

**SEEING AND HEARING SPEECH, SOUNDS, AND SIGNS:  
FUNCTIONAL MAGNETIC RESONANCE IMAGING STUDIES  
ON FLUENT AND DYSLEXIC READERS**

**Johanna Pekkola**

**Academic dissertation**

To be presented with the permission of the Faculty of Medicine of the University of Helsinki for public discussion in Auditorium E at Helsinki University of Technology (Espoo, Finland) on June 28, 2006 at 12 noon.

Helsinki University of Technology  
Department of Electrical and Communications Engineering  
Laboratory of Computational Engineering  
Espoo, Finland

Helsinki University of Technology  
Advanced Magnetic Imaging Centre (AMI)  
Espoo, Finland

University of Helsinki  
Helsinki Medical Imaging Center  
Helsinki, Finland

## **Supervisors**

Iiro P. Jääskeläinen, Professor  
Helsinki University of Technology  
Laboratory of Computational Engineering  
Espoo  
Finland

Taina Autti, Docent  
University of Helsinki  
Helsinki Medical Imaging Center  
Helsinki  
Finland

## **Preliminary examiners**

Ritva Vanninen, Professor  
Kuopio University Hospital  
Department of Clinical Radiology  
Kuopio  
Finland

Jyrki Ahveninen, Docent  
Instructor in Radiology  
Harvard Medical School  
Department of Radiology, MGH  
Charlestown, MA  
U.S.A.

## **Official opponent**

Josef P. Rauschecker, Professor  
Georgetown Institute for Cognitive and Computational Sciences  
Department of Physiology and Biophysics  
Georgetown University Medical Center  
Washington, DC  
U.S.A.

## **Distribution**

Helsinki University of Technology  
Laboratory of Computational Engineering  
P.O.BOX 9203  
FIN-02015 HUT  
FINLAND  
Tel +358 9 451 6151  
Fax + 358 9 451 4830  
<http://www.lce.hut.fi>

Online in pdf-format:  
<http://ethesis.helsinki.fi/english.html>

© Johanna Pekkola  
E-mail: [johanna.pekkola@fimnet.fi](mailto:johanna.pekkola@fimnet.fi)

ISBN 952-92-0409-4 (paperback)  
ISBN 952-10-3178-6 (PDF)

Helsinki University Press

Helsinki 2006

**To Marko and Pihla**

# Contents

Contents.....	i
List of original publications.....	iii
Abbreviations.....	iv
Abstract.....	v
Introduction.....	1
Review of the literature .....	3
Auditory speech perception .....	3
Subcortical auditory pathway .....	3
Hierarchical organization of the auditory cortex .....	4
Speech-specific auditory processing .....	6
Visual speech perception .....	7
Crossmodal processing in the auditory cortex .....	7
Feedback or feedforward modulation? .....	9
Audiovisual speech perception .....	10
Audiovisual integration and interactions .....	10
Audiovisual interactions in speech perception .....	11
Attention and speech perception .....	13
Behavioral and theoretical aspects .....	13
Attentional modulation of sound-evoked responses .....	13
Attentional modulation of visual and audiovisual speech perception.....	15
Motor system in speech perception .....	16
The motor theory of speech .....	16
Mirror neurons .....	17
Evidence of MNS-like system in humans .....	18
Dyslexia and speech perception .....	20
Language and sensory processing deficits .....	20
Functional findings .....	21
BOLD fMRI .....	22
Principles of the MR signal .....	23
Basis of BOLD signal changes .....	24
Temporal relationships between neuronal activity, CBF, and BOLD .....	25
Measurement devices and sequences .....	27
Data-analysis methods .....	28
Benefits and limitations .....	30

Aims of the study .....	32
Materials and methods .....	33
Summary of methods (I-IV) .....	33
Subjects, stimuli, and experimental designs .....	33
fMRI data acquisition .....	35
fMRI data pre-processing .....	35
Statistical methods .....	35
Experimental settings .....	36
I .....	36
II .....	37
III .....	37
IV .....	38
Results .....	39
Lip-reading activates the human primary auditory cortex (I) .....	39
Attention to visual speech modulates auditory cortex activity (II) .....	41
Stimulus-driven and attentional effects on auditory cortex activity (III) .....	42
Differences between fluent and dyslexic readers' speech perception (IV) .....	43
Discussion .....	46
Lip-reading activates the human primary auditory cortex (I) .....	46
Attention to visual speech modulates auditory cortex activity (II) .....	48
Stimulus-driven and attentional effects on auditory cortex activity (III) .....	50
Differences between fluent and dyslexic readers' speech perception (IV) .....	51
Potential limitations .....	53
Insights for future studies .....	54
Conclusions .....	56
Acknowledgements .....	58
References .....	60
Original publications .....	77

# List of original publications

This thesis is based on the following four publications which are referred to in the text by their roman numerals:

- I**      **Pekkola J**, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M: Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 2005 Feb; 16(2):125-128.
- II**     **Pekkola J**, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Sams M: Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Human Brain Mapping* 2006 June; 27(6):471-477.
- III**    Rinne T, **Pekkola J**, Degerman A, Autti T, Jääskeläinen IP, Sams M, Alho K: Modulation of auditory cortex activation by sound presentation rate and attention. *Human Brain Mapping* 2005 Oct; 26(2):94-99.
- IV**    **Pekkola J**, Laasonen M, Ojanen V, Autti T, Jääskeläinen IP, Kujala T, Sams M: Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T. *NeuroImage* 2006 Feb; 29(3):797-807.

## Contributions of the author

I was the principal author in Studies I, II, and IV, for which I carried out the fMRI measurements, analyzed the data and wrote the papers; my co-authors provided contributions at all stages. In Study III, I substantially contributed to data acquisition and analysis and participated in writing the manuscript. I had an active role in planning all the experiments included in this thesis.

## Shared publications

Publication I has previously appeared as a part of another compilation thesis (Ojanen Ville: Neurocognitive mechanisms of audiovisual speech perception. Academic dissertation for the degree of Doctor of Philosophy, Helsinki University of Technology, May 2005).

# Abbreviations

ANOVA	analysis of variance
BA	Brodmann's area
BOLD	blood oxygenation level-dependent
CBF	cerebral blood flow
CNR	contrast to noise ratio
fMRI	functional magnetic resonance imaging
EEG	electroencephalography
EPI	echo-planar imaging
ERP	evoked response potential
GLM	general linear modeling
GR-EPI	gradient echo-planar imaging
HG	Heschl's gyrus
IPL	inferior parietal lobule
ISI	inter-stimulus interval
MEG	magnetoencephalography
MMN	mismatch negativity
MMF	mismatch field
MNS	mirror neuron system
MR	magnetic resonance
MRI	magnetic resonance imaging
MSI	multisensory integration
PAC	primary auditory cortex
PET	positron emission tomography
PT	planum temporale
RF	radiofrequency
ROI	region of interest
SC	superior colliculus
SEM	standard error of mean
SMA	supplementary motor area
SOA	stimulus onset asynchrony
STG	superior temporal gyrus
STS	superior temporal sulcus
TE	time to echo
TMS	transcranial magnetic stimulation
TR	time for repetition

# Abstract

**Background:** Speech has both auditory and visual components (heard speech sounds and seen articulatory gestures): We therefore perceive it through both ear and eye. During all perception, selective attention facilitates efficient information processing and enables concentration on high-priority stimuli. Its neural mechanisms are far better known for auditory than for visual speech perception. During face-to-face conversation, auditory and visual sensory systems interact at multiple processing levels and, further, the classical motor speech regions seem also to participate in speech perception. Auditory, visual, and motor-articulatory processes may thus work in parallel during speech perception, their use possibly depending on the information available and the individual characteristics of the observer. Because of their subtle speech perception difficulties possibly stemming from disturbances at elemental levels of sensory processing, dyslexic readers may rely more on motor-articulatory speech perception strategies than do fluent readers.

**Aims:** This thesis aimed to investigate the neural mechanisms of speech perception (I, II, IV) and selective attention (II and III) in fluent (I-IV) and dyslexic (IV) readers.

**Methods:** We conducted four functional magnetic resonance imaging experiments, during which subjects perceived articulatory gestures, speech sounds, and other auditory and visual stimuli. Gradient echo-planar images depicting blood oxygenation level-dependent contrast were acquired during stimulus presentation to indirectly measure brain hemodynamic activation. These functional data were analyzed by a general linear model-based approach.

**Results:** Lip-reading elicited hemodynamic activity within the primary auditory cortex (I), and selective attention to visual speech gestures enhanced activity within the left secondary auditory cortex (II). Attention to non-speech sounds enhanced auditory cortex activity bilaterally; this effect showed modulation by sound presentation rate (III). A comparison between fluent and dyslexic readers' cerebral hemodynamic activity during audiovisual speech perception revealed stronger activation of predominantly motor speech areas in dyslexic readers during a contrast test that allowed exploration of the processing of phonetic features extracted from auditory and visual speech (IV).

**Conclusions:** The results show that visual speech perception modulates hemodynamic activity within auditory cortex areas once considered unimodal (I and II), and suggest that the left secondary auditory cortex specifically participates in extracting the linguistic content of seen articulatory gestures (II). They are strong evidence for the importance of attention as a modulator of auditory cortex function during both sound processing (III) and visual speech perception (II), and point out the nature of attention as an interactive process (influenced by stimulus-driven effects; III). Further, they suggest heightened reliance on motor-articulatory and visual speech perception strategies among dyslexic readers, possibly compensating for their auditory speech perception difficulties (IV).

**Key words:** speech perception, visual, audiovisual, attention, auditory cortex, Broca's area, mirror neuron system, dyslexia, functional magnetic resonance imaging



# Introduction

Speech is a sophisticated communication tool unique to human beings. Based on circumstantial archaeological evidence, it is likely that the Neanderthal people were already using some form of oral communication to exchange information involving abstractions approximately 100 000 years ago. Today, people around the world express their thoughts, feelings, and intentions in more than 6000 different languages. Knowledge concerning the neural mechanisms that underlie human speech perception has greatly expanded during the last several decades, following the advent of methods that enable studying the intact human brain at work.

Speech is primarily an auditory message: understood with eyes closed. Visual information in speech perception is, however, far from irrelevant (Sumbly and Pollack 1954, McGurk and MacDonald 1976). Further, visual speech (i.e., the speaker's articulatory gestures) alone can carry the necessary information, as occurs during lip-reading. Accordingly, research on the neural basis of speech perception has revealed a complex interplay of visual and auditory sensory systems during a face-to-face conversation. The central concepts of this interplay include multimodal convergence and audiovisual integration. Multimodal convergence refers to the access of inputs from different sensory modalities to the same neuronal populations. Such convergence is a prerequisite for audiovisual integration, a process that is thought to combine the auditory and visual inputs into a unified percept (Stein BE and Meredith 1993).

Recent research has suggested that, in addition to visual and auditory sensory processing, speech perception may employ neural circuits that primarily serve motor planning and execution of speech (anatomically, Broca's area and the adjacent ventral premotor cortex). Such views originally date back to the introduction of radical "motor" speech theories that suggest intended articulatory gestures (meaning the neuromotor commands to lips, tongue, and vocal tract) to mediate speech perception (Lieberman et al 1967, Lieberman and Mattingly 1985). Although motor speech theories have received substantial criticism (see Ohala 1996), the recent discovery of mirror neurons has revived interest in the basic ideas of these theories. Mirror neurons possess properties that might offer a link

between auditory and visual speech inputs and the corresponding motor representations in the brain (see Rizzolatti and Arbib 1998, Fadiga and Craighero 2003, Rizzolatti and Craighero 2004).

The brain cannot fully process all information available at a given moment. Instead, we continuously depend on selective attention to resolve the competition between simultaneous stimuli: We see and hear what we consider important or, on the other hand, what is sufficiently salient to draw our attention involuntarily (Kastner and Ungerleider 2000, Nobre 2001). How the central nervous system selects the stimulus events that enter consciousness and to which stage it processes those that do not are questions under intensive research. This research has demonstrated that selective attention continuously shapes auditory and visual sensory functions by modulating neural sensory responses at multiple processing levels (Kastner and Ungerleider 2000, Nobre 2001). The effects of attention on auditory speech perception are far better known than those on visual or audiovisual speech perception.

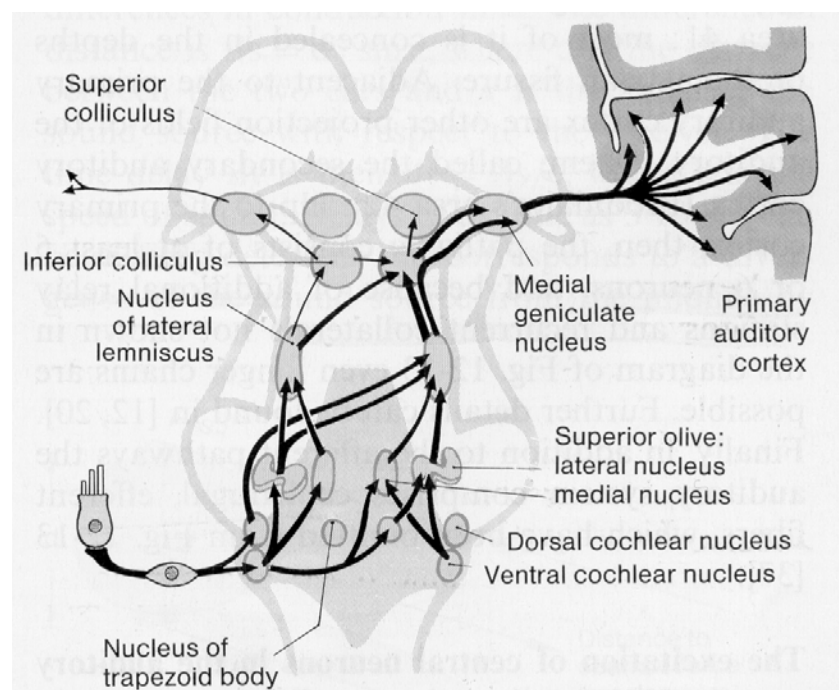
Based on all this, several parallel neural processes (auditory and visual sensory processing; motor system involvement; attention) are likely to operate during speech perception. Hypothetically, the relative weighing of auditory and visual sensory processing and motor system involvement may differ between individuals according to their individual characteristics. One candidate group that might prominently rely on motor-articulatory representations during speech perception comprises dyslexic readers who, in addition to their manifest reading difficulties, show subtle auditory and visual speech perception deficits (Godfrey et al 1981, Werker and Tees 1987, Reed 1989, de Gelder and Wroomen 1998, Hayes et al 2003) that possibly stem from disturbances in lower-level sensory processing (Stein J and Walsh 1997, Hari and Renvall, 2001).

# Review of the literature

## Auditory speech perception

### Subcortical auditory pathway

Hair cells in the inner ear cochlea convert the mechanical sound waves that enter the ear into an electrical signal. Certain locations within the cochlea selectively respond to sounds of a certain frequency; this pattern of sound frequency representations is called tonotopical organization. The auditory pathway ascends from the cochlea to the auditory cortex (Figure 1) and includes at least 5 or 6 synapses; even longer neuronal chains may emerge.



**Figure 1.** A simplified diagram of the subcortical auditory pathway. The auditory nerve bifurcates after leaving the cochlea, sending one branch into the ventral cochlear nucleus and the other into the adjacent dorsal cochlear nucleus. From the ventral cochlear nucleus, a ventral tract runs to the ipsi- and contralateral olivary complexes and ascends to the nucleus of the lateral lemniscus. From the dorsal cochlear nucleus, a dorsal tract crosses to the opposite side and ascends further to the nucleus of the lateral lemniscus. From there, the pathway proceeds through the midbrain inferior colliculus and thalamic medial geniculate nucleus to the primary auditory cortex. Illustration reprinted from: Klinke R (1989): Physiology of the sense of equilibrium, hearing and speech. In: Schmidt RF and Thews G (Eds.): Human Physiology, 2<sup>nd</sup> edition, pp. 277-305, Figure 12-13. ©Springer-Verlag Berlin Heidelberg 1983, 1989, with kind permission of Springer Science and Business Media.

The auditory pathway maintains its tonotopical organization from the cochlea up to the primary auditory cortex. At each stage are located neurons that selectively respond to pure tones of a certain frequency. As evidence of sound pattern recognition within very early auditory processing, neurons with more complex response patterns exist, however, even within the subcortical auditory nuclei (see Klinke 1989). Finally, in addition to the afferent connections visualized in Figure 1, the auditory pathway includes efferent fibers that project from the cortex into the subcortical nuclei.

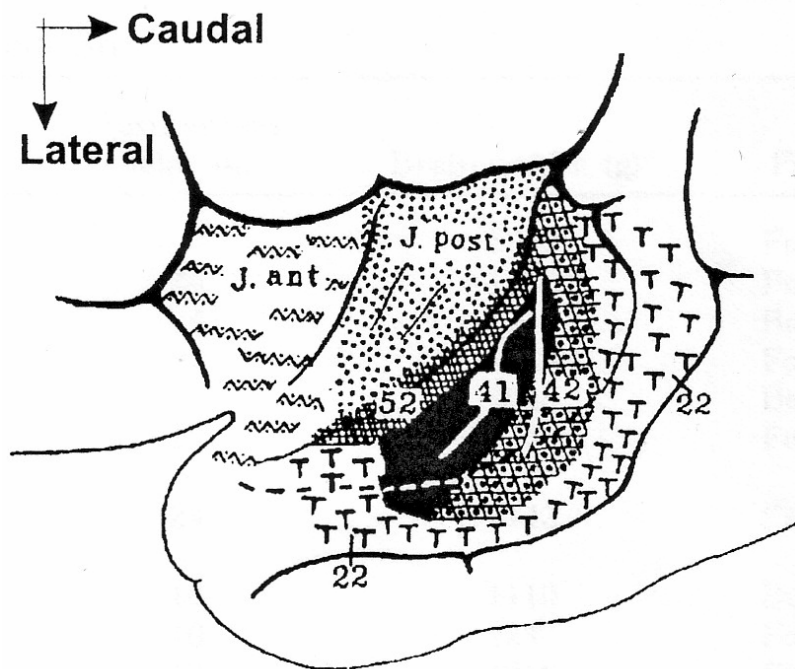
### **Hierarchical organization of the auditory cortex**

The non-human primate auditory cortex consists of three histologically and functionally distinct main regions: core, belt, and parabelt (Merzenich and Brugge 1973, Pandya and Sanides 1973, Burton H and Jones 1976, Morel et al 1993) that diverge into multiple subregions. The core (auditory koniocortex; areas A1 and R) receives its main input from the auditory thalamus (Mesulam and Pandya 1973, Burton H and Jones 1976, Morel et al 1993) and responds to pure tones by displaying a tonotopical organization that is mirror-symmetric between areas A1 and R (Brugge and Merzenich 1973, Merzenich and Brugge 1973, Morel et al 1993). The belt (auditory parakoniocortex) that immediately surrounds the core is less responsive to pure tones (Brugge and Merzenich 1973, Morel et al 1993) but responds vigorously to more complex acoustic stimuli (Rauschecker et al 1995, Tian et al 2001, Rauschecker and Tian 2004, Tian and Rauschecker 2004). Lateral belt neurons also process very complex sounds such as communicational calls for which they can show great selectivity through spectral and temporal integration (Rauschecker et al 1995, Rauschecker and Tian 2000, Tian et al 2001).

Analogous to the functional distinction in the visual sensory system (Ungerleider and Mishkin 1982), subregions of the auditory belt may serve differential functions according to their sensitivity to object identity vs. spatial features of sounds: An object feature-processing “what” stream has been suggested to involve the anterolateral belt, and a “where” stream, sensitive to spatial information, the caudomedial belt (Rauschecker and Tian 2000, Tian et al 2001, Tian and Rauschecker 2004, Rauschecker and Tian 2004). These “what” and “where” processing streams seem to extend to distinct frontal lobe areas across the third division of the auditory cortex, the parabelt, which participates in

complex auditory and associational processing (Seldon 1985, Romanski et al 1999a, 2002).

The human auditory cortex displays an organization analogous to that of non-human primates (Galaburda and Sanides 1980, Hackett et al 2001; see Figure 2). The human primary auditory cortex (PAC, auditory koniocortex) is the histological homologue of the primate auditory core (Galaburda and Sanides 1980, Hackett et al 2001). It resides within the medial two-thirds of the first Heschl's gyrus (HG) on the superior temporal plane. Its extent within HG as well as the HG's location within the superior temporal plane, however, exhibits substantial inter-individual variation (Rademacher et al 1993, 2001, Penhune et al 1996, Morosan et al 2001). PAC demonstrates the functional properties typical of a primary auditory cortex region: It shows short-latency responses to transient acoustic stimuli (Celesia 1976, Liegeois-Chauvel et al 1991) and responds to pure tones and narrow-bandwidth noise by representing the sound information in a tonotopically organized fashion (Howard MA et al 1996, Talavage et al 2000).



**Figure 2.** A view of the human supratemporal plane (Brodmann 1909). The secondary auditory cortex (area 42) surrounds the primary auditory cortex (area 41). Further centrifugally lies the auditory parabelt. The orientation of the primary auditory cortex on the supratemporal plane differs between species; e.g., the auditory core of non-human primates is rotated approximately 45° anteriorly compared to the human PAC. Illustration reprinted from: *NeuroImage* Vol 13, Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K: Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system, pages 684-701, Copyright (2001), with permission from Elsevier.

The human secondary auditory cortex (auditory association cortex, auditory parakoniocortex) is the probable anatomical and functional counterpart of the primate auditory belt (Celesia 1976, Galaburda and Sanides 1980, Liegeois-Chauvel et al 1991, Howard MA et al 2000). It covers the planum temporale (PT) on the superior temporal plane posterior to HG, the lateral-most tip of HG and the closely adjacent cortex within the superior temporal gyrus (STG) (Celesia 1976, Galaburda and Sanides 1980). The posteriormost PT, (area TpT) however, is not a histologically typical sensory cortex; it shares cytoarchitectonic similarities with the frontal Broca's area and parietal association cortex (Galaburda and Sanides 1980, Galaburda 1982). Analogously to the primate auditory belt, the human secondary auditory cortex responds more vigorously to complex acoustic stimuli than to pure tones (Creutzfeldt et al 1989, Demonet et al 1992, Zatorre et al 1992, Strainer et al 1997, Wessinger et al 2001, Hall et al 2002).

The auditory cortex thus shows a hierarchical organization in which PAC receives the main input from the auditory thalamus and sends it on via the secondary auditory cortex to higher-order associational areas; the complexity of sound features processed increases from PAC onwards. Until quite recently, PAC was viewed as a mere passive transmitter of auditory information. Animal studies have, however, made it apparent that top-down cognitive factors such as learning and attention modulate neural responses to auditory stimuli down to the level of PAC and to the level of the subcortical auditory nuclei (Fritz et al 2003, for reviews, see Suga and Ma 2003, Weinberger 2004).

### **Speech-specific auditory processing**

Since speech is a communication tool unique to humans, it may be considered likely that specific neural circuits have evolved to process it apart from other sounds. Elucidating the neural substrates for such speech-specific auditory processing is, however, not simple: it requires segregating each speech sound's intelligibility ("speechness") from its general acoustic complexity.

According to the classical observation of Carl Wernicke, left-hemisphere lesions around the posterior end of the sylvian fissure cause deteriorated speech perception, although they may have little effect on general auditory function (Wernicke's aphasia; Wernicke

1874, Seldon 1985). The classical Wernicke's area includes the PT, the posterior STG, and the adjacent temporoparietal junction. Based on these observations, the secondary auditory cortex posterior to HG has been suggested to act as a dedicated language processor. Such an interpretation has gained support from invasive (Creutzfeldt et al 1989) and non-invasive (Demonet et al 1992, Benson RR et al 2001, Belin et al 2002) studies showing that the posterior STG responds more readily to speech than to non-speech sounds. Other studies have, however, suggested that such selectivity is absent or possibly unrelated to the linguistic nature of the stimuli (Binder et al 1996, 2000, Wise et al 2001).

Comparing responses to acoustically *differing* speech and non-speech sounds does not exclude the possibility that suspected speech-specific responses derive from the general acoustic complexity of speech sounds. Studies attempting to carefully equalize the spectrotemporal properties of speech and non-speech stimuli show that largely overlapping brain regions process both types of sounds (Scott et al 2000, Voloumanos et al 2001, Joanisse and Gati 2003, Möttönen et al 2006, Narain et al 2003). In these studies, responses specific to intelligible speech have emerged within the left superior temporal sulcus (STS) and anterior temporal lobe (Scott et al 2000, Voloumanos et al 2001, Narain et al 2003, Möttönen et al 2006); the posterior secondary auditory cortex may process rapid spectral transitions not necessarily specific to speech (Jäncke et al 2002). Such recent functional data suggest that, instead of the classical temporoparietal Wernicke's area, temporal lobe regions anterior to HG might specifically participate in acoustic feature-based analysis of speech sounds.

## **Visual speech perception**

### **Crossmodal processing in the auditory cortex**

Convergence of auditory and visual inputs into the same neurons or neuronal populations is a necessary prerequisite for combining these inputs into a single perceptual object through audiovisual integration. It does not, however, in itself indicate the occurrence of integrative responses. In animals, neurons that respond to both visual and auditory stimulation have been first and most extensively characterized in the mammalian superior

colliculus (SC), a subcortical structure involved in visual spatial orienting (Stein BE and Meredith 1993, Stein BE et al 2004). Audiovisual-responsive neurons have subsequently emerged within the frontal, parietal, and temporal association cortices, including the STS (for a review, see Cusick 1997) and, recently, also within auditory sensory cortex (Schroeder and Foxe 2002, Brosch et al 2005, Ghazanfar et al 2005).

Sensory cortices can, if deprived of primary input early in development, convert to processing information from non-primary modalities (Levänen et al 1998, Nishimura et al 1999, Petitto et al 2000, Finney et al 2001, MacSweeney et al 2002a: for a review, see Bavelier and Neville 2002). Responses to stimulation via a non-primary modality in the absence of any sensory impairment, however, indicate that such multimodal convergence is one function of a normal brain (e.g. Calvert et al 1997, MacSweeney et al 2000, Campbell et al 2001, Schroeder et al 2001, Bernstein et al 2002, Schroeder and Foxe 2002). Interestingly, recent invasive studies on non-human primates have shown visually-evoked responses within auditory cortex areas once considered strictly unisensory. These responses have emerged in both core (Brosch et al 2005, Ghazanfar et al 2005) and belt (Schroeder and Foxe 2002, Brosch et al 2005, Ghazanfar et al 2005) neurons during observation of stimuli that range from simple light flashes (Schroeder and Foxe 2002, Brosch et al 2005) to the sight of a monkey face uttering communicative calls (Ghazanfar et al 2005).

Most human studies reporting auditory-cortex responses during visual stimulation have investigated visual speech perception (lip-reading). Several studies have established that, of the multisensory regions revealed by animal studies, lip-reading activates the STS and auditory cortex (Calvert et al 1997, Ludman et al 2000, MacSweeney et al 2000, Campbell et al 2001, Bernstein et al 2002, Olson et al 2002, Paulesu et al 2003). Even static images of speech gestures suffice to elicit auditory-cortex electromagnetic (Nishitani and Hari 2002) and hemodynamic (Calvert and Campbell 2003) responses. Thus, auditory cortex activation during visual speech perception is a relatively robust finding. Controversy has, however, persisted as to whether this effect is limited to the higher-order secondary auditory cortex (Bernstein et al 2002, Paulesu et al 2003) or whether it also involves PAC (Calvert et al 1997, MacSweeney et al 2000).

Auditory cortex activation during lip-reading has often been assumed to be specific to visual speech processing. Some unexpected results, however, suggest that time-variant visual signals lacking any linguistic information may also elicit auditory-cortex hemodynamic responses (Howard RJ et al 1996, MacSweeney et al 2004). Specifically, the hearing control subjects in an functional magnetic resonance imaging (fMRI) study on sign-language reading showed auditory cortex activation while watching signs and other manuo-brachial gestures regardless of whether they understood the signs' and gestures' meaning (MacSweeney et al 2004); another study reported auditory cortex hemodynamic responses during observation of non-gestural visual motion displays (Howard RJ et al 1996). Studies directly comparing auditory-cortex hemodynamic responses during observation of speech and non-speech gestures, however, suggest a degree of specialization of auditory-cortex responses to visual speech perception (Calvert et al 1997, Campbell et al 2001).

Learned associations between speech sounds and articulatory gestures form during early infancy; they probably play some role in eliciting visual speech-induced auditory cortex responses (see Bernstein et al 2004). Accordingly, the lack of auditory cortex hemodynamic activation during lip-reading in the congenitally deaf (MacSweeney et al 2002b) suggests that this effect may require exposure to speech sounds. Further, an experiment on visually observed piano playing by trained musicians and musically naïve subjects suggests that visual codes may evoke auditory-cortex responses through a learned association (Hasegawa et al 2004). The observed specificity of auditory-cortex responses to speech gestures (Calvert et al 1997, Campbell et al 2001) could thus draw from the close link between auditory and visual speech components.

### **Feedback or feedforward modulation?**

Several options exist for neural mechanisms through which visual events may elicit auditory-cortex responses: Such responses may represent a feedforward effect (arriving, for instance, via subcortical structures) or a feedback effect (via cortico-vortical connections from higher-level regions). Heavy neuronal connectivity within the central nervous system complicates the search for a route through which visual information is able to “access” the auditory cortex; virtually any two sensory cortical areas link to each

other via several alternative pathways (Rauschecker et al 1997, Felleman and Van Essen 1991, Burton H and Sinclair 1996).

Differentiating between feedback and feedforward inputs is, however, possible on the basis of their timing and laminar response profiles (Rockland and Pandya 1979, Felleman and Van Essen 1991). Visually evoked activity in the monkey auditory belt neurons has shown characteristics of a feedback response (Schroeder and Foxe 2002). In terms of the currently well documented neuronal connections, the possible sources for such feedback include the STS (Pandya et al 1969, Seltzer and Pandya 1994, Hackett et al 1998), the intraparietal sulcus (Lewis and Van Essen 2000) and the pre-frontal cortex (Romanski et al 1999a, b). Suggesting the possibility that visual and auditory cortices can also directly exchange information, two laboratories have found axonal projections from the auditory cortex to the visual striate cortex (Rockland and Ojima 2001, Falchier et al 2002); such connections have not yet, however, emerged in the reciprocal direction. When extending the interpretation of the above animal data (Schroeder and Foxe 2002) one must, however, appreciate that the same mechanisms may not fully apply to human speech perception.

## **Audiovisual speech perception**

### **Audiovisual integration and interactions**

Convergence of information from different sensory modalities to the same neuronal populations enables multisensory integration (MSI). MSI binds the inputs into a unified percept (rather than a combination of unrelated sensations); this facilitates rapid orienting and stimulus detection (Hershenson 1962, Morrell 1968, Welch and Warren 1986, Stein BE and Meredith 1993, Stein BE et al 2004).

Neurons that contribute to the MSI process have been most extensively studied in the mammalian SC (Stein BE and Meredith 1993, Stein BE et al 2004). Some of these MSI neurons fire multiplicatively during bimodal audiovisual stimulation compared to the sum of unimodal auditory and visual responses, provided that the inputs coincide in time and space. In addition, MSI neurons show response depression if the inputs occur in temporal

or spatial disparity. Their response enhancement by bimodal stimulation is greatest when the unimodal stimuli are least effective (e.g., degraded by noise; the rule of inverse effectiveness). Audiovisual-responsive neurons within the auditory cortex (Ghazanfar et al 2005) and STS (Barraclough et al 2005) have recently demonstrated similar integrative properties, although their bimodal response enhancement has represented more a deviation from the assumption of linear summation of unimodal signals rather than a true multiplicative response.

Human studies have in part adopted these MSI cell response properties to investigate audiovisual interactions as evidence of audiovisual integration. These interactions basically mean differing responses during bimodal stimulus presentation from the sum of unimodal responses; they can be either supra- or sub-additive and do not usually indicate multiplicative responses. The auditory mismatch negativity response (MMN in electroencephalographic studies; or its magnetoencephalographic counterpart MMF) provides an additional tool for audiovisual integration studies in humans. It is thought to reflect an automatic change-detection system within the auditory cortex that generates a neural response when a new sound violates the memory trace created by preceding sounds in a constant stimulus stream (for review, see Näätänen 1982, 2001; see also Jääskeläinen et al 2004a).

### **Audiovisual interactions in speech perception**

Behavioral effects attributable to audiovisual integration are very prominent in speech perception: for example, seeing the facial movements of a speaker facilitates speech perception in noisy conditions to the same extent as increasing the signal-to-noise ratio of auditory speech by 15-20 dB (Sumbly and Pollack 1954). Further, visual speech can even change the auditory percept in the case of simultaneous conflicting visual and auditory speech inputs (the McGurk illusion; McGurk and MacDonald 1976). How the central nervous system combines the visual and auditory speech inputs is a central question for speech perception research. Theories concerning it diverge into early and late integration models, according to the processing level at which the inputs are assumed to merge (before or after phonetic categorization, thought to occur within 150-200 ms from stimulus onset; Rinne et al 1999, Phillips et al 2000, Vihla et al 2000).

The multisensory STS is centrally involved in audiovisual speech integration. Well-documented audiovisual interactions during speech processing have emerged there in several electrophysiological (Callan et al 2001, 2003a, 2004a, Klucharev et al 2003, Möttönen et al 2004) and hemodynamic studies (Calvert et al 2000, Sekiyama et al 2003, Wright et al 2003, Macaluso et al 2004); comparable response amplifications during audiovisual speech perception have also emerged within the auditory sensory cortex (Calvert et al 1999). Integration of auditory and visual speech inputs has been suggested to occur primarily within STS, and this would then exert feedback modulation on auditory cortex function (Calvert 2001). Such hypothesis is plausible in the light of the currently known anatomical connections between auditory and visual cortices and STS (Pandya et al 1969, Hackett et al 1998) and the feedback nature of visual input to the auditory cortex in non-human primates (Schroeder and Foxe 2002).

The temporal scale of audiovisual interactions during speech perception makes, however, pure feedback-modulation of auditory cortex responses unlikely. Demonstrating audiovisual interactions at temporally early stages of speech processing (presumably within the auditory sensory cortex), a visual-only change in a stream of audiovisual speech stimuli evokes auditory-cortex mismatch responses (Sams et al 1991, Möttönen et al 2002, Colin et al 2002, 2004). Further, visual speech modulates auditory responses to speech sounds within 100 to 150 ms from sound onset (Klucharev et al 2003, Besle et al 2004, Möttönen et al 2004, Jääskeläinen et al 2004b, van Wassenhove et al 2005). Importantly, interactions within the auditory cortex may precede those within the STS (Möttönen et al 2004). Additionally, a recent report shows visual speech gestures as modulating subcortical auditory responses to speech sounds (Musacchia et al 2006). These observations support early integration models by showing audiovisual interactions prior to the temporal window of phonetic categorization.

Audiovisual interactions within the STS and auditory cortex are not confined to speech processing: Similar effects emerge during audiovisual perception of various other stimuli (Giard and Peronnet 1999, Pourtois et al 2000, Raji et al 2000, Calvert et al 2001, Fort et al 2002a, b, Molholm et al 2002, Lebib et al 2003). Insensitivity of the earliest cortical interactions to the inputs' phonetic congruence during speech perception (Klucharev et al 2003) suggests that they may represent integration of non-linguistic stimulus features

(related to the temporal and spatial co-occurrence of auditory and visual stimuli). Taken together, the current functional data thus indicate that visual speech gestures affect auditory speech processing at multiple stages via several feedback and feedforward mechanisms that may differ with respect to their specificity to the linguistic nature of the input.

## **Attention and speech perception**

### **Behavioral and theoretical aspects**

The world around us would sound and seem chaotic if we lacked the ability to select high-priority stimuli that enter extended processing and to ignore all others. To resolve the competition between simultaneous stimuli, we continuously depend on selective attention. It enables, for example, concentration on a friend's voice in a noisy crowd while effectively tuning out other conversations. Still, hearing someone mention one's own name can suddenly capture attention (cocktail-party phenomenon). How attention facilitates efficient information processing and suppresses irrelevant stimuli is a question under intensive research. Initial dichotic listening studies suggest that unattended auditory information is filtered out at a precognitive sensory processing stage (early selection; Broadbent 1958); later studies challenged this notion and suggested that attention operates within the constraints of short-term memory (late selection; Treisman 1964). Studies on the neurophysiological mechanisms of attention have shown that selective attention to acoustic and visual stimuli modulates the sensory cortices' function, enhancing the responses to attended stimuli (reviewed in Kastner and Ungerleider 2000, Nobre 2001). The executive system driving this modulation has been best characterized for visual spatial attention, but it is currently unclear to what extent that information applies to the auditory domain (see Nobre 2001).

### **Attentional modulation of sound-evoked responses**

Attention modulates the human auditory-cortex responses to sounds: spatial attention to sounds enhances electrical and electromagnetic neural responses originating in the contralateral auditory cortex (Hillyard et al 1973, Hari et al 1989, Rif et al 1991, Woldorff

and Hillyard 1991, Woldorff et al 1993). Moreover, auditory cortex hemodynamic activity increases when subjects attend to auditory processing in widely differing experimental settings (O’Leary et al 1996, Pugh et al 1996, Grady et al 1997, Alho et al 1999, Jäncke et al 1999, 2001, Hashimoto et al 2000, Lipschutz et al 2002, Degerman et al 2006). This is, however, not true for all studies (Zatorre et al 1999, Frith and Friston 1996). It is possible that attentional modulation of auditory cortex hemodynamic activity requires sufficiently demanding conditions such as rapid stimulus presentation rates or difficult listening tasks (for discussion, see Zatorre et al 1999).

The exact processing stage at which attention modulates human auditory-cortex responses to sounds has remained somewhat controversial. This question is relevant in the context of early vs. late selection models of attention and, additionally, in the general context of top-down modulation of auditory cortex function (this including feedback-mediated audiovisual interactions). Attentional modulation of auditory-cortex responses to sounds at post-stimulus latencies as early as 10 to 50 ms (Woldorff and Hillyard 1991, Woldorff et al 1993) suggests attentional effects during very early cortical processing and lends support to early selection models of attention. Accordingly, some hemodynamic studies utilizing group-level statistics have reported attention effects within the PAC region (Jäncke et al 2001, Alho et al 1999) and – offering more precise spatial resolution – two fMRI studies at single-subject level have suggested that attentional modulation affects PAC function (Jäncke et al 1999, Hashimoto et al 2000). Other studies, however, have shown attention effects only in the secondary auditory cortex (Grady et al 1997, Pugh et al 1996, Lipschutz et al 2002), and a recent fMRI study utilizing high-resolution surface-mapping techniques has shown attentional modulation to be limited to a lateral auditory cortex area with functional characteristics of the secondary auditory cortex (Petkov et al 2004).

In addition to active attentional modulation (a top-down cognitive effect), auditory-cortex activity during acoustic processing is subject to stimulus-driven (bottom-up) effects. Exemplifying such stimulus-driven effect, increasing the sound presentation rate strengthens auditory cortex hemodynamic activity irrespective of whether the subjects attend to the sounds; this probably occurs due to accumulation of responses to individual sounds (the rate-effect; Binder et al 1994, Frith and Friston 1996, Dhankhar et al 1997,

Rees et al 1997, Mechelli et al 2000, Tanaka et al 2000, Harms and Melcher 2002). Studying whether attentional and stimulus-driven factors modulate auditory cortex activity independently or whether they interact with each other might provide additional information concerning the neural mechanisms through which attention modulates auditory sensory function. Such questions have not, however, been studied on a wide scale.

### **Attentional modulation of visual and audiovisual speech perception**

Integration of auditory and visual speech inputs was earlier considered an “automatic,” pre-attentive process (McGurk and MacDonald 1976, Bernstein et al 2004). Two recent studies on the classical McGurk illusion, however, question this view (Tiippana et al 2004, Alsius et al 2005). These studies report that the degree to which visual articulatory gestures contribute to the audiovisual speech percept depends on whether or not the subjects attend to the articulatory gestures (Tiippana et al 2004), and that the subjects’ susceptibility to the McGurk illusion decreases if they concurrently perform an unrelated visual or auditory task (Alsius et al 2005). Thus, at least in the case of a demanding attentional load, audiovisual speech integration seems subject to attentional modulation (see Alsius et al 2005).

These observations predict that neurophysiological effects related to audiovisual integration may also depend on attention. Indeed, in a recent evoked response potential (ERP) study on audiovisual perception of non-speech stimuli, audiovisual interactions within higher-order heteromodal areas emerged only if the subjects attended to the bimodal stimuli; interactions in the sensory cortices, on the other hand, were independent of attention (Fort et al 2002c).

Little is known as to how attention may modulate auditory-cortex responses evoked by visual speech gestures occurring without concurrent speech sounds. Animal studies suggest that the behavioral relevance to an animal of some visual stimulus may determine whether the stimulus elicits an auditory cortex response (Brosch et al 2005, Ghazanfar et al 2005); such discrimination may happen via attentional facilitation. Additionally, two MMN/MMF studies indirectly imply that attention affects the processing of visual speech

signals in the auditory cortex (Möttönen et al 2002, Colin et al 2004): A change in a stream of attended speech gestures without concurrent speech sounds has elicited auditory-cortex mismatch responses (Möttönen et al 2002), whereas no such responses have occurred in a similar experiment with the subject's attention directed away from the speech gestures (Colin et al 2004). The currently existing hemodynamic studies reporting auditory cortex activity during visual speech perception have, however, not explicitly accounted for the direction of attention.

## **Motor system in speech perception**

### **The motor theory of speech**

Speech perception theories aim at understanding how the brain extracts a meaningful message from the complex flow of speech information (see Diehl et al 2004). These roughly fall into two categories: those assuming that speech perception is similar to the perception of any complex sounds and those considering speech a special stimulus, the perception of which radically differs from general auditory processing. The revised motor theory of speech (Liberman et al 1967, Liberman and Mattingly 1985) represents the latter category. Its central claim holds that the objects of speech perception are the speaker's intended speech gestures. These gestures are the physical reality behind uttering different phonemes (for instance, opening the vocal tract and rounding the lips to produce an /o/) that are represented in the brain as neuromotor commands to the lips, tongue, and vocal tract. According to the motor theory of speech, the listener converts the acoustic signal he/she receives to the corresponding motor representations, which enables phonetic categorization. As speech production and perception would thus share a common substance (the neuromotor commands), this conversion would happen automatically, requiring no learned associations or conscious thought.

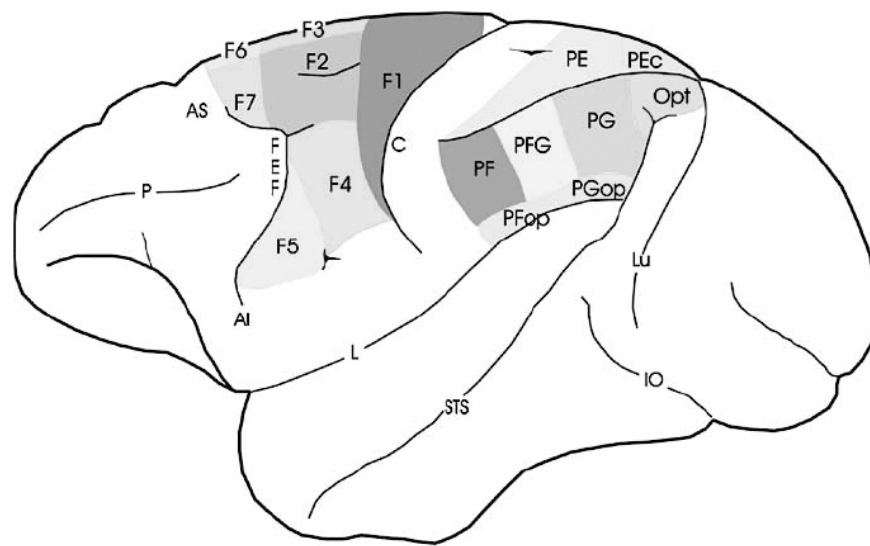
Why such motor-articulatory representations should participate in speech perception may not be intuitively clear. Phonetic categorization of speech sounds, however, includes phenomena not easily explainable by the sounds' acoustic properties alone. For example, a consonant's formant transition can differ depending on the following vowel; it may include a rising transition in one vowel context and a falling transition in another

(Liberman et al 1954). Further, these formant transitions sound different depending on their context: In isolation, the rising and falling transitions sound like glissandi or chirps that are dissimilar (Mattingly et al 1971). When embedded in speech, however, they produce an identical percept of the same consonant. The acoustic-based theories of speech perception (see Diehl et al 2004) assume that the auditory processing system manages such discrepancies. The motor speech theories, on the other hand, suggest that they are overcome by use of invariant speech gesture information to mediate phonetic categorization (Liberman et al 1967, Liberman and Mattingly 1985). Any possible neural substrates for speech perception through motor actions had, however, remained obscure until the recent discovery of mirror neurons (see Rizzolatti and Arbib 1998, Fadiga and Craighero 2003, Rizzolatti and Craighero 2004).

### **Mirror neurons**

Mirror neurons have motor properties but are also active during visual observation of actions (see Rizzolatti and Arbib 1998, Fadiga and Craighero 2003, Rizzolatti and Craighero 2004). In the monkey brain, they reside within the premotor cortex (area F5) and the rostral inferior parietal lobule (IPL) (di Pellegrino et al 1992, Gallese et al 1996, Rizzolatti et al 1996a, Fogassi et al 2005); these two areas form the core of the mirror neuron system (MNS; see Figure 3). Mirror neurons discharge both when a monkey performs an action and when it watches another monkey or a person perform the same or a similar action.

Although most mirror neurons are selective to goal-directed hand actions (di Pellegrino et al 1992, Gallese et al 1996, Rizzolatti et al 1996a), some selectively respond to communicative mouth actions (Ferrari et al 2003). Observations of mirror neurons that respond not only to the sight of an action but also to its sound (Kohler et al 2002) and ones that discharge upon observation of indirect clues of an action (Umiltà et al 2001) suggest their ability to code an action with a meaning. Mirror neurons are suggested to transform observed actions into knowledge: Seeing or hearing an action may awaken the corresponding cortical motor representation, the outcome of which is known to the observer of that action (Rizzolatti et al 2001, Rizzolatti and Craighero 2004).



**Figure 3.** Lateral view of the monkey brain. Mirror neurons reside within the inferior frontal cortex (area F5) and rostral inferior parietal lobule (area PF). The STS hosts neurons that respond to seen actions but lack the motor properties of mirror neurons (Perrett et al 1989, Jellema et al 2000). F5 and PF are directly interconnected and both receive input from the STS (Arbib and Bota 2003). The mirror neuron system thus includes two main regions (F5 and PF), to which the STS is closely connected. The probable human homologue for the monkey F5 is the pars opercularis of the inferior frontal gyrus (Petrides and Pandya 1994). This region in the left hemisphere belongs to the classical Broca's area (Broca 1861), traditionally considered as a motor speech area without perceptive functions. Illustration reprinted, with permission, from the Annual Review of Neuroscience, Volume 27 © 2004 by Annual Reviews, [www.annualreviews.org](http://www.annualreviews.org) .

### **Evidence of MNS-like system in humans**

Evidence from electromagnetic and transcranial magnetic stimulation (TMS) studies strongly supports the existence of an MNS-like neural circuit that in the human brain binds together observed actions and their motor representations (Cochin et al 1998, Fadiga et al 1995, Hari et al 1998, Strafella and Paus 2000, Baldissera et al 2001). Accordingly, several functional imaging studies show activation of regions homologous to the monkey MNS when human subjects watch actions (Grafton et al 1996, Rizzolatti et al 1996b, Iacoboni et al 1999, Nishitani and Hari 2000, 2002, Buccino et al 2001, Grèzes and Decety 2001, Perani et al 2001, Decety et al 2002, Koski et al 2002, 2003, Grèzes et al 2003, Manthey et al 2003, Wilson et al 2004) – and also when they imitate such actions (Iacoboni et al 1999, Nishitani and Hari 2000, 2002, Koski et al 2002, Grèzes et al 2003). This neural circuit could thus participate in visual speech perception and is, indeed, active, according to several studies, during lip-reading (Campbell et al 2001, Nishitani and Hari 2002, Calvert and Campbell 2003, Paulesu et al 2003). Compared to the monkey

MNS, the human MNS-like system seems to have properties that enable a wider role in interpreting and understanding actions (Fadiga et al 1995, Gangitano et al 2001, Maeda et al 2002, Patuzzo et al 2003).

Language evolution has been suggested to have proceeded through imitation of hand and mouth actions, associating them with sounds and, finally, uttering the sounds without performing the actions (see Rizzolatti and Arbib 1998, Fadiga and Craighero 2003, Rizzolatti and Craighero 2004). During this process, an auditory-motor interface within the MNS may have evolved to connect the communicative sounds with the motor representations needed to utter them (see Kohler et al 2002 for evidence of auditory-visuo-motor mirror neurons). An action would thus have become a message, as postulated by the motor speech theorists (Liberman et al 1967, Liberman and Mattingly 1985). Such evolutionary history might account for the mosaic-like organization of Broca's area where phonology, syntax, and semantics are co-represented by seemingly unrelated mouth and hand actions (Bookheimer 2002, Nishitani et al 2005).

Such ideas of "resonance" between auditory and motor neural representations during speech perception have recently gained some experimental support. TMS studies have reported that speech listening enhances the tongue and mouth muscle responses evoked by stimulation of the face motor cortex (Fadiga et al 2002, Watkins et al 2003, Watkins and Paus 2004). The magnitude of this effect corresponds to how much tongue action the heard word requires (Fadiga et al 2002). Moreover, listening to speech has elicited hemodynamic responses in the classical motor speech areas of the frontal cortex (Zatorre et al 1992, 1996, Price et al 1996, Burton MW et al 2000, Wilson et al 2004).

These findings do not suggest that speech may be perceived only or predominantly via the MNS or by utilizing motor representations. They do, however, indicate that action perception and production are closely linked and that brain areas essential to speech production also participate in speech perception. The MNS may offer a neural circuit able to process both seen articulatory gestures (via action observation) and heard speech sounds (via an association between a speech sound and its articulatory counterpart). Parallel processing streams operating on different classes of speech representations may facilitate the use of both acoustic and articulatory-gestural information during speech

processing (see Scott and Johnsrude 2003, Hickok and Poeppel 2004). Hypothetically, the use of these streams may depend on the information available and the individual characteristics of the perceiver.

## **Dyslexia and speech perception**

### **Language and sensory processing deficits**

Dyslexia is the name for a specific difficulty in learning to read and write despite normal intelligence and schooling. Although dyslexia usually manifests as difficulties in reading acquisition, it fundamentally concerns language processing at a more elemental level; its core cognitive defect involves phonological processing, meaning the use of a language's sound structure to process read or heard information (Snowling 1981, Frith 2001, Paulesu et al 2001). Dyslexic readers also show phonetic-level speech perception difficulties: Their speech sound categorizations are less sharp than those of fluent readers, as is evident in tasks that require distinguishing subtle differences between acoustically similar phonemes (Godfrey et al 1981, Werker and Tees 1987, Reed 1989).

The causes underlying dyslexic readers' speech perception and phonological processing difficulties are currently unclear. One possibility is an elemental perceptual deficit which can complicate speech sound discrimination and consequently lead to problems in phonological processing (Manis et al 1997). In accordance with such ideas, minor perceptual deficits do occur in dyslexia outside linguistic processing; they primarily involve processing of information with rapid time-variant transitions or temporally rapidly successive stimuli, or both. Such findings have emerged in auditory (Tallal and Piercy 1973, Tallal et al 1985, Hari and Kiesilä 1996, Hari and Renvall, 2001), visual (Lovegrove et al 1980, Martin and Lovegrove 1987, Eden et al 1996, Cornelissen et al 1998), and somatosensory modalities (Laasonen et al 2000, 2002). It has thus been suggested that the dyslexic readers' central nervous system is generally deficient in the processing of rapidly time-variant information: this problem may extend across sensory modalities and outside the domain of linguistic processing (Tallal et al 1985, Hari and Kiesilä 1996; for a review, see Stein J and Walsh 1997). Dyslexic readers, however, also demonstrate speech perception difficulties not readily explainable by the temporal

characteristics of the sounds; these observations suggest that speech-specific phonetic-level processing difficulties may co-exist with perceptual difficulties on a more elemental level (Mody et al 1997, McAnally et al 1997, Nittrouer et al 1999).

### **Functional findings**

Several brain research experiments have aimed at specifying the functional pathology underlying the linguistic and perceptual deficits in dyslexia. The variety of methods, study populations, and analysis strategies makes it difficult to draw a consistent picture of the results. Dyslexic readers have, however, demonstrated some characteristic responses during both linguistic processing and early-level auditory and visual perception (for reviews, see Habib 2000 and Salmelin and Helenius 2004).

During reading and phonological processing tasks, dyslexic readers in several studies have shown diminished left temporoparietal neural responses (Rumsey et al 1997, Shaywitz et al 1998, Simos et al 2000a, b, Paulesu et al 2001): This has even been suggested to represent the fundamental functional deficit in dyslexia (Temple 2002). The area with less activation in dyslexic than in fluent readers extends from the posterior STG and angular gyrus to the base of the temporal lobe, thus including parts of the classical Wernicke's area and the occipitotemporal cortex involved in visual feature analysis (for reviews, see Habib 2000 and Salmelin and Helenius 2004). Dyslexic readers have also shown loss of normal left-lateralization of temporoparietal activation during linguistic processing (Simos et al 2000a, b), and evidence of connection deficits between frontal and temporal speech processing areas (Paulesu et al, 1996, Horwitz et al 1998, Klingberg et al, 2000).

Supporting the role of auditory sensory disturbances in dyslexia, dyslexic readers show abnormal auditory-cortex responses to sounds (Kraus et al 1996, Schulte-Körne et al 1998, Nagarajan et al 1999, Kujala et al 2000, Helenius et al, 2002a, b, Renvall and Hari 2002). Studies utilizing the auditory MMN suggest that their speech sound discrimination deficits originate during early-level auditory processing (Kraus et al 1996, Schulte-Körne et al 1998). Accordingly, abnormal responses during speech listening have emerged in the dyslexic reader's auditory cortex within 100 ms from stimulus onset (Helenius et al

2002a, b; for partly conflicting results see, however, Simos et al 2000b). Such findings are not confined to speech perception: the key factor may instead be the temporal structure embedded in the acoustic stimuli (Nagarajan et al 1999, Kujala et al 2000, Renvall and Hari 2002). Conversely, positron emission tomography (PET) and fMRI studies on auditory processing in dyslexia have not produced consistent results on possible hemodynamic response abnormalities in the auditory cortex (McCrory et al 2000, Temple et al 2000, Corina et al 2001).

Functional abnormalities have also emerged in dyslexic readers' visual perception (Eden et al 1996, Salmelin et al 1996, Demb et al 1997, Helenius et al 1999), in line with behavioral observations of subtle visual sensory impairments (Lovegrove et al 1980, Martin and Lovegrove 1987, Eden et al 1996, Cornelissen et al 1998). Magnetoencephalographic studies on single-word reading have suggested a processing deficit at the level where letter-string specific responses first emerge (Salmelin et al 1996, Helenius et al 1999). As evidence of more elemental visual processing disturbances, dyslexic readers have, however, also shown diminished hemodynamic responses to moving visual displays in the primary (Demb et al 1997) and motion-specific (Eden et al 1996, Demb et al 1997) visual cortices. Further, suggesting a link between visual processing abnormalities and reading difficulties, the visual cortex activity during perception of non-linguistic moving visual stimuli has correlated with reading ability (Demb et al 1997).

Functional and behavioral data thus agree as to the fact that dyslexic readers experience subtle sensory deficits that extend beyond linguistic processing. Hypothetically, these deficits may affect how dyslexic readers weigh their use of auditory, visual, and motor-articulatory processing during speech perception.

## **BOLD fMRI**

Changes in cerebral blood flow (CBF) and blood oxygenation within the brain microvasculature (collectively known as hemodynamics) closely relate to neural activity; this allows the drawing of conclusions about brain function from study of its hemodynamic responses.

BOLD fMRI, a novel technique developed in the early 1990s, indirectly measures brain hemodynamic responses during functional activation (Ogawa et al 1990). The method is based on the fact that the magnetic properties of blood hemoglobin depend on its state of oxygenation, and that during functional activation the relative concentrations of oxygenated and deoxygenated hemoglobin within the brain microvasculature undergo change. The precise relationship between neural activity and BOLD response is the focus of active research, and many of its details are currently unclear. The following overview of BOLD fMRI will thus present only a general picture of the method. Similarly, the section on data analysis will concentrate on the strategies most frequently applied (that were also utilized in this study).

### **Principles of the MR signal**

Each hydrogen nucleus ( $H^+$ ) has a spin (a precessing magnetic moment with direction and phase). Magnetic resonance imaging (MRI) utilizes the spins of water-molecule hydrogen nuclei and the high water content of the human body to obtain images that depict body tissues with characteristic signal intensities (reflected as shades of gray in an MR image; for a description of the basis of MRI, see Haacke et al 1999).

For MRI, the subject first enters a strong external magnetic field ( $B_0$ ). This causes a small majority of the spins in the body to align in the direction of  $B_0$  (the spins' low-energy state), creating a longitudinal net magnetization vector. A short-duration radiofrequency (RF) pulse then forces this vector away from its original direction and makes the spins adapt a coherent precession phase. Consequently, the longitudinal net magnetization vector shortens, and a transverse magnetization vector emerges. After the cessation of the RF pulse, the spins spontaneously begin to return to their low-energy state and lose their phase coherence. During this relaxation process, the longitudinal net magnetization vector returns to its original state (T1 relaxation), and the transverse magnetization vector disappears (T2 relaxation).

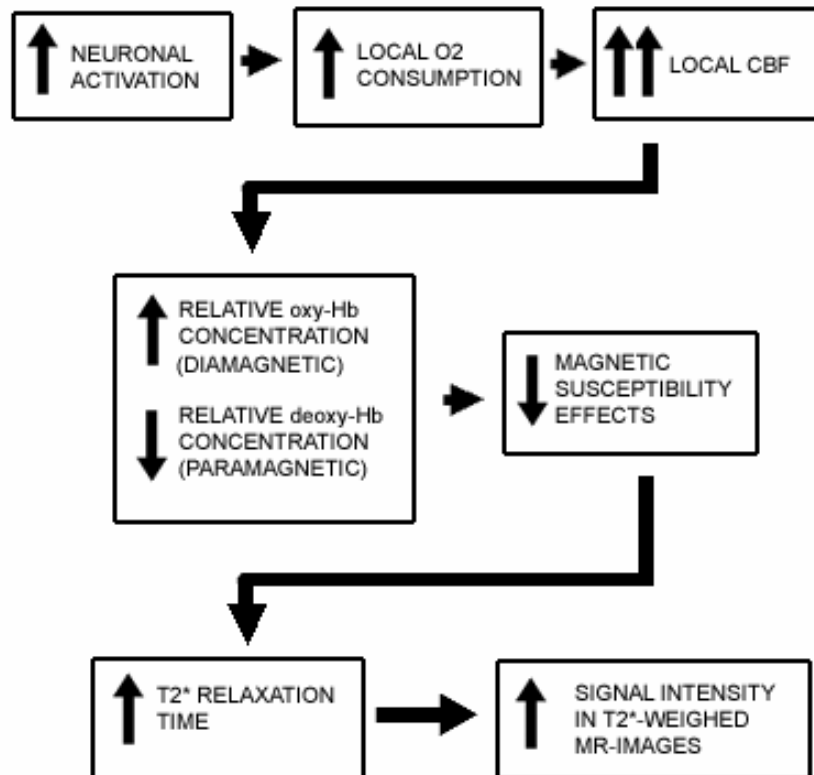
The spins emit RF energy during their relaxation; this energy constitutes the MR signal. When recorded, it characterizes the longitudinal and transverse magnetization states of a spin population at a given time. Because longitudinal (T1) relaxation is more rapid than

transverse (T2) relaxation, the time-point of signal collection can be determined to sample predominantly either the spins' longitudinal (T1-weighted imaging) or transverse magnetization state (T2-weighted imaging). The T1 and T2 relaxation times of the water spins differ between tissues, depending on the chemical surroundings of the water molecules; this forms the basis of tissue contrast in MRI. The total speed of T2 relaxation depends on two factors: the characteristic T2 relaxation time of the tissue and the local field inhomogeneity (T2\*) effects. BOLD fMRI utilizes these local T2\* effects to measure indirectly blood flow responses within the brain during functional activation.

### **Basis of BOLD signal changes**

The source of contrast in BOLD fMRI is the relative deoxyhemoglobin vs. oxyhemoglobin concentration within the brain vasculature (Ogawa et al 1990, Chen and Ogawa 1999). Hemoglobin, the oxygen-transporting protein in the blood, carries four iron-containing hemes, each of which can bind and release an oxygen molecule (Stryer 1988). The hemoglobin oxygenation state alters the magnetic properties of the heme iron. The iron in oxygenated hemoglobin (oxyhemoglobin) is diamagnetic, whereas the iron in deoxygenated hemoglobin (deoxyhaemoglobin) is paramagnetic. As a paramagnetic material, deoxyhemoglobin has a magnetic moment that polarizes in an external  $B_0$  field; the strength of this effect is termed "magnetic susceptibility." Due to its magnetic susceptibility, deoxyhemoglobin causes local inhomogeneities (T2\* effects) in the  $B_0$  field. These inhomogeneities alter the relaxation properties of the nearby water spins by shortening their T2\* relaxation times.

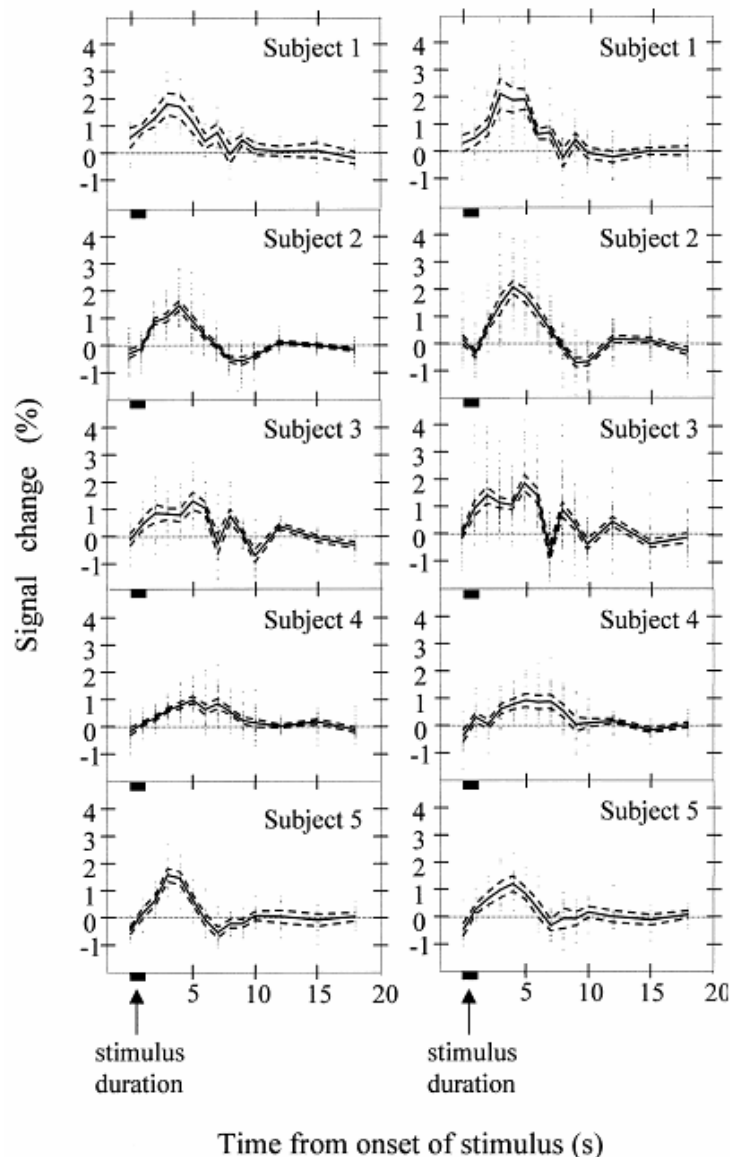
The local supply of and demand for oxygen determines the relative oxyhemoglobin and deoxyhemoglobin contents within a given brain area and thus modifies the water spins' T2\* relaxation time within that area. This allows one to measure hemodynamic responses within the brain by using MRI (see Figure 4).



**Figure 4.** The basis of BOLD signal changes. Increased synaptic activity during neuronal activation induces elevated oxygen consumption. This leads to a local CBF increase that brings an *excess* of oxygenated hemoglobin (oxy-Hb) to the activated area. Such uncoupling of local oxygen consumption and delivery leads to a relative *decrease* in paramagnetic deoxyhemoglobin (deoxy-Hb) concentration. This lengthens the  $T_2^*$  relaxation time of the nearby water spins via decreased local  $B_0$  distortions. The result is locally increased signal intensity within the activated area when the brain is imaged with MRI sequences that sample the spins'  $T_2^*$  relaxation state.

### Temporal relationships between neuronal activity, CBF, and BOLD

The coupling between neuronal activity, CBF, and BOLD response is relatively slow. After a rise in neuronal firing rate, metabolic or neurotransmitter-mediated signals relax arterial sphincters, causing local vessel dilatation and increased CBF to the activated region (see Bandettini 1999). BOLD response to a single stimulus event begins to deviate from the baseline approximately 2 to 3 s after stimulus onset, reaches its peak at about 4 to 7 s, and returns to the baseline at about 10 to 12 s (Blamire et al 1992, Hall et al 2000; see Figure 5). Transient signal intensity decreases (the initial dip and the post-stimulus undershoot; Hu et al 1999, Buxton et al 1999) both precede and follow the BOLD signal increase—these are most readily observed at high  $B_0$  fields.



**Figure 5.** The BOLD response to a sound stimulus in the auditory cortex. Illustration from: Hall DA, Summerfield AQ, Goncalves MS, Foster JR, Palmer AR, Bowtell RW: Time-course of the auditory BOLD response to scanner noise (*Magnetic Resonance in Medicine* 43:601-606). Reprinted with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc. Copyright © 2000 Wiley-Liss, Inc., A Wiley Company. Reproduced with permission of John Wiley & Sons, Inc.

Rapid stimulus repetition leads to superimposition and summing of responses to individual stimuli. During such continuous stimulation, the BOLD signal begins to deviate from the baseline approximately 2 s after the start of a stimulation period and reaches its peak at approximately 6 to 9 s. The time to return to baseline after cessation of the stimulation is longer than the rise time by about 1 to 2 s (DeYoe et al 1994, Bandettini 1999). The time-scale of BOLD signal changes (several seconds) is thus considerably slower than that of neuronal events (on the order of milliseconds).

## Measurement devices and sequences

The MRI scanner contains a supraconducting magnet, an RF coil, or coils and gradient coils. The magnet produces the strong main magnetic field ( $B_0$ ), typically of a magnitude of 1 to 3 T. The RF system produces and detects the MR signal; it transmits the RF excitation pulses and receives the signal emitted by the spins during their relaxation. The gradient coils induce three gradient magnetic fields perpendicular to each other, which enables spatial encoding of the MR signal. The signal is collected in K-space, the center of which represents low spatial frequencies (the image contrast) and the outer parts of which encode high spatial frequencies. Finally, the signal is transformed into a grayscale brain image by a mathematical method called 2D Fourier transform.

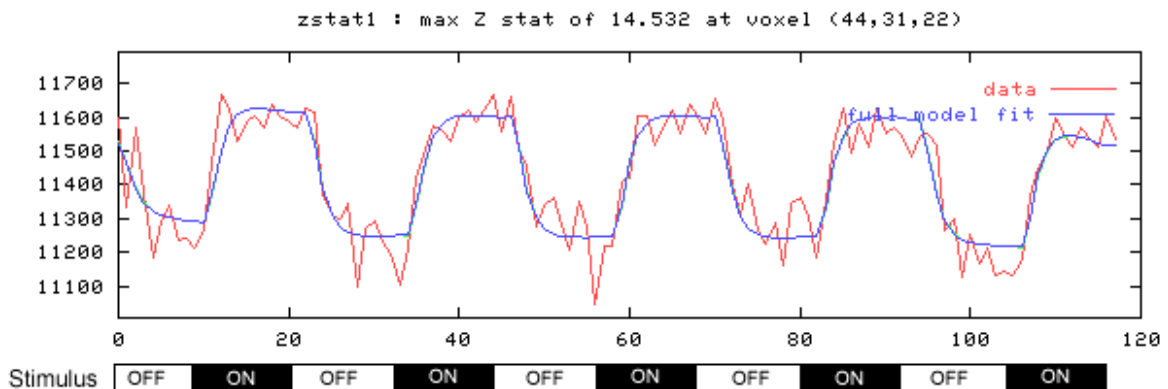
An imaging sequence is a combination of RF pulses and magnetic gradients utilized for MR signal acquisition. The need to rapidly sample the hemodynamic responses makes imaging speed essential in BOLD fMRI. Echo-planar imaging (EPI) is a sequence that collects the signal after a single excitation pulse during a rapidly switching readout gradient for spatial encoding (Cohen 1999). This enables signal collection in a contiguous raster- or spiral-like pattern in K-space and considerably reduces imaging time. Acquisition of a whole-brain image typically takes 1 to 3 seconds. Gradient echo EPI (GR-EPI) sequences in which the RF pulse is adjusted to set up a  $T_2^*$ -sensitive contrast are the ones most widely used for BOLD fMRI.

The  $B_0$  strength is an important factor in determining the magnitude of BOLD signal change (Gati et al 1999). Both the magnetic susceptibility difference between deoxyhemoglobin and the surrounding tissue (Ogawa et al 1993) and the contrast-to-noise ratio (CNR) between baseline and activated states (Gati et al 1997) increases with increasing  $B_0$  strength, creating a larger BOLD contrast between resting and activated states. The BOLD signal change differences observed between different tasks in an fMRI experiment often approach the detection limit. Thus, the use of stronger  $B_0$  provides increased sensitivity in detecting activation. Artifacts that affect data quality (chemical shift artifact, image distortion in the phase-encoding direction, susceptibility artifacts near the skull's air cavities) become, however, increasingly prominent at high fields. In part,

they compromise the theoretically linear CNR increase provided by high  $B_0$  strengths; control of them often requires specific measures.

## Data-analysis methods

A BOLD fMRI dataset is a series of T2\*-weighted brain images acquired during stimulus presentation. These images consist of voxels (“volume pixels”), the smallest distinguishable box-shaped parts of a 3D-image. Each voxel has a distinct intensity value at each time-point of the dataset. If the stimulation causes a blood-flow increase within a certain brain area, the voxels’ intensity values within that area will follow the temporal sequence of the stimulus presentation (see Figure 6); such an area is considered “activated” by the stimulation. Statistical analysis of fMRI data aims to find these activated areas at a pre-determined level of significance.



**Figure 6.** The time-course of signal intensity values within a visual-cortex voxel during lip-reading (red) and that time-course’s ideal model derived from the stimulation sequence (blue). The data fit the model well, as indicated by a high Z-statistic value. X-axis: time expressed in imaging volumes; y-axis: the voxel’s relative signal intensity. The bar at the bottom shows the stimulation sequence.

Prior to actual analysis, pre-processing is usually applied to the functional data to reduce artifacts. This may include motion correction, spatial smoothing, intensity normalization and temporal high-pass filtering (Lange 1999, Smith 2001). The statistical analysis of fMRI data usually utilizes general linear modeling (GLM). In such an analysis, an ideal model of expected intensity changes within an activated voxel is derived from the stimulus presentation sequence (for further details, see Lange 1999, Worsley 2001). This model takes into account the BOLD signal’s temporal characteristics by convolving the

stimulation sequence with a function that describes the hemodynamic response's lag and rise-time (see Figure 6). Each voxel's intensity values at each time-point of the dataset are then fitted to the model. The output is a statistical map in which a Z-value represents each voxel, describing how well that voxel's intensity profile across time corresponds to the model.

The Z-statistical maps display all positive and negative values within the imaged brain area. To determine which areas are *significantly* activated, these maps need to be thresholded at a determined significance level. The enormous number of comparisons in a typical fMRI data analysis (carried out for each voxel separately) renders this procedure vulnerable to type I errors and necessitates correcting the significance level for multiple comparisons. Several options exist for significance level correction, the choice often depending on the data and the purpose of the study. Probably the most conservative method is the Bonferroni correction, in which the significance value is divided by the number of statistical tests. A widely used and less stringent method is clustering that first applies a Z-statistic threshold to create contiguous activation clusters and then tests the significance of these clusters (not individual voxels), taking into account their spatial extent (Friston et al 1994). In focused comparisons restricted to a region that is defined by a priori hypotheses, it is appropriate to rely on significance values that are corrected within the region of interest (ROI) or, in some cases, uncorrected (Friston et al 1997).

Group-level fMRI statistics enable pooling of several subjects' data to increase statistical power and bring out robust effects that are stable across subjects; they also facilitate comparison of activations between differing groups of subjects. This usually requires that each subject's individual data are aligned into a common standard space (Talairach and Tournoux 1988, Fischl et al 1999). Group analysis results are usually reported in a standard space-based co-ordinate system which also allows comparison of results across different studies. GLM-based statistics at group level may employ a fixed-effects approach that models the fixed-effects variance (within-session across-time variances estimated in individual-level analyses), or a mixed-effects approach that additionally models the random-effects variance in the data (see Lange 1999, Woolrich et al 2004). The latter approach is suggested to be more conservative and to allow inferences concerning the population from which the subjects were drawn (Woolrich et al 2004).

## **Benefits and limitations**

Since its advent in the early 1990s, BOLD fMRI has generated great interest and is today an important tool for functional neuroscience. Other widely used methods for studying human brain function include positron emission tomography (PET), magnetoencephalography (MEG), and electroencephalography (EEG). FMRI and PET can be used to record brain hemodynamic activity, whereas MEG and EEG measure the magnetic fields and electric potentials related to neuronal activation. With PET it is also possible selectively to measure the function of different neurotransmitter systems with markers that bind to neurotransmitter receptors or to specific transporter proteins. All these methods have intrinsic benefits and limitations most prominently in relation to their spatial and temporal resolution.

The spatial resolution of BOLD fMRI clearly exceeds that of MEG and EEG and is more fine-grained than is that of PET; with BOLD fMRI it is possible to detect activation within neuronal organizations that measure only a few millimetres across (Yang et al 1996, Chen et al 1998). In contrast to MEG, subcortical structures are also reachable. The slow time-course of the BOLD response compared to that of neuronal events, however, generally prevents assessment of the temporal dynamics of neuronal activation with fMRI. The hemodynamic response in PET measurements reaches its peak even more slowly than does the BOLD response (within approximately one minute). Conversely, MEG and EEG offer temporal resolution in the order of milliseconds, which is clearly superior to fMRI or PET. Recent methodological advances, however, suggest that improving the temporal resolution of BOLD fMRI may be possible (Ogawa et al 2000).

fMRI measurements involve neither ionizing radiation nor radioactive agents. Thus, ethical problems related to radiation exposure do not limit their amount or duration, as often happens in PET experiments. The strong and rapidly changing magnetic fields during MR-imaging, however, require careful caution concerning metallic implants, foreign bodies, biostimulators, and RF energy. An additional factor to be dealt with in fMRI measurements is the loud acoustic noise that originates in the magnet's coil system during data acquisition, due to rapidly switching gradient fields. Even though hearing protection attenuates the noise to non-deleterious levels, it may still induce confounding

effects on the data. This may require modification of the experimental paradigms to ensure silence during stimulus presentation and allow control of the hemodynamic responses caused by scanner noise.

## Aims of the study

This thesis aimed to study the neural mechanisms of speech perception (I, II, IV) and selective attention (II, III) in fluent (I-IV) and dyslexic (IV) readers with BOLD fMRI. Specifically, in Studies I to IV we investigated

- 1) whether visual speech perception activates the human primary auditory cortex (I), hypothesizing that application of an individual-level analysis that circumvents problems associated with standard-space alignment enhances detection of hemodynamic responses within this anatomically variable cortical area.
- 2) whether attention to visual speech gestures affects auditory cortex activity (II), hypothesizing that attention to visual speech enhances auditory cortex hemodynamic activity compared to the situation when the subjects saw but ignored the same speech gestures.
- 3) how stimulus-driven and top-down effects (here, attention and sound presentation rate) modulate auditory cortex function during sound processing (III), hypothesizing that both will enhance auditory cortex hemodynamic activity; we further examined a possible interaction between attention and sound presentation rate.
- 4) how brain hemodynamic activity related to audiovisual speech perception differs between fluent and dyslexic readers (IV), hypothesizing that, due to the dyslexic readers' subtle perceptual deficits, they and fluent readers will recruit motor speech areas differently.

# Materials and methods

## Summary of methods (I–IV)

### Subjects, stimuli, and experimental designs

The subjects in Studies I to III were healthy native Finnish speakers with normal or corrected-to-normal hearing and vision and no linguistic or learning difficulties (see Table 1 for further details). In Study IV, the subjects were dyslexic or were fluent readers matched to the dyslexic subjects by age, sex, and handedness. The dyslexic subjects underwent a neuropsychological examination that confirmed their diagnosis and provided for each a detailed phonological processing and reading skill profile (see Study IV: Materials and methods, page 798-799 and Table 1, page 798).

We video-recorded stimulus material for Studies I to II and IV in a sound-attenuated chamber and created stimulus files from that material with video- and sound-editing programs (Purple, Praat, Sound Forge). For Study III, the stimuli were created with graphic- and sound-editing programs (Adobe Photoshop, CoolEdit). We programmed acoustic (wav files digitized at 44 100 Hz) and visual (sequences of image files, frame rate 25 Hz) stimulus presentation to occur in synchrony with fMRI data acquisition; current-pulses from the scanner guided the stimulus presentation. Prior to the scanning, the subjects familiarized themselves with the stimuli and practiced the experimental tasks outside the scanner. During scanning, they looked at visual stimuli via a mirror attached to the magnet's head coil, heard auditory stimuli through headphones, and responded to experimental tasks by pressing a button with an index finger. For details concerning the stimuli and experimental tasks in each experiment, see Table 1.

In Studies I to II we used a block design with intermittent baseline and active stimulation periods. Study III utilized a parametric design in which the sound stimuli were presented at five different rates during different blocks. For Study IV, clustered volume acquisition (Edmister et al, 1999) was applied to prevent masking of the auditory stimuli by scanner noise.

**Table 1.** Summary of methods for Studies I-IV.

	<b>Subjects and study design</b>	<b>Stimuli</b>	<b>Conditions and tasks</b>	<b>Data-analysis</b>
<b>I</b>	10 healthy volunteers (aged 21-30, right-handed; 7 males).  Block design with 30 s intermittent active and baseline blocks. Two runs with distinct active conditions ( <i>Vowels</i> and <i>Circles</i> ) and an identical <i>Baseline</i> condition.	<i>Baseline</i> : Still face image with a fixation mark on the mouth region. <i>Vowels</i> : A face silently articulating vowels (/a/, /i/, /o/, /y/ <sup>1</sup> ). <i>Circles</i> : moving circles (direction of movement: vertical, horizontal, left, right) superimposed on the mouth region of a still face image. Duration 560 ms, ISI <sup>2</sup> 420 ms.	<i>Baseline</i> : Fixate gaze on the mark. <i>Vowels</i> : lip-read the articulations; indicate whether any vowel occurs twice in succession. <i>Circles</i> : Watch the direction of the circles' movements; indicate whether two circles successively move in the same direction.	GLM-based analysis at single-subject level. ANOVA <sup>3</sup> on maximum percentage signal changes within HGs.
<b>II</b>	10 healthy volunteers (aged 21-30, right-handed; 7 males)  Block-design with 30 s intermittent active and baseline blocks. Two runs ( <i>Att_vowels</i> <sup>4</sup> and <i>Att_circles</i> <sup>5</sup> ) with identical stimuli; attention directed differentially across runs.	<i>Baseline</i> : Still face image with a fixation mark on the mouth region. <i>Att_vowels and Att_circles</i> : A face silently articulating vowels (/a/, /e/, /o/, /y/), with moving circles (direction of movement: vertical, horizontal, left, right) superimposed on the mouth region of the articulating face. Duration 560 ms, ISI 420 ms.	<i>Baseline</i> : Fixate gaze on the mark. <i>Att_vowels</i> : lip-read the articulations; indicate whether any vowel occurs twice in succession. <i>Att_circles</i> : Watch the direction of the circles' movement; indicate whether two circles successively move in the same direction.	GLM-based mixed-effects group analysis. ANOVA on maximum percentage signal changes within HGs.
<b>III</b>	12 healthy volunteers (aged 18-45, right-handed; 7 males).  Parametric block-design; 28 s blocks with differing auditory stimulus presentation rates and simultaneous visual stimulation at a steady rate. Attention shifted between visual and auditory stimuli 5 times during the experiment.	<i>Visual</i> : colored filled circles; duration 100 ms, SOA <sup>6</sup> 500-1500 ms, mean presentation rate 1 Hz. Targets: color change from red to orange at 50 ms from onset; nontargets: similar color change from yellow to orange. <i>Auditory</i> : harmonic tones (fundamental frequency 186 Hz), duration 200 ms, presentation rate 0.5, 1, 1.5, 2.5, or 4 Hz. Targets: 3% downwards frequency slide at 150 ms from stimulus onset; nontargets: similar slide upwards.	<i>Attend to pictures</i> : attend to visual stimuli; indicate the targets. <i>Attend to sounds</i> : attend to auditory stimuli; indicate the targets.	GLM-based mixed-effects group analysis. ANOVA on mean percentage signal changes within an auditory ROI.
<b>IV</b>	10 fluent readers (aged 27-31, right-handed; 6 males) 10 dyslexic readers (aged 22-34, right-handed; 6 males).  Clustered acquisition; <i>Matching</i> , <i>Conflicting</i> , and <i>Baseline</i> stimulation varied randomly in periods of 12-30 s.	<i>Baseline</i> : Still face image with fixation mark on the mouth region. <i>Matching</i> : audiovisual vowels (/a/, /i/, /o/, /y/) with matching auditory and visual content (e.g., visual /a/ + auditory /a/). <i>Conflicting</i> : vowels with conflicting auditory and visual content (e.g., visual /a/ + auditory /i/). Duration auditory 439-444 ms, visual 780 ms, ISI 100-400 ms.	<i>Matching</i> , <i>Conflicting</i> , and <i>Baseline</i> : Fixate gaze on the mark; indicate chance in stimulation type.	GLM-based mixed-effects group analysis.

1) /y/ is a vowel sound in the Finnish language; 2) inter-stimulus interval; 3) analysis of variance; 4) attend vowels; 5) attend circles; 6) stimulus onset asynchrony

## **FMRI data acquisition**

All fMRI measurements were performed with a 3T GE Signa scanner and a quadrature birdcage head coil at the Advanced Magnetic Imaging Center (Helsinki University of Technology, Espoo, Finland). Functional data were collected with a GR-EPI imaging sequence (flip angle 90°; TE 32 ms for Studies I to III and 40 ms for Study IV; TR 2500 ms for Studies I, II, and IV and 2800 ms for Study III). The in-plane resolution of the functional images was either  $3.4 \times 3.4$  mm (Studies I to III) or  $2.3 \times 2.3$  mm (Study IV), and the slice thickness either 3.4 mm (Studies I to III) or 4 mm with a 1 mm interslice gap (Study IV). Based on our pilot recordings, we collected 60 brain volumes per condition for the block-design studies (Studies I to III) and 101 volumes per condition for the one with clustered acquisition (Study IV). The duration for the functional data acquisition was 10 min (Studies I to II), 34 min (Study III), or 30 min (Study IV). For anatomic co-alignment and exploration of individual gyral anatomy, we acquired a T1-weighted fast spin echo volume with a slice prescription identical to that of the functional images but denser ( $0.9 \times 0.9$  mm) in-plane resolution, and a high-resolution (voxel size  $1 \times 1 \times 1.4$  mm) sagittal spoiled-gradient 3D-volume of the whole head.

## **FMRI data pre-processing**

The first two volumes of each run were discarded to allow signal stabilization. The following pre-processing was then applied on single-subject data: non-brain tissue extraction, motion correction (Jenkinson et al 2002), gaussian spatial smoothing with a kernel of 5 mm full-width half maximum, mean-based intensity normalization of all volumes by the same factor, and high-pass temporal filtering (Woolrich et al 2001).

## **Statistical methods**

The fMRI data were analyzed with FEAT 3.1 software, part of the FSL (FMRIB's Software Library, [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)). At single-subject level, we used a GLM-based method that prewhitens each voxel's time series with a nonparametric estimation of time series autocorrelation (Woolrich et al 2001). In Studies I to III, the model was convolved with a gamma-function (mean lag 6 s, SD 3 s) and its temporal derivative. In

Study IV, the model was not convolved, due to the sparseness of data sampling (see Study IV: Stimuli and experimental design, page 799). Instead, each block's first volume was discarded to ensure that the remaining ones sampled the hemodynamic response at its plateau. In group-level analyses, we used a mixed-effects GLM approach (Woolrich et al 2004) and thresholded the Z-statistical maps with a cluster-based method (Friston et al 1994).

The pooling of several subjects' data for group analyses can lead one to overlook activations in small and anatomically variable areas such as HG (Rademacher et al 1993, 2001, Penhune et al 1996). When we wished to focus on this particular anatomical region (I-II), we located HGs individually for each subject, adhering to anatomical landmarks (for the definition of HG, see Study I: Subjects and methods, page 126) and studied BOLD signal changes within HGs with analyses of variance (ANOVAs).

## **Experimental settings**

### **I**

The data were acquired over two runs (VOWELS and CIRCLES; for details, see Table 1 and Study I: Subjects and methods, page 126). During the active blocks of VOWELS, the subjects (N = 10) watched a face that silently articulated vowels. During the active blocks of CIRCLES, they watched a still image of that same face, with moving circles superimposed on its mouth region. The two runs had an identical baseline condition (showing only the still face image) and analogous tasks (see Table 1). We defined HGs in each subject's high-resolution MR images and assessed the extent of significant ( $Z > 2.3$ ,  $p < 0.01$ , corrected) BOLD signal changes within the medial half of HG, the location of the PAC according to cytoarchitectonical studies (Galaburda and Sanides 1980, Rademacher et al 1993). To compare HG activity between runs, we ran a stimulus type  $\times$  hemisphere ANOVA on maximum percentage signal changes within HGs.

## II

The data were collected during two runs (ATT\_VOWELS and ATT\_CIRCLES), the stimuli for which were identical (see Table 1). During the active blocks of both runs, the subjects (N = 10) watched a silently articulating face that simultaneously had a moving circle superimposed on its mouth region (for details, see Table 1 and Study II: Subjects and Methods, page 473). During the baseline blocks of both runs, they viewed a still face image. The instructions required the subjects to concentrate on either lip-reading the articulations (ATT\_VOWELS) or discriminating the direction of the circles' movements (ATT\_CIRCLES), while continuously maintaining eye fixation on a mark located in each circle's center. In data analysis, we contrasted the active blocks to the baseline within each run and compared the runs to each other in a mixed-effects group analysis. In the latter analysis, we limited the search for activation to a functionally determined ROI within the auditory cortex (see Results: caption of Figure 9, page 41). To compare HG activity between runs, we performed a stimulus type  $\times$  hemisphere ANOVA on maximum percentage signal changes within HGs at individual subject level.

## III

The subjects (N = 12) saw visual stimuli (colored circles) at a steady rate throughout the experiment and simultaneously heard harmonic tones at five different rates during the differing stimulation blocks (for details, see Table 1 and Study III: Stimulus Sequences and Behavioral Task, page 95). Both visual and acoustic stimulus streams occasionally included hard-to-detect targets. We instructed the subjects to change their focus of attention between the visual and auditory stimuli several times during the experiment and detect the targets in the modality attended to. We created a ROI of those auditory-cortex voxels that were active ( $Z > 2.33$ ,  $p < 0.05$ , corrected) across all rates and tasks (see Study III: Figure 2, on page 96). We then investigated the main effects of sound presentation rate and attention and their interaction in an ANOVA on mean percentage signal changes within that ROI.

## IV

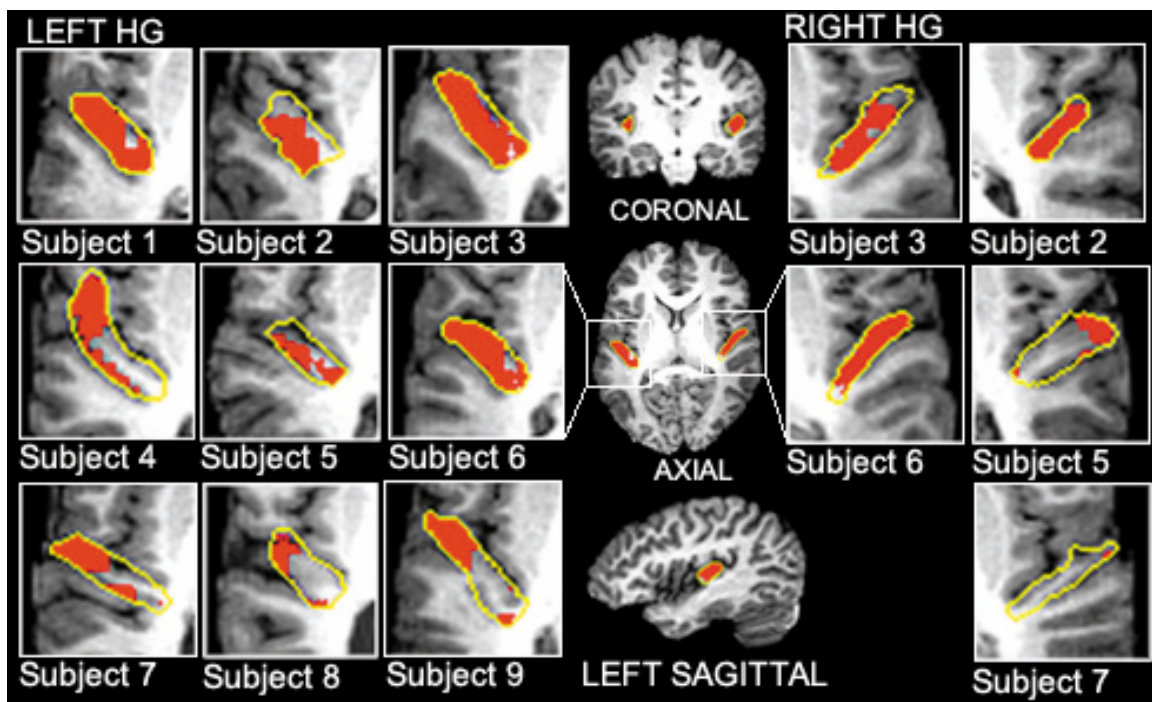
This experiment derived from our previous study (Ojanen et al 2005) that showed stronger activation during perception of conflicting than during perception of matching audiovisual phonemes within Broca's area, suggesting that this area participates in the processing of phonetic features extracted from auditory and visual speech. The reasoning behind this interpretation is based on the fact that matching audiovisual phonemes (such as auditory /a/ + visual /a/) constitute one unified percept, whereas conflicting ones (such as auditory /a/ + visual /o/) carry two separate phonetic inputs (auditory and visual) into higher-order processing areas. Thus, conflicting stimulation may cause an increased load in cortical regions that process phonetic features extracted from auditory and visual speech stimuli. In the current study, we applied this paradigm to compare brain hemodynamic activity of dyslexic (N = 10) and fluent readers (N = 10) during audiovisual speech perception.

The fMRI data were acquired with a sparse temporal sampling design (for details, see Table 1 and Study IV: Stimuli and experimental design, page 799). The experiment had three conditions: phonetically matching and phonetically conflicting audiovisual vowels, plus a baseline. The auditory and visual components of matching and conflicting vowels were spatially and temporally congruent; the stimuli differed only with respect to their phonetic congruency. In data analysis, we contrasted the active conditions to the baseline and to each other and compared signal changes between dyslexic and fluent readers. We further investigated the co-variance between the dyslexic readers' phonological processing skills and BOLD signal change strength within Broca's area, the auditory cortex, and the visual cortex.

# Results

## Lip-reading activates the human primary auditory cortex (I)

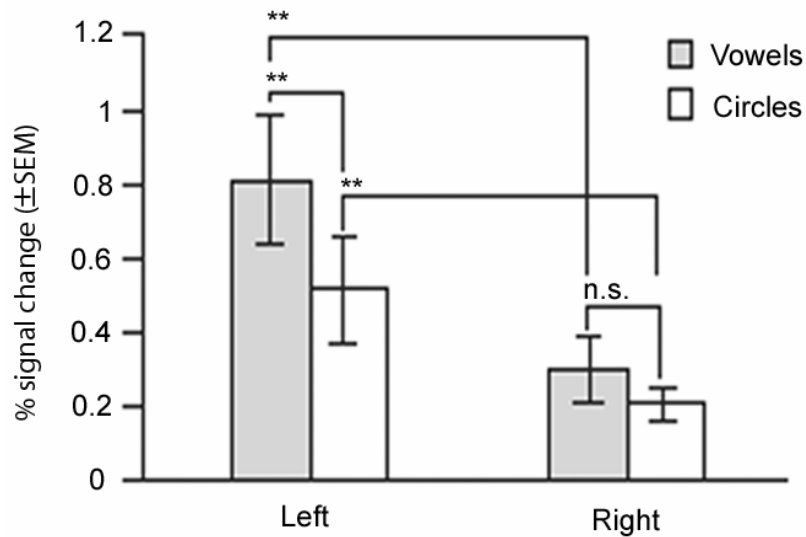
Nine subjects showed significant ( $Z > 2.3$ ,  $p < 0.01$ , corrected) signal changes within the left HG during the VOWELS run, with this activation clearly extending to its medial-most part in seven of them (Figure 7). Right-hemisphere HG activation during the VOWELS run exceeded the significance limit for five subjects. Unexpectedly, six subjects exhibited HG activation during the CIRCLES run (Table 2). ANOVA (Figure 8) showed significant main effects for stimulus type,  $F(1, 9) = 7.9$  ( $p < 0.05$ ) and hemisphere,  $F(1, 9) = 10.3$  ( $p = 0.01$ ) and a significant hemisphere by stimulus interaction,  $F(1, 9) = 14.8$  ( $p < 0.01$ ).



**Figure 7.** Red color shows significant ( $Z > 2.3$ ,  $p < 0.01$ , corrected) BOLD signal changes during lip-reading within HGs. The activation maps and HG outlines (yellow line) are collapsed into a 2D-image; the axial images' orientation follows neurological convention. Visual speech perception activated HG of 9 subjects (total  $N = 10$ ), and activation clearly extended to the anatomical location of PAC in at least 7 subjects (subjects 1-7 within the left HG and subjects 2, 3, and 6 within the right HG) Illustration reprinted with permission from Lippincott Williams & Wilkins from: Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A and Sams M: Primary auditory cortex activation by visual speech: an fMRI study at 3 T (*Neuroreport* 16(2):125-128). © (2005) Lippincott Williams & Wilkins.

**Table 2.** Number of subjects showing activation within HGs and their medial portions during VOWELS and CIRCLES runs.

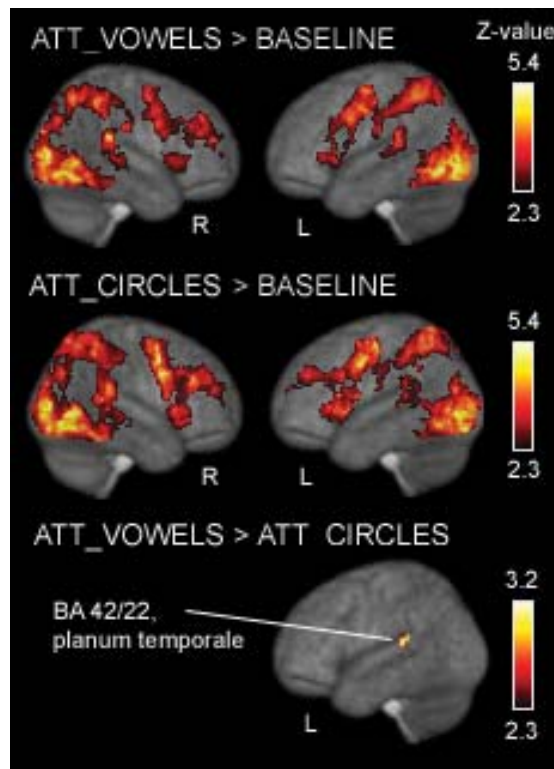
Run	LEFT HG	medial half	RIGHT HG	medial half
VOWELS	9	7	5	3
CIRCLES	6	3	1	0



**Figure 8.** Maximum percentage BOLD signal changes within HGs  $\pm$  standard error of mean (SEM). \*\*  $p < 0.001$ , n.s = non-significant. Illustration reprinted with permission from Lippincott Williams & Wilkins from: Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A and Sams M: Primary auditory cortex activation by visual speech: an fMRI study at 3 T (*Neuroreport* 16(2):125-128). © (2005) Lippincott Williams & Wilkins.

## Attention to visual speech modulates auditory cortex activity (II)

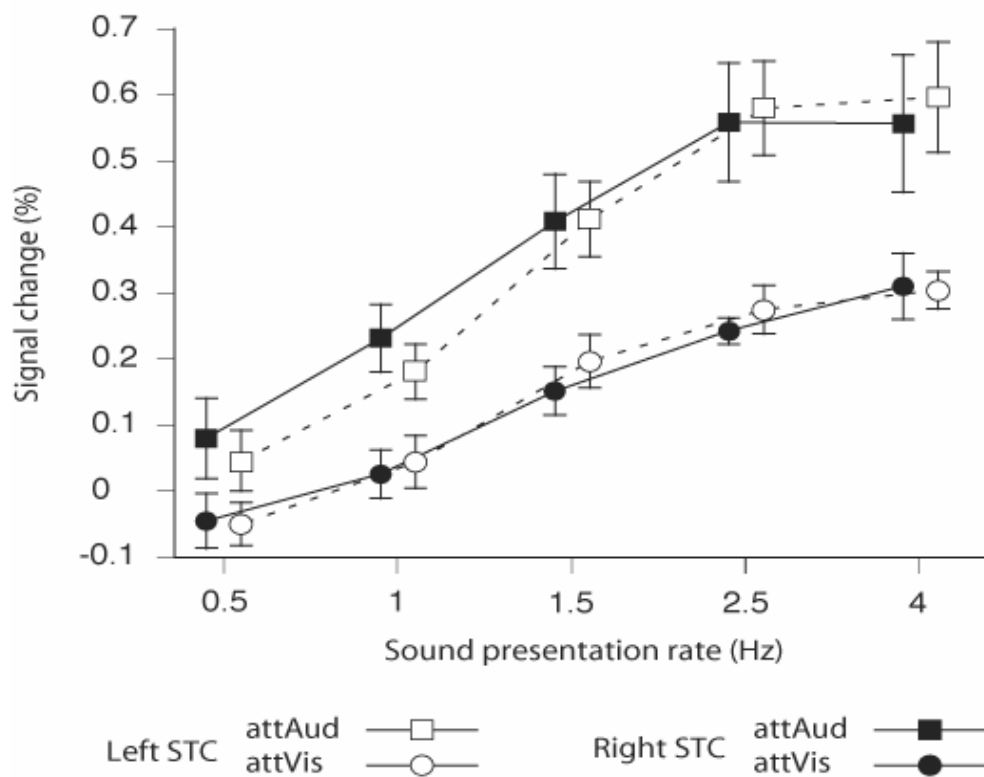
Both attentional tasks activated the bilateral auditory and visual cortices and an extensive frontoparietal network including Broca's area, its right-hemisphere homologue, the left motor cortex, bilaterally the lateral premotor cortex, parietal lobules, frontal eye fields, the supplementary motor cortex, and the anterior to middle cingulate cortex (Figure 9). Activation within the left planum temporale (secondary auditory cortex, Brodmanns area, BA, 42/22) was stronger during the ATT\_VOWELS than during the ATT\_CIRCLES run (Figure 9). No such effect emerged within the right auditory cortex or in the reverse direction (ATT\_CIRCLES > ATT\_VOWELS). The ANOVA for maximum percentage signal changes within HGs showed no significant main effects for either stimulus type or hemisphere.



**Figure 9.** Significant BOLD signal changes during ATT\_VOWELS and ATT\_CIRCLES runs and the differential auditory-cortex activation between the runs. In the contrasts between each active condition and baseline, the Z-statistic maps were thresholded at  $Z > 2.3$ ,  $p < 0.05$  (corrected across the whole acquisition volume). In the contrast between ATT\_VOWELS and ATT\_CIRCLES runs, the search was limited to the areas that survived thresholding in ATT\_VOWELS > BASELINE contrast, and the resulting statistical maps thresholded at  $Z > 2.3$ ,  $p < 0.05$  (corrected within the ROI). From: Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, and Sams M: Attention to visual speech gestures enhances activity in the left planum temporale (*Human Brain Mapping* 27:471-477). Illustration reprinted with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

### Stimulus-driven and attentional effects on auditory cortex activity (III)

Auditory cortex hemodynamic activity increased bilaterally with increasing sound presentation rate:  $F(4, 44) = 52.2$  ( $p < 0.001$ ) and during attention to sounds:  $F(1, 11) = 21.9$  ( $p < 0.01$ ). As a novel finding, we detected an interaction:  $F(4, 44) = 2.8$  ( $p < 0.05$ ) between sound presentation rate and attention in auditory cortex activity: larger attention effects occurred at higher stimulation rates (Figure 10).



**Figure 10.** Mean percent BOLD signal changes ( $\pm$  SEM) within the auditory cortex. The activity rises more steeply with increasing sound presentation rate during auditory than during visual attention. STC = superior temporal cortex, attAud = attend auditory stimuli, attVis = attend visual stimuli. From: Rinne T, Pekkola J, Degerman A, Autti T, Jääskeläinen IP, Sams M, Alho K: Modulation of auditory cortex activation by sound presentation rate and attention (*Human Brain Mapping* 26:94-99). Illustration reprinted with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

## **Differences between fluent and dyslexic readers' speech perception (IV)**

Conflicting stimulation activated the classical motor speech regions (Broca's area; left BA 44/45/6) and the supplementary motor area (SMA) more strongly than did matching stimulation (Figure 11). In this contrast, dyslexic readers exhibited more extensive activation difference than did fluent readers (Figure 11); their stronger activation during conflicting stimulation also extended to the right-hemisphere homologue of Broca's area (right BA 44/45/6), the bilateral IPL (BA 7/40), the left STS (BA 21/22), and the left ventral visual cortex (BA 19/37).

A composite contrast (conflicting vs. matching) *dyslexic* - (conflicting vs. matching) *fluent* confirmed a stronger between-condition activation difference in dyslexic than in fluent readers within areas that cover the suggested MNS regions: Broca's area, its right-hemisphere homologue, and the left rostral IPL; and, additionally, within the SMA (Figure 11).

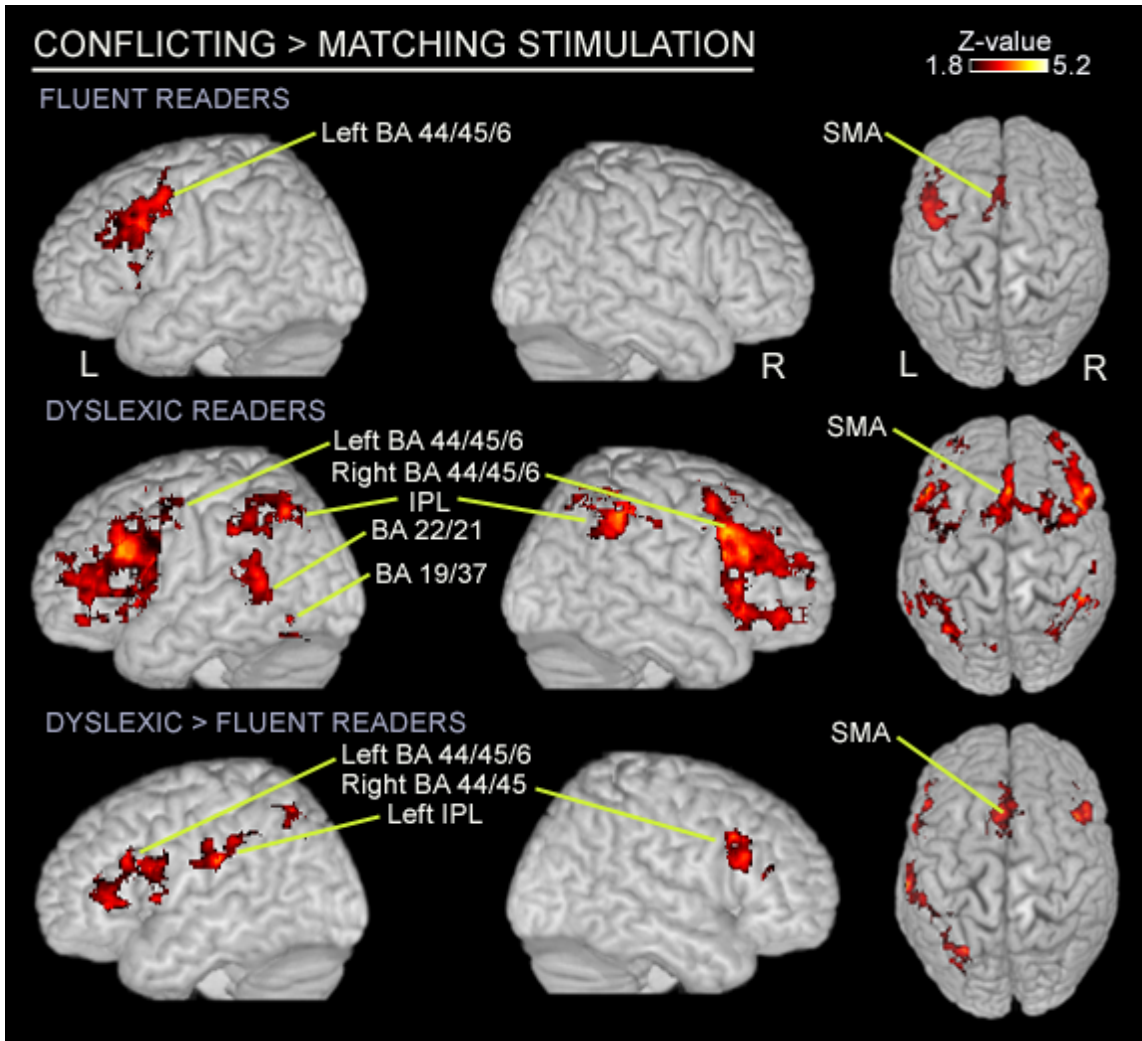
The dyslexic readers' neuropsychological index scores that reflect the three essential components of phonological processing (phonological awareness, phonological memory, and rapid naming; Torgesen et al 1994) co-varied with BOLD signal change within Broca's area, the visual cortex, and, to a lesser extent, within the auditory cortex (See Table 3).

In the comparison between matching stimulation and baseline, dyslexic readers showed stronger activation than did fluent readers in the bilateral anterior to middle auditory association cortex, insulae, basal ganglia, and the right inferior frontal and orbitofrontal cortex (See Study IV: Figure 2, page 800 and Table 2, page 801). During conflicting stimulation, they exhibited stronger activation additionally within the bilateral ventral visual cortex, supplementary motor area, anterior cingulate cortex, and cerebellar vermis (Study IV; Figure 2, page 800 and Table 2, page 801). No activation differences were detectable in the reverse direction (fluent > dyslexic readers).

**Table 3.** Clusters showing co-variance ( $p < 0.01$ , spatial extent  $> 20$  voxels) of phonological processing index scores with BOLD signal strength; number of voxels, cluster maximum Z-values and their locations. Table adapted from: *NeuroImage*, Vol 29, Pekkola J, Laasonen M, Ojanen V, Autti T, Jääskeläinen IP, Kujala T and Sams M: Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T, pages 797-807, © (2006), with permission from Elsevier.

Stimulation type	Index score	Region	Voxels	Max Z	X *	Y *	Z*	BA	left/ right	Anatomical location (Max Z)	
Matching vowels	Phonological awareness	Occipital	97	2.99	22	-52	-4	19/18	right	Lingual gyrus	
			68	2.68	-24	-86	26	19/18	left	Superior occipital gyrus	
			53	3.31	6	-78	-2	17	right	Lingual gyrus	
			33	2.55	34	-68	30	19	right	Middle occipital gyrus	
			22	2.33	24	-74	24	19	right	Superior occipital gyrus	
	Phonological memory	Frontal	32	2.81	-48	22	16	45	left	Inferior frontal gyrus**	
			22	2.28	-18	-58	2	18/19	left	Calcarine gyrus	
	Rapid naming	Occipital	68	2.65	-34	6	32	44	left	Inferior frontal gyrus***	
			59	2.49	-43	34	38	44/45	left	Middle frontal gyrus	
			155	3.8	14	-86	32	18/19	right	Cuneus	
	Conflicting vowels	Phonological awareness	Occipital	57	3.05	-14	-80	44	19	left	Superior occipital gyrus
				22	3.16	-1	-88	10	18	left	Calcarine gyrus
				92	3.14	56	-52	4	21	right	Middle temporal gyrus
				146	2.94	-30	-72	22	19/18	left	Middle occipital gyrus
96				2.77	20	-52	-6	19/18	right	Lingual gyrus	
Phonological memory		Occipital	43	3.21	-10	-66	4	17/18	left	Lingual gyrus	
			41	2.99	36	-64	32	19	right	Middle occipital gyrus	
			33	2.86	4	-82	-2	17	right	Calcarine gyrus	
			30	2.48	26	-76	28	19	right	Superior occipital gyrus	
			21	2.68	-50	24	16	45	left	Inferior frontal gyrus**	
Rapid naming	Temporal	51	4.18	66	-10	-10	21/22	right	Middle temporal gyrus		
		48	2.58	-14	-54	2	18/19	left	Lingual gyrus		
		68	2.39	-44	32	38	45	left	Middle frontal gyrus		
		38	2.48	-32	6	30	44	left	Inferior frontal gyrus***		
		35	3.42	-36	36	12	45	left	Inferior frontal gyrus**		
Rapid naming	Occipital	97	4.35	14	-88	34	19/18	right	Cuneus		
		71	3.29	-38	-68	32	19	left	Middle occipital gyrus		
		48	3.44	-6	-82	44	19	left	Superior occipital gyrus		
		35	3.71	-36	12	42	44	left	Precentral gyrus		
		Temporal	91	2.90	54	-50	4	21	right	Middle temporal gyrus	

\*MNI coordinates; \*\* trigonal part; \*\*\*opercular part



**Figure 11.** Cortical areas with stronger activation during conflicting than during matching stimulation ( $Z > 1.8$ ,  $p < 0.05$ , corrected). BA = Brodmann's area, SMA = supplementary motor area, IPL = inferior parietal lobule. Illustration reprinted from: *NeuroImage*, Vol 29, Pekkola J, Laasonen M, Ojanen V, Autti T, Jääskeläinen IP, Kujala T and Sams M: Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T, pages 797-807, © (2006), with permission from Elsevier.

## Discussion

These four fMRI studies investigated the brain hemodynamic activity of fluent and dyslexic readers during their seeing and hearing speech and non-speech stimuli. Lip-reading and, to a lesser extent, visual perception of non-speech stimuli, activated the PAC of fluent readers (I). Attention to lip-reading enhanced the activity within the left PT (II), and attentional modulation of auditory cortex activity during auditory processing interacted with sound presentation rate (III). During audiovisual speech perception, activation differences emerged between fluent and dyslexic readers within several cortical areas relevant to speech processing (IV).

### **Lip-reading activates the human primary auditory cortex (I)**

Our observations contradict traditional views as to the strict unimodality of the primary sensory cortices by showing that visual input can modulate human auditory cortex function at the lowest level of the cortical hierarchy (I). These findings agree with the initial group-level study to show auditory cortex activation during lip-reading and suggest this effect also to involve PAC (Calvert et al 1997), but are in disagreement with an fMRI (Bernstein et al 2002) and PET study (Paulesu et al 2003) that relied on standard-space-based probabilistic mapping to define HG. We suggest that this discrepancy may relate to differences in spatial resolution between these studies. Furthermore, our conducting the experiment at a higher magnetic field than before (Calvert et al 1997, Bernstein et al 2002), may have enhanced our detection of activations.

The functional relevance of these findings (I) lies in considering PAC a possible site for audiovisual interactions. This might indicate that visual speech input affects auditory processing at a pre-phonetic level, in support of the early integration theories. In accordance with such interpretations, several studies have suggested that auditory and visual information interact temporally early in the processing stream (Giard and Peronnet 1999, Fort et al 2002a, b, Molholm et al 2002, Klucharev et al 2003, Besle et al 2004,

Jääskeläinen et al 2004b, Möttönen et al 2004, van Wassenhove et al 2005), although in these studies the auditory-cortex interactions have emerged within the secondary auditory cortex, not within PAC. Further support for the involvement of PAC in audiovisual processing comes from a recent animal study (Ghazanfar et al 2005) that demonstrated audiovisual interactions within the rhesus monkey auditory core and lateral belt during observation of communicational sounds and gestures; the majority of interactive responses in that study, however, emerged within the belt.

Putting into question whether the auditory cortex activations in Study I were specific to visual speech processing was the fact that the control stimuli (moving circles) also enhanced activity within HG and the anatomical area of PAC (I). The stronger BOLD signal changes for speech gestures than for moving circles in the left-hemisphere HG and the greater number of subjects showing activation within the PAC region during visual speech perception than during the control condition (I), however, suggest a degree of specialization for speech gesture processing within the left HG and PAC. Other fMRI-studies have also reported auditory cortex activation during observation of moving visual stimuli devoid of linguistic meaning to the observer (Howard RJ et al 1996, MacSweeney et al 2004); such findings might result from audiovisual interactions not specific to speech processing (for further discussion, see page 54).

Ghazanfar and colleagues (2005) utilized stimuli quite comparable to ours (I). Showing that a natural or learned association between auditory and visual stimuli is not strictly necessary for audiovisual interactions in the auditory cortex, they detected enhancement of auditory neural responses to communicational sounds during simultaneous observation of moving geometrical forms. A great majority of these responses, however, specifically occurred during simultaneous perception of a call and the sight of a monkey uttering the call. Thus, these studies (Ghazanfar et al 2005; I) together suggest that even though non-linguistic or non-communicative visual stimuli may elicit responses in the putatively unisensory auditory cortex, the auditory cortex is to some degree selective for communicational gestures during perception both visual (I) and audiovisual (Ghazanfar et al 2005).

## **Attention to visual speech modulates auditory cortex activity (II)**

The enhancement of auditory cortex hemodynamic activity by attention to articulatory gestures (II) may account for the facilitating effect of visual speech gestures on auditory speech perception in noisy conditions (Sumbly and Pollack 1954); attending to a specific speaker's face may tune object-specific neuronal populations within the auditory cortex and enhance their responsiveness to speech sounds (see Weinberger 2004). The lack of any attentional effect within HG (II) suggests that this effect is limited to the higher-order secondary auditory cortex. It further suggests possible that the HG activation during lip-reading (I, II) may have occurred independently of voluntary attention and that the greater HG activation during lip-reading than during observation of moving geometrical forms (I) might have been due to non-linguistic stimulus features (e.g., the articulatory movements' greater visual complexity than that of the control stimuli). Since, however, the effects in fMRI studies often approach the detection limit, one must be cautious when interpreting such negative findings.

The extensive frontoparietal activation during both runs (II) covered a network reportedly involved in visual spatial attention; this network is suggested to mediate the effects of attention on the sensory cortices (Nobre 2001). The bilateral auditory cortex activation during the ATT\_CIRCLES run (II) might suggest that visual speech gestures are able to enhance auditory cortex activity involuntarily, without conscious attention. Since, however, in our previous study, observation of movement only of the circles elicited auditory cortex activation (I), we cannot here exclude the possibility that this activation (II) was due merely to the circle stimuli themselves.

The enhancement of left PT activity during attention to lip-reading (II) further suggests that this region specifically participates in the linguistic processing of visual speech gestures. The PT represents the secondary auditory cortex that is anatomically and histologically homologous to the monkey posterior belt (Celesia 1976, Galaburda and Sanides 1980). Thus, our observations (II) together with earlier studies' (Calvert et al 1997, Campbell et al 2001) suggest speech-specific processing of seen articulatory gestures to occur within the auditory sensory cortex, and they support the view that

phonetic information contained in visual speech may affect the hierarchically early stages of auditory processing.

The enhanced BOLD responses during active linguistic processing of visual speech gestures (II) also extended to the posterior-most part of the left PT and the adjacent temporoparietal junction (area TpT; Galaburda and Sanides 1980, Galaburda 1982). A special function has been suggested for this area in mediating sound- and gestural-based speech processing (Hickok and Poeppel 2004). Due to its responsiveness to several types of complex sounds, PT has been suggested to function in analyzing and segregating sound patterns and matching them to learned spectrotemporal templates (Griffiths and Warren 2002). Lesion studies demonstrate that, in addition to such auditory processing, the PT may participate in speech production: Limited lesions around the posterior end of the sylvian fissure (including the posterior PT) leave speech comprehension mostly intact while disturbing speech production by nonsemantic errors (conduction aphasia; Benson DF et al 1973, Damasio and Damasio 1980). Further, recent functional studies have reported hemodynamic responses within the posterior PT during subvocal speech production, verbal memory, and reading tasks that require maintenance of inner speech (Herholz et al 1996, Hickok et al 2000, Wise et al 2001, Hickok et al 2003, Buchsbaum et al 2005). Based on the above observations, the posterior PT and temporoparietal junction have been suggested to mediate transformations between speech sounds and their articulatory-gestural representations and to feed that information further into frontal cortex areas that process and “understand” such input (Hickok and Poeppel 2004; see also Rizzolatti and Arbib 1998, Fadiga and Craighero 2003, Rizzolatti and Craighero 2004). Our findings (II) suggest the possibility that visual speech gestures contribute to speech perception via this postulated interface between differing speech representations.

Purely acoustic feature-based speech processing, on the other hand, has been suggested to involve more anterior temporal lobe areas that also project into the frontal cortex (Scott and Johnsrude 2003, Hickok and Poeppel 2004). Analogous to the “what” and “where” streams in visual (Ungerleider and Mishkin 1982) and auditory (Rauschecker and Tian 2000, Tian et al 2001, Tian and Rauschecker 2004, Rauschecker and Tian 2004) processing, speech perception may thus proceed along parallel streams that operate with different classes of representations (Hickok and Poeppel 2004; see, however, Scott and

Johnsrude 2003 for another proposition concerning the exact anatomical location of these streams). The existence of such parallel processing streams could reconcile the long-opposed auditory-based versus motor speech theories and account for the amazing robustness of speech perception.

### **Stimulus-driven and attentional effects on auditory cortex activity (III)**

The enhancement of bilateral auditory cortex hemodynamic activity both by attention to sounds and by an increased sound presentation rate (III) agrees with earlier findings from studies investigating these effects in isolation (Hillyard et al 1973, Hari et al 1989, Woldorff et al 1993, Binder et al 1994, O'Leary et al 1996, Pugh et al 1996, Dhankhar et al 1997, Grady et al 1997, Rees et al 1997, Alho et al 1999, Jäncke et al 1999, 2001, Hashimoto et al 2000, Mechelli et al 2000, Tanaka et al 2000, Harms and Melcher 2002, Lipschutz et al 2002, Degerman et al 2006). The enhancement of the attentional effect with increasing sound presentation rate (III) indicates a dynamic interaction between stimulus-driven and top-down effects on auditory cortex activity; this finding is compatible with EEG and MEG recordings that suggest sound presentation rate to modulate both stimulus-dependent and attention-related auditory cortex responses (Rif et al 1991, Teder et al 1993).

Previous ERP studies suggest two mechanisms that may account for such an interaction. First, each attended sound elicits attention-related activity (Hillyard et al 1973, Näätänen et al 1978, Woods et al 1980); this attention-related activity to individual sounds may accumulate at rapid stimulus presentation rates during auditory attention. Second, higher presentation rates may facilitate attention itself through higher task demands or a more frequent occurrence of attended sounds (Näätänen 1982, Hansen and Hillyard 1988, Alho et al 1990, Woldorff and Hillyard 1991); these two mechanisms may also work simultaneously. In terms of neuronal events, such interaction may take place through selective tuning of the auditory cortex neurons' responsivity to attended frequencies or sound object features (see Fritz et al 2003, Weinberger 2004). Hypothetically, a similar mechanism may mediate enhancement of auditory cortex activity during attention to visual speech (II).

The current study (III) reports the average auditory cortex activation over a whole 28-s stimulation block. Studies investigating the temporal evolution of the auditory cortex fMRI signal during a stimulation period have shown that sound presentation rate modulates the block-onset, block-offset, and steady-state responses differentially (Harms and Melcher 2002) and that, throughout the block, attention-related activation increases gradually (Petkov et al 2004). These findings agree with ours (III) on the separability of stimulus-dependent and attention-related activity in the auditory cortex.

Our results (III) in part contradict those of Frith and Friston (1996) who, using a comparable experimental setting, detected neither enhancement of auditory cortex activity by attention to sounds, nor interaction between sound presentation rate and attention. Visual and auditory tasks in that experiment were, however, rather more simple than those in the current experiment (III); this fact may have enabled the subjects to divide their attention between visual and auditory stimuli. Our results (III) thus have two implications that relate to the planning of further experiments on auditory cortex function: They call for taking into account both stimulus-driven and attentional factors and making the experimental tasks sufficiently demanding to extract attentional effects.

### **Differences between fluent and dyslexic readers' speech perception (IV)**

The dyslexic readers' enhanced activity within Broca's area and the left rostral IPL (the suggested MNS regions), in a contrast that probed phonetic processing of auditory and visual speech stimuli (IV), may relate to speech processing strategies differing between dyslexic and fluent readers. Specifically, dyslexic readers may rely more on strategies that utilize articulatory-gestural speech representations than on those based on the acoustic features of speech sounds. The covariance between Broca's area activity and phonological processing skills among the dyslexic readers (IV) suggests the compensatory nature of this finding; those dyslexic readers who most effectively recruit motor-articulatory mechanisms during speech perception could also have developed the most advanced phonological processing abilities.

Such interpretation gains support from studies that report enhanced motor speech area activity during auditory processing of ambiguous or difficult speech tokens (Nakai et al

1999, Callan et al 2003a, b, Callan et al 2004b, Wang et al 2003, Golestani and Zatorre 2004). Specifically, Broca's area activity has been strengthened if noise partially masks the auditory component of audiovisual speech (Callan et al 2003a) and if a story is heard in a non-native rather than the native language (Nakai et al 1999). Differing phonetic contrasts between languages offer an additional tool to study phonetic-level processing of unclear or difficult speech sounds. Brain regions linked with the auditory-articulatory circuit (PT, the temporoparietal junction, and Broca's area; see Hickok and Poeppel 2004) have shown stronger activation during a forced-choice differentiation between heard /r/ and /l/ in native Japanese speakers (for whom this phonetic contrast is difficult) than in native English speakers (for whom it is easy; Callan et al 2004b). The same cortical regions have also displayed enhanced activity during processing of newly-learned speech sounds (Callan et al 2003b, Wang et al 2003, Golestani and Zatorre 2004) and – analogous to our observation of co-variance between Broca's area activity and phonological processing skills (IV) – Broca's area activity has correlated with native Japanese speakers' /r/ vs. /l/ identification performance (Callan et al 2004b). Together with these studies, our findings (IV) thus substantiate theories of several parallel processing streams serving audiovisual speech perception; even though predominantly acoustic-based speech processing may suffice or prevail under normal circumstances, the use of internal models of speech articulations may facilitate speech sound identification in more difficult cases.

Our findings within the visual cortex (IV) contradict previous reports of decreased visual cortex activity in dyslexic readers during various visual processing tasks (reviewed in Habib 2000, Salmelin and Helenius 2004). An MEG study on viewing human faces has, however, suggested that dyslexic readers' reduced occipitotemporal responses during visual perception may be specific to certain stimulus categories (Tarkiainen et al 2003) and possibly may not apply to speech gestures. Considering the co-variation of several phonological processing index scores with visual cortex activity (IV), we suggest that our dyslexic readers' increased ventral visual cortex activity during perception of conflicting audiovisual speech (IV) reflects behavioral observations that despite their lip-reading difficulties (de Gelder and Wroomen 1998, Hayes et al 2003), dyslexic readers still strongly rely on visual information in speech perception (Hayes et al 2003).

May, then, the differential activation between dyslexic and fluent readers in conflicting > matching contrast (IV) result from increased attention to the visual speech component during conflicting stimulation? In Study II, activation during visual speech perception extended to the bilateral inferior frontal cortex, rostral IPLs, and SMA (the areas where dyslexic readers showed stronger activation than did fluent readers in conflicting > matching contrast in Study IV), regardless of the direction of attention. No differential activation between the attentional conditions emerged within these areas, not even at very lenient significance levels ( $Z > 2.3$ , uncorrected; II, unpublished result). Further, the experimental task in IV (merely detecting a change in stimulation type) posed no great attentional demands. Thus, we consider it likely that the between-groups differential activation during processing of phonetic features extracted from auditory and visual speech stimuli (IV) was not mediated via increased attention to visual speech gestures.

Our observation of stronger lateral temporal and insular activation in dyslexic than in fluent readers during both conditions (IV) differs from reports of consistently diminished left temporoparietal and insular activation in dyslexic participants during various linguistic processing tasks (Rumsey et al 1997, Shaywitz et al 1998, Simos et al 2000a, b, Paulesu et al 2001; see Habib 2000, Temple 2002, Salmelin and Helenius 2004). The lack of previous functional imaging studies on dyslexic readers' audiovisual speech perception, however, complicates comparison of ours to others' results. We consider it possible that these differences are related to task difficulty, or that dyslexic readers' temporoparietal activation pattern during linguistic processing is essentially task-dependent.

## **Potential limitations**

Assessing the time-scale of auditory-cortex responses during visual speech perception (I to II) is impossible because of the inherent limitations of fMRI methodology. The current studies can thus provide information only concerning the hierarchical level at which visual speech affects auditory-cortex activity. Modulation of auditory cortex function during audiovisual speech perception has been suggested to occur via feedback connections from the STS (Calvert et al 2000, 2001, Calvert 2001). The temporal latencies of audiovisual interactions during speech processing, however, demonstrate that

audiovisual interactions during speech perception most probably take place at multiple processing levels via several feedback and feedforward mechanisms (Klucharev et al 2003, Besle et al 2004, Möttönen et al 2004, Jääskeläinen et al 2004b, van Wassenhove et al 2005, Musacchia et al 2006). Although animal studies suggest that visual input arrives in the auditory cortex via feedback connections (Schroeder and Foxe 2002), this may not be self-evidently applicable to human speech processing. Thus, the actual neural mechanisms through which visual input accesses the auditory processing stream during speech perception remain to be studied further.

A factor to be considered in any block-design fMRI experiment is the loud acoustic noise that originates in the scanner's coil system during rapid switching of gradient fields. Because this noise remains stable across the measurement, its effects are often considered negligible. Considering that a wide variety of sights and sounds evokes audiovisual interactions at the latency range of sensory-specific processing (Giard and Peronnet 1999, Pourtois et al 2000, Fort et al 2002a, b, Molholm et al 2002, Lebib et al 2003), this may, however, not be the case. We thus consider it possible that the auditory cortex activation during observation of visual stimuli (I to II) was based on the possibility that the visual stimuli modulated pre-existing auditory cortex activity caused by the scanner noise.

Although it may have limited the extendability of our results to the left-handed or the ambidextrous, we chose to include only right-handers. This reduced confounding factors that could arise in group-level analyses due to lessened or reversed hemispheric dominance of speech functions in left-handed subjects.

### **Insights for future studies**

Human ability to understand and produce speech results from a long evolution of the brain and the vocal tract. It is thus plausible that the visual speech perception-related functions within the posterior auditory cortex (its enhanced activity during attended lip-reading (II); possible matching of acoustic and gestural-based speech information, Hickok and Poeppel 2004) have emerged during this evolution. Taking into account the posterior belt neurons' spatial selectivity (Rauschecker and Tian 2000, Tian et al 2001, Tian and Rauschecker 2004, Rauschecker and Tian 2004), a possible "original" function for its

multisensory-responsive cells (Schroeder et al 2001, Schroeder and Foxe et al 2002, Fu et al 2003, Brosch et al 2005) may be spatial orientation to and distinguishing of objects in space by combining information from different modalities. This possibility is compatible with the observation that eye movements modulate the auditory cortex activity of non-human primates (Fu et al 2004, Werner-Reiss et al 2003). Based on that reasoning, it would be interesting to compare responses for audiovisual integration of spatial vs. object identity features of speech and non-speech stimuli within different parts of the auditory cortex.

Further, the role of predominantly motor cortical areas in speech perception awaits more detailed characterization. It might be fruitful to systematically investigate the hypothesis of such areas' enhanced involvement in the perception of unclear or degraded speech inputs (IV) and, on the other hand, compare their activity during speech perception vs. other action observation.

The suggested speech-specific processing of articulatory gestures within the auditory cortex (II) might be substantiated by further experiments in which subjects would shift their attention within a perceptual category (i.e., faces) instead of between categories (as in II). Additionally, since animal studies have shown top-down cognitive modulation of both cortical and subcortical sensory processing (Suga and Ma 2003, Weinberger 2004), it might be useful to study subcortical-level effects of selective attention and learned audiovisual associations in humans by optimized fMRI techniques.

## Conclusions

The current study suggests that the prerequisites for audiovisual interactions exist at the level of the primary auditory cortex (I) and that speech-specific processing of visual articulatory gestures takes place within the auditory sensory cortex (I, II). It places stress on the importance of attention in guiding sensory function (II, III) and provides evidence of a dynamic interaction between stimulus-driven and attentional effects on auditory cortex activity (III). It further suggests that speech perception strategies may vary according to the information available and to the individual characteristics of the observer: During speech perception, dyslexic readers may utilize the neural systems primarily serving speech production more than do fluent readers (IV).



# Acknowledgements

This thesis was done at the Laboratory of Computational Engineering (LCE) at Helsinki University of Technology. I thank LCE Professors Mikko Sams and Iiro Jääskeläinen for welcoming me to the laboratory and providing an inspiring atmosphere, excellent facilities, and important financial support for this work. I also wish to thank Professor Jouko Lampinen, head of the Laboratory, and Professors Kimmo Kaski and Jukka Tulkki for their efforts in building and developing this unique multidisciplinary research environment. I thank Professor Leena Kivisaari, Department of Diagnostic Radiology at Helsinki University Central Hospital, for her encouraging attitude and opportunities to focus on research while still keeping in touch with clinical work. I am indebted for funding to the National Institutes of Health (NIH-RO1 HD040712), the Academy of Finland (grants number 202871, 200521), the Kordelin Foundation, the state subsidy for University Hospitals, and the Memorial Fund for Pehr Oscar Klingendahl.

fMRI measurements for this work were done at the Advanced Magnetic Imaging (AMI) Centre at Helsinki University of Technology. I wish to thank Professor Riitta Hari, the scientific director of the AMI Centre, for her efforts in establishing and managing this high-standard research unit. I am also grateful to Raimo Joensuu, Veikko Jousmäki, Marita Kattelus, Antti Tarkiainen, Simo Vanni, and other AMI Centre researchers and personnel (both past and present) who contributed to the setting up the experimental fMRI laboratory and made it all work.

I am truly grateful to my supervisors Iiro Jääskeläinen and Taina Autti for their never-failing support as I toddled my way from clinical medicine to functional neuroscience. I thank Kimmo Alho, Alexander Degerman, Teija Kujala, Marja Laasonen, Riikka Möttönen, Ville Ojanen, Teemu Rinne, and Antti Tarkiainen, my co-authors at LCE, AMI, and the Department of Psychology at the University of Helsinki: They most generously contributed their time, skills, and knowledge during all phases of this work. Special thanks for *sisu* and optimism go to Ville Ojanen, who shared with me the delight and despair of getting started with fMRI. Thanks also to all the other researchers at LCE with whom I had the joy to work: Tobias Andersen, Toni Auranen, Michael Frydrych,

Aapo Nummenmaa, Laura Kauhanen, Jaakko Kauramäki, Jari Kätsyri, Janne Lehtonen, Iina Tarnanen, Kaisa Malkamäki, Kaisa Tiippana, Vasily Klucharev, and many others. I also thank Eeva Lampinen at LCE and Marja Riihimäki at Helsinki Medical Imaging Center for their help with various practical and bureaucratic matters.

I am sincerely grateful to my official reviewers Jyrki Ahveninen and Ritva Vanninen, whose insightful criticism improved this thesis beyond what I imagined likely.

With all my heart I thank my family and friends for their love and encouragement, especially my dear mom Maija whose help I could always count on and my beloved husband Marko who can create fun and magic during even the gloomiest days. Last but not least, I thank Carol Norris, my mentor and language editor par excellence, for teaching me everything I know about scientific writing.

I dedicate this book to Marko, the love of my life since age 19, and our precious little daughter Pihla.

Nurmijärvi, June 2006

A handwritten signature in black ink, consisting of several loops and a long horizontal stroke extending to the right.

Johanna Pekkola

## References

- Alho K, Lavikainen J, Reinikainen K, Sams M, Näätänen R (1990): Event-related brain potentials in selective listening to frequent and rare stimuli. *Psychophysiology* 27:73-86.
- Alho K, Medvedev SV, Pakhomov SV, Roudas MS, Tervaniemi M, Reinikainen K, Zeffiro T, Näätänen R (1999): Selective tuning of the left and right auditory cortices during spatially directed attention. *Brain Res Cogn Brain Res* 7:335-341.
- Arbib M and Bota M (2003): Language evolution: neural homologies and neuroinformatics. *Neural Netw* 16:1237-1260.
- Baldissera F, Cavallari P, Craighero L, Fadiga L (2001): Modulation of spinal excitability during observation of hand actions in humans. *Eur J Neurosci* 13:190-194.
- Bandettini PA (1999): The temporal resolution of functional MRI. In: Moonen CTW and Bandettini PA (Eds.): *Functional MRI*, pp. 205-220. Springer, New York.
- Barracough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005): Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci* 17:377-391.
- Bavelier D, Neville HJ (2002): Cross-modal plasticity: where and how? *Nat Rev Neurosci* 3:443-452.
- Belin P, Zatorre RJ, Ahad P (2002): Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res* 13:17-26.
- Benson DF, Sheremata WA, Bouchard R, Segarra JM, Price D, Gerschwind N (1973): Conduction aphasia. A clinicopathological study. *Arch Neurol* 28:339-346.
- Benson RR, Whalen DH, Richardson M, Swainson B, Clark VP, Lai S, Liberman AM (2001): Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang* 78:364-396.
- Bernstein LE, Auer ET Jr, Moore JK, Ponton CW, Don M, Singh M (2002): Visual speech perception without primary auditory cortex activation. *Neuroreport* 13:311-315.
- Bernstein LE, Auer ET Jr, Moore JK (2004): Audiovisual speech binding: Convergence or association. In: Calver GA, Spence C, Stein BE (Eds.): *Handbook of multisensory processes*, pp 203-224. MIT Press, Cambridge, MA.
- Besle J, Fort A, Delpuech C, Giard MH (2004): Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 20:2225-2234.
- Binder JR, Rao SM, Hammeke TA, Frost JA, Bandettini PA, Hyde JS (1994): Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Brain Res Cogn Brain Res* 2:31-38.
- Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW (1996): Function of the left planum temporale in auditory and linguistic processing. *Brain* 119:1239-1247.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET (2000): Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512-528.
- Blamire AM, Ogawa S, Ugurbil K, Rothman D, McCarthy G, Ellermann JM, Hyder F, Rattner Z, Shulman RG (1992): Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. *Proc Natl Acad Sci U S A* 89:11069-11073.
- Bookheimer S (2002): *Functional MRI of language: new approaches to understanding the*

- cortical organization of semantic processing. *Annu Rev Neurosci* 25:151-188.
- Broadbent DE (1958): Perception and communication. Pergamon, London.
- Broca P (1861): Remarques sur le siège de la faculté du langage articulé: suivies d'une observation d'aphemia. *Bull Soc Anat Paris* 6:330-357.
- Brodman K (1909): Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Barth JA, Leipzig.
- Brosch M, Selezneva E, Scheich H (2005): Nonauditory events of a behavioural procedure activate auditory cortex of highly trained monkeys. *J Neurosci* 20:6797-6806.
- Brugge JF, Merzenich MM (1973): Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J Neurophysiol* 36:1138-1158.
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ (2001): Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci* 13:400-404.
- Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF (2005): Reading, hearing, and the planum temporale. *NeuroImage* 24:444-454.
- Burton H and Jones EG (1976): The posterior thalamic region and its cortical projection in New World and Old World monkeys. *J Comp Neurol* 168:249-301.
- Burton H and Sinclair R (1996): Somatosensory cortex and tactile perceptions. In: Kruger L, Friedman M, Cartelette E (Eds.): Pain and Touch, pp. 105-162. Academic Press, San Diego, CA.
- Burton MW, Small SL, Blumstein SE (2000): The role of segmentation in phonological processing: an fMRI investigation. *J Cogn Neurosci* 12:679-690.
- Buxton RB, Wong EC, Frank LR (1999): The post-stimulus undershoot of the functional MRI signal. In: Moonen CTW and Bandettini PA (Eds.): Functional MRI, pp. 253-262. Springer, New York.
- Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001): Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. *Brain Res Cogn Brain Res* 10:349-353.
- Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E (2003a): Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 17:2213-2218.
- Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R (2003b): Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage* 19:113-124.
- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004a): Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J Cogn Neurosci* 16:805-816.
- Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004b): Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage* 22:1182-1194.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1997): Activation of auditory cortex during silent lipreading. *Science* 276:593-596.

- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999): Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619-2623.
- Calvert GA, Campbell R, Brammer MJ (2000): Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649-657.
- Calvert GA (2001): Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110-1123.
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001): Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage* 14:427-438.
- Calvert GA and Campbell R (2003): Reading speech from still and moving faces: the neural substrates of visible speech. *J Cogn Neurosci* 15:57-70.
- Campbell R, MacSweeney M, Surguladze S, Calvert GA, McGuire P, Suckling J, Brammer MJ, David AS (2001): Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res Cogn Brain Res* 12:233-243.
- Celesia GG (1976): Organization of auditory cortical areas in man. *Brain* 99:403-414.
- Chen W, Kato T, Zhu XH, Strupp J, Ogawa S, Ugurbil K (1998): Mapping of lateral geniculate nucleus activation during visual stimulation in human brain using fMRI. *Magn Res Med* 39:89-96.
- Chen W and Ogawa S (1999): Principles of BOLD functional MRI. In: Moonen CTW and Bandettini PA (Eds.): *Functional MRI*, pp. 103-113. Springer, New York.
- Cochin S, Barthelemy C, Lejeune B, Roux S, Martineau J (1998) Perception of motion and qEEG activity in human adults. *Electroencephalogr Clin Neurophysiol* 107:287-295.
- Cohen MS (1999): Echo-planar imaging and functional MRI. In: Moonen CTW and Bandettini PA (Eds.): *Functional MRI*, pp. 137-148. Springer, New York.
- Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P (2002): Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clin Neurophysiol* 113:495-506.
- Colin C, Radeau M, Soquet A, Deltenre P (2004): Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. *Clin Neurophysiol* 11:1989-2000.
- Corina DP, Richards TL, Serafini S, Richards AL, Steury K, Abbott RD, Echelard DR, Maravilla KR, Berninger VW (2001): fMRI auditory language differences between dyslexic and able reading children. *Neuroreport* 12:1195-1201.
- Cornelissen PL, Hansen PC, Hutton JL, Evangelineou V, Stein JF (1998): Magnocellular visual function and children's single word reading. *Vision Res* 38:471-482.
- Creutzfeldt O, Ojemann G, Lettich E (1989): Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Exp Brain Res* 77:451-475.
- Cusick CG (1997): The superior temporal polysensory region in monkeys. In: Rockland KS, Kaas JH, Peters A (Eds.): *Cerebral cortex*, pp. 435-468. Plenum, New York.
- Damasio H and Damasio AR (1980): The anatomical basis of conduction aphasia. *Brain* 103:337-350.
- Decety J, Chaminade T, Grèzes J, Meltzoff AN (2002): A PET exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage* 15:265-272.
- Degerman A, Rinne T, Salmi J, Salonen O, Alho K (2006): Selective attention to sound location or pitch studied with fMRI. *Brain Res* 1077:123-134.

- Demb JB, Boynton GM, Heeger DJ (1997): Brain activity in visual cortex predicts individual differences in reading performance. *Proc Natl Acad Sci U S A* 94:13363-13366.
- Demonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulos JL, Wise R, Rascol A, Frackowiak R (1992): The anatomy of phonological and semantic processing in normal subjects. *Brain* 115:1753-1768.
- DeYoe EA, Bandettini P, Neitz J, Miller D, Winans P (1994): Functional magnetic resonance imaging (fMRI) of the human brain. *J Neurosci Methods* 54:171-187.
- Dhankhar A, Wexler BE, Fulbright RK, Halwes T, Blamire AM, Shulman RG (1997): Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rates. *J Neurophysiol* 77:476-483.
- Diehl RL, Lotto AJ, Holt LL (2004): Speech perception. *Annu Rev Psychol* 55:149-179.
- Eden GE, VanMeter JW, Rumsey JM, Maisog JM, Woods RP, Zeffiro TA (1996): Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. *Nature* 382:66-69.
- Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999): Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp* 7: 89-97.
- Fadiga L, Fogassi L, Pavesi G, Rizzolatti G (1995): Motor facilitation during action observation: a magnetic stimulation study. *J Neurophysiol* 73:2608-2611.
- Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002): Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur J Neurosci* 15:399-402.
- Fadiga L and Craighero L (2003): New insights on sensorimotor integration: from hand action to speech perception. *Brain Cogn* 53:514-524.
- Falchier A, Clavagnier S, Barone P, Kennedy H (2002): Anatomical evidence of multimodal integration in primate striatal cortex. *J Neurosci* 22:5749-5759.
- Ferrari PF, Gallese V, Rizzolatti G, Fogassi L (2003): Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur J Neurosci* 17:1703-1714.
- Finney EM, Fine I, Dobkins KR (2001): Visual stimuli activate the auditory cortex in the deaf. *Nat Neurosci* 4:1171-1173.
- Fischl B, Sereno MI, Tootell RB, Dale AM (1999): High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272-284.
- Felleman DJ and van Essen DC (1991): Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1-47.
- Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005): Parietal lobe: from action organization to intention understanding. *Science* 308:662-667.
- Fort A, Delpuech C, Pernier J, Giard MH (2002a): Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex* 12:1031-1039.
- Fort A, Delpuech C, Pernier J, Giard MH (2002b): Early auditory-visual interactions in human cortex during nonredundant target identification. *Brain Res Cogn Brain Res* 14:20-30.
- Fort A, Delpuech C, Pernier J, Giard MH (2002c): Dynamics of cortico-cortical cross-modal operations involved in audiovisual object recognition in humans. *Cereb Cortex* 12:1031-1039.
- Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W,

- Murray MM (2002): Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J Neurophysiol* 88:540-543.
- Friston KJ, Frith CD, Dolan RJ, Mazziotta JC, Frackowiak RSJ (1997): Human brain function, p. 105. Academic Press, San Diego, CA.
- Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC (1994): Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1:210-220.
- Frith CD, Friston KJ (1996): The role of the thalamus in “top down” modulation of attention to sound. *NeuroImage* 4:210-215.
- Frith U (2001): What framework should we use for understanding developmental disorders? *Dev Neuropsychol* 20:555-563.
- Fritz J, Shamma S, Elhilali M, Klein D (2003): Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216-1223.
- Fu KG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2003): Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci* 23:7510-7515.
- Fu KM, Shah AS, O’Connell MN, McGinnis MC, Eckholt H, Lakatos P, Smiley J, Schroeder CE (2004): Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. *J neurophysiol* 92:3522-3531.
- Galaburda A, Sanides F (1980): Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* 190:597-610.
- Galaburda AM (1982): Histology, architectonics, and asymmetry of language areas. In: Arbib MA, Caplan D, Marshall JC (Eds.): *Neural models of language processes*, pp. 435-445. Academic Press, San Diego.
- Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996): Action recognition in the premotor cortex. *Brain* 119:593-609.
- Gangitano M, Mottaghy FM, Pascual-Leone A (2001): Phase-specific modulation of cortical motor output during movement observation. *Neuroreport* 12:1489-1492.
- Gati JS, Menon RS, Ugurbil K, Rutt BK (1997): Experimental determination of the BOLD field strength dependence in vessels and tissue. *Magn Reson Med* 38:296-302.
- Gati JS, Menon RS, Rutt BK (1999): Field strength dependence of functional MRI signals. In: Moonen CTW and Bandettini PA (Eds.): *Functional MRI*, pp. 277-282. Springer, New York.
- de Gelder B and Wroomen J (1998): Impaired speech perception in poor readers: evidence from hearing and speech reading. *Brain Lang* 64:269-281.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005): Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004-5012.
- Giard M, Peronnet F (1999): Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11: 473-490.
- Godfrey JJ, Syrdal-Lasky AK, Millay KK, Knox CM (1981): Performance of dyslexic children on speech perception tests. *J Exp Child Psychol* 32, 401-424.
- Golestani N and Zatorre RJ (2004): Learning new sounds of speech: reallocation of neural substrates. *NeuroImage* 21:494-506.

- Grady CL, Van Meter JW, Maisog JM, Pietrini P, Krasuski J, Rauschecker JP (1997): Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport* 8:2511-2516.
- Grafton ST, Arbib MA, Fadiga L, Rizzolatti G (1996): Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp Brain Res* 112:103-111.
- Grèzes J and Decety J (2001): Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum Brain Mapp* 12:1-19.
- Grèzes J, Armony JL, Rowe J, Passingham RE (2003): Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study. *NeuroImage* 18:928-937.
- Griffiths TD, Warren JD (2002): The planum temporale as a computational hub. *Trends Neurosci* 25:348-353.
- Haacke EM, Brown RW, Thompson MR, Venkatesan R (1999): A brief overview of MRI concepts. In: *Magnetic resonance imaging: physical principles and sequence design*, pp. 3-13. Wiley, New York.
- Habib M (2000): The neurological basis of developmental dyslexia. An overview and working hypothesis. *Brain* 123:2373-2399.
- Hackett TA, Stepniewska I, Kaas JH (1998): Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394:475-495.
- Hackett TA, Preuss TM, Kaas JH (2001): Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441:197-222.
- Hall DA, Summerfield AQ, Goncalves MS, Foster JR, Palmer AR, Bowtell RW (2000): Time-course of the auditory BOLD response to scanner noise. *Magn Reson Med* 43:601-606.
- Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield Q (2002): Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 12:140-149.
- Hansen JC and Hillyard SA (1988): Temporal dynamics of human auditory selective attention. *Psychophysiology* 25:316-329.
- Hari R, Hämäläinen M, Kaukoranta E, Mäkelä J, Joutsiniemi SL, Tiihonen J (1989): Selective listening modifies activity of the human auditory cortex. *Exp Brain Res* 74:463-470.
- Hari R and Kiesilä P (1996): Deficits of temporal auditory processing in dyslexic adults. *Neurosci Lett* 205:138-140.
- Hari R, Forss N, Avikainen S, Kirveskari E, Salenius S, Rizzolatti G (1998): Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proc Natl Acad Sci U S A* 95:15061-15065.
- Hari R and Renvall H (2001): Impaired processing of rapid stimulus sequences in dyslexia. *Trends Cogn Sci* 5:525-532.
- Harms MP and Melcher JR (2002): Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol* 88:1433-1450.
- Hasegawa T, Matsuki KI, Ueno T, Maeda Y, Matsue Y, Konishi Y, Sadato N (2004): Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Brain Res Cogn Brain Res*

- 20:510-518.
- Hashimoto R, Homae F, Nakajima K, Miyashita Y, Sakai KL (2000): Functional differentiation in the human auditory and language areas revealed by a dichotic listening task. *NeuroImage* 12:147-158.
- Hayes EA, Tiippana K, Nicol TG, Sams M, Kraus N (2003): Integration of heard and seen speech: a factor in learning disabilities in children. *Neurosci Lett* 351:46-50.
- Helenius P, Tarkiainen A, Cornelissen P, Hansen PC, Salmelin R (1999): Dissociation of normal feature analysis and deficient processing of letter-strings in dyslexic adults. *Cereb Cortex* 9:476-483.
- Helenius P, Salmelin R, Richardson U, Leinonen S, Lyytinen H, (2002a): Abnormal auditory cortical activation in dyslexia 100 msec after speech onset. *J Cogn Neurosci* 14:603-617.
- Helenius P, Salmelin R, Service E, Connolly JF, Leinonen S, Lyytinen H (2002b): Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J Neurosci* 22:2936-2944.
- Herholz K, Thiel A, Wienhard K, Pietrzyk U, von Stockhausen HM, Karbe H, Kessler J, Bruckbauer T, Halber M, Weiss WD (1996): Individual functional anatomy of verb generation. *NeuroImage* 3:185-194.
- Hershenson M (1962): Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63:289-293.
- Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, Strick PL, Ugurbil K (2000): A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci Lett* 287:156-160.
- Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003): Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci* 15:673-682.
- Hickok G and Poeppel D (2004): Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67-99.
- Hillyard SA, Hink RF, Schwent VL, Picton TW (1973): Electrical signs of selective attention in the human brain. *Science* 182:177-180.
- Horwitz B, Rumsey JM, Donohue BC (1998): Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc Natl Acad Sci U S A* 95:8939-8944.
- Howard MA, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA (1996): A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 724:260-264.
- Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Steinschneider M, Reale RA, Hind JE, Brugge JF (2000): Auditory cortex on the human posterior superior temporal gyrus. *J Comp Neurol* 416:79-92.
- Howard RJ, Brammer M, Wright I, Woodruff PW, Bullmore ET, Zeki S (1996): A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Curr Biol* 6:1015-1019.
- Hu X, Yacoub E, Le TH, Cohen ER, Ugurbil K (1999): Functional MRI signal decrease at the onset of stimulation. In: Moonen CTW and Bandettini PA (Eds.): *Functional MRI*, pp. 243-252. Springer, New York.
- Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G (1999): Cortical mechanisms of human imitation. *Science* 286:2526-2528.

- Jellema T, Baker CI, Wicker B, Perrett DI (2000): Neural representation for the perception of the intentionality of actions. *Brain Cogn* 44:280-302.
- Jenkinson M, Bannister P, Brady M, Smith S (2002): Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17:825-841.
- Joanisse MF and Gati JS (2003): Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage* 19:64-79.
- Jäncke L, Mirzazade S, Shah NJ (1999): Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci Lett* 266:125-128.
- Jäncke L, Buchanan TW, Lutz K, Shah NJ (2001): Focused and nonfocused attention in verbal and emotional dichotic listening: an fMRI study. *Brain Lang* 78:349-363.
- Jäncke L, Wüstenberg T, Sceich H, Heinze HJ. (2002): Phonetic perception and the temporal cortex. *NeuroImage* 15:733-746.
- Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW (2004a): Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci U S A* 101:6809-6814.
- Jääskeläinen IP, Ojanen V, Ahveninen J, Auranen T, Levänen S, Möttönen R, Tarnanen I, Sams M (2004b): Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *Neuroreport* 15:2741-2744.
- Kastner S and Ungerleider LG (2000): Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci* 23:315-341.
- Klingberg T, Hedehus M, Temple E, Salz T, Gabrieli JDE, Moseley ME, Poldrack RA (2000): Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron* 25:493-500.
- Klinke R (1989): Physiology of hearing. In: Schmidt RF and Thews G (Eds.): *Human physiology*, pp. 283-298. Springer-Verlag, Berlin Heidelberg.
- Klucharev V, Möttönen R, Sams M (2003): Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Res Cogn Brain Res* 18:65-75.
- Kohler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G (2002): Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297: 846-848.
- Koski L, Wohlschläger A, Bekkering H, Woods RP, Dubeau MC, Mazziotta JC, Iacoboni M (2002): Modulation of motor and premotor activity during imitation of target-directed actions. *Cereb Cortex* 12:847-855.
- Koski L, Iacoboni M, Dubeau MC, Woods RP, Mazziotta JC (2003): Modulation of cortical activity during different imitative behaviors. *J Neurophysiol* 89:460-471.
- Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB (1996): Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 273:971-973.
- Kujala T, Myllyviita K, Tervaniemi M, Alho K, Kallio J, Näätänen R (2000): Basic auditory dysfunction in dyslexia as demonstrated by brain activity measurements. *Psychophysiology* 37:262-266.

- Laasonen M, Tomma-Halme J, Lahti-Nuutila P, Service E, Virsu V (2000): Rate of information segregation in developmentally dyslexic children. *Brain Lang* 75:66-81.
- Laasonen M, Service E, Virsu V (2002): Crossmodal temporal order and processing acuity in developmentally dyslexic young adults. *Brain Lang* 80:340-354.
- Lange N (1999): Statistical procedures for functional MRI. In: Moonen CTW and Bandettini PA (Eds.): *Functional MRI*, pp. 301-335. Springer, New York.
- Lebib R, Papo D, de Bode S, Baudonniere PM (2003): Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neurosci Lett* 341:185-188.
- Levänen S, Jousmäki V, Hari R (1998): Vibration-induced auditory-cortex activation in a congenitally deaf adult. *Curr Biol* 16:869-872.
- Lewis JW, Van Essen DC (2000): Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J Comp Neurol* 428:112-137.
- Lieberman AM, Delattre PC, Cooper FS, Gerstman LJ (1954): The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs* 68:1-13.
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967): Perception of the speech code. *Psychol Rev* 74:431-461.
- Lieberman AM, Mattingly IG (1985): The motor theory of speech perception revised. *Cognition* 21:1-36.
- Liegeois-Chauvel C, Musolino A, Chauvel P (1991): Localization of the primary auditory area in man. *Brain* 114:139-151.
- Lipschutz B, Kolinsky R, Damhaut P, Wikler D, Goldman S (2002): Attention-dependent changes of activation and connectivity in dichotic listening. *NeuroImage* 17:643-656.
- Lovegrove WJ, Bowling A, Badcock D, Blackwood M (1980): Specific reading disability: differences in contrast sensitivity as a function of spatial frequency. *Science* 210:439-440.
- Ludman CN, Summerfield AQ, Hall D, Elliott M, Foster J, Hykin JL, Bowtell R, Morris PG (2000): Lip-reading ability and patterns of cortical activation studied using fMRI. *Br J Audiol* 34:225-230.
- Macaluso E, George N, Dolan R, Spence C, Driver J (2004): Spatial and temporal factors during processing of audiovisual speech: a PET study. *NeuroImage* 21:725-732.
- MacSweeney M, Amaro E, Calvert GA, Campbell R, David AS, McGuire P, Williams SCR, Woll B, Brammer MJ (2000): Silent speechreading in the absence of scanner noise: an event-related fMRI study. *Neuroreport* 11:1729-1733.
- MacSweeney M, Woll B, Campbell R, McGuire PK, David AS, Williams SCR, Suckling J, Calvert GA, Brammer MJ (2002a): Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain* 125:1583-1593.
- MacSweeney M, Calvert GA, Campbell R, McGuire PK, David AS, Williams SCR, Woll B, Brammer MJ (2002b): Speechreading circuits in people born deaf. *Neuropsychologia* 40:801-807.
- MacSweeney M, Campbell R, Woll B, Giampietro V, David AS, McGuire PK, Calvert GA, Brammer MJ (2004): Dissociating linguistic and non-linguistic gestural communication in the brain. *NeuroImage* 22:1605-1618.
- Maeda F, Kleiner-Fisman G, Pasqual-Leone A (2002): Motor facilitation while observing

- hand actions: specificity of the effect and the role of observer's orientation. *J Neurophysiol* 87:1329-1335.
- Manis FR, McBride-Chang C, Seidenberg MS, Keating P, Doi LM, Munson B, Petersen A (1997): Are speech perception deficits associated with developmental dyslexia? *J Exp Child Psychol* 66:211-235.
- Manthey S, Schubotz RI, von Cramon DY (2003): Premotor cortex in observing erroneous action: an fMRI study. *Brain Res Cogn Brain Res* 15:296-307.
- Martin F and Lovegrove W (1987): Flicker contrast sensitivity in normal and specifically disabled readers. *Perception* 16:215-221.
- Mattingly IG, Liberman AM, Syrdal AM, Halwes T (1971): Discrimination in speech and non-speech modes. *Cogn Psych* 2:131-157.
- McAnally KI, Hansen PC, Cornelissen PL, Stein JF (1997): Effect of time and frequency manipulation on syllable perception in developmental dyslexics. *J Speech Lang Hear Res* 40:912-924.
- McCrary E, Frith U, Brunswick N, Price C (2000): Abnormal functional activation during a simple word repetition task: A PET study of adult dyslexics. *J Cogn Neurosci* 12:753-762.
- McGurk H and MacDonald J (1976): Hearing lips and seeing voices. *Nature* 264:746-748.
- Mechelli A, Friston KJ, Price CJ (2000): The effects of presentation rate during word and pseudoword reading: a comparison of PET and fMRI. *Cogn Neurosci* 12:145-156.
- Merzenich MM and Brugge JF (1973): Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res* 50:275-296.
- Mesulam MM, Pandya DN (1973): The projections of the medial geniculate complex within the sylvian fissure of the rhesus monkey. *Brain Res* 60:315-333.
- Mody M, Studdert-Kennedy M, Brady S (1997): Speech perception deficits in poor readers: auditory processing or phonological coding? *J Exp Child Psychol* 64, 199-231.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Joxe JJ (2002): Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115-128.
- Morel A, Garraghty PE, Kaas JH (1993): Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol* 335:437-459.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001): Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* 13:648-701.
- Morrell LK (1968): Temporal characteristics of sensory interaction in choice reaction times. *J Exp Psychol* 77:14-18.
- Musacchia GAE, Sams M, Nicol TG, Kraus N (2006): Seeing speech affects acoustic information processing in the human brainstem. *Exp Brain Res* 168:1-10.
- Möttönen R, Krause CM, Tiippana K, Sams M (2002): Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res* 13:417-425.
- Möttönen R, Schürmann M, Sams M (2004): Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neurosci Lett* 363:112-115.
- Möttönen R, Calvert GA, Jääskeläinen IP, Matthews PM, Thesen T, Tuomainen J, Sams M (2006): Perceiving identical sounds as speech or non-speech modulates

- activity in the left posterior superior temporal sulcus. *NeuroImage* 30:563-569.
- Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM (1999): Cortical auditory signal processing in poor readers. *Proc Natl Acad Sci U S A* 96:6483-6488.
- Nakai T, Matsuo K, Kato C, Matsuzava M, Okada T, Glover GH, Moriya T, Inui T (1999): A functional magnetic resonance imaging study of listening comprehension of languages in human at 3 tesla - comprehension level and activation of the language areas. *Neurosci Lett* 263:33-36.
- Narain C, Scott SK, Wise RJS, Rosen S, Leff A, Iversen SD, Matthews PM (2003): Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb Cortex* 13:1362-1368.
- Nishimura H, Hashikawa K, Doi K, Iwaki T, Watanabe Y, Kusuoka H, Nishimura T, Kubo T (1999): Sign language 'heard' in the auditory cortex. *Nature* 397:116.
- Nishitani N and Hari R (2000): Temporal dynamics of cortical representation for action. *Proc Natl Acad Sci U S A* 97: 913-918.
- Nishitani N, Hari R (2002): Viewing lip forms: cortical dynamics. *Neuron* 36:1211-1220.
- Nishitani N, Schürmann M, Amunts K, Hari R (2005): Broca's region: from action to language. *Physiology* 20:60-69.
- Nittrouer S (1999): Do temporal processing deficits cause phonological processing problems? *J Speech Lang Hear Res* 42:925-942.
- Nobre AC (2001): The attentive homunculus: now you see it, now you don't. *Neurosci Beh Rev* 25:477-496.
- Näätänen R, Gaillard AW, Mäntysalo S (1978): Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst)* 42:313-329.
- Näätänen R (1982): Processing negativity: an evoked-potential reflection of selective attention. *Psychol Bull* 92:605-640.
- Näätänen R (2001): The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38:1-21.
- Ogawa S, Lee TM, Kay AR, Tank DW (1990): Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A* 87:9868-9872.
- Ogawa S, Menon RS, Tank DW, Kim SG, Merkle H, Ellermann JM, Ugurbil K (1993): Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys J* 64:803-812.
- Ogawa S, Lee TM, Stepnoski R, Chen W, Zhu XH, Ugurbil K (2000): An approach to probe some neural systems interaction by functional MRI at neural time scale down to milliseconds. *Proc Natl Acad Sci U S A* 97:11026-11031.
- Ohala JJ (1996): Speech perception is hearing sounds, not tongues. *J Acoust Soc Am* 99:1718-1725.
- Ojanen V, Möttönen R, Pekkola J, Jääskeläinen IP, Joensuu R, Autti T, Sams M (2005): Processing of audiovisual speech in Broca's area. *NeuroImage* 25:333-338.
- O'Leary DS, Andreasen NC, Hurtig RR, Hichwa RD, Watkins GL, Boles Ponto LL, Rogers M, Kirchner PT (1996): A positron emission tomography study of binaurally and dichotically presented stimuli: effects of level of language and directed attention. *Brain Lang* 53:20-39.
- Olson IR, Gatenby JC, Gore JC (2002): A comparison of bound and unbound audiovisual information processing in the human cerebral cortex. *Brain Res Cogn*

- Brain Res 14:129-138.
- Pandya DN, Hallett M, Kmukherjee SK (1969): Intra- and interhemispheric connections of the neocortical auditory system in the rhesus monkey. *Brain Res* 14:49-65.
- Pandya DN, Sanides F (1973): Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern. *Z Anat Entwicklungsgesch* 139:127-161.
- Patuzzo S, Fiaschi A, Manganotti P (2003): Modulation of motor cortex excitability in the left hemisphere during action observation: a single- and paired-pulse transcranial magnetic stimulation study of self- and non-self-action observation. *Neuropsychologia* 41:1272-1278.
- Paulesu E, Frith U, Snowling M, Gallagher A, Morton J, Frackowiak RSJ, Frith CD (1996): Is developmental dyslexia a disconnection syndrome? *Brain* 119:143-157.
- Paulesu E, Demonet JF, Fazio F, McCrory E, Chanoine V, Brunswick N, Cappa SF, Cossu G, Habib M, Frith CD, Frith U (2001): Dyslexia: cultural diversity and biological unity. *Science* 291:2165-2167.
- Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F (2003): A functional-anatomical model for lipreading. *J Neurophysiol* 90:2005-2013.
- di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992): Understanding motor events: a neurophysiological study. *Exp Brain Res* 91:176-180.
- Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996): Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex* 6:661-672.
- Perani D, Fazio F, Borghese NA, Tettamanti M, Ferrari S, Decety J, Gilardi MC (2001): Different brain correlates for watching real and virtual hand actions. *NeuroImage* 14:749-758.
- Perrett DI, Harries MH, Bevan R, Thomas S, Benson PJ, Mistlin AJ, Chitty AJ, Hietanen JK, Ortega JE (1989): Frameworks of analysis for the neural representation of animate objects and actions. *J Exp Biol* 146:87-113.
- Petitto LA, Zatorre RJ, Gauna K, Nikelski EJ, Dostie D, Evans AC (2000): Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proc Natl Acad Sci U S A* 97:13961-13966.
- Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004): Attentional modulation of human auditory cortex. *Nat Neurosci* 7:658-663.
- Petrides M and Pandya DN (1994): Comparative architectonic analysis of the human and the macaque frontal cortex. In: Boller F and Grafman J (Eds): *Handbook of Neuropsychology* Vol 9, pp. 17-58. Elsevier, Amsterdam.
- Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T (2000): Auditory cortex accesses phonological categories: an MEG mismatch study. *J Cogn Neurosci* 12:1038-1055.
- Pourtois G, de Gelder B, Vroomen J, Rossion B, Crommelinck M (2000): The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport* 27:1329-1333.
- Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RS, Friston KJ (1996): Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain* 119:919-931.
- Pugh KR, Shaywitz BA, Shaywitz SE, Fulbright RK, Byrd D, Skudlarski P, Shankweiler

- DP, Katz L, Constable RT, Fletcher J, Lacadie C, Marchione K, Gore JC (1996): Auditory selective attention: an fMRI investigation. *NeuroImage* 4:159-173.
- Rademacher J, Caviness VS Jr, Steinmetz H, Galaburda AM (1993): Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping and neurobiology. *Cereb Cortex* 3:313-329.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001): Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* 13:669-683.
- Raij T, Uutela K, Hari R (2000): Audiovisual integration of letters in the human brain. *Neuron* 28:617-625.
- Rauschecker JP, Tian B, Hauser M (1995): Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111-114.
- Rauschecker JP, Tian B, Pons T, Mishkin M (1997): Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol* 382:89-103.
- Rauschecker JP and Tian B (2000): Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800-11806.
- Rauschecker JP and Tian B (2004): Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91:2578-2589.
- Reed MA (1989): Speech perception and the discrimination of brief auditory cues in reading disabled children. *J Exp Child Psychol* 48:270-292.
- Rees G, Howseman A, Josephs O, Frith CD, Friston KJ, Frackowiak RSJ, Turner R (1997): Characterizing the relationship between BOLD contrast and regional cerebral blood flow measurements by varying the stimulus presentation rate. *NeuroImage* 6:270-278.
- Renvall H and Hari R (2002): Auditory cortical responses to speech-like stimuli in dyslexic adults. *J Cogn Neurosci* 14:757-768.
- Rif J, Hari R, Hämäläinen MS, Sams M (1991): Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalogr Clin Neurophysiol* 79:464-472.
- Rinne T, Alho K, Alku P, Holi M, Sinkkonen J, Virtanen J, Bertrand O, Näätänen R (1999): Analysis of speech sounds is left-hemisphere predominant at 100-150ms after sound onset. *Neuroreport* 10:1113-1117.
- Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996a): Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res* 3:131-141.
- Rizzolatti G, Fadiga L, Matelli M, Bettinardi V, Paulesu E, Perani D, Fazio F (1996b): Localization of grasp representations in humans by PET: 1. Observation versus execution. *Exp Brain Res* 111:246-252.
- Rizzolatti G, Arbib MA (1998): Language within our grasp. *Trends Neurosci* 21:188-194.
- Rizzolatti G, Fogassi L, Gallese V (2001): Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2:661-670.
- Rizzolatti G, Craighero L (2004): The mirror-neuron system. *Annu Rev Neurosci* 27:169-192.
- Rockland KS and Pandya DN (1979): Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res* 179:3-20.
- Rockland KS, Ojima K (2001): Calcarine area V1 as a multimodal convergence area. *Soc Neurosci (abstract)* 27:program no 511.20.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999a): Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131-1136.

- Romanski LM, Bates JF, Goldman-Rakic PS (1999b): Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 403:141-157.
- Romanski LM, Goldman-Rakic PS (2002): An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5:15-16.
- Rumsey JM, Nace K, Donohue B, Wise D, Maisog JM, Andreason P (1997): A positron emission tomographic study of impaired word recognition and phonological processing in dyslexic men. *Arch Neurol* 54:562-573.
- Salmelin R, Service E, Kiesilä P, Uutela K, Salonen O (1996): Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Ann Neurol* 40:157-162.
- Salmelin R and Helenius P (2004): Functional neuroanatomy of impaired reading in dyslexia. *Scientific Studies of Reading* 8:257-272.
- Sams M, Aulanko R, Hämäläinen M, Hari R, Lounasmaa OV, Lu ST, Simola J (1991): Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett* 127:141-145.
- Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001): Somatosensory input to auditory association cortex in the macaque monkey. *J Neurophysiol* 85:1322-1327.
- Schroeder CE, Foxe JJ (2002): The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res* 14:187-198.
- Schulte-Körne G, Deimel W, Bartling J, Remschmidt H (1998): Auditory processing and dyslexia: evidence for a specific speech processing deficit. *Neuroreport* 9:337-340.
- Scott SK, Blank CC, Rosen S, Wise RJS (2000): Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400-2406.
- Scott SK, Johnsrude IS (2003): The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100-107.
- Sekiyama K, Kanno I, Miura S, Sugita Y (2003): Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res* 47:277-278.
- Seldon HL (1985): The anatomy of speech perception. Human auditory cortex. In: Peters A, Jones EG (Eds.): *Cerebral Cortex Vol IV: Association and auditory cortices*, pp. 273-327. Plenum, New York.
- Seltzer B, Pandya DN (1994): Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J Comp Neurol* 343:445-463.
- Shaywitz SE, Shaywitz BA, Pugh KR, Fulbright RK, Constable RT, Mencl WE, Shankweiler DP, Liberman AM, Skudlarski P, Fletcher JM, Katz L, Marchione KE, Lacadie C, Gatenby C, Gore JC (1998): Functional disruption in the organization of the brain for reading in dyslexia. *Proc Nat Acad Sci U.S.A.* 95:2636-2641.
- Simos PG, Breier JJ, Fletcher JM, Bergman E, Papanicolaou AC (2000a): Cerebral mechanisms involved in word reading in dyslexic children: a magnetic source imaging approach. *Cereb Cortex* 10:809-816.
- Simos PG, Breier JJ, Fletcher JM, Foorman BR, Bergman E, Fishbeck K, Papanicolaou AC (2000b): Brain activation profiles in dyslexic children during non-word reading: a magnetic source imaging study. *Neurosci Lett* 290:61-65.

- Smith SM (2001): Preparing fMRI data for statistical analysis. In: Jezzard P, Matthews PM, Smith SM (Eds.): *Functional MRI: an introduction to methods*, pp 229-242. Oxford University Press, New York.
- Snowling MJ (1981): Phonemic deficits in developmental dyslexia. *Psychol Res* 43:219-234.
- Stein BE and Meredith MA (1993): *The merging of the senses*. The MIT Press, Cambridge, MA.
- Stein BE, Jiang H, Stanford TE (2004): Multisensory integration in single neurons of the midbrain. In: Calvert GA, Spence C, Stein BE (Eds.): *The handbook of multisensory processes*, pp. 243-264. The MIT Press, Cambridge, MA.
- Stein J and Walsh V (1997): To see but not to read; the magnocellular theory of dyslexia. *Trends Neurosci* 20:147-152.
- Strafella AP, Paus T (2000): Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport* 11:2289-2292.
- Strainer JC, Ulmer JL, Yetkin FZ, Hughton VM, Daniels DL, Millen SJ (1997): *Am J Neuroradiol* 18:601-610.
- Stryer L (1988): Oxygen-transporting proteins: myoglobin and haemoglobin. In: Stryer L: *Biochemistry*, pp. 143-171. W.H. Freeman and Company, New York.
- Suga N and Ma X (2003): Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat Rev Neurosci* 4:783-794.
- Sumbly WH, Pollack I (1954): Visual contribution to speech intelligibility in noise. *J Acoustic Soc Am* 26:212-215.
- Talairach J and Tournoux P (1988): *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging*. Thieme, Stuttgart.
- Talavage TM, Ledden PJ, Benson RR, Rosen BR, Melcher JR (2000): Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear Res* 150:225-244.
- Tallal P and Piercy M (1973): Defects of non-verbal auditory perception in children with developmental aphasia. *Nature* 241:468-469.
- Tallal P, Stark RE, Mellits ED (1985): Identification of language-impaired children on the basis of rapid perception and production skills. *Brain Lang* 25:314-322.
- Tanaka H, Fujita N, Watanabe Y, Hirabuki N, Takanashi M, Oshiro Y, Nakamura H (2000): Effects of stimulus rate on the auditory cortex using fMRI with “sparse” temporal sampling. *Neuroreport* 11:2045-2049.
- Tarkiainen A, Helenius P, Salmelin R (2003): Category-specific occipitotemporal activation during face perception in dyslexic individuals: an MEG study. *NeuroImage* 19:1194-1204.
- Teder W, Alho K, Reinikainen K, Näätänen R (1993): Interstimulus interval and the selective-attention effect on auditory ERPs: “N1 enhancement” versus processing negativity. *Psychophysiology* 30:71-81.
- Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JDE (2000): Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proc Natl Acad Sci U S A* 97:13907-13912.
- Temple E (2002): Brain mechanisms in normal and dyslexic readers. *Curr Opin Neurobiol* 12, 178-183.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001): Functional specialization in rhesus monkey auditory cortex. *Science* 292:290-293.

- Tian B and Rauschecker JP (2004): Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 92:2993-3013.
- Tiippana K, Andersen TS, Sams M (2004): Visual attention modulates audiovisual speech perception. *Eur J Cogn Psychol* 16:457-472.
- Torgesen JK, Wagner RK, Rashotte CA (1994): Longitudinal studies of phonological processing and reading. *J Learn Disabil* 27:276-286.
- Treisman A (1964): Monitoring and storage of irrelevant messages in selective attention. *J Verb Learn Verb Beh* 3:449-594.
- Umiltà MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, Rizzolatti G (2001): I know what you are doing. A neurophysiological study. *Neuron* 31:155-165.
- Ungerleider LG, Mishkin M (1982): Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJW (Eds.): *Analysis of visual behaviour*, pp. 549-586. The MIT Press, Cambridge, MA.
- Vihla M, Lounasmaa OV, Salmelin R (2000): Cortical processing of change detection: dissociation between natural vowels and two-frequency complex tones. *Proc Natl Acad Sci U.S.A.* 97:10590-10594.
- Voloumanos A, Kiehl KA, Werker JF, Liddle PF (2001): Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cogn Neurosci* 13:994-1005.
- Wang Y, Sereno JA, Jongman A, Hirsch J (2003): fMRI evidence for cortical modification during learning of mandarin lexical tone. *J Cogn Neurosci* 15:1019-1027.
- van Wassenhove V, Grant KW, Poeppel D (2005): Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 25:1181-1186.
- Watkins KE, Strafella AP, Paus T (2003): Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41:989-994.
- Watkins K and Paus T (2004): Modulation of motor excitability during speech perception: the role of Broca's area. *J Cogn Neurosci* 16:978-987.
- Weinberger NM (2004): Specific long-term memory traces in primary auditory cortex. *Nat Rev Neurosci* 5:279-290.
- Welch RB and Warren DH (1986): Intersensory interactions. In: Boof KR, Kaufman KR, Thomas JP (Eds.): *Handbook of Perception and Human Performance, Vol 1: Sensory Processes and Perception*, pp. 1-36. Wiley, New York.
- Werker JF and Tees RC (1987): Speech perception in severely disabled and average reading children. *Can J Psychol* 41:48-61.
- Werner-Reiss U, Kelly KA, Underhill AM, Groh JM (2003): Eye position affects activity in primary auditory cortex of primates. *Curr Biol* 13:554-562.
- Wernicke C (1874): *Der aphasische symptomkomplex*. Breslau: Cohn, Weigert.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001): Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1-7.
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004): Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701-702.
- Wise RJS, Scott SK, Blank C, Mummery CJ, Murphy K, Warburton EA (2001): Separate neural subsystems within 'Wernicke's area'. *Brain* 124:83-95.
- Woldorff MG and Hillyard SA (1991): Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr Clin Neurophysiol* 79:170-191.

- Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE (1993): Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc Natl Acad Sci U S A* 90:8722-8726.
- Woods DL, Hillyard SA, Courchesne E, Galambos R (1980): Electrophysiological signs of split-second decision-making. *Science* 207:655-657.
- Woolrich MW, Ripley BD, Brady M, Smith SM (2001): Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage* 14:1370-1386.
- Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM (2004): Multi-level linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage* 21:1732-1747.
- Worsley KJ (2001): Statistical analysis of activation images. In: Jezzard P, Matthews PM, Smith SM (Eds): *Functional MRI: An introduction to methods*, pp. 251-267. Oxford University Press, New York.
- Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003): Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034-1043.
- Yang X, Hyder F, Shulman RG (1996): Activation of single whisker barrel in rat brain localized by functional magnetic resonance imaging. *Proc Natl Acad Sci U S A* 93:475-478.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992): Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846-849.
- Zatorre RJ, Meyer E, Gjedde A, Evans AC (1996): PET studies of phonetic processing of speech: review, replication and reanalysis. *Cereb Cortex* 6:21-30.
- Zatorre RJ, Mondor TA, Evans AC (1999): Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage* 10:544-554.