



Studiengang Physik

Diplomarbeit

**Lateralisierung von Sprache und
mögliche auditorische
Grundlagen. Eine fMRT-Studie.**

vorgelegt von
Hagen Wierstorf

Betreuer: Dr. Stefan Uppenkamp
Erster Gutachter: Prof. Dr. Dr. Birger Kollmeier
Zweiter Gutachter: Juniorprof. Dr. Jesko Verhey

Fakultät V, Institut für Physik
Carl von Ossietzky-Universität Oldenburg

Oldenburg, im November 2008

Inhaltsverzeichnis

1	Einleitung	9
2	Sprachverarbeitung im Kortex	13
2.1	Probleme bei der Lokalisierung mit bildgebenden Verfahren	13
2.2	Zwei Verarbeitungszweige	14
2.3	Mögliche auditorische Grundlagen der Lateralisierung	18
2.4	Der <i>Pop-Out-Effekt</i>	20
3	Rotierte Sprache	23
3.1	Einführung	23
3.2	Die einzelnen Schritte der Verzerrung	24
3.3	Verständlichkeit von rotierter Sprache	27
3.3.1	Rotierte Phoneme	28
3.3.2	Wörter und Sätze	32
3.3.3	Was bestimmt die Diskrimination?	34
4	Psychoakustik des Pop-Out-Effektes	37
4.1	Einführung	37
4.2	Versuchsablauf	38
4.3	Ergebnisse	40
4.4	Diskussion	47
5	Untersuchung der Sprachverarbeitung mit fMRT	51
5.1	Einführung	51
5.2	Versuchsablauf	51
5.3	Ergebnisse	56
5.3.1	Mittelwerte	56
5.3.2	Einzelergebnisse	62
5.4	Diskussion	64
6	Zusammenfassung	69
A	Anhang	71
A.1	Blesser Skript	71
A.2	Formanten	72

Inhaltsverzeichnis

A.3 Anatomie und Begriffserklärung	74
A.4 fMRT-Ergebnisse	75
Literaturverzeichnis	77

Abbildungsverzeichnis

1.1	Gehirn des von Paul Broca obduzierten Patienten	10
2.1	<i>Ventraler</i> und <i>dorsaler</i> Sprachverarbeitungsweig	16
2.2	Visuelle Demonstration des <i>Pop-Out-Effektes</i> I	21
2.3	Visuelle Demonstration des <i>Pop-Out-Effektes</i> II	22
3.1	Spektrogramm natürlicher und rotierter Sprache	23
3.2	Spektrum und Zeitsignal eines Tonkomplexes	24
3.3	Die einzelnen Schritte des Rotationsalgorithmus	25
3.4	Langzeitspektrum der natürlichen und der rotierten Sprache	26
3.5	Spektrum und Zeitsignal eines rotierten Tonkomplexes	27
3.6	Spektrum des natürlichen und des rotierten Vokals /a/	29
3.7	Vokaldreieck für rotierte Vokale	31
4.1	Verständlichkeit der rotierten Sprache in Abhängigkeit von ihrem Abstand zur Sprache	40
4.2	Lernkurven für das Verstehen von rotierter Sprache	42
4.3	Verständlichkeit der rotierten Sprache in Abhängigkeit des dargebotenen Satzes I	44
4.4	Verständlichkeit der rotierten Sprache in Abhängigkeit des dargebotenen Satzes II	45
4.5	Verständlichkeit der rotierten Sprache in Abhängigkeit des dargebotenen Satzes III	46
4.6	Verständlichkeit der rotierten Sprache in Abhängigkeit von ihrem Abstand zur Sprache für ausgewählte Sätze	48
5.1	Händigkeit der Versuchspersonen	55
5.2	Verständlichkeit der einzelnen Stimuli im Magnetresonanztomographen	57
5.3	Sprachverarbeitung im Kortex I – „Sprache“ minus „unverständliche rotierte Sprache“	58
5.4	Sprachverarbeitung im Kortex II – „verständliche rotierte Sprache“ minus „unverständliche rotierte Sprache“	59
5.5	Sprachverarbeitung im Kortex III – Unterschied zwischen natürlicher Sprache und verständlicher rotierter Sprache	60

5.6	Sprachverarbeitung im Kortex IV – Sprache im Rauschen	61
5.7	Lateralisierung der Sprachverarbeitung und Händigkeit bei den einzelnen Versuchspersonen	64
A.1	Formanten der natürlichen bzw. rotierten Vokale e, i und u	72
A.2	Formanten der natürlichen bzw. rotierten Vokale o, ä, ü und ö	73
A.3	Anatomie des Gehirns	74
A.4	Gehirnaktivierungen I: Schall – Stille und SR – S	75
A.5	Gehirnaktivierungen II: S – Stille und Rauschen – Stille	76
A.6	Gehirnaktivierungen III: RS2 ⁽⁰⁾ – Stille und RS1 ⁽⁰⁾ – Stille	76

Tabellenverzeichnis

3.1	Identifikation rotierter Vokale	30
4.1	Darbietungsreihenfolge während der psychoakustischen Messung .	39
5.1	Stimulusbedingungen für die Messung im Magnetresonanztomogra- phen	52
5.2	Darbietungsreihenfolge während der Messung im Magnetresonanztomographen	53
5.3	Sprachverarbeitung im Kortex – „verständliche rotierte Sprache“ minus „unverständliche rotierte Sprache“	59
5.4	Lateralisierung der Sprachverarbeitung und Händigkeit bei den ein- zelnen Versuchspersonen	63
A.1	Abkürzungen anatomischer Bezeichnungen	74

1 Einleitung

Whereas elementary functions of a tissue can, by definition, have a precise localization in particular cell groups, there can of course be no question of the localization of complex functional systems in limited areas of the brain or of its cortex.

Alexander R. Lurija, 1976

Ein hoch differenziertes Sprachsystem ist Basis sowohl unserer menschlichen Kommunikation als auch unseres Zusammenlebens. Durch seine Komplexität und Einzigartigkeit ist es Gegenstand intensiver Forschung. Dabei ergeben sich Fragestellungen aus verschiedenen Fachbereichen, beispielsweise der Evolutionsforschung, der Sozialwissenschaften oder der Linguistik. Aus physikalischer und neurowissenschaftlicher Sicht stellt sich insbesondere die Frage nach dem Unterschied von Sprache zu anderen akustischen Reizen und dem Aufbau der Sprachverarbeitung im menschlichen Gehirn. Ziel der vorliegenden Studie ist es, zur genaueren Kenntnis dieses Aufbaus beizutragen.

Die neurowissenschaftliche Untersuchung der Sprachverarbeitung war stets den vorherrschenden Trends des Fachgebietes in Bezug auf die Lokalisierbarkeit von Gehirnfunktionen unterworfen. Diese bewegten sich zwischen den beiden Extrema der ganzheitlichen Verarbeitung durch das Gehirn und der strikten Lokalisierung von Funktionen in bestimmten Neuronenverbänden. In neuerer Zeit ist besonders durch die Verwendung von funktioneller Magnetresonanztomographie wieder eine weniger kritische Haltung gegenüber strikter Lokalisierung zu beobachten. Aus diesem Grund soll mit dem einleitenden Zitat von Alexander R. Lurija daran erinnert werden, dass es für die komplexen und vielseitigen höheren Funktionen des menschlichen Gehirns unwahrscheinlich ist, dass jede von diesen einem eigenen Neuronenverband zugeordnet werden kann. Vielmehr bilden sich zusammenhängende funktionelle Netzwerke, deren einzelne Bestandteile in verschiedenen Netzwerken mitwirken können. Darüber hinaus können sie bei Beschädigung teilweise durch eine Verarbeitung in anderen Bereichen kompensiert werden.¹

¹Um ein einfaches, aber beeindruckendes Beispiel für solch eine Funktionskompensation zu geben, sei hier folgender Fall vorgestellt: Fallen bei einem Patienten Teile des Kleinhirns aus, in denen der automatische Bewegungsablauf des Gehens abgespeichert ist, ist dieser vorerst nicht mehr in der Lage zu gehen. Er kann diesen Vorgang jedoch zurückerlangen, indem er gezwungen wird, den Bewegungsablauf bewusst durchzuführen und dadurch Bereiche in der Großhirnrinde

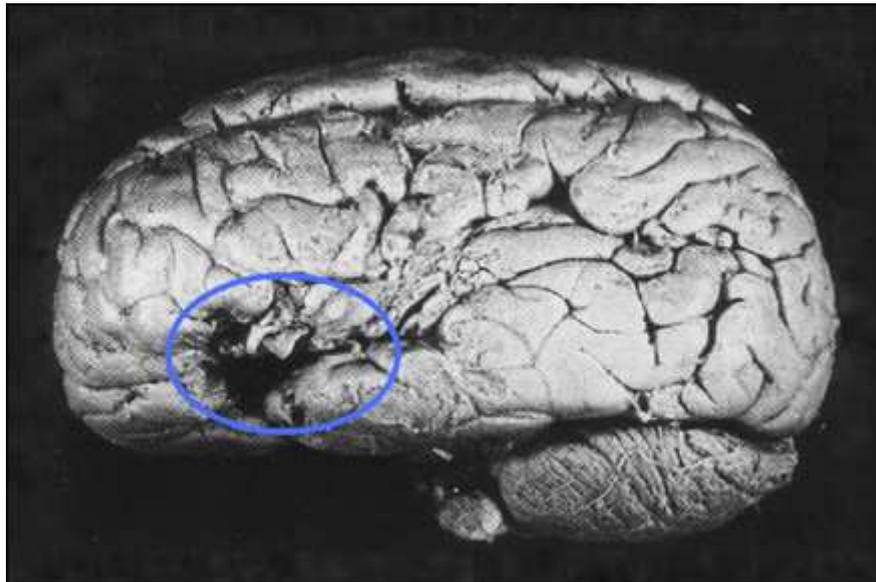


Abbildung 1.1: Das obduzierte Gehirn des französischen Patienten von Paul Broca, der nur noch in der Lage war, die Silbe „tan“ zu sprechen. In dem mit einem blauen Kreis markierten Gebiet, ist eine Läsion des linken *Gyrus frontalis inferior* zu erkennen. Dieser Bereich ist nach diesem Fund als Broca-Areal benannt worden und wird mit der Sprachproduktion in Verbindung gebracht. (Rorden und Karnath, 2004)

Die ersten Ergebnisse zur Lokalisierung der Sprachverarbeitung im menschlichen Kortex stellen die Befunde von Paul Broca dar. Dieser hatte um das Jahr 1860 einen Patienten behandelt, der nur noch die Silbe „tan“ aussprechen konnte. Trotzdem konnte er an ihn gerichteten Text verstehen und versuchte, sich nur mit Hilfe der Satzmelodie und seiner einzigen verbliebenen Silbe zu verständigen. Nach dessen Tod nahm Broca eine Autopsie seines Patienten vor und stellte eine Beschädigung des Kortex im linken *Gyrus frontalis inferior*² fest. Dieser Bereich wird seitdem mit der Produktion von Sprache in Verbindung gebracht und als Broca-Areal bezeichnet. Eine Beschädigung dieses Bereiches und die daraus resultierenden Beeinträchtigungen der Sprachproduktion wird als Broca-Aphasie bezeichnet.

1874 veröffentlichte Carl Wernicke eine Arbeit, in der er eine sensorische Aphasie beschrieb. Diese beinhaltet, dass Patienten noch fließend sprechen können,

dafür zu aktivieren. Dies kann auf beeindruckende Weise gezeigt werden, indem dem Patienten kleine Hindernisse in den Weg gelegt werden, über die er seine Beine heben muss. Mit diesem von außen induzierten Trick ist er wieder in der Lage zu gehen. Solche äußeren Hilfsstellungen konnten vielfach zur Rehabilitation verwendet werden. (Lurija, 1993)

²Erläuterungen zu den anatomischen Begriffen finden sich im Anhang A.3.

aber Probleme mit Wortbedeutungen haben und so ihr Verständnis zunehmend beeinträchtigt wird. Als Konsequenz daraus weist ihre Sprache zunehmend Phantasiewörter auf. Die sensorische Aphasie ging einher mit einer Schädigung einer Region im linken *Gyrus temporalis superior*, welche daraufhin als Wernicke-Areal bezeichnet worden ist. Zudem beschrieb Wernicke als erster das Phänomen, das später als Leitungsbahnaphasie bekannt wurde. Diese zeichnet sich dadurch aus, dass im Extremfall nur das Nachsprechen beeinträchtigt ist. Er vermutete diese Läsion in einem verbindenden Bereich zwischen Wernicke- und Broca-Areal.

Durch diese beiden Arbeiten wurde eine Fülle von Aphasiestudien in Gang gesetzt, die unter anderem durch die beiden Weltkriege mit ihren vielen Kopfverletzten und zu behandelnden Aphasiepatienten Auftrieb erhielt.

Allerdings gibt es bei den Untersuchungen zur Aphasie und den daraus gewonnenen Schlussfolgerungen zur Sprachverarbeitung im Kortex Probleme, die bis heute nicht gelöst werden konnten. Zum einen sind die Läsionen des Kortex durch die Natur bedingt, zum Beispiel durch Schlaganfälle, und können aus offensichtlichen ethischen Gründen nicht künstlich herbeigeführt werden. Läsionen können daher meistens nicht in einem eng umgrenzten Gebiet lokalisiert werden, sondern größere Bereiche des Gehirns sind häufig in Mitleidenschaft gezogen. Zum anderen setzt bei vielen Patienten sehr schnell eine Verbesserung der Sprache ein, was auf eine schnelle Umorganisation des Sprachsystems schließen lässt. Daraus ergibt sich das Problem, dass oft gar nicht der Verlust der Sprache durch die Läsionen des Kortex untersucht wird, sondern vielmehr die Sprachverarbeitung durch das sich neu formierte funktionelle System auf kortikaler Ebene. Auf aktuelle Ergebnisse aus Aphasie-Studien und deren mögliche Deutung im Hinblick auf die Sprachverarbeitung im Kortex wird im nächsten Kapitel eingegangen.

In neuerer Zeit haben bildgebende Verfahren wie etwa Positronen-Emissions-Tomographie (PET) oder funktionelle Magnetresonanztomographie (fMRT) die Möglichkeit geschaffen, Sprachverarbeitung *in vivo* und am gesunden Kortex zu untersuchen. Bei PET wird dem Probanden ein Radiopharmakon verabreicht, welches positive Betastrahlung emittiert. Dieses wird so gewählt, dass es durch den erhöhten Stoffwechselverbrauch in den aktiveren Gehirnregionen angereichert wird. Die frei werdenden Positronen wechselwirken mit den an diesen Stellen befindlichen Elektronen, indem sie in zwei entgegengesetzt fliegende Photonen zerstrahlen. Diese Photonen können mit einer Gamma-Kamera detektiert werden und aus ihrer räumlichen und zeitlichen Verteilung kann auf die Quelle ihrer Aussendung geschlossen werden. Bei fMRT werden die Spins der Atomkerne durch ein starkes äußeres Magnetfeld ausgerichtet und durch ein hochfrequentes Wechselfeld zum Präzedieren um die Feldrichtung angeregt. Nach Abschalten des Wechselfeldes klingt die Präzession ab. Dieser Vorgang unterscheidet sich je nach Gewebeart in seiner Länge. Verschieden starke Aktivierungen im Gehirn führen mit einem zeitlichen Versatz von ungefähr 5 Sekunden zu einer unterschiedlich starken Verteilung

von Hämoglobin im Blut. Das Hämoglobin führt auf diese Weise zu einer lokalen Änderung der Abklingzeit und ermöglicht daher eine funktionelle Messung der Gehirnaktivität.

Die bildgebenden Verfahren haben allerdings nicht – wie vielleicht zu vermuten wäre – zu einer schnellen und endgültigen Klärung des funktionellen Systemes der Sprachverarbeitung im Kortex geführt. Es wurde eine Vielzahl unterschiedlicher Ergebnisse beobachtet, die über große Teile des Kortex verstreut lagen. Die Ursachen dieser widersprüchlichen Forschungsergebnisse und mögliche Auswege aus dieser Situation werden in Abschnitt 2.1 erörtert.

Wie bereits angedeutet, sind das Broca- und das Wernicke-Areal nur in der linken Gehirnhälfte lokalisiert worden. Die meisten Forschungsergebnisse sprechen auch dafür, dass die Sprachverarbeitung größtenteils in der linken Gehirnhälfte stattfindet. Uneinigkeit besteht jedoch darin, wie stark dies der Fall ist und ab welchem Punkt in der Hierarchie der Sprachverarbeitung diese Lateralisierung beginnt. Gleichmaßen stellt sich die Frage, was die physiologische Grundlage für die Lateralisierung darstellt.

Um diese Fragen erhellen zu können, werden in dieser Studie mögliche Verarbeitungszweige der Sprache im Kortex und mögliche Grundlagen für ihre Lateralisierung untersucht. Zu diesem Zweck wird die Verarbeitung des Sinngeltes von Sprache im Kortex mit Hilfe eines gezielten Einsatzes verzerrter Sprache in einer fMRT-Messung lokalisiert. Hierbei wird der Umstand ausgenutzt, dass die Verständlichkeit der verzerrten Sprache von ihrem Darbietungskontext abhängig ist und ein und derselbe akustische Sprachstimulus auf diese Weise mal verständlich und mal unverständlich präsentiert werden kann.

Zunächst soll aber ein Überblick über die Theorien der Sprachverarbeitung im menschlichen Kortex und deren Lateralisierung gegeben werden. Anschließend wird auf die Eigenschaften der verwendeten Sprachverzerrung eingegangen und ihre Verständlichkeit in Abhängigkeit ihres Darbietungskontextes psychoakustisch untersucht. Abschließend wird das durchgeführte fMRT-Experiment vorgestellt und die Ergebnisse werden diskutiert.

2 Sprachverarbeitung im Kortex

2.1 Probleme bei der Lokalisierung mit bildgebenden Verfahren

Soll die Sprachverarbeitung mit funktioneller Bildgebung wie PET oder fMRT untersucht werden, stellt sich das Problem der Trennung dieser Verarbeitung von rein auditorischen Vorverarbeitungen. Die Isolierung bestimmter Stimuluseigenschaften wird in der Regel über geschickt gewählte Kontraste realisiert. Mit einem Kontrast wird der statistisch signifikante Unterschied zwischen den Aktivierungen im Kortex für zwei unterschiedliche Stimulusbedingungen bezeichnet. Die ersten Vergleichsstimuli, die im Zusammenhang mit der Sprache zur Erzeugung von Kontrasten gewählt wurden, waren meistens einfache akustische Stimuli wie Töne, Rauschen oder sogar Stille (vgl. Binder et al., 2000). Da diese Stimuli sich jedoch in vielen akustischen Eigenschaften von Sprache unterscheiden, ist es schwierig die beobachteten Aktivierungen auf reine Sprachverarbeitung zurückzuführen und von einer allgemeinen auditorischen Verarbeitung zu trennen. So ist ein Hauptmerkmal der Studien mit sich akustisch deutlich von der Sprache unterscheidenden Vergleichsstimuli, dass sie bilaterale Aktivierungen gezeigt haben (vgl. Scott und Wise, 2004). Es ist daher der Versuch unternommen worden, einen Vergleichsstimulus zu finden, der akustisch Sprache sehr nahe kommt, aber trotzdem nicht verstanden werden kann. Eine häufig verwendete Variante stellt dabei rückwärts abgespielte Sprache dar (Binder et al., 2000). Diese hat jedoch den Nachteil, dass sie ein anderes zeitliches Verhalten als Sprache hat. Das Zeitsignal von Sprache weist keine symmetrische Struktur auf, so dass es bei seiner zeitlichen Invertierung zu Lauten kommen kann, die nicht mehr artikuliert werden können. Deshalb bleibt bei rückwärts abgespielter Sprache ein großer Unterschied zu normaler Sprache bestehen. Aus diesem Grund ist es wiederum schwierig, die gewonnenen Aktivierungen zu interpretieren.

Eine andere Vorgehensweise stellt die Verwendung von Fremdsprachen dar. Für diese sind ähnliche akustische Eigenschaften wie bei der Muttersprache zu erwarten, da es auch hier Vokale und Konsonanten gibt. Außerdem kann die Artikulation einer Fremdsprache – zumindest bis zu einem gewissen Alter – erlernt werden. Auf der anderen Seite gibt es Unterschiede in den phonologischen Regeln, die trotzdem für akustische Unterschiede sorgen können. Dazu gehört beispielsweise die Frage, wie viele Konsonanten auf einen Vokal folgen dürfen. Außerdem besteht ein

größerer Bereich der Überschneidung mit den Sprachstimuli, da natürlich auch in Fremdsprachen einzelne Phoneme identifiziert werden oder bekannt sein können. In diesem Fall ist es daher ebenfalls nicht einfach, die gewonnenen Kontraste zu interpretieren.

Einen Schritt in Richtung der Lösung dieses Problems stellt die Verwendung von rotierter Sprache (Blesser, 1972) als Vergleichsstimulus zu Sprache dar. Auf diese Weise ist es Scott et al. (2000) gelungen, Aktivierungen durch die Sprache nur im linken Temporallappen außerhalb des auditorischen Kortex zu zeigen. Eine Reihe von weiteren Studien (Narain et al., 2003; Spitsyna et al., 2006; Awad et al., 2007; Sabri et al., 2008) hat ebenfalls rotierte Sprache als Vergleichsstimuli benutzt und ähnliche Ergebnisse erzielt. Die genauen Eigenschaften der rotierten Sprache werden in Kapitel 3 näher beschrieben. An dieser Stelle kann jedoch zusammenfassend festgestellt werden, dass bei ihr sowohl der spektro-temporale Informationsgehalt als auch der zeitliche Ablauf und die Tonhöhe der Sprache intakt bleibt.

Weiterhin gilt es, die linguistische Komplexität der Stimuli bei den einzelnen Studien zur Sprachverarbeitung zu beachten. Scott et al. (2000) haben kurze Sätze verwendet und eine Aktivierung im anterioren Temporallappen gefunden. Binder et al. (2000) hingegen haben mit einzelnen Wörtern eher Bereiche im mittleren und posterioren Teil des *Gyrus temporalis superior* gefunden. So gilt es mittlerweile als gesichert, dass zusammenhängende Sätze weiter anterior im Temporallappen zu Aktivierungen führen als zufällige Wortlisten (Mazoyer et al., 1993).

Es ist zu erkennen, dass auch durch den Einsatz von funktionellen bildgebenden Verfahren die Sprachverarbeitung im menschlichen Kortex nicht einfach zu untersuchen ist. Es muss sowohl eine angemessene Vergleichsbedingung als auch die Komplexität der Sprache bedacht werden. Zudem haben Sabri et al. (2008) nachgewiesen, dass die Aufmerksamkeit der Probanden einen größeren Einfluss auf die beobachteten Aktivierungen haben kann als dies etwa Scott und Wise (2004) noch angenommen haben. Die Rolle der Aufmerksamkeit wird in Kapitel 5 in Zusammenhang mit der durchgeführten fMRT-Messung noch eingehender behandelt werden.

Im nächsten Abschnitt sollen einige Theorien vorgestellt werden, die aus den Ergebnissen der Bildgebung und der Aphasiestudien versuchen, ein konsistentes Bild der Sprachverarbeitung im menschlichen Kortex zu formen.

2.2 Zwei Verarbeitungszweige

Die vielen unterschiedlichen Ergebnisse zur Sprachverarbeitung, sei es durch Läsionsstudien oder bildgebende Verfahren, weisen auf einen wichtigen Punkt bei der Sprachverarbeitung hin: die hohe Parallelität in der Verarbeitung. Das Verstehen von Sprache ist unter sehr vielen verschiedenen Bedingungen möglich: unterschied-

liche Sprecher, variierende Dialekte, Sprache im Störgeräusch oder verzerrte Sprache. Deshalb kann die Sprachverarbeitung nicht durch ein einfaches Umwandeln akustischer Merkmale in phonetische Symbole beschrieben werden. Vielmehr ist davon auszugehen, dass eine gleichzeitige Verarbeitung unterschiedlichster Merkmale geschieht, die sich gegenseitig beeinflussen können und schließlich zu einem „Verstehen“ der Sprache führen. Mit hoher Wahrscheinlichkeit verhält es sich so, dass je komplizierter dieser Prozess des Verstehens abläuft, desto mehr und höhere Funktionen im Kortex an der Sprachverarbeitung beteiligt sind. Ein eindrucksvolles Beispiel hiervon stellt die Fähigkeit dar, bestimmte Dialekte nach wenigen Minuten verstehen zu können.

Aus Studien zur auditorischen Verarbeitung ist bekannt, dass die Verarbeitung von akustischen Stimuli hierarchisch organisiert ist. Dabei werden komplexere Signale weiter lateral zum primären auditorischen Kortex verarbeitet, bis zu einem Bereich in dem rechten anterioren *Sulcus temporalis superior* für Melodien (Patterson et al., 2002). In der auditorischen Verarbeitung zeigen sich somit Analogien zu der hierarchischen Verarbeitung im visuellen Kortex (Zeki, 2000). Beachtenswert ist allerdings die Tatsache, dass die einzelnen Verarbeitungsschritte im visuellen Kortex zumeist räumlich voneinander getrennt werden können. Dies trifft im auditorischen Kortex nur eingeschränkt zu.

Aus der Vielzahl der Ergebnisse zur Sprachverarbeitung, aus konzeptionellen Überlegungen und der Analogie zum visuellen Kortex haben eine Reihe von Autoren eine Aufspaltung in einen *dorsalen* und einen *ventralen* Verarbeitungszweig für die Sprache vorgeschlagen (Scott und Wise, 2004; Hickok und Poeppel, 2004, 2007; Davis und Johnsrude, 2007). Dem *ventralen* Zweig werden dabei eher semantische und syntaktische Aufgaben zugeschrieben. Der *dorsale* Zweig stellt hingegen eine Verbindung zwischen den wahrgenommenen Sprachstimuli und ihrer Artikulation her. Über die genaue Lokalisation der einzelnen Zweige im Kortex herrscht jedoch noch Unklarheit.

Davis und Johnsrude (2007) gehen davon aus, dass sich der *ventrale* Verarbeitungszweig vom auditorischen Kortex aus in der linken Hemisphäre in anteriorer Richtung auf dem Temporallappen ausbreite und Verbindungen zu anterioren Gebieten im Frontallappen besitze. Diese Regionen werden von ihnen vor allem mit einer akustischen Repräsentation des Sprachsignals in Verbindung gebracht. Je weiter die Repräsentation sich in anteriorer Richtung befinde, desto länger seien die Zeiträume, über die das Zeitsignal aufintegriert wird. Diese These ist besonders durch die Beobachtung, dass Sätze weiter anterior als zufällige Wortlisten verarbeitet werden, bekräftigt. Damit gehen sie davon aus, dass in diesem Gebiet alle Prozesse stattfinden, die ein längeres Zeitfenster benötigen. Dazu gehört die Verarbeitung der Satzmelodie, der Syntax oder der lexikalen Segmentierung. Der *dorsale* Verarbeitungszweig diene hingegen der Verbindung der Sensorik mit der Motorik. Dies führt dazu, dass die wahrgenommene Sprache hier in phonologische

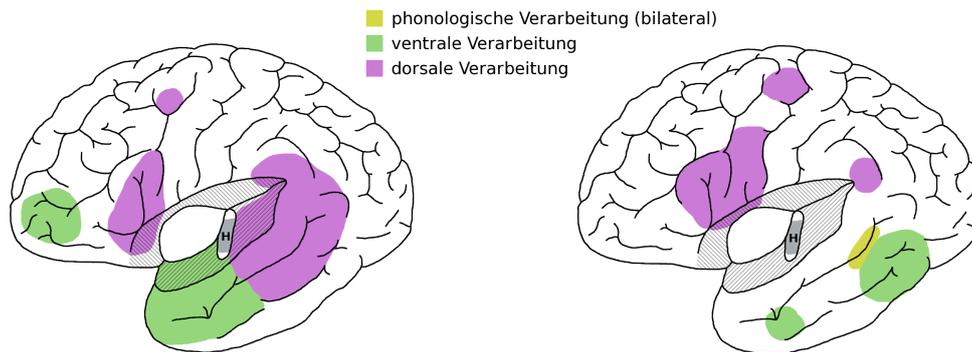


Abbildung 2.1: Links: Ventraler und dorsaler Verarbeitungszweig nach Davis und Johnsruide (2007). Als Untermenge sind darin die Vorschläge von Scott und Wise (2004) enthalten. Rechts: Ventraler und dorsaler Verarbeitungszweig nach Hickok und Poeppel (2007). Zusätzlich schlagen diese noch eine bilaterale phonologische Vorverarbeitung vor. Der *ventrale* Zweig wird in allen Theorien mit dem Sprachverstehen und der *dorsale* Zweig mit der Sprachartikulation in Verbindung gebracht. In Grau ist jeweils der primäre auditorische Kortex in der Heschl'schen Querwindung (H), dem *Gyrus temporalis transversus*, eingezeichnet.

Kategorien zerlegt werden kann, die aufgrund ihrer möglichen Artikulierbarkeit erkannt werden. Diesen Bereich ordnen sie in der linken Hemisphäre dem posterioren Teil des Temporallappens, dem *Gyrus angularis*, dem *Gyrus supramarginalis*, dem Broca-Areal, und dem Prämotor- und Motorkortex zu. Diese Zuordnung wird größtenteils durch bildgebende Versuche zu phonologischen Kategorien und Verbindungen im Kortex, wie sie in Primaten gefunden worden sind, bestätigt.

Scott und Wise (2004) schreiben den einzelnen Verarbeitungszweigen ähnliche Funktionen zu und lokalisieren sie auch an ähnlichen Orten im Kortex. Der *ventrale* Verarbeitungszweig sei für die Umwandlung des akustischen Sprachsignals in eine sinnhafte Bedeutung verantwortlich. Sie haben ihn in einem PET-Experiment (Scott et al., 2000) und einer fMRT-Messung (Narain et al., 2003) jeweils mit rotierter Sprache als Vergleichsstimulus im anterioren Temporallappen lokalisiert. Gleichzeitig sprechen Ergebnisse von Patienten mit progressiver Aphasie für diese Lokalisierung. Diese Patienten verlieren mit der Zeit ihren gesamten Wortschatz. In zwei weiteren Studien wurde dieser Bereich sowohl in dem Zusammenhang von Sprachverstehen und Sprachproduktion (Awad et al., 2007) als auch in dem Zusammenhang von Sprachverstehen und Lesen (Spitsyna et al., 2006) bestätigt, wobei nicht alle Aktivierungen nur in der linken Hemisphäre zu finden waren. Ein kleiner Bereich im bilateralen anterioren Temporallappen zeigt außerdem mehr Aktivierung für Sätze als für zufällige Wortlisten (Mazoyer et al., 1993).

Der *dorsale* Verarbeitungszweig wird auch von Scott und Wise (2004) als eine Verbindung der Sensorik mit der Sprachartikulation angesehen. Dieser Bereich

erstreckt sich in der linken Hemisphäre in posteriorer Richtung vom auditorischen Kortex aus den Temporallappen entlang und reicht bis in den *Gyrus supramarginalis* mit Verbindungen zum Motorkortex. Wie aus Aphasiestudien mit Patienten, die Wortfindungsschwierigkeiten zeigten oder nicht mehr in der Lage waren, Sprache zu verstehen, hervorgeht, ist dieser Bereich allerdings auch mit Wortkodierungen befasst. Scott und Wise gehen jedoch nicht ausführlicher auf den dorsalen Verarbeitungszweig ein, da sie sich darauf konzentrieren, Belege für die Verarbeitung im anterioren Temporallappen zu finden. Diese Region ist erst in aktuelleren Studien mit der Sprachverarbeitung in Verbindung gebracht worden und als solche umstritten. Bei einer Reihe von Aphasiestudien bleibt das Sprachverständnis trotz Schädigung dieses Bereiches intakt. Allerdings führt eine beidseitige Läsion in fast allen Fällen zu einem Verlust des Sprachverständnisses. Dies könnte dafür sprechen, dass bei einseitigem Verlust die andere Hemisphäre diesen Ausfall kompensieren kann.

Hickok und Poeppel (2007) haben in der neuesten Variante ihrer Modellvorstellung auch einen Bereich im anterioren Temporallappen aufgenommen, welcher in ihrem vorherigen Modell (Hickok und Poeppel, 2004) fehlte. Ansonsten weist ihre Vorstellung jedoch noch deutliche Unterschiede zu den in dieser Studie bisher vorgestellten auf. Sie betonen die Rolle des posterioren Temporallappens in der Verarbeitung von Semantik und haben daher Teile davon in den *ventralen* Verarbeitungszweig mit aufgenommen (Abb. 2.1, rechts). Zusätzlich haben sie einen weiteren Verarbeitungszweig für die phonologische Kodierung eingeführt, der bilateral im posterioren *Sulcus temporalis superior* zu finden sein soll. Allerdings sollen alle Gebiete, die zwischen dem posterioren *Sulcus temporalis superior* und dem *Gyrus temporalis transversus* liegen, nur mit der Verarbeitung akustischer Stimuli beschäftigt sein. Der *ventrale* Verarbeitungszweig stellt auch bei ihnen eine Verbindung zur Artikulation der Sprache dar und ist in der linken Hemisphäre über den *Gyrus supramarginalis* mit dem Broca-Areal und weiteren Motor-Arealen verbunden.

Zusammenfassend soll noch einmal festgehalten werden, welche Ergebnisse als weitestgehend gesichert gelten können: Die Sprachverarbeitung findet größtenteils in der linken Hemisphäre statt, eine ausführlichere Diskussion hierzu wird im nächsten Abschnitt folgen. Dabei ist sie in hohem Maße parallel und benötigt akustische Repräsentationen des Satzes. Es gibt eine Verbindung zum Broca-Areal und dem Motorkortex. Diese verläuft posterior zum auditorischen Kortex und erstreckt sich über die Verbindung zwischen Temporal-, Okzipital- und Parietallappen. In diesem *dorsalen* Zweig findet eine Verarbeitung der Sprache auf Grund ihrer Artikulierbarkeit statt, die höchstwahrscheinlich genutzt wird, um sie in kategoriale Einheiten zu zerlegen. Es herrscht Uneinigkeit darüber, was diese Einheiten alles darstellen, ob es sich nur um die Grundeinheiten der Sprache

handelt, wie etwa Phoneme oder Silben¹, oder doch gleich um ganze Wörter. Gesichert ist jedoch, dass dieser Verarbeitungszweig Verbindungen mit dem *ventralen* Verarbeitungszweig aufweist (Spitsyna et al., 2006), der damit in der Lage ist, komplexere Zusammenhänge wie die Semantik oder Syntax von Sätzen zu kodieren. Beeinflusst wird er dabei durch Aktivitäten im anterioren Frontallappen (Davis und Johnsrude, 2007).

2.3 Mögliche auditorische Grundlagen der Lateralisierung

Schon die frühen Aphasiestudien von Broca und Wernicke haben nahe gelegt, dass die Verarbeitung von Sprache im Kortex nicht symmetrisch stattfindet, sondern zur linken Hemisphäre lateralisiert ist. In Kombination mit den Ergebnissen aus neueren Aphasiestudien ergibt sich folgendes Bild: Bei einer Läsion beider Temporallappen oder bei Läsion im linken Temporallappen lassen sich Störungen bei der Sprachverarbeitung feststellen, nicht jedoch bei einer alleinigen Läsion in der rechten Hemisphäre (Scott und Wise, 2004). Im zuletzt genannten Fall sind jedoch Störungen in der Verarbeitung von Musik zu beobachten (Johnsrude et al., 2000).

Ergebnisse aus bildgebenden Verfahren lassen sich bezüglich der Lateralisierung nicht so leicht interpretieren, da es ohne genügend angepasste Vergleichsstimuli stets zu starker bilateraler Aktivierung im auditorischen Kortex kommt. Ein möglicher Lösungsansatz wäre, die Untersuchung auf Bereiche im Kortex außerhalb des auditorischen Kortex zu beschränken, von denen eine Aktivität während der Sprachverarbeitung bekannt ist. Nachteil dieser Methode ist allerdings, dass nur eine bestimmte Unterfunktion der Sprachverarbeitung untersucht würde.

Bei den Untersuchungen zur Lateralisierung von Sprache hat sich ein Zusammenhang mit der Händigkeit der Probanden gezeigt. Dieser ist allerdings deutlich geringer als vielleicht zu vermuten wäre. So haben Pujol et al. (1999) bei 50 Linkshändern und 50 Rechtshändern mit Hilfe von fMRT versucht, ihre Sprachlateralisierung zu bestimmen. Zu diesem Zweck wurde eine stille Wortgenerierungsaufgabe verwendet und ausschließlich die Aktivität im Broca-Areal, dem dorsolateralen prefrontalen Kortex und dem Prämotorkortex untersucht. Dabei zeigte sich bei keinem der Rechtshänder² eine Aktivität ausschließlich in der rechten Hemisphäre

¹Phoneme erscheinen uns durch unseren regelmäßigen Umgang mit geschriebener Sprache als die natürlichen Spracheinheiten. Es hat sich jedoch gezeigt, dass Menschen, die das Alphabet nicht beherrschen, große Schwierigkeiten darin haben, einzelne Wörter zu segmentieren (Pettersson et al., 2000). Beim Aussprechen verschmelzen benachbarte Phoneme stets zu neuen eigenen Lauten, so dass davon auszugehen ist, dass größere Einheiten – wie etwa Silben – viel eher als die kleinste Einheit von gesprochener Sprache fungieren.

²Bestimmt mit dem *Edinburgh Handedness Inventory* (Oldfield, 1971)

und nur bei 10% der Linkshänder war dies der Fall. Umgekehrt zeigten 96% der Rechtshänder und 74% der Linkshänder ausschließlich Aktivität in der linken Hemisphäre. Zu vergleichbaren Ergebnissen kamen Knecht et al. (2000) mit einer anderen Messmethode, der transkranialen Doppler-Ultraschographie. Folglich ist eine Abhängigkeit der Sprachlateralisierung von der Händigkeit zu beobachten, mit der Einschränkung, dass auch bei den meisten Linkshändern eine Lateralisierung der Sprache in die linke Hemisphäre besteht.

Eine weitere interessante Frage stellt der Beginn der Lateralisierung dar, werden akustische Reize doch zunächst immer bilateral im auditorischen Kortex verarbeitet. Für die Lateralisierung der Musikverarbeitung in die rechte Hemisphäre konnte diese Frage bereits geklärt werden. Patterson et al. (2002) haben gezeigt, dass die Kodierung einer sich nicht ändernden Tonhöhe bilateral im *Gyrus temporalis superior* stattfindet, während durch Einführen einer Melodie aus unterschiedlichen Tonhöhen, der rechte posteriore *Sulcus temporalis superior* zusätzlich aktiviert wird.

Bei der Lateralisierung der Sprachverarbeitung ist dieser Beginn der Lateralisierung jedoch noch nicht geklärt. Die Läsionen sind häufig viel zu großflächig, als daraus Schlussfolgerungen gezogen werden könnten. Die Ergebnisse aus den bildgebenden Verfahren werden hingegen in zweierlei Hinsicht interpretiert. Zum einen gibt es Hinweise, dass einzelne Phoneme noch bilateral verarbeitet werden (Uppenkamp et al., 2006). Auf der anderen Seite hat es in letzter Zeit eine ganze Reihe von Studien gegeben, die gezeigt haben, dass es schon bei der Verarbeitung einfacher akustischer Stimuli, die noch keine Sprachinformationen enthalten, zu einer Lateralisierung in der Verarbeitung kommt. Poeppel und Hickok, die in ihrem Modell zur Sprachverarbeitung von einer ersten bilateralen Verarbeitungsstufe für phonologische Informationen ausgehen, nehmen an, dass die Verarbeitung im sekundären auditorischen Kortex in den unterschiedlichen Hemisphären auf unterschiedlichen Zeitskalen abläuft. Dabei kommt im linken *Gyrus temporalis superior* ein Zeitfenster von 25-50 ms zum Einsatz und auf der rechten Seite eines mit einem Zeitfenster von 200-300 ms (Boemio et al., 2005). Anschließend empfängt der linke *Sulcus temporalis superior* mehr von dem Eingangssignal aus dem linken *Gyrus temporalis superior*. Für den rechten *Sulcus temporalis superior* gilt entsprechend das Umgekehrte. Mit diesem Ergebnis erklären Poeppel und Hickok die Lateralisierung der Melodieverarbeitung.

Zatorre et al. (2002) gehen einen Schritt weiter und behaupten, die Lateralisierung der Verarbeitung von sich auf schnellen Zeitskalen ändernden Signalen (Zatorre und Belin, 2001; Schönwiesner et al., 2005; Jamison et al., 2006) in die linke Hemisphäre sei eine ausreichende Erklärung für die Lateralisierung von Sprache. Zu beachten bleibt, dass die Ergebnisse für die Lateralisierung von sich schnell ändernden Signalen nicht unumstritten sind (Scott und Wise, 2004). Allerdings gibt es auch eine Reihe von anatomischen Hinweisen, welche diese These

begünstigen, zum Beispiel die unterschiedliche Anordnung und Myelinisierung der Neurone im linken und rechten primären auditorischen Kortex und Ergebnisse aus Untersuchungen zur Frequenzabstimmung einzelner Neurone (Liégeois-Chauvel et al., 2001). Diese Beobachtung haben Zatorre et al. (2002) zu ihrer *Trade-Off-Hypothese* veranlasst. Dabei gehen sie davon aus, dass auf Grund der Unschärfe-Beziehung zwischen der Zeit und der Frequenz, die beiden Hemisphären nicht beide Eigenschaften mit derselben feinen Auflösung bearbeiten können. Daher habe sich eine Spezialisierung der linken Hemisphäre für sich zeitlich schnell ändernde Signale und eine Spezialisierung der rechten Hemisphäre für sich in kleinen Frequenzschritten ändernde Signale herausgebildet. Wobei unklar bleibt, wieso dies zu einer Lateralisierung und nicht zu zwei parallelen Verarbeitungswegen in den beiden Hemisphären führen sollte.

Ebenso erscheint die Hypothese von Zatorre et al. in Bezug auf die Sprachlateralisierung recht unwahrscheinlich. So handelt es sich bei Sprache um ein komplexes akustisches Signal, bei dem es nicht nur auf die schnellen zeitlichen Änderungen ankommt. Eine parallele Verarbeitung aller akustischer Eigenschaften erscheint außerdem mehr als wahrscheinlich. Ein zusätzliches Problem kommt zum Vorschein, wenn Zatorre et al. ausführlich argumentieren, dass für die Sprachverarbeitung nur tonale Änderungen in der Größenordnung einer Oktave nötig sind, wie sie bei der Satzmelodie zu beobachten sind. Diese Argumentation mag für ihre Muttersprache – das Englische – noch zutreffen. Auf der anderen Seite muss von ihrer These ein Großteil der Menschheit ausgeschlossen werden, denn in den tonalen Sprachen spielen deutlich geringere Tonhöhenunterschiede in der Größenordnung von Halbtonschritten eine sinnunterscheidene Rolle.³

In der vorliegenden Studie wird mit Hilfe des im nächsten Abschnitt vorgestellten *Pop-Out-Effektes* die Verarbeitung des Sinngehaltes der Sprache untersucht. Außerdem gibt es mit der in Kapitel 3 vorgestellten rotierten Sprache einen Vergleichsstimulus, der akustisch und vor allem zeitlich derart an die unverzerrte Sprache angepasst ist, dass auf diesem Gebiet kein Unterschied in der akustischen Verarbeitung zu erwarten ist. Nach der Hypothese von Zatorre et al. (2002) wäre auf der Verarbeitungsstufe für die rotierte Sprache die Lateralisierung schon abgeschlossen. Es sollte daher kein Unterschied mehr zwischen diesem Kontrast und der Verarbeitung des Sinngehaltes der Sprache geben. Die Hypothese dieser Studie lautet daher, dass bei der Verarbeitung des Sinngehaltes der Sprache eine eindeutige Lateralisierung in die linke Hemisphäre zu beobachten sein wird. Sollte sich allerdings bei dem Vergleich der Verarbeitung der unverzerrten Sprache mit der rotierten Sprache noch immer eine bilaterale Aktivierung zeigen, wäre dies ein eindeutiger Widerspruch zu der These von Zatorre et al. (2002). Es wäre dann eher ein Hinweis auf eine später einsetzende Lateralisierung der Sprachverarbeitung, die

³Beim von fast einer Milliarde Menschen gesprochenem Hochchinesisch (auch Mandarin genannt) handelt es sich etwa um eine tonale Sprache.

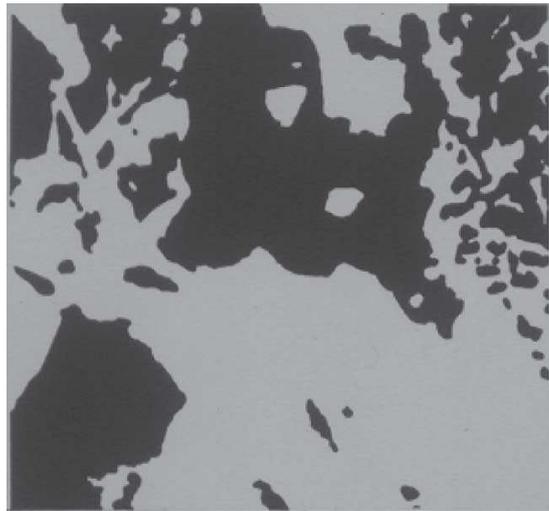


Abbildung 2.2: Visuelle Demonstration des *Pop-Out-Effektes*: das Bild ist so stark verzerrt, dass nur noch schwarze und graue Flächen zu erkennen sind und kein Objekt mehr. Nach Betrachtung von Abbildung 2.3 sollte in diesem Bild allerdings auch deutlich ein Objekt zu sehen sein. (Ahissar und Hochstein, 2004)

nicht auf den akustischen Merkmalen von Sprachsignalen beruhen würde.

2.4 Der Pop-Out-Effekt

Eine große Herausforderung für die Untersuchung der Sprachverarbeitung mit bildgebenden Verfahren stellt die Suche nach einem möglichst sprachähnlichen Stimulus dar. Das Problem kann jedoch auch von einer anderen Seite gelöst werden. Es ist nämlich durchaus möglich, für akustisch identische Sprachstimuli unterschiedliche Wahrnehmungen zu erzielen. So sind Versuchspersonen in der Lage, verzerrte und für sie unverständliche Sprache durch entsprechendes Kontextwissen zu verstehen. Dieses Kontextwissen steuert in einem *top-down*-Prozess die Wahrnehmung, und sie sind der festen Überzeugung, die entsprechenden Wörter zu verstehen. Dabei macht es auf die Probanden den Eindruck, als ob diese plötzlich hervorspringen würden. Deshalb soll dieser Effekt im Folgenden mit dem Begriff *Pop-Out-Effekt* bezeichnet werden (Davis et al., 2005).

Eine Beschreibung davon lässt sich schon bei Cherry (1957) finden:

D.B. Fry has given a remarkable demonstration of the way in which a priori knowledge bears upon recognition. He has made a gramophone recording of two men holding a conversation, but with their speech so artificially distorted that not a word can be recognized. After one playing of the record, the listener is informed that the speakers are discussing the subject of buying a



Abbildung 2.3: Visuelle Demonstration des *Pop-Out-Effektes*: Hinweis auf das Objekt in Abbildung 2.2. Nach Betrachtung dieses Bildes sollte der „Mann mit Bart“ auch in jener Abbildung problemlos zu sehen sein (Ahissar und Hochstein, 2004).

new suit; they refer to their tailors, the price of clothes, styles, et cetera. The record is then played a second time, and most listeners are able to follow the conversation. The words jump out at one.

Die einfachste Methode diesen Effekt hervorzurufen besteht darin, nach einem verzerrten Sprachsignal das gleiche Signal unverzerrt und anschließend wieder verzerrt darzubieten. Bei der letzten Darbietung wird dann auch das verzerrte Signal verstanden. Eine Demonstration dieses Effektes ist unter <http://medi.uni-oldenburg.de/download/diplom/hw08/> zu finden. Bei der für die Demonstration verwendeten Verzerrung handelt es sich um Sinussprache (Remez et al., 1981). Dabei werden die ersten drei Formanten des Sprachsignals durch Sinustöne angenähert und zu einem Signal zusammengesetzt. Der Effekt ist allerdings unabhängig von der verwendeten Verzerrungsmethode und funktioniert auch mit Sprache in Rauschen, rauschangeregter Sprache (vgl. Davis und Johnsruide, 2007) und, wie in dieser Studie noch gezeigt wird, auch mit rotierter Sprache.

Der Kontext muss jedoch nicht immer ein akustischer sein. Liegt der unverzerrte Satz ausschließlich in gedruckter Form vor, funktioniert der Effekt genauso. Ebenso kann es ausreichen, wie im Beispiel von Cherry (1957), wenn der Sinnzusammenhang, also z.B. eine Zusammenfassung der kommenden Aussage oder die Einbettung in einen Erzählstrang, gegeben ist. Weiterhin haben Davis et al. (2005) durch einen Vergleich zwischen geschriebener und akustischer Darbietung des unverzerrten Satzes gezeigt, dass das Verständnis der verzerrten Sprache bei ihrer Darbietung nach der unverzerrten Version mit hoher Wahrscheinlichkeit nicht durch das Merken oder Einprägen akustischer Cues funktioniert. Es ist also davon

auszugehen, dass das Hervortreten der Wörter durch *top-down*-Prozesse erklärt werden kann, welche die Ebenen der akustischen und phonologischen Wahrnehmung beeinflussen oder teilweise ignorieren (Davis und Johnsrude, 2007).

Somit ist mit Hilfe des *Pop-Out-Effektes* die Möglichkeit gegeben, kontextabhängig einen identischen akustischen Stimulus in unverständlicher und in verständlicher Form zu präsentieren.

Der *Pop-Out-Effekt* soll in dieser Studie benutzt werden, um die Sprachverarbeitung und ihre Lateralisierung näher untersuchen zu können. Um eine möglichst gute Vergleichbarkeit mit bisherigen Studien (Scott et al., 2000; Sabri et al., 2008) zu erreichen und auch bei den akustischen Stimuli möglichst eng an wirklicher Sprache zu bleiben, soll der Effekt dazu im Folgenden mit rotierter Sprache hervorgerufen werden. Dazu wird im anschließenden Kapitel zunächst näher in die rotierte Sprache eingeführt, um danach den *Pop-Out-Effekt* für die rotierte Sprache in einem psychoakustischem Experiment näher zu quantifizieren und ihn abschließend in der fMRT-Messung der Sprachverarbeitung zu benutzen.

3 Rotierte Sprache

3.1 Einführung

Die Idee der Sprachverzerrung durch eine Spiegelung oder auch Rotation des Spektrums wurde das erste Mal von Blesser (1972) angewandt. Dazu wurde das Sprachspektrum tiefpassgefiltert, durch eine Multiplikation mit einem Sinus gespiegelt und die obere Hälfte durch einen erneuten Tiefpass unterdrückt. In Abbildung 3.1 sind die Spektrogramme des bei 5,7 kHz tiefpassgefilterten Satzes (S) „Achte auf die Autos“ und der verzerrten Version desselben Satzes (RS) dargestellt. Es ist zu erkennen, dass das verzerrte Signal um eine Frequenz von 3 kHz rotiert worden ist. Diese Rotation kommt einer Spiegelung gleich, es wird aber zusätzlich noch eine Verstärkung der hohen Frequenzen vor der Spiegelung durchgeführt. Dies ist deutlich an der Energie des rotierten „s“ zu erkennen.

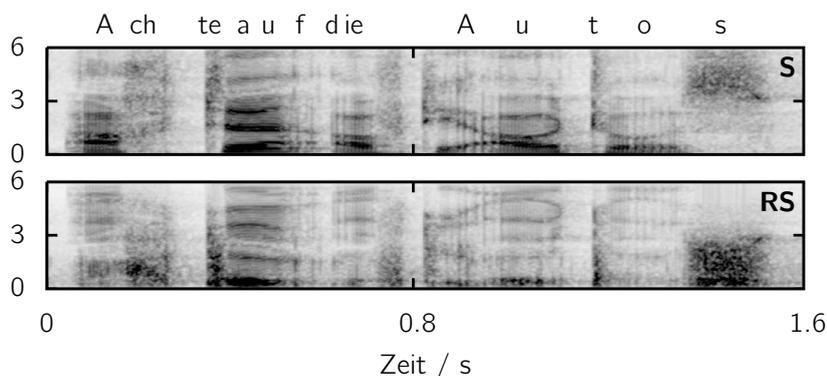


Abbildung 3.1: Spektrogramm des bei 5,7 kHz tiefpassgefilterten Satzes „Achte auf die Autos“. Es ist sowohl seine unverzerrte Version (S), als auch seine rotierte Version (RS) zu sehen. Zudem sind die einzelnen Phoneme ihrem zeitlichen Auftreten entsprechend auf der oberen Achse aufgetragen. Für die farbliche Kodierung wurde die vierte Wurzel des Spektrogramms aufgetragen, um einen guten Kontrast zu erzielen.

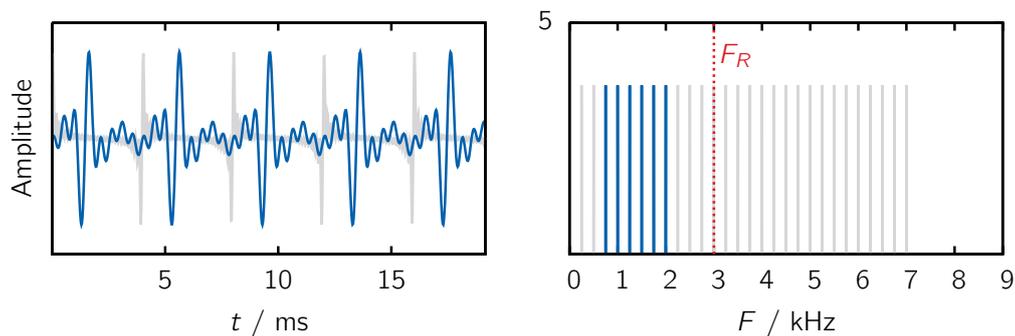


Abbildung 3.2: *Blau:* Tonkomplex mit einer Grundfrequenz von 250 Hz, bestehend aus sechs Komponenten zwischen 750 und 2000 Hz. *Hellgrau:* breitbandige Ergänzung (250-7000 Hz) des Tonkomplexes mit gleicher Grundfrequenz. *Rot:* Frequenz $F_R = 3000$ Hz, um die rotiert werden soll. Auf der linken Seite ist jeweils das Zeitsignal dargestellt und auf der rechten Seite das Spektrum.

3.2 Die einzelnen Schritte der Verzerrung

Die Rotation des Sprachspektrums wurde mit Matlab durchgeführt. Die verwendete Funktion ist im Anhang A.1 ausgeführt, dabei handelt es sich um eine Abänderung der digitalen Version nach Scott et al. (2000).¹ An dieser Stelle sollen die einzelnen Verarbeitungsschritte an Hand eines Beispiels näher erläutert werden.

Als Ausgangssignal für die Rotation wird ein Tonkomplex gewählt, der aus sechs Komponenten besteht und in Abbildung 3.2 in Blau dargestellt ist. Zur Veranschaulichung der Filter ist zusätzlich noch eine breitbandige Version des Tonkomplexes in Hellgrau eingezeichnet. Der Tonkomplex hat eine Grundfrequenz von 250 Hz, beginnt bei 750 Hz und endet bei 2000 Hz. Er soll um eine Frequenz F_R von 3000 Hz rotiert werden, welche in Rot eingezeichnet ist.

Die hauptsächlichen Schritte des Rotationsalgorithmus sind die folgenden:

Schritt 1

Tiefpassfilterung des Ausgangssignals bei einer Frequenz von $0.95 \cdot 2F_R$. Dazu wird ein elliptischer Filter sechster Ordnung verwendet. Gefiltert wird vorwärts und rückwärts, um eine Null-Phasen-Filterung zu erreichen. Die auf diese Weise gefilterten Signale werden in den folgenden Versuchen immer als unverzerrte „Original“-Sprachsignale verwendet. Auf diese Weise ist garantiert, dass sie sich spektral nur durch die Rotation von den verzerrten Signalen unterscheiden und nicht zusätzlich durch ihre spektralen Breiten. Die Verständlichkeit der Sprach-

¹Die Originalsoftware von Scott ist unter <ftp://ftp.phon.ucl.ac.uk/pub/matlab/Blessier.zip> herunterzuladen und stellt eine digitale Nachampfindung der analogen Verzerrung von Blessier dar.

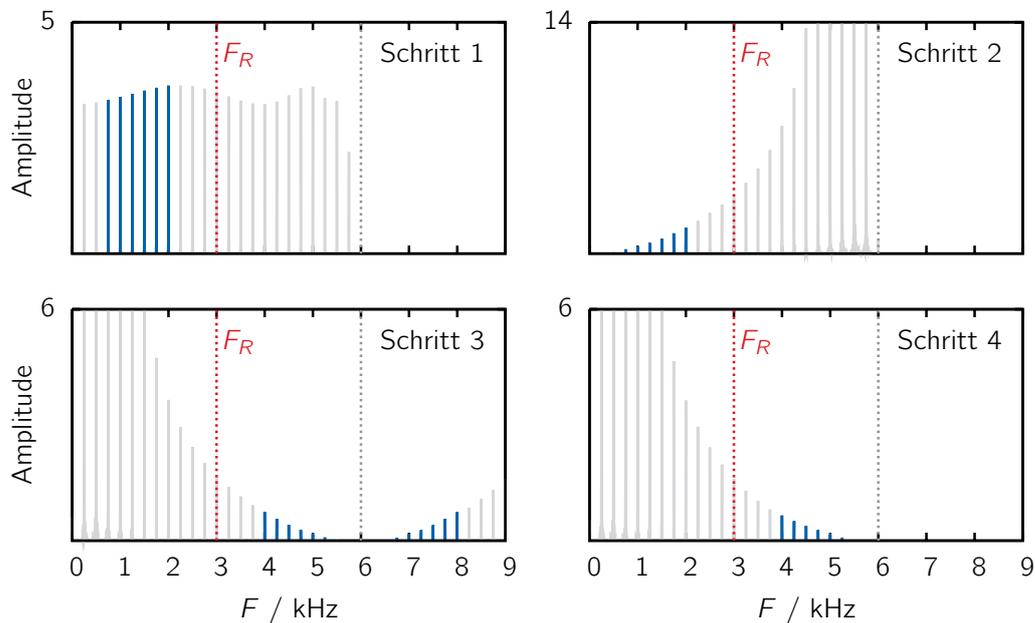


Abbildung 3.3: Die Auswirkungen der einzelnen Schritte des Rotationsalgorithmus auf die in Abbildung 3.2 vorgestellten Tonkomplexe. Die einzelnen Schritte sind dabei: (1) Tiefpassfilterung, (2) Verstärkung der Energie bei hohen Frequenzen, (3) Multiplikation mit Sinussignal und (4) erneute Tiefpassfilterung. In Rot ist die Frequenz F_R , um die rotiert werden soll und in Dunkelgrau $2F_R$ dargestellt.

signale wird durch diesen Schritt nicht beeinträchtigt. Im Folgendem wird mit „dem Ausgangssignal“ oder „dem unverzerrten Sprachsignal“ immer auf dieses tiefpassgefilterte Sprachsignal Bezug genommen.

Schritt 2

Anpassung des Spektrums, um sicher zu stellen, dass die rotierte Sprache ein ähnliches Langzeitspektrum erhält wie die unverzerrte Sprache. Da das Langzeitspektrum von Sprache (siehe Abbildung 3.4) sein Maximum unterhalb von 1 kHz aufweist und danach abfällt, läuft die Anpassung auf eine Hochpassfilterung vor der Spiegelung hinaus. Mit diesem Vorgehen soll erreicht werden, dass das Langzeitspektrum nach der Spiegelung dem von unverzerrter Sprache ähnlich ist. Für diese Hochpassfilterung wurde ein Filter nach Byrne et al. (1994) genommen, der ebenfalls in Abbildung 3.4 dargestellt ist. Als digitalisierte Version wurde die von Scott et al. (2000) zur Verfügung gestellte verwendet.² Byrne hatte die Langzeitspektren für unterschiedliche Sprachen untersucht und war zu dem Schluss gekommen,

²zu finden unter <ftp://ftp.phon.ucl.ac.uk/pub/matlab/Blessor.zip>

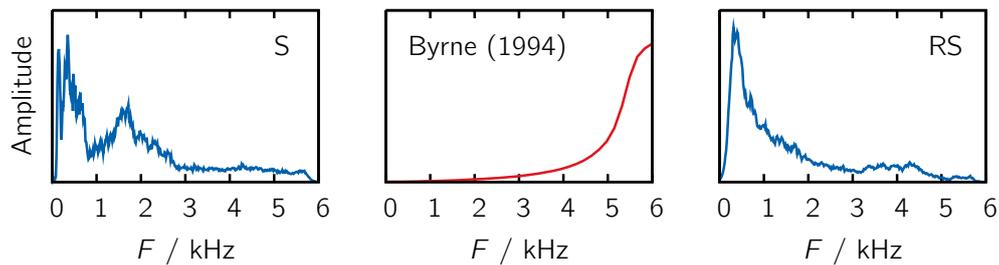


Abbildung 3.4: Links: Geglättetes mittleres Langzeitspektrum der unverzerrten Sprachsignale aus dem GÖSA-Sprachmaterial (mit einem Tiefpassfilter auf 5,7 kHz begrenzt). Mitte: Filter nach Byrne et al. (1994) zum Ausgleichen des Sprachspektrums vor dem Rotieren, um ein ähnliches Langzeitspektrum zu erhalten. Rechts: Geglättetes mittleres Langzeitspektrum der rotierten Sprachsignale aus dem GÖSA-Sprachmaterial.

dass sie sich so sehr gleichen, dass von einem allgemeinem Langzeitspektrum für alle Sprachen und Sprecher ausgegangen werden kann. Folglich bedarf es keiner speziellen Anpassung des Filters an das verwendete Sprachmaterial.

Als Sprachmaterial wurden in dieser Studie Sätze aus dem Göttinger Satztest (GÖSA, Kollmeier und Wesselkamp (1997)) verwendet. In Abbildung 3.4 sind deren mittlere Langzeitspektren vor und nach der Verzerrung aufgetragen. Das von Byrne ermittelte durchschnittliche Langzeitspektrum hat einen ähnlichen Verlauf wie das für die verzerrte Sprache in dem rechten Graphen in Abbildung 3.4 dargestellte: es enthält einen fast monotonen Abfall ab einem Maximum um 500 Hz. Bei dem Langzeitspektrum für die unverzerrten GÖSA-Sätze fällt somit auf, dass es einen Einbruch kurz vor 1 kHz aufweist.

Schritt 3

Multiplikation des Signals mit einem Sinus der Frequenz F_R :

$$RS' = \sin(2\pi F_R t) \cdot S' \quad (3.1)$$

Dies ist der Schritt, bei welchem die Spiegelung stattfindet. Mit S' soll das schon ausgeglichene unverzerrte Sprachsignal bezeichnet werden und mit RS' das rotierte aber noch nicht wieder tiefpassbegrenzte Signal. Wie bei Schritt 3 in Abbildung 3.3 zu erkennen ist, handelt es sich bei RS' um ein um die Frequenz $2 \cdot F_R$ gespiegeltes Signal. In diesem Signal sind alle Frequenzen $F_i^{S'}$ aus S' auf die Frequenzen $F_i^{RS'} = 2F_R - F_i^{S'}$ und $F_j^{RS'} = 2F_R + F_j^{S'}$ abgebildet worden.

Diese Beziehung führt dazu, dass nach dem nächsten Schritt alle Frequenzen aus S in RS um eine Achse bei der Frequenz F_R gespiegelt erscheinen.

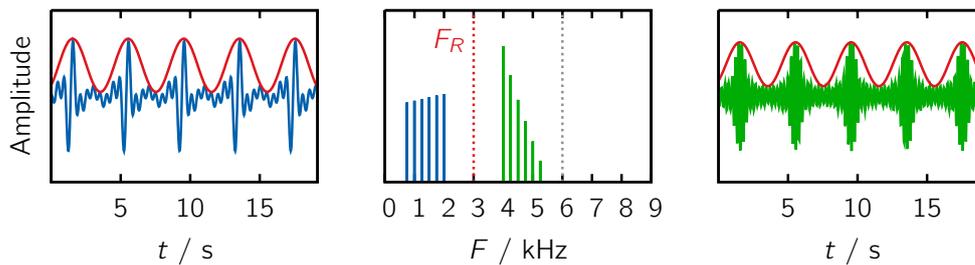


Abbildung 3.5: *Links:* Zeitsignal des tiefpassgefilterten Tonkomplexes. Dieser befindet sich im Spektrum (Mitte) ebenfalls auf der linken Seite. In Rot ist die Einhüllende mit einer Frequenz von 250 Hz eingezeichnet. *Mitte:* Spektren der beiden Tonkomplexe: links von F_R das Ausgangssignal nach der Tiefpassfilterung und rechts von F_R nach der Rotation. *Rechts:* Zeitsignal des rotierten Tonkomplexes. In Rot ist die Einhüllende eingezeichnet, die wieder eine Frequenz von 250 Hz aufweist.

Schritt 4

Hier erfolgt eine erneute Tiefpassfilterung bei einer Frequenz von $0.95 \cdot 2F_R$ wie in Schritt 1. Auf diese Weise wird der obere Teil des durch die Multiplikation entstandenen symmetrischen Spektrums entfernt. Es verbleibt ein Spektrum, dessen Frequenzen um eine Achse bei $F_R = 3000$ Hz gespiegelt worden sind, so dass ein neuer Tonkomplex mit Komponenten zwischen 4000 und 5250 Hz entstanden ist. Da die Spiegelung keinen Einfluss auf den Abstand zwischen den einzelnen Komponenten hat, beträgt auch die Grundfrequenz des rotierten Tonkomplexes 250 Hz. Allerdings sind die Frequenzen der einzelnen Komponenten des rotierten Tonkomplexes nicht notwendigerweise ganzzahlige Vielfache der Grundfrequenz, da sie von der verwendeten Rotationsfrequenz abhängen. Nur für den hier gezeigten Fall, dass die doppelte Rotationsfrequenz ($2F_R$) selber ein Vielfaches der Grundfrequenz ist, bleibt auch dies erhalten.

In Abbildung 3.5 sind sowohl die Zeitsignale des unverzerrten und des verzerrten Tonkomplexes als auch ihre beiden Spektren dargestellt. Durch die Rot eingezeichnete Einhüllende in den Zeitsignalen ist die unveränderte Grundfrequenz der beiden Signale angedeutet.

3.3 Verständlichkeit von rotierter Sprache

Nachdem das Sprachsignal rotiert worden ist, stellt sich die Frage, inwiefern die Verständlichkeit der Sprache durch die Rotation beeinträchtigt wird. Hierzu sollen zunächst einzelne Aspekte wie zum Beispiel der Einfluss der Rotation auf Phoneme, untersucht werden. Später wird zu komplizierteren Einheiten wie ganzen Sätzen übergegangen.

Wie im vorherigen Abschnitt gezeigt, ändert die Rotation der Sprache nichts an ihrer Grundfrequenz. Daraus folgt, dass die Satzmelodie und die Tonhöhe des Sprechers erhalten bleibt. Gleiches gilt selbstverständlich für die zeitliche Reihenfolge der Sprache, da diese durch die rein spektrale Verzerrung unangetastet bleibt. Folgende Eigenschaften bleiben daher ebenfalls erhalten: stimmlos - stimmhaft, Plosiv - kein Plosiv, Frikativ - kein Frikativ. Der Ausgleich des Langzeitspektrums führt darüber hinaus zu einer gleichbleibenden Lautheit der rotierten Sprachsignale.

Im Gegensatz dazu bleiben die Vokale nicht erhalten, da ihre Formantstruktur durch die Rotation verändert wird und sie zum Teil wie andere Vokale der gleichen oder einer fremdartigen Sprache klingen können. Auch die Übergänge zwischen einzelnen Formanten bleiben nicht erhalten. Um diesen Änderungen detaillierter auf den Grund zu gehen, wird im folgenden Abschnitt der Einfluss der Rotation auf einzelne Phoneme betrachtet.

3.3.1 Rotierte Phoneme

Vokale

Die Wahrnehmung von Vokalen hängt von unterschiedlichen Eigenschaften ab, die nicht alle spektraler Natur sein müssen. So hat z.B. die Länge eines Vokals einen Einfluss auf seine Wahrnehmung. Da die Rotation der Vokale einzig ihr Spektrum verzerrt, sollen im Folgendem nur spektrale Eigenschaften Betrachtung finden. Um Vokale spektral klassifizieren zu können, reicht es in der Regel (speziell für westliche Sprachen) aus, ihre ersten beiden Formanten oder das Verhältnis aus diesen zu betrachten (Peterson und Barney, 1952). Daher stellt sich die Frage nach der Auswirkung der Verzerrung auf eben diese Formanten. Hierbei zeigt sich, dass die Verzerrung stark von der gewählten Rotationsfrequenz F_R abhängt, da die Formanten jeweils bei dieser Frequenz gespiegelt werden.

Im Gegensatz zum hier vorliegenden Versuch hatte Blesser (1972) eine Frequenz von 1600 Hz gewählt. Dies führte dazu, dass es bei ungeübten Versuchspersonen zu sehr vielen Verwechslungen zwischen Vokalen kam, die vorne und solchen, die hinten artikuliert werden. Blessers Ergebnis lässt sich mit der entsprechenden Verschiebung der Formanten erklären. So werden z.B. die ersten beiden Formanten des /i:/ (271, 2232 Hz)³ nach der Rotation denen des /u:/ ähnlich:

$$\begin{aligned} F_1^- &= 2F_R - F_3^{/i:/} = 222 \approx 295 = F_1^{/u:/} \\ F_2^- &= 2F_R - F_2^{/i:/} = 968 \approx 813 = F_2^{/u:/} \end{aligned} \quad (3.2)$$

³Als Referenz für die Formanten der Vokale ist im Folgenden die erste deutsche Stimme der espeak-Software genommen worden (<http://espeak.sourceforge.net>, Version 1.31). Die wav-Dateien der Vokale sind zudem unter <http://medi.uni-oldenburg.de/download/diplom/hw08/> erreichbar.

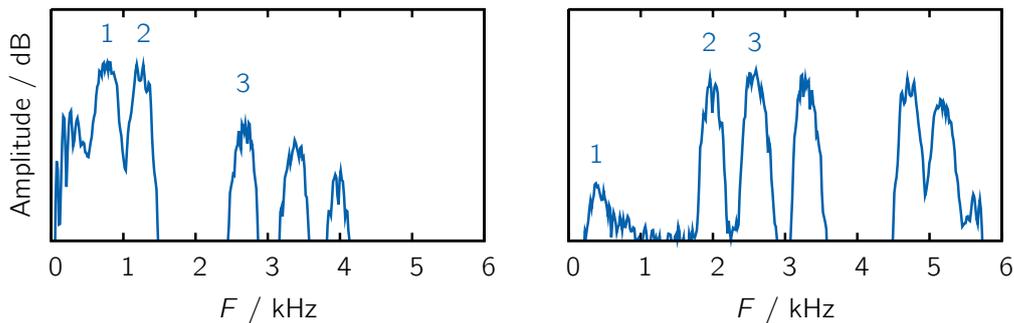


Abbildung 3.6: Das Spektrum des Vokals /a/ (links) und des rotierten Vokals /a/ (rechts). Zur besseren Darstellung ist jeweils über 20 benachbarte Punkte geglättet worden. Die ersten drei Formanten sind jeweils durchnummeriert worden.

Dabei bezeichnet F_j^- die Formanten des rotierten /i/. Zu einem vergleichbaren Ergebnis kamen Azadpour und Balaban (2008). Sie hatten um eine Frequenz von 1500 Hz rotiert, wobei die Versuchspersonen ein rotiertes /i/ und /ε/ als /o/, ein rotiertes /u/ als /i/ und ein rotiertes /a/ und /o/ als /ε/ identifizierten.

In dieser Studie führt die deutlich höhere Rotationsfrequenz⁴ von $F_R = 3000$ Hz zu einem anderen Ergebnis bei den rotierten Vokalen. Der vorher jeweils fünfte Formant wird bei allen Vokalen zum zweiten Formanten und der vorher sechste zum ersten, wodurch dieser jeweils nur sehr schwach ausgeprägt ist (siehe beispielhaft den Vokal /a/ in Abbildung 3.6 und Anhang A.2 für die restlichen Vokale). Der sechste Formant ist in der Regel bei den unverzerrten Vokalen nicht mehr identifizierbar. Er wird nur durch die zunehmende Verstärkung bei hohen Frequenzen während der Verzerrung als erster Formant des rotierten Vokals sichtbar. Um zu überprüfen, ob dieser noch wahrnehmbar ist, sind die rotierten Vokale von 15 Versuchspersonen identifiziert worden. Dazu wurden ihnen die um 3000 Hz rotierten Vokale⁵ diotisch über Kopfhörer dargeboten. Aufgabe der Versuchspersonen war es auf einem Zettel jeweils den Vokal zu notieren, den sie meinten gehört zu haben. Die Ergebnisse sind in Tabelle 3.1 aufgeführt. Es zeigen sich für alle Vokale relativ eindeutige perzeptive Zuordnungen. Lediglich das rotierte /i:/ (↔) ist ungefähr gleich häufig als fünf unterschiedliche Vokale wahrgenommen worden und zeigt daher kein klares Zuordnungsmaximum. Daraus kann gefolgert werden,

⁴Die Begründung für die Verwendung einer höheren Rotationsfrequenz beruht zum einen darauf, dass für die Messung im Magnetresonanztomographen möglichst breitbandige Stimuli verwendet werden sollten, um so höhere Aktivierungen zu erzielen. Außerdem sollten die Stimuli möglichst sprachähnlich sein und gleichzeitig für die Verwendung des *Pop-Out-Effektes* geeignet sein. Siehe Kapitel 4 für eine empirische Begründung.

⁵Die Vokale sind ebenfalls mit der ersten deutschen Stimme der *espeak*-Software generiert worden (<http://espeak.sourceforge.net>, Version 1.31). Die wav-Dateien der Vokale sind unter <http://medi.uni-oldenburg.de/download/diplom/hw08/> erreichbar.

	/a/	/ε/	/i:/	/o:/	/u:/	/ε:/	/y:/	/ø:/
æ	0	2	0	0	0	1	4	8
e	0	3	1	0	0	3	7	1
i	2	3	0	0	0	4	3	3
o	0	3	12	0	0	0	0	0
u	0	0	2	0	0	0	11	1
ä	0	5	10	0	0	0	0	0
ü	0	10	3	0	0	2	0	0
ö	0	11	2	0	0	0	2	0

Tabelle 3.1: Zuordnung der rotierten Vokale (Schriftzeichen, æ ≐ rotiertes /a/) durch 15 Versuchspersonen zu Vokalkategorien (in Phonemnotation). Es ist die jeweilige Anzahl an Zuordnungen zu einer Vokalkategorie eingetragen. Die höchste Anzahl an Zuordnungen zu einer Vokalkategorie ist für jeden rotierten Vokal durch Fettdruck gekennzeichnet.

dass der erste Formant bei den meisten rotierten Vokalen noch wahrnehmbar ist. Um diese These zu bekräftigen, wurden die ersten beiden Formanten der rotierten Vokale in ein Vokaldreieck eingetragen (Abbildung 3.7). Damit sollte überprüft werden, ob ihre dortige Position mit den Zuordnungen durch die Versuchspersonen übereinstimmt. Das Vokaldreieck selbst wurde aus den unverzerrten Versionen der Vokale bestimmt. Hierzu wurden ihre Spektren über das gesamte Zeitsignal gebildet. Die Ausdehnung der einzelnen Vokale ist aus der spektralen Breite der jeweiligen Formanten bestimmt und als gestrichelte Ellipse eingetragen worden. Für die Formanten der rotierten Vokale wurden keine Ausdehnungen betrachtet, sondern die jeweils höchste Amplitude als Formantposition angenommen. Es ist zu erkennen, dass die rotierten Vokale sich ausnahmslos in der Nähe des /ε/ befinden. Trotzdem wurden nicht alle von ihnen als /ε/ erkannt, sondern oft als der Vokal, der in Richtung ihrer Abweichung vom Mittelpunkt des /ε/ anzutreffen ist. Gleichzeitig lässt sich eine Übereinstimmung mit dem Vokaldreieck erkennen, welche die Schlussfolgerung zulässt, dass der erste Formant der rotierten Vokale noch ausreichend wahrnehmbar war.

Nichtsdestotrotz nahm die Mehrheit der Versuchspersonen die rotierten Vokale nicht als Vokale wahr, d.h. sie wären nicht von sich aus auf die Idee gekommen, die Stimuli einzelnen Vokalen zuzuordnen. Aus diesem Grund bleibt die Zuordnung zu einem bestimmten Vokallaut nicht immer konstant. So hat etwa Warren (1999) gezeigt, dass die Wahrnehmung sinnfreier Äußerungen sehr verschieden sein kann und stark vom Kontext abhängt.

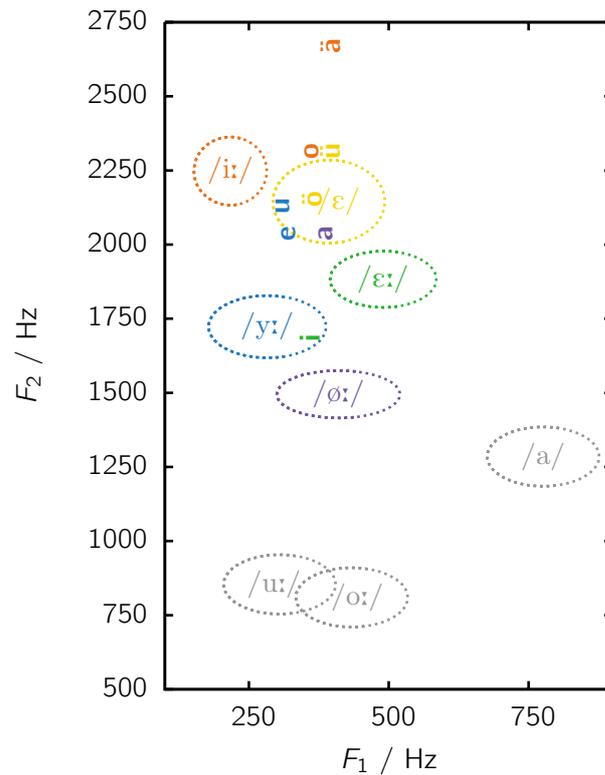


Abbildung 3.7: Dargestellt ist der Abstand der einzelnen Vokale zueinander an Hand ihrer ersten beiden Formanten F_1 und F_2 . Die gestrichelten Kreise stellen das Gebiet der unverzerrten Vokale dar (die genaue Bestimmung dieses Vokaldreiecks wird im Text beschrieben). Die rotierten Vokale sind mit dem um 90° gedrehtem Schriftzeichen für ihren Ausgangsvokal bezeichnet und in der Farbe ihrer Wahrnehmung kodiert. Beispielsweise fällt das rotierte /i:z/, mit der Bezeichnung -- in das Gebiet des /y:z/, wurde aber von der Mehrzahl der Versuchspersonen als /ε:z/ wahrgenommen.

Konsonanten

Die Wahrnehmung von Konsonanten ist ebenso wie bei den Vokalen von ihrem Umfeld abhängig, wobei festgehalten werden muss, dass die Wahrnehmung der Vokale flexibler ist. Dies beruht auf der Tatsache, dass diese nicht so starken Kategorien zugeordnet sind wie die Konsonanten (Patel, 2008), d.h. ein Vokal kann leichter an einen vorgegebenen Kontext angepasst werden.

Die Diskriminationsleistung für Konsonanten, welche in eine Silbe eingebettet sind (z.B. „pan“ und „tan“ oder „girl“ und „curl“), lag bei Blesser (1972) um 90%, wobei eine Unterscheidung für die beiden Konsonantengruppen „Quelle“ und „Ort“ zu treffen ist. Die „Ort“-Gruppe lag bei ungefähr 80% Diskriminationsleistung, während die „Quellen“-Gruppe 95% erreichte. Diese Ergebnisse erforderten zudem

ein Training von mindestens 45 Minuten und waren stark davon abhängig, ob das Wahrgenommene als Sprache erkannt worden ist. Im selben Zusammenhang wurde die Zuordnung der Konsonanten untersucht. Hier zeigte sich, dass nur 36% der Konsonanten richtig erkannt worden waren.

Zu den in dieser Studie verwendeten Konsonanten, die um eine Frequenz von $F_R = 3000$ Hz rotiert worden sind, liegen keine Diskriminations- oder Zuordnungsergebnisse vor. Es werden nur ganze Sätze verwendet, und die Ergebnisse zu den Konsonanten spielen bei deren Wahrnehmung nur eine untergeordnete Rolle, wie sich in den nächsten Abschnitten zeigen wird.

3.3.2 Wörter und Sätze

Bevor im nächsten Abschnitt die Ergebnisse von Blesser (1969) für die Verständlichkeit von rotierten Sätzen und Wörtern vorgestellt werden sollen, wird zunächst kurz auf den von ihm gewählten Versuchsablauf eingegangen.

Blessers Experiment sah vor, dass jeweils zwei Versuchspersonen versuchen sollten, mit Hilfe der rotierten Sprache zu kommunizieren. Dazu wurde ihre gesprochene Sprache der jeweils anderen Person verzerrt über Kopfhörer dargeboten. Zum Abschluss dieses 45-minütigen „Gesprächs“ führte Blesser zusätzlich verschiedene Identifikations- und Diskriminationstests mit den Versuchspersonen durch. Insgesamt fanden 20 dieser Gespräche mit anschließenden Tests je Versuchsperson statt.

Wörter

Zur Überprüfung der Wortverständlichkeit wurden den Versuchspersonen ein- und zweisilbige Wörter vorgespielt (die zweisilbigen Wörter waren dabei immer aus zwei Wörtern zusammengesetzt wie z.B. „*blackboard*“ oder „*horseshoe*“). Anschließend sollten die Versuchspersonen notieren, welche Wörter sie verstanden hatten. Es zeigte sich, dass das Verstehen von einzelnen Wörtern auch nach dem letzten Versuchstag nicht über 14% für die einsilbigen und nicht über 9% für die zweisilbigen Wörter lag. Richtig erkannt wurde zwar die Anzahl der Silben, die Betonung und die Konsonant-Vokal-Abfolge, aber bei der „Ort der Artikulations“-Eigenschaft der Vokale und Konsonanten kam es zu häufigen Verwechslungen. Die Versuchspersonen waren nur selten in der Lage, ein Wort zu verstehen. Besser sahen die Ergebnisse aus, bei denen die Wörter in einem gewissen semantischen Kontext dargeboten worden sind. Wurde die Kategorie mit angegeben, aus der die Wörter stammen könnten, so waren Erkennungsleistungen bis zu 70% möglich. Selbiges wurde z.B. für die Kategorie „Ortsnamen, Staaten und Länder“ erreicht. Für eine weitere Darbietung der Wörter im Kontext wurden Sätze verwendet, welche die Versuchspersonen auf einem Blatt Papier stehen hatten. Dieses lag während des Experimentes vor ihnen, allerdings fehlte jeweils ein Wort. Hier

wurde eine Erkennungsleistung von 25% erreicht, wobei es häufig zu semantischen Verwechslungen kam. So wurde z.B. „Frankreich“ anstatt „Paris“ gehört und umgekehrt. Gleichzeitig waren sich die Versuchspersonen ihrer Fehler nicht bewusst, sondern davon überzeugt, wirklich das falsche Wort gehört zu haben.

Diese Ergebnisse von Blesser lassen zusammenfassend darauf schließen, dass es bei der Wahrnehmung von rotierter Sprache stark auf den Kontext ankommt und häufig die Wörter verstanden werden, die der Versuchsperson als erstes ins Bewußtsein kommen und in den semantischen Zusammenhang passen.

Sätze

Neben den Tests mit einzelnen Wörtern hat Blesser (1969) auch das Verstehen ganzer Sätze untersucht. Diese Sätze bestanden im Durchschnitt aus acht Wörtern. Auch hier fand eine Aufteilung in eine kontextabhängige und eine kontextunabhängige Messung statt. In der kontextunabhängigen Messung wurden den Versuchspersonen mehrere zusammenhanglose Sätze dargeboten und sie mussten notieren, was sie verstanden hatten. Im Gegensatz dazu wurde den Versuchspersonen in der kontextabhängigen Messung gesagt, worum es in der folgenden kurzen Erzählung gehen wird. Außerdem bildeten die Sätze eine zusammenhängende Geschichte.

Bei der kontextunabhängigen Messung waren die Ergebnisse der Versuchspersonen zunächst nicht besser als dies für das Verstehen einzelner Wörter der Fall war. Sie erreichten lediglich eine Verständlichkeit von 5%. Interessanterweise haben sie aber die syntaktische Struktur der Sätze einwandfrei erkannt. So wurde z.B. der Satz „*Hoist the load to your left shoulder*“ von einer Versuchsperson als „*Turn the page to the next lesson*“ erkannt. Ebenso stimmte die Segmentierung der Sätze in einzelne Wörter stets mit dem Originalsatz überein, obwohl durch die zu erwarteten Phonemverwechslungen des öfteren eine andere Segmentierung zu erwarten gewesen wäre. Daraus lässt sich schließen, dass der Prosodie des Satzes eine wichtige Rolle bei der Erkennung seiner syntaktischen Struktur und seiner Segmentierung zukommt (zumindest in Abwesenheit von semantischen Eigenschaften). Im Gegensatz zu den kontextlosen Wörtern war bei den Ergebnissen zu den kontextunabhängigen Sätzen eine deutliche Lernkurve zu erkennen, und die Ergebnisse hatten sich bei der letzten Messung auf 35% Erkennungsleistung gesteigert.

Ein anderes Bild zeigte sich, wenn der Kontext, über den die Sätze handeln, bekannt war. In diesem Fall waren nicht nur die Erkennungsleistungen mit 50-60% deutlich besser, sondern es dominierte auch die Semantik über die Syntax. Bei dieser Messung wurde nicht mehr in allen Fällen die Syntax richtig erkannt, sondern diese teilweise so verändert und konstruiert, dass der Inhalt in den Kontext passte. Auf den dargebotenen rotierten Satz „*I have a special two-week trip to Europe*“ waren z.B. folgende Antworten zu vernehmen:

*I have a travel plan for you.
I had expected to take a trip to Europe.*

Dies scheint ein allgemeines Ergebnis für Sprache zu sein, so haben z.B. Matys et al. (2005) gezeigt, dass *bottom-up*-Segmentierungsmerkmale nur benutzt werden, wenn keine *top-down*-Informationen vorhanden sind, also der Sinn einer Äußerung nicht bekannt ist.

Darüber hinaus hat Blesser (1969) gezeigt, dass es keine Korrelation zwischen den Ergebnissen für die Verständlichkeit von Phonemen und der von Sätzen gibt. Die Verständlichkeit des Satzes ergibt sich vielmehr aus seiner Semantik, seiner Syntax, seiner Prosodie und dem Kontext, in dem er dargeboten wird. Aus diesen Gründen kann nicht davon ausgegangen werden, dass Sprache in einem einfachen hierarchischen Modell verarbeitet wird, sondern dass eine parallele Verarbeitung mit vielen *top-down*-Prozessen stattfindet.

3.3.3 Was bestimmt die Diskrimination?

Rotierte Sprache wird von Versuchspersonen, die mit ihr noch nicht vertraut sind, nicht als Sprache wahrgenommen. Trotzdem sind sie in der Lage, rotierte Vokale konsistent Vokalkategorien zuzuordnen (siehe Abschnitt 3.3.1) und sie auf diese Weise in phonologische Kategorien zu unterscheiden. Andererseits können sie natürlich auch akustische Unterschiede zwischen den rotierten Vokalen unterscheiden. Es stellt sich daher die Frage, welches Kriterium die Versuchspersonen zur Diskrimination von rotierter Sprache heranziehen.

Azadpour und Balaban (2008) haben diese Frage untersucht, indem sie 20 Versuchspersonen, welche vorher noch nie rotierte Sprache gehört hatten, eine Diskriminationsaufgabe haben durchführen lassen. Dabei wurden den Versuchspersonen drei Stimuli vorgespielt, woraufhin sie entscheiden sollten, ob der dritte identisch war mit dem ersten oder dem zweiten. In einem weiteren Versuch haben sie 10 andere Personen rotierte Silben mit entsprechenden Buchstaben versehen lassen. Aus der Prozentzahl der jeweiligen Identifikationen haben die Autoren phonologische Abstände zwischen den rotierten Vokalen errechnet. Ebenso wurde aus der Differenz der jeweiligen ersten drei Formanten ein akustischer Abstand berechnet. Eine Korrelationsanalyse hat anschließend gezeigt, dass zur Erklärung der Diskriminationsleistung bei rotierter Sprache weder der phonologische noch der akustische Abstand alleine ausreichend ist. Daher ist davon auszugehen, dass die Versuchspersonen beide Kategorien zur Diskrimination benutzt haben.

Azadpour und Balaban (2008) hatten in ihrer Studie eine Rotationsfrequenz F_R von lediglich 1500 Hz gewählt, damit die rotierten Stimuli möglichst sprachfern klingen. Trotzdem haben die Versuchspersonen unbewusst phonologische Kategorien benutzt, um die Stimuli unterscheiden zu können. Daraus lässt sich schließen,

dass auch rotierte Sprache schon Teile der Sprachverarbeitung im Kortex in Anspruch nimmt (siehe auch Diskussion zu den Ergebnissen der fMRT-Messung in Kapitel 5.4).

Nach diesen allgemeineren Betrachtungen zu der Verständlichkeit von rotierter Sprache soll im nächsten Kapitel die Verständlichkeit der rotierten Sprache als Träger des *Pop-Out-Effektes* quantifiziert werden.

4 Psychoakustik des Pop-Out-Effektes

4.1 Einführung

Wie in Kapitel 2.4 bereits vorgestellt beschreibt der *Pop-Out-Effekt* die Tatsache, dass durch ein entsprechendes Kontextwissen auch stark verzerrte Sprache verstanden werden kann. Ihr Verständnis springt durch dieses Wissen geradezu aus dem normalerweise nicht zu verstehenden Signal hervor und lässt auch erst wieder nach einer bestimmten Zeitspanne nach. In einem klassischen Experiment zum *Pop-Out-Effekt* wird erst die verzerrte Sprache (VS1) dargeboten, dann die klar verständliche Form (S) und anschließend wieder die verzerrte Sprache (VS2).¹ Davis und Johnsrude (2007) gehen davon aus, dass die Leistung des (kontextabhängigen) Verstehens bzw. des Lernens des Inhalts von VS2 durch das akustische Kurzzeitgedächtnis erbracht wird. Es ist bisher aber unklar, wie die Verständlichkeit von VS2 von ihrem zeitlichen Abstand zu der klar verständlichen Version derselben Äußerung abhängt. Dies soll in dieser Studie näher untersucht werden.

Gleichzeitig stellt sich die Frage, wie der generelle Lernprozess der rotierten Sprache abläuft und ob der *Pop-Out-Effekt* einen Einfluss auf diesen hat. Davis et al. (2005) haben den *Pop-Out-Effekt* für rauschangeregte Sprache untersucht und dabei einen großen Lerneffekt festgestellt. Ihre Versuchspersonen waren in der Lage, nach der Darbietung von 30 rauschangerechten Sätzen 60% der Wörter zu verstehen, während es bei den ersten beiden Sätzen nur etwa 20% der Wörter waren. Zudem wurde die Trainingsleistung stark durch den *Pop-Out-Effekt* beeinflusst. Wurde nach jedem der 30 Sätze einmal seine unverzerrte Version (S) erneut gefolgt von der verzerrten Version (VS2) vorgespielt, kam es zu dem *Pop-Out-Effekt*. Zusätzlich steigerte dies den Trainingseffekt und die Versuchspersonen erkannten nach 30 Sätzen über 75% der Wörter von VS1. Ein weiteres Ergebnis ihrer Studie besagt, dass die Verständnisseleistungen für VS1 stark von den entsprechenden Sätzen abzuhängen schienen, da es zu konsistenten Antworten der Versuchspersonen für die einzelnen Sätze kam. Diese wichen zum Teil stark von

¹Ein Beispiel hierfür ist unter <http://medi.uni-oldenburg.de/download/diplom/hw08/> zu finden. Hierbei handelt es sich um eine Verzerrung durch Sinussprache (Remez et al., 1981), bei der die ersten drei Formanten durch Sinustöne nachgebildet werden.

dem Mittelwert über alle Sätze für die Verständlichkeit von RS1 ab.

Mit dem psychoakustischen Experiment sollen daher folgende drei Fragestellungen zum *Pop-Out-Effekt* beantwortet werden:

- 1) Wie sehen die Ergebnisse mit rotierter Sprache als Träger des Effektes aus?
- 2) Wie hängt der Effekt vom Abstand der Stimuli zueinander ab?
- 3) Gibt es Lerneffekte beim Verständnis von VS1 durch diesen Effekt?

4.2 Versuchsablauf

Stimuli

Bei der vorliegenden Studie wurde eine erweiterte Fassung des Göttinger Satztestes (GÖSA, Kollmeier und Wesselkamp, 1997) als Sprachmaterial verwendet. Diese enthält 230 Sätze, die 1,2-2,9 s lang sind und jeweils aus drei bis sieben Wörtern bestehen. Dieser Satztest ist auf eine gleiche Schwierigkeit des Sprachverstehens im Rauschen konzipiert worden und wird in der Sprachaudiometrie verwendet.

Die einzelnen Sätze wurden bei 5,7 kHz tiefpassgefiltert. Die nach diesem Schritt entstandenen Sätze werden im Folgenden mit S bezeichnet und stellen die unverzerrte Version der Sprache dar. Aus diesen Sätzen wurde jeweils eine rotierte Version erstellt, die im Folgenden mit RS (bzw. RS1 und RS2) bezeichnet werden soll. Dazu wurden die Sätze S hochpassgefiltert und mit einem Sinus bei einer Frequenz von 6 kHz multipliziert. Anschließend wurden sie erneut bei 5,7 kHz tiefpassgefiltert und als rotierte Version RS abgespeichert. In einem letzten Schritt wurden die RMS-Werte von S und RS angepasst (siehe Abschnitt 3.2 für eine genaue Beschreibung der einzelnen Schritte). In Abbildung 3.1 sind die beiden Spektrogramme für S und RS eines ausgewählten Satzes zu sehen. In Abbildung 3.4 finden sich darüber hinaus die über alle Sätze gemittelten Langzeitspektren.

Darbietung

Das klassische *Pop-Out-Effekt*-Experiment besteht aus der Darbietungsreihe verzerrte Sprache (VS1), Sprache (S), verzerrte Sprache (VS2), wobei es sich bei VS1 und VS2 jeweils um identische Stimuli handelt, die aus S erzeugt worden sind. Da es sich bei den verzerrten Stimuli um rotierte Sprache handelt, wird dies im Folgenden immer mit RS1, S, RS2 abgekürzt werden. Um die Abhängigkeit von der Zeit bzw. Störeinflüssen zwischen den Darbietungen zu messen, sind zwischen den einzelnen Darbietungen von RS1, S und RS2 eines Satzes andere Sätze dargeboten worden. Die genaue Vorgehensweise ist in Tabelle 4.1 skizziert.

Auf RS1 folgten 0-30 andere Sätze, dann S und nach 1, 2, 3, 5 oder 8 weiteren Sätzen RS2. Die anderen Sätzen wurden zufällig ausgewählt und es konnte sich jeweils um die Versionen RS1, S oder RS2 der anderen Sätze handeln. Die

...	S ₂₆	RS2 ₂₈	S ₂₃	RS1 ₂₄	S ₂₄	RS2 ₂₃	RS1 ₂₅	RS2 ₂₄	S ₂₁	RS2 ₂₆	RS2 ₂₁	...
...	S ₂₆	RS2 ₂₈	S ₂₃	RS1 ₂₄	S ₂₄	RS2 ₂₃	RS1 ₂₅	RS2 ₂₄	S ₂₁	RS2 ₂₆	RS2 ₂₁	...
...	S ₂₆	RS2 ₂₈	S ₂₃	RS1 ₂₄	S ₂₄	RS2 ₂₃	RS1 ₂₅	RS2 ₂₄	S ₂₁	RS2 ₂₆	RS2 ₂₁	...
...	S ₂₆	RS2 ₂₈	S ₂₃	RS1 ₂₄	S ₂₄	RS2 ₂₃	RS1 ₂₅	RS2 ₂₄	S ₂₁	RS2 ₂₆	RS2 ₂₁	...

Tabelle 4.1: Skizzierung des Messablaufs. Der Messablauf ist hier künstlich in vier Wiederholungen aufgespalten worden, um die Abstände besser darstellen zu können. In der Messung selbst wird das Ganze nur einmal durchlaufen. S steht für unverzerrte Sätze, RS1 für rotierte Sätze mit Darbietung vor S und RS2 für rotierte Sätze mit Darbietung nach dem entsprechendem S. Ein gleicher Index steht für den gleichen Satz. Es soll hier die Messung der Verständlichkeit von RS2 in Abhängigkeit vom Abstand zum dazugehörigem S veranschaulicht werden. In Rot sind dazu die zum jeweils gemessenen RS2 gehörigen gleichen Sätze markiert und mit Blau alle Sätze, die zwischem diesem RS2 und dem dazugehörigem S liegen. Der Abstand zwischen RS2 und S wird dann in Anzahl der dazwischen liegenden Stimuli angegeben. So ist der Abstand für RS2₂₃ zwei, der für RS2₂₄ ebenfalls zwei, der für RS2₂₆ acht und der für RS2₂₁ eins.

Messung der einzelnen *Pop-Out-Effekte* wurde also verzahnt durchgeführt.² Dabei wurde nach jedem dargebotenen Satz seine Verständlichkeit überprüft, indem die Versuchspersonen nach jedem Satz die Wörter wiederholen mussten, die sie meinten verstanden zu haben. Erst daraufhin wurde der nächste Satz dargeboten.

Eine Messung dauerte ungefähr 30 Minuten und es wurden darin alle fünf verschiedenen Abstände 18 Mal dargeboten, was einer Gesamtanzahl von 270-310 präsentierten Sätzen pro Messung entspricht.³

Insgesamt musste jede Versuchsperson zwei Messungen an unterschiedlichen Tagen durchführen, wobei vor der ersten Messung ein kurzes Training stattfand. Das Training glich in seinem Ablauf der Messung, nur dass als Abstände 0, 1, 2 und 3 gewählt worden sind. Hierbei heißt ein Abstand von 0, dass RS2 direkt auf S folgte. Als Trainingsmaterial wurden zufällig 30 Sätze aus dem Oldenburger Satztest (Wagener et al., 1999) gewählt.

Durchführung

Die Sätze wurden mit einer Abtastrate von 44100 Hz erzeugt und über D/A-Wandler und Kopfhörerverstärker diotisch über einen Kopfhörer HDA200 in ei-

²Dies wurde durch die für diese Messung entwickelte Software gewährleistet. Diese ist auf Anfrage beim Autor erhältlich.

³Durch die komplizierte Verzahnung der einzelnen Abstände, die zudem zufällig erfolgen sollte, war es vereinzelt nötig nicht RS2 darzubieten, wodurch die Gesamtzahl der dargebotenen Sätze variieren konnte.

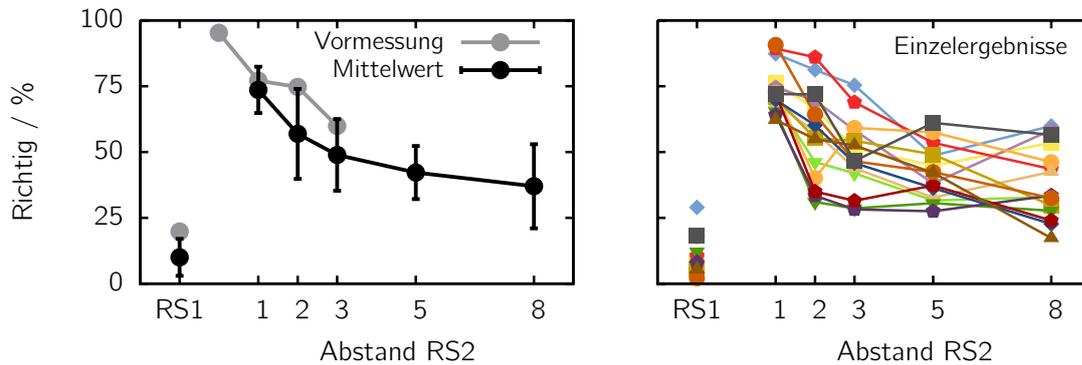


Abbildung 4.1: Ergebnisse für die Verständlichkeit der rotierten Sprache in Abhängigkeit von ihrem Abstand zu S. RS1 steht für die rotierte Sprache, die vor S dargeboten worden ist und RS2 für die rotierte Sprache, die nach S dargeboten worden ist. Mit dem Abstand ist die Anzahl an anderen Stimuli bezeichnet, welche zwischen S und RS2 dargeboten worden sind. Im linken Graphen ist in Grau der Mittelwert über eine Vormessung an drei Versuchspersonen eingetragen und in schwarz der Mittelwert mit Standardabweichung über 15 Versuchspersonen der Hauptmessung. Im rechten Graphen sind die Einzelergebnisse der 15 Versuchspersonen der Hauptmessung eingetragen.

ner doppelwandigen Hörkabine dargeboten.⁴ Die Lautstärke der Sätze betrug ungefähr 65 dB SPL. An der Messung nahmen 15 normalhörende Versuchspersonen mit einem mittlerem Alter von 25 ± 3.2 Jahren teil. Ihnen wurde eine Aufwandsentschädigung für ihre Teilnahme gezahlt.

4.3 Ergebnisse

In Abbildung 4.1 ist im linken Teil als schwarze Linie der Mittelwert mit Standardabweichung über alle 15 Versuchspersonen eingetragen. Der Prozentwert für die Verständlichkeit bezieht sich dabei auf die Verständlichkeit einzelner Wörter. Das heißt wenn zwei von zehn Wörtern verstanden worden sind, entspricht dies einer Verständlichkeit von 20%. Mit RS1 ist das Ergebnis für die Verständlichkeit der rotierten Sprache vor der unverzerrten Version des gleichen Satzes angegeben. Sein Wert beträgt 10%. Daneben sind die Ergebnisse für RS2, d.h. der rotierten Sprache nach der verständlichen Version, in Abhängigkeit ihres Abstandes zu dieser verständlichen Version aufgetragen. Für einen Abstand von 1, welcher einem anderem Stimulus zwischen S und RS2 entspricht, liegt die Verständlichkeit bei ungefähr 74%. Diese fällt zu höheren Abständen zunächst stärker, später schwächer

⁴RME Analog/Digital Interface ADI-8 DS; Tucker-Davis Technologies HB7 Headphone Driver; Sennheiser HDA200.

ab. Die Verständlichkeit passiert bei einem Abstand von ungefähr 3 die 50%-Marke und liegt bei einem Abstand von 8 noch bei 37%.

In Grau sind die Ergebnisse für eine von zwei Vormessungen eingetragen. An dieser Vormessung nahmen drei Versuchspersonen teil, die sich alle nicht mehr an der Hauptmessung beteiligten. Bei dieser hier eingetragenen Vormessung wurden etwas andere Abstände als im Hauptexperiment untersucht, nämlich 0 (RS2 direkt nach S), 1, 2 und 3. Interessant ist hier das Ergebnis für die Verständlichkeit des später nicht mehr untersuchten Abstandes von 0. Dieses liegt bei ungefähr 94%. Da die Ergebnisse für einen Abstand von 3 in der Vormessung noch über 50% lagen, sollten in der Hauptmessung auch höhere Abstände untersucht werden, wodurch aber auf Grund von in der Anzahl begrenztem Sprachmaterial der Abstand 0 nicht mehr gemessen worden ist.

In einer zweiten Vormessung ist für drei Probanden die Verständlichkeit für rotierte Sprache mit einer Rotationsfrequenz von 4 kHz bestimmt worden. Dabei zeigte sich, dass die Verständlichkeit für RS1 mit 29% deutlich höher lag, als dies für eine Rotationsfrequenz von 3 kHz der Fall wäre. Da für das Experiment im Magnetresonanztomograph eine möglichst geringe Verständlichkeit von RS1 wichtig ist, wurde für das Hauptexperiment die niedrigere Rotationsfrequenz gewählt.

Im rechten Teil von Abbildung 4.1 sind die Einzelergebnisse aus der Hauptmessung für alle 15 Versuchspersonen dargestellt. Dabei handelt es sich jeweils um den Mittelwert über ihre beiden Einzelmessungen. Es ist zu erkennen, dass eine breite Streuung für die verschiedenen Abstände existiert. Diese reicht von 28,5 Prozentpunkten für den Abstand 1 bis hin zu 55 Prozentpunkten für den Abstand 2. Dabei war es häufig so, dass Versuchspersonen, die bei einem Abstand relativ hohe oder niedrige Werte für die Verständlichkeit erzielten, dies auch bei den anderen Abständen wiederholten. Ebenfalls eine große Streuung ergab sich für die Verständlichkeit von RS1. Dort reichten die Ergebnisse von 2% bis hin zu 29%.

Lerneffekte

Davis et al. (2005) haben einen großen Lerneffekt für rauschangeregte Sprache entdeckt, daher soll auch hier untersucht werden, ob sich die Verständlichkeit von RS1 bei den Versuchspersonen mit der Zeit ändert. In Abbildung 4.2 ist dazu die Anzahl der verstandenen Wörter aus RS1 in Abhängigkeit ihrer Darbietungsposition in der Gesamtmessung dargestellt. Insgesamt sind fast 200 Sätze mit jeweils 3-7 Wörtern dargeboten worden. In Hellgrau ist die Anzahl an insgesamt dargebotenen Wörtern aufgetragen, in Dunkelgrau die Anzahl der daraus verstandenen Wörter. Die blaue Linie gibt die Anzahl der verstandenen Wörter in Bezug auf die Anzahl der dargebotenen Wörter in Prozent an. Die kleine Lücke nach 99 Darbietungen markiert das Ende der ersten Messung und den Beginn der zweiten. In Rot ist ebenso wie in Blau die Anzahl der verstandenen Wörter in Prozent dargestellt, allerdings in diesem Fall nur für die zweite Messung.

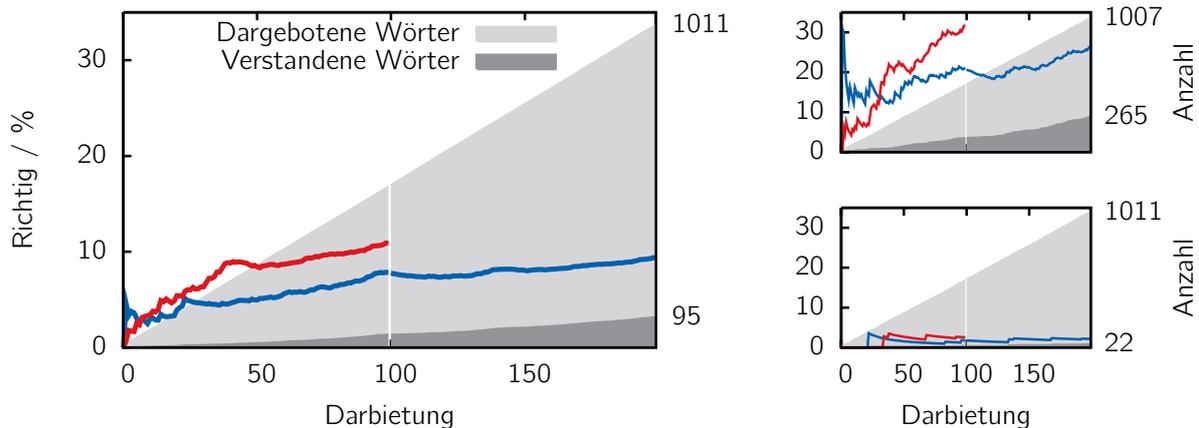


Abbildung 4.2: Lernkurven für das Verstehen von rotierter Sprache ohne Kontextwissen. Es ist sowohl der Mittelwert über alle 15 Versuchspersonen (großes Teilbild), sowie die Versuchsperson mit dem besten (oben rechts) bzw. dem schlechtesten Ergebnis (unten rechts) dargestellt. In hellem Grau ist die Anzahl der dargebotenen Wörter aufgetragen, deren maximaler Wert auf der rechten Ordinate notiert ist. In dunklem Grau ist die Anzahl der davon verstandenen Wörter aufgetragen. Deren maximaler Wert ist auch auf der rechten Ordinate notiert. In Blau ist jeweils die Anzahl verstandener Wörter im Verhältnis zur Anzahl der dargebotenen Wörter in Prozent aufgetragen. Die Prozentwerte sind auf der linken Ordinate notiert. Es sind die Ergebnisse für die erste (bis 100 Darbietungen) und die zweite Messung (ab 100 Darbietungen) gemeinsam aufgetragen. In Rot sind die verstandenen Wörter in Prozent für die zweite Messung alleine dargestellt. Zum besseren Vergleich beginnt die rote Linie wie die erste Messung bei 0.

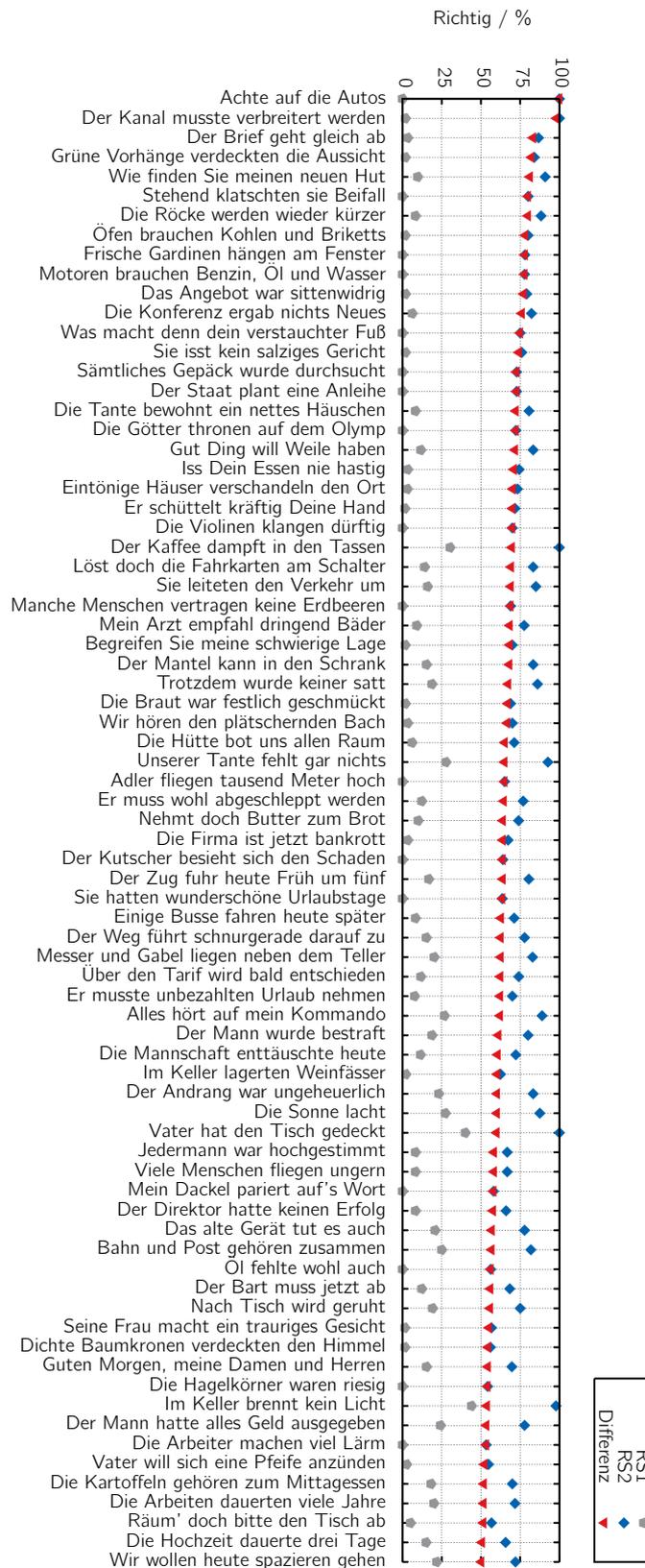
Der große Graph stellt den Mittelwert über alle 15 Versuchspersonen dar. Nach 198 Darbietungen sind insgesamt 1011 Wörter vorgespielt worden, von denen 95 Wörter verstanden worden sind. Dies entspricht einer Verständlichkeit von 9,4%. Am Ende der ersten Messung lag dieser Wert noch bei 7,8%, womit eine leichte Verbesserung zu beobachten ist. Auffällig ist zudem das leichte Einknicken der blauen Kurve zu Beginn der zweiten Messung. Dies weist darauf hin, dass die Versuchspersonen bei der zweiten Messung von neuem mit dem Lernen beginnen mussten. Zu beobachten ist diese Entwicklung auch deutlich an der roten Kurve, welche die Ergebnisse der zweiten Messung darstellt. Sie beginnt am Anfang noch unterhalb der blauen Kurve der ersten Messung, steigt dann jedoch deutlich schneller an und endet bei 10,9%. In den kleineren Graphen auf der rechten Seite in Abbildung 4.2 sind die beiden Messergebnisse für die Versuchsperson mit den besten Ergebnissen für die Verständlichkeit und der Versuchsperson mit den schlechtesten Ergebnissen für die Verständlichkeit dargestellt. Bei der Versuchsperson mit dem schlechtesten Ergebnis ist es insgesamt schwer, von einem Lerneff-

fekt zu reden. Sie hat am Ende der Messung insgesamt erst 22 von 1011 Wörtern (2,2%) verstanden. Bei der Versuchsperson mit dem besten Ergebnis zeigt sich hingegen ein ähnlicher Verlauf wie beim Mittelwert. Nach der ersten Messung lag die Verständlichkeit der rotierten Sprache bei 21%, nach der zweiten Messung bei 26,3%. Wird die zweite Messung unabhängig von der ersten betrachtet, ist eine deutlich stärkere Steigung zu sehen und die Verständlichkeit am Ende der Messung beträgt 31,6%. Insgesamt gab es noch zwei weitere Versuchspersonen, die eine ähnlich gute Verständlichkeit von RS1 aufwiesen. Alle anderen Versuchspersonen lagen hingegen mit ihren Ergebnissen unterhalb des Mittelwertes.

Satzabhängigkeit

Die Ergebnisse für die Verständlichkeit der rotierten Sprache können außer in Abhängigkeit von ihrem Abstand zu S auch noch in Abhängigkeit vom Inhalt des dargebotenen Satzes betrachtet werden. Dazu sind in den Abbildungen 4.3 - 4.5 alle dargebotenen Sätze und ihre über alle Versuchspersonen gemittelte Verständlichkeit in Prozent aufgetragen. In Blau ist die über alle Abstände zwischen 1 und 8 gemittelte Verständlichkeit von RS2 und in Grau die von RS1 dargestellt. Die Differenz zwischen diesen beiden, RS2 – RS1, ist in Rot eingetragen. Nach diesen Werten für die Differenz sind die Sätze absteigend sortiert worden. Mit „Achte auf die Autos“ ist genau ein Satz vorhanden, für den in der Situation RS1 von allen Versuchspersonen kein einziges Wort und in der Situation RS2 alle verstanden worden sind (und dies unabhängig vom Abstand zu S). Ebenso gibt es noch eine Reihe weiterer Sätze, die in der Situation RS2 nahezu immer verstanden worden sind. Allerdings sind diese auch in der Situation RS1 oft zu 25% oder mehr verstanden worden. Zu dieser Gruppe gehört zum Beispiel der Satz „Vater hat den Tisch gedeckt“. Zwei Sätze sind bereits in der RS1 Situation zu über 50% verstanden worden. Dabei handelt es sich um die Sätze „Hast du den Kuckuck gehört“ und „Das Vergnügen findet morgen statt“. Am unteren Ende des Graphen sind Sätze zu finden, die in keiner der beiden Situationen zu verstehen waren, wie z.B. „Schnupfen stört uns natürlich sehr“. Ebenfalls am unteren Ende des Graphen befinden sich Sätze, die in der Situation RS2 nicht verstanden worden sind, Teile davon jedoch in der Situation RS1. Interessant war die Beobachtung, dass die Versuchspersonen sich in der kontextfreien Situation RS1 bisweilen selbst einen Kontext gesucht haben. So haben mehrere Versuchspersonen auf den Satz „Ist denn noch nicht Feierabend“ in der Situation RS1 mit dem aus der Werbung bekannten Satz „Ich bin doch nicht blöd“ geantwortet. Ähnliche Fälle ergaben sich zuweilen aus den zuvor dargebotenen Sätzen. Wenn die Versuchspersonen auf den Satz in der Situation RS1 auf solch eine Weise falsch antworteten, dann haben sie auch verstärkt in der Situation RS2 mit genau demselben falschen Satz geantwortet, obwohl sie den richtigen unverzerrten Satz inzwischen gehört hatten.

Abbildung 4.3: Verständlichkeit von RS1 (grau) und RS2 (blau) in Abhängigkeit des dargebotenen Satzes. In Rot ist die Differenz RS2 – RS1 zwischen den beiden eingetragenen. Nach dieser sind die Sätze absteigend sortiert.



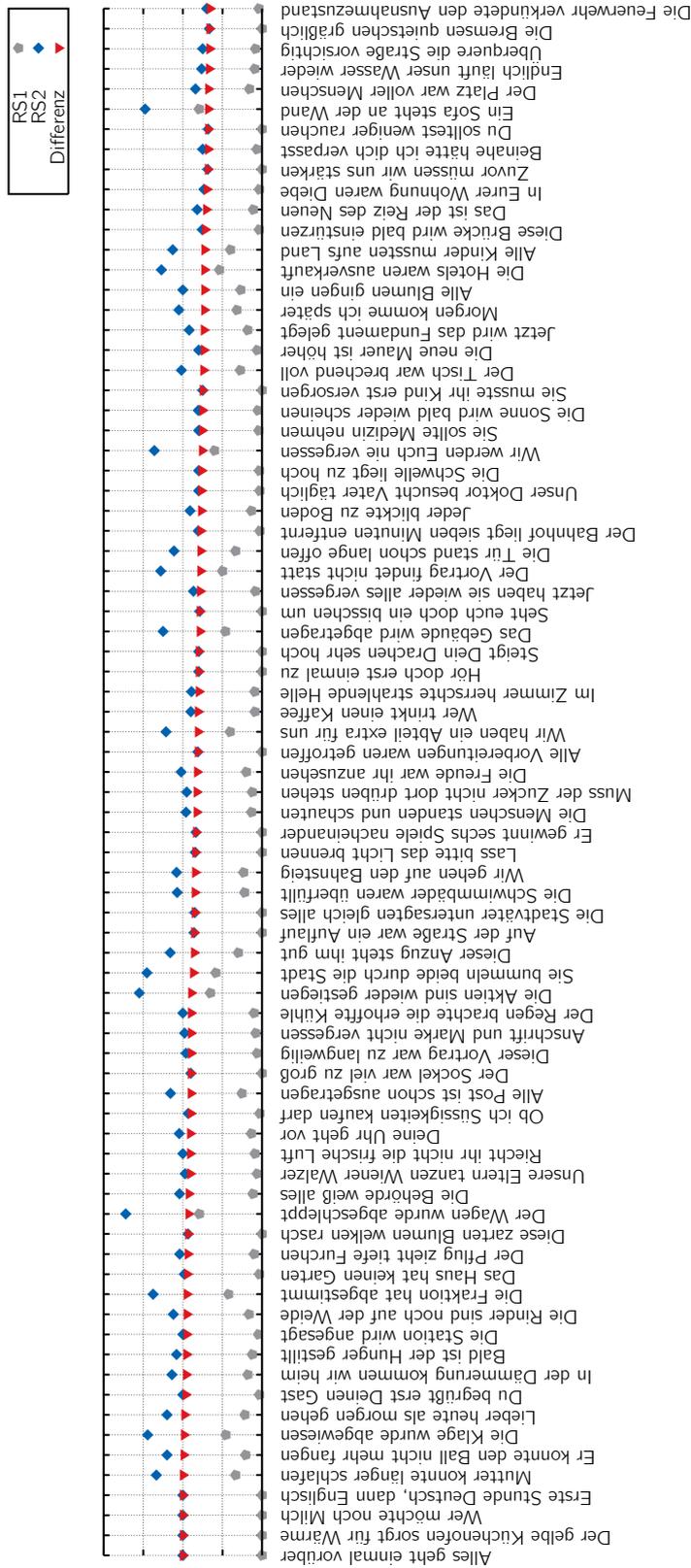
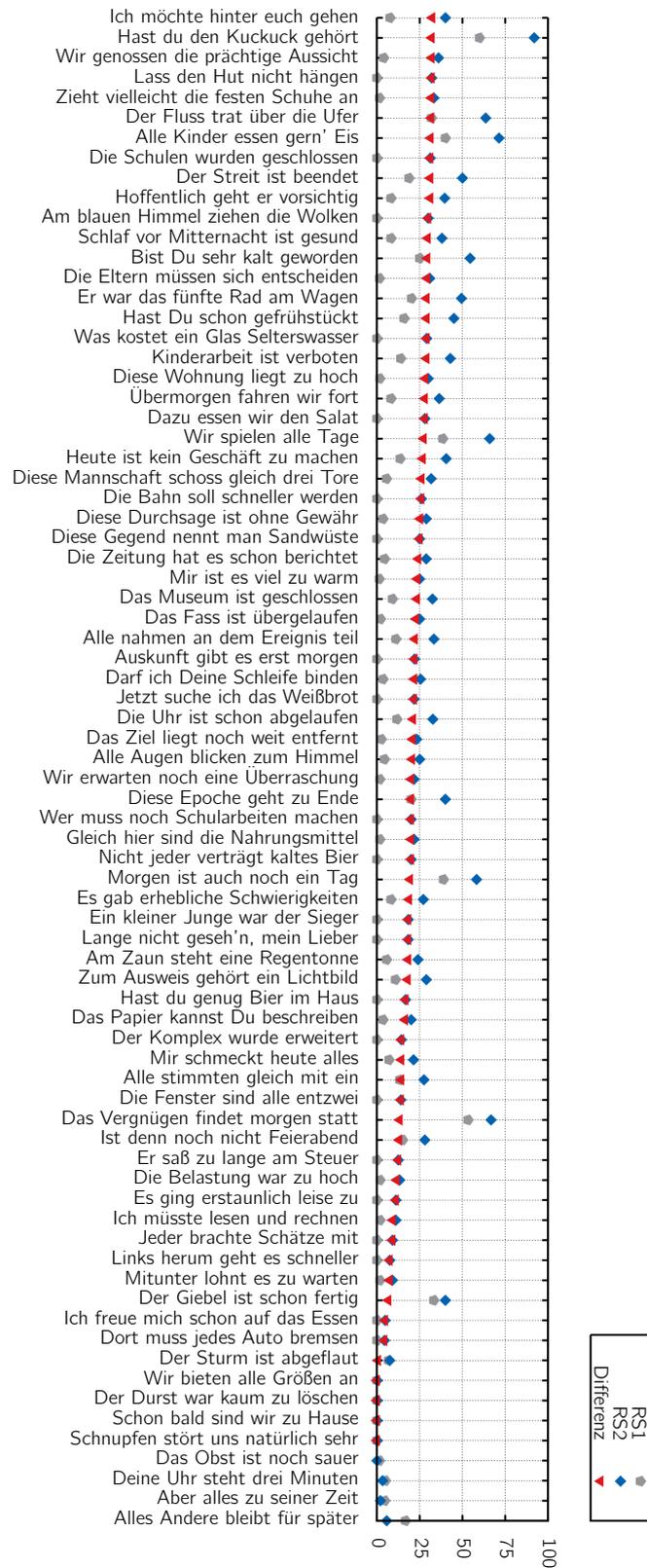


Abbildung 4.4: Verständlichkeit von RS1 (grau) und RS2 (blau) in Abhängigkeit des dargebotenen Satzes. In Rot ist die Differenz RS2 – RS1 zwischen den beiden eingetragen. Nach dieser sind die Sätze absteigend sortiert.

Abbildung 4.5: Verständlichkeit von RS1 (grau) und RS2 (blau) in Abhängigkeit des dargebotenen Satzes. In Rot ist die Differenz RS2 – RS1 zwischen den beiden eingetragen. Nach dieser sind die Sätze absteigend sortiert.



4.4 Diskussion

Ziel war es, mit Hilfe des *Pop-Out-Effektes* Stimuli zu finden, die in der Situation VS1 nicht verständlich, in der Situation VS2 jedoch sehr wohl verständlich sind. Die Ergebnisse zu der Messung mit rotierter Sprache als verzerrte Sprache zeigen, dass diese Stimuli die Anforderungen erfüllen. In der Situation RS1 ist die rotierte Sprache nur zu ungefähr 10% verständlich, in der Situation RS2 hingegen zu ungefähr 95%.

Zudem ist die zeitliche Abhängigkeit des *Pop-Out-Effektes* durch die zwischen S und RS2 liegende Darbietung anderer Stimuli untersucht worden. Auf diese Weise ist unter anderem ein Punkt mit einer Sprachverständlichkeit von ungefähr 50% für die rotierte Sprache ermittelt worden. Um den Abfall der Verständlichkeit durch die Darbietung anderer Stimuli zu erklären, erscheint es zunächst sinnvoll, die Darbietungsdauer der anderen Stimuli genau zu bestimmen. Die Zeit, die zwischen der Darbietung von S und RS2 vergangen ist, hängt zum einen natürlich von der Anzahl der dazwischen dargebotenen Stimuli ab, zum anderen aber auch von der Antwortgeschwindigkeit der einzelnen Versuchspersonen nach jedem Satz. Die Bestimmung der Zeit für die jeweiligen Messungen haben ergeben, dass für eine Darbietung von einem Satz mit nachfolgender Antwort im Durchschnitt 7,3 s vergehen. Das heißt bei einem Abstand von 8 (was 8 anderen Sätzen zwischen S und RS2 entspricht) vergeht zwischen der Darbietung von S und der von RS2 fast eine Minute.

Werden diese Werte mit der Literatur zum auditorischen Gedächtnis verglichen, ergibt sich folgendes Bild: Für das Abklingen des auditorischen Kurzzeitgedächtnisses ergeben sich Zeiten zwischen 2-20 s (Baddeley, 1997). Die Sprachverständlichkeit für RS2 zeigte eine größere Steigung in dem Bereich mit einem Abstand von 0-3, wobei ein Abstand von 3 ungefähr 22 s entspricht. Das heißt es ist wahrscheinlich, dass es in diesem Bereich zu deutlich besseren Ergebnissen kommt als für größere Abstände, da die Probanden noch in der Lage waren, die Sätze in ihrem auditorischen Kurzzeitgedächtnis zu speichern. Ab einem Abstand von 3 findet ein deutlich geringerer Abfall der erzielten Sprachverständlichkeit statt, was für eine Benutzung des auditorischen Langzeitgedächtnisses für diesen Bereich spricht. Es hat sich auch durch vereinzelte Beobachtungen gezeigt, dass die Sprachverständlichkeit für einen erst nach einer Woche wieder dargebotenen Satz in der Kondition RS2 deutlich besser ist als für einen vorher noch nie gehörten (RS1).

Lerneffekte

Im Gegensatz zu den Ergebnissen von Davis et al. (2005) für rauschangeregte Sprache ist für die rotierte Sprache kein großer Lerneffekt erkennbar. Die Verständlichkeit von RS1 verbessert sich zwar mit der Zeit, bleibt aber bei den

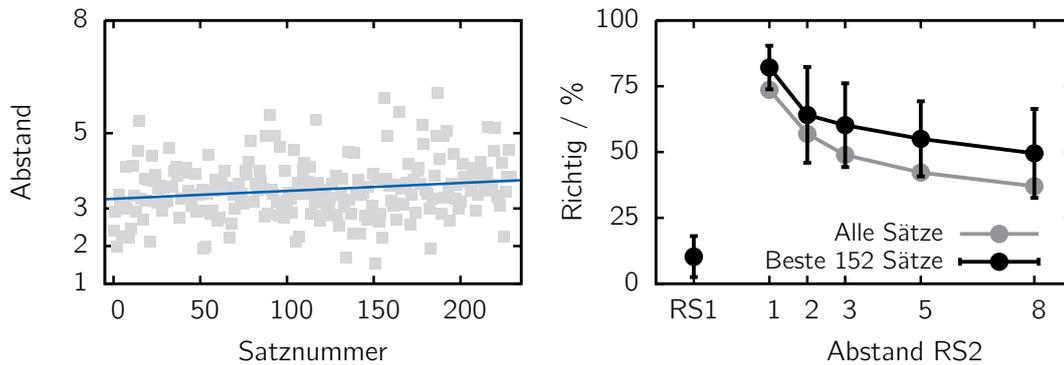


Abbildung 4.6: Links: Die grauen Quadrate markieren den bei dem jeweiligem Satz dargebotenen mittleren Abstand. Die Durchnummerierung der Sätze erfolgte nach der Sortierung ihrer Differenz in der Verständlichkeit zwischen RS2 und RS1. Das heißt, bei einer niedrigen Nummer war diese Differenz besonders groß und umgekehrt. Die blaue Linie stellt eine mit der *least-squares*-Methode an die mittleren Abstände angepasste Gerade dar. Rechts: Verständlichkeit der Sätze in Prozent. Dargestellt ist jeweils der Mittelwert über alle 15 Versuchspersonen, in Grau über alle 230 dargebotenen Sätze, in schwarz über die 152 Sätze mit der größten Differenz in der Verständlichkeit zwischen RS1 und RS2.

meisten Versuchspersonen unterhalb von 10%, während sie für die rauschangeregte Sprache bereits nach 30 dargebotenen Sätzen bei über 60% lag (Davis et al., 2005).

Die Lernkurve für die zweite Messung beginnt am Anfang unterhalb jener der ersten Messung, was jedoch leicht mit dem vor der zweiten Messung fehlendem Training erklärt werden kann. Gleichzeitig spricht dies dafür, dass das einmal gelernte Verständnis der rotierten Sprache nicht unbedingt dauerhaft sein muss. Andererseits wird die Lernkurve aus der ersten Messung durch eine größere Steigung schnell eingeholt und überholt. Dies deutet darauf hin, dass es einen Unterschied macht, ob eine Versuchsperson vorher schon einmal rotierte Sprache gehört hat oder nicht.

Die Versuchspersonen im Magnetresonanztomographen werden 76 mal RS1 als Stimulus vorgespielt bekommen und das gleiche Training erhalten wie die Versuchspersonen, welche an diesem Experiment teilgenommen haben. Daher ist davon auszugehen, dass der bei der Messung im Magnetresonanztomographen zu beobachtende Lerneffekt nicht größer als 4 Prozentpunkte sein wird und somit vernachlässigt werden kann.

Satzabhängigkeit

Davis et al. (2005) haben ebenfalls berichtet, dass die Ergebnisse für das Verständnis

von rauschangeregter Sprache von den einzelnen Sätzen abhinge, allerdings hatten sie es nicht systematisch ausgewertet. Überraschend ist zudem, wie stark die Abhängigkeit ist. Auf der einen Seite gibt es einen Satz, der den Idealfall des *Pop-Out-Effektes* darstellt und in der Situation RS1 komplett unverständlich und in der Situation RS2 komplett verständlich ist. Wobei letzteres erstaunlich ist, da es für jeden Abstand, also auch dem maximalen Abstand von 8, der Fall ist. Auf der anderen Seite gibt es mehrere Sätze, von denen kein Wort verstanden worden ist, bei denen der *Pop-Out-Effekt* folglich überhaupt nicht funktioniert.

Hierfür eine Erklärung zu finden ist schwierig, da es sich bei ganzen Sätzen um zu komplexe Gebilde handelt, um leicht feststellen zu können, welche Verarbeitungsschritte zum Verständnis der einzelnen Sätze notwendig sind. Ebenso lässt sich schwer abschätzen, wie leicht oder schwierig diese einzelnen Schritte sind. Es ist aber anzunehmen, dass die Erklärung für die unterschiedliche Verständlichkeit von RS2 nicht auf zu erwartende Verwechslungen für einzelne Phoneme zurückzuführen sein wird, da schon Blesser (1969) gezeigt hat, dass bei ausreichendem Kontext die Verständlichkeit ausschließlich auf bestimmte semantische und syntaktische Eigenschaften zurückzuführen ist. Die zu Grunde liegende akustische und phonematische Wahrnehmung wird dann dementsprechend verändert, um mit diesen in Übereinstimmung zu gelangen.

Für die Verständlichkeit von RS1 könnten hingegen die Phonemverwechslungen eine Rolle spielen. Hier ist der vorhandene Kontext deutlich geringer als bei RS2. Ebenso könnte der syntaktische Aufbau der Sätze eine Rolle spielen, da die Versuchspersonen oft in der Lage sind, die syntaktische Struktur des Satzes zu erkennen. Eine Eigenschaft, die sie hierfür sicherlich verwenden könnten, stellt die intakte Prosodie der Sätze dar. Diese wird auch eine wichtige Rolle bei dem Verständnis von RS2 spielen. Für RS2 kommt es für die Versuchspersonen darauf an, sich an den richtigen Satz zu erinnern. Dieses Erinnern hängt höchstwahrscheinlich davon ab, wie gut oder schlecht eine Satzmelodie einzuprägen ist.

Eine andere einfache Erklärungsmöglichkeit für den beobachteten Zusammenhang zwischen der Verständlichkeit und der Satznummer stellt eine Abhängigkeit vom Abstand von RS2 zu S dar. Dies wäre der Fall, falls die Messung schlecht ausbalanciert wäre und für „Achte auf die Autos“ fast immer ein Abstand von 1 gemessen worden wäre, während für einen nie verstandenen Satz von RS2 immer ein Abstand von 8 gemessen worden wäre. Um diesen Umstand auszuschließen, wurden für alle Sätze die dargebotenen Abstände überprüft und der mittlere Abstand jedes einzelnen Satzes in Abbildung 4.6 eingetragen. Es ist zu erkennen, dass der mittlere Abstand von dem ersten Satz („Achte auf die Autos“) deutlich größer als 1 ist. Ebenso entspricht der mittlere Abstand des am schlechtesten verstandenen Satzes nicht 8. Zu beobachten ist allerdings ein leichtes Ansteigen des mittleren Abstandes in Richtung der schlechter verstandenen Sätzen hin. Um diese Tatsache genauer zu quantifizieren ist eine Gerade mit Hilfe der *least-squares*-Methode

an die mittleren Abstände angepasst worden. Das Ergebnis liefert einen Anstieg des mittleren Abstandes von 3,3 für den besten Satz hin zu 3,7 für den letzten Satz (beide Angaben mit einem Fehler in der Größenordnung $\pm 0,1$). Bei diesem äußerst geringen Anstieg des Abstandes ist nicht davon auszugehen, dass er für die vorgefundenen großen Unterschiede in der Verständlichkeit in Betracht gezogen werden könnte.

Für die Messung im Magnetresonanztomographen werden insgesamt 152 verschiedene rotierte Sätze benötigt (siehe Abschnitt 5.2). Da hierbei der Unterschied zwischen verständlicher und unverständlicher rotierter Sprache gemessen werden soll, werden für diese Messung nur die 152 Sätze mit der größten Differenz $RS2 - RS1$ benutzt. Dies entspricht den ersten 152 Sätzen aus den Abbildungen 4.3 - 4.5. Durch eine Beschränkung auf die Sätze mit einer möglichst großen Differenz zwischen $RS2$ und $RS1$ ändert sich natürlich auch das Resultat für die Verständlichkeit in Abhängigkeit vom Abstand. Im rechten Graphen von Abbildung 4.6 ist der Mittelwert der Verständlichkeit begrenzt auf die 152 ausgewählten Sätze in schwarz eingetragen. In Grau ist der Mittelwert für alle 230 Sätze zu sehen. Deutlich wird, dass für alle dargebotenen Abstände eine Verbesserung der Verständlichkeit zu beobachten ist. Die Verständlichkeit für einen Abstand von 1 hat sich auf ungefähr 82% verbessert. Nach einem Abstand von 2 fällt die Kurve zudem weniger stark ab und erreicht erst bei einem Abstand von 8 einen Wert von 50% Verständlichkeit. Eine Verbesserung der Verständlichkeit für die Situation $RS1$ wurde jedoch nicht erzielt. Daraus lässt sich folgern, dass die größere Differenz zwischen $RS2$ und $RS1$ bei den 152 Sätzen nur auf Grund einer besseren Verständlichkeit von $RS2$ zu Stande kommt und nicht wegen einer gesunkenen Verständlichkeit von $RS1$.

Nachdem in diesem Versuchsteil das Verhalten der rotierten Sprache in unterschiedlichen Situationen des *Pop-Out-Effektes* untersucht worden ist, wird im nächsten Kapitel dieses Wissen benutzt, um mit Hilfe des *Pop-Out-Effektes* die Sprachverarbeitung im menschlichen Kortex genauer zu untersuchen.

5 Untersuchung der Sprachverarbeitung mit fMRT

5.1 Einführung

Untersucht werden soll die Sprachverarbeitung im Kortex und ihre Lateralisierung mit Hilfe von funktioneller Magnetresonanztomographie (fMRT). Hierbei wird der *Pop-Out-Effekt* verwendet, um die Verarbeitung des Sinngehaltes von Sprache zu lokalisieren. Dies ist möglich durch den Vergleich der Gehirnaktivität von akustisch identischen Stimuli, die sich auf Grund dieses Effektes aber in ihrer Verständlichkeit unterscheiden. Weiter soll der Verarbeitungsunterschied zwischen natürlicher Sprache und verzerrter Sprache untersucht werden. Weicht dieser Kontrast in der Lateralisierung von dem vorherigen ab, ist dies ein Ergebnis, das gegen die Hypothese der Lateralisierung der Sprachverarbeitung aufgrund von akustischen Merkmalen sprechen würde (vgl. Abschnitt 2.3).

5.2 Versuchsablauf

Stimuli

Es sollten akustisch identische Stimuli mit unterschiedlicher Verständlichkeit dargeboten werden. Eingesetzt wurde hierfür der in Kapitel 4 untersuchte *Pop-Out-Effekt* mit durch Rotation des Leistungsspektrums verzerrter Sprache. Dabei wurden die Abstände zwischen den Stimuli so gewählt, dass eine parametrisch veränderte Verständlichkeit der rotierten Sprache von etwa 10%, 50% und 95% erzielt wurde. 95% Verständlichkeit entsprach einem Abstand von 0 Konditionen und 50% einem Abstand von 5 Konditionen der rotierten Sprache (RS2) zur Sprache (S). Unter einem Abstand von x Konditionen ist die Anzahl x an Konditionen bezeichnet, die zwischen der Darbietung zur Erzielung des *Pop-Out-Effektes* von S und RS2 abgespielt werden (siehe auch Abschnitt 4.2). Eine Verständlichkeit von 10% wurde bei RS1 erzielt. Um die Komplexität der Sprachverarbeitung von schwer verständlicher, verzerrter Sprache kontrollieren zu können ist außerdem Sprache im Rauschen als Vergleichsstimulus in das Experiment aufgenommen worden.

Als Konditionen für die funktionelle Messung im Magnetresonanztomographen kamen die neun in Tabelle 5.1 vorgestellten Stimuli zum Einsatz. Für die Konditio-

Abkürzung	Stimulus Beschreibung	Verständlichkeit / %
	Stille	
	Rauschen	
SR	Sprache im Rauschen	50
RS1 ⁽⁰⁾	Rotierte Sprache	10
S ⁽⁰⁾	Sprache	100
RS2 ⁽⁰⁾	Rotierte Sprache (Abstand: 0 Konditionen)	95
RS1 ⁽⁵⁾	Rotierte Sprache	10
S ⁽⁵⁾	Sprache	100
RS2 ⁽⁵⁾	Rotierte Sprache (Abstand: 5 Konditionen)	50

Tabelle 5.1: Stimuluskonditionen für die Messung im Magnetresonanztomographen. Mit RS1 wird stets die rotierte Sprache bezeichnet, die vor der unverzerrten Version desselben Satzes (S) dargeboten wird. Daher ist deren Verständlichkeit auch deutlich geringer. Mit RS2 wird die rotierte Sprache bezeichnet, welche nach der unverzerrten Version dargeboten wird. Diese ist durch den *Pop-Out-Effekt* deutlich besser zu verstehen, wobei diese Verständlichkeit vom Abstand (Zahl in der hochgestellten Klammer) zu S abhängt. Die Werte für die Verständlichkeit sind ungefähre Werte und stammen aus psychoakustischen Messungen in der Hörkabine (siehe Kapitel 4).

nen mit den gleichen hochgestellten Indizes, also z.B. RS1⁽⁵⁾, S⁽⁵⁾ und RS2⁽⁵⁾, sind jeweils dieselben Sätze aus dem Sprachmaterial gewählt worden. Entnommen wurden die Sätze dabei dem Göttinger Satztest (GÖSA, Kollmeier und Wesselskamp, 1997). Die 152 Sätze daraus, die den größten Unterschied in der psychoakustischen Messung zwischen den Situationen RS1 und RS2 aufwiesen, wurden für die rotierte Sprache und die Sprache verwendet (Abschnitt 4.3). Aus den übrigen 78 Sätzen wurden 76 zufällig ausgewählt und für die Sprache im Rauschen verwendet. Da ein *sparse-imaging*-Paradigma¹ (Hall et al., 1999) mit einer TR² von 9 s für die funktionelle Messung gewählt werden sollte, ergab sich daraus eine Zeit für die Stimulusdarbietung von ungefähr 6 s. Die Länge der einzelnen Sätze aus dem Göttinger Satztest beträgt 1,2–2,9 s. Daher wurden pro Stimulus immer zwei Sätze mit einer Pause von 200 ms zwischen diesen verwendet. Das bedeutet, dass aus den ausgewählten 152 Sätzen für die Messung jeder Versuchsperson zufällig 76 verschiedene Satzpaare erzeugt wurden. Ebenso wurden aus den 76 Sätzen für die Kondition „Sprache im Rauschen“ jeweils zufällig 38 Satzpaare erzeugt.

¹Das heißt, die Darbietung der Stimuli und die Aufnahmen durch den Tomographen finden zeitlich getrennt statt, so dass der Lärm des Scanners nicht die Verständlichkeit der Stimuli beeinträchtigt und die Verarbeitung im auditorischen Kortex verfälscht.

²*time of repeat*, der Abstand zwischen dem jeweiligem Anfang von zwei aufeinander folgenden Phasen der Bildaufnahme des Magnetresonanztomographen



Tabelle 5.2: Skizzierung des Messablaufs im Magnetresonanztomographen. SR steht für „Sprache im Rauschen“, RS1 für „Rotierte Sprache“, S für „Sprache“ und RS2 für die verständliche „Rotierte Sprache“, die nach S dargeboten wird. Die unteren Indizes bezeichnen jeweils das verwendete Satzpaar, die oberen den Abstand zwischen S und RS2. Mit den schwarzen Querbalken ist ein Messvorgang durch den Scanner gekennzeichnet. In Rot sind beispielhaft drei zusammenhängende Darbietungen markiert, die einen *Pop-Out-Effekt* ermöglichen.

Das Rauschen für diese Kondition wurde ebenfalls dem GÖSA entnommen. Dieses wurde auch benutzt, um die reine Rauschkondition zu erzeugen. Bei dieser wiederum wurden zwei 2.1 s lange Rauschsignale ebenfalls mit einer Pause von 200 ms aneinander gehängt. Bei der Kondition „Stille“ wird kein Signal abgespielt.

Bei dem Versuchsablauf im Magnetresonanztomographen herrschte mit 9 s ein unterschiedlicher zeitlicher Abstand zwischen den einzelnen Konditionen im Vergleich zu den ungefähr 7,2 s während der psychoakustischen Messung. Daher ist mit 5 Konditionen ein geringerer Abstand zwischen S und RS2 gewählt worden, um eine Verständlichkeit von 50% für RS2 zu erzielen.

Darbietung

Die Darbietungsreihenfolge der neun Konditionen sollte zufällig sein, wobei durch die Anwendung des *Pop-Out-Effektes* mit einem Abstand von 0 Konditionen und einem Abstand von 5 Konditionen bestimmte Vorgaben in der Stimulus-Reihenfolge eingehalten werden mussten. Jede Kondition sollte insgesamt 38 Mal dargeboten werden, was einer Gesamtzahl von 342 Stimuli pro Messung bei einer Versuchsperson entspricht. Die 342 möglichen Positionen für einen Stimulus sind zunächst in 38 Teile mit einer Länge von 9 Positionen aufgeteilt worden. In die 38 Teile ist jeweils zufällig ein unverzerrter Sprachstimulus verteilt worden. Anschließend wurde die dazu gehörige rotierte Sprache desselben Satzes einmal in einem Abstand von 0-4 davor und einmal je nach Kondition³ in einem Abstand von 0 oder 5 dahinter angeordnet. Die übrigen Konditionen wurden zufällig auf die restlichen Positionen verteilt. Die Reihenfolge der Konditionen war für alle Versuchspersonen identisch, allerdings wurden die einzelnen Sätze jeweils randomisiert ausgewählt.

Die Messung wurde in drei Abschnitte von jeweils 114 Stimuli aufgeteilt, so dass die Probanden im Magnetresonanztomographen zwei kleine Pausen während der Messung hatten. Nach jedem Stimulus setzte die 2,7 s lange Messphase des Tomographen ein. Während dieser Zeit haben die Versuchspersonen auf die Frage „Hast Du die Sprache verstanden?“ per Knopfdruck mit „Ja“ oder „Nein“ geantwortet. Diese Frage wurde von ihnen nach jeder Kondition, also auch nach

³0 für RS₂⁽⁰⁾ und 5 für RS₂⁽⁵⁾.

„Rauschen“ und „Stille“, beantwortet. Zwei Ziele sollten hierdurch erreicht werden. Zum einen sollte die Aufmerksamkeit der Versuchspersonen auf einem konstant hohen Niveau gehalten werden. Dieses war besonders bei den schwer oder gar nicht zu verstehenden rotierten Sprachstimuli von Vorteil, da die Probanden während ihrer Darbietung unaufmerksamer in Bezug auf deren Verständlichkeit hätten werden können. Andererseits konnte auf diese Weise das Verständnis der einzelnen Stimuli abgeschätzt werden.

Der Messablauf stellte sich für die Versuchsperson damit wie folgt dar: Zunächst wurde sie mündlich und schriftlich über die Messung aufgeklärt und hat den *Edinburgh Handedness Inventory*-Fragebogen zur Händigkeit (Oldfield, 1971) und eine Einverständniserklärung ausgefüllt. Danach wurde ein Training mit Sätzen aus dem OLSA absolviert, entsprechend dem Vorgehen in der psychoakustischen Messung (siehe Abschnitt 4.2). Anschließend fanden die drei Durchläufe im Magnetresonanztomographen statt, welche jeweils ungefähr 18 Minuten dauerten. Dazwischen gab es eine ein- bis zweiminütige Pause für die Versuchsperson, in der sie in dem Tomographen verweilte. Abschließend wurde in etwa 14 Minuten ein anatomisches Bild von ihrem Gehirn erstellt.

Durchführung

Die Messungen erfolgten im Klinikum Kreyenbrück. Dabei wurden die Stimuli mit einer Abtastrate von 44100 Hz abgespielt. Das Training mit dem OLSA fand vorab im Warteraum für den Magnetresonanztomographen statt. Dazu wurden die Stimuli über einen Laptop mit externer Soundkarte diotisch über einen Kopfhörer dargeboten.⁴ Alle Magnetresonanz-Messungen wurden mit einem Tomographen Siemens Sonata 1,5 Tesla durchgeführt. Um den Lärm des Scanners von den akustischen Stimuli zu trennen, wurde ein *sparse-sampling*-Paradigma (Hall et al., 1999) verwendet. Die Stimuli wurden den Probanden diotisch über einen MR-kompatiblen Kopfhörer (MRconform GmbH Magdeburg) bei einem Pegel von 70 dB SPL dargeboten. Die Bildaufnahme des Magnetresonanztomographen erfolgte mit Hilfe einer EPI-Sequenz mit einer TR von 9 s und einer TE von 63 ms. Dabei wurden 21 Schichten mit einer Voxelgröße von 1,95x1,95x3 mm³ aufgezeichnet.

Versuchspersonen

An dem Versuch nahmen 16 Personen teil, von denen eine von der weiteren Auswertung ausgeschlossen wurde. Der Grund war ein asymmetrischer Kopfhörersitz, welcher deutlich unterschiedliche Aktivierungen in beiden primären auditorischen Kortizes verursachte. Das mittlere Alter der verbleibenden Versuchspersonen betrug 24,4 Jahre, wobei die jüngste Versuchsperson 18 und die älteste 35 Jahre alt

⁴Soundkarte: Roland - Edirol UA-25. Kopfhörer: AKG K271.

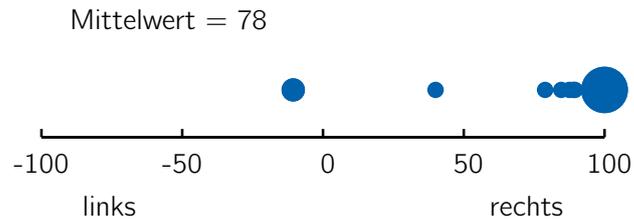


Abbildung 5.1: Ergebnisse des *Edinburgh Handedness Inventory* (Oldfield, 1971) aller Versuchspersonen. Der Flächeninhalt der einzelnen Kreise ist proportional zu der Anzahl an Versuchspersonen mit dem gleichen Ergebnis. Der kleinste Kreis steht für eine Anzahl von einer Versuchsperson. Ein Index von -100 steht für vollständig links-, und ein Index von 100 für vollständig rechtshändig. Die Ergebnisse sind zusätzlich in der Tabelle 5.4 zu finden.

war. Von den 15 Personen waren zehn männlich und fünf weiblich. In Abbildung 5.1 sind die Ergebnisse des *Edinburgh Handedness Inventory* dargestellt. Es ist zu erkennen, dass zwölf Versuchspersonen stark rechtshändig sind. Lediglich zwei Versuchspersonen erreichten einen negativen Index. Dieser war allerdings so gering, dass sie nach diesem Test als beidhändig einzustufen sind. Gleiches gilt für die eine Versuchsperson aus dem positiven Wertebereich, die 40 Punkte erreichte.

Datenauswertung

Die Datenauswertung wurde mit SPM5 durchgeführt.⁵ Dabei wurden zunächst einige Vorverarbeitungsschritte ausgeführt. Zunächst ist eine Bewegungskorrektur der Aufnahmen aus dem Scanner vorgenommen worden, um auszuschließen, dass es zu statistischer Signifikanz in den Gehirnaktivitäten durch kleine Bewegungen der Versuchspersonen kommen konnte. Anschließend wurden die Aufnahmen der Gehirne der einzelnen Versuchspersonen auf das Standard-Gehirn⁶ von SPM5 normiert, damit die einzelnen Versuchspersonen miteinander verglichen und Mittelwerte berechnet werden konnten. Zur Verbesserung des Signal-zu-Rausch-Verhältnisses wurden die einzelnen Bilder abschließend mit einem Gaußkernel mit einer Halbwertsbreite von 6 mm geglättet. Die hämodynamische Antwort ist voxelweise mit einem linearen Regressions-Modell ermittelt worden. Nach diesem Schritt wurde dann für die Kontraste zwischen zwei Konditionen in jedem Voxel der beiden Konditionen ein T-Test durchgeführt. Auf diese Weise konnte eine Abweichung der Aktivierung zwischen den beiden Konditionen festgestellt werden.

⁵<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>

⁶Ein gemittelttes Gehirn über 152 Personen vom *Montreal Neurological Institute*. Koordinatenangaben bei den Ergebnissen beziehen sich dann immer auf diesen MNI-Raum.

Dafür musste ein FWE⁷ korrigierter p -Wert von $< 0,05$ erreicht werden. Zusätzlich wurden nur Gebiete aus mindestens 8 zusammenhängenden Voxeln als signifikant angezeigt.

Für die Berechnung der Mittelwerte ist ein *fixed-effects*-Modell berechnet worden. Die Darstellung der Kontraste erfolgt für die Mittelwerte jeweils auf dem normierten anatomischen Bild einer Versuchsperson.

5.3 Ergebnisse

Die Präsentation der Ergebnisse wird sich im Folgenden auf die Kontraste beschränken, die wichtig für die zu untersuchende Fragestellung sind. Weitere Kontraste, wie etwa zwischen „Schall“ und „Stille“, werden in Anhang A.4 präsentiert.

Zu Beginn werden die Mittelwerte über alle 15 Versuchspersonen gezeigt. Dabei wird hauptsächlich auf die Lokalisation der Sprachverarbeitung und ihre Lateralisierung eingegangen. Außerdem werden Unterschiede in der Verarbeitung von rotierter und natürlicher Sprache untersucht. Abschließend wird auf die Einzelergebnisse der Versuchspersonen eingegangen und die Lateralisierung der Sprachverarbeitung in Abhängigkeit der Händigkeit der einzelnen Probanden untersucht.

5.3.1 Mittelwerte

Sprachverständlichkeit

Durch das Abfragen nach jedem Stimulus, ob die Versuchspersonen den Satz verstanden hätten, konnte die Verständlichkeit der einzelnen Stimuli abgeschätzt werden. Für die ersten drei Versuchspersonen ist dieser Wert wegen technischer Probleme leider nicht verfügbar. In Abbildung 5.2 ist daher der Mittelwert über 12 Versuchspersonen dargestellt. Für die Sprache im Rauschen und für die rotierte Sprache ist in Grau zusätzlich der Mittelwert über die 15 anderen Versuchspersonen eingezeichnet, die zuvor an der Messung zur Sprachverständlichkeit für die benutzten Sätze in der Hörkabine teilgenommen hatten. Dieser Mittelwert ist Abbildung 4.6 entnommen worden. Es ist zu erkennen, dass die Stimuli der Konditionen „Stille“ und „Rauschen“ wie zu erwarten nicht verstanden werden konnten, während die beiden Konditionen mit unverzerrter Sprache zu fast 100% verstanden worden sind. Bei den Abschätzungen zur unverständlichen rotierten Sprache RS1 gab es keine Abweichungen zur Messung aus der Hörkabine. Die Sprachverständlichkeit betrug hier jeweils etwa 10%. Ebenfalls keine Abweichungen ergaben sich für RS2⁽⁰⁾. Hier betrug die geschätzte Verständlichkeit ungefähr 95%. Unterschiede waren lediglich für die Konditionen SR und RS2⁽⁵⁾ zu erken-

⁷Family Wise Error.

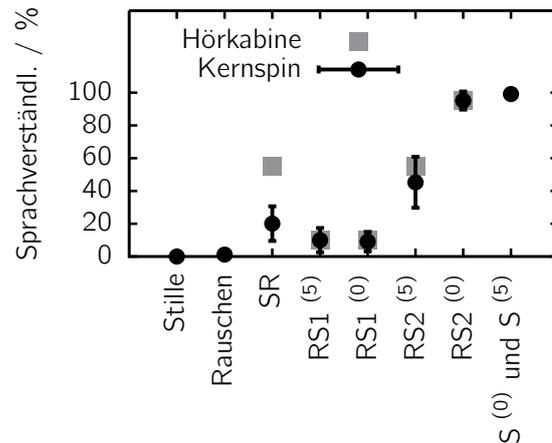


Abbildung 5.2: Verständlichkeit der einzelnen Stimuli. In Grau ist für die Stimuli SR bis RS2⁽⁰⁾ der Mittelwert der Sprachverständlichkeit über alle in der Hörkabine gemessene Versuchspersonen aufgetragen. Die Standardabweichung ist hierbei nicht mit eingezeichnet und ist für die rotierte Sprache in Abbildung 4.6 zu finden. In schwarz ist der Mittelwert mit Standardabweichung der geschätzten Sprachverständlichkeit über 12 Versuchspersonen aufgetragen. Hierzu haben diese Versuchspersonen während der Messung im Magnetresonanztomograph nach jedem Stimulus auf die Frage „Haben Sie die Sprache verstanden?“ mit „Ja“ oder „Nein“ geantwortet.

nen, wobei besonders die geschätzte Sprachverständlichkeit von nur 18,6% für die Sprache im Rauschen stark von den 55% aus der Hörkabine abwich.

Die Ergebnisse der „Ja“-„Nein“-Frage ergeben wie erwähnt nur eine Schätzung der Sprachverständlichkeit. Dabei gilt, dass diese Selbsteinschätzung der Probanden zum Teil deutlich über den Werten einer Identifikationsmessung liegen kann (Klotz und Kumpf, 1995). Unter einer Identifikationsmessung ist eine Messung zu verstehen, in der die Versuchspersonen die verstandenen Wörter wiederholen müssen, wie dies in der Messung in der Hörkabine (siehe Kapitel 4) geschehen ist.

Sprachverarbeitung

Im folgenden Abschnitt werden die Mittelwerte über alle 15 Versuchspersonen gezeigt. Für die Bestimmung der Mittelwerte ist ein *fixed-effects*-Modell gerechnet worden (siehe Abschnitt 5.2). Die Gehirnaktivierungen, die zur Kontrolle des Experimentes dienten und keine Aussagen auf die untersuchte Fragestellung gaben, sind im Anhang A.4 zu finden.

In Abbildung 5.3 ist in Rot die Gehirnaktivität für den Kontrast „Sprache“ minus „unverständliche rotierte Sprache“ ($S - RS1$) dargestellt. Dabei wurden für S die beiden Sprachkonditionen $S^{(0)}$ und $S^{(5)}$ zusammengefasst. Gleiches gilt für $RS1$ in Bezug auf $RS1^{(0)}$ und $RS1^{(5)}$. In Gelb ist der Kontrast zwischen der unverzerrten

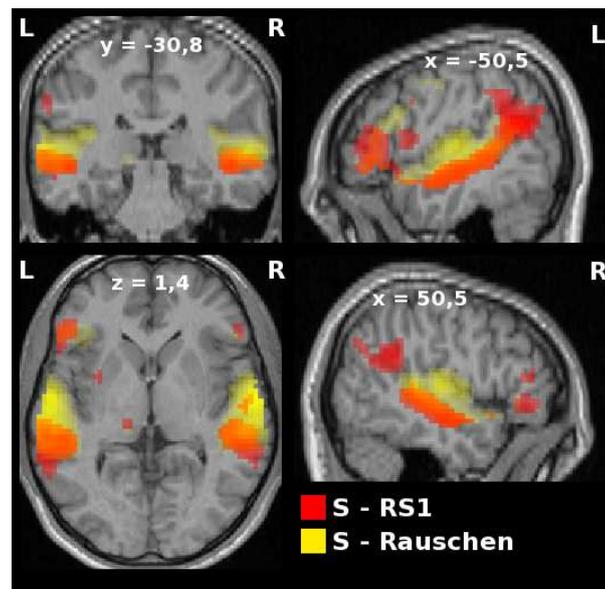


Abbildung 5.3: Rot: Gehirnaktivität für den Kontrast „Sprache“ minus „unverständliche rotierte Sprache“ (S – RS1). Gelb: Gehirnaktivität für den Kontrast „Sprache“ minus „Rauschen“ (S–Rauschen). Bereiche, in denen sich die Aktivität überlagert, sind durch eine Vermischung der Farben angedeutet. L steht für links und R für rechts. Ist nur einer der beiden Buchstaben angegeben, so wird von dieser Seite auf das Gehirn geschaut. Die jeweiligen Koordinatenangaben x, y oder z beziehen sich auf die Lage der jeweiligen Schnittebene im MNI-Koordinatenraum.

Sprache (S) und Rauschen dargestellt.

Es ist zu erkennen, dass es zu bilateraler Aktivität für den Kontrast S – RS1 kommt, wobei diese außerhalb des mittleren Teils des *Sulcus temporalis superior* stärker in der linken Hemisphäre ausgeprägt ist. In der linken Gehirnhälfte ist ein großes zusammenhängendes Aktivierungsgebiet zu sehen. Dieses erstreckt sich vom sekundären auditorischen Kortex ausgehend einmal in die posteriore Richtung hin zum posterioren *Sulcus temporalis superior* und endet im inferioren Parietallappen im *Gyrus angularis* und *Gyrus supramarginalis*. In anteriore Richtung verläuft die Aktivierung über den anterioren *Sulcus temporalis superior* bis in den frontalen Kortex (*Gyrus frontalis inferior*). Bei dem Kontrast S – Rauschen ist zudem noch eine deutliche Aktivierung im primären auditorischen Kortex und weiten Teilen des *Gyrus temporalis superior* zu sehen. Die Gebiete aus dem vorherigen Kontrast sind ebenfalls aktiv. Allerdings ist in beiden Hemisphären eine geringere Gehirnaktivität im *Gyrus supramarginalis* zu beobachten.⁸

⁸Unter <http://medi.uni-oldenburg.de/download/diplom/hw08/> ist eine animierte Fahrt durch alle Schichten des Gehirns für verschiedene Kontraste zu finden, um auf diese Weise mehr Aktivierungsmuster darstellen zu können, als dies in einer Abbildung möglich ist.

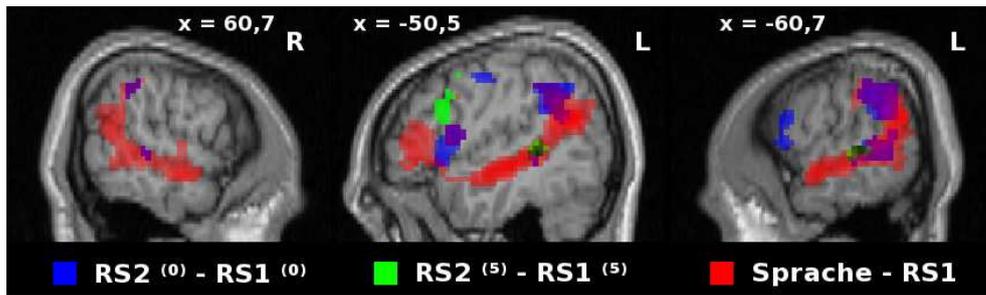


Abbildung 5.4: Sprachverarbeitung im Kortex. In Rot ist der Kontrast zwischen allen unverzerrten Sprachstimuli (S) und allen rotierten, unverständlichen Sprachstimuli (RS1) eingezeichnet (vgl. Abb. 5.3). Die beiden anderen Kontraste wurden jeweils zwischen akustisch identischer rotierter Sprache, jedoch mit unterschiedlicher Verständlichkeit gebildet. Bei dem in Blau eingezeichneten Kontrast betrug die Verständlichkeit von RS2 95% (Abstand 0), bei dem in Grün dargestellten 50% (Abstand 5). Die Verständlichkeit von RS1 betrug in allen drei Fällen 10%. Es ist mit x jeweils die Koordinate des Schnittes im MNI-Raum angegeben.

Ort	p (FWE)	Z-Wert	Voxel	x/mm	y/mm	z/mm
L <i>Gyrus supramarginalis</i>	0	Inf	286	-60	-48	27
L <i>Gyrus praecentralis</i>	0	6,67	19	-51	-9	45
L <i>Gyrus frontalis inferior</i>	0	7,84	130	-54	9	9
L <i>Sulcus temporalis superior</i>	0	7,51	131	-61	-42	0
R <i>Gyrus supramarginalis</i>	0,001	5,42	11	60	-42	39
R <i>Sulcus temporalis superior</i>	0,001	5,42	13	63	-30	0

Tabelle 5.3: Signifikante Aktivierungen für den Kontrast RS2⁽⁰⁾-RS1⁽⁰⁾. Es ist der anatomische Ort des jeweiligen Maximums eines aktivierten Gebietes angegeben. Seine korrigierter p -Wert, sein Z-Wert, die Anzahl aktiver Voxel und seine Position im MNI-Raum. L steht für links, R für rechts.

In Abbildung 5.4 sind Aktivierungen des Kortex nur auf Grund unterschiedlicher Verständlichkeit derselben akustischen Stimuli zu sehen. Dazu ist in Rot erneut der Kontrast S – RS1 aus Abbildung 5.3 dargestellt. In Blau ist der Kontrast zwischen der zu 95% verständlichen rotierten Sprache (RS2⁽⁰⁾) und der zu 10% verständlichen rotierten Sprache (RS1⁽⁰⁾) aufgetragen. Der grüne Bereich resultiert aus dem gleichen Kontrast, nur dass die rotierte Sprache RS2⁽⁵⁾ lediglich zu ungefähr 50% verständlich war.

Wird der in Blau eingetragene Kontrast betrachtet (vgl. Tabelle 5.3), ist zu erkennen, dass in der linken Hemisphäre vier voneinander getrennte Gebiete aktiv sind. In der rechten Hemisphäre ist ebenfalls eine schwache Aktivität zu erkennen,

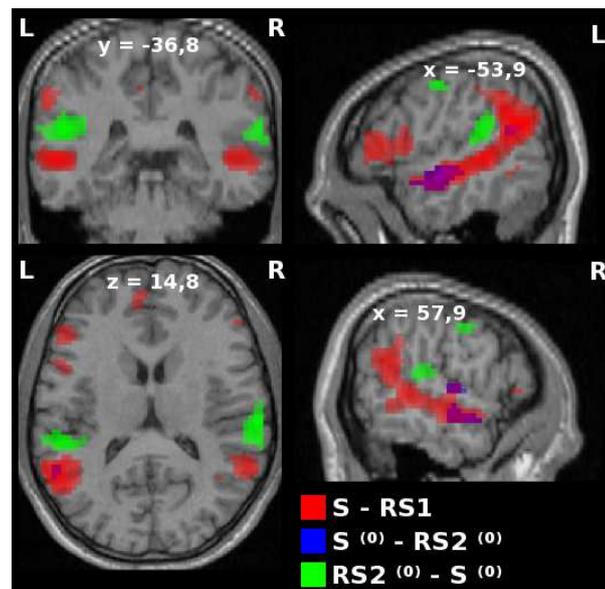


Abbildung 5.5: Rot: Kontrast zwischen der natürlichen Sprache und der unverständlichen rotierten Sprache ($S - RS1$). Blau: Signifikant stärkere Aktivierungen für die natürliche Sprache verglichen mit der zu 95% verständlichen rotierten Sprache ($S - RS2^{(0)}$). Grün: Signifikant stärkere Aktivierungen für die zu 95% verständliche rotierte Sprache verglichen mit der natürlichen Sprache ($RS2^{(0)} - S$). Mit x ist jeweils die Schnittebene im MNI-Raum angegeben.

allerdings nur an zwei der vier Stellen. Dabei handelt es sich bei den auf beiden Seiten aktiven Gebieten um einen kleinen Teil des posterioren *Sulcus temporalis superior*, der in das Wernicke-Areal fällt und um den *Gyrus supramarginalis* im inferioren Parietallappen. In der linken Hemisphäre ist zusätzlich eine Aktivität im inferioren Frontallappen in einem Teil des Broca-Areals zu beobachten. Zudem ist eine Aktivität im primären Motorkortex festzustellen. Die bilateralen Aktivierungen liegen jeweils in dem Gebiet, das auch bei dem Kontrast zwischen der unverzerrten Sprache und dem Rauschen aktiv war. Die Aktivierung im Motorkortex ist bei diesem Kontrast jedoch nicht gefunden worden. Außerdem ist eine kleine Abweichung im Aktivierungsmuster im frontalen Kortex zu erkennen, mit einer Verschiebung der Aktivierung für den Fall $RS2^{(0)} - RS1^{(0)}$ in Richtung Broca-Areal.

Bei einer Verringerung der Sprachverständlichkeit von $RS2$ ist zu erkennen, dass die Aktivität insgesamt zurückgeht (grün). Sie ist nur noch in der linken Hemisphäre zu beobachten. Gleichzeitig ist sie auf die Gebiete Wernicke-Areal und Prämotorkortex beschränkt.

Weitere interessante Kontraste ergeben sich aus der Fragestellung nach einem Unterschied in der Verarbeitung der unverzerrten und somit verständlichen Sprache und der verständlichen rotierten Sprache. Dazu sind in Abbildung 5.5 die

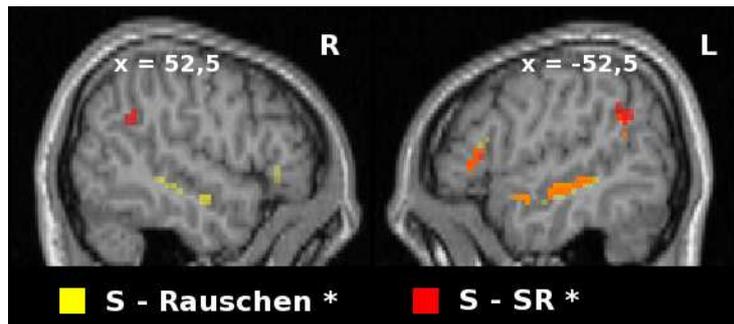


Abbildung 5.6: Dargestellt sind die Kontraste „Sprache“ minus „Rauschen“ (S – Rauschen, gelb) und „Sprache“ minus „Sprache im Rauschen“ (S–SR, rot) jeweils mit dem Kontrast „Rauschen“ minus „Stille“ maskiert (*). Mit x ist jeweils die Schnittebene im MNI-Raum angegeben.

Kontraste $RS2^{(0)} - S^{(0)}$ und $S^{(0)} - RS2^{(0)}$ eingetragen. Zusätzlich ist der Kontrast „Sprache“ minus „unverständlicher rotierter Sprache“ (S – RS1) dargestellt. Deutlich wird, dass der Kontrast $RS2^{(0)} - S^{(0)}$ zu einer Aktivierung im bilateralen *Planum temporale* führt. Außerdem ist eine Aktivierung im Motorkortex in der linken Hemisphäre zu beobachten. Im Vergleich der natürlichen Sprache mit der ähnlich verständlichen, rotierten Sprache fällt auf, dass diese eine stärkere Aktivität im anterioren *Sulcus temporalis superior* in beiden Hemisphären bewirkt. Zudem ist in der linken Hemisphäre ein Bereich mit stärkerer Aktivität am posterioren Ende des Temporallappens, am Übergang zum Parietallappen zu finden. In der rechten Hemisphäre zeigt ein Bereich im *Planum polare* mehr Aktivität für die unverzerrten Sprachstimuli.

Sprache im Rauschen

Neben der rotierten Sprache kam mit Sprache im Störgeräusch noch ein weiterer Sprachstimulus zum Einsatz, dessen Verständlichkeit deutlich unterhalb von 100% lag. Bei dem Störgeräusch handelte es sich um Rauschen, welches auch noch als selbständige Kondition dargeboten worden ist.

Wird der Kontrast „Sprache im Rauschen“ minus „Sprache“ (SR – S) betrachtet, finden sich zwei Regionen im Frontallappen, die eine Aktivität zeigen. Die eine davon ist nur in der rechten Hemisphäre und im inferioren Frontallappen angesiedelt. Die andere im *Gyrus frontalis medialis*.

Ebenso kann noch betrachtet werden, ob mehr Aktivität im Gehirn für Sprache als für Sprache im Rauschen zu beobachten ist. Der direkte Kontrast zwischen diesen beiden Bedingungen liefert ein ähnliches Bild wie der zwischen Sprache und Rauschen. Um eine differenziertere Betrachtung zu ermöglichen, ist daher im Folgenden die Aktivierung, die nur durch das Rauschen zu Stande gekommen

ist, ausgeblendet worden. Hierzu wird der zu beobachtende Kontrast mit dem Kontrast „Rauschen“ minus „Stille“ maskiert. Die Ergebnisse werden in Abbildung 5.6 präsentiert. Dabei werden die Kontraste S – Rauschen (gelb) und S – SR (rot) mit der beschriebenen Maskierung dargestellt. Es sind bei beiden Kontraste ähnliche Aktivierungen zu sehen, wie dies schon bei den bisherigen Kontrasten mit verständlicher Sprache der Fall war. Folglich ist eine Aktivierung im *Sulcus temporalis superior* und teilweise im *Gyrus temporalis medius* zu erkennen. Diese ist allerdings nur für den Fall S – Rauschen * bilateral zu finden. Sie ist zudem weiter anterior als im Fall RS2 – RS1. Im *Gyrus supramarginalis* sind bilaterale Aktivierungen zu beobachten, welche für den Fall S – SR * stärker ausgeprägt sind. Ebenso lässt sich insgesamt eine stärkere Aktivität in der linken Hemisphäre feststellen. Im inferioren Frontallappen ist nur für den Kontrast S – Rauschen * eine bilaterale Aktivierung zu finden. Auch diese stärker in der linken Hemisphäre ausgeprägt. Sie liegt – wie schon im Fall S – RS1 – weiter anterior im Vergleich mit RS2 – RS1.

5.3.2 Einzelergebnisse

Durch die Verwendung von 9 verschiedenen Konditionen in der Messung ist die Anzahl an Wiederholungen der einzelnen Konditionen pro Versuchsperson mit 38 relativ gering. Folglich war bei einigen nur bei den Kontrasten S – Stille, S – Rauschen und RS – Rauschen ein signifikanter Unterschied in der Gehirnaktivität festzustellen. Für den Kontrast RS2⁽⁰⁾ – RS1⁽⁰⁾ ergab sich lediglich bei einem einzigen Probanden ein signifikanter Unterschied in der Aktivierung. Die Interpretation dieses Kontrastes ist daher auf der Ebene einzelner Versuchspersonen nicht möglich.

Um dennoch die Lateralisierung auf Ebene einzelner Versuchspersonen untersuchen zu können, wird der Kontrast S – RS1 im nächsten Abschnitt betrachtet.

Lateralisierung

Bei dem Kontrast S – RS1 zeigen 9 der 15 Versuchspersonen eine signifikante Aktivierung. Die anderen 6 Probanden sind aus der Auswertung herausgenommen worden. Um eine Aussage über die Lateralisierung treffen zu können, ist die Anzahl an aktiven Voxeln in der jeweiligen Hemisphäre bestimmt worden. Dabei wurde nicht das ganze Gehirn betrachtet, sondern die Temporallappen und die lateralen Bereiche ($|x| > 15$) des Frontal- und des Parietallappens. Die Ergebnisse für die 9 Versuchspersonen für die Aktivität in den jeweiligen Hemisphären werden in Tabelle 5.4 präsentiert.

Zur genaueren Quantifizierung der Lateralisierung ist ein Lateralisierungsindex nach der folgenden Formel berechnet worden:

$$\text{Lateralisierungsindex} = 100 \frac{\text{Voxel rechts} - \text{Voxel links}}{\text{Voxel rechts} + \text{Voxel links}}$$

Proband	Aktivierung links/Voxel	Aktivierung rechts/Voxel	Lateralisierungsindex	Händigkeit
VP3	116	71	-24	90
VP5	666	162	-61	100
VP6	65	135	35	100
VP7	27	20	-15	100
VP8	118	152	13	40
VP11	10	275	93	100
VP12	486	44	-83	79
VP13	979	176	-67	88
VP15	72	35	-35	-11
Summe	2539	1070	-41	78
Mittelwert	1531	881	-26	78

Tabelle 5.4: Lateralisierung bei 9 Versuchspersonen. Die Aktivierung ist in aktiven Voxeln in einer ausgewählten Region in der jeweiligen Hemisphäre für den Kontrast S – RS1 eingetragen. Daraus ist der Lateralisierungsindex mit $100 \frac{\text{Voxel rechts} - \text{Voxel links}}{\text{Voxel rechts} + \text{Voxel links}}$ berechnet worden. 100 steht für eine vollständige Lateralisierung in die rechte und -100 für eine vollständige Lateralisierung in die linke Hemisphäre. Zusätzlich werden die Ergebnisse für die Händigkeit der Probanden dargestellt. Die Zahlenwerte haben hier die entsprechende Bedeutung. Außerdem ist die Summe über alle aktiven Voxel der Probanden gebildet und daraus der Lateralisierungsindex berechnet worden. Der Eintrag für den Mittelwert stammt aus dem Kontrast S – RS1 des *fixed-effects*-Modells (siehe Abschnitt 5.3.1).

Zusätzlich sind die Ergebnisse des *Edinburgh Handedness Inventory* für die Händigkeit der einzelnen Probanden eingetragen worden. Die beiden untersten Einträge in der Tabelle stellen einmal die Summe über alle aktiven Voxel der einzelnen Versuchspersonen und die aktiven Voxel für das *fixed-effects*-Modell für den Mittelwert dar.

In Abbildung 5.7 ist der Lateralisierungsindex in Abhängigkeit der Händigkeit der Versuchspersonen aufgetragen. Es ist zu erkennen, dass 5 der 7 Rechtshänder die Aktivität für den Kontrast S – RS1 in die linke Hemisphäre lateralisiert haben. Gleichzeitig ist bei 2 rechtshändigen Versuchspersonen die stärkste Lateralisierung in die rechte Hemisphäre anzutreffen. Die beiden beidhändigen Probanden weisen mit den Lateralisierungsindizes 13 und -35 eher eine beidseitige Aktivierung auf.

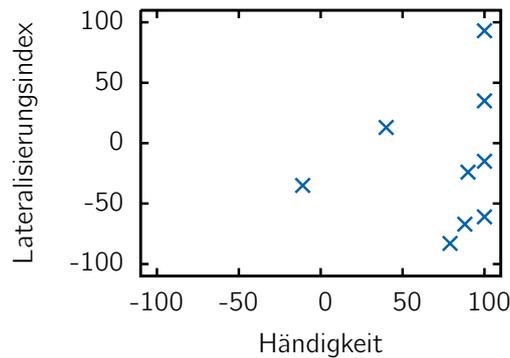


Abbildung 5.7: Zusammenhang zwischen der Lateralisierung der Gehirnaktivität für den Kontrast S – RS1 und der Händigkeit der einzelnen Versuchspersonen. Jedes Kreuz steht für eine Versuchsperson. Die Händigkeit ist mit dem *Edinburgh Handedness Inventory* bestimmt worden. Für die Bestimmung des Lateralisierungsindex siehe die Beschreibung im Text.

5.4 Diskussion

Sprachverarbeitung

Bei dem Kontrast S – RS1 zeigt sich ein Bild der Sprachverarbeitung, wie es in der Einleitung (vgl. auch Davis und Johnsrude, 2007) umschrieben wurde: Ein *ventraler* Verarbeitungszweig, der vom *Sulcus temporalis superior* und *Gyrus temporalis medius* ausgehend in deren anteriore Teile verläuft und im prefrontalen Kortex endet; ein *dorsaler* Verarbeitungszweig, der vom *Sulcus temporalis superior* und *Gyrus temporalis medius* ausgeht und im *Gyrus angularis* und *Gyrus supramarginalis* mündet. Hier beginnend ist er laut den vorgestellten Theorien mit dem Broca-Areal im *Gyrus frontalis inferior* verbunden (vgl. Abbildung 5.3). Ob diese Verbindung tatsächlich vorliegt, konnte mit der angewandten Messmethode nicht untersucht werden. Es ist allerdings auch Aktivierung im Broca-Areal gefunden worden.

Eine Aktivität im primären auditorischen Kortex ist nicht zu erkennen. Dies spricht dafür, dass die rotierte Sprache ein geeigneter Vergleichsstimulus ist, um die Sprachverarbeitung zu untersuchen. Der Kontrast „Sprache“ minus „Rauschen“ zeigt hingegen eine deutliche Aktivierung in diesem Bereich. Frühere Studien fanden für den Kontrast S – RS1 meistens deutlich kleinere Aktivierungsbereiche, die auf den Temporallappen und den *Gyrus supramarginalis* beschränkt waren (Scott et al., 2000; Narain et al., 2003). Es besteht daher die Möglichkeit, dass die in diesem Versuch von den Probanden während der Messung durchzuführende Aufgabe zu der größeren Aktivität geführt hat. Dies müsste in einer weiteren Messung überprüft werden, in der auch Durchläufe ohne Aufgabe aufgenommen

werden.

Im darauf folgenden Schritt wurde mit akustisch identischen Stimuli die Sprachverarbeitung untersucht (RS2 – RS1). Dabei zeigte sich, dass diese Untersuchung mit Hilfe von rotierter Sprache möglich ist. Vier Regionen in der linken Hemisphäre ergaben sich, von denen mindestens drei mit der Verarbeitung des Sinngeltes von Sprache in Verbindung gebracht werden können: Teile des Wernicke-Areals im *Sulcus temporalis superior*, der *Gyrus supramarginalis* und Teile im inferioren Frontallappen, die im Broca-Areal liegen. Der vierte Bereich befindet sich im Motorkortex und deutet auf die besondere Verarbeitung von rotierter Sprache im Vergleich zu natürlicher Sprache hin. Diese besondere Verarbeitung wird weiter unten noch diskutiert werden.

Dehaene-Lambertz et al. (2005) haben ebenfalls den Unterschied zwischen der Wahrnehmung von akustisch identischen Stimuli als Sprache oder als Nicht-Sprache untersucht. Sie verwendeten dazu Silben aus Sinussprache (Remez et al., 1981) und machten sich den Umstand zu nutze, dass die Versuchspersonen diese zu Beginn der Messung nicht verstanden hatten. Zu einem späteren Zeitpunkt haben sie jedoch spontan in einen Sprachmodus „gewechselt“ und konnten die Stimuli verstehen. Durch diese Darbietung konnten die Autoren die Verständlichkeit ihrer Stimuli nicht wie im vorliegenden Versuch genau kontrollieren. Trotzdem ist es ihnen gelungen, ebenfalls Aktivierungen im *Sulcus temporalis superior*, *Gyrus supramarginalis* und im Frontallappen zu finden. Daher ist davon auszugehen, dass es sich bei den drei oben aufgezählten Gebieten tatsächlich um zentrale Areale für die Verarbeitung des Sinngeltes von Sprache handelt. Ein weiterer wichtiger Punkt der dafür spricht, ist die zu beobachtende Reduzierung der Aktivität in diesen Bereichen, wenn die Verständlichkeit von RS2 verringert wird (vgl. Abb. 5.4).

Keine Aussagen können mit Hilfe der durchgeführten fMRT-Messung in Bezug auf die Art der Verarbeitung in den einzelnen Verarbeitungszweigen getroffen werden. Dazu müssten Stimuli benutzt werden, die sich in ihrer linguistischen Komplexität stärker unterscheiden.

Die Lokalisation der Verarbeitung des Sinngeltes ist im vorliegendem Versuch ausschließlich durch verständliche rotierte Sprache geschehen. Daher stellt sich die Frage, ob es einen Unterschied in der Verarbeitung von natürlicher Sprache und der zu 95% verständlichen rotierten Sprache gibt. In Abbildung 5.5 ist zu erkennen, dass es für die verständliche rotierte Sprache zwei Gebiete in jeder Hemisphäre gibt, in denen sie zu stärkerer Gehirnaktivität als unverzerrte Sprache führt. Eines davon befindet sich im *Planum temporale*. Es ist zu vermuten, dass diese Region mit der Kodierung des auditorischen Gedächtnisses in Verbindung gebracht werden kann, da dieses selbstverständlich für die Kondition RS2 stärker beansprucht wird als für die Kondition S. Das andere Gebiet ist im primären Motorkortex zu orten und zwar an identischer Position wie für den Kontrast RS2⁽⁰⁾ – RS1⁽⁰⁾. Da es

auch im letztgenannten Fall zu beobachten war, ist davon auszugehen, dass die Verständlichkeit eine Voraussetzung für die Aktivität in diesem Gebiet darstellt. Ein Grund für die beobachtete Aktivität wird daher in einem innerlichen Nachsprechen der verständlichen rotierten Sprache zu finden sein, um auf diese Weise ihren Inhalt entschlüsseln zu können. Eine Vorgehensweise, die für das Verständnis der unverzerrten Sprache nicht notwendig ist.

Der umgekehrte Kontrast ($S^{(0)} - RS2^{(0)}$) zeigte eine starke Aktivierung hauptsächlich im anterioren Temporallappen. Da dort die Verarbeitung syntaktischer und semantischer Eigenschaften von Sätzen vermutet wird (Mazoyer et al., 1993), ist davon auszugehen, dass dies bei der Kondition RS2 weniger der Fall war. Dieses Ergebnis legt die Vermutung nahe, dass für das Verstehen der Kondition RS2⁽⁰⁾ teilweise die Erinnerung an den Satz verwendet werden kann und er nicht vollständig neu verarbeitet werden muss.

Lateralisierung

Durch die Verwendung unterschiedlicher Kontraste war es möglich, sowohl frühere als auch spätere Stationen in der Sprachverarbeitung zu erfassen. Durch den Kontrast zwischen der Kondition „Sprache“ (S) und „unverständliche rotierte Sprache“ (RS1) konnten frühe Bereiche in der Sprachverarbeitung untersucht werden. Die Signale waren in diesem Fall nicht identisch, aber akustisch aufeinander abgestimmt. Durch den Kontrast zwischen der Kondition „verständliche rotierte Sprache“ (RS2) und „unverständliche rotierte Sprache“ (RS1) konnten Bereiche in der Sprachverarbeitung untersucht werden, die den Sinngehalt der Sprache kodieren. Dies war möglich, da die Stimuli akustisch identisch waren.

Gezeigt hat sich dabei, dass die frühe Verarbeitung größtenteils bilateral stattfindet. Dies trifft zum einen auf den *Sulcus temporalis superior* zu, den auch Hickok und Poeppel (2007) als eine bilaterale Stufe in ihr Modell eingebaut hatten. Gleiches gilt aber auch für den *Gyrus supramarginalis* und den inferioren Frontallappen. Die späteren Verarbeitungen, insbesondere das Kodieren des Sinngehaltes der Sprache, werden hingegen vorwiegend in der linken Hemisphäre durchgeführt. Die Ergebnisse sind insofern überraschend, da mit Ausnahme des Modells von Hickok und Poeppel (2007) alle anderen Modelle zur Sprachverarbeitung von einer Lateralisierung der gesamten Verarbeitung in die linke Hemisphäre ausgehen. Einzig für Linkshänder werden Ausnahmen gemacht. Weiter unten wird daher noch diskutiert werden, wie stark der Einfluss der Händigkeit der Versuchspersonen auf die vorliegenden Ergebnisse war.

Die beobachteten Ergebnisse klären einen weiteren wichtigen Punkt, nämlich die Anwendbarkeit der *Trade-Off-Hypothese* von Zatorre et al. (2002) als Erklärung für die Lateralisierung der Sprachverarbeitung. Dabei gehen Zatorre et al. davon aus, dass schnelle zeitliche Änderungen in akustischen Signalen zu einer bevorzugten Verarbeitung in der linken Hemisphäre führen würden. Dies nehmen

sie dann als Startpunkt und Ursache der Sprachlateralisierung an. In der vorliegenden Studie konnte gezeigt werden, dass die Lateralisierung erst bei einem relativ späten Punkt in der Verarbeitungshierarchie einsetzt. Die Lateralisierung kann somit nicht ausschließlich durch Unterschiede in der auditorischen Verarbeitung in beiden Hemisphären erklärt werden. Die *Trade-Off-Hypothese* reicht folglich nicht als Begründung aus. Einige Teilaspekte lassen sich damit beschreiben, wie zum Beispiel die eher in der rechten Hemisphäre stattfindende Kodierung der Sprachmelodie. Diese kann jedoch bei entsprechender Aufgabenstellung ebenfalls in Regionen der linken Hemisphäre stattfinden (Zatorre und Gandour, 2008). Damit zeigt sich auch, dass die Lateralisierung ebenso wie Prozesse auf auditorischer Ebene durch *top-down*-Prozesse beeinflusst werden kann. Es überrascht daher nicht, wenn Zatorre in der Veröffentlichung mit Gandour 2008 ebenfalls Abstand von seiner ursprünglichen These nimmt.

Die Auswertung der Lateralisierung bei den einzelnen Versuchspersonen zeigt, dass es unwahrscheinlich ist, dass die beim Mittelwert beobachtete bilaterale Verarbeitung durch die unterschiedliche Händigkeit der Probanden erklärt werden kann. Es zeigten sogar zwei rechtshändige Probanden die stärkste Lateralisierung in die rechte Hemisphäre. Gleichzeitig ergab sich für alle Versuchsperson für den Kontrast S – RS1 eine Aktivierung in beiden Hemisphären, auch wenn sie natürlich unterschiedlich stark ausgeprägt war. Es ist daher davon auszugehen, dass die unterschiedliche Händigkeit der Versuchspersonen keinen signifikanten Einfluss auf die Ergebnisse des *fixed-effects*-Modell zur Händigkeit hatte.

Sprache im Rauschen

Für die Verwendung von Sprache im Störgeräusch gilt es zu beachten, dass der Scanner eine Verschlechterung des Signal-zu-Rausch-Abstandes bewirkte. Dies hatte sich insbesondere daran gezeigt, dass die Verständlichkeit für diese Kondition im Magnetresonanztomographen deutlich geringer war, als es in der Hörkabine noch der Fall gewesen war.

Für den Kontrast S – SR oder auch S – Rauschen hat sich gezeigt, dass diese ebenfalls dafür verwendet werden können, späte Stationen der Sprachverarbeitung zu lokalisieren. Dazu mussten diese Kontraste allerdings mit der Aktivierung durch das Rauschen (Rauschen – Stille) maskiert werden. Es kam dabei zu ähnlichen Ergebnissen wie durch den Kontrast zwischen der verständlichen rotierten Sprache und der unverständlichen rotierten Sprache.

Einer der Gründe für die Verwendung von Sprache im Rauschen als eine Kondition in der vorliegenden Studie war die Idee, einen Kontrollstimulus für die schwierige Verständlichkeit von rotierter Sprache zu haben. Auf diese Weise sollte eine Region im Gehirn lokalisiert werden können, die mit schlechter Verständlichkeit eines Sprachsignals korreliert ist. Um dies zu bewerkstelligen, wird der Kontrast SR – S mit dem Kontrast RS⁽⁵⁾ – S⁽⁵⁾ verglichen. Die Verständlichkeit der Kondi-

tion RS2⁽⁵⁾ war der Verständlichkeit der Sprache im Rauschen ähnlich. Bei dem durchgeführten Vergleich lässt sich eine Region im *Gyrus frontalis medialis* identifizieren, die in beiden Fällen aktiv war. Diese Region kann daher mit der Verarbeitung von schlecht verständlicher Sprache in Verbindung gebracht werden, unabhängig von dem Grund für diese Verschlechterung.

6 Zusammenfassung

In der vorliegenden Studie ist ein Netzwerk zur Sprachverarbeitung im Gehirn lokalisiert worden. Ein Bereich davon konnte identifiziert werden, für die Kodierung des Sinngehaltes der Sprache verantwortlich zu sein.

Um diese Ergebnisse erzielen zu können, sind zunächst die Schwierigkeiten eines geschickten Vergleichsstimulus für die Untersuchung der Sprachverarbeitung mit bildgebenden Verfahren diskutiert worden. Dabei hat sich rotierte Sprache in bisherigen Studien als geeignetes Instrument gezeigt (Scott et al., 2000; Narain et al., 2003). Das Phänomen des *Pop-Out-Effektes*, bei dem die Verständlichkeit von verzerrter Sprache durch ihre Darbietung vor oder nach dem entsprechenden unverzerrten Satz abhängt, wurde verwendet, um dasselbe akustische Signal mal verständlich und mal unverständlich darzubieten zu können. In einem psychoakustischen Experiment wurde anschließend evaluiert, wie die rotierte Sprache sich dabei verhielt. Zu gleich wurde untersucht, wie ihre Verständlichkeit parametrisch durch Variation des Abstandes zur unverzerrten Sprache verändert werden konnte. Es zeigte sich, dass sich mit der rotierten Sprache und dem *Pop-Out-Effektes* Verständlichkeiten zwischen 10% und 95% erzielen lassen.

Auf Grundlage dieser Ergebnisse wurde anschließend ein fMRT-Experiment entworfen, in dem zu 10%, 50% und 95% verständliche rotierte Sprache zum Einsatz kam. Ergebnis dieses Experimentes war das oben erwähnte Netzwerk der Sprachverarbeitung. Außerdem konnte gezeigt werden, dass sich die Lateralisierung der Sprachverarbeitung in die linke Hemisphäre erst auf relativ hoher Ebene vollzieht. Es ist daher nicht davon auszugehen, dass die Sprache allein auf Grund ihrer akustischen Merkmale und einer unterschiedlichen Verarbeitung auf Ebene des auditorischen Kortex in die linke Hemisphäre lateralisiert ist.

In naher Zukunft soll untersucht werden, ob es möglich ist, das fMRT-Experiment soweit zu vereinfachen, dass die Messzeit für einzelne Versuchspersonen auf ungefähr 20 Minuten reduziert werden kann. Gleichzeitig soll dabei das Ziel verfolgt werden, auch für den Kontrast zwischen der rotierten Sprache mit unterschiedlicher Verständlichkeit statistisch signifikante Ergebnisse für die einzelnen Versuchspersonen zu erzielen. Wenn dies gelingen sollte, wäre damit die Gelegenheit geschaffen, die Sprachverarbeitung bei einzelnen Probanden genau und schnell zu lokalisieren. Dies würde eine interessante Methode sowohl für den Einsatz in der Klinik als auch für die weitere Forschung auf dem Gebiet der Sprachverarbeitung darstellen.

A Anhang

A.1 Blesser Skript

```
function blesser(MaxFreq, InputFile, LowpassOutputFile, RotatedOutputFile)
% RotFreq = MaxFreq/2
% Lese Signal ein
[x,SampFreq] = wavread(InputFile);
x = x(:,1);
x = [x' zeros(1,(100/1000)*SampFreq)]';

% Schritt 1 - Tiefpassfilterung
lowpass_filter = 0.95 * MaxFreq;
[blpf, alpf] = ellip(6, 0.5, 35, lowpass_filter/(SampFreq/2));
signal1 = filtfilt(blpf, alpf, x);
% RMS-Bestimmung
n_samples=length(signal1);
input_rms = norm(signal1)/sqrt(length(signal1));

% Schritt 2 - Ausgleich des Langzeitspektrums
b = my_fir2(mf/2, SampFreq, 'pre', 256);
signal2 = fftfilt(b, signal1);

% Schritt 3 - Modulation
modulator = sin(2.0*pi*MaxFreq*linspace(1, n_samples, n_samples)/SampFreq)';
signal3 = signal2.*modulator;

% Schritt 4 - Tiefpassfilterung
signal4 = filtfilt(blpf, alpf, signal3);

% RMS-Angleichung
signal4 = signal4 * (input_rms/(norm(signal4)/sqrt(length(signal4))));
[signal4, signal1] = no_clip2(signal4, signal1);
% Speichere das tiefpassgefilterte unverzerrte und das rotierte Signal
wavwrite(signal1, SampFreq, LowpassOutputFile)
wavwrite(signal4, SampFreq, RotatedOutputFile)
```

A.2 Formanten

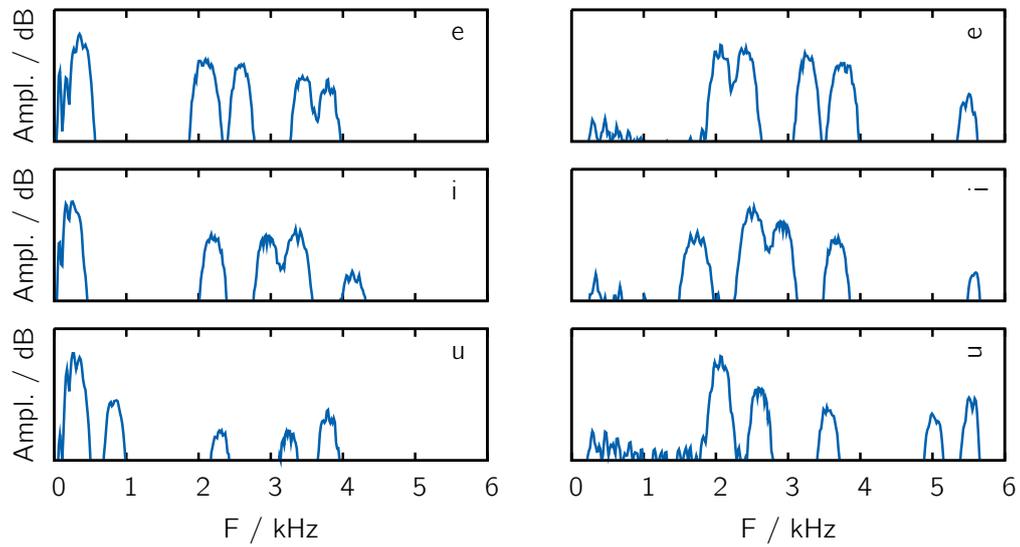


Abbildung A.1: Links: Spektren der unverzerrten Vokale. Rechts: Spektren der rotierten Vokale. In beiden Fällen ist eine Formantstruktur zu erkennen.

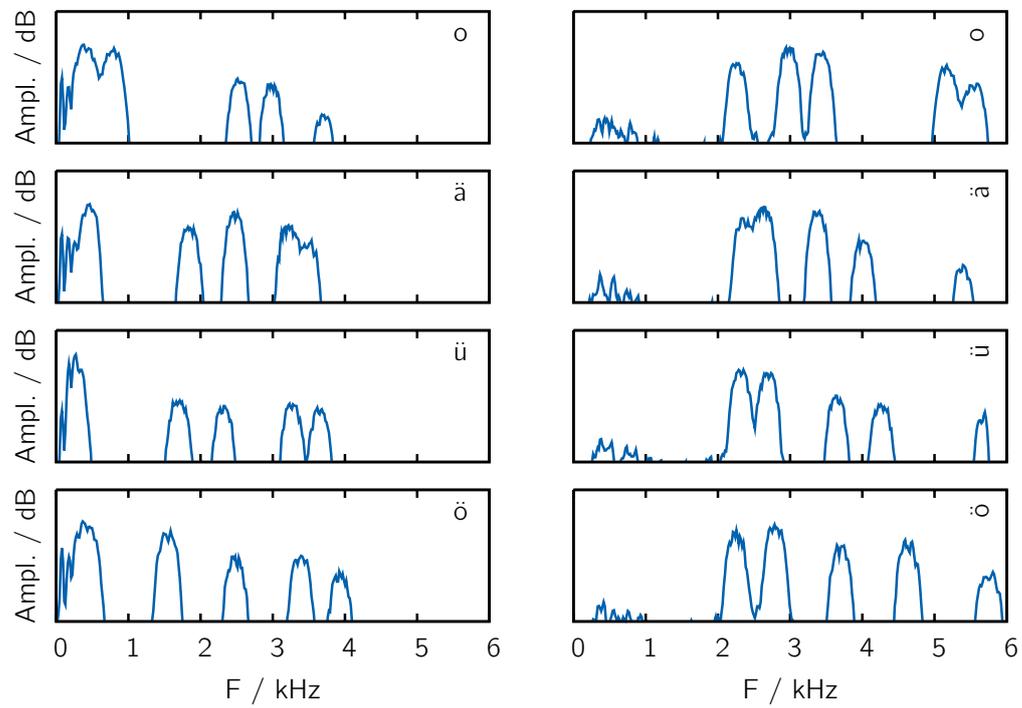


Abbildung A.2: Links: Spektren der unverzerrten Vokale. Rechts: Spektren der ro-
tierten Vokale. In beiden Fällen ist eine Formantstruktur zu erkennen.

A.3 Anatomie und Begriffserklärung

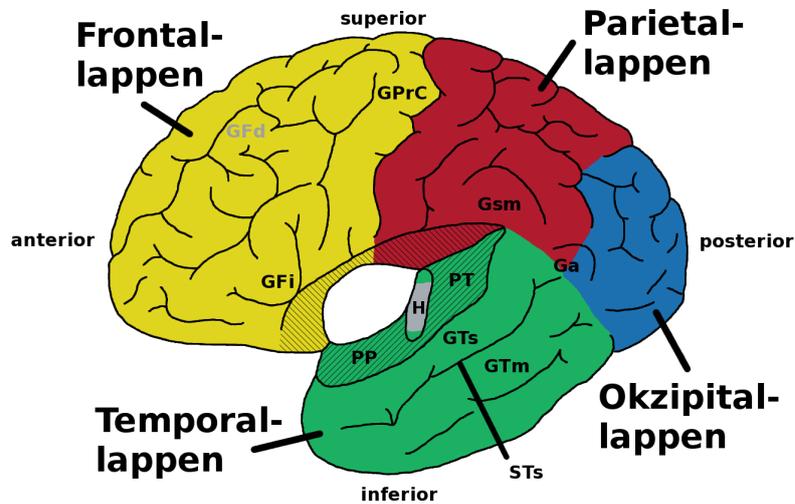


Abbildung A.3: Blick auf die linke Gehirnhemisphäre. Es sind jeweils die in dieser Arbeit verwendeten anatomischen Begriffe an ihren entsprechenden Positionen eingezeichnet. Die in Grau eingetragene Bezeichnung GFd befindet sich an der betreffenden Stelle im Inneren des Gehirns und nicht auf der hier dargestellten Oberfläche. Eine Erklärung der Abkürzungen ist Tabelle A.1 zu entnehmen. Die Gehirnlappen sind ebenso wie der primäre auditorische Kortex (grau) farbig gekennzeichnet.

Ga	<i>Gyrus angularis</i>	GTs	<i>Gyrus temporalis superior</i>
GFi	<i>Gyrus frontalis inferior</i>	H	<i>Gyrus temporalis transversus</i>
GFd	<i>Gyrus frontalis medialis</i>	PP	<i>Planum polare</i>
GPrC	<i>Gyrus praecentralis</i>	PT	<i>Planum temporale</i>
GSm	<i>Gyrus supramarginalis</i>	STs	<i>Sulcus temporalis superior</i>
GTm	<i>Gyrus temporalis medius</i>		

Tabelle A.1: Alle in dieser Studie vorkommenden anatomischen Bezeichnungen und ihre Abkürzungen, mit denen sie in Abbildung A.3 eingetragen sind.

A.4 fMRT-Ergebnisse

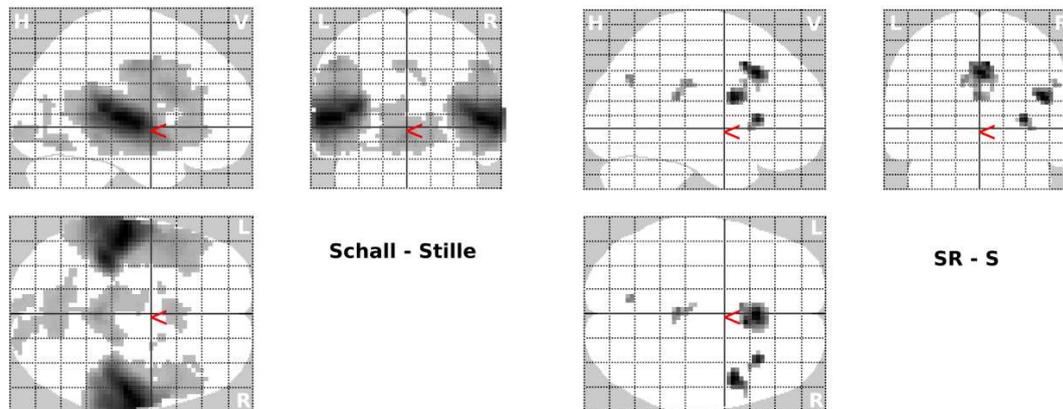


Abbildung A.4: *Links:* Kontrast zwischen „Schall“ (allen Konditionen, in denen Schall dargeboten worden ist) und „Stille“. *Rechts:* Kontrast zwischen den Konditionen „Sprache im Rauschen“ und „Sprache“. Es sind in den drei Ansichten des Gehirns jeweils die Aktivierungen über das gesamte Gehirn aufsummiert (*glasbrain view*).

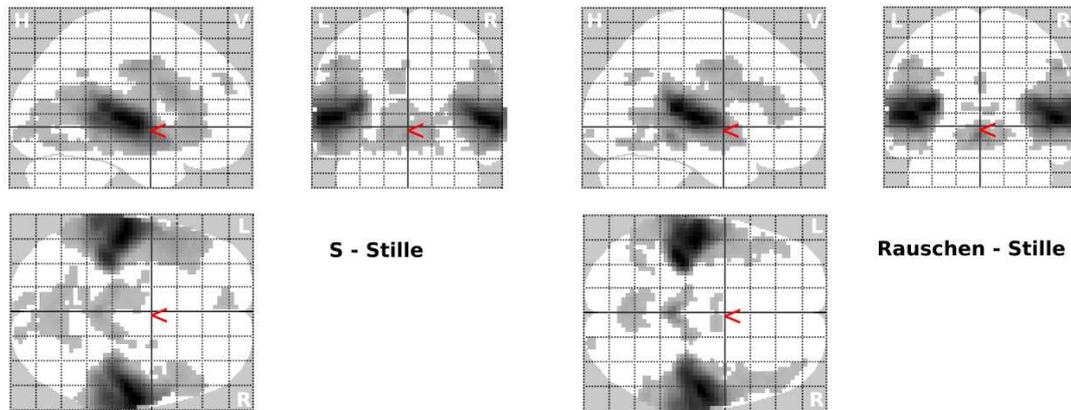


Abbildung A.5: Links: Kontrast zwischen den Konditionen „Sprache“ und „Stille“. Rechts: Kontrast zwischen den Konditionen „Rauschen“ und „Stille“. Es sind in den drei Ansichten des Gehirns jeweils die Aktivierungen über das gesamte Gehirn aufsummiert (*glasbrain view*).

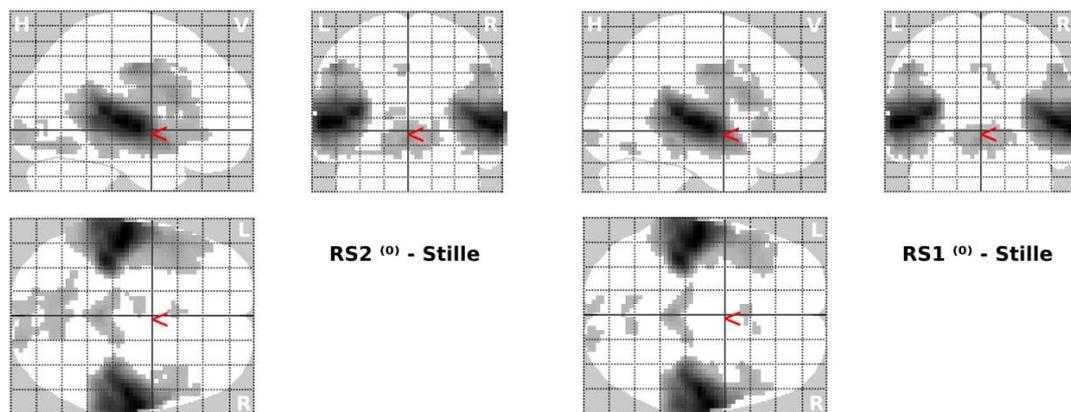


Abbildung A.6: Links: Kontrast zwischen den Konditionen „verständliche rotierte Sprache“ und „Stille“. Rechts: Kontrast zwischen den Konditionen „unverständliche rotierte Sprache“ und „Stille“. Es sind in den drei Ansichten des Gehirns jeweils die Aktivierungen über das gesamte Gehirn aufsummiert (*glasbrain view*).

Literaturverzeichnis

- M. Ahissar und S. Hochstein.
The reverse hierarchy theory of visual perceptual learning.
Trends Cogn Sci, 8(10):457–464, 2004.
- M. Awad, J. E. Warren, S. K. Scott, F. E. Turkheimer und R. J. S. Wise.
A common system for the comprehension and production of narrative speech.
J Neurosci, 27(43):11455–11464, 2007.
- M. Azadpour und E. Balaban.
Phonological representations are unconsciously used when processing complex, non-speech signals.
PLoS ONE, 3(4):e1966, 2008.
- A. D. Baddeley.
Human memory: theory and practice.
Psychology Press, London, 1997.
- J. R. Binder, J. A. Frost, T. A. Hammeke, P. S. Bellgowan, J. A. Springer, J. N. Kaufman und E. T. Possing.
Human temporal lobe activation by speech and nonspeech sounds.
Cereb Cortex, 10(5):512–528, 2000.
- B. Blesser.
Perception of Spectrally Rotated Speech.
PhD thesis, Massachusetts Institute of Technology, 1969.
- B. Blesser.
Speech perception under conditions of spectral transformation: I. phonetic characteristics.
J Speech Hear Res, 15:5–41, 1972.
- A. Boemio, S. Fromm, A. Braun und D. Poeppel.
Hierarchical and asymmetric temporal sensitivity in human auditory cortices.
Nat Neurosci, 8(3):389–395, 2005.
- D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C. Lui, J. Kiessling, M. Nasser Kotby, N. H. A. Nasser, W. A. H. El Kholy, Y. Nakanishi et al.

An international comparison of long-term average speech spectra.
J Acoust Soc Am, 96:2108–2120, 1994.

E. C. Cherry.

On Human Communications.

M.I.T. Press, Cambridge, Massachusetts, 1957.

M. H. Davis und I. S. Johnsrude.

Hearing speech sounds: top-down influences on the interface between audition and speech perception.

Hear Res, 229(1-2):132–147, 2007.

M. H. Davis, I. S. Johnsrude, A. Hervais-Adelman, K. Taylor und C. McGettigan.
Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences.

J Exp Psychol Gen, 134(2):222–241, 2005.

G. Dehaene-Lambertz, C. Pallier, W. Serniclaes, L. Sprenger-Charolles, A. Jobert und S. Dehaene.

Neural correlates of switching from auditory to speech perception.

Neuroimage, 24(1):21–33, 2005.

D. A. Hall, M. P. Haggard, M. A. Akeroyd, A. R. Palmer, A. Q. Summerfield, M. R. Elliott, E. M. Gurney und R. W. Bowtell.

„Sparse“ temporal sampling in auditory fMRI.

Hum Brain Mapp, 7(3):213–223, 1999.

G. Hickok und D. Poeppel.

Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language.

Cognition, 92(1-2):67–99, 2004.

G. Hickok und D. Poeppel.

The cortical organization of speech processing.

Nat Rev Neurosci, 8(5):393–402, 2007.

H. L. Jamison, K. E. Watkins, D. V. M. Bishop und P. M. Matthews.

Hemispheric specialization for processing auditory nonspeech stimuli.

Cereb Cortex, 16(9):1266–1275, 2006.

I. S. Johnsrude, V. B. Penhune und R. J. Zatorre.

Functional specificity in the right human auditory cortex for perceiving pitch direction.

Brain, 123(Pt 1):155–163, 2000.

- B. Klotz und W. Kumpf.
Selbstgeschätzte und tatsächliche Verständlichkeit sprachaudiometrischen Testmaterials.
Laryngorhinootologie, 74(10):591–593, 1995.
- S. Knecht, B. Dräger, M. Deppe, L. Bobe, H. Lohmann, A. Flöel, E. B. Ringelstein und H. Henningsen.
Handedness and hemispheric language dominance in healthy humans.
Brain, 123(Pt 12):2512–2518, 2000.
- B. Kollmeier und M. Wesselkamp.
Development and evaluation of a german sentence test for objective and subjective speech intelligibility assessment.
J Acoust Soc Am, 102(4):2412–2421, 1997.
- C. Liégeois-Chauvel, K. Giraud, J. M. Badier, P. Marquis und P. Chauvel.
Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of the human auditory cortex.
Ann N Y Acad Sci, 930:117–132, 2001.
- A. R. Lurija.
The working brain: An introduction to neuropsychology.
Penguin Books, London, 1976.
- A. R. Lurija.
Romantische Wissenschaft.
Rowohlt, Hamburg, 1993.
- S. L. Mattys, L. White und J. F. Melhorn.
Integration of multiple speech segmentation cues: a hierarchical framework.
J Exp Psychol Gen, 134(4):477–500, 2005.
- B. M. Mazoyer, N. Tzourio, V. Frak, A. Syrota, N. Murayama, O. Levrier, G. Salamon, S. Dehaene, L. Cohen und J. Mehler.
The cortical representation of speech.
J Cogn Neurosci, 5(4):467–479, 1993.
- C. Narain, S. K. Scott, R. J. S. Wise, S. Rosen, A. Leff, S. D. Iversen und P. M. Matthews.
Defining a left-lateralized response specific to intelligible speech using fMRI.
Cereb Cortex, 13(12):1362–1368, 2003.
- R. C. Oldfield.
The assessment and analysis of handedness: the edinburgh inventory.
Neuropsychologia, 9(1):97–113, 1971.

- A. D. Patel.
Music, Language, and the Brain.
Oxford University Press, New York, 2008.
- R. D. Patterson, S. Uppenkamp, I. S. Johnsrude und T. D. Griffiths.
The processing of temporal pitch and melody information in auditory cortex.
Neuron, 36(4):767–776, 2002.
- G. E. Peterson und H. L. Barney.
Control methods used in a study of the vowels.
J Acoust Soc Am, 24(2):175–184, 1952.
- K. M. Petersson, A. Reis, S. Askelöf, A. Castro-Caldas und M. Ingvar.
Language processing modulated by literacy: a network analysis of verbal repetition in literate and illiterate subjects.
J Cogn Neurosci, 12(3):364–382, 2000.
- J. Pujol, J. Deus, J. M. Losilla und A. Capdevila.
Cerebral lateralization of language in normal left-handed people studied by functional MRI.
Neurology, 52(5):1038–1043, 1999.
- R. E. Remez, P. E. Rubin, D. B. Pisoni und T. D. Carrell.
Speech perception without traditional speech cues.
Science, 212(4497):947–949, 1981.
- C. Rorden und H.-O. Karnath.
Using human brain lesions to infer function: a relic from a past era in the fMRI age?
Nat Rev Neurosci, 5(10):813–819, 2004.
- M. Sabri, J. R. Binder, R. Desai, D. A. Medler, M. D. Leitl und E. Liebenthal.
Attentional and linguistic interactions in speech perception.
Neuroimage, 39(3):1444–1456, 2008.
- M. Schönwiesner, R. Rübsem und D. Y. von Cramon.
Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex.
Eur J Neurosci, 22(6):1521–1528, 2005.
- S. K. Scott und R. J. S. Wise.
The functional neuroanatomy of prelexical processing in speech perception.
Cognition, 92(1-2):13–45, 2004.

- S. K. Scott, C. C. Blank, S. Rosen und R. J. S. Wise.
Identification of a pathway for intelligible speech in the left temporal lobe.
Brain, 123 Pt 12:2400–2406, 2000.
- G. Spitsyna, J. E. Warren, S. K. Scott, F. E. Turkheimer und R. J. S. Wise.
Converging language streams in the human temporal lobe.
J Neurosci, 26(28):7328–7336, 2006.
- S. Uppenkamp, I. S. Johnsrude, D. Norris, W. Marslen-Wilson und R. D. Patterson.
Locating the initial stages of speech-sound processing in human temporal cortex.
Neuroimage, 31(3):1284–1296, 2006.
- K. C. Wagener, B. T. und B. Kollmeier.
Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests.
Zeitschrift für Audiologie, 38:4–15, 1999.
- R. M. Warren.
Auditory Perception.
Cambridge University Press, 1999.
- C. Wernicke.
Der aphasische Systemkomplex. Eine psychologische Studie auf anatomischer Basis.
M. Cohn & Weigert, Breslau, 1874.
- R. Zatorre, P. Belin und V. Penhune.
Structure and function of auditory cortex: music and speech.
Trends Cogn Sci, 6(1):37–46, 2002.
- R. J. Zatorre und P. Belin.
Spectral and temporal processing in human auditory cortex.
Cereb Cortex, 11(10):946–953, 2001.
- R. J. Zatorre und J. T. Gandour.
Neural specializations for speech and pitch: moving beyond the dichotomies.
Philos Trans R Soc Lond B Biol Sci, 363(1493):1087–1104, 2008.
- S. Zeki.
Inner Vision: An Exploration of Art and the Brain.
Oxford University Press, USA, February 2000.

Danksagung

Zunächst möchte ich mich bei Prof. Birger Kollmeier und Juniorprof. Jesko Verhey dafür bedanken, dass sie mir die Möglichkeit gegeben haben, diese spannende Diplomarbeit bei Ihnen schreiben zu dürfen. In gleichen Maßen bedanke ich mich bei meinem Betreuer Stefan Uppenkamp. Er war ebenfalls sehr interessiert an dem Thema und stand stets mit fachlichem Rat und als Diskussionspartner zur Verfügung.

Darüber hinaus gilt auch ein großer Dank der gesamten Arbeitsgruppe „Medizinische Physik“, in der ich sehr gerne mitgearbeitet habe.

Ein extra Dankeschön möchte ich Doro für die vielen Diskussionen und Jörn, Jörg-Hendrik, Hendrik und Frank für die Rechenknechtung aussprechen. Außerdem gilt der Dank noch meinen weiteren Büropartnern für unsere angenehme Zeit und Thomas und Sven für die Zeit zwischen den Büroaufenthalten.

Und zu guter letzt noch ein dickes Dankeschön an Kristina für ihr strenges Korrekturlesen und die moralische Unterstützung.

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe.

Hagen Wierstorf