

Zusammenfassung / Abstract

Audiovisual integration during speech perception

Der Fakultät für Naturwissenschaften
der Otto-von-Guericke-Universität Magdeburg

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

am 15. August 2008

eingereichte Dissertation,

vorgelegt von Dipl.-Psych. Gregor R. Szycik

Zusammenfassung

Das Sprechen und das Hören von Sprache sind essentiell für das soziale Dasein. Während normaler Kommunikation von Angesicht zu Angesicht empfangen wir die Sprache durch zwei sensorischen Kanäle: den visuellen und den auditorischen. Die visuelle Information beteiligt sich im wesentlichen Ausmaß an Sprachwahrnehmung, indem sie die Verständlichkeit der Sprache in geräuschvollen Umgebungen verbessert. Darüber hinaus kann künstliche audiovisuell inkongruente Sprache zu neuen Wahrnehmungen führen, die weder der auditorischen noch der visuellen Information entsprechen, wie es durch den McGurk-Effekt deutlich wird. Die hier vorliegende Doktorarbeit beschäftigt sich mit dem Verstehen und Lokalisieren der neuronalen Prozesse, die der Integration von visueller und auditorischer Information während der Sprachwahrnehmung zugrunde liegen.

Im ersten Kapitel wurde ein Überblick über die relevante empirische und theoretische Literatur zur audiovisuellen Sprachwahrnehmung gegeben. Das erste Kapitel erklärt die Gegebenheiten, die die audiovisuelle Wahrnehmung der Sprache begleiten und determinieren. Auch aktuell wichtige Theorien zur audiovisuellen Integration während der Sprachwahrnehmung werden beschrieben insbesondere mit Blick auf die zugrunde liegenden neuronalen Prozesse.

Das zweite Kapitel beschreibt ein funktionelles kernspintomographisches (fMRT) Experiment zur Replizierung und Erweiterung früherer bildgebender Ergebnisse und zielt auf die Frage der neuralen Implementierung audiovisueller Integration während der Verarbeitung der Sprache. Die Daten wurden sowohl im Standardraum mittels voxelweiser Gesamthirnanalyse als auch mit Hilfe eines Ansatzes analysiert, bei dem die Hirnaktivierungen innerhalb von vorbestimmten auf individueller Anatomie basierenden Auswertungsfeldern erhoben wurden. Die voxelweise Analyse im Standardraum ergab ein weit verteiltes neuronales Netzwerk mit gesteigener Aktivität für inkongruente Stimulation, das den hinteren Teil des linken Sulcus temporalis superior, die Broca Region und deren rechthemisphärisches Pendant umfasst. Mittels der auf Auswertungsfeldern basierenden Analyse ist es möglich gewesen, Unterschiede in der Aktivität zwischen den Brodmann Arealen 44 und 45 innerhalb der Broca Region für inkongruente Stimulation zu finden.

Das Kapitel drei beschäftigt sich mit der Frage der audiovisuellen Sprachverarbeitung und rauschgestörten Umgebungen. Zur Aufdeckung der relevanten Hirnareale wurde ein schrittweiser bildgebender Ansatz angewendet. Zuerst ist eine funktionelle

Lokalisierungsmessung verwendet worden. Diesbezüglich kam ein Blockdesign mit auditorischer, visueller und audiovisueller Sprache als Stimuli zum Einsatz. Im zweiten Schritt wurde ein langsames ereigniskorreliertes bidgebendes Design in Verbindung mit einer so genannten sparse sampling Methode verwendet. Durch die Verbindung beider experimentellen Schritte konnten funktionelle Hirnareale im Bereich des hinteren Sulcus temporalis superior identifiziert werden. Diese zeigten signifikante Aktivierung für die Interaktion der audiovisuellen Kongruenz und der Verständlichkeit der Sprache.

Im vierten Kapitel wird eine ereigniskorrelierte 7 Tesla fMRT Studie zur Klärung der audiovisuellen Fusion während der Sprachwahrnehmung beschrieben. Die Analyse der Daten erlaubte die Untersuchung von Hirnarealen, die an Verarbeitung und Fusionierung der audiovisuellen Sprache beteiligt sind und Regionen, die mit Verarbeitung der auditorischen und visuellen Differenzen innerhalb der audiovisuell präsentierten Sprache beschäftigt sind. Fusionierte inkongruente Stimuli riefen stärkere Hirnaktivität hervor als natürlich gesprochene kongruente Silben. Betroffene Areale schlossen den kaudalen Teil des Sulcus temporalis superior und verschiedene frontale und parietale Hirnregionen ein. Erfolgreiche Fusion der audiovisuellen Sprache führt zu Aktivierung des kaudalen Teils der Region um Sulcus temporalis superior und Lobulus parietalis inferior rechts. Darüber hinaus konnten Hemisphärendifferenzen in der funktionellen Organisation der Region um Sulcus temporalis superior gefunden werden. Die linke Region zeigte zwei Aktivierungskluster für Verarbeitung von auditorischen und visuellen Differenzen während die linke Region beide Funktionen in einem Cluster beherbergte.

Ziel der fMRT Studie, die im Kapitel fünf beschrieben wird, war die Analyse der neuronalen Grundlagen für Defizite der Patienten mit Schizophrenie in audiovisueller Sprachverarbeitung. Zu diesem Zweck wurden eine Gruppe der Patienten mit Schizophrenie und eine Kontrollgruppe untersucht. Die statistische Auswertung der Daten erfolgte varianzanalytisch mit zwei Hauptfaktoren: audiovisuelle Kongruenz und Gruppenzugehörigkeit. Die Hirnaktivität der Patientengruppe unterschied sich von der Kontrollgruppe und spiegelte sich in einer Interaktion der beiden Haupteffekte wider. Die betroffenen Hirnareale lagen vorwiegend in der rechten Hemisphäre und schlossen die Pars opercularis, den Sulcus frontalis medius und den Gyrus temporalis superior ein. Zusätzlich ist die Interaktion bilateral im Bereich des Gyrus fusiformis und des Nucleus accumbens beobachtet worden. Es wird vorgeschlagen, dass die Defizite der Patienten mit Schizophrenie

in audiovisueller Integration während der Sprachwahrnehmung als Dysfunktion des motorischen Sprachsystems der rechten Hemisphäre zu verstehen sind.

Abstract

Producing of and listening to speech sounds is our daily business and essential for our social being. In normal face-to-face communication, we perceive speech both via visual and auditory sensory channels. Visual information significantly contributes to speech comprehension by improving the intelligibility in noisy environments. Furthermore, audiovisually incongruent artificial speech may lead to novel percepts that neither matches the auditory nor the visual information as evidenced by the McGurk effect. The present thesis is devoted to the understanding and localising of the neural processes underlying the integration of visual and auditory information during speech perception.

In the first chapter, a review of the relevant empirical and theoretical literature on audiovisual speech perception is given. The first chapter elucidates the circumstances attending and determining audiovisual speech perception. Chapter two describes a functional magnetic resonance imaging (fMRI) experiment conducted to replicate and extend earlier neuroimaging findings and addressing the neural implementation of audiovisual integration during speech processing. Data were analysed both in standard space by voxel-wise whole brain analysis and using a regions of interest approach based on individual anatomy. Chapter three addresses the issue of audiovisual speech processing in noisy environments. To reveal the respective brain areas a step-wise functional neuroimaging approach was used. First, a functional localizer session using a block design with auditory speech, visual speech, and audiovisual speech was conducted. In a second session slow-event-related imaging in conjunction with sparse sampling was used. In the fourth chapter, an event related 7 Tesla fMRI study addressing the issue of audiovisual binding during speech perception is described. Data analysis enabled investigation of both brain sites involved in processing and fusing of audiovisual speech and regions engaged in processing of auditory and visual differences within audiovisually presented speech. The goal of the study described in chapter five was to investigate the neural basis of the deficit in audiovisual speech processing in schizophrenic patients using fMRI. Concerning this matter a group of patients with schizophrenia and a control group participated in this experiment.