher Signale nachzuweisen ist, bzw. wie sich die Verarbeitungsprozesse bei Abweichungen in der einen und der anderen Dimension zueinander verhalten, wurde in Experiment II nachgegangen.

**Experiment II: Stimminformation und phonetische Verarbeitung**

In der Literatur wird sowohl für einem Sprecherwechsel (Titova & Näätänen, 2001) als auch für einen Phonemwechsel (z.B. Alho, Winkler, Escera, Huotilainen, Virtanen, Jääskelainen et al., 1998b; Diesch & Luce, 1997) eine Mismatch-Negativierung berichtet. Bisher wurden diese Effekte jedoch immer in separaten Studien untersucht und nie direkt miteinander verglichen (kein Within-Subject-Design). Es ist daher unbekannt, wie sich die Mismatch-Prozesse bei einer phonetischen und einer sprecherbezogenen Abweichung zueinander verhalten. Eine bedeutsame Fragestellung ist dabei, ob es sich um serielle oder parallele Prozesse handelt, und ob die Stimulus-Dimensionen unabhängig voneinander oder integrativ verarbeitet werden. Dieser Fragestellung war Experiment II gewidmet. Dazu wurden im Oddball-Paradigma nicht nur Abweichungen der Phoneme (Ph) und der Sprechstimme (Sp), sondern auch eine zweidimensionale Abweichung beider Faktoren (PhSp) präsentiert. Die ereigniskorrelierten magnetischen Felder wurden mit dem MEG aufgezeichnet und im Rahmen einer Dipolmodellierung analysiert. Die Auswertung zeigte weder im Hinblick auf die Lokalisation noch auf die bilaterale Hemisphärenverteilung des Mismatch-Effekts signifikante Unterschiede zwischen den Bedingungen. In den Latenzen der Mismatch Reaktion zeigte sich ebenfalls kein Unterschied zwischen den eindimensionalen Devianten (Ph, Sp), was als ein erster Hinweis auf die parallele (d.h. zeitgleiche) Verarbeitung der beiden Informationsarten interpretiert werden kann. Von besonderer Bedeutung war für die Fragestellung des Experiments II die zweidimensionale Devianz (PhSp). In Bezug auf die Dipolstärke waren hier keine Unterschiede zu den eindimensionalen Abweichungen zu beobachten. Diese Subadditivität deutet bereits auf eine integrative Verarbeitung der Dimensionen hin. Weitere Unterstützung

erfährt dieser Befund durch die Ergebnisse der Latenzanalyse. So zeigte sich eine deutliche Verringerung der Latenz, sowie eine Reduktion ihrer Varianz für die gleichzeitige Abweichung von Phonem *und* Stimme im Vergleich zu den eindimensionalen Devianzen (Phonem *oder* Stimme). Ein ähnliches Muster ist auf der Verhaltensebene in den Reaktionszeiten bei der Diskrimination (same-different-judgement) zu finden. Auch dort ist für die Paarung aus Stimme und Wort bei der Diskrimination ein Redundanzeffekt zu beobachten, d.h. kürzere und weniger variable Reaktionszeiten (Mullenix & Pisoni, 1990; Wood, 1974). Nach den Grundannahmen des Speeded Classification Paradigmas (Garner, 1974) deutet dies auf eine nicht-unabhängige Verarbeitung zweier Dimensionen hin. Die neurophysiologischen Daten des Experiments II unterstützen diese Annahme für die präattentive Verarbeitung.

Wenn jedoch in dieser frühen Perzeptionsphase eine integrative Verarbeitung der Dimensionen stattfindet, so läßt sich daraus die Frage ableiten, zu welchem Zeitpunkt eine Trennung der Verarbeitungsprozesse stattfindet und ob diese neurophysiologisch lokalisiert werden kann. Geht man davon aus, daß eine frühe integrative Verarbeitung auch zu integrativen Langzeitrepäsentationen (*Templates*) führt, wie sie z.B. im Rahmen der Exemplar-Modelle (z.B. Johnson, 1997) postuliert werden, so ließen sich damit Effekte wie z.B. das bessere Verstehen von maskierter Sprache für bekannte als für unbekannte Stimmen (Nygaard et al., 1994) erklären: Das Wiedererkennen von Sprachlauten ist einfacher, wenn Testreiz und Musterrepräsentation viele gemeinsame akustische Details aufweisen. Evidenz für einen Normalisierungsprozeß im Sinne einer Transformation auf eine rein linguistische/symbolische Ebene, wie ihn abstraktionistische Modelle der Sprachverarbeitung postulieren (z.B. Ladefoged & Broadbent, 1957; Sussman, 1986), lieferte Experiment II hingegen nicht. In einer aktuellen, noch unpublizierten Studie, deren erste Daten im Internet verfügbar gemacht wurden, versuchen Orrin, Nusbaum & Small (2001) den Normalisierungsprozeß im auditorischen Kortex oder im limbischen System zu lokalisieren. Diese Studie hat aber das intrinsische Problem, daß auch akustische Unterschied zwischen den Stimuli, d.h. zwischen Wortlisten mit und ohne Sprecherwechsel, und möglicherweise

damit verbundene unspezifische Aufmerksamkeitsprozesse abgebildet werden. Diese dürften von einer "reinen" Sprechernormalisierung nur schwer zu dissoziieren sein.

Unabhängig davon, ob sich ein solcher Normalisierungsprozeß nachweisen und lokalisieren läßt, oder andere, eher exemplar-basierte Mechanismen die Dekodierung der Sprachlaute besser erklären, so zeigt insbesondere Experimentalserie III, daß das semantische Verstehen von einem Wechsel der Sprechstimme nicht betroffen ist.

#### Experimentalserie III: Stimminformation und semantisches Satzverstehen

Die folgenden drei Experimente (Experimente III a-c) untersuchten, inwieweit die Stimminformation einen Einfluß auf das semantische Satzverstehen hat. Es wurde in diesen Experimenten folgendes Stimulusmaterial verwendet: Als Kontrollbedingung wurden semantisch kongruente Sätze dargeboten, die von einem männlichen Standard-Sprecher (Std) geäußert wurden. Daneben wurden Sätze präsentiert, die eine Abweichung beinhalteten. Diese Abweichung bestand entweder in einer semantischen Verletzung auf dem satzfinalen Wort (*"Der Vegetarier hat das frische Obst gebügelt."*) oder in einem Sprecherwechsel, wobei dann das satzfinale Wort von einem anderen männlichen Sprecher (M) oder einer weiblichen Sprecherin (W) geäußert wurde. Auch eine Kombination aus Sprecherwechsel und semantischer Verletzung wurde dargeboten. Mit diesen Sätzen wurde zunächst ein Reaktionszeitexperiment mit zwei Probandengruppen durchgeführt (Experiment IIIa). Die erste Probandengruppe hatte die Aufgabe, unabhängig von der semantischen Kongruenz einen möglichen Sprecherwechsel zu detektieren. Die Detektion des Sprecherwechsels war deutlich einfacher, wenn dieser einen Wechsel im Geschlecht des Sprechers umfaßte, als wenn zwei männliche Sprecher voneinander zu unterscheiden waren. Weiterhin zeigte sich eine Verzögerung der Antwortzeiten für semantisch inkongruente Sätze, obwohl die semantische Verarbeitung für die Durchführung der Experimentalaufgabe irrelevant war. Die zweite Probandengruppe sollte die semantische Kongruenz unabhängig von

einem möglichen Sprecherwechsel beurteilen. Hier zeigte sich ebenfalls eine schnellere Reaktion für semantisch kongruente als für inkongruente Sätze. Im Vergleich zur Standard-Bedingung (kein Sprecherwechsel), war für Sätze, die sowohl einen Sprecherwechsel als auch eine semantische Inkongruenz beinhalteten, eine schnellere Detektionsleistung zu beobachten.

Wichtige Ergebnisse der Reaktionszeitstudie sind also (1) die schneller Detektion des Sprecherwechsels, wenn dieser einen Wechsel des Geschlecht des Sprechers umfaßt; (2) das Auftreten des semantischen Priming-Effekts (d.h. kürzere Reaktionen bei der Beurteilung kontextuell angemessener als unangemessener Wörter vgl. Marslen-Wilson & Tyler, 1980) auch dann, wenn die semantische Verabeitung nicht verlangt war, sondern die Experimentalaufgabe in der Sprecherdiskrimination bestand; (3) die schneller Detektion einer semantischen Verletzung, wenn diese mit einem Wechsel des Sprechers einherging. Dieser letzte Befund deutet darauf hin, daß die Stimmverarbeitung einen Einfluß auf die semantische Beurteilung im Reaktionszeitexperiment zu haben scheint. Grundsätzlich scheint das Verarbeitungssystem aber auf die Dekodierung des semantischen Inhalts angelegt zu sein, da die präsentierten Sätze unabhängig von der Experimentalaufgabe, stets auch inhaltlich interpretiert wurden.

Es wurde mit denselben Stimulusbedingungen eine EKP-Studie durchgeführt (Experiment IIIb), in der ebenfalls ein möglicher Sprecherwechsel detektiert werden sollte. Obwohl der semantische Inhalt für diese Experimentalaufgabe irrelevant war, trat ein N400-Effekt als Korrelat der erschwerten semantischen Verarbeitung inkongruenter Wörter (Kutas & Hillyard, 1980a,b) auf. Damit kann die semantische Verarbeitung im Kontext einer Sprecherdiskriminationsaufgabe als obligatorischer Prozeß gekennzeichnet werden. Der N400-Effekt trat jedoch in gleicher Weise für alle Sprecherbedingungen (Std, M, W) auf, war also unabhängig vom Vorhandensein eines Sprecherwechsels. In Reaktion auf einen Wechsel der Sprechstimme war in einem früheren Zeitbereich ein N200/P300-Komplex zu beobachten. Es wird davon ausgegangen, daß dieser EKP-Komplex aus Subkomponenten besteht, die einerseits die auditorische Wahrnehmung des Sprecherwechsels reflektieren (MMN, P3a), andererseits aber auch die Relevanz und die

Schwierigkeit der Stimuli im Hinblick auf die Experimentalaufgabe (N2b, P3b; siehe hierzu auch Pritchard, 1991; Pritchard, Shappell & Brandt, 1991; Comchero & Polich, 1999, 1998). Um die aufgabenbedingten Effekte auszuschließen, wurde das EKP-Experiment ohne explizite Experimentalaufgabe wiederholt (passives Zuhören).

In diesem Experiment (Experiment IIIc) zeigte sich im frühen Zeitbereich eine MMN, die eine stärkere Amplitude für den Wechsel vom Standardsprecher zur weiblichen Stimme aufwies als bei dem Wechsel zur devianten männlichen Stimme. Auch eine P3a-Komponente, die den physikalischen Kontrast und möglicherweise einen damit verbundenen Orientierungsprozeß widerspiegelt, war ebenfalls nur für den Wechsel zu weiblichen Stimme zu beobachten. Dieses Ergebnis ist nicht nur in Übereinstimmung mit dem Reaktionszeitexperiment, bei dem eine eine schnellere Reaktion für den salienteren Stimmenkontrast zu beobachten war, sondern zeigt darüber hinaus, daß eine iterative Stimuluspräsentation im Oddball-Paradigma offenbar nicht notwendig ist, um eine MMN für eine abweichende Stimme zu evozieren. Eine Verletzung der Standardsprecher-Repräsentation, wie sie durch den Satzbeginn aufgebaut wird, führt offenbar ebenfalls zu einer Mismatch Negativierung.

Im Vergleich der semantisch kongruenten und inkongruenten Sätze traten zwei Negativierungen auf: Eine N250, die mit semantischen Selektionsprozessen in Verbindung gebracht wird (van den Brink, Brown & Hagoort, 2001) und ein N400-Effekt. Diese unterschieden sich jedoch nicht zwischen den Sprecherbedingungen. Damit zeigte sich im EKP weder bei der aktiven Detektion des Sprecherwechsels (Experiment IIIb) noch beim passiven Zuhören (Experiment IIIc) eine Interaktion von Semantik und Sprechstimme. Die Effekte aus dem Verhaltensexperiment werden daher auf die spezifischen Anforderungen des Reaktionszeitparadigmas zurückgeführt. Die ereigniskorrelierten Potentiale weisen hingegen darauf hin, daß beim einfachen Zuhören die Stimminformation ohne Verarbeitungskosten für die semantische Integration verarbeitet wird. Die Unabhängigkeit der N400-Komponente gegenüber dem unerwarteten Sprecherwechsel bedeutet allerdings nicht, daß die Information einer (Standardsprecher-Repräsentation kei-

nen Einfluß auf die Satzinterpretation haben kann, wenn sie implizite Kontextinformation trägt. In einer EKP Studie von Van Berkum, Haagoort & Brown (1999) wurden den Probanden Kontextinformationen in Form kurzer Texte präsentiert. Der letzte Satz des jeweiligen Textes war in Bezug auf diese Kontextinformation entweder kongruent oder inkongruent. Für die Wörter, die eine semantisch inkongruente Information lieferten, fanden die Autoren einen N400-Effekt. Kontextinformation moduliert also die semantische Integration und damit auch die N400-Komponente. Da nun auch die Sprecheridentität ein Teil der Kontextinformation in eine Kommunikationssituation ist, kann die Vorhersage getroffen werden, daß in einem Experiment selbstreferenzielle Aussagen eines Sprechers, die mit dem, was der Hörer über den Sprecher weiß, inkompatibel sind, ebenfalls einen N400-Effekt auslösen.[3] Solche Effekte sind aber fundamental verschieden von den auditorischen Verarbeitungsprozessen bei einer plötzlichen Stimmdevianzen, wie sie in der vorliegenden Arbeit untersucht wurden. Diese sind eher vergleichbar mit den Studien von Kutas & Hillyard (1980a) in der visuellen Domäne, die zeigten, daß ein plötzlicher Wechsel des Schrifttyps keinen Einfluß auf den N400-Effekt beim Lesen eines Satzes hat. Sowohl in der Schrift als auch bei der Stimmwahrnehmung ist das Perzeptionssystem offenbar in der Lage, solche physikalischen Abweichungen effektiv auszugleichen, um das Verstehen des Kommunikationsinhalts zu gewährleisten.

[3]Dazu ist nicht nötig, daß die Identität des Sprechers bekannt ist. Der Effekt sollte z.B. auch bei Verletzungen geschlechtsbezogener Stereotypen auftreten. So sollten in dem Satz *"Morgen kaufe ich mir ein neues Kleid."* das Wort *"Kleid"* eine größere N400 Komponente evozieren, wenn der Satz von einer männlichen, statt einer weiblichen Stimme geäußert wurde.

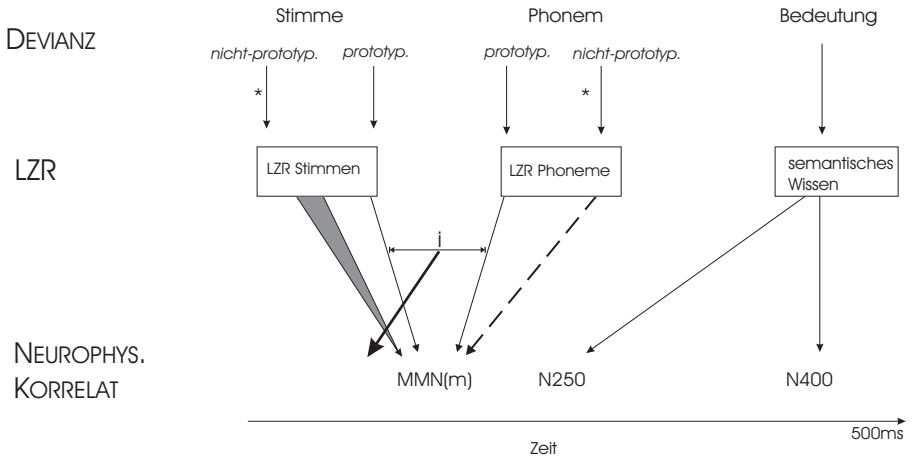



Abbildung 11.1: Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformation - Zusammenfassung der Experimente I-III: In der Experiementalserie I wurden im Oddball-Paradigma prototypische Stimmen als Devianten präsentiert, sowie nicht-prototypische, die eine Verletzung (dargestellt durch $*$) der Langzeitrepräsentation (LZR) natürlicher Stimmen bewirken. Die Verletzung der LZR durch eine nicht-prototypische Stimme führt zu einer stärkeren MMNm, als bei einer prototypischen Stimmdevianten. Die Interaktion zwischen Stimmwahrnehmung und phonetischer Perzeption war Untersuchungsgegenstand in Experiment II. Die Abbildung zeigt die qualitativen Unterschiede in den Mismatch-Reaktionen auf deviante Stimmen und Phoneme. Eine nicht-prototyische Stimme führt zu einer verstärkten MMNm (Experiment I), ein nicht-prototypisches Phonem hingegen zu einer schwächeren MMNm (vgl. Näätänen et al., 1997, hier ist dieser schwächere Mismatch-Effekt als gestrichelte Linie dargestellt). Experiment II zeigte, daß die MMN für die zweidimensionale Abweichung (Wort *und* Stimme) nicht unabhängig ist, sondern auf einer integrativen Verarbeitung (i) der Reizdimensionen basiert, die zu einer früheren Mismatch-Reaktion als bei den eindimensionalen Devianzen (Wort *oder* Stimme) führt. Die Experimente IIIb und c zeigten schließlich, daß eine Inkongruenz von Wortbedeutung und semantischem Satzkontext zu zwei EKP-Komponenten führt: zu einer N250, die als Korrelat einer erschwerten lexikalischen Selektion des Wortes im mentalen Lexikon interpretiert wird und zu einer N400 die eine erschwerte semantische Integration abbildet. Die Experimente liefern keine Hinweise auf eine Abhängigkeit dieser späteren Komponenten von frühen auditorischen Wahrnehmungsprozessen.

## 11.4 Resumée und Ausblick

Abbildung 11.1 faßt die wichtigsten Ergebnisse der vorliegenden Arbeit zusammen. In der Experiementalserie I wurde die Mismatch Reaktion als spezifischer Prozeß, d.h. als abhängig von Langzeitrepräsentationen gekennzeichnet. Experiment II zeigte daß die Wahrnehmung einer phonetischen Abweichung, die einen frühen Prozeß der Sprachverarbeitung darstellt, von der Stimminformation nicht unabhängig ist: bei einer zweidimensionale Abweichung (Phonem *und* Stimme) findet eine integrative Verarbeitung der Reizdimensionen statt. Hingegen deuten die Experimente IIIb und c darauf hin, daß Prozesse des semantischen Satzverstehens bzw. Komponenten, die diese Prozesse reflektieren, von den frühen auditorischen Wahrnehmungsprozessen nicht beeinflußt werden. Somit wurde die auditorische Verarbeitung von Stimminformation in der vorliegenden Arbeit auf ihre neuronalen Grundlagen hin untersucht und in Bezug zur Verarbeitung der linguistisch enkodierten Nachricht gesetzt. Im folgenden Abschnitt sollen nun mögliche weiterführenden Fragestellungen, die sich aus den Ergebnissen der Arbeit ergeben, skizziert werden.

Da gezeigt wurde, daß die akustischen Besonderheiten der menschlichen Stimme schon in der präattentiven Verarbeitung vom auditorischen System erkannt werden und zu Abweichungen zu einer spezifischen neuronalen Reaktion führen, können die hier vorgestellten Verfahren beispielsweise dazu genutzt werden, um diese Prototypikalitätseffekte näher zu bestimmen. So kann in weiteren Studien getestet werden, welche akustischen Parameter für die Natürlichkeit synthetischer/manipulerter Sprachsignale von Bedeutung sind bzw. welche Parameterkonfigurationen als natürliche Varianz akzeptiert werden.

Ein weiteres wichtiges Ergebnis der vorgestellten Experimente ist der Hinweis auf eine konfigurale/holistische Verarbeitung von Stimminformation. Parallelen zur visuellen Wahrnehmung von Gesichtern sind bereits aufgezeigt worden. Es wäre sicherlich lohnenswert, die Gemeinsamkeiten von Gesichts- und Stimmerkennung genauer zu untersuchen, um die allgemeinen Verarbeitungsprinzipien zu verstehen. Dies kann zum Beispiel durch die Untersuchung der ontogeneti-

schen Entwicklung beider Fähigkeiten geschehen. Die Studie von Mann et al. (1979) zeigt, daß für die Stimmen- und für die Gesichtererkennung erst im Alter von etwa 16 Jahren das Leistungsniveau der Erwachsenen erreicht wird. Diese Daten beruhen aber allein auf behavioralen Experimenten. Über ontogenetische Veränderungen in der neuronalen Verarbeitung ist nichts bekannt. Zudem sind meines Wissens weder bei Kindern noch bei Erwachsenen Geschlechtsunterschiede bei der Verarbeitung von Stimminformation untersucht worden. Gerade im Hinblick auf diesen letzten Aspekt mag es interessant sein, verschiedene Kulturräume zu vergleichen. Ein weiteres Forschungsfeld ist die neuronale Stimmverarbeitung bei Personen, für die Stimmen eine besondere Rolle spielen, wie z.B. Blinde. Im Vergleich zu Normalprobanden wäre dort eine andere funktionale Belegung kortikaler Areale zu vermuten, ähnlich wie sie für die Verarbeitung von Gebärden bei Gehörlosen berichtet wurde (Neville et al., 1998; Nishimura et al., 1999). Schließlich bleibt zu sagen, daß die Stimme nicht nur zur sprachlichen Kommunikation dient, sondern auch ein musikalisches Ausdrucksmittel ist. Im Zuge des zur Zeit herrschenden Interesses and der neurokognitiven Verarbeitung von Musik im Hinblick auf deren strukturelle Prozesse mag auch der Vergleich von Sing- und Sprechstimme bzw. von Singstimme und Instumentalklang weitere Einsicht in die auditorische Verarbeitung akustischer Reize bringen (vgl. Levy et al., 2001).

Die hier aufgeführten, z.T. recht unterschiedlichen weiterführenden Fragestellungen zeigen in ihrer Gesamtheit einmal mehr: Es handelt sich bei der neuronalen Verarbeitung der Stimminformation um ein relativ neues Untersuchungssgebiet und *"very little is actually known about voice perception" (J.Kreiman, 1997, S. 99)*.

# Teil V

# APPENDIX

# Anhang A

# Experimentalserie I

| Stimulus | Natürlichkeit | Wohlklang |
|---|---|---|
| M | 1.73 (1.03) | 1.73 (0.96) |
| W | 1.47 (0.52) | 1.80 (0.86) |
| G | 3.86 (0.92) | 3.93 (0.45) |
| S | 3.80 (0.94) | 4.46 (0.92) |

Tabelle A.1: Mittelwerte und Standardabweichungen für die Beurteilung der Natürlichkeit und des Wohlklangs der Stimuli (Experiment Ia und Ib) auf einer 1-5 Skala (1=normal/angenehm; 5=künstlich/unangenehm).

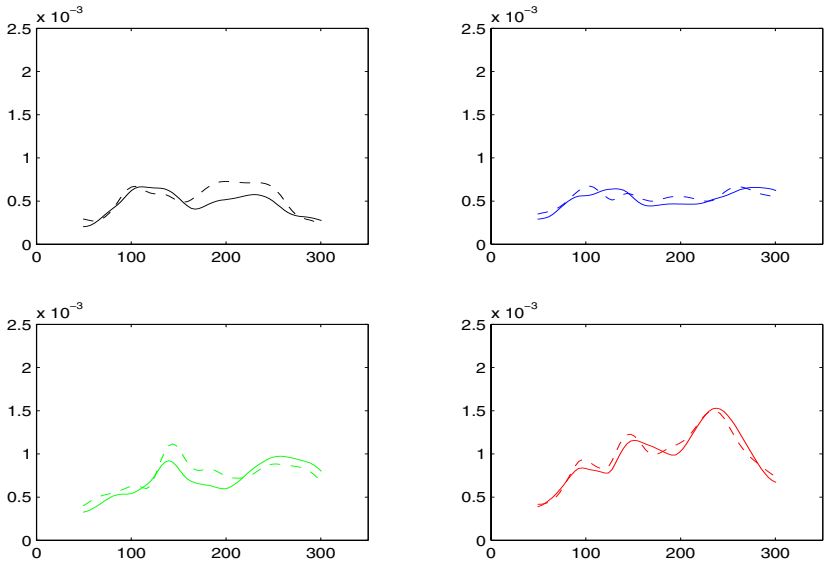



Abbildung A.1: Experiment Ia: Stromdichteverlauf für jede Hemisphäre relativ zum Stimulusbeginn (durchgezogene linie markiert die linke Hemisphäre, gestrichelte Linie markiert die rechte Hemisphäre; links oben = M_std; rechts oben = W ; links unten = G; rechts unten = S).

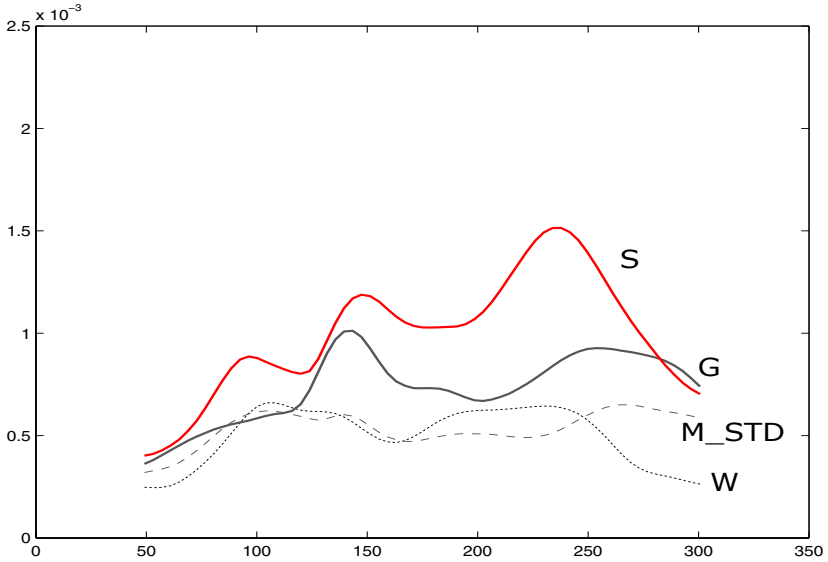



Abbildung A.2: Experiment Ia:Experiment Ia: Stromdichteverlauf aller Bedingungen (gemittelt über die Hemisphären..

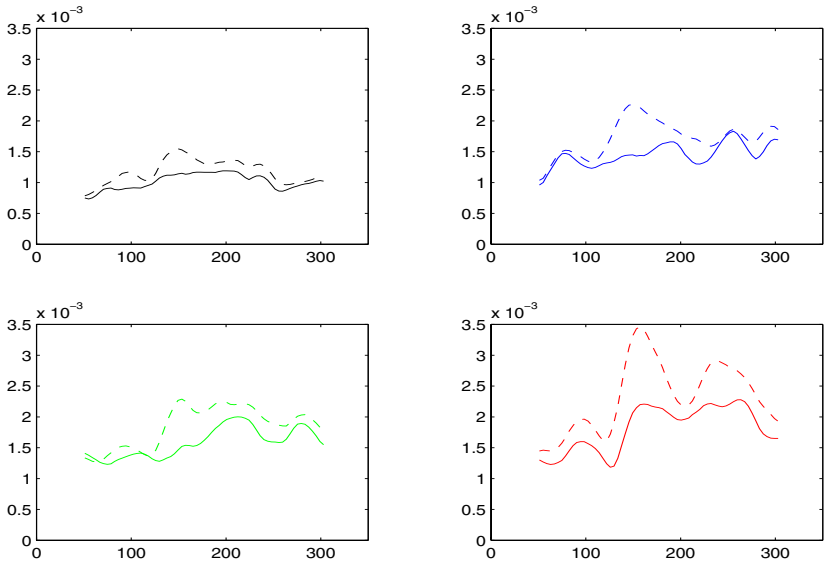



Abbildung A.3: Experiment Ib: Stromdichteverlauf für jede Hemisphäre relativ zum Stimulusbeginn (durchgezogene Linie markiert die linke Hemisphäre, gestrichelte Linie markiert die rechte Hemisphäre;links oben = M_std; rechts oben = W ; links unten = G; rechts unten = S).

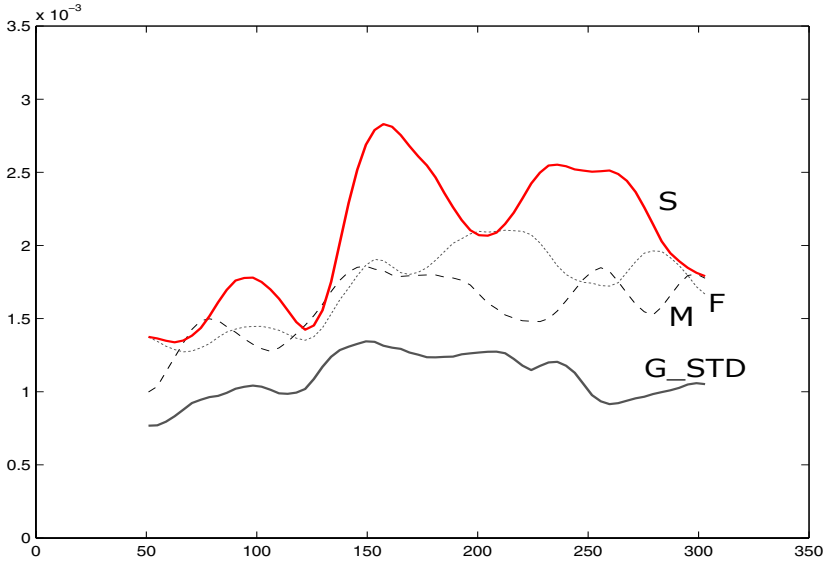



Abbildung A.4: Experiment Ib:Experiment Ia: Stromdichteverlauf aller Bedingungen (gemittelt über die Hemisphären).

# Anhang B

# Experiment II

| Stimulus | Plosiv [ms] | Vokal [ms] | Onset Frikativ [ms] | Dauer gesamt [ms] |
|---|---|---|---|---|
| Std (*"Test"*) | 55 | 121 | 171 | 383 |
| Ph (*"Dach"*) | 12 | 131 | 153 | 397 |
| Voc (*"Test"*) | 59 | 140 | 196 | 384 |
| PhVoc (*"Dach"*) | 8 | 145 | 153 | 378 |

Tabelle B.1: Stimuluseigenschaften I: Dauer des Plosivs und Vokals, Beginn des Frikativs (gemessen relativ zum Stimulusbeginn) und Gesamtdauer der Stimuli.

| Stimulus | Sprecher | F0 [Hz] | F1 [Hz] | F2 [Hz] | F3 [Hz] |
|---|---|---|---|---|---|
| Std (*"Test"*) | männl. | 82 | 470 | 2386 | 3878 |
| Ph (*"Dach"*) | männl. | 80 | 806 | 1653 | 4616 |
| Voc (*"Test"*) | weibl. | 167 | 390 | 2476 | 4462 |
| PhVoc (*"Dach"*) | weibl. | 167 | 736 | 1361 | 2897 |

Tabelle B.2: Stimuluseigenschaften II: mittlere Grudfrequenz, und Formant 1 - Formant 3.

# Anhang C

# Experimentalserie III

| | | Gruppe 1 Sprecherunterscheidung | | Gruppe 2 Semantische Beurteilung | |
|---|---|---|---|---|---|
| **SEMANTIK** | **SPRECHER** | **% korr.** | **RT** | **% korr.** | **RT** |
| kongruent | Std1 | 99 [2.1] | 663 [134] | 92 [6.0] | 609 [165] |
| | M | 89 [20.1] | 640 [166] | 97 [5.0] | 664 [199] |
| | Std2 | 100 [1.1] | 559 [165] | 96 [3.8] | 617 [157] |
| | W | 98 [3.0] | 452 [154] | 96 [4.2] | 613 [138] |
| inkongruent | Std1 | 96 [6.0] | 710 [151] | 95 [4.5] | 723 [163] |
| | M | 94 [10.4] | 608 [163] | 97 [3.8] | 683 [145] |
| | Std2 | 99 [1.1] | 599 [160] | 97 [3.1] | 723 [124] |
| | W | 99 [1.5] | 470 [175] | 96 [3.8] | 681 [132] |

Tabelle C.1: Mittelwerte und Standardabweichungen der Antwortquoten sowie der Reaktionszeiten bei der Sprecherunterscheidung und der semantischen Beurteilung

| Bedingung | Sprecherunterscheidung inkongruent - kongruent (Effekt in ms) | semantische Beurteilung inkongruent - kongruent (Effekt in ms) |
|---|---|---|
| S1 | 47 [46] | 114 [60] |
| M | -32 [60] | 19 [81] |
| S2 | 40 [45] | 105 [63] |
| W | 18 [35] | 69 [54] |

Tabelle C.2: Mittelwerte und Standardabweichungen der Semantik-Effekte, gemessen als Differenz der Reaktionenzeiten bei kongruenten und inkongruente Sätzen.

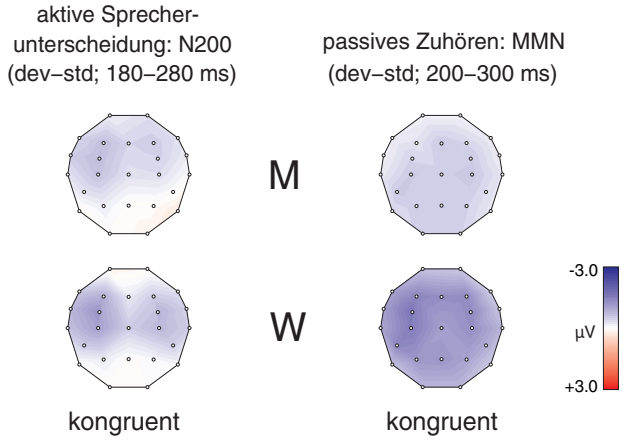



Abbildung C.1: Experiment IIIb (aktive Sprecherunterscheidung und IIIc (passives Zuhören): Potentialfelddifferenzen (Bedingungen M-Std und W-Std) für semantisch kongruente satzfinale Partizipien.

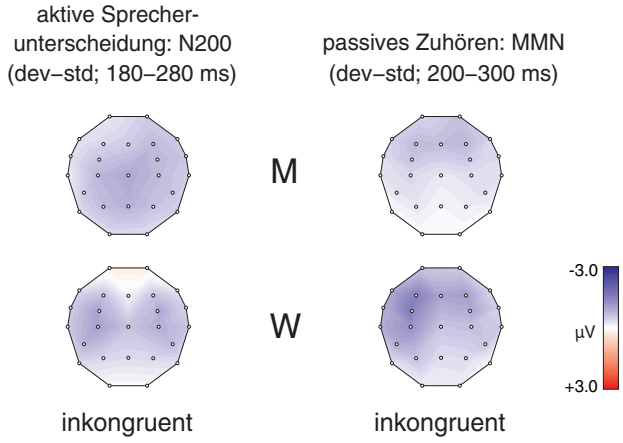



Abbildung C.2: Experiment IIIb (aktive Sprecherunterscheidung und IIIc (passives Zuhören): Potentialfelddifferenzen (Bedingungen M-Std und W-Std) für semantisch inkongruente satzfinale Partizipien.

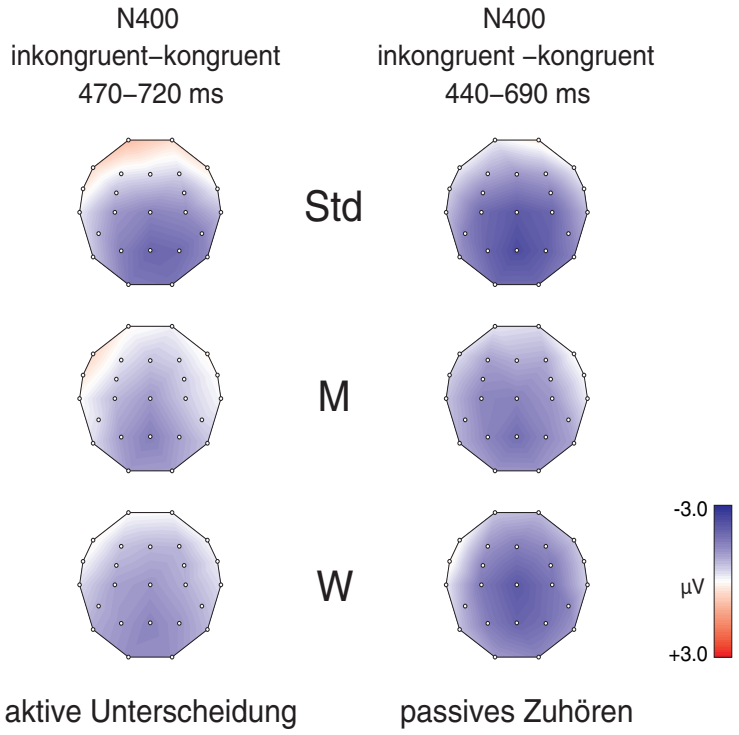



Abbildung C.3: Experiment IIIb (aktive Sprecherunterscheidung und IIIc (passives Zuhören): Potentialfelddifferenzen für semantisch inkongruente - kongruente satzfinale Partizipien der drei Sprecherbedingungen.

# Stimulusmaterial

### Trial kongruent

Die Polizei hat den schlauen Dieb gefaßt.
Die Feuerwehr hat den heißen Brand gelöscht.
Der Vegetarier hat das frische Obst gegessen.
Der Geselle hat ein lustiges Lied gepfiffen.
Der Einbrecher hat das wertvolle Gold gestohlen.
Die Mitschüler haben das dicke Kind gehänselt.
Der Pfarrer hat das kleine Kind getauft.
Der Bäcker hat das frische Brot gebacken.
Der Bräutigam hat seine schöne Braut geheiratet.
Die Putzfrau hat das Treppenhaus geputzt.
Der Chef hat den langen Bericht gelesen.
Die Buben haben den kalten Saft getrunken.
Der Forscher hat das seltene Insekt gefangen.
Der Hund hat das freche Kind gebissen.
Der Koch hat die dicke Wurst gebraten.
Der Feriengast hat sich ein Fahrrad gemietet.
Die Wölfe haben das frische Blut gerochen.
Der Chor hat ein fröhliches Lied gesungen.
Die Braut hat ein weißes Kleid getragen.
Die Wespe hat das kleine Kind gestochen.
Die Witwe hat das alte Grab gepflegt.
Der Reiter hat das edle Pferd gestriegelt.
Der Autor hat einen neues Roman geschrieben.
Die Kundin hat einen schicken Hut gekauft.
Die Fußballfans haben ein lautes Lied gegrölt.
Der Pianist hat ein bekanntes Stück gespielt.
Die Hausfrau hat das weiße Hemd gebügelt.
Der Weihnachtsmann hat ein großes Geschenk gebracht.
Das Verkehrsschild hat die Vorfahrt geregelt.
Der Rennfahrer hat den großen Preis gewonnen.
Die Kumpanen haben den selbstgebrannten Schnaps gesoffen.
Die Schülerin hat ein langes Gedicht gelernt.
Der Verehrer hat einen kleinen Blumenstrauß gepflückt.
Das fleißige Zimmermädchen hat das Bett gemacht.
Die Gemeinde hat das gesammelte Geld gespendet.
Die Verbündeten haben einen heiligen Eid geschworen.
Der Archäologe hat den versteckten Schatz gefunden.
Die Handwerker haben die Wand gestrichen.
Der Kriminalbeamte hat einen Schuß gehört.
Der Händler hat das kostbare Gewürz gemahlen.
Die Nachbarsfrau hat das dumme Gerücht geglaubt.
Der Autofahrer hat das verkehrsregelnde Schild gesehen.
Der Schüler hat den langweiligen Unterricht geschwänzt.
Der Musiker hat sein wertvolles Instrument gestimmt.
Die Großmutter hat den löchernen Strumpf gestopft.
Der Künstler hat ein gelungenes Bild gemalt.
Die Kühe haben das grüne Gras gefressen.
Der Fischer hat sein zerrissenes Netz geflickt.

## Trial inkongruent

Die Polizei hat den schlauen Dieb gegrölt.

Die Feuerwehr hat den heißen Brand gespielt.

Der Vegetarier hat das frische Obst gebügelt.

Der Geselle hat ein lustiges Lied getragen.

Der Einbrecher hat das wertvolle Gold geregelt.

Die Mitschüler haben das dicke Kind gewonnen .

Der Pfarrer hat das kleine Kind gesoffen.

Der Bäcker hat das frische Brot gelernt.

Der Bräutigam hat seine schöne Braut gepflückt.

Die Putzfrau hat das Treppenhaus gelesen.

Der Chef hat den langen Bericht geputzt.

Die Buben haben den kalten Saft gefangen.

Der Forscher hat das seltene Insekt getrunken.

Der Hund hat das freche Kind gestrichen .

Der Koch hat die dicke Wurst gepflegt.

Der Feriengast hat sich ein Fahrrad gemahlen.

Die Wölfe haben das frische Blut geglaubt.

Der Chor hat ein fröhliches Lied gesehen.

Die Braut hat ein weißes Kleid gepfiffen.

Die Wespe hat das kleine Kind gestriegelt.

Die Witwe hat das alte Grab gebraten.

Der Reiter hat das edle Pferd gestimmt.

Der Autor hat einen neues Roman gemalt.

Die Kundin hat einen schicken Hut geflickt.

Die Fußballfans haben ein lautes Lied gefaßt.

Der Pianist hat ein bekanntes Stück gelöscht.

Die Hausfrau hat das weiße Hemd gegessen .

Der Weihnachtsmann hat ein großes Geschenk geschwänzt.

Das Verkehrsschild hat die Vorfahrt gestohlen.

Der Rennfahrer hat den großen Preis gehänselt.

Die Kumpanen haben den selbstgebrannten Schnaps getauft .

Der Klassenbeste hat ein langes Gedicht gebacken.

Der Verehrer hat einen kleinen Blumenstrauß geheiratet.

Das fleißige Zimmermädchen hat das Bett gespendet.

Die Gemeinde hat das gesammelte Geld gemacht.

Die Verbündeten haben einen heiligen Eid gefunden.

Der Archäologe hat den versteckten Schatz geschworen.

Die Handwerker haben die glatte Wand gebissen.

Der Kriminalbeamte hat einen Schuß gestopft.

Der Händler hat das kostbare Gewürz gemietet.

Die Nachbarsfrau hat das dumme Gerücht gerochen.

Der Autofahrer hat das verkehrsregelnde Schild gesungen.

Der Schüler hat den langweiligen Unterricht gebracht.

Der Musiker hat sein wertvolles Instrument gestochen.

Die Großmutter hat den löchernen Strumpf gehört.

Der Künstler hat ein gelungenes Bild gefressen.

Die Kühe haben das grüne Gras geschrieben.

Der Fischer hat sein zerrissenes Netz gekauft.

## Filler kongruent

Ina fuhr gestern in Urlaub.

Heute hat Frank keine Lust zu arbeiten.

Kann Anja das Rennen gewinnen?

Lucy ist schon sehr gewachsen.

Der Fisch zappelte an der Angel.

Johnny mußte alles alleine organisieren.

Traurig ging der Mann über die Straße.

Der große Künstler gab sich die Ehre.

Peter hatte es Anna versprochen.

Warum schreibt Anne keine Briefe mehr?

Die Antwort des Beamten war sehr unfreundlich.

Rosie war das dickste Mädchen in der Klasse.

Jessica hat vielleicht ein Problem.

Carsten wollte sofort nach Hause rennen.

Die Zwillinge lachten über den Vater.

Judith könnte mehr aus sich machen.

Die Blätter fallen schon von den Bäumen.

Hans und Maria können sich nicht leiden.

Jan kann sich einfach nicht entscheiden.

Das Mittagessen wird ganz kalt.

Jutta fährt nächste Woche nach Helsinki.

Vom Zelten waren sie nicht sehr begeistert.

Janna hat einen starken Heuschnupfen.

Der Kapitän verläßt das Schiff als letzter.

Wer wird unseren neuen Computer kaufen?

Jessica hat die Vase fallen lassen.

Thorsten wäscht sich fast nie.

Julika will auf keinen Fall verlieren.

Alle freuten sich auf die Kanufahrt.

Der Graben war unüberwindbar.

Petra wird sich ein neues Auto kaufen.

Silvia will jetzt bald nach Hause gehen.

Der Schneemann schmilzt in der Sonne.

Der März ist der erste Frühlingsmonat.

Der Gärtner will die Pflanzen gießen.

Der Tennisspieler wird das Match verlieren.

Das Zaun ist frisch lackiert worden.

Die Blumen duften sehr gut.

Suse hat die Prüfung zum Glück bestanden.

Sie haben ein Picknick im Grünen gemacht.

Viele Häuser in der Stadt sind sehr verfallen.

Am Sonntag sind alle in die Kirche gegangen.

Der Tee ist zum Trinken noch zu heiß.

Katrin ist die Treppe hinabgestürzt.

Karl bejubelte den Sieg seines Teams.

Man darf die Hoffnung nicht so schnell aufgeben.

Er hatte gestern seinen Job verloren.

Viele Köche verderben den Brei.

Die Zeit im Internat war für Tom sehr schwer.

Johann ist heute ausgesprochen gut gelaunt.

Der Zaun ist vom Nachbarn gestrichen worden.

Sie haben die Mauer innerhalb weniger Tage errichtet.

Die Süßigkeiten waren sofort aufgegessen worden.

Die Operation erfolgte ohne Komplikationen.

Der Stromausfall kam völlig überraschend.

Im November ist alles grau und trübe.

Hellhäutige Menschen bekommen leicht Sonnenbrand.

Paris ist eine Reise wert.

Heute hat es stark geregnet.

Der Teppich ist schon lange nicht mehr gesaugt worden.

Franziska kann noch nicht schwimmen.

Die Politiker debattieren schon den ganzen Nachmittag.

Der Boden ist zum Schlafen etwas zu hart.

Die kleine Wiebke buddelt im Sandkasten.

Stefan hat einen großen Fisch geangelt.

Natascha will unbedingt Spanisch lernen.

Die Tische stehen zu weit auseinander.

Im Park standen uralte Bäume.

Justus möchte gerne Astronaut werden.

Darf ich Dir etwas Gesellschaft leisten?

Die Wäsche ist schon fast getrocknet.

Berlin ist die neue Hauptstadt Deutschlands.

Frank war noch nie auf dem Münchner Oktoberfest.

Der Kommissar wird den Fall bald aufgeklärt haben.

Schottland ist von London aus in wenigen Stunden zu erreichen.

Die Praktikantin wurde plötzlich sehr nervös.

Vorhin traf ich meine Freundin in der Stadt.

Sie spielt sehr gut Klavier.

Die Lebensmittel sind alle verdorben.

Kann Hans das Fahrrad reparieren?

Neulich sah Jens ein lustiges Theaterstück.

Diese Tabletten sind für Kinder zu stark.

Die Firma musste Konkurs anmelden.

Philipp hat große Angst vorm Fliegen.

Wird unser Verein ein Abschlußfest feiern?

Die Weihnachtsfeier war sehr gelungen.

Der Verräter wurde zu lebenslanger Haft verurteilt.

Der Mond wird heute nacht gut zu sehen sein.

Wenigstens haben wir noch das ganze Wochenende für uns.

Die Sekretärin hat sich den Finger eingeklemmt.

Auch Marco wird irgendwann Freunde finden.

Dein Bekannter sollte mal wieder zum Frisör gehen.

Die Entdeckung der Biologen war von größter Wichtigkeit.

Es gibt eine gute und eine schlechte Neuigkeit.

Wann klingelt endlich das Telefon?

Segeln ist bei diesem Wetter viel zu gefährlich.

## Filler inkongruent

Ina fuhr gestern in die Küche.

Heute hat Frank keine Lust zu atmen.

Kann Anja das Rennen schlafen?

Lucy ist schon sehr gekleidet.

Der Fisch zeichnet an der Angel.

Johnny mußte alles im Fliegen organisieren.

Jessica hat vielleicht ein Argentinien.

Carsten wollte sofort ins Frühstück rennen.

Die Zwillinge öffnen den Vater.

Judith könnte mehr aus sich hüpfen.

Die Blätter fallen schon von den Augen.

Hans und Maria können sich nicht jaulen.

Jan kann sich einfach nicht verdienen.

Das Mittagessen hat gezappelt.

Jutta fährt letzte Woche nach Helsinki.

Alle freuten sich auf den Schrebergarten.

Janna hat einen aufregenden Heuschnupfen.

Der Kapitän raucht das Schiff als letzter.

Wer wird unseren neuen Computer verdrehen?

Traurig ging der Mann unter die Straße.

Der große Künstler gab sich die Hoffnung.

Peter hatte das Aufkommen Anna versprochen.

Warum schreibt Anne keine Tische mehr?

Die Soße des Beamten war sehr unfreundlich.

Rosie war die dickste Phantasie in der Klasse.

Karl bejubelte den Körper seines Teams.

Man darf die Hoffnung nicht so schnell träumen.

Er hatte gestern seinen Jahrmarkt verloren.

Viele Köche verderben den Fußgängerüberweg.

Die Zeit im Kochtopf war für Tom sehr schwer.

Johann ist heute ausgesprochen wetterfest gelaunt.

Der Zaun ist vom Nachbarn bekannt gemacht worden.

Sie haben die Mauer innerhalb weniger Jahrhunderte errichtet.

Die Süßigkeiten waren sofort aufgeblasen worden.

Die Operation erfolgte ohne Theaterbesuch.

Im November ist das Geschenk grau und trübe.

Der Stromausfall kam völlig verdorben zur Tür herein.

Hellhäutige Menschen bekommen leicht Flügel.

Jessica hat die Vase anrufen lassen.

Thorsten schüttelt sich fast nie.

Julika will auf keinen Fall pochen.

Vom Zelten waren sie nicht sehr verbittert.

Der Graben war sehr hoch.

Petra wird sich ein betrogenes Auto kaufen.

Silvia will jetzt bald nach Hause tragen.

Der Schneemann quietscht in der Sonne.

Der März ist der erste Frühlingsbaum.

Der Gärtner will die Vorträge gießen.

Der Tennisspieler wird das Alter verlieren.

Das Zaun ist frisch gelacht worden.

Die Blumen schnarchen sehr gut.

Suse hat die Prüfung zum Glück gehandelt.

Sie haben ein Picknick im Grübeln gemacht.

Viele Häuser in der Schule sind sehr verfallen.

Am Sonntag sind alle in das Weltall gegangen.

Der Tee ist zum schreien noch zu heiß.

Katrin ist die Wanderung hinabgestürzt.

Vorhin traf ich meine Seefahrt in der Stadt.

Sie spielt sehr gut Insel.

Die Lebensmittel sind alle gekniffen.

Kann Hans das Fahrrad sammeln?

Neulich sah Jens ein salziges Theaterstück.

Diese Tabletten sind für Kinder zu heiter.

Die Firma musste Kaviar anmelden.

Philipp hat große Chemie vorm Fliegen.

Wann klingelt endlich das Tulpenbeet?

Die Weihnachtsfeier war sehr gebraucht.

Paris ist eine Hälfte wert.

Heute hat es stark geradelt.

Der Teppich ist schon lange nicht mehr gesalbt worden.

Franziska kann noch nicht tief hören.

Die Politiker maskieren sich schon den ganzen Nachmittag.

Der Boden ist zum Schlafen etwas zu pubertär.

Die kleine Wiebke buddelt im Untersuchungsausschuß.

Stefan hat eine große Baustelle geangelt.

Natascha will unbedingt Spanisch krönen.

Die Tische stehen zu weit auf dem Auszubildenden.

Im Park standen uralte Waschmaschinen.

Justus möchte gerne Astronaut lösen.

Darf ich Dir etwas Gesellschaft weben?

Die Wäsche ist schon fast geworfen.

Berlin ist die neue Tischdecke Deutschlands.

Frank war noch nie auf dem Münchner Schiffsdeck.

Der Kommissar wird den Fall bald gesummt haben.

Schottland ist von London aus in wenigen Eiszeiten zu erreichen.

Die Praktikantin wurde plötzlich sehr gesüßt.

Der Verräter wurde zu lebenslanger Haft vertagt.

Der Mond wird heute nacht gut zu riechen sein.

Wenigstens haben wir noch die ganze Färbung für uns.

Die Sekretärin hat sich den Kaffeefilter eingeklemmt.

Auch Marco wird irgendwann eine Störung finden.

Dein Bekannter sollte mal wieder zum Frisör weinen.

Die Entdeckung der Biologen war von größter Nachbarschaft.

Es gibt eine gute und eine schlechte Nagelprobe.

Wird unser Hochzeitstag ein Abschlußfest feiern?

Segeln ist bei diesem Wetter viel zu steil.

# Literaturverzeichnis

Aaltonen O., Niemi P., Nyrke T. & Tuhkanen M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, 24:197–207.

Abberton E. & Fourcin A. (1978). Intonation and speaker identification. *Language and Speech*, 21(4):305–318.

Alain C., Woods D. & Covarrubias D. (1997). Activation of duration-sensitive auditory cortical fields in humans. *Electroencephalography and Clinical Neurophysiology*, 105:531–39.

Alho K., M. C., A. L., Huotilainen M., Virtanen J., Aulanko R. et al. (1998a). Hemispheric lateralization in preattentive processing of speech sounds. *Neuroscience letters*, 258(1):211–224.

Alho K., Winkler I., Escera C., Huotilainen M., Virtanen J., Jääskelainen I. et al. (1998b). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings. *Psychophysiology*, 35:211–224.

Alho K., Woods D., Algazi A. & Näätänen R. (1992). Intermodal selective attention. Effects of attentional load on processing of auditory and visual stimiuli in central space. *Electroencephalography and Clinical Neurophysiology*, 82:356–368.

Alku K., Sivonen P., Palomäki K. & Tiitinen H. (2001). The periodic structure of

vowel sounds is reflected in human electromagnetic brain responses. *Neuroscience Letters*, 298:25–28.

Anwander A., Kuhn M., Reitzinger S. & Wolters C. (2001). A parallel algebraic multigrid solver for finite element method based source localization in the human brain. *Publikationen des Max Planck Instituts für Mathematik in den Naturwissenschaften*, 35:preprint.

Aulanko R., Hari R., Lousna O., Näätänen R. & Sams M. (1993). Phonetic invariance in the human auditory cortex. *NeuroReport*, 4:1356–1358.

Banich M. & Heller W. (1998). Evolving perspectives on lateralization of function. *Current Directions in Psychological Science*, 7(1):1–2.

Belin P. & Zatorre R. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience*, 3(10):965–966.

Belin P., Zatorre R. & Ahad P. (2002). Human temporal -lobe response to vocal sounds. *Cognitive Brain Research*, 13:17–26.

Belin P., Zatorre R., Lafalle P., Ahad P. & Pike B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(20):309–312.

Belin P., Zilbovicius M., Crozier S., Thivard L. & Fontaine A. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience*, 10(4):536–540.

Bentin S., Kutas M. & Hillyard S. (1993). Electrophysiological evidence for task-effects on semantic priming in auditory word processing. *Psychophysiology*, 30:161–169.

Besson M., Kutas M. & Van Petten C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects. *Journal of Cognitive Neuroscience*, 4(2):132–149.

Binder J., Frost J., Hemmeke T., Bellogwan P., Springer J., Kaufman J. et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10(5):512–528.

Binder J., Frost J., Hemmeke T., Cox R., Rao S. & Prieto T. (1997). Human brain language areas identified by fMRI. *The Journal of Neuroscience*, 17:353–362.

Binder J., Frost J., Hemmeke T., Rao S. & Cox R. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain*, 119:1239–47.

Birbaumer N. & Schmidt R. (1990). *Biologische Psychologie*. Berlin etc.: Springer-Verlag.

Bladon R., Henton C. & Pickering J. (1984). Towards an auditory theory of speaker normalization. *Language and Communication*, 4:59–69.

Bobes M., Valdes-Sosa M. & Olivares E. (1994). An ERP study of expectancy in face processing. *Brain and Cognition*, 26:1–22.

Boff K., Kauffman L. & Thomas J.H. (Hg.) (1986). *Handbook of perception and human performance*. New York: Wiley.

van den Brink D., Brown C. & Hagoort P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 Effects. *Journal of Cognitive Neuroscience*, 13(7):967–985.

Brown R. (1980). *Auditory speaker recognition*. Hamburg: Helmut Buske Verlag.

Buchanan T., Lutz K., Mirsazade S., Specht K., Shah N., Zilles K. et al. (2000). Recognition of emotional speech prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research*, 9:227–238.

Celsis P., Boulanouar K., Ranjeva J., Berry I., Nespoulous J. & Chollet F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supermarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage*, 285(9):135–144.

Church B. & Schacter D. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning Memory and Cognition*, 20(3):521–533.

Chwilla D., Brown C. & Hagoort P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32:274–285.

Coles M. (1995). Event-related brain potentials: an introduction. In M. Rugg & M. Coles (Hg.), *Electrophysiology of Mind*, Oxford: Oxford University Press, 1–26.

Comchero M. & Polich J. (1998). P3a, perceptual distinctiveness, and stimulus modality. *Cognitive Brain Research*, 7:41–48.

Comchero M. & Polich J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, 110:24–30.

Conolly J. & Phillips N. (1994). Event-related potential components reflect phonological, semantic processing of the terminal words of spoken sentences. *Journal of Cognitive Neuroscience*, 6:256–266.

Conolly J., Phillips N. & Forbes K. (1995). The effects of phonological semantic features of sentence-ending words on visual event-related potentials. *Neurophysiology*, 94:276–287.

Courchesne E., Courchesne R. & Hillyard S. (1978). The effect of stimulus deviation on P3 waves to easily recognized stimuli. *Neuropsychologia*, 16:189–199.

Creem S. & Proffit D. (2001). Defining the cortical visual systems: 'What', 'Where' and 'How'. *Acta Psychologica*, 107:43–68.

Csepe V., Pantev C., Hoke M., Hampson S. & Ross B. (1992). Evoked magnetic responses of the human auditory cortex to minor pitch changes - localization of the mismatch field. *Electroencephalography and Clinical Neurophysiology*, 84(6):538–548.

Czigler I. & Csibra G. (1990). Event-related potentials in a visual discrimination task: Negative waves related to detection and attention. *Psychophysiology*, 27:669–676.

Czigler I., Csibra G. & Csontos A. (1992). Age and inter-stimulus interval effects on event-related potentials to frequent and infrequent auditory stimuli. *Biological Psychology*, 33:195–206.

Danemann M. & Carpenter P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19:450–466.

Deacon D., Breton F., Ritter W. & Vaughan J.H. (1998). Automatic change detection: Does the brain use representations of individual stimulus features or gestalts. *Psychophysiology*, 35:413–419.

Diesch E., Eulitz C., Hampson S. & Ross B. (1996). The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain and Language*, 53:143–168.

Diesch E. & Luce T. (1997). Magnetic mismatch fields elicited by vowels and consonants. *Experimental Brain Research*, 116(1):139–152.

Démonet J., Chollet F., Samsay S., Cardebat D., Nespoulous H., Wise R. et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115:1753–1768.

Doehring D. & Batholomeus B. (1971). Laterality effects in voice recognition. *Neuropsychologia*, 9:425–430.

Donchin E. & Coles M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11:406–417.

Donchin E., Spencer K. & Dien J. (1997). The varieties of deviance experience: ERP manifestations of deviance processors. In van Boxtel; G.J.M. & B. K.B.E. (Hg.), *Brain and Behavior: Past, Present and Future*, Tilburg, NL: University Press, 67–91.

Duffy S. & Pisoni D. (1992). Comprehension of synthetic speech produced by rule: a review and theoretical implication. *Language and Speech*, 35(4):351–389.

Eckert H. & Laver J. (1994). *Menschen und ihre Stimmen*. Weinheim: Beltz - Psychologie Verlags Union.

Eggermont J. (2001). Between sound and perception: reviewing the search for a neural code. *Hearing Research*, 157:1–42.

Elberling C., Bak C., Kofoed B., Lebech L. & Saermark K. (1982). Auditory magnetic fields from the human cerebral cortex: Location and strength of an equivalent current dipole. *Acta Neurologica Scandinavia*, 65:553–569.

Elbert T. (1993). Slow cortical potentials reflect the regulation of cortical excitability. In *Slow potential changes in the human brain*, New York: Plenum Press, 235–251.

Escera C.a. (2000). Involuntary attention and distractability as evaluated with event-related brain potentials. *Audiology and Neuro-Otology*, 5:151–166.

Fant G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Ferguson A., Zhang X. & Stroink X. (1994). A complete linear discretisation for calculating the magnetic field using the boundary element method. 41:459.

Fischler I., Bloom P., Childers D., Arroyo A. & Perry N. (1984). Brain potentials during sentence verification - late negativity and long-term-memoy strength. *Neuropsychologia*, 22(5):559–568.

Friederici A., Pfeifer E. & Hahne A. (1993). Event-related potentials during natural speech processing: effects of semantic morphological and syntactic violations. *Cognitive Brain Research*, 183–192.

Friederici A., Wang Y. & Maess B. (2001). Unraveling the N400: The hemispheric interplay during semantic processing. *Annual Report of the Max Planck Institute of Cognitive Science 2000*, 41–42.

Friedman D., Cycowicz Y. & Gaeta H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, 25:355–373.

Garner W. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Giard M., Lavikainen J., Reinikainen K., Perrin F., Bertrand O., Pernier J. et al. (1995). Separate representations of stimulus frequency, intensity and duration in auditory stimulus deviants. *Journal of Cognitive Neuroscience*, 7:133–143.

Giard M., Perrin F., Pernier J. & Boiuchet P. (1990). Brain generators implicated in the processing of auditory stimulus deviance; a topographic event-related potential study. *Psychophysiology*, 27:627–640.

Goggin J., Thompson C., Sturbe G. & Simental L. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, 19(5):448–458.

Goldinger S. (1997). Words and voices - perception and production in an episodic lexicon. In K. Johnson & J. Mullenix (Hg.), *Talker Variability in Speech Processing*, San Diego: Academic Press, Inc., 33–66.

Goldinger S., Kleider H. & Shelley E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory and Cognition*, 27(2):328–338.

Goldinger S., Pisoni D. & Logan J. (1991). On the locus of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17:152–162.

Gomes H., Bernstein R., Ritter W., Jr. Vaughan H. & Miller J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, 34:712–716.

Good C., Johnsrude I., Ashburner J., Henson R., Friston K. & Frackowiak R. (2001). Cerebral symmetry and the effects of sex and handedness on brain

structure: A voxel-based morphometric analysis of 465 normal adult human brains. *NeuroImage*, 14:685–700.

Green K., Tomiak G. & Kuhl P. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, 59(5):675–692.

Hagoort P. & Brown C. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, 38:1518–1530.

Hahne A., Schröger E. & Friederici A. (2002). Segregating early syntactic and physical processes in auditory sentence comprehension. *NeuroReport*, 13(3).

Haley K.L. & Ohde R.N. (1996). Stimulus uncertainty and speaker normalization processes in the perception of nasal consonants. *Phonetica*, 53:185–199.

Halgren E., Baudena P., Heit G., Clarke J. & Marinkovich K. (1994). Spatio-temporal stages in face and word processing. *Journal of Physiology*, 88:1–50.

Halgren E., Marinkovich K. & Chauvel P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalography and Clinical Neurophysiology*, 106:156–164.

Hancock P., Bruce V. & Burton A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Science*, 4(9):330–337.

Hari R., Aittoniemi K., Jarvinen M., Katila T. & Varpula T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain: Localization of neural generators. *Experimental Brain Research*, 40:237–240.

Hari R., Hämäläinen M., Ilmionemi R., Kaukoranta E., Alho K., Näätänen R. et al. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips. *Neuroscience Letters*, 50:127–132.

Hari R., Levänen S. & Raji T. (2000). Timing of human cortical functions during cognition: role of MEG. *Trends in Cognitive Sciences*, 4(12):455–462.

Hecker M. (1971). *Speaker Recognition*. Washington: ASHA Monographs Number 16.

Helenius P., R S., Service E. & Connolly J. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, 212:1133–1142.

Hickok G. & Poeppel D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Science*, 4(4):131–138.

Hillyard S., Squires K., Bauer J. & Lindsay P. (1971). Evoked potential correlates of auditory signal detection. *Science*, 172:1357–1360.

Hämälainen M. & Ilmoniemi R. (1994). Interpreting magnetic field of the brain: minimum norm estimates. 32(1):183.

Hämäläinen M. & Sarvas J. (1989). Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. 36:165–171.

Holcomb P. & Neville H. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and cognitive processes*, 5(4):281–312.

Holcomb P. & Neville H. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, 19(4):286–300.

Ikeda K., Hayashi A., Hashimoto S., Otomo K. & Kanno A. (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical differences. *Neuroscience Letters*, 321:133–136.

Imaizumi S., Mori K., Kiritani S., Kawashima R. & Sugiura M. (1997). Vocal identification of speaker and emotion activates different brain regions. *NeuroReport*, 8(12):2809–2812.

Ivry R. & Robertson L. (1998). *The Two Sides of Perception*. Cambridge Mass.: MIT Press.

Jacobsen T. & Schröger E. (2001). Is there pre-attentive memory based comparison of pitch? *Psychophysiology*, 38:723–727.

Jaramillo M., Ilvonen T., Kujala T., Alku P., Tervaniemi M. & Alho K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cognitive Brain Research*, 12:459–466.

Jasper H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 20:371–375.

Johnson K. (1997). Speech perception without speaker normalization. In K. Johnson & J. Mullenix (Hg.), *Talker Variability in Speech Processing*, San Diego: Academic Press, Inc., 145–165.

Johnson R.J. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, 30:90–97.

Kasai K., Yamad H., Kamio S., Nakagome K., Iwami A., Fukuda M. et al. (2001). Brain lateralization for mismatch response to across- and within-category changes of vowels. *NeuroReport*, 12(11):2467–2471.

Kekoni J. (1998). Processing of stimulus repetition and change in the somatosensory system: Recordings of electrical and magnetic brain responses.

Kekoni J., Hämäläinen H., McCloud V., Reinikainen K. & Näätänen R. (1996). Is the somatosensory N250 related to deviance discrimination or conscious target detection? *Electroencephalography and Clinical Neurophysiology*, 100:115–125.

Kimura D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex*, 3:163–178.

Knecht S., Dräger B., Deppe M., Bobe L., Lohmann H., Flöel A. et al. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain*, 123(12):2512–2518.

Knight R. (1996). Contribution of human hippocampal region to novelty detection. *Nature*, 383:256–259.

Knösche T. (1997). Solutions of the neuromagnetic inverse problem: An evaluation study.

Kotz A., Meyer M., Alter K., Besson M., von Cramon Y. & Friederici A. (in press). On the lateralization of emotional prosody: An event-related functional MR study. *Brain and Language*.

Koyama S., Gunji A., Yabe H., Oiwa S., Akahane-Yamada R., Kakigi R. et al. (2000). Hemispheric lateralization in an analysis of speech sounds - Left hemisphere dominance replicated in Japanese subjects. *Cognitive Brain Research*, 10(1-2):119–124.

Kreiman J. (1997). Listening to voices - theory and practice in voice perception research. In K. Johnson & J. Mullenix (Hg.), *Talker Variability in Speech Processing*, San Diego: Academic Press, Inc., 85–108.

Kreiman J. & Van Lancker D. (1988). Hemispheric specialization for voice recognition: Evidence from dichotic listening. *Brain and Language*, 34:247–252.

Kutas M. & Federmeier K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4:463–470.

Kutas M. & Hillyard S. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, 11:99–116.

Kutas M. & Hillyard S. (1980b). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language*, 11:354–373.

Kutas M. & Hillyard S. (1984). Brain potentials during reading reflect word expectancy amd semantic association. *Nature*, 307:161–163.

Ladefoged P. & Broadbent D. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29:98–104.

Landis T., Buttet J., Assal G. & Graves R. (1982). Dissociation of ear preference in monaural word and voice recognition. *Neuropsychologia*, 20(3):501–504.

Langner G., Sams M., Heil P. & Schulze H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetencephalography. *Journal of Comparative Physiology*, 181(6):665–676.

Laver J. (Hg.) (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver J. & Mackenzie-Beck J. (2001). Unifying principles in the description of voice, posture and gesture. In C. C., I. Guatitella & S. Santi (Hg.), *Oralite and Gestualite - Interactions et comportements multimodaux dans la communication*, l'Harmattan, 15–24.

Lavner Y., Gath I. & Rosenhouse J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication*, 30(1):9–26.

Levänen S., Hari R., McEvoy L. & Sams M. (1993). Respones of the human auditory cortex to changes in one vs two stimulus features. *Experimental Brain Research*, 97(1):177–183.

Levy D.D., Granot R. & Bentin S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *NeuroReport*, 12(12):2653–2657.

Lewis M. & Johnston R. (1997). The Thatcher-Illusion as a test of configurational disruption. *Perception*, 26:225–227.

Lurito J. & Dzemidzic M. (2001). Determination of cerebral hemisphere languange dominance with functional magnetic resonance imaging. *Neuroimaging Clinics of North America*, 11(2):355–363.

Mackie K., Dermody P. & R. K. (1987). Assessment of evaluation measures for processed speech. *Speech Communication*, 6:309–316.

Mann V., Diamond R. & Carey S. (1979). Development of voice recognition: Parallels with face recognition. 27:153–165.

Marslen-Wilson W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25:71–102.

Marslen-Wilson W. & Tyler L. (1980). The temporal structure of spoken language understanding. *Cognition*, 8:1–71.

Martin C., Mullenix D.B.and Pisoni D. & Summers W. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning Memory and Cognition*, 15(4):676–684.

Mathiak K., Hertrich I., Lutzenberger W. & Ackermann H. (1999). Preattentive processing of consonant vowel syllables at the level of the supratemporal plane: a whole-head magnetoencephalography study. *Cognitive Brain Research*, 8:251–257.

Mathiak K., Hertrich I., Lutzenberger W. & Ackermann H. (2000). Encoding of temporal speech features (formant transients) during binaural and dichotic stimulus application: a whole-head magnetencephalography study. *Cognitive Brain Research*, 10:125–131.

McCallum W., Farmer S. & Pocock P. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 59:477–488.

McPherson W. & Holcomb P. (1999). An electroohysiological investigation of semantic with pictures of real objects. *Psychophysiology*, 36:53–65.

Mendoza E., Valencia N., Munoz J. & Trijillo H. (1996). Differences in voice quality between men and women: Use of the long-term average spectrum. *Journal of Voice*, 10(1):59–66.

Milner A. & Goodale M. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Mishkin M. (1979). Analogous neural models for tactual and visual learning. *Neuropsychologia*, 17(2):139–151.

Mäkelä A., Mäkinen V., Nikkilä, Ilmoniemi R.J. & Tiitinen H. (2001). Magnetoencephalographic (MEG) localization of the auditory N400m: effects of stimulus duration. *NeuroReport*, 12(2):249–253.

Mosher J. & Leahy R. (1998). Recursive MUSIC - a framework fopr EEG and MEG source localization. *IEEE Transactions on Biomedical Engineering*, 45(11):1342–1354.

Mullenix J. (1997). On the nature of perceptual adjustments to voice. In K. Johnson & J. Mullenix (Hg.), *Talker Variability in Speech Processing*, San Diego: Academic Press, Inc., 67–85.

Mullenix J., Johnson K., Topcu-Durgun M. & Farnsworth L. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America*, 98:3080–3095.

Mullenix J. & Pisoni D. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, 47:379–390.

Mullenix J., Pisoni D. & Martin C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85(1):365–378.

Nakasato N., Fujita S., Seki S., Kawamura T., Matani A., Tamura I. et al. (1995). Functional localization of bilateral auditory cortices using an MRI-linked whole-head magnetoencephalography (MEG) system. *Electroencephalography and Clinical Neurophysiology*, 94:183–190.

Neppert J. & Pétursson M. (1992). *Elemente einer akustischen Phonetik. (3.Aufl.)*. Hamburg: Helmut Buske Verlag.

Neville H., Bavelier D., Corina D., Rauschecker J., Karni A., Lalwani A. et al. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proceedings of the National Academy of Sciences of the United States of America*, 95(3):922–929.

Nishimura H., Hashikawa K., Doi K., Iwaki T., Watanabe Y., Kusuoka H. et al. (1999). Sign language 'heard' in the auditory cortex. 397:116.

Nix A., Mehta G., Dye J. & Cutler A. (1993). Phoneme detection as a tool for comparing perception of natural and synthetic speech. *Computer, Speech and Language*, 7:211–238.

Nobre A. & McCarthy G. (1994). Language-related ERPs: Scalp distribution by word type and semantic priming. *Journal of Cognitive Neuroscience*, 6:233–255.

Nobre A. & McCarthy G. (1995). Language-related field potentials in the anterior-medial temporal lobe II: Effects of word type and semantic priming. *Journal of Neuroscience*, 15:1090–1098.

Nolan F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Novak G., Ritter W. & Vaughan J. (1992). The chronometry of attention-modulated processing and automatic mismatch detection. *Psychophysiology*, 29:412–430.

Näätänen R. (1992). *Attention and Brain Function*. Hillsdale: Lawrence Erlbaum Associates.

Näätänen R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38:1–21.

Näätänen R. & Escera C. (2000). Mismatch negativity: clinical and other applications. *Audiology and Neuro-Otology*, 5:105–110.

Näätänen R., Gaillard A. & Mäntysalo S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42:313–329.

Näätänen R., Lehtokoski A., Lennes M., Cheour M., Houtilainen M., Ilvonen A. et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(30):432–434.

Näätänen R. & Picton T. (1986). N2 and automatic versus controlled processes. *Electroencephalography and Clinical Neurophysiology*, Suppl. 38:169–186.

Näätänen R., Tervaniemi M., Sussman E., Paavilainen P. & Winkler I. (2001). Primitive intelligence in the auditory cortex. *Trends in Neuroscience*, 24(5):283–287.

Nusbaum H., Francis A. & Henly A. (1995). Measuring the naturalness of synthetic speech. *International Journal of Speech Technology*, 1:7–19.

Nygaard L., Sommers M.S. & Pisoni D.B. (1994). Speech perception as a talker-contingent process. *Pychological Science*, 5(1):42–46.

O'Halloran J., Isenhart R., Sandmann C. & Larkey L. (1988). Brain responses to semantic anommally in natural, continuous speech. *International Journal of Psychophysiology*, 6:243–254.

Oldfield R. (1971). The assessmentand analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9:97–113.

Opitz B. (1999). *Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße*. Dissertation, Leipzig.

Opitz B., Mecklinger A., von Cramon D. & Kruggel F. (1999). Combinig electrophysiological and hemodynamic measures of the auditory oddball. *Psychophysiology*, 36:142–147.

Orrin S., Nusbaum H. & Small S. (2001). Cortical neurology Laboratory: Talker normalization. *unreviewed internet publication*, http://angularis.bsd.uchicago.edu/ ∼fmri/Studies/Study_Talker_Normalization/.

Owen A., Herrod N., Menon D., Clark J., Downey S., Carpenter T. et al. (1999). Redefining the functional organization of working memory processes within human lateral prefrontal cortex. *European Journal of Neuroscience*, 11:567–574.

Paavilainen P., Simola J., Jaramillo M., Näätänen R. & Winkler I. (2001a). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, 38:359–365.

Paavilainen P., Valppu S. & Näätänen R. (2001b). The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: 1+1=2 but 1+1+1<3. *Neuroscience Letters*, 301:179–182.

Palmeri T., Goldinger S. & Pisoni D. (1993). episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning Memory and Cognition*, 19(2):309–328.

Pantev C., Elbert T., Ross B., Carsten E. & Terhart E. (1996). Binaural fusion and the representation of virtual pitch in the human auditory cortex. *Hearing Research*, 100:164–170.

Papanicolaou (1998). *Fundamentals of functional brain imaging - a guide to the methods and their applications to psychology and behavioral neuroscience*. Lisse, NL: Sweets & Zeitlinger.

Penhune V., Zatorre R., MacDonald J. & Evans A. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probalilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, 6:661–672.

Phillips M., Pellathy T., Marantz A., Yellin E., Wexler K., Poeppel D. et al. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, 12(6):1038–1055.

Pihan H., Ackermann H. & Altenmüller E. (1997). The cortical procesing of perceived emotion: a DC potential study on affective prosody. *NeuroReport*, 8:623–627.

Poeppel D. & Roberts T. (1996). Effects of vowel pitch and task demands on latency and amplitude of the auditory evoked M100. *Brain and Cognition*, 32(2):156.

Poeppel D. & Rowley H. (1996). Magnetic source imaging and the neural basis of dyslexia. *Annals of Neurology*, 40(2):137–138.

Poeppel D., Yellin C., Rowley H. & Roberts T. (1997). Processing of vowels in the supratemproal auditory cortex. *Neuroscience Letters*, 221:145–148.

Pritchard W. (1991). Psychophysiology of P300. *Psychophysiology*, 89:506–540.

Pritchard W., Shappell S. & Brandt M. (1991). Psychophysiology of N200/N400: A review and classification scheme. *Advances in Psychophysiology*, 4:43–106.

Reite M., Sheeder J., Teale P., Richardson D., Adams M. & J. S. (1995). MEG based brain laterality: Sex differences in normal adults. *Neuropsychologia*, 33(12):1607–1616.

Remez R., Fellowes J. & Rubin P. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23(651):651–666.

Rif J., Hari R., Hämäläinen M. & Sams M. (1991). Auditory attention affects 2 different areas in the human supratemporal cortex. *Electroencephalography and Clinical Neurophysiology*, 79(6):464–472.

Riley E. & Sackheim H. (1982). Ear asymmetry in recognition of unfamiliar voices. *Brain and Cognition*, 1:245–258.

Rinne T., Alho K., Alku P., Holi M., Sinkkonen J., Virtanen J. et al. (1999). Analysis of speech sounds is left-hemisphere pre-dominant 150 ms after sound onset. *NeuroReport*, 10(56):1113–1117.

Rinne T., Alho K., Ilmoniemi R., Virtanen J. & Näätänen R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. *NeuroReport*, 10(56):1113–1117.

Ritter W., Paavilainen P., Lavikainen J., Reinikainen K., Alho K., Sams M. et al. (1992). Event-related potentials to repetition and change of auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 83:306–321.

Roberts T., Ferrari P. & Poeppel D. (1998). Latency of evoked neuromagnetic M100 reflects perceptual and acoustic stimulus attributes. *NeuroReport*, 9(14):3265–3269.

Roberts T., Ferrari P., Stufflebeam S. & Poeppel D. (2000). Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights toward perception. *Journal of Clinical Neurophysiology*, 17(2):114–129.

Roberts T. & Poeppel D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *NeuroReport*, 7(6):1138–1140.

Rösler F. (1982). *Hirnelektrtische Korrelate kognitiver Prozesse*. Berlin etc.: Springer-Verlag.

Rugg M. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory and Cognition*, 18:367–379.

Sachs L. (1997). *Angewandte Statistik. 8. Aufl.* Berlin: Springer-Verlag.

Sams M. & Hari R. (1991). Magnetoencephalography in the study of human auditory information processing. *Annals New York Academy of Sciences*, 620:102–117.

Sams M., Paavilainen P., Alho K. & Näätänen R. (1985). Auditory Frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 62:437–448.

Sandridge S. & Boothroyd A. (1996). Using naturally produced speech to elicit the mismatch negativity. *Journal of the American Academy of Audiology*, 7(2):105–112.

Schacter D. & Church B. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory and Cognotion*, 18(5):913–930.

Scherg M. & Berg P. (1991). Use of prior knowledge in brain electromagnetic source analysis. *Brain Topography*, 4:143–150.

Scherg M. & von Cramon D. (1985). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, 62:32–44.

Scherg M., Vajsar J. & Picton T. (1989). A source analysis of the late human auditory evoked potential. *Journal of Cognitive Neuroscience*, 1:336–355.

Schirmer A., Kotz S. & Friederici A. (in press). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*.

Schröger E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology*, 32:55–65.

Schröger E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments and Computers*, 30(1):131–145.

Schröger E. & Wolff C. (1996). Mismatch response to changes in sound location. *Neuro-Report*, 7:3005–3008.

Schweinberger S. (2001). Human brain potentials correlates of voice priming and voice recognition. *Neuropsychologia*, 39:921–3–936.

Scott S., Blakd C., Rosen S. & Wise J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123:2400–2406.

Senkfor A. & Van Petten C. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24(4):1005–1025.

Shamma S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5(8):340–348.

Sharborough F., Chatrian G., Lesser R., Lüders H., Nuwer M. & Picton T. (1991). American Electroencephalographic Society guidelines for standards electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8:200–202.

Shtyrov Y., Kujala T., Lyytinen H., Kujala J., Ilmoniemi R.J. & Näätänen R. (2000a). Lateralization of speech processing in brain as indicated by mimatch negativity and dichotic listening. *Brain and Cognition*, 43:392–398.

Shtyrov Y., Kujala T., Palca S., Ilmoniemi R.J. & Näätänen R. (2000b). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemisphere. *NeuroImage*, 12:657–663.

Slembek E. (1998). Frauenstimmen hören - ein Vergleich verschiedener Kulturen.

Sonntag G. (1999). *Evaluation von Prosodie*. Dissertation, Bonn.

Springer S. & Deutsch G. (1995). *Linkes Rechtes Gehirn*. Heidelberg etc.: Springer-Verlag.

Squires N., Squires K. & Hillyard S. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38:387–401.

Steinhauer K., Alter K. & Friederici A. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2:191–196.

Stevens A. & Schwartzreich E. (2000). Dissociating message and messenger: fMRI of word and voice recognition. *Journal of Cognitive Neuroscience*, 33B(Suppl. S 2000).

Stufflebeam S., Poeppel D., Rowley H. & Roberts T. (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *NeuroReport*, 9(15):91–94.

Sussman H. (1986). A neural model of vowel normalization and representation. *Brain and Language*, 28:12–23.

Sutton S., Braren M., Zubin J. & John E. (1965). Evoked potential correlates of stimulus uncertainty. *Science*, 150:1187–1188.

Syrdal A. & Gopal H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79:1086–1100.

Takegata R., Paavilainen P., Näätänen R. & Winkler I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, 266:109–112.

Tallal P., Miller S. & Fitch R. (1993). Neurobiological bases of speech: A case of the preeminence of temporal processing.

Tartter V. (1984). Laterality differences in speaker and consonant identification in dichotic listening. *Brain and Language*, 23:74–85.

Taylor R. (1953). The CLOZE-Procedure. *Journalism Quarterly*, 30:415.

Tervaniemi M., Medvedev S., Alho K., Pakhomov S., Roudas M., van Zuijen T. et al. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Human Brain Mapping*, 10:74–79.

Thompson P. (1980). Margret Thatcher: a new illusion. *Perception*, 9:483–484.

Tian B., Teser D., Durham A., Kustov A. & Rauschecker J. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, 313(5515):290–293.

Tiitinen H., Sivonen P., Alku P. & Virtanen J. Näätänen R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Cognitive Brain Research*, 8:355–363.

Titova N. & Näätänen R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neuroscience Letters*, 308:63–65.

Ungerleider L. & Mishkin M. (1982). Two cortical visual systems. In D.J. Ingle & M. Goodale (Hg.), *Analysis of visual behavior*, Cambridge Mass.: MIT Press, 549–586.

Van Berkum J., Haagoort P. & Brown C. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6):657–671.

Van Dommelen W. (1990). Acoustic parameters in human speaker recognition. *Language and Speech*, 33(3):259–272.

Van Lancker D., Cummings J., Kreiman J. & Dobkin B. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex*, 24:195–209.

Van Lancker D. & Kreiman J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, 25(5):829–834.

Van Lancker D., Kreiman J. & Cummings J. (1989). Voice perception deficits: Neuroanatomical correlates of Phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, 11(5):665–674.

Van Petten C. & Kutas M. (1991). Influences of semantic and syntactic context on open and closed class words. *Memory and Cognition*, 19:95–112.

Verleger R. (1988). Event-related potentials and cognition: a critique of the context-updating hypothesis and an alternative interpretation of the P3. *Behavioral and Brain Sciences*, 11:343–356.

Wang J., Williamson S. & Kaufman L. (1992). Magnetic source images determined by a lead-field analysis - the unique minimum-norm least-squares estimation. 39(7):665–675.

Wilson F., O'Scalaidhe S. & Goldman-Rackic P. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, 260:1955–1958.

Winkler I., Kujala T., Tiitinen H., Sivonen P., Alku P., . et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36:638–642.

Wise R., Scott S., Blank C., Mummery C., Murphy K. & E.A. W. (2001). Separate neural subsystems within 'Wernicke's Area'. *Brain*, 124:83–95.

Woldorff M., Hackley S. & Hillyard S. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, 28:30–42.

Wolff C. & Schöger E. (2001). Human pre-attentive auditory change-detection with single , double, and triple deviations as revealed by mismatch negativity additivity. *Neuroscience Letters*, 311:37–40.

Wood C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics*, 15:501–508.

Woods D., Alain C., Covarrubias D. & Zaidel O. (1995). Middle latency auditory evoked potentials to tones of different frequencies. *Hearing Research*, 85:65–75.

Wright R., Frisch S. & Pisoni D. (1997). Speech Perception. *Research on spoken language processing*, 21:1–48.

Yabe H., Tervaniemi M., Reinikainen K. & Näätänen R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, 8:1971–1974.

Young A., Hellawell D. & D.C. H. (1987). Configurational information in face perception. *Perception*, 16:747–759.

Zatorre R. (1997). Cerebral correlates of human auditory processing: perception of speech and musical sounds. In J. Syka (Hg.), *Acoustical Signal Processing in the Central Auditory System*, New York: Plenum Press.

Ziegler W., Jochmann A. & Zierdt A. (1999). Assessment of Auditory Word Comprehension in Aphasia. In P. Maassen B. Groenen (Hg.), *Pathologies of Speech and Language - Advances in Clinical Phonetics and Linguistics*, London: Whurr Pblishers Ltd., 229–235.

Ziegler W. & Kraus B. (1999). Phonological and talker related information interfere in aphasic word comprehension. In E. Witruk & T. Lachmann (Hg.), *Abstracts of the International Conference on Basic Mechanisms of Language and Language Disorders*, Leipzig: Leipziger Universitätsverlag, 74–75.

## MPI Series in Cognitive Neuroscience

1 Anja Hahne

Charakteristika syntaktischer und semantischer Prozesse bei der auditiven Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien

2 Ricarda Schubotz

Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion

3 Volker Bosch

Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale

4 Jorge Jovicich

An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain

5 Rosemary C. Dymond

Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations

6 Stefan Zysset

Eine experimentalpsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie

7 Ulrich Hartmann

Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes

8 Bertram Opitz

Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße

9 Gisela Müller-Plath
Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen

10 Thomas Jacobsen

Characteristics of processing morphological structural and inherent case in language comprehension

11 Stefan Kölsch

Brain and Music: A contribution to the investigation of central auditory processing with a new electrophysiological approach

12 Stefan Frisch

Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen

13 Markus Ullsperger

The role of retrieval inhibition in directed forgetting – an event-related brain potential analysis

14 Martin Koch

Measurement of the Self-Diffusion Tensor of Water in the Human Brain

15 Axel Hutt

Methoden zur Untersuchung der Dynamik raumzeitlicher Signale

16 Frithjof Kruggel

Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie

17 Anja Dove

Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselparadigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie

18 Karsten Steinhauer

Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache

19 Silke Urban

Verbinformationen im Satzverstehen

20 Katja Werheid

Implizites Sequenzlernen bei Morbus Parkinson

21 Doreen Nessler

Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition

22 Christoph Herrmann

Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse

23 Christian Fiebach

Working Memory and Syntax during Sentence Processing. A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging

24 Grit Hein

Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten

25 Monica de Filippis

Die visuelle Verarbeitung unbeachteter Wörter. Ein elektrophysiologischer Ansatz.

26 Ulrich Müller

Die katecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen

27 Kristina Uhl

Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information

28 Ina Bornkessel

The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation