Audiovisual Speech Perception in Dyslexia

Thomas Kaltenbacher

1st edition (e-book)
Author: Thomas H. Kaltenbacher, PhD, M.phil.
Cover illustration: Elfriede Breninger
 Layout: TK
Proof Editing: Birgit Breninger, PhD

Cite as:
Kaltenbacher, T. (2018). *Audiovisual Speech Perception in Dyslexia.* Salzburg: Salzburg Institute of Reading Research Publishing (SIRR).

Birgit and Anaïs

Preface

This book is devoted to the search for a psycholinguistic theory that might be able to explain the origin of the pre-lexical linguistic deficit in dyslexia. Audiovisual speech and visual speech perception lie at the heart of approaching the linguistic structure that seems to be most affected in dyslexics, across all languages and writing systems. The topic as such has fascinated me ever since I was introduced to AV-speech as a graduate student. The famous McGurk effect together with the remarkable research it has nurtured ever since its – quite accidental – discovery in the 70s has highlighted bimodal speech perception and all phonemic, visemic and, as I would like to suggest, also graphemic aspects of language processing. By the time I set out to explore AV-speech perception, I had already presented various McGurk eliciting video clips to hundreds of students, mostly undergraduate students of linguistics and cognitive neuroscience. Like most researchers and scholars who use the McGurk effect for demonstration, I was astounded by the robustness of the phenomenon. To me the most intriguing fact, however, has always been why the McGurk effect cannot be elicited in certain subjects (a tentative percentage of 25%). Researchers over the past five decades have investigated the McGurk illusion and various modifications thereof, from temporal modulation to blurring of images, from – animated – talking heads to semi-hollow models and naturalistic videos in high definition with a millisecond-true temporal alignment.

So, for the empirical part of the book, I designed my own experimental paradigm permitting the tracking of subjects' abilities to process phonological stimuli in visual- only, acoustic- only and audiovisual conditions. For this purpose, I went to great lengths to elevate the video quality to the latest technical standard. Not only did we use professional digital video recording equipment and a studio with perfect lighting, but we also employed the most reliable audio recording and editing machinery that a phonetics lab could provide. Designing and improving the stimuli was strangely rewarding, even after the umpteenth take on 'baga' and 'paka' syllables. The obtained stimulus set was later on also used in a subsequent fMRI study and yielded further back-up for the results discussed in the final chapters of this book.

When we study dyslexia in the audiovisual domain, a simple phonological task involving visual, i.e. graphemic material must be included – not only to ascertain that a phonological deficit is present, but also to verify a visual processing impairment in script. This undisputedly critical phonological task was tested with pseudoword reading – one of the strongest predictors of the reading deficit that constitutes dyslexia.
The aim of this book was to examine the perception of audiovisual speech in adolescent and adult dyslexic subjects and detect correlations between deficient audiovisual and phonological processing. Based on current models of audiovisual speech processing a conceptual and analytical framework of bimodal speech perception is provided that manages to explain such correlations.

I owe thanks to a great number of people who have supported me throughout my endeavours of audiovisual speech research, who I cannot all list here. However, the following people do have my everlasting gratitude:
Birgit Breninger, who is not only my partner in life and in research, but also the most empathic and unwavering beacon of light when academic life turns sour and the light at the end of the tunnel seems to have been switched off.
Peter Hummer, who is to blame for it all, having introduced me not only to phonetics, phonology, psycholinguistics, neurolinguistics and clinical linguistics, but also to dyslexia research outside of the reading and writing domain.
Ulli Kipman, whose help with calculations and statistics has been invaluable to our research, across all disciplines and experimental designs.
Christian Felder, who is second to none when it comes to filming, photographing, lighting, editing and – partying in Munich.

Content:

List of Figures

List of Tables

1. Introduction

Do Be Do Be Do
(Frank Sinatra, 1966)

Pata Pata
(Miriam Makeba, 1967)

Ba Da Ga
(Harry McGurk and John MacDonald, 1976)

For many years, the concept of dyslexia has been associated with research surrounding various hypotheses of potential causal deficits. Researchers rely on the World Health Organization's ICD-10, definition of developmental dyslexia and relate to it generally as "[…] a specific difficulty in the acquisition of reading and writing in spite of preserved general intelligence, learning opportunity, motivation or sensory acuity" (e.g. Silani et al., 2005, p. 2453).

In the innumerable attempts to investigate the complex nature and causes of this phenomenon, various hypotheses have been proposed over the past decades. Some researchers have suggested causal explanations that are closely linked to the linguistic aspects of reading and writing, whereas others have chosen more general cognitive approaches to explain dyslexia through basal cognitive or biological impairments.

Among the linguistically oriented approaches, the most prominently discussed hypothesis is the phonological deficit hypothesis, which has been outlined by many researchers (among others by: Marshall, Ramus, & van der

Lely, 2011; M. Snowling, Bishop, & Stothard, 2000; M. Snowling & Stackhouse, 2006; M. J. Snowling, 2006; M. J. Snowling & Hulme, 2005; Vellutino, 1981; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Lower linguistic level approaches have attempted to complement the phonological deficit hypothesis, as, for example, in the highly controversial low-level auditory deficit hypothesis introduced by Tallal (1980) in which temporal order judgements are said to be impaired.

A recent and more promising explanation was put forward by Giraud and Poeppel (Giraud & Poeppel, 2012), who discuss *speech sampling deficits* in dyslexia. Investigations concerning the underlying cause of the phonological deficit have also been analysed from a perspective of speech perception problems (Baart, de Boer-Schellekens, & Vroomen, 2012; Blomert & Mitterer, 2004; Mody, Studdert-Kennedy, & Brady, 1997; Serniclaes, Van Heghe, Mousty, Carre, & Sprenger-Charolles, 2004) as well as from a phonological representations perspective (Ramus & Szenkovits, 2008; Szenkovits & Ramus, 2005; Van Orden & Kloos, 2005)

Some of the biological approaches include the visual deficit and visual attention deficits (Valdois et al., 2011; Valdois, Bosse, & Tainturier, 2004; Valdois, Lassus-Sangosse, & Lobier, 2012; Vidyasagar, 2004; Vidyasagar & Pammer, 2010) as well as theories that subsume learning deficits, memory deficits and visual attention deficits caused by a magnocellular deficit as devised by John Stein (J. Stein, 2001, 2008).

For this specific purpose here, the core linguistic deficit, which underlies the reading and writing problems in dyslexia is examined further, in order to contribute scientifically to the not yet sufficiently researched area of audiovisual speech processing in dyslexic subjects. The ability to simultaneously (i.e. bimodally) process audiovisual speech alongside the ability to process unimodal visual and acoustic speech signals lies at the heart of the eye-tracking and behavioural research presented here.

When it comes to audiovisual speech processing, the McGurk effect plays a vital role. The processing of speech sounds, as well as the integration of speech sounds and visual speech components has long intrigued researches and scientists from various disciplines. Already as early as 1954, Sumby and Pollack investigated the influence of visual information in speech perception. The discovery of the McGurk effect in 1976 (Harry McGurk & John Mac Donald), also known as the McGurk illusion, can be regarded as 'the dawn of multimodal language processing' and studies concerning the McGurk effect are still prominently featured in the cognitive neurosciences.

As a matter of fact, the McGurk effect is said to have been discovered by accident. As far as the story goes, technicians had created test items by dubbing audio syllables onto talking faces and when the items were presented, the researchers saw videos of syllables they had not recorded. Eventually, they realised that perception of the items would change, depending on whether they looked at the videos or only listened to them (Harry McGurk cited by Massaro, 1998, p. 25). McGurk and Mac Donald thus found out that a face producing a visual <ga> dubbed with an acoustic <ba> does not result in some <bga> or <gba> co-articulated percept of the syllable but leads most people (in whose phonetic inventory bilabial, dental and palatal voiced plosives occur) to perceiving <da>. Researchers like Dominic W. Massaro have investigated the McGurk effect in all its complexity, aiming at eliciting as many McGurk percepts as possible in various languages and measuring precisely within which time frames the McGurk effect can be triggered. The robustness of the effect serves as a strong evidence for audiovisual speech processing, and the situations in which the effect is not be triggered will be discussed further in chapter 2.

Powerful illusions, such as the McGurk effect, in which the organism integrates perceptual processes unconsciously, have been referred to as 'afference binding' (Morsella & Bargh, 2011). The observer is unaware of the processes of any intersensory interaction leading to the outcome of the perception of <da>. Phenomena, such as the McGurk effect, the pupillary reflex (Morsella, Gray, Krieger, & Bargh, 2009), interactions among modular processing of colour, motion and shape in visual perception (L. J. Bernstein & Robertson, 1998; Holcombe, 2009; Masson, Dodd, & Enns, 2009; Zeki & Bartels, 1999) and interactions among modular processing of diverse cues in depth perception (Harris & German, 2008; Hochberg, 1998), amongst others, demonstrate vividly that intersensory crosstalk can occur unconsciously.

Audiovisual speech perception has been studied ever since by researchers with various disciplinary backgrounds ranging from phoneticians to neurocognitive scientists, vying for explanations of what has been called 'the struggle of the parts problem' (Mayr, 2001): if parts of an organism evolved at many different times, then how can they all be brought together to work in such a seamless fashion? Massaro (C. S. Campbell & Massaro, 1997; Massaro, 1998; Massaro & Jesse, 2009), among many others, has devoted substantial work to the perception of talking faces. Selected aspects of Massaro's Fuzzy Logical Model of Perception and its neural underpinnings will be discussed in more detail in chapter 2. In this chapter, the revised motor theory of speech (Liberman & Mattingly, 1985), crucial for the outcome of this book will also be outlined together with the pertinent theories on speech perception (cf. Blomert & Mitterer, 2004; Hickok & Poeppel, 2004; Nusbaum, 2011; Poeppel, Idsardi, & van Wassenhove, 2009).

The term 'audiovisual integration' was introduced in this context to describe the phenomenon of the merging of the two modalities and has been thoroughly investigated by various researchers ever since (for an overview, cf. R. Campbell, 2009). Sumby and Pollack's findings on visual contribution to speech intelligibility in noise (1954) together with phenomena like the ventriloquism illusion have set the stage for audiovisual speech research. The illusion that a 'talking' puppet's visual gestures complement the 'invisible' articulatory gestures of the puppeteer, was first described by Radeau and Bertelson in 1974. Additionally the famous saying "read my lips", which is attributed to various American presidents from Richard Nixon to George W. Bush Sr. may certainly have encouraged researchers to investigate further the visual influence on speech. Especially in such political contexts it might be considered rewarding to test the McGurk effect with the talking faces of people who have been proven deceitful in their words in order to find out whether trust plays a role in speechreading. On a more linguistic note, susceptibility of subjects with a (speech-) reading disorder or a face processing disorder (commonly known as 'prosopagnosia') to the McGurk effect or the ventriloquism illusion might be a promising venture.

Some studies (Campanella & Belin, 2007; de Gelder & Vroomen, 1998; de Gelder et al., 2005; Yehia, Kuratate, & Vatikiotis-Bateson, 2002) also take into consideration how the talker's face including the face's kinematics may influence and even enhance speech reading. These aspects have been taken into account when the experimental stimuli for this practical part were designed (cf. chapter 4).

Apart from the plethora of research concerning itself with the interaction between visual and auditory processes and in particular with the McGurk effect, the results of two pilot studies have contributed significantly to the generation of the four viability hypotheses underlying the empirical part of this book. In my first pilot study (unpublished M.A. thesis, 2005) adult dyslexic subjects (n= 20) and age and education matched controls (n= 14) were tested for their susceptibility to the McGurk effect (see chapter 2.). The results indicated that dyslexics are not as susceptible to the McGurk effect as controls are, which led me to suspect an audiovisual speech integration deficit in dyslexia. This would explain the poor performance of dyslexics in such tasks. In the consecutive pilot study by Widerin (Widerin, 2007) at the University of Salzburg, dyslexic subjects were shown a video of a talking face uttering short sentences in natural language and were then asked to repeat the sentences. The auditory signal was masked with noise as suggested by various studies (cf. Bastien-Toniazzo, Stroumza, & Cavé, 2009; Hayes, Tiippana, Nicol, Sams, & Kraus, 2003; Sams, Aulanko, et al., 1991; Sumby & Pollack, 1954) forcing subjects to rely on the visual features of the talking face, without the processing of which it would not be possible to correctly repeat the sentence. Subjects also received semantic support by means of three drawings that provided the core information of the semantic content of the stimuli. These drawings were displayed serially at the appropriate sentence constituent position. With the help of these drawings subjects were able to understand and repeat the utterances, albeit incompletely: subjects who relied solely on the drawings would always have some words missing in their repetitions, whereas subjects focussing on the talking face would be able to repeat back the utterances in full. Subjects who displayed a preference for the semantic help, i.e. the drawings rather than the talking face, were suspected to either be unaware of their 'complementary' audiovisual speech perception mode or to ignore the talking face in this auditorily challenging condition through some intuitive knowledge about their inability to speechread. The latter was suspected in dyslexic subjects who, as eyetracking revealed, preferred to look at the drawings and repeated back the utterance with correct yet incomplete content or looked at the talking face without profiting from the visual information, which resulted in incorrect repetitions of the sentences.

Based on the assumption that the core symptom of dyslexia, i.e. the reading and writing deficit, must have a – potentially subtle – underlying speech signal processing deficit, I devised an experimental paradigm in order to examine such processes on the phonemic syllable level and on a phonemic pre-lexical level, as well as on a lexical level. At the beginning, the following three hypotheses are challenged in order to prepare the ground for the specific-deficit hypotheses of the experiments conducted for this work:

1. The null hypothesis: Dyslexic subjects and non-dyslexic controls are equally susceptible to the McGurk effect. A robust McGurk effect indicates that audiovisual integration of speech signals is intact. In order to ascertain that subjects and controls fixate the visual aspects of the audiovisual signal i.e. the speechreading areas of the talking face as identified by Massaro and Jesse (2009, p. 24), subjects' eye movements are recorded with an eyetracker.
2. In addition, an amendment to the null hypothesis is tested: Dyslexic subjects and non-dyslexic controls perform equally well in a visual only speechreading task. This means that subjects and controls both identify the speechreadable (= lipreadable) aspects of visual speech signals, such that subjects are able to repeat back what a talking face articulates.
3. Eventually, an extension of the null hypothesis concerning the correlation between participants' lipreading abilities (prelexical) and their pseudo-word processing in the pseudoword task is put forward and examined: Dyslexic subjects should display difficulties in the pseudoword task whereas non-dyslexic subjects are supposed to display good pseudoword processing skills (for the pseudoword problem cf. e.g. Goswami et al., 2011; Ramus & Ahissar, 2012; M. J. Snowling, 2006; Vellutino et al., 2004). No negative correlation between the visual only task and poor pseudoword task performance for dyslexic subjects is expected here.

These unspecific deficit hypotheses will be confronted with an *alternative hypothesis*, namely the specific-deficit hypothesis that dyslexic subjects differ significantly from non-dyslexic controls in their susceptibility to the McGurk effect. In this case it is expected that dyslexic subjects are either not susceptible to the McGurk effect at all, or that the effect is not as robustly elicited as in non-dyslexic controls. Dyslexic subjects, who are not susceptible to McGurk items will show a strong tendency to process the audio signal and ignore the visual speech signal. In order to rule out that the McGurk effect was not triggered by subjects' non-fixations of the speechreading areas in a talking face, all trials where subjects' fixations lay outside the areas of interest were excluded from the sample.

This is the gist of the empirical work of the book:

First, it will be shown that the null hypotheses do not hold. Both, the insensitivity of dyslexics to the McGurk-Effect as well as the negative correlation between the visual only task and poor pseudoword task for dyslexic subjects call for a specific explanation in terms of multimodal integration. The explanation is framed in terms of the following hypotheses that are to replace the null-hypotheses:

i. Dyslexic subjects and non-dyslexic controls differ significantly in their susceptibility to the McGurk effect. A robust McGurk effect indicates that audiovisual integration of speech signals is intact. Weak responses to the McGurk effect in dyslexic subjects indicate a multimodal, i.e. audiovisual integration problem.
ii. Dyslexic subjects perform significantly worse than non-dyslexic controls in a visual only speechreading task. This means that subjects have great problems identifying the speechreadable (= lipreadable) aspects of visual speech signals, such that subjects are not able to repeat back what a talking face articulates.
iii. A strong correlation between the visual only task and poor pseudoword task performance for dyslexic subjects is expected.

In order to be able to test for these hypotheses, a representative sample of participants needed to be examined. Hence, an experimental setup was designed that allowed testing of 71 participants (dyslexic subjects and controls matched for age and education). The dyslexic subjects presented with an official diagnosis according to the above mentioned WHO definition of dyslexia. Stimuli comprised audiovisually congruent speech material, audiovisually incongruent items triggering a McGurk effect, as well as audio only and visual only stimuli. Upon presentation of the items, subjects' eye movements were recorded with an SR research Eyelink 1000 eyetracking system to ascertain that subjects would look at the speech relevant aspects. In addition to the audiovisual eyetracking paradigm, a pseudoword reading task from an established test battery was used (Moll & Landerl, 2010).

Dyslexic subjects were expected to show poorer performance in the audiovisually incongruent condition, as well as in the visual only (= speechreading) condition and in the pseudoword reading task. The alleged subtle speech processing deficit should help explain the poor performance in the audiovisual tasks and fit the recent renaissance of motor speech theories (cf. Nusbaum, 2011). The correlation between subjects' poor performance in speechreading and pseudoword reading tasks will be discussed further, following the concept of the phonological deficit hypothesis in chapter 3 and with the help of current speech processing theories outlined in detail in chapter 2.

In the experimental part of this book, McGurk items across the various constraints were used, including syllables, highly frequent lexical items and phonologically complex pseudowords. The ability of subjects to lipread in a visual-only condition, and to auditorily process in an acoustic-only condition as well as in an audiovisual condition (both congruent and incongruent) was tested.

The ability to process audiovisual stimuli simultaneously, thereby allowing effects like the McGurk effect to be triggered seems to be a universal linguistic ability functioning on all linguistic levels (cf. R. Campbell, 2009). The McGurk effect items were created because the effect is a robust phenomenon, which means, it can still be triggered, even when subjects are familiar with the effect. It can also be elicited when subjects are shown a McGurk video from a distance of up to 15meters, where the mouth is no longer clearly visible (cf. Jordan & Sergeant, 2000) and even when the video is blurred (cf. Munhall, Kroos, Jozan, & Vatikiotis-Bateson, 2004). These facts are quite important when one deals with audiovisual integration because it suggests that audiovisual integration cannot be suppressed (unless the respective brain area is numbed, as for example by TMS, see: Beauchamp, Nath, & Pasalar, 2010).

Another important facet of audiovisual integration that is relevant for this study is that the McGurk effect also has an impact on children (cf. McGurk & MacDonald, 1976), even prelinguistic babies (cf. D. Burnham & Dodd, 2004; Kuhl & Meltzoff, 1982). In McGurk and MacDonald's original study (1976, p. 747), the susceptibility to the effect was not strong in young school-children (susceptibility: 64%). It showed a strong effect in adults (98%) and a fairly strong effect in pre-schoolers (81%). The findings that prelinguistic babies seem to be sensitive to the McGurk effect is crucial for this research here, since it suggests that the McGurk effect does occur at a certain age where the phonological development is not yet complete and the phoneme inventory is not yet established.

If infants at risk for dyslexia show no susceptibility to the McGurk effect – can this perhaps be an indicator for oncoming phoneme- grapheme problems? One behavioural study has tested childrens' susceptibility to the McGurk effect and has shown, that dyslexic children are not only less susceptible to the audiovisual illusion, but that they are also poorer lip-readers (Bastien-Toniazzo et al., 2009).

It is known that the McGurk percept depends on the phoneme inventory of the perceiver or as Campbell (R. Campbell, 2009) claims, it depends on the fact that "[…] a seen speech event usually maps onto several (acoustically defined) phonological categories" (2009, p. 137). When referring to Auer and Bernstein (1997) Campbell remarks that visually confusable phonemes – the 'visemes' – can be considered to constitute phonemically equivalent classes (PECs) (R. Campbell & MacSweeney, 2012).

Studies of the McGurk effect across various languages support this idea (for an overview cf. Denis Burnham & Dodd, 1996): while speakers of English and Greek may perceive either an alveolar plosive or an interdental fricative as a consequence of the auditory bilabial plosive dubbed onto the visual velar or palatal plosive, speakers of German or Dutch will simply perceive the alveolar plosive, due to the absence of phonemic interdental fricatives in their mother tongue. Hence the McGurk percept might perhaps also be used in adult second language (L2) acquisition to measure the increase in the L2's phoneme inventory.

Would the early stage of language acquisition in young school kids, who according to the McGurk study in 1976 showed the smallest susceptibility to the effect (64%), be able to explain this reduced effect? Are visemes perhaps not considered by children in a short phase in language acquisition, due to the novel acquisition of graphemes and their convergence with existing phonemes? If this is the case, it could explain the reduced effect. Consequently, in scripts that do not use graphemes, the conflicting phoneme – viseme problem should not occur. Hence, the McGurk effect should be triggered throughout the preschool phase into the phase of scripture acquisition. One study that has examined this context suggests that Japanese adults are less influenced by visual information than English adults (cf. Sekiyama &

Burnham, 2008). In recent research on language acquisition, Sekiyama and Burnham (2008) suggest that this difference starts to emerge between 6 and 8 years. Campbell (2009, pp. 137) comments on the fact that the number of PECs varies interindividually, depending on an individual's speech reading skill. Following Auer and Bernstein, Campbell also states that: "The reason why a relatively small number of PECs can suffice for identifying individual spoken words is that most words in English are relatively unique in their segmental and syllabic structure." (ibid., p. 137)

The amount of attention required when audiovisual processing takes place, and whether this attention is automatic or not, has also been thoroughly discussed, for example, by Soto-Faraco et al. (Soto-Faraco, Navarra, & Alsius, 2004). In the context of dyslexia research, attention deficits have been suspected to correlate with reading and writing difficulties. It is common practice to include a standardized intelligence test in the diagnostic process of dyslexia (e.g. Wechsler Intelligence Scales or Kaufman Assessment Batteries) that includes subtests for working memory. Some researchers focus on attention deficits and their causal relevance for dyslexia research (Lobier, Zoubrinetzky, & Valdois, 2012; Valdois et al., 2004; Vidyasagar & Pammer, 2010). For this research no empirical test of subjects' attention was conducted, since the method of eyetracking warrants whether subjects' eyes rest on the speech reading areas – the jaw and lips. If information from seen speech is not processed even though subjects looked at the talking face at the relevant moment, it is not considered to be due to lack of attention. Studies by Tiippana et al. (K. Tiippana, Andersen, & Sams, 2004) and Alsius et al. (2005) report that the McGurk effect can be influenced to the level of reduction when visual (Tiippana et al., 2004) and auditory (Alsius et al., 2005) distractors are applied to a McGurk stimulus. It remains unclear whether the decrease in susceptibility to the McGurk effect is the result from additional attentional workload or from a perceptual overload. Alsius (2005, p. 841) confirms that the McGurk effect is even stronger when the auditory signal is somewhat degraded. In the empirical research of this study any form of distraction has been carefully avoided – subjects were tested within quiet, distractor-free surroundings and auditory stimuli were presented via loudspeakers positioned directly in front of the subjects in order to prevent any auditory distraction.

In order to provide a coherent overview, this book is structured in five major parts. After the first introductory part, chapter two presents a state-of-the-art report of audiovisual speech processing research, which includes pertinent theories and models of speech perception. Following this pertinent research report, the theoretical conceptions and the neurolinguistic framework for audiovisual speech processing will be introduced and discussed. The neural networks, language pathways and linguistic processes required for audiovisual speech are discussed alongside selected theoretical approaches to speech processing. The concept and the requirements of a psycholinguistic theory that can explain audiovisual speech processing abilities and a deficit in audiovisual speech processing complete this part.

An overview of the recent research that has investigated the potential causes for dyslexia will also be given and their relevance to dyslexia as a language impairment will be discussed in chapter 3. Among the various deficit hypotheses striving to explain the cause of dyslexia, the phonological deficit hypothesis is the most linguistically-oriented hypothesis. Since dyslexia's dominant symptoms constitute the linguistic problem of a reading and writing deficit, it seems promising to investigate these linguistic aspects further.

Furthermore, the development from phonological awareness to reading skills will be discussed in relation to the development of a phonological system in pre-school children and the enhancement of this system in school children who learn to read and write. The linguistic development from audiovisual integration to speechreading together with its impact on phonological and lexical reading processes will be carefully outlined in this section. Speech impairments and developmental language deficits linked to dyslexia will be dealt with in regard to the phonological and audiovisual integration deficit.

The chapters four and five of this book are the empirical chapters that focus on the dyslexic subjects' ability for audiovisual speech integration and the susceptibility to the McGurk effect forms. In chapter four the empirical research that has been conducted is presented in full. The experimental paradigm that allows researchers to test dyslexic subjects' ability to process phonological stimuli in an audiovisual, a visual-only and an audio-only condition is introduced. The creation of the stimulus set, the method of testing subjects with eyetracking and the psychological and linguistic tests that are generally used to identify dyslexic subjects will be explained. The novelty of the experimental design and the various stimulus conditions are described in detail. Rather than reading phonological items from an alphabetic stimulus set, subjects were asked to lip-read, while subjects' eye movements were monitored in order to warrant that subjects actually focussed on the talking faces.

The conception of this task allowed drawing a conclusion concerning the deficient audiovisual speech perception from subjects' reading and lip-reading abilities. The audiovisual speech processing task, which involved audiovisual congruent and audiovisual incongruent items (= McGurk items), should test subjects' ability to integrate two speech signals and – in case of a non-susceptibility to the McGurk effect – determine their preference for either the acoustic or the visual channel.

In the subsequent chapter, chapter five, the obtained data are presented. The computations of the data and the statistical methods used to interpret the data are explained. The discussion of results in relation to the initial hypotheses leads to a critical interpretation of the findings and their implications for current research. Here the null hypothesis is refuted and the alternative hypothesis is confirmed. The identified audiovisual processing deficit and its correlation with poor phonological skills is discussed in regard to current theories of the reading and writing deficit that is dyslexia.

The actual nature of the suspected audiovisual speech integration deficit and the explanatory power of this deficit in the phonological deficit hypothesis implies an improvement in future diagnostics of dyslexia and perhaps even in therapeutic approaches. Further suggestions concerning testing audiovisual speech processing abilities with other imaging techniques aside from eyetracking are also given in this section, followed by the bibliography and appendices.

2. Audiovisual Speech Perception

Science is nothing but perception.
Plato (427-347 B.C.)

Alvin Liberman put forward the intriguing question: “Why is Speech so Much Easier than Reading and Writing?“ (1998, p. 5). Taking into account the ample recent literature on speech perception and speech reading, the immediate reply to Liberman nowadays must be: “Is it really?”

This chapter will give an overview of pertinent theories on speech perception and speech reading and their underlying neural processes. Seeing as these are sensory, motor and sensorymotor processes, it will also be discussed, how and where these skills are represented in the brain alongside with how they interact when speech perception and production are acquired in typically developing children. The chapter commences with an introduction to speech processing in general and a subsection 2.1.1 on Liberman and Mattingly’s revised theory of motor speech (Liberman & Mattingly, 1985). Liberman and Mattingly were the first to be convinced that more than one modality is involved in the process of speech perception and their explanation of the McGurk effect will, hence, be outlined in some detail. In subchapter 2.1.2 a section on the neural basis of speech perception will follow. Among the most relevant theoretical approaches in speech perception is also that of Poeppel and Hickock, therefore their speech perception model will be discussed in detail (Hickok & Poeppel, 2004; Poeppel et al., 2009). A short excursus on visual speech in section 2.2, explaining the key aspects of the visual speech signal (and explaining how and why "Video Killed the Radio Star" (T. Horne, Downes, & Wooley, 1979), will thereafter lead us to audio-visual speech processing. Subchapter 2.3.1 introduces the topics concerning the robustness of AV-integration, the benefits of audiovisual speech (see subchapter 2.3.2), audiovisual speech and lexical representations (see subchapter 2.3.3) and, finally, the brain locations of audiovisual speech (see subchapter 2.3.4). From a discussion concerning recent research and the upsurge of new theories and models of audiovisual speech perception in section 2.4 we will proceed to audiovisual speech processing research linked with dyslexia research in section 2.5 and a conclusive suggestion for further research.

Based on the assumption that (audiovisual) speech processing correlates with the processing of written language, a few comments will be propounded throughout the discussion of (audiovisual) speech perception models and theories regarding how the interaction between speech processing and speech representation might find an equivalent in literacy.

2.1. Speech Perception

It was not until the 1950's that speech perception was considered from a perspective inclusive of visual speech processing. Owing to Sumby and Pollack, visual aspects of speech signals suddenly became interesting (Sumby & Pollack, 1954). However, research focussing exclusively on the auditory aspects still dominated the speech perception science for another two decades until McGurk and MacDonald showed in 1976 that the visual aspects of speech can, literally, not be ignored (McGurk & MacDonald, 1976). I use the term ‚literally' here, to already hint at the influence that becoming literate may have on audiovisual speech perception (see chapter 3.4 for a discussion on how learning to read may impact perceptual processes).

Undoubtedly, the auditory aspect of speech perception is essential to the perception of a speech signal, enabling us to communicate without seeing the face of the interlocutor, whereas communicating without hearing the interlocutor is an infinitely greater challenge.

2.1.1 The Motor Theory of Speech

The core aspect of Liberman and Mattingly's revised motor theory of speech is that phonetic information must be processed in a „module" i.e. a "biologically distinct system" (1985, p. 1) that identifies phonetic gestures of a speaker and translates them into phonetic categories. This module comprises the "lawful relationship between the gestures and the acoustic patterns" (ibid.). The phonetic structure may thusly be processed without causing the listener to translate from what Liberman and Mattingly call "preliminary auditory impressions" (ibid.). This phonetic module would "compete with other (non speech specific representation, comment TK) modules for the same stimulus variations", which makes the motor theory of speech an indispensable element for audiovisual speech integration models (cf. also Blomert & Mitterer, 2004).

Liberman and Mattingly constitute that the original motor theory was motivated by findings that synthetic speech is unintelligible unless an "invariant phonetic percept" was included (1985, p. 2). Perception of motor invariants would, therefore, "depend on a specialized phonetic mode" (ibid.). One of the theory's key aspect remains the idea that

> "[…] objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations.

> These gesture commands are the physical reality underlying the traditional phonetic notions (…) tongue backing, lip rounding, jaw raising." (ibid.)

Another proposition of the theory is that speech perception and speech production are 'intimately linked' because they share the same set of 'invariants', i.e. invariant gestures that encompass the basic (segmental) phonetic gestures and their phonological bundles of features (1985, p. 3). Liberman and Mattingly's suggestion that this link is 'innately specified' has been drawn upon later by Hickock & Poeppel who conclude that already

> "[…] infants must shape their articulatory gestures in a way that matches the phonetic structure of the language they are exposed to; yet the primary input to this motor learning task is acoustic […] therefore there must be some mechanism for using auditory input to shape motor output"(Hickok & Poeppel, 2004, p. 68).

One suggestion that may follow from these two viewpoints is that phonemic awareness is not fully acquired until the child has become a reader/ writer, because phonemic representations are strongly influenced by graphemic, i.e. alphabetic representations. The phonological skills, on the other hand, would indeed be developed earlier.

Poeppel et al. devised the central aspects of speech perception to lie in the extraction of phonologically distinctive features from the acoustic input (Poeppel et al., 2009). Poeppel et al. also incorporated the idea that timing plays a crucial role in speech perception. Liberman and Mattingly used the required precision in timing gestures simultaneously – as in coarticulation processes – to explain how the acoustic signal "is influenced by several gestures at the same time" (Liberman & Mattingly, 1985, p. 4). This relation between gesture and signal is said to be "peculiar to speech" (ibid.). In coarticulation, they argue, gestures may be represented differently in different (phonetic) contexts, which can be identified by investigating formant transitions in different contexts. If these transitions were translated into synthesized waveforms (as in synthetic speech experiments) they would not sound like speech any more and listeners would not be able to identify the phonetic categories of the respective sounds. Voice onset time (VOT) and lenis/ fortis features may yet serve as another example how crucial temporal aspects of signals are. A native speaker of German from the Middle-Bavarian dialect region, for example, will find it nigh impossible to distinguish between the French boisson [b] and poisson [p], due to lack of aspiration in the French voiceless bilabial plosive. In Middle- Bavarian German, it is not the difference in VOT that makes the voiced and the voiceless bilabial plosives distinguishable, but the aspects of lenification/ fortification (cf. Moosmueller & Ringen, 2004).

For Liberman and Mattingly speech perception can only be explained through the 'specialization for phonetic gestures' that must take place in early language acquisition phases – they refer to a "biologically based link between perception and production" (1985, p. 6.).

But what then makes the motor theory 'motor'? Liberman and Mattingly argue that the 'proper object of phonetic perception' would have to be a motor event and consequently, it would be a prerequisite that to facilitate speech the motor system evolved in order to control the organs of the vocal tract (1985, p. 7). Liberman & Mattingly further hypothesise on neural structures required for speech perception (ibid). Naturally, without the availability of imaging technologies, they were only able to speculate that the motor system is one of special 'perceiving systems' or 'modules', which must be represented on the neural level in the 'special structures', expounded on later by various researchers (e.g. Hickok & Poeppel, 2004). Taking into consideration the speaker's intention and the listener's processing of the speaker's signal, Liberman & Mattingly suggest that "speech somehow informs listeners about phonetic intentions of the talker" (1985, p.9). The interesting notion here is the existence of a 'perceiving' module answering to the representations of the talker's 'phonetic intentions'. What Liberman & Mattingly do not address in their theory is the question of lexical access and its influence on speech perception, which is crucial in the most recent theories (e.g. Hickok & Poeppel, 2004; Poeppel et al., 2009). Liberman and Mattingly further fail to address the problem of speech errors and speech error repair mechanisms in the listener. In speech errors, the listener must immediately (and usually does so quite effortlessly) correct the slip of the tongue produced by the speaker – in Liberman and Mattingly's terminology, these would have to be referred to as something like 'erroneous phonetic intentions'. In phonological speech errors, e.g. 'Brizal' instead of 'Brazil' the acoustic repair process must undoubtedly include rapid lexical access. Even though the motor theory revised considers phonetic and phonemic issues, a potential lexical influence is not discussed.

Liberman & Mattingly suggest that a 'stripped' linguistic signal, that is, phonetic units that have been reduced to 'sine-wave analogues of speech' would still be identified by a listener, who considers the signal to be speech relevant (1985, p. 12). In dyslexia research it has long been argued (most prominently by P. Tallal, 1980) that dyslexics show poor auditory abilities when it comes to discriminating categories – hence the name categorical perception deficit. Blomert & Mitterer were the first to examine this perception deficit in natural vs. synthetic speech (Blomert & Mitterer, 2004). Their core finding in relation to the motor theory of speech perception was that it makes a fundamental difference whether perception abilities are tested with natural speech signals. Liberman and Mattingly seem to have suspected this as they claim "that there is simply no way to define a phonetic category in purely acoustic terms" – rather the acoustic, natural speech signal would "serve only as a source of information about the gestures", which "[…] would properly define the category" (1985, p. 12).

If a speech perception deficit in dyslexia were to be identified, coarticulation processes would be another aspirant for thorough testing. In many developmental speech impediments, coarticulation processes are deficient in children (cf. chapter 3.2). However, it must be noted that not all articulatory deficits result in speech perception deficits, just as dyslexia does not seem to correlate with speech disorders (see: Liberman & Mattingly, 1985, p. 24). In the motor theory, coarticulation and the 'resulting overlap of phonetic information in the acoustic pattern' is assumed to be a processing sequence of discrete phonetic gestures realized by the respective articulators (1985, p. 14). The theory suggests further "that an equally efficient perceptual process might use the resulting acoustic pattern to recover the discrete gestures" (ibid.). If, in contrast to Tallal's hypothesis, the decoding or 'recovery' of these gestures is impaired in dyslexia, one can hypothesize that this impairment is causal to the poor phonological representations in dyslexics. Poor phonological representations have been considered causal for dyslexia by many researchers (cf. Bishop & Snowling, 2004; Brady, Braze, & Fowler, 2011; Castles & Coltheart, 2004; M. Snowling et al., 2000).

Liberman & Mattingly's revised motor theory also offers an explanation of the McGurk effect. They introduce the term "duplex perception", which stands for "a single acoustic stimulus" being "processed simultaneously by the phonetic and auditory modules to produce perception of the two distal objects: a phonetic gesture and a sound (i.e. a non-linguistic sound, comment TK)" (1985, p.17). The McGurk effect would be similar in that it combines two speech signals, albeit from different modalities, visual and acoustic, and merges these signals into one "coherent perception of a distal event" (ibid.). Interestingly, the abilities of prelinguistic infants (cf. D. Burnham & Dodd, 2004; Kuhl & Meltzoff, 1982) concerning audiovisual speech perception point to the robustness of audiovisual integration in speech signals, but not if McGurk like items are created with pure tones. The motor theory of speech would consequently attribute these findings to the perception-production link that is already present in infants. Once again, Liberman & Mattingly find proof herein, that phonetic perception is "perception of gesture" (1985, p. 21). Also, they recognize how "the child sometimes mistakes the phonological significance of the gesture" which is irrelevant for language acquisition as long as what the child perceives "captures the systematic nature of its relation to sound" (1985, p. 25). In Poeppel et al. (2009) this approach is drawn upon and strongly modified to include all temporal aspects of speech perception as well as interfering pre- lexical and lexical units. They sum up how speech perception is a 'multi-time resolution process' with perceptual analyses occurring at two time scales. According to their hypothesis there are "two principal time windows within which a given auditory signal (speech or non-speech) is processed" (Poeppel et al., 2009, p. 258). There is a time window of approx. 20-80ms within which segmental and subsegmental cues are processed and the segmental 'order' is identified, Poeppel et al give the example of "pest" vs. "pets" (ibid.). At the time scale of 150-300ms, suprasegmental and syllabic phenomena should be processed. Poeppel at al. argue that auditory signals are consequently processed in time windows of different size (ibid). They also point out that these two temporal integration windows are responsible for the fact that signals are analysed in a discontinuous rather than a continuous fashion.

At this point, one has to raise the question at what stage in language acquisition it all happens. Liberman and Mattingly argue that prelinguistic infants are already able to categorize phonetic distinctions like adults (1985, p. 24). Also, up to the age of one, this ability is valid for all speech sounds, but disappears for those that do not occur in the mother tongue(s):

> "[…] the sensitivity of infants to the acoustic consequences of linguistic gestures includes all those gestures that could be phonetically significant in any language, acquisition of one's

> native language being a process of losing sensitivity to gestures it does not use. […] the phonetic mode, and the perception-production link it incorporates, are innately specified." (1985, p. 24).

Whether innate or not, these observations certainly explain, why phonetic categories that become phonemically relevant for languages learnt later in life cause the language learner such problems, both in perception and production.

Concerning phonemic awareness and the acquisition of phonology, the motor theory suggests that once a child has appreciated the gestural source of sound, it will master the phonetic structure (Liberman & Mattingly, 1985, p. 25). Even though children make (the most stunning) phonological mistakes, they are able to perceive and identify correctly the systematic relations of sounds to meaningful units. The theory promulgates that "Further constraints become available as experience with the phonology of a particular language reduces the inventory of possible gestures and provides information about the phonotactic and temporal restrictions on their occurrence" (1985, p. 27). It ignores, however, the relevance of meaningful phonological units for the process of lexical attribution to certain phonemic/ phonological chains, which is essential for access to the phonological route in reading models such as Colthearts dual route model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

Liberman and Mattingly (1985, p.27) also pick up the notion of 'modularity' as introduced by Fodor (Fodor, 1983), and attribute speech perception to a unique module a "[…] piece of neural architecture that performs special computations required to provide central cognitive processes with representations of objects or events belonging to a natural class" (ibid.). According to them, this class or 'domain' "[…]controls perceptual processes", which are not cognitive, hence the processor is unaware of the computations, which leads them to agree with Fodor in suggesting that "[…] the perception of language is neither cognitive nor auditory" (1985, p. 28). Of course, they still refer to speech perception, not higher language levels. They conclude that "[…] speech perception uses all the information in the stimulus that is relevant to phonetic structure: every cue proves to be an actual cue" (ibid.). In Liberman and Mattingly's view the McGurk effect combines "[…] relevant optical information […] with relevant acoustic information" resulting in a "[…] coherent phonetic percept, in which […] the bimodal nature of the stimulation is not detectable." (ibid.). However, to date, it has not been shown that speakers who are not susceptible to 'duplex percpets' or McGurk items are in any way suffering from a speech perception deficit. What Liberman and Mattingly refer to as 'auditory suppression' has later been identified as excitatory and inhibitory mechanisms which are also present in audiovisual speech perception (cf. Massaro & Jesse, 2009 on multimodal neural circiuts). Liberman and Mattingly's conclusion that, in duplex conditions, the module makes a mistake, does no longer hold. Subjects, who report either the visual or the acoustic aspect of McGurk items, obviously perform some operation that allows them to solve the conflict caused by the ambiguity of the two signals. It remains to be explained, how this mechanism works and whether there exists a neural network where audiovisual speech is processed in a McGurk manner as opposed to a non- McGurk manner. Liberman and Mattingly further claim, that one could compare "how the set of possible gestures is specified for the perceiver" to Chomsky's universal constraints for syntactical and phonological forms (1985, p. 29), but refrain from this idea, as the "[…] knowledge of the acoustic-phonetic properties of the vocal tract, unlike other forms of tacit knowledge, seems to be totally inaccessible" (ibid.). If that is true, it is bad news for all second language learners, who struggle to master new phonemes.

2.1.2 Speech Perception on the Neural Level

What happens Where? Evocative of Hickock and Poeppel's ventral and dorsal pathway model this question may be answered with the help of Poeppel et als' scheme for the operations required in speech perception: the auditory input is first encoded as the waveform at the auditory periphery, from where it is analysed in the auditory pathway according to time and frequency. Poeppel et al. speculate about a 'phonological primal sketch', an intermediate device that maps the spectro-temporal representation of the acoustic signal onto the lexical-phonological representation of the word (cf. Poeppel et al., 2009, p. 253). In the final stage the segments, consisting of the individual, distinctive features are represented.

Poeppel et als' 'phonological primal sketch' (PPS) may be a categorical representation, i.e. because it consists of temporal windows of different sizes – it retains acoustic properties which differ according to length: short for segments, longer for syllables (2009, p. 253). In other words these two time frames, comprise a 'phonetic' segmental time frame with a level of up to one feature per segment and a 'phonological' representation for syllable- level generalizations. Again, the question is, what would the notable consequences of a perceptive speech impairment at this level constitute? The segmental level would represent categorical perception - but a deficit here does not altogether explain a phonological deficit. A serialistation problem on the syllabic level i.e. a processing deficit within the syllabic temporal primitive, for instance consonant cluster reduction would result in phonological problems. Here, one may pose the question if the 'phonological primal sketch' might be the function that is impaired in dyslexics.

In Poeppel et als' scheme the locus of this PPS is the superior temporal cortex, imbedded – probably – in Hickok & Poeppel's ventral pathway: STG → STS being the "relevant part of the cortex to construct an interface representation" (Poeppel et al., 2009, p. 253). However, the phonological deficit hypothesis has been questioned strongly in recent publications, most prominently by Frank Ramus (Landerl et al., 2013; Ramus, 2001, 2004, 2013; Ramus, Marshall, Rosen, & van der Lely, 2013; Ramus & Szenkovits, 2008). Even the existence of the phoneme per se has been questioned as early as the 1980s (cf. Kaye, 1989 on 'The death of the Phoneme'). The question remains, if phonemes are not existent and features are exclusively relevant in speech perception what does the grapheme correspond to then and what acoustic (or visual, or both) aspect of speech influences its representation?

Poeppel at al. offer three steps, required for the transformation of acoustic signals to internal representations, which are most important in this context here (2009, p. 254):

> (i) Multi-time resolution processing in auditory cortex as a computational strategy to fractionate the signal into appropriate ‘temporal primitives’ commensurate with processing the auditory input concurrently on a segmental and a syllabic scale
> (ii) analysis- by- synthesis as a computational strategy to linking top-down and bottom-up operations in auditory cortex
> (iii) the construction of abstract representations (distinctive features) that form the computational basis for both lexical representation and transforming between sensory and motor coordinates in speech processing.

Poeppel et al. (2009) embed these processes into the following anatomical network:

> “The initial cortical analysis of speech occurs bilaterally in core and surrounding superior auditory areas. Subsequent computations (typically involving lexical-level processing) are largely left lateralized (with the exception of the analysis of pitch change; the analysis of voice; the analysis of syllable-length signals), encompassing the STG, anterior and posterior aspects of the STS as well as inferior frontal, temporo – parietal and inferior temporal structures. This listing shows that practically all classical, peri-Sylvian language areas are

> implicated in some aspects of the perception of speech. […] the processing of speech at initial stages is robustly bilateral, at least at the level of core and surrounding STG."
> "The[…] primary (core) auditory cortex builds high-fidelity representations of the signal, and surrounding non primary areas differentially 'elaborate' this signal by analysing it on different time scales." (pp. 255-256)

Poeppel et al. (2009, p. 256f.) also relate to the established "where/how" and "what" parallel pathways of vision when they talk about the auditory pathways that Hickok and Poeppel have introduced in 2004. In this model, the "what" pathway, mapping sound to meaning, is the ventral pathway involving aspects of the temporal lobe – which aspect of it is responsible for lexical access is, however, not yet clear. The dorsal pathway supposedly plays a role in speech processing i.e. sound to articulation mapping and comprises temporo-parietal, parietal and frontal areas. Again, if a motor problem in dyslexia is to be detected, the responsible brain area that plays a role in transformation from auditory to motor coordinates would be temporo- parietal. Poeppel et al. also attribute speech processing to Broca's area and they highlight in this context recent findings that have 'challenged the view that Broca's area is principally responsible for production tasks or syntactic tasks and have reinvigorated the discussion of a 'motor' contribution to speech perception'. They also point out, how these novel research developments pertain to recent mirror neuron approaches and the motor theory of speech.

The timing of speech perception plays an important part as well. If we look at the timing of speech perception, we find some undisputed benchmarks (Kenneth N. Stevens, 2002). Poeppel et al. consider speech perception as a 'multi-time resolution process' with perceptual analyses occurring at two time scales. As pointed out earlier, they rely on "two principal time windows within which a given auditory signal (speech or non-speech) is processed" (2009, p. 249). They consider these two temporal integration windows responsible for signals being analysed in a discontinuous rather than a continuous fashion. The signal would consequently be “sampled” within these two windows with one spectral analysis of rapid temporal changes, reflecting glottis action and the other processing narrow band frequencies as in formants. To the author’s knowledge, none of the experiments quoted by Poeppel et al. to endorse this hypothesis were carried out with dyslexic subjects. An experiment like the one conducted by Lu et al. (Lu, Liang, & Wang, 2001), who tested for minimum stimulus onset asynchrony (SOA) by varying this asynchrony between clicks, would be of particular interest to be tested with dyslexic subjects. Such an experiment should include a paradigm as devised by Saberi & Perrott’s (1999, cited by Poeppel et al.), in which they showed that if sentence slices were reversed in direction, it would - up to 50ms segment duration - not significantly affect sentence intelligibility.

2.2 Visual Speech

How important is facial information in speech perception and is the visual speech signal alone sufficient for communication? Undoubtedly, when visual information is available, it is beneficial for the listener, but which parts of the talking face are really reliable, or in other words, where does the listener have to look, in order to profit from a talking face? Once again, the ‘viseme’, which was introduced earlier, becomes important here: the term represents the ‘visual’ phoneme, which in script corresponds to the grapheme. Massaro (Massaro, 1988, 1998, 2012; Massaro & Bosseler, 2006) uses the term to describe visual perceptual categories, comparable to acoustic categories, which, coincidentally play a vital role in the auditory processing deficit as discussed by various researchers to be the underlying deficit in dyslexia (cf. chapter 3.2). The visemes correspond to the phoneme equivalence class (cf. Auer & Bernstein, 1997).

Only recently, Lynne E. Bernstein has discussed research on visual phonetic perception and commented on how accurate visual word recognition is facilitated (L. Bernstein, E., 2012). She argues that even though visual speech stimuli are phonetically 'impoverished', the phonetic information is not so reduced that it would result in complete unintelligibility (cf. L. Bernstein, E., 2012, p. 32). In congenitally deaf subjects the accuracy of word identification was reported to reach accuracy levels of 48-85% in word recognition tasks (L. Bernstein, E., 2012, p. 23). Nevertheless, one mustn't neglect two important aspects: An eyetracking study suggests that congenitally deaf subjects perceive visual speech in their L1, that is, in sign language, within a much larger visual field than hearing learners of sign language do (Landsgesell, 2011). Another study tested deaf individuals' peripheral attention and suggested that the visual language processing field is much larger, thereby allowing to process more information than hearing subjects would process – the authors relate that to the reorganization of cross-modal brain areas) (Bavelier, Dye, & Hauser, 2006). Secondly, word recognition studies always have the lexical component, a variable that may only be controlled in pseudoword tasks, such as the ones I have used here. The lexical equivalent to Auer and Bernstein's phonemically equivalent classes are termed LECs (Auer & Bernstein, 1997; L. Bernstein, E., 2012) LECs are defined as "[…] the set of words rendered notationally identical by re-transcribing words in a lexicon in terms of a set of PECs" (L. Bernstein, E., 2012, p. 27). Dominic W. Massaro and Alexandra Jesse carefully describe which aspects of a talking face are relevant (Massaro & Jesse, 2009, p. 24ff.). On the basis of their descriptions and suggestions, the areas of interest were defined for the speech perception experiments in the empirical part of this book. In the eyetracking analysis of the audiovisual stimuli, fixation had to lie within these AIs, or the trial would be removed from data analysis, because it could not be guaranteed that visual information from the talking face was being processed. There is no doubt that movement of the lips counts as one of the most important factors of visual speech - hence the terms lip-reading - by laymen and scientists alike. Closed lips will convey that a bilabial consonant is about to be produced, the lip rounding that may accompany closed lips provides information about the following vowel. The acoustically similar nasals [m] and [n] can immediately be distinguished when the visual aspects are included. Labiodental and dental sounds have a high visual identification potential, as do movements of the tongue towards the alveolar and dental region. Places of articulation towards the back of the mouth (i.e. the soft palate and the velum) are not considered to be visually discernable, depending, of course, also on the preceding or following sound (cf. Tuomainen, Andersen, Tiippana, & Sams, 2005 on McGurk items in Finnish). Also, lowering of the jaw contributes visually to sound identification. Hence, the dynamics of articulation additionally comprise a visual aspect that may enhance an acoustic signal, facilitating bimodal speech processing.

2.3 Describing Audiovisual Speech Perception

Speech perception is multimodal. Sensory signal processing aside, it is beyond doubt that there are motor aspects involved as well. And, it can be shown through visual speech processing that visual aspects are involved as well. It may also be considered a fact, that speech perception involves modalities such as acoustic processing of the speech signal on a phonemic level and visual processing of moving articulators, given that these are visible and more or less synchronous unlike modern telecommunication functions such as skype – as opposed to the ‘good old’ telephonic signal transmission, with the acoustic downside of omitting various bandwidths. Suprasegmental modalities such as prosody and emotion are processed in audiovisual speech as well as hand gestures, and emotional face gestures (Massaro & Jesse, 2009).

Audiovisual speech integration has intrigued researchers for quite some time and has gained renewed interest with techniques such as EEG and fMRI. As mentioned earlier, the McGurk effect is a robust phenomenon, which means, it can still be triggered, even when subjects have realized how it is created. This again shows, that audiovisual speech is a highly automatized process that occurs as soon as both modalities are available. The fact that it is not possible for the vast majority of listeners to ignore one of the two input signals in the McGurk effect trigger signals, supports theories of audiovisual speech integration happening at the basal neural level and is beneficial across all linguistic levels.

Ruth Campbell discusses studies that suggest how "all linguistic levels are susceptible to visual influence" and comments on how "cortical correlates of seen speech suggest that [...] 'auditory speech regions' are activated by seen speech" (R. Campbell, 2009, p. 133f.). She also introduces two main modes of audiovisual speech processing, a 'complementary mode, whereby vision provides information more efficiently than hearing' and a 'correlated mode, whereby vision partially duplicates information about dynamic articulatory patterning' (ibid.). By these two modes she means that vision provides information about 'some aspects of the speech event' that is hard to hear, and - given that the speech readable face portions are visible- this will activate the 'complementary mode'. For her 'correlated mode', the 'temporo-spectral signature' of the speech stream is decisive, as it will activate regions of similar dynamic patterns across both audible and visible channels.

2.3.1 The Robustness of AV- Integration

The McGurk effect is not just robust. Audiovisual integration even tolerates some signal deterioration, as for example when seeing the talking face from a greater distance, and when the face is blurred or not fixated foveally, but parafoveally or when the acoustic signal is degraded, for instance, through masking noise or when the signal is synthesized (Baart et al., 2012; Bartlett et al., 2000; Blomert, Mitterer, & Paffen, 2004; C. S. Campbell & Massaro, 1997; Massaro, 1998; Kaisa Tiippana, Puharinen, Mottonen, & Sams, 2011). The processing and simultaneous integration of the two modalities still occurs when listeners are aware of the creation of McGurk items, or when they are instructed to ignore the visual or the acoustic input, suggesting that listeners they cannot suppress audiovisual integration of speech signals (Massaro, 1998; Massaro & Jesse, 2009). As discussed earlier, even some temporal misalignment is tolerated by listeners, albeit the McGurk effect in consonant-vowel (CV) items becomes weaker very quickly when the acoustic plosive is temporally misaligned to the visually perceived parting of the lips in bilabial or alveolar plosives.

Massaro and Jesse comment on various robust aspects of McGurk items, highlighting how acoustic /da/ and visual /ba/ lead to a combined /bda/ response, whereas the acoustic /ba/ and visual /da/ does not lead to a combined but a fused, i.e. McGurk – /da/ response (Massaro & Jesse, 2009:22). In their article on speechreading in language impaired children, Meronen et al. (2013) cite various Finnish McGurk studies, suggesting that due to the Finnish phoneme inventory, the voiced McGurk effect cannot be triggered (cf. Andersen, Tiippana, & Sams, 2004; Hayes et al., 2003; Meronen, Tiippana, Westerholm, & Ahonen, 2013; Saalasti, Tiippana, Katsyri, & Sams, 2011; Kaisa Tiippana et al., 2011). According to these studies, for Finnish listeners the McGurk /ta/ response for audio /pa/ and visual /ka/ is not as robust as for audio /pa/ and visual /ta/. They also suggest that this results from visually highly discernable places of articulation for /k/ and /t/, which is also shown in lip-reading tasks. However, in the naturalistic i.e. not over-articulated stimuli used here, it cannot be confirmed that visual /k/ and /t/ are highly distinguishable. Quite to the contrary, in a CV item such as the naturalistically articulated pseudoword items /pelami/ /telami/ /kelami/, the visual aspects of /t/ and /k/ were hardly discerned as different by subjects and controls.

The fact the audiovisually incongruent items also cause strong Mismatch Negativity (MMN) responses (cf. Colin et al., 2002; Mottonen, Krause, Tiippana, & Sams, 2002; Sams, Aulanko, et al., 1991) endorses the idea that perceptual information unfolds in a temporally sensitive way, hence the McGurk effect decreases as temporal misalignment increases. Moreover, in some sounds, visually relevant information for sound identification is available before the acoustic information. This is, for instance, true for the voiceless bilabial plosive, and also occurs in coarticulation processes. Coarticulation, incidentally, was one of the crucial aspects in Liberman and Mattingly's revised motor theory (1985: pp. 13-15). One could, of course, argue, that the sound inventory of the language in question is essential to this dimension. In the middle Bavarian version of German, which was the native language to all participants in the experiments, there is no phonemic onset difference between [p] and [b]. The temporal aspect of audiovisual speech and the advantage of some visual aspects in identifying speech are discussed by Massaro and Jesse in some detail (2009, p. 23). They suggest, in alignment with other researchers (Greenberg & Arai, 2004; Lewald & Guski, 2003; van Wassenhove, Grant, & Poeppel, 2007)that "[…] Accurate recognition performance […] actually improves […] when visual information leads by about 80-120ms", and that "[…] early arriving visual place of articulation information might 'prime' (Greenberg & Arai, 2004, p. 1068) speech representations that share this place of articulation", which constitutes a significant advantage over unisensory speech processing (Massaro & Jesse, 2009, p. 23).

2.3.2 The Benefits of Audiovisual Speech

Campbell (2009, p.139f.) provides two reasons for the superiority of audiovisual processing to auditory processing: one reason is the clear visual contrast in some segments, which fosters disambiguation in acoustically confusable segments. The other reason, according to Campbell is that various features of an utterance are perceived both, by eye and ear. The dynamics of speech production are said to be responsible for the correlation between audible and visible patterns – an explanation that again pays tribute to the motor theory of speech (Liberman & Mattingly, 1985).

Presence and processing of the face is relevant in communication due to various factors, for instance, emotion is better decoded, so are turn taking cues in dialogue and - most importantly for this kind of experiment - the presence of a talking face leads to better intelligibility of communication (Massaro & Jesse, 2009, p. 19). As Sumby and Pollack have demonstrated as early as 1954, the presence of the visual speech signal enhances the listener's ability to process speech signals drastically. Massaro and Jesse build on Sumby and Pollack's 1954 findings who compared the audiovisual benefit to a 15 dB change in signal to noise ratio (SNR), as confirmed by Grant and Seitz in their 2000 study of visible speech cues in auditory detection of sentences (Massaro & Jesse, 2009, p. 20).

While Massaro and Jesse are convinced that auditory speech is generally more beneficial than visual speech – to use their terminology – 'visible speech', they also state that the "[…] audiovisual recognition benefit emerges from both the complementary and redundant nature of visual and auditory speech information" (ibid.). Hence, the benefit of audiovisual information is, that whenever one modality is disambiguating or more informative than the other, for instance, the visual aspects of the bilabial nasal or the alveolar nasal and the acoustics of voicing, the complementing nature of both modalities improves speech perception significantly (ibid.). What has not yet been examined to the author's knowledge but might be of some relevance, is to investigate how beneficial audiovisual speech is to adolescent/ adult learners of a second language, since the ability to profit from visemes may be exclusively reserved for infant learners (Kuhl & Meltzoff, 1982; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Massaro & Bosseler, 2006). In other words, do adult learners of a language profit as much from the complementary nature of audiovisual speech, for instance in noisy surroundings: a conference conversation during break with background noise or – more lexically challenging – a chat in a pub on some random topic? Massaro and Jesse refer to this, stating that "[…] the size of the audiovisual benefit […] depends on the distribution of information within and between these two modalities, or […] on the degree of redundancy, complementarity and audiovisual uniqueness" (2009, p. 21). Furthermore, the audiovisual benefit is subject to individual differences in the listeners and his or her ability to process the information (ibid.).

It will become obvious in the discussion of subjects' performance in the experiment tasks - the unisensory tasks (acoustic only & visual only) and the multimodal tasks (audiovisually congruent & incongruent) - that some individual variability has to be taken into account. Since I have worked with language- impaired subjects (dyslexics) and have hypothesised about a deficit in visual and audiovisual speech perception, Massaro and Jesse's remark concerning perceivers' audiovisual processing abilities corroborates my concern. Unaware of the current dyslexia deficit/ cause debate they argue that one would "[…] need to know if a poor result is due to poor integration of other processing factors, such as limited working memory capacities or difficulties in application of linguistic knowledge, or if it is due to less auditory and/ or visual information" (Massaro & Jesse, 2009, p. 21 relying on Grant et al., 1998,). Hence, they propose two reasons why audiovisual speech integration might be impaired or prevented. These reasons might on the one hand, fortify propagators of the working memory deficit and phonological deficit hypothesis in dyslexia. On the other hand, they might also be exploited by propagators of the visual spatial attention deficit and propagators of the temporal- auditory deficit (cf. Andersen et al., 2009 for visual spatial attention and Tallal, 2006 for auditory deficit). Both these modality deficits would impinge upon audiovisual integration of speech signals.

2.3.3 Audiovisual Speech and Lexical Representations

In audiovisual speech research, it is also of interest how lexical access is influenced by visual speech, even though in McGurk and MacDonald's original study the strength of the effect overrules lexical expectations (McGurk & MacDonald, 1976). This aspect was, of course, carefully considered when the stimuli for the empirical part of this book were created (cf. chapter 4.2). A lexical benefit of audiovisual speech integration engulfs the acoustic ability to distinguish visually confusable phonemes – the 'visemes' and the visual ability to distinguish acoustically confusable phonemes supporting the idea of the afore mentioned phonemically equivalent classes, the 'PECs'. Campbell (2009, p. 137f.) comments on the fact that the number of PECs varies interindividually, depending on an individual's speech reading skill. She also, very much in accord with Auer & Bernstein (1997), states that: "The reason why a relatively small number of PECs can suffice for identifying individual spoken words is that most words in English are relatively unique in their segmental and syllabic structure" (2009, p. 137). In addition, Bernstein refers to 'homophenous words', which sound differently, but share the same visemes, for instance closed lips suggest bilabial plosives or the nasal (L. Bernstein, E., 2012, p. 26). Such homophenous words cause the lip-reader difficulty, but they may receive some disambiguation from lexical processes such as context.

For purveyors of lexical representation theories, who consider lexical representations a tight link to theories of speech perception, such as Poeppel et al. 2009 this should be good news. In Poeppel et als' view, words are represented in the brain as a series of segments, "[...] each of which is a bundle of distinctive features that indicate the articulatory configuration underlying the phonological segment"(Poeppel et al., 2009, p. 250). They commit to phonological generalizations being stated over features and not holistic phonemes (ibid.). The central aspect of speech perception should therefore lie in the extraction of distinctive features from the acoustic input. Poeppel et als' link to phonological theories consists of the "[…] fact that the elements of phonological organization can be interpreted as articulatory gestures with distinct acoustic consequences" which would require "[…] a tight and efficient architectural organization of the speech system in which speech production and perception are intimately connected through the unifying concept of distinctive features" (2009, p. 251). But how would an impairment of this 'concept of distinctive features' present itself? Substitutions of features with other features is a well known symptom in child language impairments (cf. chapter 3.2), the severity of the impairment could be measured according to how far the feature classes are apart, for instance, an alveolar fricative being substituted with a palatal stop would be more disconcerting than an alveolar fricative being substituted with a postalveolar fricative. Tallal tried to implement categorical speech perception impairments as a cause for dyslexia, however this attempt has not seen 'high ground' in dyslexia research (cf. P. Tallal, 1980; Tallal, 2006).

Speech impairments have been identified in dyslexic children and children at risk for dyslexia (as discussed later in chapter 3.2). However, the majority of dyslexics does not present with obvious speech impairments that would point to a problem of discriminating distinct features and yet, some (subtle and hitherto not identified) deficit in the distinctive feature representation of dyslexics seems to be the best option for tackling the linguistic symptoms of dyslexia. Poeppel et als' hypothesis that distinctive features are "both the basis for speech representations and have acoustic representations" may serve this purpose best (2009, p. 251). They adapted Stevens & Halle's (K. N. Stevens & Halle, 1967), 'analysis by synthesis' algorithm, which is applied in the analysis of incoming speech (ibid.). This 'analysis by synthesis' algorithm might have an equivalent in the synthesising problem of speech sounds represented as graphemes in alphabetic scriptures that cause dyslexics problems. If we speculate that a synthesising problem exists in dyslexia, how would it present itself?

On a speech perception and articulation basis, we ought to find a serialisation problem probably induced by a temporal processing deficit, i.e. the process should be slower. There is already some evidence in dyslexia research for this, as discussed in detail in chapter 3. Such a temporal deficit should be measurable even though it would not be overt to naïve listeners. The synthesis deficit could also be useful in explaining why audiovisual integration is slower, if not impaired in dyslexics – and of course it could explain why dyslexics should not be susceptible to the McGurk effect.

Poeppel at al. (2009, p. 262f.) adopt an 'analysis by synthesis' or 'perception driven by predictive bonding based on internal forward models' approach to speech perception and lexical access. Their 'forward synthesis of candidate representations' is central to the word recognition by distinctive feature processing mechanism. Pseudoword processing, which has been identified to always challenge dyslexics (Ramus, 2001, 2004; M. J. Snowling, Goulandris, & Stackhouse, 1994; Stackhouse, 2006), should be facilitated by a 'mini-lexicon' of valid syllables and all possible parses in a first pass analysis. The 'phonological primal sketch' of the segmental time sequence is based on the 'major' phoneme classes and gives a neighbourhood of words matching the detected 'landmark sequence'. According to Poeppel et al. this primal sketch includes the information required to ('broadly') classify (certain) prosodic characteristics such as number of moras and syllables in the word (2009, p. 263).

While the motor theory of speech by Liberman and Mattingly (1985) did not specifically discuss speech perception in regard to lexical access, Massaro and Jesse clearly note that "[…] a full account of spoken word recognition should also consider the role of visual speech information" (2009, p. 23). The crucial role lexical representations play in speech perception can be demonstrated as listeners immediately try to match the incoming percept with stored lexical items. In order to rule out that lexical access computations and lexical identification processes would interfere with subjects' processing of the stimuli in the experiments conducted for this book, the target items I used here were exclusively non-lexical, i.e. nonsense syllables and nonsense words (see chapter 4.2). This would also prevent that subjects became influenced in their decisions by lexical equivalent classes (LECs, see Massaro & Jesse, 2009, p. 24), which are the visual aspects of highly confusable words. The LECs are based on the afore mentioned PECs (phoneme equivalence classes, cf. Auer & Bernstein, 1997). In trying to identify words that include or consist of highly confusable visual phonemes (i.e. visemes), listeners would therefore make use of or rely on their lexical knowledge. Massaro and Jesse use the phrase "[…] lexical permissibility constraints might greatly reduce the difficulty in identifying a word" (2009, p. 24) to underline how LECs influence speech perception. For this very reason prelexical items on the phonemic/ visemic level were used in the experimental paradigm instead of actual lexical items. This is owed to the idea that lexical items must be expected to impinge upon the process of speech perception on higher linguistic levels. Massaro and Stork (1998, p. 237) offer a striking example, how such lexical processes may interfere, when the acoustic nonsense words 'bab', 'paup', 'po', 'brive', become audiovisually fused with the visual input of 'gag', 'kauk', 'ko', 'grive', resulting in the McGurk sentence "My dad taught me to drive"; (I underlined the fusion phonemes in the sentence for better understanding).

Unfortunately, neither the afore mentioned phonemic equivalent classes (PECs) nor the lexical representation theories can sufficiently explain why the McGurk effect should work well with pseudowords that do not have a lexical representation.

2.3.4 Brain Locations of Audiovisual Speech:

Lesion studies e.g. Campbell, 1997, who described a motor-blind patient with bilateral damage to the lateral occipital cortices comprising V5 (C. S. Campbell & Massaro, 1997), do not show a uniform picture but support the presumption for an audiovisual integration area. The 'locus' of the McGurk effect has been examined carefully by several researchers with various imaging technologies such as fMRI (cf. for fMRI studies: Andersen et al., 2004; Calvert, Hansen, Iversen, & Brammer, 2001; Pekkola et al., 2006; Pekkola et al., 2005) and MEG/ EEG (cf. Colin et al., 2002; Sams & Hari, 1991; Sams, Kaukoranta, Hamalainen, & Naatanen, 1991). ERP studies conducted by van Wassenhove investigated the tolerance for temporal asynchronies in McGurk stimuli that have been discussed earlier (van Wassenhove, 2009; van Wassenhove, Grant, & Poeppel, 2005; van Wassenhove et al., 2007). FMRI studies have used Mc Gurk items with healthy subjects (Pekkola et al., 2006) suggesting where the 'Mc-Gurk brain area' might be found. In their 2012 study, Szycik et al. investigated which brain regions were activated by the McGurk effect in healthy subjects using a 7Tesla fMRI scanner (Szycik, Stadler, Tempelmann, & Munte, 2012).

According to the findings of these researchers, the 'locus', where the McGurk effect happens in the brain can therefore be reduced to the following areas, albeit researchers do not refer to these areas in unison: The superior temporal sulcus (STS), including the supramarginal gyrus the inferior parietal lobule, the precentral gyrus the superior frontal gyrus, Heschl's gyrus and the middle temporal gyrus.

An original overview of research offered by Campbell (2009) is listed below and – where applicable – supplemented by recent findings and commented on where necessary for the purpose of this book. Campbell's overview was chosen as it includes spatial and temporal aspects of audiovisual speech processing:

> (i) "Speech-reading in the absence of any auditory input activates auditory cortex…include activation within core regions of primary auditory cortex (A1). […] parts of the superior temporal plane adjoining the upper part of the superior temporal gyrus are activated by silent speech... auditory cortex within Heschl's gyrus might be activated by seen silent speech" (but see also Stevenson & James, 2009).
> (ii) "Speech reading tends to generate left-lateralized or bilateral activation […] in contrast to the usual finding for other face actions, […] which tend to show more extensive right- lateralized activation."
> (iii) "The middle and posterior parts of the superior temporal gyrus, including the posterior superior temporal sulucs (pSTS)."
> (iv) (Left) pSTS can show differential activation for congruent and incongruent audiovisual speech. […] Inhibitory activation for audio-visual compared with unimodal input can be observed in other parts of the superior temporal gyrus, and for incongruent audiovisual pairings within pSTS. […] pSTS is a primary binding site for audiovisual speech processing.
> (v) Inferior frontal regions, including Broca's region, and extending into anterior parts of the insula, are activated by speech reading. (R. Campbell, 2009, p. 140).

In Campbell's recapitulation of current audiovisual speech processing theories, she also draws upon the idea of a "how" and "what" stream, which originate in the primary auditory cortex and activate the left perisylvian regions (superior, temporal, and inferior frontal) (R. Campbell, 2009, p. 144f.). The "what" stream is supposed to run anteriorly along the upper surface of the temporal lobe and reacts to semantic aspects of an utterance. The "how" stream is expected to run dorsally through the superior temporal gyrus to the temporo-parieto-frontal junction and reacts to segmental aspects of speech. These two streams have attributed functions such as "aligning the segmental specifications of speech whether it is planned, produced or perceived", which – again – fits into central aspects of the motor theory of speech (R. Campbell, 2009, p. 144). Since both these streams project into Broca's area, which is always a 'popular candidate' for activation when speech is produced and perceived – particularly relevant for seen speech because of the supposed mirror neurons in Broca's area (cf. Andersen et al., 2004) – the speech processing and/ or producing network finds its beginning or end here. Campbell also points out that several studies have confirmed how Broca's area is "involved primarily in the selection of speech acts for production" and "is especially active in processing seen speech […] even when no overt speech action is required" but she, like many, remains critical of the 'mirror neuron' hypothesis (2009, p. 145).

The idea that mental representations of speech sounds are not segment sized units but built from distinctive features is widely acknowledged in phonological theory today. If these features provide the connection between articulation (the equivalent to motor action in Liberman & Mattingly's 1987 reprise of the motor theory of speech) and perception then features are dually represented as motor gestures and acoustic patterns. In Levelt's 1999 'Blueprint of the speaker' (Levelt, 1999) and in speech production & perception models like the TRACE model (McClelland and Elman, 1986) or the Logogen model (cf. de Bleser et als' adaptation in LeMO, 2004) this representation finds its equivalent in "phonological loops" and respective pathways input lexica, which are, however not linked in a direct route, therefore not as clearly pronounced as for example in Hickok & Poeppel's pathway systems. (Hickock and Poeppel, 2004; Hickock and Poeppel, 2000).

Could speech impairments be explained with specific impairments, for instance, too slow temporal processing of one or both of Hickok and

Poeppel's time windows? There is ample speculation and quite some crucial evidence that temporal processing disorders result in profound speech impairments (F. E. Gibbon, 1999; P. Tallal, 1980). It is therefore required to accommodate the processing disorders assumed to occur in dyslexia, in a model of speech perception.

Poeppel et al. (Poeppel et al., 2009) admit that it is unclear what happens during signal desynchronization and the McGurk effect elicited in desynchronized conditions is far from being as robust as in the synchronized version used in the empirical part of this book. If dyslexics' audiovisual integration is impaired as hypothesised, asynchronous stimulus material would be hard to control. This is the reason why audiovisual items with even the slightest lags were eradicated from the item sample in this study.

However, van Wassenhove et als' 2005 findings are crucial for the understanding of temporal processing in speech perception. Equally essential are their insights that listeners tolerate asynchronies up to 200ms, which is in balance with mean syllable duration of 200ms (cf. also Poeppel et al 2008; Greenberg, 2005). It would be very interesting to establish how large dyslexics' temporal asynchrony tolerance windows might be in relation to the currently widespread phonological deficit hypothesis that suspects the reading deficit on the syllabic level.

2.4 Theories and Models of Audiovisual Speech Perception

A noteworthy critique of the motor theory of speech is put forward by Massaro (1998; Massaro & Jesse, 2009), who disagrees with the notion of a 'phonetic module specific to auditory as well as visual speech', in which audiovisual integration takes place through the gestural representations of audio and visual speech (2009, p. 27). Massaro and Jesse suggest that "[…] there is no need to postulate a special processing module for speech", seeing as there is some evidence that "[…] the same processes involved in other pattern recognition domains can also account for the integration of multiple sources of information in speech perception" (ibid.) Massaro introduced a new model for speech processing, which will be discussed subsequently. The motor theory of speech has, however, also found support in audiovisual speech research and some of its fundamental principles are reborn in recent theories of motor speech, like the one propagated by Nusbaum (2011).

2.4.1 New Motor Theories of Speech

Howard C. Nusbaum's 2011 approach to communication focuses on the interaction between listeners and speakers and analyses this interaction also from such a perspective as to that 'specific social goals and motives' are satisfied (Nusbaum, 2011, p. 668). Nusbaum additionally comments on the idea that mirror neurons in the human brain would function as a motor system response to an observed action in another individual, an idea that stems from Rizzolatti & Craighero, 2004; Rizzolatti et al., 2001 (cited in: Nusbaum, 2011, p. 668). If the idea of the mirror neurons persists in speech science, it might be because of the explanation they offer relating to visual speech, for instance, when it comes to imitating gestures, that is, visually perceived gestures. This idea is not new (Kuhl & Meltzoff, 1982; Meltzoff & Borton, 1979; Meltzoff & Moore, 1977, 1979) but it gives rise to new motor theories of speech. In Rizzolatti & Arbib's (1999) evolutionary theory of language, which is based on manual meaning making gesture, they also suggest that the observation of another individual's linguistic gesture action leads to a representation of these actions (Rizzolatti & Arbib, 1999).

Nusbaum (2011) also refers to Liberman and Mattingly's motor theory of speech in support for what he calls 'functional linkage between production and perception of biologically significant signals' and an 'interaction among the underlying neural systems' (2011, p. 670). A new motor theory of speech would also have to take into account the aspect that mirror neuron research may back theories that language perception (Nusbaum does not solely refer to speech in this context) and language production rely on one motor system. When phonemes and syllables (i.e. sound patterns, legal to the language in question) are processed, resulting in recognizable speech patterns, the speech production model of Friederici et al. and Caplan (Caplan, 1996; Friederici, 2006) is applicable to explain this process. Hickock and Poeppel's 'ventral stream' (Hickok & Poeppel, 2004) also builds upon this notion that speech processing 'progresses from auditory patterns to phonological analysis', which would happen in the superior temporal gyrus (STG) to lexical processing in more posterior regions (Wernicke's area) to syntactic processing (Broca's area).

When Nusbaum refers to Hickock and Poeppel's 2004 paper, he states how "[…] speech perception is mediated by networks that are similar to the ventral/dorsal pathway in vision." (Nusbaum, 2011, p. 671). The speech perception pathway consists of the sound to meaning ventral stream (primary auditory cortex, ventrally and laterally to the posterior inferior temporal cortex) which projects to the anterior superior temporal sulcus, where sentence processing takes place. The dorsal stream in Hickock and Poeppel's model, projects from auditory areas via the parietal cortex to the inferior frontal gyrus (IFG) and the promotor cortex. It is attributed to language development and word learning as well as to storing information in a phonological buffer (for short term memory), and to the phonological loop in models such as the logogen model, for instance in de Bleser's adaptation of the logogen model in the clinical test battery LeMo (De Bleser, 2004). Complex constructs of ideas like the motor theory of speech (Liberman & Mattingly, 1985) and the 'analysis by synthesis' approach (Stevens & Halle, 1967) have speech perception always include the motor system. Nusbaum argues that in neuroimaging, ventral premotor activity should be visible, when a talking face is being processed (2011, p. 673). Skipper, Nusbaum and

Small (2005) showed in their study of listeners hearing and seeing others talk as opposed to only hearing them talk, how ventral premotor activity increases in the hearing and seeing condition (correlating with increased activity in superior temporal cortex regions). As the intelligibility of the talking faces improved (visemes), the premotor activity grew larger. This suggests that premotor activity is related to visible phonetic information. Nusbaum concludes that in McGurk stimuli, increased activity in the ventral premotor region should not only be expected but also explain where and how the effect is created. In this context he also mentions Stevens and Halle (1967), whose analysis by synthesis approach provides evidence for motor systems being involved in disambiguating similar acoustic phonetic segments. The motor system, as surmised by Skipper, Nusbaum and Small (2005), consitutes an active part of speech processing.

As sensory information, both from visual and acoustic representations, is decoded into motor representations in the premotor cortex, these representations might feedback into the sensory system. Hence, in order to fully encompass all linguistic information, the listener would have his sensory and motor cortices interact throughout the perception process. In Skipper, Wassenhove, Nusbaum and Small's study (Skipper, van Wassenhove, Nusbaum, & Small, 2007), they examined neural activity patterns during McGurk stimuli perception resulting in a BOLD time course response. For the McGurk stimuli (audio [pa] dubbed with visual [ka]) they found that activation in the ventral premotor region would best accommodate the McGurk percept of [ta]. An acoustic [pa] would be best attributed to the supermarginal gyrus, which shifts to the premotor region in case of a McGurk percept. The visual [ka] would be best represented in the middle occipital gyrus, where the actual mouth articulation is processed, and shifting from there to the premotor region for McGurk percept. For Nusbaum this is consistent with the hypothesis that "visual and auditory information, when fed to the premotor cortex, give rise to an activity pattern consistent with the McGurk illusion which then may interact with sensory cortices resulting in a final activity pattern consistent across all regions with the McGurk percept" (2011, p. 673). This would fit an "active theory of perception for which the lack of invariance between acoustic patterns and phonetic categories is resolved by an interaction between articulatory knowledge in the premotor representation of speech"(ibid.). Further evidence for cross-modal interactions of the articulatory knowledge providing phonetic constraint type can be found in Bensmaia, Killebrew & Craig (2006).

Nusbaum proceeds to ask whether premotor activity might also serve as a constraint in the absence of a visual input. Here, he relegates to Liberman and Mattingly's revised motor theory, in which motor knowledge is relevant in any case, that is, with or without visible articulator movement. Studies by Skipper et al. and Wilson et al. (Skipper et al., 2005; Wilson, Saygin, Sereno, & Iacoboni, 2004) also argue for a sensorymotor process as the basis of speech perception.

2.4.2 The Fuzzy Logic Model of Audiovisual Speech

Another very recent pattern recognition model that incorporates the neural processes, which might underlie audiovisual speech integration, is the fuzzy logic model of speech perception by Massaro (Massaro, 1998; Massaro & Jesse, 2009). Pattern recognition accounts are the basis of Massaro's 1998 Fuzzy Logical Model of Perception (FLMP), which was reviewed and discussed by Massaro & Jesse in 2009. This model suggests that speech perception considers all available sources of information, i.e. acoustic and visual, and follows 'domain-general' processes for 'domain specific' signals (2009, p. 27). Visual signals are processed as contributing information, which is then integrated with all other available information. For Massaro and Jesse "[…] Integration is a general algorithm that applies to all available sources of information." (2009, p. 28). The FLMP processes all sources of information, also lexical and content information as well as structural information such as phonological, syntactic, semantic and pragmatic. It is most noteworthy here that the FLMP evaluates all information separately and integrates it subsequently, based on fuzzy values rather than binary values. Pattern recognition is achieved when the information matches "prototypes for each possible alternative as stored in long term memory" (ibid.). Information matches a 'prototype', by which they mean "[…] summary descriptions of the best exemplars of a category" to a certain degree, "fuzzy truth values ranging between zero and one" (ibid.). Regarding the McGurk effect, the 0.5 value represents the ambiguity of support for the respective categories here. The integrated percept would represent this ambiguity between visual and acoustic signal.

The underlying neural mechanisms for the FLMP's algorithm build upon findings from Meredith (Meredith, 2002) and Stein and Meredith (B. E. Stein & Meredith, 1992). Massaro and Jesse suggest three types of neural representation for the integration process of auditory and visual information, the first being 'Sensory penetration' (2009, p. 29), which refers to one modality impacting on the other, in this case the visual input, which activates the location usually attributed to acoustic information processing. The second type of multimodal integration, 'simple feed-forward convergence', implies that simultaneous auditory and visual speech signals activate a third location reacting to both of these modalities (ibid.). The third type of neural representation is activated by simultaneous incongruent audiovisual input and behaves like an integration process but lacks a shared location where the two signals are integrated. Massaro and Jesse hypothesise that this signal is fed forward to other cortical areas, where the synchronous fusion might eventually take place, but they do not specify these areas (ibid.).

2.5. Linking Audiovisual Speech to Dyslexia Reaserch

The link to written language, which is particularly relevant to the hypotheses and research questions raised in this book, may be established by taking a closer look at Massaro's 1987 Bayesan rule scheme concerning McGurk items. This constitutes that the principle required for the combination of audition and vision should be the same in, for instance, the combination of written and heard or written and spoken syllables. Could some kind of McGurk effect therefore be triggered when subjects hear <ba> syllables while they (quietly) read <ga> syllables? If such a scripture induced McGurk effect could be triggered it would be worthwhile testing it with dyslexic subjects in whom the classical McGurk effect could not be elicited.

Since reading is itself such a highly automated process, shown by such phenomena as the Stroop effect, it does not seem likely, that these two modalities would be susceptible to McGurk effects. However, Sams et al. (Sams, Mottonen, & Sihvonen, 2005) have shown, that a strong McGurk effect can be triggered when subjects hear <pa> while the subjects themselves are asked to (simultanesously) produce <ka> syllables and see their reflection in a mirror. A weaker, yet significant effect could even be elicited in a condition where subjects uttered the syllables and heard incongruent syllables via headphones without visual feedback (cf. Sams et al., 2005). This condition would also be worthwhile testing with language impaired subjects to establish whether their audiovisual integration is intact.

In all the audiovisual integration experiments that we have conducted so far, significant McGurk effects were only robustly triggered when the temporal alignment of the visual and auditory syllables was astute. The video editing software Adobe Premiere Pro is one of the few programs that enable audio alignment in milliseconds rather than video frames (25/30 frames per second). Upon editing the audiovisual material, it was obvious that only a precise alignment with a tolerance ranging from -30 ms to +40 ms reliably elicits McGurk effects. Virginie van Wassenhove et al. (2007) came up with a considerably higher tolerance in subjects in their McGurk experiments (-30ms to +170ms) although their data showed that the effect gets weaker as the asynchrony becomes larger. Van Wassenhove's studies (2008; 2005, 2007; 2007) contribute to the 'multisensory interaction' approach, which serves as a fundamental basis in explaining av integration. Her studies show that the visual input elicits internal abstract representations, which predict the possible audio targets so that an audio lag is easier to process than a visual lag. The visual information 'processability' also varies in quality according to what Massaro (1998) considers speech readable (visemes) i.e. visible features (bilabial, labiodental, dental, interdental, alveolar, lip rounding,), which elicit possible candidate targets. It comes as no surprise then that bilabials show the most rapid and precise synthesis and temporal facilitation.

From articulation therapy studies with visual feedback techniques such as electropalatography (EPG); see Gibbon for an overview (2009) comes evidence for the profound qualities of visemes: the sounds that are most difficult in therapy are the posterior sounds that cannot be seen (e.g. velar and palatal stops and fricatives). With visual feedback techniques such as EPG or Ultrasound that make places of articulation visible to the patient, the motor skills in producing the target sounds are drastically improved. From this follows that so-called 'therapy resistant' sounds may be produced correctly after three to five therapy sessions and become automatized after only twelve weeks (F. Gibbon et al., 1998; F. Gibbon, Hardcastle, & Dent, 1995). In hearing impaired patients (see, for example, Kaltenbacher, Krotzer, Hummer, & Leyrer, 2012) and language development impaired children such therapies also result in better sensory judgement of speech sounds that pre- therapy could not be produced or distinguished.

2.6. Implications for Further Research

A question that has not yet been satisfactorily answered is, how non- susceptibility to the McGurk effect can be explained (cf. also Hickok & Poeppel, 2004). This question also requires an explanation why a listener chooses one of the two input modalities for his/her percept. Massaro and Jesse raise two questions concerning the nature of visual speech: "[…] whether visual speech contributes to perception through featural cues […] or through configurational cues […]", by which they must mean 'analytic', visual detail and 'configurational', holistic patterns (Massaro & Jesse, 2009, p. 26). They also discuss the static vs. dynamic informational benefit of visual speech, by asking "[…] whether static information or dynamic information plays a role in speech reading or audiovisual speech perception" (ibid.). While both, featural and configurational aspects contribute significantly to speech perception, static information promotes speech recognition and dynamics of visual speech contribute to the visual percept (Massaro & Jesse, 2009, p. 26). Undoubtedly, audiovisual speech is beneficial, because it may disambiguate the signal, and because it is robust, i.e. some distortion, both in the acoustic and in the visual signal, is tolerated. Distortion may come in the form of acoustic masking, visual blurring, or temporal misalignment of the two signals. However, some people seem to be 'immune' against the McGurk effect and the question remains why this is so.

A McGurk item may yield various percepts: e.g. from the visual <pa> <ka> <ka> that is presented with the acoustic [pa] [pa] [ka], the fused percept becomes /pa/ /ta/ /ka/, the actual McGurk effect showing a truly combined percept of the incongruent middle syllable. If only the visual signal is processed, the unisensory percept is /ka/ for the middle syllable and if only the acoustic signal is processed the percept is /pa/. What does this tell us about the listener's speech processing, when only the visual signal is perceived? What does it mean, when only the acoustic signal is perceived? While the first question remains unanswered for the time being, an answer to the second question, why only the acoustic input is processed, will be provided in the next chapter – at least for dyslexic subjects.

3. Dyslexia and Beyond: Reading and Speechreading

Stan: "What's the matter?"
Ollie: "Didn't you read it?"
Stan: "Yeah, but I wasn't listening"

Stan Laurel & Oliver Hardy, *Beau Hunks* (1931)

While most children acquire written language trough regular educational instruction, some 5-10% of all children[1], have serious difficulty in acquiring the skill of reading and writing (cf. M. Snowling & Stackhouse, 2006; Sprenger-Charolles, Colé, & Serniclaes, 2006 suggest up to 5% of all children). Up to the end of the 19th century, very little was known about the specific problem in written language acquisition that became known as dyslexia. In German speaking countries, it was not until the 1970s that dyslexia was identified as a specific language problem, independent of intellectual capabilities of children (cf. Spezialbibliographie der Universität Trier, 1985). The last five decades, however, have witnessed a remarkable research interest in dyslexia and with the advent of novel imaging technologies supplementing existent behavioural methods, new insights into dyslexia have been constantly gained by multidisciplinary researchers in neuroscience labs across the world.

---

[1]Other sources, e.g. Zvia Breznitz, consider a percentage of 10-15% or even up to 17,5% (Breznitz, 2008; R. I. Nicolson & A. Fawcett, 2008; S. E. Shaywitz, 1998).

In addition to the definition of dyslexia provided earlier, the World Health Organization also offers a more detailed definition in the International Classification of diseases, ICD-10, section F 81.0:

> "The main feature is a specific and significant impairment in the development of reading skills that is not solely accounted for by mental age, visual acuity problems, or inadequate schooling. Reading comprehension skill, reading word recognition, oral reading skill, and performance of tasks requiring reading may all be affected. Spelling difficulties are frequently associated with specific reading disorder and often remain into adolescence even after some progress in reading has been made. Specific developmental disorders of reading are commonly preceded by a history of disorders in speech or language development. Associated emotional and behavioural disturbances are common during the school age period." (World Health Organization, 2008)

Nevertheless, the WHO does not suggest what might cause these difficulties, hence the pivotal question for researchers trying to come to terms with dyslexia has long been how to identify the cause(s) of dyslexia.

Children typically learn to read around the age of six, and the majority of kids learn to read and write without strenuous effort. As they learn to read, children will apprehend how letters and sounds relate to each other and how specific sounds are then linked to print. However, in order to master written language skills, children must first become aware of structural aspects in their mother tongue. From here on we will refer to such aspects as 'awarenesses', which will be discussed in more detail in the first subsection of this chapter. In the subsequent chapter 3.2, it will be outlined how these awarenesses need to be applied – and adjusted – when written language is processed or produced. Neural mechanisms in the reading brain, with special regard to differences in typically developed reading skills and dyslexic readers, will be introduced and analyzed in the next section. The crucial question, which linguistic impairments and deficits may cause dyslexic reading behaviour, will be critically outlined and discussed in the fourth subchapter. In order to provide a conclusive overview for this very context, some of the pertinent causal suggestions and deficit theories are examined and further contrasted and debated. The principal engagement will be with theories that explain the linguistic aspects of the phenomenon dyslexia. Finally, the existent link between reading and speechreading, as previously introduced in chapter 2, is established and the theoretical assumptions underlying this connection between poor reading skills and speechreading abilities will be adduced. The approach of testing dyslexics' speechreading abilities in order to formulate an audiovisual speech deficit hypothesis will conclusively round off chapter three.

3.1. Linguistic Prerequisites for Reading

Learning to read and write is achieved by most children as effortlessly as they master oral skills in their native language. When children acquire these oral skills, more accurately referred to as 'meta-linguistic awareness', they do not only discover that words consist of segments and that these segments can be shuffled around to create new words with different meanings. They also uncover that oral communication spans from word meanings and ambiguities to pragmatic aspects and that spoken language follows a set of discrete rules: phonological, morpho(phono)logical and (morpho)syntactic structures. Even though it has recently been demonstrated in pertinent research that literacy has an influence on the previously acquired verbal skills (cf. Dehaene, 2013; Dehaene et al., 2010; C. A. Fowler, 2011), it goes without saying that the linguistic structures of the first language(s) are a prerequisite to written language acquisition. Mattingly even referred to reading as being 'parasitic' of spoken language (Mattingly, 1972). Nowadays, it is also beyond doubt that phonemic awareness is one of the fundamental precursors to learning how to read and write. But what constitutes phonemic awareness? And how does the child, having learned to talk, learn to read and write? The following subsection provides an overview of selected theoretical approaches, which try to tackle the issue how spoken language is translated into written language and how written language is translated into spoken language.

Nowadays, the terms 'phonemic awareness' and 'phonological awareness' are often used interchangeably in reading research. In spite of several linguists and phonologists, as, for example, Noam Chomsky, Morris Halle, Jonathan Kaye, Carol A. Fowler, Anne E. Fowler and others being quite critical of the term 'phoneme' and what it represents in phonological theory, there remains a need to describe and characterize phonological units which serve as the building blocks of spoken language (Chomsky & Halle, 1968; A. E. Fowler & Swainson, 2004; C. A. Fowler, 2011; Kaye, 1989). In reading research the term 'phoneme' generally refers to classes of phonetic segments used to distinguish words (cf. M. Snowling & Stackhouse, 2006; M. J. Snowling & Hulme, 2005; Sprenger-Charolles et al., 2006). These phonetic segments can be characterized by their 'features'. For instance, the features of /p/ are voiceless, bilabial and stop. In middle Bavarian dialects /p/ is pronounced lenis at the word-initial position, making 'Packerl' and 'Backerl' homophonous ('parcel' vs. 'cheek'). In standard German, word initial /p/ is fortis/ aspirated, facilitating the distinction between /b/ and /p/ thereby rendering the two sounds 'phonemic' (for a detailed discussion cf. Moosmueller & Ringen, 2004). Carol Fowler relates phonemes to graphemes, when she suggests in a simplified explanation that "[…] consonant and vowel phonemes are discrete from one another, and are invariant in their featural attributes. Entities with these characteristics are what the letters of an alphabetic writing system represent more or less directly depending on the writing system" (2011, p. 6). A phonological unit would hence be composed of segments, which can either consist of 'pronounceable feature packets' ('elements') or combinations of elements as suggested by Kaye (1989, p. 160). In children the ability to extract phonemes from words may be playfully assessed through games such as 'I spy with my little eye, something beginning with 't'…'. If a child finds objects whose denominations actually begin with the sound [t], it can be considered as established and present in the phoneme inventory[2]. The production of spoonerisms and sound omitting games also attests awareness of segments. For rhyming games, as well as for the ability to memorise nursery rhymes, the next higher level of phonological awareness would be needed, i.e. identification of nuclei and codas. This aspect can be easily assessed in young children through language play.

The phonological units of the respective language constitute the phonological inventory. Depending on language variations, i.e. various accents that are spoken in the language learner's environment such as in the aforementioned example for middle Bavarian, this phonological inventory needs to store representations of ambiguous pronunciation elements ('homophones'), which Kaye explains as follows: "[…] All languages contain homophones (e.g., English sea and see) […] speakers will make adjustments for such mergers without having to learn the new pronunciations as new words." (1989, p. 25). Through homophones that become 'heterophonic' in another accent of the same language, potentially represented in the orthography through different graphemes, the beginning reader/ writer becomes aware of the sound change. This phenomenon, of which the learner is quite unaware, can only result in a reorganisation of the phonological inventory and may serve as the most basic proof that literacy influences and alters the representation of language units, both, in the 'phonemic' and the 'visemic' modality. Another facet that highlights the importance of visual aspects in speech is the

---

[2] Naturally, this game also draws on lexical knowledge and object naming. Therefore, when the 'I spy' game is played at the breakfast table, a correct answer to what begins with 't' would also be 'breakfast', if, the sound in question were pronounced like the homophone 'tea' and the letter [ti:] rather than the [t].

point that has already been raised in the previous chapter, namely, meaningful information drawn from visual speech gestures. Even though lipreading occurs predominantly as a discrete process of which the listener is unaware, it facilitates information processing. If one considers the basic example of a child asking his or her parent for permission to do something, one will find that the child will be able to draw information concerning the anticipated answer also by carefully watching the parents mouth: while the child declares the desire to, for example be allowed to have another go in the merry go round, the parent might already start to prepare the oral gesture of lip-spreading as required in the German articulation of the answer "nein" (no). Hence while the child argues the point to convince the parent, the parent's answer may already be predictable through the prepositioned lips.

The acquisition of phonemes and consequently 'phonemic awareness' is a task that children have to master discretely, owing to the fact that when children learn new words, they tend to learn them as a meaningful word form, i.e. lexical units, rather than phonological units. C.A. Fowler sees the difficulty in childrens' acquisition of phonemic awareness in relation to A. Liberman's concept of a 'speech perception brain module', which is not overtly available to the language user, who cannot "introspect on the workings of the module" (2011, pp. 7-8). She also points to the difficulty that arises when phenomena like coarticulation make it tough for the child to identify the individual segments of a word. According to her, this difficulty is not always remedied through graphemic representations, as "[…] even languages with very regular and consistent alphabetic writing systems will have these ambiguous segmental properties, the letters of the alphabet only can come close to mapping in a one-to-one way to the basic phonological entities of the spoken language" (2011, p. 8). There is ample evidence promoting the crucial role of phonology in normal reading acquisition – as well as in impaired phonological processing at the core of reading impaired dyslexics, which will be discussed in later chapters (for an overview, see also Sprenger-Charolles et al., 2006).

Another set of covert rules that the child acquires automatically lies within the field of morphology and morphophonology. Morphemic awareness is also considered a fundamental prerequisite for the mastery of reading and writing. Drawing on Chomsky and Halle's phonological theory, Fowler posits that English spellings reflect 'phonological (near-) systematicities' which result in spellings that tend to be 'morphophonemic'. The reader's benefit is hence to be descried in such forms as the inflectional suffixes –s and –ed that can be easily detected in writing (C. A. Fowler, 2011, p. 10). But how does this benefit the beginning reader? The resultant asset for the reader thereby is twofold: firstly, the pre-reading child profits from the knowledge that graphemes map onto something meaningful, i.e. if a word is elongated, for instance by means of a suffix, the written word would also have to be longer. Secondly, once morphological awareness is established, it assists the reader in pronouncing the "[…] morphologically complex form in the neutral- suffix form" and helps him/her "[…] know how pronunciation changes in non- neutral suffixes" (ibid.). To recapitulate, Fowler states that morphological and phonemic awareness precedes morphophonological accuracy, which is a "[…] strong predictor of word decoding, but affects reading comprehension only indirectly through the effect of decoding on comprehension" (2011, p. 10).

This was another reason, why the McGurk effect items of the empirical part of this book comprised nonwords (e.g. /pelami/ or /mabali/) and single syllable sequences (e.g. /pataka/ or /badaga/), because they are independent of morphophonological accuracy. Such audiovisual material is therefore ideal for testing phonemic awareness in visual speech perception as there is no higher- level linguistic interference. McGurk items that involve context free lexical items may be susceptible to word frequency effects.

McGurk items with pseudohomophones and pseudowords would be suitable to assess morphophonological awareness, by adding typical inflectional affixes.

3.2 Learning to Read: from Phonological Awareness to Reading

Once the child has acquired the fundamental oral prerequisites of his or her mother tongue, written language mastery poses the next challenge. Children may develop an interest in written language long before school. This interest may be triggered by story books that are read to the children, or through written or drawn 'secret signs' that encompass a certain meaning, or through writings, which become recognizable to children through various modalities, e.g. the colour, size and font of a text logo. Also, the beginning letter of the child's name is often known, and his/her personal name is recognized. Proper nouns with which the child is familiar might follow the recognition of the personal name, especially in languages where proper nouns are capitalized. Furthermore, it is frequently noticed that children attempt to write their name or copy letters from the written forms of known objects or names. Cultural aspects prompting the importance of writing and reading, e.g. writing a wish list for Santa Claus (in German speaking countries the "Christkind" or the "Weihnachtsmann") or being aware that Santa has a list of all children, also encourage children to concern themselves with reading and writing. Especially with the help of computer keyboards, children may often learn letters before they receive instruction concerning sounds and their representations as letters.

Reading researchers have dedicated decades of research to the question how exactly the oral skills are mapped onto written language and what the fine-grained structure of this mapping process is. Therefore many models of reading development and of reading processes, as they might occur in the brain, have been expounded over the past decades. Some of the more promi-

nent and timely models will be briefly outlined and discussed here, before we devote more time to the neurological processes and neural networks involved in reading. Unsurprisingly, all reading models include phonological aspects at one stage or another and it will be shown that the fundamental role of phonological factors needs to be considered by all reading models. In a prominent model by Linnea Ehri, the above mentioned ability of pre-reading children to identify visual aspects of writing is considered in a 'pre-alphabetic phase' that transgresses via a 'partial alphabetic phase' to the full alphabetic phase and reaches the sight word reading phases (cf. Ehri, 2005). Another popular model along the same line is that of Uta Frith, in which she considers three stages, a 'logographic stage' for word recognition through the visual aspect identification, e.g. the beginning letter, an 'alphabetic stage' (phonological) which comprises phoneme to grapheme attribution and a 'final orthographic stage' which relies on the automatisation of phonological recoding and thus fosters sight word recognition (cf. Frith, 1985). A concise overview of the various protagonists' theoretical approaches to reading development as devised in phase or stage models is given by Ehri
(Ehri, 2005, p. 139).

The models by Frith and Ehri mentioned above are both able to explain phenomena such as recognition of text logos and word identification based on visual aspects such as beginning letter/ capital letter. They also include the fluent reading stage, which accomodates automatisation and sight word reading. The automatisation of reading is a trait that can also be measured by eyetracking. This technology is capable of tracking eye movements at a sampling rate of 1000 pictures per second and it can test reading speed as well as the width of a reader's perceptual span. With the help of eyetracking research it has been affirmed that young schoolchildren display eye movements that resemble adult readers' once they have mastered the automatisation stage (cf. Bergmann & Wimmer, 2008; Hawelka & Wimmer, 2005;

Hutzler & Wimmer, 2004; Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004; Rayner, 1998; Rayner, Juhasz, & Pollatsek, 2005; Rayner, Pollatsek, & Binder, 1998). Eyetracking further helped to uncover how the eye movements of dyslexic subjects differ from typically developed reading eye movements. The reading eye movement data obtained for this book show how adult and adolescent dyslexic subjects may have to resort to phonological recoding when the reading material becomes too difficult, as for example in pseudoword reading tasks (see figure 2, p.61).

In reading research, the notion of 'modelling the reading' process has also been crucial over the past decades. Reading models do not only have to yield reliable simulation results, but they also need to be able to incorporate impaired reading processes. A simplified model, drawn upon by many dyslexia researchers, is the triangle model of Seidenberg and McClelland. It has often been adapted among others by Snowling, with the well known two pathways, a phonological and a semantic pathway, which explain how words are translated into writing or translated from writing into speech (M. J. Snowling, 2006, p. 5).

This model, which has later been implemented as a connectionist model, is a precursor of dual route models, as the Coltheart model discussed in the following section. The triangle model suggests a simple phonological mapping of letters and sounds, resulting in a recoding procedure in reading. Via synbook of the segments the word is pronounced and its meaning is, literally 'deciphered'. Along a semantic mapping, word meanings may be extracted directly from the orthographical representation allowing its immediate pronunciation. It is considered a connectionist model in the sense that connections represent neural network connections. These connections are 'either-or' connections. The triangle model and its many successors stand in stark contrast to dual route models, which provide the pronunciation of a word via two, not necessarily independent routes. In dual route models, the reader uses a phonological route to attribute graphemes to phonemes and 'spell' out words, whereas the direct (or lexical) route relies on the visual word representation. This requires that the two above mentioned 'awarenesses', phonemic/ phonological and morphophonological awareness, have been established in the pre-reader. By means of these two awarenesses, the 'parameters' are set for the two reading routes, as proposed by Coltheart et al. (Coltheart et al., 2001). The dual route approach to reading yields the explanatory power to relate the reading of difficult words i.e. novel word or loanwords with challenging spelling or pronunciation such as "chanson" in English or in German, and of course reading of pseudowords. For such tasks the phonological route is quintessential as it allows the reader to phonologically recode/map the corresponding phonemes to the graphemes. The expected reading behaviour, as can be shown through eyetracking fixation measurements, may range from letter by letter recoding to the identification of larger phonological units. In novel words and novel pseudowords, the pronunciation of words should not be possible through the lexical route. Although it may be attempted at first to identify a novel word or pseudoword via the lexical route, one should always find eye movements that point to phonological recoding. The two fixation protocols below, taken from the pseudoword task of the empirical eyetracking research for this book, demonstrate such reading eye movements:

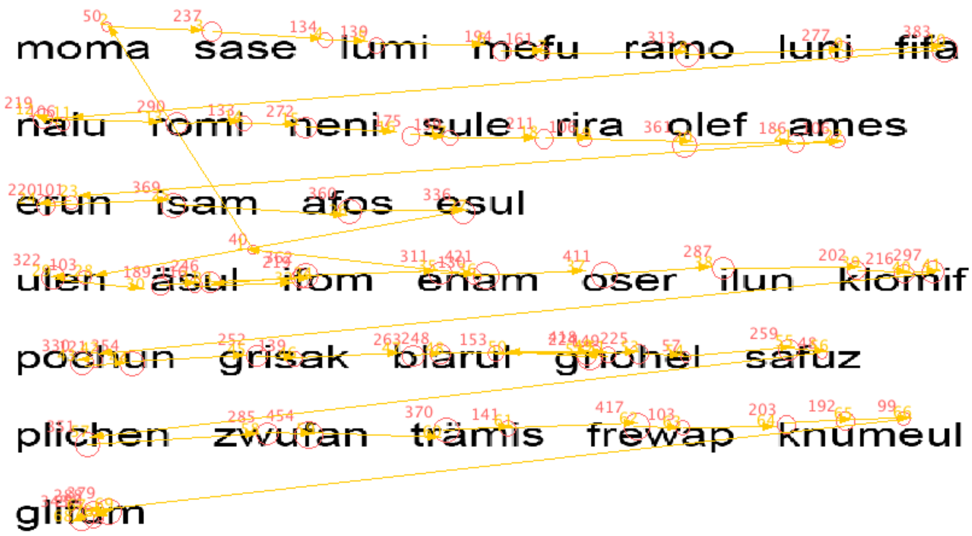


Figure 1: Eye movements of a non- reading impaired adult reading pseudowords. Fixation protocol created by the author, pseudoword items from Moll & Landerl (Moll & Landerl, 2010).

Figure 1 shows the (raw data) fixation protocol of an adolescent normal reader in an aloud pseudoword reading task (circles with numbers represent the fixations and their duration in ms, arrows represent the ballistic eye movements, the saccades). The reader in figure 1 made no errors in pronouncing the unknown pseudowords correctly. It can be seen that this reader is capable of identifying phonological units, sometimes he is even able to read the word within one sight word fixation.

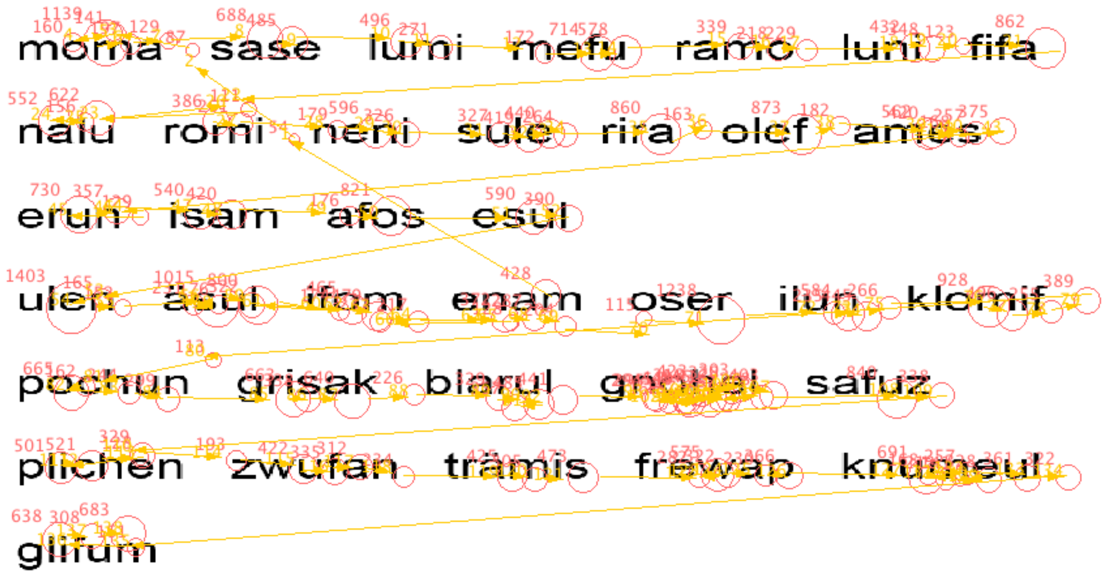


Figure 2: Eye movements of a dyslexic subject reading pseudowords. Fixation protocol created by the author, pseudoword items from Moll & Landerl (Moll & Landerl, 2010).

Figure 2 shows the fixation protocol (raw data) of an age matched dyslexic subject. This reader has to resort to phonological recoding at all times and the size of the phonological units (the perceptual span) is much smaller compared to the reader in figure 1.

For a more detailed overview concerning eye movements of beginning readers and dyslexic readers within the dual route approach, the research groups of Peter de Jong and Heinz Wimmer are suggested here (Peter F. de Jong, Bitter, van Setten, & Marinus, 2009; P. F. de Jong, Seveke, & van Veen, 2000; Martens & de Jong, 2006; van den Boer, de Jong, & Haentjens-van Meeteren, 2012; Wimmer et al., 2010).

Van Orden and Kloos also address the dual route issues in their discussion focussing on phonology and reading. In this context they claim that homophone errors, which are made by skilled readers may reflect use of both reading routes (Van Orden & Kloos, 2005). One would expect that the correction of reading mistakes which skilled readers produce and correct without effort, i.e. pronunciation mistakes and semantic errors, suggests that during reading the 'phonological loop' - as suggested in the Logogen model (cf. De Bleser, 2004) - is activated just as the semantic system is persistently being accessed during reading facilitating the choice of the most suitable candidate for the context.

Furthermore, Van Orden and Kloos comment on a feedback effect that has to do with auditory lexical decision: "[…] the word pint in an auditory lexical decision is spoken, but ambiguity in how pint's spelling might be pronounced slows down the lexical decision time – even though no spellings appear in the experiment." (Van Orden & Kloos, 2005, p. 74). Another effect that they consider highly relevant for the interaction between spelling and phonology in written and spoken items is that an item sharing the same onset phoneme as the previous one (their example is /k/ in 'cushion' and in 'coffee') is activated faster verbally when the graphemes are also the same. In the case of 'kidney' and 'camel', the /k/ onset does not constitute such a beneficial activation. Van Orden and Kloos conclude from this that spoken word production is influenced by the graphemic representation of the word resulting both in a feedforward and feedback interaction between phonology and spelling (ibid.). This finding serves the hypotheses well that audiovisual integration is not only an active component of reading and writing, but a prerequisite for skilled reading and writing appertaining to both routes. An impairment in audiovisual "awareness" of phonemes, graphemes and visemes should therefore also precipitate poor performance in pseudoword tasks as well as in automatized naming tasks.

3.3 The Reading Brain and the Dyslexic Brain

In the previous subchapter on audiovisual speech, the brain locations of audiovisual speech perception and of speechreading were discussed (subsection 2.3.4). This chapter will further ascertain the brain localizations for reading as well as the differences that may be found in normal readers and dyslexics. Pertinent research suggests that reading and reading acquisition processes may be ascribed to specific brain areas with very little variation across writing systems and cultural aspects of language representation. The most prominent research in regard to this academic field will be outlined to have resulted in what has been referred to as a 'reading brain map'.

3.3.1 Some Initial Neurological Perspectives

When it comes to the neurobiological realm, it is now commonly accepted that a cognitive weakness, such as, for example, poor phonological processing, is most likely to be caused, not by one, but by several different neural irregularities (cf. R. I. Nicolson & A. J. Fawcett, 2008; J. Stein, 2008). Hence, the case that one of these neural abnormalities is able to account for all instances, can most likely be discarded. When considering the utterly complex and inherently inter-dependent systems constituting the brain, the search for the 'one same underlying reason why' has to be given up in favour of a more versatile approach capable to account for a host of intricately linked possible causes.

Equivalent to other fields of cognitive science dealing with complex findings, the idea of interdisciplinarity, which implies pooling various kinds of expertise and making concessions to other perspectives, has become central in dyslexia research. It is consequently not surprising then, that the more 'proximal' explanations are supplemented by deeper 'distal' ones, which again implicate a certain 'removedness', by which the obvious link to the problem, as for example to reading, becomes less apparent. When it comes to dyslexia, examining and studying the distal factors which predict the three criterial difficulties for dyslexia – reading, writing and spelling - can offer us potent insights into how the brain copes with them. One of the main discoveries concerning the visual system has c ertainly been that the different qualities of visual targets are not analysed one after the other but simultaneously by separate, parallel pathways working simultaneously. This separation starts already in the retina and the two major kinds of ganglion cells, whose axons project the visual information back to the brain are: the larger magnocellular cells (m- cells) and the more numerous but smaller parvocells (p-cells). The main reading related function of the visual m- cells and p- cells is as follows:

> "[…] the m- system provides the brain with rapid information about the location and broad outline of letters and words primarily for directing attention and eye movements to each one if necessary, as is required when learning to read a new word. But parallel processing by the p-system is required to identify the fine details of letters, for example to distinguish between 'm' and 'n' or 'c' and 'e'" (J. Stein, 2008, p. 59).

Related to this magnocellular theory, one finds ample research that suggests that the underlying deficit in dyslexia might be attributed to a visual deficit or a visual attention deficit. This idea has received lavish attention (cf. Lobier et al., 2012; Peyrin et al., 2012; Valdois et al., 2011; Valdois et al., 2004; Vidyasagar, 2004; Vidyasagar & Pammer, 2010).

Over the last couple of years a fair amount of evidence has been accumulated, showing that the m- ganglion cells of many dyslexics exhibit significantly slower responses (cf. Pammer & Wheatley, 2001). Since the m- system is highly sensitive to low spatial and high temporal frequency stimuli, contrast sensitivity to such stimuli is more likely to be lower in dyslexics (cf. Lovegrove, Bowling, Badcock, & Blackwood, 1980) and the visual motion area V5/MT, which receives most of its input from the m- system, has shown to be less activated by visual motion stimuli in dyslexics (cf. Eden et al., 1996) resulting in a lower sensitivity to visual motion (cf. Cornelissen, Richardson, Mason, Fowler, & Stein, 1995). Furthermore the fast track m- inputs to the dorsal stream are for controlling eye movements as well as mastering the focus of visual attention (J. Stein, 2008, p. 62) and several studies have shown that dyslexics are poor at serial visual research and disproportionately affected by distractors (Facoetti, Turatto, Lorusso, & Mascetti, 2001; Iles, Walsh, & Richardson, 2000). Consequentially a reasonable amount of studies have supported the idea that dyslexics often display m- cell weakness (for an overview, see J. Stein, 2008). Interestingly enough, we also find large neurons having rapid temporal processing capacities in the auditory pathways, which have been called 'auditory magnocells' due to them being recognised by magnocellular specific antibodies (J. Stein, 2008, p. 66). In this context Stein raises the possibility "that impaired auditory and visual temporal processing are among a number of possible causes of impaired phonological processing" and that "most dyslexics, for example, have both lowered visual motion sensitivity and lowered sensitivity to changes in auditory frequency" (ibid.).

Nevertheless, these large magnocellular neurons are not confined to the visual and auditory system but can also be found in the somatosensory, the memory and the motor systems throughout the brain. Based on this, Stein proposes the idea that the root cause of dyslexia might be an impaired development of all these magnocellular systems (J. Stein, 2008, p. 67) and he relishes the opportunity to further suggest that the development of all the temporal processing skills (visual, auditory, motor and memory) which are necessary for reading are, most likely, controlled by genes like KIAA 0319 (J. Stein, 2001). Hence different alleles might affect different individuals more in one system then in another, which would explain the difference in visual and auditory systems being affected, i.e. one dyslexic is weaker visually, the other auditorily.

As we have discussed up to now, reading, writing and spelling require complex coordination processes between different brain regions and the cerebellum has, in this context, long been established as the key brain region for coordination of actions (cf. Eccles, Ito, & Szentagothai, 1967; Ito, 1984). However, the cerebellum almost always works in conjunction with other brain areas (cf. R. I. Nicolson & A. J. Fawcett, 2008, p. 80). Key features of the cerebellum are the great plasticity of the system (Holmes, 1922) as well as its role in automaticity (Brindley, 1964) and its crucial part in most cognitive skills associated with speech (Desmond & Fiez, 1998). The compelling idea, that the cerebellum really does play such a vital role in cognition inspired Nicolson et al. (1995) who thereafter conceived their cerebellar deficit hypothesis, a 'causal explanation' of dyslexia:

> "Cerebellar impairment would therefore be predicted to cause, by direct and indirect means, the 'phonological core deficit' that has proved such a fruitful explanatory framework for many aspects of dyslexia. Of course the central role of the cerebellum in skill automatization also provides a principled explanation of automatization deficit, the second major 'cognitive' explanation. It also provides a natural explanation of the more recent 'double deficit' hypothesis" (R. I. Nicolson & A. J. Fawcett, 2008, p. 86).

One major challenge, however, has been that the complex inter-dependent systems of the brain do not lend themselves easily to simple distinctions that would once and for all designate the roles of cause and correlate. All in all, a vast amount of empirical evidence has been produced over the last 20 years pointing to the idea that the cerebellum might be one of the key brain structures affected in dyslexia (for a summary see: R. I. Nicolson & A. J. Fawcett, 2008).

3.3.2 The Reading Brain

One question regarding the reading brain, however, became more and more vital: What are the cortical structures and neural pathways involved in reading? By means of modern imaging techniques notable insights into the neurobiology of reading have been gained over the last years. Research conducted with the help of PET (positron emission tomography), fMRI (functional magnetic resonance imaging), and MEG (magnetoencephalogrpahy) predominantly resulted in studies of reading behaviour in normal adult readers compared with dyslexic and/ or poor adult readers. Comparably few studies included young children learning to read and even fewer studies included clinical intervention trial studies (for an overview, see Pugh et al., 2006). The neural architecture of reading is supposed to be consistent with basic reading system models like the one outlined by Coltheart (see below). It suggests the basic reading architecture to be a foundation for the Dual Route model as redrawn from Baron's reading system architecture (Coltheart, 2005, p. 8).

But what does the typical adult readers' cortical system integrating phonological, orthographical and semantic aspects of words look like? Pugh et al. begin with a rudimentary representation of brain regions involved in 'reading related' tasks (i.e. phonological, orthographic and semantic (lexical) tasks). According to their model, the reading areas of the brain can roughly be found within the language dominant, left hemisphere comprising a posterior, occipitotemporal (ventral stream) and a temporoparietal (dorsal stream) as well as the anterior inferior frontal gyrus system. In Pugh et als.' overview, the ventral system in the inferior occipitotemporal fusiform area extends to anterior middle and inferior temporal gyri (Pugh et al., 2006). Stanislas Dehaene, who, among others, has devoted extensive imaging research to this particular area (cf. Cohen et al., 2000; Dehaene, 2013; Dehaene et al., 2010), when outlining its role in reading acquisition, has also regarded the left fusiform region to serve as a (pre-) semantic, i.e. sublexical, visual word form area (VWFA). The visual word form area, which Dehaene calls the brain's 'letter box', reacts to words in skilled readers, whereas poor readers or illiterates and patients with lesions in the said area show activations to other visual stimuli, such as faces, objects or checkerboard drawings (for a lesion study, see Gaillard et al., 2006).

Pseudowords trigger activation in skilled readers predominantly in the left supramarginal gyrus, which advances the view that the temporoparietal region is crucial for phonological analyses (Pugh et al., 2006) with substantial evidence that phonemic representation projects into the planum temporale (Dehaene, 2013; Dehaene et al., 2010). In the anterior region the inferior frontal gyrus (IFG) is split into a) the posterior part and b) an anterior part (cf. Pugh et al., 2006). Here, activation occurs when a) phonological recoding takes places, involving phonological working memory (which has been shown to often be deficient in dyslexics) and syntactic processing (ibid.). The anterior part of the IFG b) reacts to semantic (lexical) processing (ibid.).

So what is the cortical pathway of the processing of written words (in unimpaired adult readers) and is it the same for all writing systems? Dehaene claims that across all cultures and "[…] all over the world, the same brain regions activate to decode a written word." (Dehaene, 2009, p. 7). Researchers are widely in agreement, which neural networks and neural pathways activate in the processing of written language. Sprenger-Charolles (2006, p. 15ff) provides a solid overview of such areas and their pathways. Following her, there is, for instance, agreement that the primary cortical area for vision is Brodmann's area 17, from which information is passed on to the secondary visual area, Brodmann's area 18. From there visual information, just like language representation follows a dorsal and a ventral pathway. Similar to speech, the reading process seems to be left-hemisphere dominant (cf. Pugh et al., 2006, p. 27f; Sprenger-Charolles et al., 2006, p. 15). The fusiform gyrus (BA 37 which borders the VWFA) is part of the ventral pathway, in charge of written word form access. The phonological route runs along the dorsal pathway via BA 39, the gyrus angularis and BA 40, the supramarginal gyrus, and project to BA 22 (Wernicke's area), where written word forms are translated into meanings. The anterior component in the IFG consists of Broca' area, BA 44 and 45 and 47, connecting to both pathways to enable phonological and syntactical processes.

3.3.3 How Learning to Read Changes Cortical Networks

Dehaene has authored two, critically acclaimed studies on the subject how learning to read influences brain plasticity (Dehaene, 2013; Dehaene et al., 2010). In his 2010 article, Dehaene comments on how the acquisition of reading changes the brain's anatomy as well as activation patterns in the brain. A very important recent finding of Dehaene is also, how literacy alters auditory speech processing through the development of phonemic awareness (Dehaene et al., 2010, p. 1359). The important aspect of this development is that through literacy, the usually fully developed speech processing mechanisms in preschool children change, in order to set the stage for phoneme-grapheme conversions. Children who learn to read must therefore attribute another language processing modality to their auditory speech processing abilities and this visual modality might contrast with the existing visual speech processing (i.e. speechreading) ability that develops alongside auditory speech perception (see also: Blomert, 2011 on reading acquisition and new functions in the neural networks for speech processing). Blomert (ibid.) also refers to new audiovisual grapheme-phoneme associations that the beginning reader has to acquire, a feature that results in extensive adaptation in the network linking speech processing and visual object recognition. The rivalry – Dehaene calls it 'cooperative' or 'competitive effects' – between these two visual modalities might explain why children are less susceptible to

McGurk items at this very stage, when they need to accommodate two visual speech processing strategies (Dehaene et al., 2010 ibid.).

Researchers such as Pugh et al. corroborate the theory that during reading acquisition, the dorsal and anterior systems are predominant (Pugh et al., 2006, p. 27). This coincides with the idea that, in all reading models discussed earlier, sight word reading appears only in later stages. Shaywitz et al. found that typically developing children under the age of 10.5 years draw more on the dorsal and anterior systems, whereas children older than 10.5 years already show increased ventral system engagement, as found in skilled readers (cited in Pugh et al., 2006, pp. 27-28). Dehaene locates the brain site which starts responding to orthographic stimuli in beginning readers within the left occipito temporal cortex, the so- called visual word form area (WVFA) where also faces, houses and other objects are usually processed (Dehaene et al., 2010 ibid.). FMRI studies, as cited in Sprenger Charolles et al., provide further evidence that children of 10.7 years mean age show less activation in the left fusiform gyrus region than adults as the children are not yet quick enough to process orthographic and phonological information (Sprenger-Charolles et al., 2006, p. 16). Other studies reveal that grapho-phonological skills in 9-12 year olds do not yield as strong activations of angular gyrus regions as in adults, suggesting that automaticity has not manifested itself yet. For Dehaene it is clear that literacy implies fundamental changes within this VWFA. Before literacy and in illiterates the VWFA is only activated by recognition and identification of faces, houses, objects (such as tools) and letter strings together with false font strings and moving checkerboards (Dehaene et al., 2010, p. 1360). The fact that Dehaene also found in literates that the left temporal and frontal language areas are activated by written materials perfectly fits the linguistic- based dyslexia cause hypothesis (Dehaene et al., 2010, p. 1360).

In Dehaene et als' study, spoken language processing was also examined in literate, illiterate and late literates, with the notable result that literacy also seems to have an impact on the processing of spoken sentences. In good readers, left STS and bilateral MTG showed reduced activations which Dehaene et al. attribute to a 'facilitation of speech comprehension in literate participants' – also mirrored by stronger left lateralized responses in the planum temporale, a classical site for phonological coding and simultaneous processing of congruent phonemes and graphemes (Dehaene et al., 2010, p. 1362). This planum temporale activation, known to be absent in dyslexics, corroborates Dehaene et als' idea that the PT is a 'prime candidate for the enhanced phonemic processing that accompanies reading acquisition' (Dehaene et al., 2010, p. 1362).

In close proximity to the visual word form area, Dehaene et al. identified another region activated in auditory lexical decision tasks, including words and pseudowords: in literate, but not illiterate subjects the lateral inferior temporal cortex, just beneath the VWFA, reacted to spoken words and pseudowords as well as to written string items (Dehaene et al., 2010, p. 1363). Activiation in these areas were interpreted as a reaction to 'an orthographic code' (ibid.). Together with the planum temporale activation and the VWFA as well as the adjacent IFG area activation, Dehaene et al. provide evidence that the mastery of orthography influences speech processing by activating the above mentioned orthographic code. They conclude thusly that literacy changes phonological representations. Dehaene et als' conclusions are therefore threefold: Literacy 'boosts' visual cortex organization through enhanced responses in the VFWA and early visual cortex (occipital) responses (Dehaene et al., 2010, p. 1364). Literacy is responsible for the activation of the left lateral language areas through written sentences. Hence, literacy 'refines' speech processing by enhancing a phonological region in the planum temporale and by making an 'orthographic code' accessible in a top down process (ibid.).

3.3.4 The Dyslexic Brain

How then, are the processing circuits in dyslexic readers different from normal readers? And where exactly do we find such differences? Studies that focus on voxel based morphometry reveal that significant differences are already found in grey matter differences in the left hemisphere as well as in 'usual suspect' sites (Silani et al., 2005). Most significantly, Silani et al. identified:

> "[…] a cortical structural disorganization of the cortex with both reduction and increases of 'grey matter' [...]" as well as that "[...] the left middle temporal region was the site of maximal difference in brain activation in normal and dyslexic subjects, and here an area of reduced grey matter density was observed which can be interpreted as a regional atrophy. However, this was surrounded by a more posterior region of relative augmentation of grey matter that spanned downwards into inferior temporal cortex." (Silani et al., 2005, p. 2458).

Functional imaging studies clearly reveal that there is a profound difference between reading impaired subjects and typical readers concerning the activation patterns in the three above mentioned sites, dorsal, ventral and anterior. Shaywitz et al., for instance, have confirmed disruptions in the left hemisphere sites for phonological processing, the posterior dorsal and ventral regions (B. A. Shaywitz, Lyon, & Shaywitz, 2006). In their meta-analysis of functional abnormalities in the dyslexic brain studies, Richlan et al. identified reduced activations in the posterior reading areas, relating to reduced network activity in phonological and orthographic linking tasks

(Richlan, Kronbichler, & Wimmer, 2009). The idea of a functional connectivity problem across the major reading areas has received considerable support recently (cf. Pugh et al., 2006, p. 29f). Dehaene found significant differences within the VWFA. Whereas good readers show strong visual word form area activation for words, activation for words cannot be triggered in dyslexics.

More recent fMRI studies also advocate a visual-phonological disconnection, which provides further evidence for the functional connectivity problem in dyslexia (cf. Schurz et al., 2013). Furthermore, such imaging studies reveal that compensational activation processes may be found in dyslexics. Phonological processing tasks, for instance, yield greater IFG and prefrontal dorso-lateral site activation and contra-lateral, i.e. right hemisphere activation in mirrored regions suggests a lack of the asymmetric, typical left hemisphere dominance. Additionally, reduced left hemisphere activation in the posterior regions correlating with right hemisphere posterior regions activation is interpreted as a compensatory letter by letter processing strategy (for an overview of these studies, cf. Pugh et al., 2006, p. 30f)

The phonological, orthographic and semantic (lexical) time line has also been of special interest to researchers in recent years (for a brief overview cf. Sprenger-Charolles et al., 2006, p. 15). The most striking finding might be that, in the fusiform gyrus area, activity is obtained after 200 ms, suggesting that orthographic information processing is the first activity in reading. With 300ms post onset activation in the superior temporal gyrus and 400 ms onset in temporo-parietal regions, the timeline corresponds with the processing models outlined above (cf. Sprenger-Charolles et al., 2006, p. 15 for relevant ERP studies). In audiovisual language processing, research on the Mismatch Negativity phenomenon has been investigated with TD and dyslexic children, among others by Sams et al. (Sams, Kaukoranta, et al., 1991; Tuomainen et al., 2005). Low susceptibility to MMN in dyslexic subjects and poor orthographic skills would seem to occur within the same time frame, of around 200 ms - an interesting parallel that will be discussed further in the conclusive remarks of the empirical results section. With the sublexical McGurk stimuli used in the audiovisual task, a third aspect of temporal processing within the time window of 200ms has to be included in that discussion.

3.4 Phonological Deficits and Reading – Language processing deficits and Reading

The idea that developmental dyslexia seems to be best characterised as a specific, i.e. 'exclusive' phonological deficit, arbitrarily accompanied by sensorimotor syndromes has been quite appealing to some researchers (cf. Snowling; Ramus 2003 but see also Ramus 2008 and 2013 for a change of the phonological approach). The paradox that a large proportion of dyslexics also present with sensory and/or motor deficits, which nevertheless seem to be less prominent than the phonological deficit (cf. Tallal; Breznitz; Ramus 2003 for an overview), has puzzled the dyslexia research community over the last decades. A rather antagonistic approach – specific phonological deficit or general sensorimotor dysfunction – when approaching this multifaceted condition has given way to a more interconnected theoretical approximation. As posited in this very research context here, this search of 'the direct cause' has to be abandoned in favour of a more inclusive theory covering various aspects and interdependencies of dyslexia.

Dylsexia researchers such as Margaret Snowling and Joy Stackhouse have closely examined the relation of speech, language and dyslexia (M. Snowling & Stackhouse, 2006). In this context Snowling considers children's oral language abilities the 'foundation for later developing literacy skills and finds that phonological impairments such as poor phonological awareness, rapid naming skills and verbal short and long term memory are the predominant trigger for most reading difficulties, including the reading impairment that is dyslexia (M. J. Snowling, 2006). For a multi-center study on reading difficulties across various orthographies consult the work of Landerl and colleagues, who did not find strong evidence for memory aspects (Landerl et al., 2013).

I have already outlined two of the three preeminent theories when it comes to developmental dyslexia: the magnocellular (auditory and visual) theory and the cerebellar theory. It is now vital to frame the third approach: the phonological theory, which predicates, that dyslexics have a specific impairment in the representation, storage and/or retrieval of speech sounds:

> "It (the phonological theory) explains dyslexics' reading impairment by appealing to the fact that learning to read an alphabetic system requires learning the grapheme-phoneme correspondence, i.e. the correspondence between letters and constituent sounds of speech. If these sounds are poorly represented, stored or retrieved, the learning of grapheme-phoneme correspondences, the foundation of reading for alphabetic systems will be affected accordingly" (Ramus et al., 2003, p. 842).

Most of the eminent theorists agree on the central and causal role of phonology. Hence, when it comes to dyslexia, they promulgate a straightforward link between a cognitive deficit and the behavioural problem(s) in question. The most vigorous version of the phonological theory assumes that the cognitive deficit is specific to phonology. A phonological awareness deficit should be noticeable in alphabetic scripts when the reading of non-words, also called pseudoword or nonsense words, is significantly impaired, which is why the empirical research presented in chapter 4 included a pseudoword reading task and used pseudowords in all perception tasks (visual only, acoustic only and audiovisual). Rapid automatized naming tasks show whether subjects' verbal memory access is within the range of typical readers. A study conducted by Snowling et al. (2003) showed that 4 year olds with slow language development had difficulties with phonological awareness at the age of 6 years and were also at high risk of becoming poor readers at the age of 8 years (Snowling, Gallagher and Frith, 2003 discussed in: M. J. Snowling, 2006, pp. 10-11). Support for the phonological awareness deficit is also provided by Liliane Sprenger-Charolles et al. who discuss studies on phonological priming in adults and in typically developing children as opposed to dyslexic children (Sprenger-Charolles et al., 2006, pp. 11-14). Findings that dyslexic readers gain little to nothing from phonological primes in naming and lexical decision tasks thereby also support the phonological deficit hypothesis.

A most challenging approach would be to defy the specificity of the phonological deficit and claim it to be secondary to a more basic auditory or visual deficit, as theorists propagating the rapid auditory processing (cf. P. Tallal, 1980) and propagators of the visual theory (cf. Vidyasagar & Pammer, 2010) have done. However, while there is still some debate whether categorical auditory perception deficits, as suggested by Tallal and others, result in difficulties to acquire reading, there is abounding evidence that phonological language impairments lead to reading difficulties (on categorical deficits cf. Benasich & Tallal, 1996, 2002; Paula Tallal, 1980), see Robertson et al. for speech perception difficulties that distinguish language and reading impairments in children (Robertson, Joanisse, Desroches, & Ng, 2009). Hence, some exponents of auditory and visual theories, now concur that visual and auditory disorders are part of a more general magnocellular dysfunction (cf. Ramus et al., 2003, p. 843).

Considering the other major theories, i.e. the cerebellar and magnocellular theories, which try to explain the phenomenon, or better: the phenomena of dyslexia, it is, however, problematic to reduce the complexities of dyslexia to a mere phonological deficit. It has been advocated that the disorder is much more extended and the phonological deficit is just one aspect or consequence of it. And this is to zoom in on one of the major weaknesses: the phonological theory in its current form is not able to explicate the occurrence of sensory and motor disorders in dyslexics. None of the above mentioned theories is however without weaknesses: the cerebellar theory is not able to account for sensory disorders either, however, Fawcett and Nicolson (2001), one of its main proponents, have managed to put forward the idea of distinct cerebellar and magnocellular dyslexia subtypes. The magnocellular theory, with its unique appeal in being able to cover all manifestations of dyslexia, had to face criticism due to several failures to replicate findings of auditory disorders (Heath, Hogben, & Clark, 1999; Hill, Bailey, Griffiths, & Snowling, 1999; McArthur & Hogben, 2001) and obtained results which are inconsistent with the idea that the auditory deficit lies in 'rapid' auditory processing (hence in magnocellular processing), e.g.: with some tasks 'rapid' auditory processing is found to be in perfect working order whilst with others 'slow' auditory processing is flawed (McAnally & Stein, 1996; Reed, 1989; Schulte-Korne, Deimel, Bartling, & Remschmidt, 1998a, 1998b; Share, Jorm, Maclean, & Matthews, 2002). Ramus et al. (2003, p. 844) sum up the weaknesses perfectly:

> "[…] the phonological theory suffers from its inability to explain the sensory and motor disorders that occur in a significant proportion of dyslexics, while the magnocellular theory suffers mainly from its inability to explain the absence of sensory and motor disorders in a significant proportion of dyslexics. The cerebellar theory presents both types of problems".

In addition, it seems to be problematic that quite many studies on dyslexia suffer from a heavy sampling bias when it comes to the socio-economic status of the dyslexic sample (very often university students are recruited), sample size (usually quite small) and a small array of tasks which usually focuses on only one modality (cf. Ramus, 2003).

The double deficit hypothesis (DDH), originally conceived by Bowers and Wolf in 1993, is still of keen interest to a select number of the research community (see Wolf & O'Brien, 2006 for a recapitulation). The DDH suggests three possible aspects occurring in poor readers, namely poor readers with phonological deficits but without naming speed deficits, poor readers with naming speed deficits but no phonological deficits and poor readers suffering from both deficits (Wolf & O'Brien, 2006, pp. 10,11).

Following the phonological deficit hypothesis (Bradley & Bryant, 1983; M. J. Snowling et al., 1994; Vellutino, 1981), the development of the automatisation deficit hypothesis was an influential explanatory construct for dyslexia, outlined by Nicolson and Fawcett (1990) in the nineties. It proposed that dyslexic children have difficulties making skills automatic – this refers to them not having to think consciously anymore of how to do it. Drawing on the idea that the critical aspect of learning a skill like reading is to make it automatic, the hypothesis that dyslexic children would have difficulty in automatizing any skill (might it be cognitive or motor) became prominent. Early findings strongly suggested that dyslexic children had difficulties in automatizing skills and had therefore to concentrate and focus harder to achieve regular levels of performance (cf. A. J. Fawcett & Nicolson, 1992). Particularly intriguing in this context is Blomert's review of ERP studies focussing on Mismatch Negativity paradigms with letter speech sounds, in which he also provides a detailed discussion concerning lack of automation in the letter-speech sound domain (Blomert, 2011). When it comes to difficulties in automatizing skills, Fawcett and Nicolson use the analogy of driving in a foreign country: "one can do it, but it requires continual effort and is stressful and tiring over long periods. On our account, life for a dyslexic child is like always living in a foreign country" (R. I. Nicolson & A. Fawcett, 2008, p.

196). It soon became obvious that dyslexic children have to concentrate extra hard on even the simplest skills:

> "(…) the dyslexic children appeared to have greater difficulty in blending existing skills into a new skill, and their performance after extensive practice (such that the skill was no longer improving noticeably) was slower and more error-prone. In other words, they were simply less skilled; their 'quality' of automatized performance was lower. It seems reasonable, therefore, to argue that this group of dyslexic children have difficulties both with the initial proceduralisation of skill, and with the 'quality' of skill post-training" (R. I. Nicolson & A. Fawcett, 2008, p. 198) .

Nicolson and Fawcett further apply the differences between types of memory onto learning and begin to distinguish between declarative learning – mostly the learn ing of facts, which can be split up into episodic learning (learning via experiencing situations) and semantic learning (learning more general information pertaining to meanings) – and procedural learning – divided into three sub-types: statistical learning, skill learning and conditioning (ibid., p. 200). They claim that it is perfectly possible for dyslexic children to display difficulties in all five types of learning. But given that dyslexia is associated with normal levels of intelligence, which is usually measured via declarative tasks, it might be more reasonable to affiliate dyslexia with one or more of the three sub-types of procedural learning.

By drawing on Doya (Doya, 1999), Nicolson and Fawcett (2008, p. 201) introduce the concept of 'unsupervised' learning, driven mainly by environmental input than by a goal-directed process. The cerebral cortex is said to be endowed with this special ability. It is a sort of pattern recognition (an alternative term for statistical learning) of which the 'connectionist learning' capability is a significant example, since a fair share of our early sensory learning, such as the effortless ability to learn to recognise auditory or visual (language) patterns, is predicated on it. Moreover, it is crucial to prioritise more important processes. This happens via the dopaminergic system of the basal ganglia and is referred to as 'reinforcement learning' (ibid. 201). The third competence is defined as 'supervised learning', in which a 'training signal' somehow reflects the difference between the desired output and the planned output, with the cerebellum being the vital and perhaps the only brain structure to do this: "(…) with the error signal being provided via the climbing fibers from the inferior olive, and the patterned sensory information coming in via the thalamus" (R. I. Nicolson & A. J. Fawcett, 2008, p. 202). According to Nicolson and Fawcett (ibid.) the aforementioned three types of procedural learning fall into these three categories:

1. statistical learning being unsupervised (achieved in the cerebral cortex)
2. conditioning requires reward and can hence be considered a form of reinforcement learning (some involvement of the basal ganglia is necessary) and
3. skill learning, involving comparison of the actual with the desired outcome (supervised learning is needed).

Thereafter Nicolson and Fawcett draw on Ullman (2004), who applies the procedural/declarative categorisation to language skills and propounds that the 'Declarative Memory System' influences the 'mental lexicon':

> "The 'Procedural Memory System' underpins the 'mental grammar' – the learning of new rule-based procedures that govern the regularities of language – together with the learning of new skills and the control of established sensori-motor and cognitive habits. It comprises the basal ganglia; frontal cortex, in particular Broca's area and premotor regions; parietal cortex; superior temporal cortex

> and cerebellum […] Ullman proposes that the declarative memory (DM) and procedural memory (PM) systems form a dynamically interacting network which yields both cooperative and competitive learning and processing, leading to a *see-saw effect,* such that a dysfunction of one system leads to enhanced learning in the other, or that learning in one system depresses functionality of the other" (R. I. Nicolson & A. J. Fawcett, 2008, p. 205).

In terms of learning processes then, not only language-related processes are affected, but many of the procedural learning processes for motor and language skills function at this low level. From Nicolson and Fawcett's typography for learning difficulties it can be concluded that dyslexia is indeed distinguishable from other language impairments such as SLI. Could the motor-corticocerebellar procedural learning skills impairment be reflected in poor motor speech skills resulting in poor audiovisual speech perception skills?

Joy Stackhouse confirms that poor readers present significantly often with speech and language difficulties (Stackhouse, 2006). She hints at a 'subtle' speech and language problem that may persist in older dyslexic children (2006, p. 16). This 'subtle' difficulty is not further elaborated, but matches the topic of this book, namely, that speech and language representations involve more than acoustic speech perception, articulatory production and visual (i.e. read) processing. A hitherto not thoroughly studied discrete audiovisual language processing deficit, originating from poor acoustic and visual segment representation would result in a serious language impairment that persists throughout life. Breznitz' Asynchrony Theory suggests that "[…] dyslexia is the outcome of the failure to *synchronize* the various grain entities activated during the reading process." (Breznitz, 2008, p. 11). She talks about a lag or 'gap' in the processing speed ('SOP: speed of processing') of information that travels between the different brain areas involved in decoding words and being responsible for the prevention of accurate synchronization. The severity of the word decoding failure increases, the wider this 'gap' is. Also central to Breznitz' theory is that there are differences in processing speeds between the brain areas, and a lack of coordination among the reading network's regions. These two deficits together may account for the fact that integration is prevented (cf. Breznitz, 2008, p. 12). In her studies, Breznitz found that dyslexic readers were slower than controls in visual and auditory speed processing tasks and identified a certain ansynchrony between the posterior and anterior brain areas (Breznitz, 2008). These findings are in accord with Heinz Wimmer's suggestion, that dyslexics present with a visual- phonological disconnection in the visual word form area as well as with a reduced connectivity between the VWFA and the inferior frontal regions explaining both, the phonological decoding problems and the 'speed impairment' (Wimmer, 2013). It is also the most potent explanation why the dyslexics in the empirical audiovisual tasks in the experiment of this book did not present with strong audiovisual integration and appeared to be poor lipreaders .

3.5 Reading and Speechreading

As this book aims at establishing links between reading and speechreading, landmark studies that have discussed dyslexia and auditory processing deficits, dyslexia and visual speech processing deficits and dyslexia in regard to audiovisual deficits need to be briefly outlined here. Insight into very early acoustic discrimination deficits in speech sound categorisation tasks is provided by van Leeuwen et al., who tested two month old infants at risk for dyslexia in a mismatch negativity paradigm (MMN) with /bak/ vs. /dak/ items (van Leeuwen et al., 2008). The study was designed to substantiate the idea that dyslexics present with auditory, visual and (neuro-) motor deficits and van Leeuwen et al. investigated whether, in infants at risk for dyslexia, a persistent difficulty in distinguishing between stop consonants in the syllable onset can be ascertained. Studies using EEG to examine MMN items allow to measure neural responses to acoustically deviant speech and non-speech signals (cf. Kujala & Naatanen, 2001; Nagarajan et al., 1999; Paul, Bott, Heim, Eulitz, & Elbert, 2006). In van Leeuwen et al.'s 2007 study, designed to elicit MMNs with manipulated consonant-vowel-consonant (CVC) items, they found that two month old control infants already showed robust MMNs, whereas the at-risk-for-dyslexia subject group did not. In order to endorse these findings, especially the performance of the at-risk subject group, where the estimate that dyslexia would manifest itself was below 50%, van Leeuwen et al. replicated their experiment with a larger sample in 2008. The subject group comprised 82 infants (two months old) and among other inclusion criteria it was required that the infant had one dyslexic parent and at least one dyslexic first degree family member. The controls were 57 infants (two months old). As in the previous study, the at-risk infants displayed hardly any MMNs, which the authors interpreted as powerful support of the assumption that an auditory temporal processing deficit underlies a phonological deficit responsible for reading impairments.

It is now high time we took a closer look at research concerning speechreading abilities, in language impaired subjects, to be specific, in reading impaired subjects. Since the focus of most dyslexia researchers has predominantly been on the immediate reading and writing problems that the children present with, very little research has been dedicated to the examination of speechreading in dyslexic subjects. Hence it is vital to give a brief overview of these selected few reading studies centring on speech reading and audiovisual integration. In the previous section it was suggested that reading acquisition alters the representation of visual aspects, specifically word forms, faces and objects (Dehaene, 2013; Dehaene et al., 2010). Now, studies that apply imaging technologies such as fMRI to reading research will be used to draw parallels between reading and speech processing.

A study aimed at bridging the gap between speechreading and reading research was conducted by Hayes et al. (Hayes et al., 2003; Nagarajan et al., 1999): They presented unimodal (visual and auditory) and bimodal (audio-visually congruent and incongruent, i.e. McGurk) stimuli to normal learning children (n=10) and children with learning disabilities (n=13). By adding noise (12dB SNR) to the audiovisual tasks, they intended to assess how noise influences learning impaired subjects. The most relevant finding of this study was that normal learning children differed significantly from learning impaired children, who demonstrated poor unimodal visual accuracy and even poorer audiovisual integration skills: When McGurk items were masked with 12 dB SNR noise, learning impaired subjects hardly showed the fused McGurk perception of items but responded to the visual only aspect. Controls showed more McGurk responses in the masked task. However, in the unmasked condition, Hayes et al. found no significant difference in the performance of both subject groups (Hayes et al., 2003). In order to explain the weak McGurk responses in the learning impaired group, Hayes et al. also performed brainstem responses to speech sounds that have long been regarded as deficient, thereby explaining why the acoustic aspect in McGurk items might not be considered by learning impaired subjects (cf. de Gelder & Vroomen, 1998; Kraus et al., 1996 on auditory processing deficits). However, this task did not yield significant differences between the two subject groups in Hayes et als' (2003) study. The differences that they found for audiovisual integration in 'challenging listening conditions' may emanate from auditory and visual deficits, or from a synchronisation deficit of the two modalities. Both options are currently investigated in dyslexia research (Hayes et al., 2003, p. 49). Amongst other approaches that aim to explain this phenomenon, one finds the idea that dyslexics activate different cortical areas in audiovisual tasks, such as increased activation in Broca's area but weak activation in Wernicke's area, the gyrus angularis and the striate cortex (cf. S. E. Shaywitz et al., 1998). Speech motor areas and Broca's area might also be activated via the mirror neuron system (Kohler et al., 2002). The above mentioned functional disconnection approach (Richlan et al., 2009) and Breznitz' cortical asynchrony hypothesis (Breznitz, 2008) would also explain the characteristics of an audiovisual integration deficit. Taking into consideration that several cortical and subcortical mechanisms may be involved in impaired audiovisual perception, Hayes et als.' conclusion is that learning impaired children's performance in audiovisual tasks does reflect impaired multisensory integration but needs further research (Hayes et al., 2003, p. 49). In their outlook they advance the idea that perceptual and neurophysiological studies examining audiovisual speech perception in normal and impaired readers would foster alternative strategies for reading instruction based on the individuals perceptual strengths. One behavioural study that relied on Hayes et als.' paradigm and their findings investigated the McGurk effect in dyslexic children in comparison to age matched and reading age matched children (Bastien-Toniazzo et al., 2009). The authors found that dyslexic children (n=12) were less susceptible to the McGurk percept when presented with audiovisually incongruent /apa/ vs. /aka/ items compared to reading age matched (n=12) and age matched controls (n=12) (Bastien-Toniazzo et al., 2009, pp. 11-12). They also found that dyslexics did not display any problems in processing unimodal auditory stimuli, but that they were poorer lipreaders in unimodal visual tasks. These findings, albeit from a small sample without control where the subjects actually looked in the visual and audiovisual tasks, support the findings of this book (see chapter 5.3)

A test for (visual) speechreading in adults (TAS, Test of Adult Speechreading) was introduced by Mohammed et al. (Mohammed, Campbell, Macsweeney, Barry, & Coleman, 2006). In the accompanying study, they examined adult speechreading in profoundly prelingual deaf subjects, dyslexic subjects and unimpaired controls. Their approach to the relationship between reading and speechreading is based on phonological awareness and speechreading: impairments in central processes of speech and language underlie dyslexia. The subjects of this study included 21 profoundly deaf subjects (whose speechreading performance is not discussed here), 21 dyslexic adults and 21 controls. The dyslexic subjects were considered to have compensated for reading abilities. Testing of visual speech abilities was carried out, among other tasks, with minimal pairs comparable to the items used in the subsequent empirical part. Similar to the findings gained from the lipreading tasks of the experiment underlying this book, Mohammed et al. found that dyslexics were poorer speechreaders than controls, a result that the authors interpreted as a falsification of the prediction that dyslexics rely more on visible speech. They consequently suggested that the correlation of phonological awareness with reading and speechreading "may be mediated by speech-based (phonological) representations" (Mohammed et al., 2006, p. 629).

In an fMRI study with 10 highly educated dyslexics and 10 matched adults, Pekkola et al. investigated how subjects' brain responses to matching and conflicting audiovisual speech differed between dyslexics and controls, expecting to find such differences within Broca's area and premotor cortex (as the motor speech areas) and sensory specific auditory and visual cortices (Pekkola et al., 2006). The conflicting audiovisual stimuli were not McGurk type items but vowel stimuli that either matched, i.e. visual /a/ and acoustic /a/ or incongruent e.g. visual/a/ and acoustic /y/ (cf. Pekkola et al., 2006, p. 799). Whereas the two subject groups performed equally well when identifying incongruent items, and also showed increased activity in the incongruent condition, the brain responses of dyslexics and controls differed, both in matching and conflicting items. In the dyslexic group stronger activations were found for the incongruent items bilaterally in ventral visual cortex, the supplementary motor areas, the anterior gyrus cinguli and the cerebellar vermis (Pekkola et al., 2006, p. 800).

Hence, the crucial findings of this study were that dyslexics showed significant activation differences and a co-variance between activations and phonological processing skills assessed in phonological awareness tasks. The stronger activations of speech motor areas was interpreted by Pekkola et al. to reflect dyslexics' dependability on motor-articulatory and visual strategies in audiovisual speech processing. Also, in dyslexic subjects, the bilateral lingual and fusiform gyri were activated more strongly. These findings concur with my hypothesis that motor speech processes are essential for visual speech processing and audiovisual speech processing. Furthermore, Pekkola et al. also refer to speech motor theories in that "[…] the perceived articulatory objects are intended articulatory movements (i.e. neuromotor commands to the lips, tongue, and vocal cords)" (Pekkola et al., 2006, p. 803). Thus, if motor representations of articulatory gestures are strongly activated through motor speech area activation, Pekkola et al. conclude that

> "[…] dyslexic readers' more widespread and stronger activation in the conflicting < matching contrast […] encompassing the motor speech areas, their right-hemisphere homologues and supplementary motor areas, may reflect their heightened use of sub-vocal motor-articulatory strategies during phonetic processing of audiovisual speech" (ibid.).

These findings may provide some support for the visual and the auditory deficit hypothesis in dyslexia. The heightened dependency on motor-articulatory and visual speech processing is tentatively referred to by Pekkola et al. as "[…] compensatory mechanisms to overcome linguistic perceptual difficulties." (Pekkola et al., 2006, p. 804). An fMRI study that is currently conducted by the author tests adult poor readers and controls and their susceptibility to the McGurk effect. This study was designed to identify whether poor readers, who do not display a McGurk effect, show the same activation patterns as unimpaired controls in the motor speech areas and the anterior fusiform gyrus extending to the VWFA. Preliminary results indicate a tendency that dyslexics show stronger, perhaps compensatory response activations, uni- and bilaterally in these regions (Kaltenbacher et al., in prep.).

Concluding this chapter now, it could be drawn from pertinent research that the reading brain and the speech reading brain activates areas that serve both purposes. It was also argued that phonological processes underlying reading, visual speech and speechreading are reflected in, and consequently also interdependent of, motor speech functions. In the empirical chapter that is to follow this one, the experimental design that allowed testing of dyslexic subjects and controls for their ability to lipread, to process speech and audiovisually congruent and incongruent (McGurk) stimuli will suggest a novel approach to assessing speech processing skills in reading research.

4. Experimental Research in Audiovisual Speech Perception

Syllables govern the world.
George Bernard Shaw (1856-1950)

In the following chapter the experimental design used to test the aforementioned hypotheses will be outlined in more detail. For the purpose of testing visual and audiovisual speech, a new, more advanced set of stimuli was conceived by the author in order to meet the requirements for visual and audiovisual speech perception as well as to generate more reliable data from a linguistic point of view. In furtherance of testing how beneficial audiovisual speech truly is, 'naturalistic', i.e. not over articulated (cf. Massaro & Cohen, 1983), non- synthesized speech stimuli had to be devised (following the findings of Blomert & Mitterer, 2004). Besides, items had to be created, from which subjects would not be able to infer the correct answers, let us say, for example, from context or word frequency (Blomert et al., 2004; Massaro & Cohen, 1994), in order to properly test visual speech reading abilities. The theoretical bedrock for the creation of such stimuli has already been laid in the previous chapters and in the following subsection 4.2 on stimulus creation, the focus will be on the practical realisation of stimuli designed to examine subjects' speechreading, lipreading, auditory processing and pseudoword reading abilities. Thereafter the data acquisition method including the eye-tracking apparatus and the methodology of the experimental paradigm's implementation will be outlined. Finally, subjects and controls and their assessment of dyslexic vs. non- reading impaired status will be specified.

4.1 Hypotheses: The Audiovisual Deficit Hypothesis

The empirical part of this book was designed to test against two null hypotheses and for an alternative hypothesis. As the null hypotheses were expected not to hold, a specific explanation in terms of a multimodal integration deficit was formulated. This explanation is framed in terms of the three hypotheses that are to replace the null-hypotheses. For the reader's convenience, these hypotheses are listed here once again:

i. Dyslexic subjects and non-dyslexic controls differ significantly in their susceptibility to the McGurk effect. A robust McGurk effect indicates that audiovisual integration of speech signals is intact. Weak responses to the McGurk effect in dyslexic subjects indicates at a multimodal, i.e. audiovisual integration problem.
ii. Dyslexic subjects perform significantly worse than non-dyslexic controls in a visual only speechreading task. This means that subjects have great problems identifying the speechreadable (= lipreadable) aspects of visual speech signals, such that subjects are able to repeat back what a talking face articulates.
ii. A strong correlation between the visual only task and poor pseudoword task performance for dyslexic subjects is expected.

As will be demonstrated in the following chapter (chapter 5), the behavioural data gained from this experiment supported these hypotheses, and hence an explanation for the dyslexic subjects' non-susceptibility to the McGurk effect and correlations between poor lipreading and poor pseudoword reading is then provided in terms of a multimodal, motor-sensory speech processing deficit, which I refer to as 'the audiovisual deficit'.

4.2. Stimulus Set: Conditions, Types and Stimulus Creation

The complete set of stimuli is annexed in Appendix A. The test items used for the visual, auditory and audiovisual conditions consisted of context free- lexical and sub-lexical items, i.e. syllables, real words and non-lexical pseudowords. For the lip-reading condition, stimuli needed to be created that enabled subjects to identify the uttered syllables/ nonsense words/ words based on the ability to identify the visible aspects of the articulatory gestures. These 'visemes' (see chapter 2.2. on visual speech) should be clearly discernable, albeit there is some ambiguity: a bilabial closure could be interpreted as /m/ as well as /p/ or /b/.[3] There is also some visual similarity and ambiguity in /t/, /d/, /n/, /l/ and /g/, /k/, whereby the distinction between dental and velar plosives was expected to be the most difficult one, as the place of articulation for the velar plosives is, of course, not visible in normal articulation. The vowels used in the set of stimuli should be clearly distinguishable, also allowing some variation, for instance between /i/ and /e/ and /o/ and /u/. In order to accommodate these variations and ambiguities and thereby to facilitate the coding of the answers given by the subjects, answers to the visual items were typed according to the following numerical code:

> 0= correct answer, i.e. all aspects were repeated back correctly.
>
> 1= partly correct, i.e. either not all the visual gestures were repeated back correctly or some were left out, or most aspects were identified, for example, the visually presented stimulus is /apa-ata-aka/, and the answer given is /apa-ta-ka/ or /apa-aka-aka/.
>
> 2= partly incorrect, i.e. most aspects were not identified correctly, as, for example, the stimulus /apa-ata-aka/, elicited the answer /amata/

[3] In experienced lip-readers, the visual distinction of voiced and unvoiced consonants is facilitated in male speakers by the movement of the Adam's apple. However, this aspect was not deemed relevant for the present study.

3= incorrect, i.e. subjects reported that they could not identify anything or gave an answer in which none of the aspects were correctly identified, not even the item length, as, for example, the answer /ala/ to /apa-ata-aka/.

In the auditory condition, the same syllable/ nonsense word/ word items were used. It was expected that subjects would not find it difficult to identify the auditory items. Subjects' answers to these auditory items were coded into a binary code: 0= correct; 1= incorrect. Only if the subject's answer was completely correct, i.e. the item was repeated back in the exact same way as it was played to the subject.

The audiovisual items included congruent and incongruent (= McGurk) items. For the audiovisually congruent items the same binary answer code was used as already applied in the auditory condition. The audiovisually congruent items were considered to be the easiest items, whereas the McGurk items were expected to reveal the difficulties formulated in the hypotheses. For the incongruent, McGurk triggering, items the following numerical code was used:

0= correct, i.e. the McGurk effect was elicited.
1= the acoustic part of the McGurk item is perceived, i.e. the effect could not be elicited.
2= the visual aspect of the McGurk item is perceived, i.e. the effect could not be triggered. 3 = the answer is not correct, neither visual nor acoustic aspects are repeated back correctly by the subject.

For the pseudoword reading task, used to assess subjects' pseudoword reading skills, the items were taken from Kristina Moll and Karin Landerl's *Salzburger Lese- Rechtschreibtest II* (Moll & Landerl, 2010). Following Moll and Landerl's procedure, subjects have to read out aloud nonsense words. These words have to be read as quickly and as precisely as possible. Subjects have one minute to accomplish the task. The result obtained consists of the number of correctly read nonsense words. Subjects' performance in this task can be evaluated according to the norm tables provided in this standardized reading assessment (cf. manual of the SLRT II Moll & Landerl, 2010). The pseudoword items were adapted for the eyetracking paradigm by transferring the nonsense word lists into eyetracking stimuli.

4.2.1 Stimulus Types

Two types of syllabic items were created: vowel- consonant- vowel (VCV) and consonant-vowel (CV) syllables. The VCV items included voiced plosives and open vowels: /aba/, /ada/, and /aga/; voiceless, non- aspirated plosives: /apa/, /ata/, and /aka/; and voiceless, aspirated plosives /ap^{h}a/, /at^{h}a/ and /ak^{h}a/. The CV items included voiced plosives /ba/, /da/, /ga/, voiceless, non- aspirated plosives:/pa/, /ta/, /ka/, and voiceless, aspirated plosive /p^{h}a/, /t^{h}a/, /k^{h}a/. The voiceless, aspirated plosive versions were included because plosive aspiration is a common feature in German, especially in standard German. The VCV and CV items were designed to be presented both, in isolation (e.g. the item /pa/ presented separately), and as the building blocks for the polysyllabic items (e.g. /pa/-/ta/-/ka/ and /apa/-/ata/-/aka/ items). These items were presented in all four conditions: visual only, acoustic only, audiovisual congruent and audiovisual incongruent (McGurk items).

The real word items, consisted of minimal pairs allowing McGurk percepts. Only words with voiceless plosives were used. These words comprised the verbs: *picken, ticken, kicken* (to pick, to tick and to kick), and the nouns: *Pasten, Tasten, Kasten* (ointments, key(s), cabinet/box). The real word items served also as instruction items before the actual tests. For the actual data analysis the audiovisually incongruent (McGurk) real word items were excluded, seeing that lexical item processing is context/ memory dependent (see chapter 2.3.3 on audiovisual speech processing and lexical representations). Real word items were also presented as individual items (e.g. *kicken*) and as items with three words (e.g. *Pasten*, *Tasten*, *Kasten*) in all four conditions.

Trisyllabic nonsense words (i.e. pseudowords – PW) were designed to follow German phonology and included voiced and voiceless plosives in the word onset and in word internal, syllable onsets. The nonsense words were: /pelami/, /telami/, /kelami/, /mabali/, /madali/, /magali/, /barulo/, /darulo/, /garulo/. In all nonsense words the stress was on the second (middle) syllable, all sounds in the nonsense words were carefully articulated and schwa sounds were avoided. Like all other items, the nonsense words were also developed to be presented both, in isolation and in combination, including McGurk triggering av-combinations.

The audiovisual incongruent (i.e. McGurk) items were subcategorized in five categories, which were analysed individually as well as collectively in the data analysis. Note that the lexical items occur in this categorization as well, although they were not considered for the final discussion of the data. McGurk stimuli were thus distinguished in accordance to voicing/ devoicing of the plosives as well as to the syllable position of the plosives, i.e., CV and VCV:

avmganlaut+voice (= audiovisually incongruent, i.e. McGurk triggering in the syllable/word onset)	Code Number: 4
avmganlaut-voice (= audiovisually incongruent, i.e. McGurk triggering in the syllable /word onset)	Code Number: 5
avmgintervok+voice (= audiovisually incongruent, i.e. McGurk triggering intervocalic)	Code Number: 6
avmgintervok-voice (= audiovisually incongruent, i.e. McGurk triggering intervocalic)	Code Number: 7
avmglexanlaut-voice (= audiovisually incongruent, i.e. McGurk triggering in the word onset)	Code Number: 0

Key: avmganlaut+voice: /ba/ + /ga/; /barulo/ + /garulo/
avmganlaut-voice: /pa/+ /ka/; /pelami/ + /kelami/
avmgintervok+voice: /aba/ + /aga/; /mabali/ + /magali/
avmgintercok-vopice: /apa/ + /aka/; /ap^{h}a/ + / ak^{h}a/
avmglexanlaut-voice: /picken/ + /kicken/; /pasten/ + /kasten/

In the audiovisual condition, a total of 27 items i.e. 9x3 stimuli of the following stimuli was included:

acoustic /ap^{h}a/and visual /ak^{h}a/ → McGurk /at^{h}a/
acoustic /apa/ and visual /aka/ → McGurk /ata/
acoustic /aba/ and visual /aga/ → McGurk /ada/
acoustic /ba/ and visual /ga/ → McGurk /da/
acoustic /pa/ and visual /ka/ → McGurk /ta/
acoustic /barulo/ and visual /garulo/ → McGurk /darulo/
acoustic /mabali/ and visual /magali/ → McGurk /madali/
acoustic /pelami/ and visual /kelami/ → McGurk /telami/
acoustic /pasten/ and visual /kasten/ → McGurk /tasten/: lexical item that was not considered in the final data discussion.

With these stimulus types, a stimulus of 114 items was devised of which the first 7 served as instruction items.

4.2.2 Stimulus Creation

A German native speaker, with professional speaker and voice-over artist training (the author), recorded all audiovisual items in two recording sessions. Recordings took place in two sound proofed studios with speaker booths and optimized lighting conditions. Lighting was arranged such that the articulatory gestures produced by the speaker would not be hit by any shadows. The recordings consisted of audiovisual material that was recorded with a Canon EOS 5D Mark II camera, with 21.1 megapixel, full frame digital single lens reflex (SLR) camera technology capable of video recording in full HD with 1920 × 1080 pixel resolution. Simultaneously, the audio signal was recorded into a hard-disc recorder (Zoom H4) at a sampling rate of 44,1 kHz through a professional speaker microphone, a Sure PG 48: Frequency: 70 Hz – 15 kHz, Sensitiviy: -56 dBV/Pa/ 1,6 mV/Pa, see technical description below:

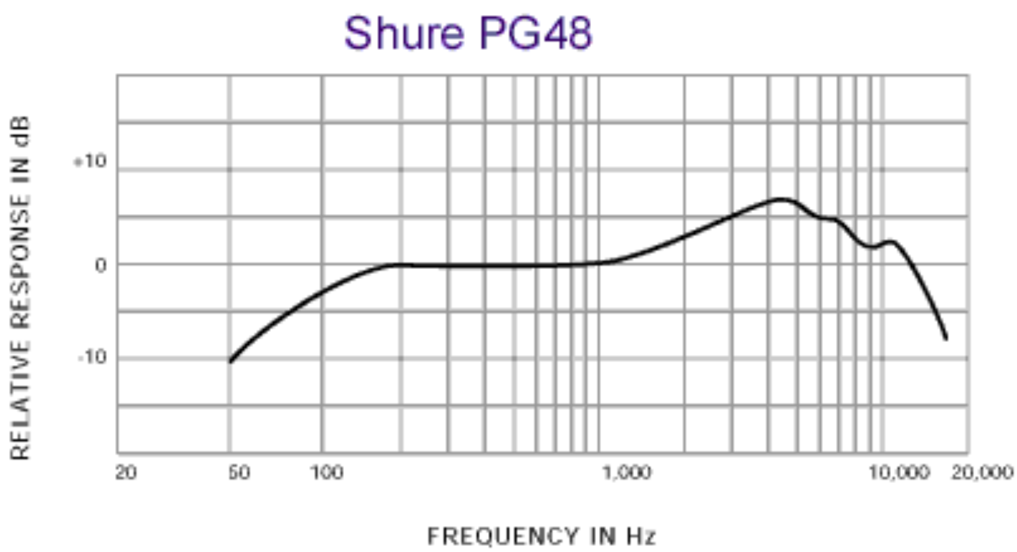


Figure 3: Technical details of the speaker microphone used for the recording of the audio channel. Taken from: www.shure.de (SHURE, 2009-2014).

In order to prevent any distraction of subjects by detectable head movement or emotion in the speaker's face, great care was taken to keep the face's expression neutral and to keep the head as still as possible during the recording. For this purpose, a mirror was positioned opposite the speaker and a rectangular crosshair was drawn around the speaker's face to allow him to monitor head position and head movement. The utterances were not over-articulated and the voice was also kept within the same frequency and amplitude for all items.

A filmset clapper was used to ascertain synchrony of the audio and video channels. The raw audio and video material was then processed in the Adobe Premiere Pro ™ (Adobe, 2010) software, which, at the time, was the only video editing software that allowed audio editing in milliseconds rather than in video frames (25 or 30 frames per second). For McGurk items, it is essential that audio editing can be done within milliseconds, in order to prevent lags between the acoustic and visual onset of the McGurk triggering plosives. The video files, which originally were in .mov format and the audio files, which were in WAV format were thereafter edited by a trained digital video editor to create the building blocks, i.e. the phonological items for the three tasks. For audiovisually incongruent items (McGurk items), the visual plosives /g/ and /k/ were dubbed with auditory plosives /b/ and /p/ in order to trigger the (German) McGurk percepts /d/ and /t/. Dubbing had to be carefully arranged, in order to warrant that the acoustic onset of the plosive was synchronous with the visual onset of the plosives. The single syllable and single word/ pseudoword items were edited to play the item immediately, i.e. without delay. The visual and audiovisual items showed the still face for 50ms before the items were uttered. The single syllable items had a duration of approx. 450ms for the CV and approx. 770ms for the VCV items. Single word items had a duration of approx. 800ms, the pseudoword items had a duration of approx. 1200ms. For the items that consisted of three stimuli displayed in succession, a 360ms gap of black screen/ silence was inserted, so that subjects would be able to identify three different items in the respective trials. The final video files were in .avi format, which is required by the eyetracking presentation software Experiment Builder, were .avi. files need to be split into a video and an audio channel for the actual eyetracking recording sessions.

4.3 Experiment Design: Method, Apparatus, Subjects

The empirical paradigm consisted of two eyetracking experiments. Upon arrival of the subjects the purpose of the study, the technology involved and the eyetracking apparatus was explained to them. Following this introduction, written informed consent was obtained from each participant by filling in a form declaring that they agreed to take part in the experiment and to give permission to the pseudonymized use of all collected data. With controls, who had never undergone dyslexia screenings, a reading and writing assessment was performed (following Wimmer et al., 2010). These assessments were either performed individually or as a group assessment. They took place in seminar rooms at the respective schools and/or universities. The eyetracking experiments were performed in four different, quiet and darkened rooms in order to keep distraction noise to a minimum.

In the first experiment, subjects were exposed to the speech perception stimuli. In the second experiment, which was carried out immediately after the first one, subjects performed the pseudoword reading task. All stimuli were programmed to function with an SR Research, Eyelink 1000 desktop eyetracker. The speech perception items were programmed in the Experiment Builder software, which adapted the stimulus set to the eyetracking setting, allowing the presentation of video and audio files (SR-Research, 2011b). In order to present audiovisual stimuli in this software, the .avi files that had been produced with the video- editing program needed to be split into .wav and .xvd files. Depending on the stimulus condition, the acoustic .wav file, or the video .xvd file or the audiovisual combination of .wav and .xvd (congruent and incongruent) were presented.

The stimulus set consisted of a total of 114 items, 7 of which served as introductory items (see Appendix A for the complete stimulus set). The visual, acoustic and audiovisual items were presented in a pseudo- randomized order (obtained via the RAND function in MC Excel). The eyetracking data for the audiovisual and visual only tasks and the pseudoword reading task were recorded monocularly (tracking subjects' left eyes) at 1000 Hertz (i.e. 1 picture/ms) with a 35mm lens. For the audiovisual experiment, subjects had to sit in a chair without armrests and were asked to position their forehead against a headrest. A nine-point calibration with subsequent validation was performed on a 19" 4:3 Belinea monitor.

The eyelink system consisted of a host PC (Dell Optiplex 755 DT Office PC with inbuilt PCI Card for eyetracking camera and Ethernet communication with display PC) and a display PC (both supplied by SR Research), custom built (Intel 2.83 GHz Core2 Quad with 4 GB DDR3 memory and NVidia MSI N440GT 1GB DDR3 VRAM graphics cards as well as an ASIO compatible Sound Blaster sound card for 2 ms audio output). The acoustic signal was deploy at a maximum of 75dB via two Sony (SRS-A35 active speaker system) active loudspeakers. In order to record subjects' responses to the stimuli, subjects were asked to repeat each visual, acoustic and audiovisual stimulus exactly as they 'understood' it. Their replies were recorded with a Sony ICD-SX712 audiorecorder. In addition, subjects' answers were written down on the respective participant form. Each recording session took approximately 60 minutes.

Altogether, 71 subjects and controls were recruited for the eyetracking experiments and were reimbursed ten Euros for their participation in the experiment. The pseudonymized data of all subjects and controls can be found in Appendix C. Out of 36 recruited controls, two participated in the experiments, but their data had to be excluded due to the fact that their first language (L1) was not German, hence it was not entirely clear, what kind of McGurk effects could be triggered via their phonemic inventory. Another two controls' recording sessions had to be discontinued because of eyetracking system crashes. A further control's experimental data were discarded as she did not show up for the reading and writing assessment completed by all other controls (the assessment described by Wimmer et al., 2010). Due to the exclusion of the aforementioned five controls, the number of controls that completed all tasks successfully made up 31. From 36 recruited subjects three were lost: two because of eyetracking calibration failure and one due to an eyetracking system crash. Hence, eyetracking data were collected from 33 subjects and 31 controls, who were all mono-lingual native speakers of German within a dialect region called 'Middle Bavarian'. Subjects (age range: 14 years 4 months – 28 years, mean age: 19 years 5months; 24 males, 7 females) were recruited from two German schools (one grammar school with a one junior high school branch and one vocational school, i.e. a polytechnic institute). All subjects reported a history of dyslexia and had been previously diagnosed as dyslexic by educational psychologists. Their dyslexic status had been officially recognized by the German educational authorities, hence their orthographic mistakes are not to be counted in written assignments. All of the subjects had undergone dyslexia therapy at some point in their educational career. The dyslexic subjects who were recruited via two German schools, had previously granted informed consent through their support teachers and were verbally informed about the nature of the experiments, before they were asked to consent to the experiment and the pseudonymized use of all the collected data. The controls were age matched (age range: 14 years 8 months – 28 years, mean age: 20 years 7months, 24 females, 7 males) and were recruited from two schools (one grammar school and one evening school preparing students for university education) as well as from university courses. For the controls from the high school, a written permission from the local education board had been obtained. All controls underwent an additional 1 minute reading fluency test, where subjects are confronted with a list of sentences and have to decide as fast as possible whether the sentences are true or false by ticking an answer box (as devised and used by Wimmer et al., 2010). Furthermore, controls had to complete a writing assessment for adults (Kersting & Althoff, 2004). These two tests ascertained that controls had no reading or writing problems. Only when the results from the reading fluency test and the writing assessment confirmed the absence of reading and writing problems, controls were invited to participate in the two eye-tracking experiments. The results obtained will be presented and discussed in the subsequent chapter 5.

5. Empirical Research: Results and Discussion

Science never solves a problem without creating ten more.
George Bernard Shaw (1856-1950)

The aim of the experiments I described in the previous section was to ascertain whether dyslexic subjects were as good lip-readers (visual speech) as controls and whether dyslexics responded like controls to audiovisually incongruent speech stimuli (McGurk effect items). Poor lip-reading was hypothesized to correlate with poor pseudoword reading and with few to no elicitations of the McGurk effect. In order to test dyslexics' and controls' visual speech and audiovisual speech reading abilities as well as their pseudoword reading skills, 71 subjects and controls were recruited and tested, which resulted in a sample of 31 data sets for the control group and 33 data sets for the dyslexic experiment group. Eyetracking was used to warrant that subjects and controls fixated the relevant areas of interest of a talking face (speech reading AIs) during those tasks, from which information could be obtained. The presentation of the results and discussion of the latter, this chapter does not only provide a critical view on researching audiovisual speech perception with "real", naturalistic stimuli, but does not fall short to discuss the limitations of such stimuli as well as going to lengths why this is the only kind of speech perception that makes sense in the real world…

5.1 Results

With the 64 valid subject and control data sets, preliminary data analyses were performed in order to identify participants, whose speech perception data needed to be excluded due to lack of fixations on the relevant speech-reading facial parts. For this very purpose, all behavioural data were encoded in a data matrix that allowed to type subjects' responses to the speech perception tasks (see Appendix B). This initial data analysis was accompanied by an eyetracking data analysis with SR Research's Data Viewer software (SR-Research, 2011a), with which subjects' eye movements were first scrutinized for fixations on the speechreading areas of interest ("face.ias"):

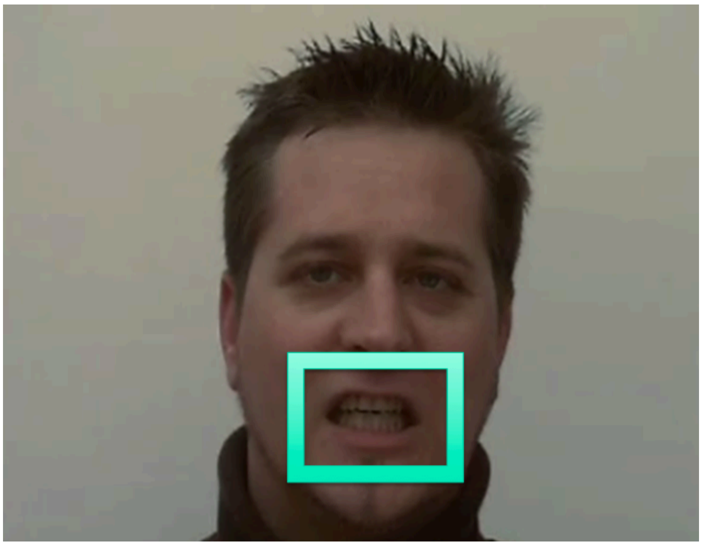


Figure 4: Area of Interest "speechreading" as used for the preliminary data analysis (stimulus created by author).

When subjects' fixations did not lie within this area of interest, the stimulus item was coded as 'missing' and excluded from the data set. For this purpose the trials were grouped following the command 'edit trial grouping by condition' into auditory (code 3 – the auditory only condition did, of course, not have any relevant eye movement data) audiovisual (code 1 – av) and visual trials (code 2 – v). Then the default interest area *AI Speechrieading.ias* was applied. These trial data were then included in the data output report. The variables included in this report were: *recording session label*, *IA* (interest area) *dwell time*, *IA ID* (interest area identification) and *IA label* (IA label, i.e. the default AI 'Speechreading'), *trial index, IA fixation count, video, wav, condition, count* and *fixceck* (to ascertain that fixations ocurred at all)'. The audio only trials were, of course, excluded from the eyetracking data evaluation, as subjects had to repeat what had been presented acoustically, rendering eye movements irrelevant for this task.

The output report was translated into excel where the data were checked for missing fixations in the AIs for the relevant items (all McGurk items, all visual only items and all audiovisually congruent items). The participants' trials were matched onto the respective stimulus items. These data were then copied into the data matrix that included the record of participants' responses of their percepts. In the data matrix every trial was labelled according to the given answer and the corresponding eye movements. Trials with missing eye movement data or missing repetition were labelled '9=missing'. According to these labels, the data were computed. For convenience, these labels are reproduced again here:

Auditory and audiovisual speech labels:

0= correct

1= incorrect

Visual Speech labels:

0= correct answer, i.e. all aspects were repeated back correctly.

1= partly correct, i.e. either not all the visual gestures were repeated back correctly or some were left out, or most aspects were identified. In the statistical analyses, these responses were encoded as correct.

2= partly incorrect, i.e. most aspects were not identified correctly. In the statistical analyses, these responses were encoded as correct. In the statistical analyses, these responses were encoded as incorrect.

3= incorrect, i.e. either no answer was given at all or an answer was provided with none of the aspects being correctly identified, not even the item length.

Audiovisually incongruent (McGurk) labels:

0= correct, i.e. the McGurk effect was elicited.

1= the acoustic part of the McGurk item is perceived, i.e. the effect could not be elicited. In the statistical analyses, these responses were encoded as incorrect.

2= the visual aspect of the McGurk item is perceived, i.e. the effect could not be triggered. In the statistical analyses, these responses were encoded as incorrect.

3 = the answer is not correct, neither visual nor acoustic aspects are repeated back correctly by the subject. These responses were encoded as incorrect.

With the McGurk items, it was noticeable that there was only one control, who perceived the visual aspect (answer label 2), when the effect could not be triggered. When the effect was not triggered, the response was always the acoustic aspect in dyslexic subjects. It is also noteworthy that in some control subjects the McGurk items yielded 100% McGurk percepts. In the majority of dyslexic subjects no McGurk effects were elicited.

5.2 Statistical Analyses

On the basis of this data matrix, the following statistical procedures were computed (the statistics syntax is to be found in Appendix D):

For each task (visual, auditory, audiovisual and audiovisually incongruent/ McGurk) participant's percentage of correct responses (accuracy) was assessed. A 4x2 Analysis of variance (ANOVA) was performed including task as within-subjects-factor and group as between-subjects-factor (dyslexic/non dyslexic). Across tasks, dyslexic participants performed worse than non dyslexic participants (main effect of group: $F(1)= 58.57$, $p < .001$). A significant main effect of task on accuracy ($F(3,60) = 173.71$, $p < .001$) was observed. In both groups, accuracy was highest in auditory and audiovisual stimuli and worst in McGurk stimuli. Furthermore, a significant interaction between task and group ($F(3,60)=14.68$, $p < .001$) was found. To resolve this interaction post hoc univariate group comparisons for each task was performed (see tables 1 & 2). Group differences were most pronounced for visual items, followed by McGurk stimuli, and least pronounced in audiovisual stimuli. In auditory items difference was not significant.

Since dyslexic participants performed significantly worse in the pseudoword-task than non-dyslexic participants ($t(62)= 10.59$, $p < .001$), in a second step pseudoword-reading-performance was entered as a covariate. The main effects of task ($F(3,59)= 39.82$, $p < .001$) and group ($F(1)= 16.33$, $p < .001$) remained significant.

Post-hoc analysis then revealed also the largest group difference for visual items, followed by McGurk items, followed by auditory items while performance in audio-visual items was comparable between groups when pseudoword-reading was controlled for.

As susceptibility to the McGurk effect correlated with lipreading skills ($r= .54$, $p= .000$), a univariate analysis of covariance (ANCOVA) was consequently calculated, comparing dyslexic subjects' and controls' susceptibility to McGurk items controlling for lipreading skills. The obtained analysis suggests a significant difference between groups in McGurk susceptibility when lipreading skill is controlled ($F(1)= 6.46$, $p < .05$).

UniAN(C)OVAs

	SG	CG	df	F_{group}	η^2_{group}	F_{group_pw}	$\eta^2_{group\_pw}$
McGurk Stimuli	0.16 (0.18)	0.48 (0.30)	1	26.38***	.298	4.60*	.070
Audiovisual Stimuli	0.96 (0.08)	0.99 (0.02)	1	6.66*	.097	1.63	.026
Visual Stimuli	0.59 (0.24)	0.89 (0.09)	1	45.21***	.422	14.27***	.190
Auditory Stimuli	0.89 (0.10)	0.93 (0.08)	1	2.29	.036	4.15*	.064

** $p < .05$, ** $p < .01$, *** $p < .001$*

_pw: controlled for pseudoword-reading ability

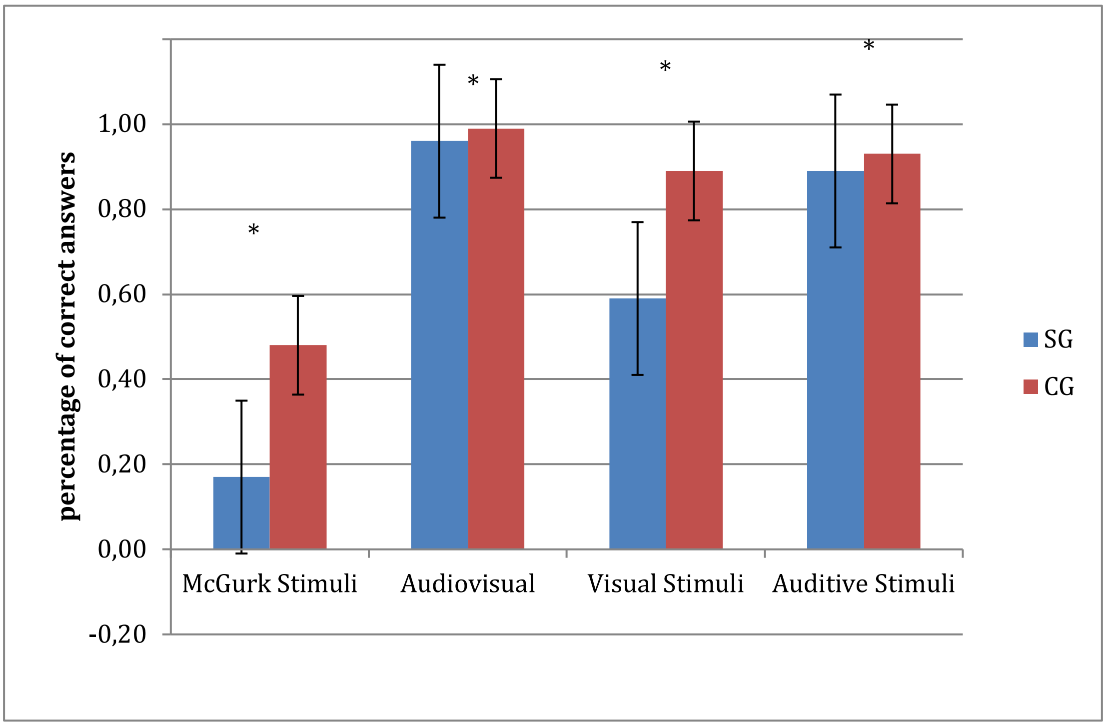
F_{group}: Main Effect of group

η^2_{group}: Effect size (Eta square) for group differences

F_{group_pw}: Main Effect of group controlling for pseudoword reading ability

$\eta^2_{group\_pw}$: Effect size (Eta square) for group differences controlling for pseudoword reading ability

Table 1: Subjects' and controls' performance across all tasks when pseudoword reading is controlled.



*significant

Table 2: Bar charted responses to speech perception tasks

As suggested in the previous chapter, where the McGurk stimulus set was introduced and outlined, the audiovisually incongruent items were grouped according to the position of the McGurk trigger and voicing/devoicing. Furthermore, the McGurk items that consisted of lexical items, i.e. the real word items, were computed in isolation and were not considered in the audiovisual deficit which I shall discuss in the following section.

Hence, in the statistical analysis, participant's percentage of correct responses (accuracy) was assessed for each kind of McGurk-items A 5x2 Analysis of variance (ANOVA) was performed including the McGurk-task as within-subjects-factor and group as between-subjects-factor (dyslexic/non dyslexic). The different kinds of McGurk items are reproduced here for convenience:

avmganlaut+voice (= adiovisually incongruent, i.e. McGurk trigering in the word onset)	Code Number: 4
avmganlaut-voice (= adiovisually incongruent, i.e. McGurk trigering in the word onset)	Code Number: 5
avmgintervok+voice (= adiovisually incongruent, i.e. McGurk trigering intervocalic)	Code Number: 6
avmgintervok-voice (= adiovisually incongruent, i.e. McGurk trigering intervocalic)	Code Number: 7
avmglexanlaut-voice (= adiovisually incongruent, i.e. McGurk trigering in the word onset)	Code Number: 0

Key: avmganlaut+voice: /ba/ + /ga/; /barulo/ + /garulo/
avmganlaut-voice: /pa/+ /ka/; /pelami/ + /kelami/
avmgintervok+voice: /aba/ + /aga/; /mabali/ + /magali/
avmgintercok-vopice: /apa/ + /aka/; /ap^{h}a/ + / ak^{h}a/
avmglexanlaut-voice: /picken/ + /kicken/; /pasten/ + /kasten/

Across McGurk-tasks dyslexic subjects performed worse than controls (main effect of group: F(1)= 26.13, p < .001). In both groups, accuracy was lowest in McGurk_7 tasks and best in McGurk_0 tasks. We observed no significant interaction between task and group (F(4,59)=0.69, p > .05) but a significant main effect of task (F (4,59) = 4.67, p < .01).

Group differences were significant for all McGurk_tasks (t(62)> 3.04). They were most pronounced in McGurk_7 tasks and least in McGurk_0 tasks.

Since dyslexics performed significantly worse in pseudoword-tasks than controls (t(62)= 10.59, p < .001), in a second step pseudoword-reading-performance was entered as a covariate. The main effect of group (F(1)= 4.39, p < .05) remained significant.

However, post-hoc analysis revealed that performance in McGurk_0 stimuli, McGurk_4 stimuli and McGurk_5 stimuli was comparable between groups when pseudoword-reading is controlled.

UniAN(C)OVAs

	SG	CG	df	F_{group}	η^2 group	F_{group_pw}	η^2 $_{group\_pw}$
McGurk _0	0.28 (0.32)	0.53 (0.28)	1	9.24**	.130	0.65	.011
McGurk _4	0.14 (0.21)	0.49 (0.37)	1	22.91***	.270	3.47	.054
McGurk _5	0.16 (0.23)	0.51 (0.36)	1	21.09***	.254	3.33	.052
McGurk _6	0.16 (0.24)	0.49 (0.37)	1	18.00***	.225	4.13*	.063
McGurk _7	0.11 (0.17)	0.39 (0.27)	1	25.25***	.252	6.15*	.092

** p < .05, ** p < .01, *** p < .001*

_pw: controlled for pseudoword-reading ability

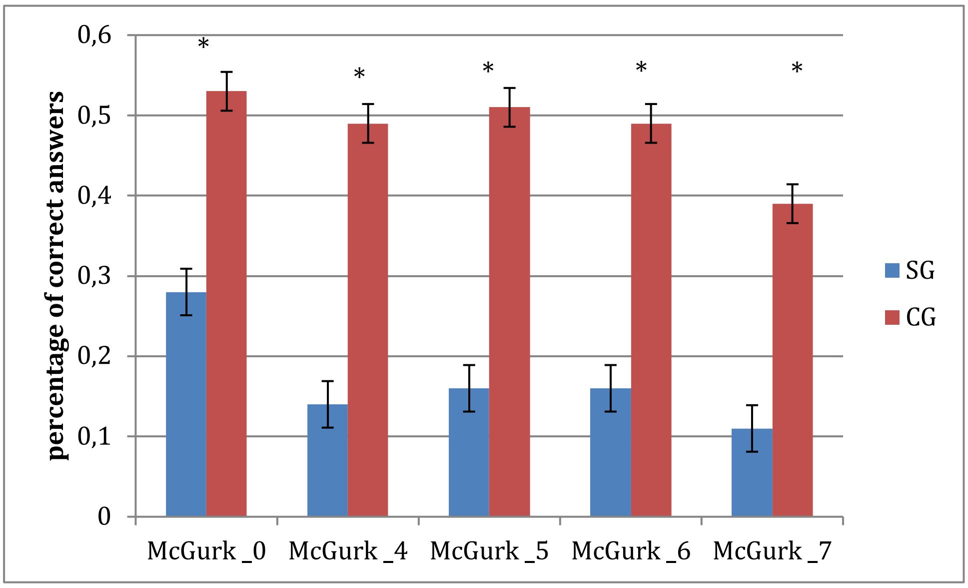
$F_{group:}$ Main Effect of group

η^2 $_{group:}$ Effect size (Eta square) for group differences

$F_{group\_pw:}$ Main Effect of group controlling for pseudoword reading ability

$\eta^2_{group\_pw:}$ Effect size (Eta square) for group differences controlling for pseudoword reading ability

Table 3: Responses to the various McGurk items



* significant

Table 4: Bar charted responses to McGurk items

Consequently, all sub-lexical McGurk item types (4-7) were included in one single group, which then compared the combined percentage of correct answers of this items group to the lexical McGurk items (McGurk_0).

The same calculation, a 2x2 analysis of variance (ANOVA), was performed including McGurk-tasks (McGurk_0 vs. McGurk_4-7) as within-subjects-factor and group as between-subjects-factor (dyslexic/non dyslexic). Again, dyslexics performed worse across sub-lexical McGurk-tasks than controls (main effect of group: $F(1)= 20.90$, $p < .001$). In both groups, accuracy was lowest in McGurk_4-7 (non lexical) tasks and best in McGurk_0 (lexical) tasks. No significant interaction between task and group ($F(1,62)=0.96$, $p > .05$) could be observed, although there was a significant main effect of task ($F(1,62) = 6.84$, $p < .05$). Group differences were higher for sublexical McGurk (4-7) tasks (33%; $t(62)= 5.01$, $p < .001$) compared to lexical McGurk (0) tasks (25%, $t(62)= 3.04$) and significant in both cases.

UniAN(C)OVAs

	SG	CG	df	F_{group}	η^2_{group}	F_{group_pw}	$\eta^2_{group\_pw}$
McGurk _0	0.28 (0.32)	0.53 (0.35)	1	9.23**	.130	0.65	.011
McGurk _4-7	0.14 (0.18)	0.47 (0.32)	1	25.89***	.295	4.92*	.075

** p < .05, ** p < .01, *** p < .001*

_pw: controlled for pseudoword-reading ability

F_{group}: Main Effect of group

η^2_{group}: Effect size (Eta square) for group differences

F_{group_pw}: Main Effect of group controlling for pseudoword reading ability

$\eta^2_{group\_pw}$: Effect size (Eta square) for group differences controlling for pseudoword reading ability

Table 5: Responses to lexical and non-lexical McGurk items

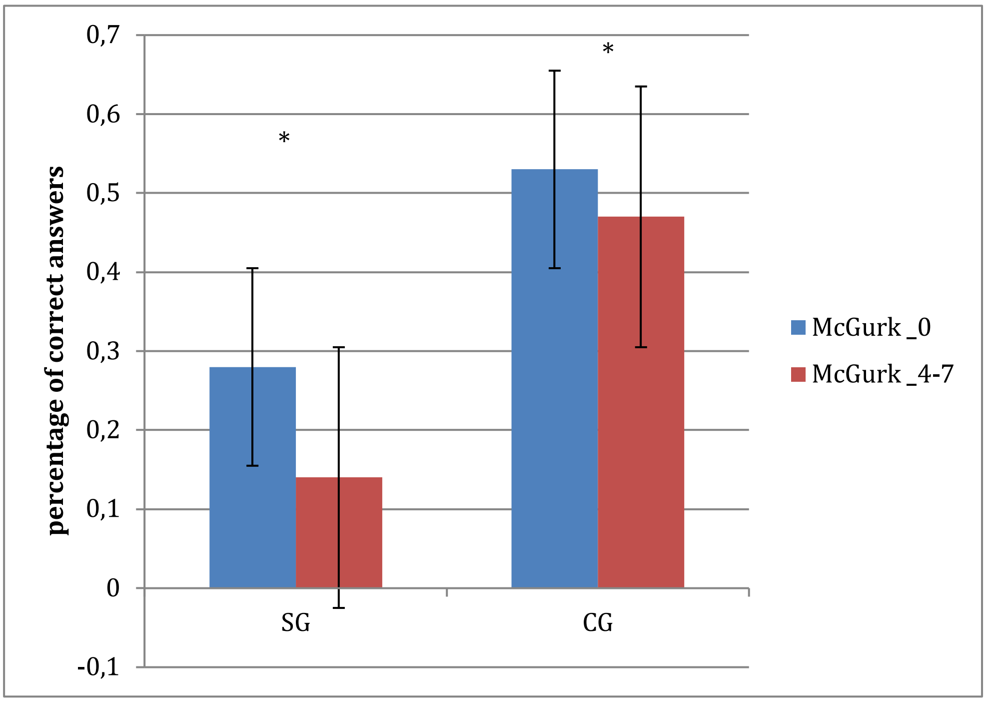


Table 6: Bar charted responses to lexical and sub-lexical items

As depicted above in table 6, lexical items triggered more McGurk responses across both groups of subjects. In chapter 2.3.3, however, it was discussed that lexical items cannot be easily controlled for word frequency and subjects' anticipation for three different lexical items, which is why these items were not considered further.

5.3 General Discussion

The experiments I conducted as well as the theoretical framework outlined in this book were based on the assumption that at the core of the linguistic symptoms of dyslexia lies a – potentially subtle – speech signal processing deficit. Overall, the results of the experiments provide strong evidence that in dyslexics, audiovisual speech perception and visual speech perception are poorly developed skills compared to unimpaired controls. The results further revealed that the phonological abilities required to read pseudowords seem to be generally impaired across the dyslexic subjects of the sample tested here. Hence, the phonological deficit which results in poor pseudoword reading, is clearly present in the adult subjects that were recruited for this research, as could be expected from the findings of various other researchers who studied dyslexics' phonological reading skills (cf. Marshall et al., 2011; M. Snowling et al., 2000; M. Snowling & Stackhouse, 2006; M. J. Snowling, 2006; M. J. Snowling & Hulme, 2005; Vellutino, 1981; Vellutino et al., 2004). In dyslexia research, speech perception problems beyond the scope of auditory processing deficits have long been dealt with from a phonological perspective (cf. Baart et al., 2012; Blomert & Mitterer, 2004; Mody et al., 1997; Serniclaes et al., 2004).

This study demonstrated that subjects and controls benefitted equally from the bimodal presence of visual and acoustic information in the congruent speech perception tasks, showing a ceiling effect across all stimulus types. However, in unimodal conditions, there is a significant difference between subjects' and controls' lip-reading abilities. Thus, the expected strong correlation between the visual only task and poor pseudoword task performance for dyslexic subjects was confirmed, thereby supporting the hypothesis that subjects have great problems identifying the speechreadable (= lipreadable) aspects of visual speech signals. While some researchers have not found evidence that lip-reading was impaired in adult dyslexics (Baart et al., 2012) and in dyslexic children (Bastien-Toniazzo et al., 2009), the result here clearly implies that lip-reading is comparatively poorer in the dyslexic subjects who took part in these experiments.

The fact that dyslexics and controls perform equally well in perceiving phonological units and words in the auditory- only condition refutes the attempt to complement the phonological deficit hypothesis with poorer acoustic perception skills, for instance caused by low-level auditory deficits as suggested by Tallal (1980), although subjects' non susceptibility to McGurk tasks may support her approach that there might be a structural temporal impairment.

The McGurk items yielded the expected results, i.e. they verify the hypothesis that dyslexic subjects and non-dyslexic controls differ significantly in their susceptibility to the McGurk effect. However, controls' susceptibility to McGurk items was nowhere near as strong as found in some McGurk studies with adult subjects (cf. McGurk & MacDonald, 1976; Sams et al., 2005) but is very much in line with results from more recent studies (Nath & Beauchamp, 2012; Skipper et al., 2007; Szycik et al., 2012).

In summary, the results of the current study revealed that the stimulus set and the experiments into which they were incorporated could successfully be applied to testing the formulated hypotheses and to verify my hypotheses. A specific avenue is therefore required to formulate the aforementioned audiovisual deficit and its implications for further research. Moreover, an explanation is required, why the McGurk effect cannot be triggered robustly in dyslexics and what implications this might have for reading research.

5.4 Specific Discussion and Conclusive Remarks

What is the impact of these findings on current dyslexia research? In the introduction (p. 7.), I have argued that weak responses to the McGurk effect in dyslexic subjects indicate a multimodal, i.e. audiovisual integration problem, whereas a robust McGurk effect would indicate that audiovisual integration of speech signals is intact. Also, the alleged subtle speech processing deficit should help explain the poor performance in the audiovisual tasks and fit the recent renaissance of motor speech theories (cf. Nusbaum, 2011). So, let us return to the characteristic features of audiovisual speech perception: In chapter 2, we have looked at the nature of audiovisual speech, including the sensory- motor basis for speech perception. In this context I commented on the neural patterns involved in audiovisual speech perception and referred to the study by Skipper et al., who had identified neural activity patterns during McGurk stimuli perception (Skipper et al., 2007). They had found activation in the ventral premotor region for the McGurk percept of [ta] and attributed the auditory percept [pa] to the supermarginal gyrus, which would shift to the premotor region in case of a McGurk percept. They consider the visual [ka] to be best represented in the middle occipital gyrus, where the actual mouth articulation is processed, from whence it shifts to the premotor region in a McGurk percept. If this shift does not occur, the McGurk effect can thereupon not be triggered. Could the explanation for dyslexics' non- susceptibility and acoustic-only, i.e. /pa/ and /ba/ responses to the McGurk effect thus be based on a motor 'undernourishment' of the relevant visual, i.e. sensory, areas? The poor lip-reading scores of subjects in the sample of this study would also confirm poor visual processing skills of visual speech gestures and these gestures would be less distinct or absent due to the lack of motor representation. Hence, the poor visual speech processing skills, would be traceable to poor motor representations (not skills), explaining both, poor lip-reading and indifference to McGurk items. It is rather an interesting view, when the locus of the visual representation of articulatory gestures, as identified by Skipper et al. (2007), is compared to Dehaene's reading circuits in the brain. In this comparison we can see regions activated when syllable videos are watched, linked closely to Dehaene's reading brain circuits. It is particularly noteworthy here that we can identify the same left hemisphere regions involved in reading and in lip-reading. Weak and/or absent activations in these regions in dyslexics might therefore suggest these areas to be possible loci for a visual speech processing and audiovisual integration deficit. This is in line with Nusbaum's view that "[…] visual and auditory information, when fed to the premotor cortex, give rise to an activity pattern consistent with the

McGurk illusion which then may interact with sensory cortices resulting in a final activity pattern consistent across all regions with the McGurk percept" (Nusbaum, 2011, p. 673). It would also match his explanation of an "[…] active theory of perception for which the lack of invariance between acoustic patterns and phonetic categories is resolved by an interaction between articulatory knowledge in the premotor representation of speech"(ibid.). I have discussed earlier, how Nusbaum considers premotor activity a constraint in the absence of a visual input, which he relates to Liberman and Mattingly's revised motor theory, in which motor knowledge is relevant in any case, that is, with or without visible articulator movement.

In line with these suggestions, Giraud and Poeppel (Giraud & Poeppel, 2012), recently discussed speech sampling deficits in dyslexia. In audio-visual language processing, research on the Mismatch Negativity phenomenon has been investigated with typically developing and dyslexic children, among others by Sams et al. (Sams, Kaukoranta, et al., 1991; Tuomainen et al., 2005). Low susceptibility to MMN in dyslexic subjects and poor orthographic skills would seem to occur within a time frame of around 200 ms. Following Poeppel et als.' 'two time windows' approach for speech perception, how could we explain the absent MMN in dyslexics? According to their hypothesis there are "two principal time windows within which a given auditory signal (speech or non-speech) is processed" (Poeppel et al., 2009, p. 258). The time window of approx. 20-80ms computes segmental and subsegmental cues and processes and the segmental 'order'. Here, one would locate the ability to acoustically distinguish between consonant sounds and small clusters (remember Poeppel et als' example of "pest" vs. "pets"). If a temporal speech sound processing deficit occurs within this time window, the subsequent computations may already be disturbed. At the time scale of 150-300ms, suprasegmental and syllabic phenomena should be processed, which again may be inhibited by a delay from the first time window processes. This delay or disruption might well explain the temporal processing problem within the time window of 200ms. Lexical items may be easier to process as there is also semantic and contextual as well as word frequency help. With the sub-lexical McGurk stimuli used in the audiovisual task, this contextual interference was avoided.

What I have referred to as a delay here has already been suggested by Breznitz' Asynchrony Theory, in which she considers dyslexia a problem in synchronising the temporal aspects of the reading process (Breznitz, 2008, p. 11). Breznitz' also recognizes differences in processing speeds between the brain areas, and a lack of coordination among the reading network's regions. Her 'lag' or 'gap' in the processing speed of information that travels between the different brain areas involved in decoding words and being responsible for the prevention of accurate synchronization. Wimmer's (2013) research suggests that at the core of dyslexia lies a functional disconnectivity problem in the visual word form area as well as reduced connectivity between the VWFA and the inferior frontal regions explaining both, the phonological decoding problems and the 'speed impairment'. With these temporal processing and connectivity problems in mind, it will be enthralling to follow the direction of audiovisual research that covers temporal processing problems often attributed to dyslexics.

If one takes a closer look at visual and audiovisual speech perception, another striking parallel unfolds: Learning to read brings on a cortical reorganization of the neural networks that process language (cf. Dehaene et al., 2010) and alters the representation of visual aspects, specifically word forms, faces and objects (Dehaene, 2013; Dehaene et al., 2010). Yet, learning to read also results in a reorganization of the smallest meaningful spoken linguistic units, whether we call them phonemes or by a different term (C. A. Fowler, 2011). Through the acquisition of graphemes and the subsequent need to master potential ambiguity in written language as well as to profit from their potential disambiguating function in spoken language, the beginning reader gains insight into a hitherto unknown structural aspect of language. In other words, phonemic awareness is developed early on through motor and sensory skills, is furthermore influenced by the sensory aspects of visual speech, and is subject to alteration as literacy is developed. If this insight/ alteration is prevented because the visual aspects of speech have not been developed in early infancy, causing unawareness of visemes, then a visual counterpart to sound units is never developed. If a subtle motor dysfunction thwarts the ability to imitate articulatory gestures, then the necessary basis for motor-sensory skills is forestalled.

Another comparison between the audiovisual speech network and the reading network may support this. As described earlier, Dehaene's considers the VWFA the locus for processing word forms, faces and objects (Dehaene, 2013), and for Campbell, the area VT, adjacent to and overlapping with the VWFA it is a central region for audiovisual integration, making it one of the relevant parts of a 'McGurk' network.

An fMRI study by the author and others is currently in progress to test dyslexic subjects and unimpaired controls with the single syllable items of this study (Kaltenbacher, Pfleiderer, Bühner, Hummer, & Breninger, forthcoming). This fMRI study will allow insight into the dyslexic brain's audiovisual speech processing.

5.5 Outlook: Audiovisual Deficits, Future Diagnostics and a Therapeutic Approach

By means of this book' experiments, a link has been established between deficient audiovisual integration of speech stimuli and dyslexia. Adolescent and adult subjects had been chosen for these experiments, as reading acquisition in these subjects had already been completed and their reading behaviour was not likely to change. The next step will now be to bridge the gap between reading research and speech reading research with beginning readers and pre-reading children. Stackhouse has already shown that poor readers present significantly often with speech and language difficulties (Stackhouse, 2006). She had also surmised a 'subtle' speech and language problem that may persist in older dyslexic children. In order to identify a beginning audiovisual integration deficit, a new set of test items is required. For such a venture, a childrens' version of the visual, auditory, audiovisual and McGurk items will be created. Furthermore, MMN stimuli and categorical speech perception items must be included in such a test battery, with the advantage that both these technologies are also "child-friendly". Motor speech skills may be easily assessed through child-friendly imaging technologies such as Ultrasound Tongue Imaging (cf. Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007; Bernhardt et al., 2008; Modha, Bernhardt, Church, & Bacsfalvi, 2008).

The aim to create such a test battery would not simply be to design a diagnostic tool allowing to screen speech perception in three conditions (with an option of adding masking conditions, such as noise in acoustic or blurring in visual items) but to foster a therapeutic approach that enables beginning readers to master the challenge of acquiring a new linguistic modality and successfully adapting an old one. Therefore future work will be dedicated to the design of a multimodal speech processing assessment (MSPA) for pre-reading children as well as reading impaired children.

Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *Am J Speech Lang Pathol, 16*(2), 128-139. doi:10.1044/1058-0360(2007/017)

Adobe. (2010). Adobe Premiere Pro CS5. München: Adobe Systems GmbH.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current biology : CB, 15*(9), 839-843. doi:10.1016/j.cub.2005.03.046

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Brain research Cognitive brain research, 21*(3), 301-308. Retrieved from <Go to ISI>://MEDLINE:15511646

Auer, E. T., Jr., & Bernstein, L. E. (1997). Speechreading and the structure of the lexicon: computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *J Acoust Soc Am, 102*(6), 3704-3710. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9407662

Baart, M., de Boer-Schellekens, L., & Vroomen, J. (2012). Lipread-induced phonetic recalibration in dyslexia. *Acta Psychol (Amst), 140*(1), 91-95. doi:10.1016/j.actpsy.2012.03.003

Bartlett, C. W., Flax, J., Yabut, O., Hirsch, L., Li, W., Tallal, P., & Brzustowicz, L. (2000). A Genome-scan for linkage of Specific Language Impairment: Report on chromosomes, 1, 6, and 15. *Am J Hum Genet, 67*(4), 309-309. Retrieved from <Go to ISI>://WOS:000089400701709

Bastien-Toniazzo, M., Stroumza, A., & Cavé, C. (2009). Audio-visual Percpetion and Integration in Developmental Dyslexia: An Exploratory Study Using the McGurk Effect. *Current psychology letters, 25*(3), 2-14.

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends Cogn Sci, 10*(11), 512-518. doi:10.1016/j.tics.2006.09.006

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci, 30*(7), 2414-2417. doi:10.1523/JNEUROSCI.4865-09.2010

Benasich, A. A., & Tallal, P. (1996). Auditory temporal processing thresholds, habituation, and recognition memory over the 1st year. *Infant Behavior & Development, 19*(3), 339-357. doi:Doi 10.1016/S0163-6383(96)90033-8

Benasich, A. A., & Tallal, P. (2002). Infant discrimination of rapid auditory cues predicts later language impairment. *Behav Brain Res, 136*(1), 31-49. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12385788

Bensmaia, S. J., Killebrew, J. H., & Craig, J. C. (2006). Influence of visual motion on tactile motion perception. *J Neurophysiol, 96*(3), 1625-1637. doi:10.1152/jn.00192.2006

Bergmann, J., & Wimmer, H. (2008). A dual-route perspective on poor reading in a regular orthography: evidence from phonological and orthographic lexical decisions. *Cogn Neuropsychol, 25*(5), 653-676. doi:10.1080/02643290802221404

Bernhardt, M. B., Bacsfalvi, P., Adler-Bock, M., Shimizu, R., Cheney, A., Giesbrecht, N., . . . Radanov, B. (2008). Ultrasound as visual feedback in speech habilitation: exploring consultative use in rural British Columbia, Canada. *Clin Linguist Phon, 22*(2), 149-162. doi:10.1080/02699200701801225

Bernstein, L., E. (2012). Visual speech perception. In G. Bailly, P. Perrier, & E. Vatikiotis-Bateson (Eds.), *Audiovisual Speech Processing* (pp. 21-39). Cambridge: Cambridge University Press.

Bernstein, L. J., & Robertson, L. C. (1998). Independence between illusory conjunctions of color and motion with shape following bilateral parietal lesions. Psychol. Sci. 9, 167–175. *Psychol Sci, 9*, 167-175.

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: same or different? *Psychol Bull, 130*(6), 858-886. doi:10.1037/0033-2909.130.6.858

Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *Neuroimage, 57*(3), 695-703. doi:10.1016/j.neuroimage.2010.11.003

Blomert, L., & Mitterer, H. (2004). The fragile nature of the speech-perception deficit in dyslexia: natural vs synthetic speech. *Brain Lang, 89*(1), 21-26. doi:10.1016/S0093-934X(03)00305-5

Blomert, L., Mitterer, H., & Paffen, C. (2004). In search of the auditory, phonetic, and/or phonological problems in dyslexia: context effects in speech perception. *J Speech Lang Hear Res, 47*(5), 1030-1047. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15603461

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read[mdash]a causal connection. *Nature, 301*(5899), 419-421. Retrieved from http://dx.doi.org/10.1038/301419a0

Brady, S. A., Braze, D., & Fowler, C. A. (2011). *Explaining Individual Differences in Reading: Theory and Evidence*: Psychology Press.

Breznitz, Z. (2008). The Origin of Dyslxia: The Asynchrony Phenomenon. In G. Reid, A. Fawcett, F. Manis, & L. Siegel (Eds.), *The SAGE Handbook of Dyslexia* (pp. 11-29): SAGE Publications.

Brindley, G. S. (1964). The Use made by the Cerebellum of the Information that it receives from the Sense Organs. *International Brain Research Organization Bulletin, 3*(80).
Burnham, D., & Dodd, B. (1996). Auditory-Visual Speech Perception as a Direct Process: The McGurk Effect in Infants and Across Languages. In D. Stork & M. Hennecke (Eds.), *Speechreading by Humans and Machines* (Vol. 150, pp. 103-114): Springer Berlin Heidelberg.
Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220. doi:10.1002/dev.20032
Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage, 14*(2), 427-438. doi:10.1006/nimg.2001.0812
Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn Sci, 11*(12), 535-543. doi:10.1016/j.tics.2007.10.001
Campbell, C. S., & Massaro, D. W. (1997). Perception of visible speech: influence of spatial quantization. *Perception, 26*(5), 627-644. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9488886
Campbell, R. (2009). The processing of audio-visual speech: empirical and neural bases. In B. C. J. Moore, L. K. Tyler, & W. D. Marslen-Wilsen (Eds.), *The Perception of Speech: from sound to meaning* (2007/09/11 ed., pp. 133-150). Oxford: Oxford University Press.
Campbell, R., & MacSweeney, M. (2012). Brain bases for seeing speech:fMRI studies of speechreading. In G. Bailly, P. Perrier, & E. Vatikiotis-Bateson (Eds.), *Audiovisual Speech Processing* (pp. 76-103). Cambridge: Cambridge University Press.
Caplan, D. (1996). *Language: Structure, Processing, and Disorders*. Cambridge: MIT Press.
Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*(1), 77-111. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14711492
Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York,: Harper & Row.
Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123 (Pt 2)*, 291-307. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10648437
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*(4), 495-506. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11955994
Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 6-23): Blackwell.
Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol Rev, 108*(1), 204-256. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11212628
Cornelissen, P., Richardson, A., Mason, A., Fowler, S., & Stein, J. (1995). Contrast sensitivity and coherent motion detection measured at photopic luminance levels in dyslexics and controls. *Vision research, 35*(10), 1483-1494.
De Bleser, R. (2004). *LEMO - Lexikon modellorientiert: Einzelfalldiagnostik bei aphasie, Dyslexia und Dysgraphie. Entscheiden, Nachsprechen: Diagnostikbände, Band 3* (Vol. 3): Elsevier, Urband & Fischer.
de Gelder, B., & Vroomen, J. (1998). Impaired speech perception in poor readers: evidence from hearing and speech reading. *Brain Lang, 64*(3), 269-281. doi:10.1006/brln.1998.1973
de Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophr Res, 72*(2-3), 195-203. doi:10.1016/j.schres.2004.02.013
de Jong, P. F., Bitter, D. J. L., van Setten, M., & Marinus, E. (2009). Does phonological recoding occur during silent reading, and is it necessary for orthographic learning? *J Exp Child Psychol, 104*(3), 267-282. doi:http://dx.doi.org/10.1016/j.jecp.2009.06.002
de Jong, P. F., Seveke, M. J., & van Veen, M. (2000). Phonological sensitivity and the acquisition of new words in children. *J Exp Child Psychol, 76*(4), 275-301. doi:10.1006/jecp.1999.2549
Dehaene, S. (2009). *Reading in the Brain: The New Science of How We Read.* New York: Penguin Books.
Dehaene, S. (2013). Inside the letterbox: how literacy transforms the human brain. *Cerebrum, 2013*, 7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23847714
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704307/pdf/cer-07-13.pdf
Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., . . . Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science, 330*(6009), 1359-1364. doi:10.1126/science.1194140

Desmond, J. E., & Fiez, J. A. (1998). Neuroimaging studies of the cerebellum: language, learning and memory. *Trends Cogn Sci, 2*(9), 355-362.
Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw, 12*(7-8), 961-974. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12662639
Eccles, J. C., Ito, M., & Szentagothai, J. (1967). *The Cerebellum as a Neuronal Machine*. Nwe York: Springer.
Eden, G. F., VanMeter, J. W., Rumsey, J. M., Maisog, J. M., Woods, R. P., & Zeffiro, T. A. (1996). Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. *Nature, 382*(6586), 66-69. doi:10.1038/382066a0
Ehri, L., C. (2005). Development of Sight Word REading: Phases and Findings. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 135-154): Blackwell.
Facoetti, A., Turatto, M., Lorusso, M. L., & Mascetti, G. G. (2001). Orienting of visual attention in dyslexia: evidence for asymmetric hemispheric control of attention. *Exp Brain Res, 138*(1), 46-53.
Fawcett, A. J., & Nicolson, R. I. (1992). Automatisation deficits in balance for dyslexic children. *Percept Mot Skills, 75*(2), 507-529. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1408614
Fawcett, A. J., & Nicolson, R. I. (2001). Dyslexia: The Role of the Cerebellum. In A. J. Fawcett (Ed.), *Dyslexia: Theory and Good Practice* (pp. 89-105). London: Whurr.
Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*: MIT Press.
Fowler, A. E., & Swainson, B. (2004). Relationships of naming skills to reading, memory, and receptive vocabulary: evidence for imprecise phonological representations of words by poor readers. *Annals of Dyslexia, 54*(2), 247-280. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15741938
Fowler, C. A. (2011). How Theories of Phonology May Enhance Understanding of the Role of Phonology in Reading Development and Reading Disability. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining Individual Differnces in Reading: Theory and Evidence* (pp. 3-19): Psychology Press.
Friederici, A. D. (2006). The neural basis of language development and its impairment. *Neuron, 52*(6), 941-952. doi:10.1016/j.neuron.2006.12.002
Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading*. London: Lawrence Erlbaum Associates.

Gaillard, R., Naccache, L., Pinel, P., Clemenceau, S., Volle, E., Hasboun, D., . . . Cohen, L. (2006). Direct intracranial, FMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron, 50*(2), 191-204. doi:10.1016/j.neuron.2006.03.031

Gibbon, F. (2009). Instrumental Analysis of Articulation in Speech Impairment. In M. J. Ball, M. R. Perkins, N. Müller, & S. Howard (Eds.), *The Handbook of Clinical Linguistics* (pp. 311-331). Oxford: Blackwell Publishing.

Gibbon, F., Crampin, L., Hardcastle, B., Nairn, M., Razzell, R., Harvey, L., & Reynolds, B. (1998). CleftNet (Scotland): a network for the treatment of cleft palate speech using EPG. *Int J Lang Commun Disord, 33 Suppl*, 44-49. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10343662

Gibbon, F., Hardcastle, B., & Dent, H. (1995). A study of obstruent sounds in school-age children with speech disorders using electropalatography. *Eur J Disord Commun, 30*(2), 213-225. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7492852

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *J Speech Lang Hear Res, 42*(2), 382-397. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10229454

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci, 15*(4), 511-517. doi:10.1038/nn.3063

Goswami, U., Wang, H. L., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *J Cogn Neurosci, 23*(2), 325-337. doi:10.1162/jocn.2010.21453

Gould, L., Cummine, J., & Borowsky, R. (2012). The Cognitive Chronometric Architecture of Reading Aloud: Semantic and Lexical Effects on Naming Onset and Duration. *Front Hum Neurosci, 6*. doi:10.3389/fnhum.2012.00287

Greenberg, S., & Arai, T. (2004). What are the essential cues for understanding spoken language? *Ieice Transactions on Information and Systems, E87d*(5), 1059-1070. Retrieved from <Go to ISI>://WOS:000221445100003

Harris, J. M., & German, K. J. (2008). Comparing motion induction in lateral motion and motion in depth. *Vision research, 48*(5), 695-702. doi:10.1016/j.visres.2007.11.018

Hawelka, S., & Wimmer, H. (2005). Impaired visual processing of multi-element arrays is associated with increased number of eye movements in dyslexic reading. *Vision research, 45*(7), 855-863. doi:10.1016/j.visres.2004.10.007

Hayes, E. A., Tiippana, K., Nicol, T. G., Sams, M., & Kraus, N. (2003). Integration of heard and seen speech: a factor in learning disabilities in children. *Neurosci Lett, 351*(1), 46-50. Retrieved from <Go to ISI>://MEDLINE:14550910

Heath, S. M., Hogben, J. H., & Clark, C. D. (1999). Auditory temporal processing in disabled readers with and without oral language delay. *J Child Psychol Psychiatry, 40*(4), 637-647. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10357169

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99. doi:10.1016/j.cognition.2003.10.011

Hill, N. I., Bailey, P. J., Griffiths, Y. M., & Snowling, M. J. (1999). Frequency acuity and binaural masking release in dyslexic listeners. *J Acoust Soc Am, 106*(6), L53-58. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10615710

Hochberg, J. (1998). Gestalt Theory and its Legacy: Organization in Eye and Brain, in Attention and Mental Representation. In J. Hochberg (Ed.), *Perception and Cognition at Century's End: History, Philosophy, Theory* (pp. 253-304): Elsevier Science.

Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. *Trends Cogn Sci, 13*(5), 216-221. doi:10.1016/j.tics.2009.02.005

Holmes, G. (1922). Clinical Symptoms of Cerebellar Disease and their Interpretation. *Lancet, 1*, 1177-1237.

Horne, J., W. (Writer). (1931). Beau Hunks.

Horne, T., Downes, G., & Wooley, B. (1979). Video killed the Radio Star. On *The Buggles: The Age of Plastic*. s.l.: Island Records Ltd.

Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain Lang, 89*(1), 235-242. doi:10.1016/S0093-934X(03)00401-2

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition, 91*(3), 273-296. doi:10.1016/j.cognition.2003.09.006

Iles, J., Walsh, V., & Richardson, A. (2000). Visual search performance in dyslexia. *Dyslexia, 6*(3), 163-177. doi:10.1002/1099-0909(200007/09)6:3<163::aid-dys150>3.0.co;2-u

Ito, M. (1984). *The Cerebellum and Neural Control*. New York: Raven Press.

Jordan, T. R., & Sergeant, P. (2000). Effects of distance on visual and audiovisual speech recognition. *Lang Speech*.

Kaltenbacher, T., Krotzer, E., Hummer, P., & Leyrer, M. (2012). *EPG Intervention in an adult cochlear implant user*. Paper presented at the 14th Conference of the International Clinical Phonetics and Linguistics Association 2012, Cork, Ireland.

Kaltenbacher, T., Pfleiderer, B., Bühner, B., Hummer, P., & Breninger, B. (forthcoming). *Hearing Lips and Seeing What? An fMRI study of the McGurk effect in Dyslexia. Forthcoming.*

Kaye, J. (1989). *Phonology: A Cognitive View*. Hillsdale: Lawrence Erlbaum Associates.

Kersting, M., & Althoff, K. (2004). *Rechtschreibungstest (RT)*. Göttingen: Hogrefe.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*(5582), 846-848. doi:10.1126/science.1070311

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science, 273*(5277), 971-973. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8688085

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*(4577), 1138-1141. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7146899

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606-608.

Kujala, T., & Naatanen, R. (2001). The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neurosci Biobehav Rev, 25*(6), 535-543. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11595273

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppanen, P. H., Lohvansuu, K., . . . Schulte-Korne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *J Child Psychol Psychiatry, 54*(6), 686-694. doi:10.1111/jcpp.12029

Landsgesell, M. (2011). *Blickpräferenz in der Gebärdensprache: eine empirische Studie am Eyetracker. .* Unpublished Master Thesis: University of Salzburg.

Levelt, W. J. M. (1999). Producing spoken language: a blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83-122). Oxford England ; New York: Oxford University Press.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporary disparate auditory and visual stimuli. *Cognitive Brain Research, 16*(3), 468-478. doi:Doi 10.1016/S0926-6410(03)00074-0

Liberman, A. M. (1998). Why is Speech so Much Easier than Reading and Writing? . In M. Joshi & C. Hulme (Eds.), *Reading and Spelling: Development and Disorder* (pp. 5-18). Hillsdale, NY: Erlbaum Associates.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4075760

Lobier, M., Zoubrinetzky, R., & Valdois, S. (2012). The visual attention span deficit in dyslexia is visual and not verbal. *Cortex, 48*(6), 768-773. doi:10.1016/j.cortex.2011.09.003

Lovegrove, W. J., Bowling, A., Badcock, D., & Blackwood, M. (1980). Specific reading disability: differences in contrast sensitivity as a function of spatial frequency. *Science, 210*(4468), 439-440.

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci, 4*(11), 1131-1138. doi:10.1038/nn737

Marshall, C. R., Ramus, F., & van der Lely, H. (2011). Do children with dyslexia and/or specific language impairment compensate for place assimilation? Insight into phonological grammar and representations. *Cogn Neuropsychol, 27*(7), 563-586. doi:10.1080/02643294.2011.588693

Martens, V. E., & de Jong, P. F. (2006). The effect of visual word features on the acquisition of orthographic knowledge. *J Exp Child Psychol, 93*(4), 337-356. doi:10.1016/j.jecp.2005.11.003

Massaro, D. W. (1988). Ambiguity in perception and experimentation. *J Exp Psychol Gen, 117*(4), 417-421. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2974865

Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*: MIT Press.

Massaro, D. W. (2012). Speech perception and reading: two parallel modes of understanding language and implications for acquiring literacy naturally. *Am J Psychol, 125*(3), 307-320. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22953690

Massaro, D. W., & Bosseler, A. (2006). Read my lips: The importance of the face in a computer-animated tutor for vocabulary learning by children with autism. *Autism, 10*(5), 495-510. doi:10.1177/1362361306066599

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *J Exp Psychol Hum Percept Perform, 9*(5), 753-771. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6227688

Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *J Exp Psychol Hum Percept Perform, 20*(6), 1107-1128. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7844509

Massaro, D. W., & Jesse, A. (2009). Audiovisual speech perception and word recognition. In G. Gaskell & G. Altmann (Eds.), *The Oxford Handbook of Psycholinguistics* (pp. 19-35). Oxford: Oxford University Press.

Massaro, D. W., & Stork, D. G. (1998). Speech Recognition and Sensory Integration. *Am Sci*(86), 236-244.

Masson, M. E., Dodd, M. D., & Enns, J. T. (2009). The bicycle illusion: sidewalk science informs the integration of motion and shape perception. *J Exp Psychol Hum Percept Perform, 35*(1), 133-145. doi:10.1037/0096-1523.35.1.133

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* (pp. 133-147): MIT Press.

Mayr, E. (2001). *What Evolution Is.* London: Weidenfeld & Nicolson.

McAnally, K. I., & Stein, J. F. (1996). Auditory temporal coding in dyslexia. *Proc Biol Sci, 263*(1373), 961-965. doi:10.1098/rspb.1996.0142

McArthur, G. M., & Hogben, J. H. (2001). Auditory backward recognition masking in children with a specific language impairment and children with a specific reading disability. *J Acoust Soc Am, 109*(3), 1092-1100. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11303923

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1012311

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature, 282*(5737), 403-404. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/503219

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 75-78. doi:10.1126/science.198.4312.75

Meltzoff, A. N., & Moore, M. K. (1979). Interpreting "imitative" responses in early infancy. *Science, 205*(4402), 217-219. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/451596

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Brain Res Cogn Brain Res, 14*(1), 31-40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12063128

Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *J Speech Lang Hear Res, 56*(1), 211-221. doi:10.1044/1092-4388(2012/11-0270)

Modha, G., Bernhardt, B. M., Church, R., & Bacsfalvi, P. (2008). Case study using ultrasound to treat /r. *Int J Lang Commun Disord, 43*(3), 323-329. doi:10.1080/13682820701449943

Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: auditory processing or phonological coding? *J Exp Child Psychol, 64*(2), 199-231. doi:10.1006/jecp.1996.2343

Mohammed, T., Campbell, R., Macsweeney, M., Barry, F., & Coleman, M. (2006). Speechreading and its association with reading among deaf, hearing and dyslexic individuals. *Clin Linguist Phon, 20*(7-8), 621-630. doi:doi:10.1080/02699200500266745

Moll, K., & Landerl, K. (2010). *SLRT II - Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT).* Berlin: Hans Huber Verlag.

Moosmueller, S., & Ringen, C. (2004). Vocie and Aspiration in Austrian German Plosives. *Folia Linguistica, 38*(1-2), 43-62. doi:10.1515/flin.2004.38.1-2.43

Morsella, E., & Bargh, J. A. (2011). Unconscious Action Tendencies: Sources of 'Un-Integrated' Action. In J. Decety & J. T. Cacioppo (Eds.), *Handbook of Social Neuroscience* (pp. 335-347). Oxford: OUP.

Morsella, E., Gray, J. R., Krieger, S. C., & Bargh, J. A. (2009). The essence of conscious conflict: subjective effects of sustaining incompatible intentions. *Emotion, 9*(5), 717-728. doi:10.1037/a0017121

Mottonen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res, 13*(3), 417-425. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11919005

Munhall, K. G., Kroos, C., Jozan, G., & Vatikiotis-Bateson, E. (2004). Spatial frequency requirements for audiovisual speech perception. *Percept Psychophys, 66*(4), 574-583. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15311657

Nagarajan, S., Mahncke, H., Salz, T., Tallal, P., Roberts, T., & Merzenich, M. M. (1999). Cortical auditory signal processing in poor readers. *Proc Natl Acad Sci U S A, 96*(11), 6483-6488. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10339614

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage, 59*(1), 781-787. doi:10.1016/j.neuroimage.2011.07.024

Nicolson, R. I., & Fawcett, A. (2008). *Dyslexia, Learning and the Brain.* Boston: MIT Press.

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: a new framework for dyslexia research? *Cognition, 35*(2), 159-182. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2354611

Nicolson, R. I., & Fawcett, A. J. (2008). Learning, Cognition and Dyslexia. In G. Reid, A. Fawcett, F. Manis, & L. Siegel (Eds.), *The SAGE Handbook of Dyslexia* (pp. 192-211): SAGE Publications.

Nicolson, R. I., Fawcett, A. J., & Dean, P. (1995). Time estimation deficits in developmental dyslexia: evidence of cerebellar involvement. *Proc Biol Sci, 259*(1354), 43-47. doi:10.1098/rspb.1995.0007

Nusbaum, H. C. (2011). Language and Communication. In J. Decety & J. T. Cacioppo (Eds.), *The Oxford Handbook of Social Neuroscience* (pp. 668-682): Oxford University Press.

Pammer, K., & Wheatley, C. (2001). Isolating the M(y)-cell response in dyslexia using the spatial frequency doubling illusion. *Vision research, 41*(16), 2139-2147.

Paul, I., Bott, C., Heim, S., Eulitz, C., & Elbert, T. (2006). Reduced hemispheric asymmetry of the auditory N260m in dyslexia. *Neuropsychologia, 44*(5), 785-794. doi:10.1016/j.neuropsychologia.2005.07.011

Pekkola, J., Laasonen, M., Ojanen, V., Autti, T., Jaaskelainen, I. P., Kujala, T., & Sams, M. (2006). Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T. *Neuroimage, 29*(3), 797-807. doi:10.1016/j.neuroimage.2005.09.069

Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport, 16*(2), 125-128. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15671860

Peyrin, C., Lallier, M., Demonet, J. F., Pernet, C., Baciu, M., Le Bas, J. F., & Valdois, S. (2012). Neural dissociation of phonological and visual attention span disorders in developmental dyslexia: FMRI evidence from two case reports. *Brain Lang, 120*(3), 381-394. doi:10.1016/j.bandl.2011.12.015

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2009). Speech perception at the interface of neurobiology and linguistics. In B. J. C. Moore, L. K. Tyler, & W. D. Marslen-Wilson (Eds.), *The Perception of Speech: from Sound to Meaning.* (pp. 250-274). Oxford: Oxford University Press.

Pugh, K. R., Sandak, R., Frost, S. J., Moore, D., Rueckl, J. G., & Mencl, W. E. (2006). Neurobiological Studies of Skilled and Impaired Reading: A Work in Progress. In G. D. Rosen (Ed.), *The Dyslexic Brain* (pp. 22-47). Mahway, New Jersey: Lawrence Erlbaum Associates, Publishers.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *Q J Exp Psychol, 26*(1), 63-71. doi:10.1080/14640747408400388

Ramus, F. (2001). Dyslexia. Talk of two theories. *Nature, 412*(6845), 393-395. doi:10.1038/35086683

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Curr Opin Neurobiol, 13*(2), 212-218. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12744976

Ramus, F. (2004). Neurobiology of dyslexia: a reinterpretation of the data. *Trends Neurosci, 27*(12), 720-726. doi:10.1016/j.tins.2004.10.004

Ramus, F. (2013). What's the point of neuropsychoanalysis? *Br J Psychiatry, 203*, 170-171. doi:10.1192/bjp.bp.113.127217

Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: the difficulties of interpreting poor performance, and the importance of normal performance. *Cogn Neuropsychol, 29*(1-2), 104-122. doi:10.1080/02643294.2012.677420

Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: towards a multidimensional model. *Brain, 136*(Pt 2), 630-645. doi:10.1093/brain/aws356

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain, 126*(Pt 4), 841-865. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12615643

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Q J Exp Psychol (Hove), 61*(1), 129-141. doi:10.1080/17470210701508822

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol Bull, 124*(3), 372-422. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9849112

Rayner, K., Juhasz, B., J., & Pollatsek, A. (2005). Eye Movements During Reading. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 79-98): Blackwell.

Rayner, K., Pollatsek, A., & Binder, K. S. (1998). Phonological codes and eye movements in reading. *J Exp Psychol Learn Mem Cogn, 24*(2), 476-497. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9530845

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *J Exp Child Psychol, 48*(2), 270-292. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2794857

Richlan, F., Kronbichler, M., & Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: a quantitative meta-analysis of neuroimaging studies. *Hum Brain Mapp, 30*(10), 3299-3308. doi:10.1002/hbm.20752

Rizzolatti, G., & Arbib, M. A. (1999). From grasping to speech: imitation might provide a missing link: reply. *Trends Neurosci, 22*(4), 152. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10203851

Robertson, E. K., Joanisse, M. F., Desroches, A. S., & Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Dev Sci, 12*(5), 753-767. doi:10.1111/j.1467-7687.2009.00806.x

Saalasti, S., Tiippana, K., Katsyri, J., & Sams, M. (2011). The effect of visual spatial attention on audiovisual speech perception in adults with Asperger syndrome. *Experimental brain research Experimentelle*

Hirnforschung Experimentation cerebrale, 213(2-3), 283-290. Retrieved from <Go to ISI>://MEDLINE:21660467
Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett, 127*(1), 141-145. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1881611
Sams, M., & Hari, R. (1991). Magnetoencephalography in the study of human auditory information processing. *Ann N Y Acad Sci, 620*, 102-117. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2035937
Sams, M., Kaukoranta, E., Hamalainen, M., & Naatanen, R. (1991). Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. *Psychophysiology, 28*(1), 21-29. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1886961
Sams, M., Mottonen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Res Cogn Brain Res, 23*(2-3), 429-435. doi:10.1016/j.cogbrainres.2004.11.006
Schulte-Korne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998a). Auditory processing and dyslexia: evidence for a specific speech processing deficit. *Neuroreport, 9*(2), 337-340. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9507979
Schulte-Korne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998b). Role of auditory temporal processing for reading and spelling disability. *Percept Mot Skills, 86*(3 Pt 1), 1043-1047. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9656305
Schurz, M., Kronbichler, M., Crone, J., Richlan, F., Klackl, J., & Wimmer, H. (2013). Top-down and bottom-up influences on the left ventral occipito-temporal cortex during visual word recognition: An analysis of effective connectivity. *Hum Brain Mapp*. doi:10.1002/hbm.22281
Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Dev Sci, 11*(2), 306-320. doi:10.1111/j.1467-7687.2008.00677.x
Serniclaes, W., Van Heghe, S., Mousty, P., Carre, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *J Exp Child Psychol, 87*(4), 336-361. doi:10.1016/j.jecp.2004.02.001
Share, D., Jorm, A., Maclean, R., & Matthews, R. (2002). Temporal processing and reading disability. *Read Writ, 15*(1-2), 151-178. doi:10.1023/A:1013876606178
Shaywitz, B. A., Lyon, G. R., & Shaywitz, S. E. (2006). The role of functional magnetic resonance imaging in understanding reading and dyslexia. *Dev Neuropsychol, 30*(1), 613-632. doi:10.1207/s15326942dn3001_5

Shaywitz, S. E. (1998). Dyslexia. *New England Journal of Medicine, 338*(5), 307-312. doi:doi:10.1056/NEJM199801293380507
Shaywitz, S. E., Shaywitz, B. A., Pugh, K. R., Fulbright, R. K., Constable, R. T., Mencl, W. E., . . . Gore, J. C. (1998). Functional disruption in the organization of the brain for reading in dyslexia. *Proceedings of the National Academy of Sciences, 95*(5), 2636-2641. Retrieved from http://www.pnas.org/content/95/5/2636.abstract
SHURE. (2009-2014). PG48 Dynamisches Mikrofon. Retrieved from http://www.shure.de/produkte/mikrofone/pg48
Silani, G., Frith, U., Demonet, J. F., Fazio, F., Perani, D., Price, C., . . . Paulesu, E. (2005). Brain abnormalities underlying altered activation in dyslexia: a voxel based morphometry study. *Brain, 128*(Pt 10), 2453-2461. doi:10.1093/brain/awh579
Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage, 25*(1), 76-89. doi:10.1016/j.neuroimage.2004.11.006
Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex, 17*(10), 2387-2399. doi:10.1093/cercor/bhl147
Snowling, M., Bishop, D. V., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *J Child Psychol Psychiatry, 41*(5), 587-600. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10946751
Snowling, M., & Stackhouse, J. (2006). *Dyslexia: Speech and Language. A Practionier's Handbook.* (2nd ed.). London: Whurr.
Snowling, M. J. (2006). Language skills and learning to read: the dyslexia spectrum. In M. J. Snowling & J. Stackhouse (Eds.), *Dyslexia: Speech and Language. A Practionier's Handbook.* (2nd ed., pp. 1-14). London: Whurr.
Snowling, M. J., Goulandris, N., & Stackhouse, J. (1994). Phonological constraints on learning to read: evidence from single-case studies of reading difficulty. In C. Hulme & M. J. Snowling (Eds.), *Reading development and dyslexia* (pp. 86-104). London: Whurr.
Snowling, M. J., & Hulme, C. (2005). *The Science of Reading: A Handbook*: Blackwell.
Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition, 92*(3), B13-23. doi:10.1016/j.cognition.2003.10.005
Sprenger-Charolles, L., Colé, P., & Serniclaes, W. (2006). *Reading Acquisition and Developmental Dyslexia*. Hove: Psychology Press.
SR-Research. (2011a). Data Viewer Version 1.11.1: SR Research.

SR-Research. (2011b). Experiment Builder Version 1.10.165: SR Research.
Stackhouse, J. (2006). Speech and spelling difficulties: what to look for. In M. J. Snowling & J. Stackhouse (Eds.), *Dyslexia: Speech and Language. A Practionier's Handbook.* (2nd ed., pp. 15-35). London: Whurr.
Stein, B. E., & Meredith, M. A. (1992). *The Merging of the Senses*. Cambridge: MIT Press.
Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia, 7*(1), 12-36. doi:10.1002/dys.186
Stein, J. (2008). The Neurobiological Basis of Dyslexia. In G. Reid, A. Fawcett, F. Manis, & L. Siegel (Eds.), *The SAGE Handbook of Dyslexia* (pp. 53-76): SAGE Publications.
Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *J Acoust Soc Am, 111*(4), 1872-1891. doi:doi:http://dx.doi.org/10.1121/1.1458026
Stevens, K. N., & Halle, M. K. (1967). Remarks on analysis-by-synthesis and distinctive features. In W. Wathen-Dunn (Ed.), *Models for the Perception of Speech and Visual Form: Proceedings of a Symposium* (pp. 88-102). Cambridge: MIT Press.
Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *J Acoust Soc Am, 26*(2), 212-215. Retrieved from http://dx.doi.org/10.1121/1.1907309
Szenkovits, G., & Ramus, F. (2005). Exploring dyslexics' phonological deficit I: lexical vs sub-lexical and input vs output processes. *Dyslexia, 11*(4), 253-268. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16355747
Szycik, G. R., Stadler, J., Tempelmann, C., & Munte, T. F. (2012). Examining the McGurk illusion using high-field 7 Tesla functional MRI. *Front Hum Neurosci, 6*, 95. doi:10.3389/fnhum.2012.00095
Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang, 9*(2), 182-198. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7363063
Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain Lang, 9*(2), 182-198. doi:http://dx.doi.org/10.1016/0093-934X(80)90139-X
Tallal, P. (2006). What happens when 'dyslexic' subjects do not meet the criteria for dyslexia and sensorimotor tasks are too difficult even for the controls? *Dev Sci, 9*(3), 262-264. doi:Doi 10.1111/J.1467-7687.2006.00487.X
Tiippana, K., Andersen, T., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology, 16*, 457-472.
Tiippana, K., Puharinen, H., Mottonen, R., & Sams, M. (2011). Sound location can influence audiovisual speech perception when spatial attention is

manipulated. *Seeing and perceiving, 24*(1), 67-90. Retrieved from <Go to ISI>://MEDLINE:21406157

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition, 96*(1), B13-22. Retrieved from <Go to ISI>://MEDLINE:15833302

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition, 92*(1-2), 231-270. doi:10.1016/j.cognition.2003.10.008

Universität Trier. (1985). *Bibliographie zur Legasthenie: Literatur aus den deutschsprachigen Ländern; Spezialbibliographie*: Zentralstelle für psychologische Information und Dokumentation an der Universität Trier.

Valdois, S., Bidet-Ildei, C., Lassus-Sangosse, D., Reilhac, C., N'Guyen-Morel, M. A., Guinet, E., & Orliaguet, J. P. (2011). A visual processing but no phonological disorder in a child with mixed dyslexia. *Cortex, 47*(10), 1197-1218. doi:10.1016/j.cortex.2011.05.011

Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: review of evidence for a selective visual attentional disorder. *Dyslexia, 10*(4), 339-363. doi:10.1002/dys.284

Valdois, S., Lassus-Sangosse, D., & Lobier, M. (2012). Impaired letter-string processing in developmental dyslexia: what visual-to-phonology code mapping disorder? *Dyslexia, 18*(2), 77-93. doi:10.1002/dys.1437

van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2012). Lexical decision in children: sublexical processing or lexical search? *Q J Exp Psychol (Hove), 65*(6), 1214-1228. doi:10.1080/17470218.2011.652136

van Leeuwen, T., Been, P., van Herten, M., Zwarts, F., Maassen, B., & van der Leij, A. (2008). Two-month-old infants at risk for dyslexia do not discriminate /bAk/ from /dAk/: A brain-mapping study. *J Neurolinguistics, 21*(4), 333-348. doi:http://dx.doi.org/10.1016/j.jneuroling.2007.07.004

Van Orden, G. C., & Kloos, H. (2005). The Question of Phonology and Reading. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 61-78): Blackwell Publishing.

van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philos Trans R Soc Lond B Biol Sci, 364*(1525), 1815-1830. doi:10.1098/rstb.2009.0023

van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of subjective time perception within and across senses. *PLoS One, 3*(1), e1437. doi:10.1371/journal.pone.0001437 van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A, 102*(4), 1181-1186. doi:10.1073/pnas.0408949102

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607. doi:10.1016/j.neuropsychologia.2006.01.001

van Wassenhove, V., & Nagarajan, S. S. (2007). Auditory cortical plasticity in learning to discriminate modulation rate. *J Neurosci, 27*(10), 2663-2672. doi:10.1523/JNEUROSCI.4844-06.2007

Vellutino, F. R. (1981). *Dyslexia: Theory and Research*: Mit Press.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *J Child Psychol Psychiatry, 45*(1), 2-40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14959801

Vidyasagar, T. R. (2004). Neural underpinnings of dyslexia as a disorder of visuo-spatial attention. *Clin Exp Optom, 87*(1), 4-10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14720113

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: a deficit in visuo-spatial attention, not in phonological processing. *Trends Cogn Sci, 14*(2), 57-63. doi:10.1016/j.tics.2009.12.003

Widerin, B. (2007). *Speechreading bei Dyslexie (Lippenlesen bei auditiver Maskierung) : eine empirische Studie am Eyetracker.* University of Salzburg: Unpublished Master Thesis.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat Neurosci, 7*(7), 701-702. doi:10.1038/nn1263

Wimmer, H. (2013). *Dyslexia in regular orthographies: manifestation, cognition, brain functions*. Paper presented at the 3rd Oxford-Kobe Symposium: The Neurobiologigal Basis of Dyslexia in Alphabetic, Syllabic and Logographic Scripts, St. Catherine's College, Oxford, UK.

Wimmer, H., Schurz, M., Sturm, D., Richlan, F., Klackl, J., Kronbichler, M., & Ladurner, G. (2010). A dual-route perspective on poor reading in a regular orthography: an fMRI study. *Cortex, 46*(10), 1284-1298. doi:10.1016/j.cortex.2010.06.004

Wolf, M., & O'Brien, B. (2006). From the Sumerians to Images of the Reading Brain: Insights for Reading Theory and Intervention. In G. D. Rosen (Ed.), *The Dyslexic Brain* (pp. 5-20). Mahway, New Jersey: Lawrence Erlbaum Associates, Publishers.

World Health Organization. (2008). *The International Statistical Classification of Diseases and Related Health Problems 2008: ICD-10*: World Health Organization.

Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics, 30*(3), 555-568. doi:http://dx.doi.org/10.1006/jpho.2002.0165

Zeki, S., & Bartels, A. (1999). Toward a theory of visual consciousness. *Conscious Cogn, 8*(2), 225-259. doi:10.1006/ccog.1999.0390