Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners

The Whole is More than the Sum of its Parts

Dörte Hessler

university of groningen
faculty of arts
CLCG

The work reported in this thesis has been carried out under the auspices of the School of Behavioral and Cognitive Neurosciences (BCN) and the Center for Language and Cognition Groningen (CLCG). Publication of this thesis was financially supported by the University of Groningen, the Stichting Afasie Nederland (SAN) and BCN.

Groningen Dissertations in Linguistics 98
ISSN 0928-0030
ISBN 978-90-367-5224-4

Cover Design based on work from Martin Prout
(http://www.contextfreeart.org/gallery/view.php?id=2675)
Printed by Wöhrmann Print Service, Zutphen, The Netherlands
Document prepared with LaTeX 2ε and typeset in pdfTeX.

RIJKSUNIVERSITEIT GRONINGEN

Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners

The Whole is More than the Sum of its Parts

Proefschrift

ter verkrijging van het doctoraat in de
Letteren
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
donderdag 15 december 2011
om 16:15 uur

door

Dörte Annette Heßler

geboren op 14 oktober 1981
te Cuxhaven, Duitsland

Promotor:	Prof. dr. Y.R.M. Bastiaanse
Copromotor:	Dr. R. Jonkers
Beoordelingscommissie:	Prof. dr. R. De Bleser Prof. dr. P. Mariën Prof. dr. W. Ziegler

ISBN: 978-90-367-5224-4

Acknowledgments

First and foremost, I want to thank Roel Jonkers and Roelien Bastiaanse for their help and guidance throughout my PhD project. Roelien was the one who believed in my plans and made sure I could come to Groningen in the first place. As my promotor, she played an immense role during the complete project. Although she was not my daily supervisor, she was always interested in my work and read abstracts, papers, chapters, and finally the thesis fast and thoroughly. Roelien, thank you for your valuable advise and your trust!

Roel was an unmissable member of the team. In weekly meetings we discussed the ups and downs of the project and together we always found a way to handle things. Furthermore, he also lend his face and voice for my stimuli. Besides that, Roel also always helped me to put things into perspective. Therefore, we supplemented our discussions with the topics soccer (both active and passive), vacation plans, and other seemingly unimportant stuff. Thus, I want to thank Roel for being the best supervisor and co-promoter one could wish for, not only on a professional, but also on a personal level.

Ria De Bleser, Peter Mariën, and Wolfram Ziegler agreed to be in my reading committee. I want to thank them for their willingness to read my thesis, attend my defense, and present at the accompanying workshop.

This thesis would consist only of the theoretical parts if it hadn't been for the people who participated in my research studies (some even in more than one) and those who helped to carry them out. I am very grateful for their commitment. I also want to thank the revalidation centers Beatrixoord and Revalidatie Friesland for their cooperation.

Laurie I would like to thank for her help with the ERP experiments. When

I started this project, I planned ERP research without knowing anything practical about it. Thanks to Laurie and the colleagues at the NIC, this has quite changed during the last four years.

Research is not possible without the colleagues who listen to new ideas and help making sense out of the data. Through the years these were various people and I would like to thank all of them: my fellow PhD candidates Jantien, Eleonora, Julia, Tuba, Maria, Katrien, Marlous, Hayo, Harwintha, Tom, Laura, Rimke, and Ellie; the visitors Marie, Vasiliki, Nicky, and Olga; the research assistants Femke and Sanne and of course the "regular colleagues" Roelien, Roel, Laurie, Gerard, and Ben.

Furthermore, there has always been excellent support from various people. The audiovisual service department, especially Callista, helped tremendously with the recording and editing of the stimuli. Wyke was the first person I met when I came here and she made sure I'd find my way around easily. Throughout the years she has always been there for questions and support. The staff at the secretariat of the cluster Nederlands always had an open door and ear for the more practical issues, thanks for that.

The educational part of my PhD program was taken care of by BCN. I followed a lot of their courses and learned quite a bit, especially about neighboring disciplines. Furthermore, I want to thank BCN for providing funding, allowing me to visit additional conferences. I also want to thank BCN and the Stichting Afasie Nederland (SAN) for their contributions to the printing costs of this thesis.

But life is not only about work and during my time here I had various roommates, who all made coming to the office something to look forward to. Therefore I would like to thank Rasmus, Marlous, Jacky, Proscovia, Harwintha, Barbara, Jelena, Çağrı, Valerio, and Charlotte for their ability to keep a good working atmosphere with just the correct amount of breaks for chatting.

During most of my PhD-time I had my office on the corridor of "Alfa-Informatica". I felt very much at home there and want to thank all previous and current "inhabitants" for adopting me as part of their group. They shared their cakes with me, took me to lunch and group outings, and finally even gave me a small position to bridge the time until my new "real" job started. A huge "thank you" to all of you for being so "gezellig".

Several colleagues also had a big contribution to my social life, which really started when I joined the "'AI-team" in the weekly pub-quiz in early 2008.

Around the same time I also joined the CLCG sports group in which an uncountable number of people participated. I would like to thank all of them for making sure we could continue every year.

In the meanwhile I joined another sports-group: the SPR Zumba. I really enjoy it and appreciate everybody who also joined, as it is much more fun with a larger group. Rimke, Marieke, Laura, Audrey, and Kashy: thanks for making sure that I stay fit!

GRASP, which is now split in the two organizations GRIN and Gopher, also had a very important role in my PhD-life. Especially during the first two years I went to almost every activity they organized and I thankfully remember all of them. Together with the foreign guest club, they made sure I saw more of the Netherlands than the Harmony building.

During my time as a PhD candidate, I got more and more interested in "the other side of research", the organizational and representational tasks behind the scenes. At that time there was no proper PhD representation in our faculty. I want to thank Eva, Simone, Theisje, Frigga, and Nynke for establishing the PhD council of the Graduate School for Humanities together with me. Rudolf and Rimke filled in when Eva and I left. Thanks to them and Corien, who joined after I left, the work is still kept up. I'm unbelievably happy with how everything progressed. As part of the PhD council I also attended regular meetings with representatives from other councils and I want to thank them for the fruitful discussions we had in those "OPO meetings".

Organizing the 30th TABU Dag together with Diana, Martijn, Myrte, and Alexandra was another valuable experience. I really enjoyed the time we spent together and am still very proud of the result. Furthermore, all the experiences we made in organizing an event like this, made it much easier for me to organize smaller workshops later on. Thank you for being such an efficient team and for making it fun as well.

During the last year Barbara, Çağrı, Peter, and I formed a group to monitor our process in writing, discuss common issues and just have a chat with people in the same phase of their PhD. Thanks Barbara, Çağrı, and Peter for doing this together with me and thanks to the "schildpad group" for being an inspiration.

Of course there was a time before coming to Groningen... My interest in research was really developed while working on my Diploma thesis in Potsdam. Nicole Stadie had a huge impact on that. Not only did she make sure I carried out my research properly, but she also sent me to two conferences to present the results. I enjoyed the conferences a lot, met other researchers, and decided that I wanted to stay in academia, at least for a while. Nicole and Ria De Bleser supported me with that and so I finally got the position in Groningen.

Going to Groningen was hard in the beginning, as I left a lot of dear ones behind. Anna, Annemarie, Daniela, Francis, Franziska, Manuela, Theresa, Yvonne, and I formed first the PIP-, then the PIDA-, and finally the PIFES-group, more or less regular gettogethers to discuss our studies, theses, and life in general. And still when I visit Potsdam, we try to get as many of us together for dinner or brunch, usually accompanied by partners and (in the meanwhile) a lot of children. I am very happy to have you as friends!

I would like to thank Kashy for proofreading my thesis. If you find any errors left, I probably ignored her advice or changed things after she was done. So please, do not blame her! Gideon helped with the layout of the cover of this thesis. I am very happy to have someone with a great eye for design in my extended family. Thanks for your advice!

Of course, I cannot forget to thank Rimke and Kostadin for being my paranimfs. I am very happy to have them at my side during the defense, but of course they already helped before that! With the amount of things that need to be taken care of prior to a PhD defense (correcting the Dutch summary, making sure everybody gets the books, that there is a reception and a party, buying a nice outfit etc), it's good to have friends that help, so thank you Rimke and Kostadin for everything!

There are only a few people left to thank (except for all the ones I forgot) and those are the most import ones: First and foremost my parents. Not only did they support me during my studies, but also going "home" always really feels like coming home. Of course also my sister and her family contribute to that. As you cannot choose your family, I must say that I am very grateful for my luck in getting the one I have. Thank you all for your support and love.

I also want to thank my "in-laws-to-be", Luit and Gertie, for making me feel welcome since the first time we met and of course for rescheduling the yearly trip to Australia so they can be at my defense.

Finally, I want to thank Daniël. I could write pages about the reasons to be thankful, but I'll keep it short: Thank you for being yourself. I could not possibly imagine a better partner than you.

Groningen, November 1, 2011

Contents

List of Figures

List of Tables

Chapters Based on Submitted and Accepted Peer Reviewed Publications

Chapter 2

Hessler, D., Jonkers, R. and Bastiaanse, R. (2010). The influence of phonetic dimensions on aphasic speech perception. *Clinical Linguistics and Phonetics* 24, 980-996.

Chapter 3

Hessler, D., Jonkers, R. and Bastiaanse, R. (2011). Processing of audiovisual stimuli in aphasic and non-brain-damaged listeners. *Aphasiology*. Advance online publication. doi:10.1080/02687038.2011.608840

Chapter 5

Hessler, D., Jonkers, R., Stowe, L. and Bastiaanse, R. (subm.). When the whole is more than the sum of its parts - audiovisual processing of phonemes investigated with ERPs. *Brain and Language*.

Prolegomenon

Comprehension of spoken language is achieved without much effort for most of us. It is an automatic process, consisting of many steps, the complexity of which is only realized when comprehension fails. This is often the case for individuals who have suffered from brain damage resulting in aphasia. Only if it is disturbed, all the steps necessary for the comprehension of a single word become evident: starting from the filtering and analysis of speech sounds out of the incoming sound stream, followed by the short term storage, the identification and recognition of lexical items, and the retrieval of a word's meaning.

Often, when studying or describing this process, the focus lies merely on auditory information, disregarding another form of information usually available to listeners: the information gained from the articulatory movements via speechreading.[1]

It has been argued previously that the possibility to see a speaker's face helps to understand speech when there are adverse conditions, for example several people speaking at the same time or background noise (Sumby & Pol-

[1] Terminology is not consistent in the literature and both 'speechreading' and 'lipreading' have been used. In this thesis the term 'speechreading' is used because the visual input received is not restricted to the lips, but rather covers the lower face, neck, and the shoulders. This terminology has been suggested by Campbell, Dodd, and Burnham (1998) in order to state clearly that more than just lip information is taken into account and to stress that what is read is indeed natural speech.

lack, 1954). Information from speechreading is, however, not only integrated in adverse conditions, but also when the auditory input is clear. Thus, seen speech seems to play a role during comprehension. This has been shown by experiments conducted by McGurk and MacDonald (1976), which led to the discovery of the famous McGurk effect: dubbed incongruent auditory and visual information (e.g. auditory /ba/ dubbed on visual /ga/) are integrated and yet another syllable, comprising features of both input syllables is perceived (e.g. /da/). Speechreading is also known to have a beneficial effect on the comprehension abilities of aphasic participants suffering from a speech perception disorder (cf. Buchman, Garron, Trost-Cardamone, Wichter, & Schwartz, 1986; Shindo, Kaga, & Tanaka, 1991).

Models of both auditory and audiovisual speech perception have been postulated. In these models, not speech sounds (phonemes) are the basic unit of processing, but rather smaller entities, phonetic features, are assumed. In this thesis, four experimental studies are described, which investigate several aspects of auditory and audiovisual speech perception by non-brain-damaged and aphasic listeners. Within those chapters three main issues are addressed with different methodologies:

(1) The first issue addressed in the current thesis is the processing of phonemes and their contrasts. Special attention will be paid to the role of phonetic features in processing. It will be investigated whether processing depends on the phonetic feature distinguishing phonemes, that means whether the theoretically assumed differences between features can be supported by behavioral and neurophysiological data. The processing accuracy for various features will be compared for a group of aphasic listeners in order to find out whether all features are equally difficult to process. Furthermore, an attempt will be made to determine the level of the deficit in aphasia: it will be investigated whether automatic processing (representing the competence level) is affected or whether the deficits are limited to later, performance related processing levels.

(2) A second focus is on the influence of speechreading on comprehension. The way the described advantage of audiovisual speech is manifested will be studied: whether all phonetic dimensions benefit equally or whether the advantage is due to a specific dimension. Another aim is to find out whether audiovisual speech perception is more than a mere addition of auditory and visual information.

(3) Thirdly, several predictions made on the basis of the TRACE model of audiovisual speech perception (Campbell, 1988, 1990) will be evaluated. This model forms the basis of the studies carried out. It will be discussed whether it can explain all findings of the conducted behavioral and neurophysiological experiments and if not, how it needs to be extended.

Below, an overview of the chapters and their contribution to the main research focus of this thesis is given.

Chapter 1 provides background information on auditory and audiovisual speech perception. First, the importance of phonetic features is discussed. These subphonemic entities play an important role in the current studies and in models of speech perception. Two of these models will be introduced: the Logogen model (e.g. Morton, 1969; Howard & Franklin, 1988) and the TRACE model (McClelland & Elman, 1986). The Logogen model serves as a basis to explain auditory comprehension deficits in aphasia. The TRACE model provides more details about speech perception, and is, hence, the basis of the experimental studies. As speech perception is not only based on auditory information, but also on seen speech, chapter 1 also discusses the evidence in favor of multimodality in speech perception, such as the McGurk effect (McGurk & MacDonald, 1976). Moreover, four types of models explaining audiovisual perception will be introduced. Finally, an overview of studies on audiovisual perception in aphasia is provided.

Chapter 2 is the first experimental chapter. It describes a study on auditory and audiovisual processing of the three phonetic dimensions in aphasic listeners. In this study, several issues regarding the perception of phonemes are addressed within a discrimination of nonwords paradigm. Pairs of nonwords were presented auditorily or audiovisually to two groups of participants: a group of non-brain-damaged listeners and a group of aphasic participants with a deficit in speech sound perception. The results of both groups will be compared to establish whether the aphasic listeners show an impairment on this task. It will be evaluated with regards to which kind of phonetic contrasts (number and type of phonetic dimensions differing) participants have most difficulties. Furthermore, the influence of speechreading will be investigated, not only for the overall performance, but also per phonetic dimension in order to find out whether improvement is limited to one specific dimension and if so to which.

In the second experimental chapter, chapter 3, a study on audiovisual perception and the McGurk effect is presented. A nonword identification task is described with four conditions: auditory, visual (seen speech), audiovisual and McGurk type stimuli. Three aphasic and fourteen non-brain-damaged listeners participated. Their results will be discussed on the basis of accuracy and reaction times. Next to a comparison of each of the aphasic participants with the control group, another focus will lie on the evaluation of the benefits of audiovisual processing. Finally, the occurrence of the McGurk effect and the reaction times associated with various answer types will be discussed in order to highlight potential differences in audiovisual integration strategies between aphasic and non-brain-damaged listeners.

In chapter 4, the event-related potentials technology (ERP) is introduced. The following two experiments make use of this method. Before discussing the studies in chapters 5 and 6, a background on electroencephalograms (EEGs) and event-related potentials is given. In this overview, it is explained how ERPs are derived from the ongoing EEG signal and how the placement of electrodes is standardized. Furthermore, three ERP components that are found in active oddball designs are discussed: the mismatch negativity (MMN), the N2b, and the P3. While the MMN is related to automatic processing, the N2b and the P3 are associated with conscious, task-relevant detection of a mismatch. This chapter, also, discusses studies that made use of the MMN to investigate aphasic auditory processing and audiovisual perception in non-brain-damaged listeners.

Chapter 5 describes an ERP study with thirteen non-brain-damaged participants. An active oddball task was administered while the EEG was recorded. Participants were asked to push a button whenever they encountered a deviant stimulus in a sequence of repeating 'standard' stimuli. This experiment was split in four sub-experiments with the presentation modalities 'pure tones', 'auditory syllables', 'visual (speechreading) syllables', and 'audiovisual syllables'. Two deviants were used in the first three sub-experiments. In the audiovisual sub-experiment, an additional McGurk type deviant was introduced. While the results from the pure tones and visual syllables sub-experiments serve as a baseline measure, the research questions focus on the other conditions. The aim of this study is to find out whether different phonemic contrasts also elicit distinct activation patterns, such that neural processing resembles the differences in subphonemic entities that are assumed by models of speech perception. A second issue addressed in this study is an evaluation of the neural differences between auditory and audiovisual perception. While it has been shown behaviorally that speechreading aids comprehension, this chapter addresses the brain correlates of this beneficial effect. In order to deepen the insight into audiovisual integration, not only congruent but also incongruent McGurk type stimuli form part of the analysis.

In chapter 6, the experiment described in chapter 5 is conducted with aphasic participants. First it will be established, based on the 'pure tone' results, whether the paradigm as it was used for non-brain-damaged listeners is reliable for aphasic participants. The activation patterns for the 'pure tones' differed substantially from those of the non-brain-damaged participants, hindering the analysis of the brain responses to speech stimuli. The results to the speech and speechreading related conditions will, nonetheless, be discussed shortly on the basis of visual inspection.

This thesis is concluded, in chapter 7, with a general discussion of all findings. In this last chapter, the three main points which were raised in the beginning of this prolegomenon are addressed again. Evidence from the four experimental studies is combined to discuss the outcomes concerning those issues and to come to final conclusions.

CHAPTER 1

Auditory and Audiovisual Speech Perception

Comprehending spoken language is a complex task, consisting of a number of processing steps. Various models have been proposed to describe language comprehension. This chapter will discuss how processing is described in terms of the Logogen model (e.g. Morton, 1969; Howard & Franklin, 1988). The aphasia diagnostic battery "Psycholinguistic Assessments for Language Processing in Aphasia (PALPA)" (Kay, Lesser, & Coltheart, 1992) is based on this model. Therefore it forms the basis of the following description of aphasic disorders of speech perception. For the current studies it is however necessary to describe speech perception in more detail than is done in the Logogen model. The TRACE model (McClelland & Elman, 1986), which will be discussed subsequently, provides more details about the pre-lexical steps of processing than the Logogen model and is therefore of particular interest to the current study of phoneme perception.

Speech perception is, however, a multimodal process using not only auditory but also visual information (Rosenblum, 2008). Therefore this chapter will also focus on the multimodality of speech perception, discussing evidence for this claim, introducing models describing multimodal processing and finally addressing audiovisual processing in aphasia.

1.1 Auditory speech perception

As mentioned above, two models of single word processing will be discussed in this section: the Logogen model (e.g. Morton, 1969; Howard & Franklin, 1988) and the TRACE model (McClelland & Elman, 1986). Both models describe the processing of phonemes and words. Phonemes are speech sounds that are employed to form meaningful contrasts between words. They are built by phonetic features of three distinct dimensions. The importance of phonetic dimensions is therefore discussed first, before describing the models in more detail.

1.1.1 Phonetic dimensions

Phonetic features define the different phonemes of a language (Chomsky & Halle, 1968). These features can be categorized into the three (distinctive) phonetic dimensions 'place of articulation', 'manner of articulation', and 'voicing'. A combination of these dimensions uniquely identifies each phoneme. Changes in one phonetic dimension will lead to a different phoneme: changing, for example, the 'place of articulation' from bilabial to alveolar would transform a /p/ into a /t/. But also broader contrasts between two phonemes (changes in two or three of the phonetic dimensions) are possible: /p/ and /z/ for example are distinguished by 'voicing', 'place of articulation', and 'manner of articulation'. It is therefore not sufficient to classify two words or syllables as different in one phoneme, as this distinction can be formed by differences in one, two or all three distinctive dimensions. Rather the number and type of the phonetic dimensions deviating should be mentioned as well.

The phonetic characteristics of the dimension 'voicing' are manifested differently across languages: the distinction 'voiced' versus 'voiceless' is made by differences in voice onset time (VOT). According to Lisker and Abramson (1964) the voicing distinction in English is achieved by contrasting the onset of voicing at the release of the lips (for /b/) with an onset of voicing up to 100ms later (/p/). Therefore the distinction in English is rather one between voiceless and voiceless-aspirated. For other languages as Dutch and Hungarian however, Lisker and Abramson (1964) found that the voiceless /p/ was produced by having the onset of voicing aligned with the release of the lips, while the voiced counterpart /b/ was produced with a voice onset at least 50ms prior to the lip-release.

A distinction in 'voicing' therefore refers to different phonetic distinctions in different languages, which is important when comparing English and Dutch data. In Dutch and Hungarian, however, the distinction between 'voiced' and 'voiceless' is accomplished by similar VOT patterns.

1.1.2 Logogen model

The Logogen Model (e.g. Morton, 1969; Howard & Franklin, 1988) describes single word comprehension and production in the oral and written modality. This means that understanding, speaking, reading and writing single words are covered by the model. As this thesis focuses on auditory language perception and its disorders, only this part of the model will be described here. The Logogen model consists of different modules, which all fulfill specific tasks. In Howard and Franklin's (1988) version of the Logogen model the modules involved in language comprehension are the 'auditory analysis', the 'auditory input buffer', the 'phonological input lexicon', and the 'semantic system'. These modules are connected by routes as shown in Figure 1.1.

Auditory analysis

The auditory analysis system is responsible for the recognition and analysis of incoming speech. This is done by filtering the speech sounds out of the background noise, for example enabling a listener to comprehend speech while listening to music. The system needs to account for the perception of various input types, such as different voices and even accents. Therefore an abstract unit is needed in the recognition process. The input is analyzed and categorized into phonemes by extracting phonetic features out of the incoming speech stream. By combining the (distinctive) phonetic features the phonemes can be recognized, discriminated and identified.

When hearing, for example, the word 'pea', the auditory analysis thus filters the speech sounds from the background noises and identifies the phonemes /p/ and /i:/ by extracting the phonetic features [plosive], [bilabial], and [voiceless] for /p/ followed by the features [high], [front], [unrounded] for /i:/.

Auditory input buffer

The auditory input buffer is an auditory short term memory system. It holds on to the segments that result from the auditory analysis and their ordering. The input is then matched upon the syllable structure of the relevant language.

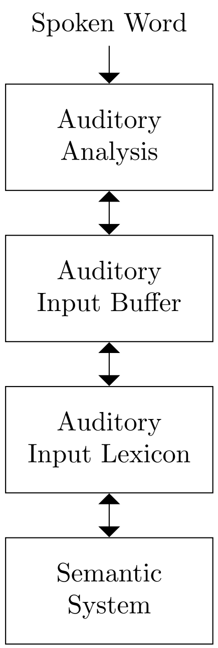


Figure 1.1: Schematic outline of single word comprehension (adapted from Howard and Franklin, 1988).

The buffer holds on to the elements that have been analyzed until they are recognized by the following module, the auditory input lexicon.

In the example of the word ‘pea’, the buffer stores the output of the auditory analysis, the phonemes /p/ and /i:/, and their ordering. Also it is assigned the syllable structure ‘CV’. The resulting entry /pi:/ is stored in the buffer until the next processing step, the recognition in the auditory input lexicon, is completed.

Auditory input lexicon

The auditory input lexicon is a long term memory system storing word forms (but not their meanings). In the auditory input lexicon, the incoming sound stream is compared with the stored entries. At this point in processing, words can be distinguished from nonwords, as the latter do not have an entry in the lexicon. It is assumed that phonologically related words are stored closely together, resulting in co-activation for related words: the more incoming sound information is processed the higher the activation level of the target raises while the activation for phonologically related competitors decreases. Ultimately, the activation of the competitors diminishes and the target is recognized by reaching threshold. During this process, there is a constant exchange of information between the auditory input lexicon and the auditory input buffer.

The example word ‘pea’ is recognized as a word at this point in processing. The input received from the auditory input buffer (/pi:/) activates, next to the entry of ‘pea’, also entries of words that are phonologically related, such as ‘bee’, ‘glee’, ‘pie’ and ‘peek’. However, only the target ‘pea’ receives enough activation to reach the necessary threshold to be selected. The matching meaning for the word form ‘pea’ will be looked up in the next module, the semantic system.

Semantic system

Once the phonological representation is activated in the auditory input lexicon, the identified word is forwarded to the semantic system. This is a long term storage for the meaning of words. It is not clear whether one semantic system (as proposed by Howard and Franklin (1988) and Ellis and Young (1988), among others) is sufficient. Other authors (e.g. Levelt, 1989) propose a distinction between a verbal and a non-verbal semantic system.

In the semantic system words are ordered by meaning. Words from the same semantic category (e.g. animals, tools or vegetables) or with related meanings are stored closely together. When a target is activated by an item in the input lexicon, related words also receive a certain amount of co-activation (depending on their similarity), while the activation for the target is highest. In this way, the correct word is finally understood. That means that for the example word ‘pea’ also related words such as ‘lentil’, ‘bean’, and ‘spinach’ receive activation. The activation of ‘pea’ is however highest, which finally leads to the understanding of the word ‘pea’.

1.1.3 Auditory speech perception in aphasia

In the previous section, speech perception as it occurs in non-brain-damaged listeners has been described. Brain damage, for example due to a stroke, can cause language impairments, called aphasia. In aphasia, production and comprehension of spoken and written language can be affected. Which modalities are impaired differs between individuals. Each processing level can be affected independently from the others, leading from rather isolated deficits to global deficits affecting all modalities. Language problems can therefore manifest themselves very differently. Deficits of auditory comprehension have been linked to the different modules of the Logogen model (Howard & Franklin, 1988; Franklin, 1989). Franklin (1989) has given an overview of the different types of comprehension disorders with accompanying case descriptions to underline the distinctions between different processing components.

Deficit in the auditory analysis component

A deficit in the auditory analysis of speech leads to a disorder or inability to analyze the incoming speech. This disorder has been called 'word-sound deafness' (Franklin, 1989). It was first described by Kussmaul (1877), who called it 'pure word deafness' because the patient he described did not suffer from other aphasic symptoms. Word-sound deafness can be diagnosed with auditory discrimination tasks in which participants have to report whether two auditory, phonologically related stimuli (words or nonwords) are the same or different. In word-sound deafness, problems are restricted to linguistic material, while there are no problems in discriminating or identifying non-linguistic auditory stimuli. In her study, Franklin (1989) presented three patients with difficulties in auditory discrimination: two with only mild impairments, and one with a severe impairment. The latter was considered to suffer from 'word-sound deafness'. Franklin (1989) did not classify the two patients with mild impairment due to the mildness of their deficit.

Patients with a disorder in auditory analysis will present with severe comprehension problems, as their inability to identify and discriminate speech sounds prevents them from (correct) further linguistic processing. It was, however, reported that aphasic patients perform better in word than in nonword discrimination (Caplan & Aydelott-Utman, 1994). Cattell (1886) described a single patient and showed, for the first time, that processing of words is superior to processing of nonwords: letters were named faster when presented within a word than when presented within a nonword (word-superiority-effect). This general proof of word-superiority together with the findings of Caplan and Aydelott-Utman (1994) indicates that there is a lexical influence even at this early processing stage. Furthermore, aphasic individuals with a disorder of the auditory analysis will have more severe problems in discriminating items differing by less phonetic features. The more features are different, the easier the discrimination task becomes (Blumstein, Baker, & Goodglass, 1977; Blumstein, 1994). It is also important which dimension differs between phonemes (Saffran, Marin, & Yeni-Komshian, 1976; Blumstein et al., 1977; Caplan & Aydelott-Utman, 1994; Csépe, Osman-Sági, Molnár, & Gósy, 2001). Several factors, which have a beneficial influence on the performance, have been described. Among those are the use of context, slowed speech and the possibility to see the speakers face, thus the possibility to gain speechreading information (Buchman et al., 1986; Shindo et al., 1991). Of these, only the latter factor was also successfully utilized in treatment studies (Gielewski, 1989; Morris, Franklin, Ellis, Turner, & Bailey, 1996; Grayson, Hilton, & Franklin, 1997; Hessler & Stadie, 2008).

In the previous section, the processes involved in the comprehension of the word 'pea' have been described. If an aphasic patient with a deficit in the auditory analysis perceives the sound stream belonging to the word 'pea', he has problems in the extraction of the features. Therefore instead of [plosive], [bilabial], and [voiceless], leading to the phoneme /p/, he might extract only the features [plosive] and [bilabial] correctly and therefore forwards the wrong phoneme to the buffer, namely /b/, the voiced counterpart of /p/. This influences all further processing steps as the incorrect input to the buffer cannot be corrected later on in processing. Even if all other modules function flawlessly, 'pea' will not be comprehended correctly, but rather 'bee' will be perceived.

Deficit in the auditory input buffer

A disorder of the auditory input buffer results in problems with holding on to the analyzed sounds. This leads to problems in the following steps of processing. A disorder of the input buffer is identified by growing difficulties with increasing stimulus length. This can be seen, for example, in (word or nonword) repetition tasks. Problems concerning the order of phonemes are characteristic as well. There can be misunderstandings based on a permutation of phonemes. Usually, fewer problems occur with initial than with final sounds.

For the example task of comprehending 'pea' this means that the correct order of the phonemes /p/ and /i:/ is not stored. Therefore the information made available to the input lexicon could be /i:p/ rather than /pi:/.

Deficit in the auditory input lexicon

A deficit in the auditory input lexicon leads to problems in the activation of lexical entries. It is possible that related words are activated instead of the target. The more frequent a target is (i.e. the more often it occurs in the language) the easier it is to access the lexical entry. Lexical decision tasks can detect this disorder. In these tasks, participants listen to a stimulus, which can be either an existing word or a nonword. Words can vary in frequency of occurrence, while nonwords can differ in similarity to existing words. Patients with a deficit in the auditory input lexicon will have difficulties in this task, which are more profound for low-frequency words than for high-frequency words. A disorder of this processing component has been called 'word-form deafness' by Franklin (1989). Out of her nine patients, three presented with problems in auditory lexical decision, while auditory discrimination was not impaired. Those patients were considered to suffer from a deficit in the auditory input lexicon and therefore from 'word-form deafness' (Franklin, 1989).

In the example of the word 'pea' a patient with word-form deafness would have trouble in accessing the correct word form. It is possible that 'pea' is not recognized as a word at all. Another possibility is that a competitor, for example the phonological neighbor 'knee', is selected instead of the target.

Deficit in access to the semantic system

A disorder in accessing the semantic system (from the auditory input lexicon) leads to difficulties in comprehending the meaning of the target word. However, this problem with retrieving the meaning is limited to auditory comprehension. If no other deficits exist, accessing the meaning can be accomplished if the word is presented in the written modality. Impairments in the access to the semantic system can be determined by comparing the results of auditory and written semantic tasks as synonym judgment or word-picture-matching. If the impairment is limited to the access and does not concern the semantic system itself, difficulties will only occur in the auditory version of these tasks. Franklin (1989) called this phenomenon 'word-meaning deafness' and presented one patient who had problems in an auditory synonym judgment task, but not in the written version. As his performance was not impaired in auditory discrimination or lexical decision tasks this modality-specific deficit cannot be attributed to deficits in the components described before and must, therefore, result from a deficit in accessing the semantic system.

A patient with word-meaning deafness would have trouble in comprehending the target word 'pea', while he is aware that it is an existing word. The access to the semantic system is impaired, which leads to problems in retrieving the correct meaning. Instead of the target the patient might select the meaning of a related word, such as 'lentil'. These problems occur only in auditory comprehension. Reading, writing and orally producing the word do not form a problem.

Deficit in the semantic system

A disorder in the semantic system itself is central to all language modalities; speaking, reading, writing and listening are all effected. Therefore semantic errors should occur in every modality. The parameters 'concreteness' and 'imageability' have a major influence of the performance of patients with a deficit in the semantic system. Concrete or highly imageable words are considered to be easier to process than abstract words. A task like synonym judgment has the advantage that abstract words can be included to compare them to concrete words. Franklin (1989) distinguished between two types of central semantic deficits: general semantic deficit and abstract semantic deficit: in the first, no effect of concreteness is found while in the second, only abstract words are problematic. For each subtype, Franklin (1989) presented one case. Both patients had no difficulties in auditory discrimination or lexical decision, but presented with severe problems in auditory as well as written synonym judgment. One of them showed an effect of concreteness while the other did not. Therefore, Franklin (1989) came to the conclusion that the first suffered from an abstract semantic deficit while the second had a general semantic deficit.

A patient with a general semantic deficit will have comparable difficulties understanding the word 'pea' as one with word-meaning deafness. However there are also problems in understanding the written word and producing the word. The heard or written word 'pea' could be comprehended as 'lentil' or 'bean'. Also when asked to name a picture of a 'pea', the patient could produce a semantic paraphasia, naming it 'lentil'. As the example 'pea' is a concrete word, it should not form a difficulty for patients with an abstract semantic deficit. An abstract word such as 'peace', however, can lead to comprehension problems for patients with either a general semantic deficit or an abstract semantic deficit.

In Table 1.1 below, an overview of the different levels of impairment is provided. Furthermore the table lists the attributed symptoms and the terminology that has been used to refer to the disorder.

Table 1.1: Levels of impairment in auditory comprehension. Terminology taken from Franklin (1989).

Functional deficit	Symptoms	Terminology
Auditory Analysis	Impairment in discriminating and/or identifying speech sounds, often improved by speechreading	Word Sound Deafness
Auditory Input Buffer	Problems to perceive the correct order of sounds, difficulties increase with length	
Auditory Input Lexicon	Difficulties in recognizing and comprehending spoken words	Word Form Deafness
Access to the Semantic System	Problems in comprehending spoken words, leading to semantic errors, restricted to auditory modality	Word Meaning Deafness
Semantic System	Difficulties in all semantic tasks, independent of modality	General Semantic Deficit
	Difficulties in all semantic tasks, independent of modality, but specific for abstract words	Abstract Semantic Deficit

1.1.4 TRACE model

While the Logogen model (Howard & Franklin, 1988) is a valuable tool in the diagnosis of aphasic disorders, a model more specifically aimed at speech perception is necessary to embed the current research theoretically. The TRACE model of language processing[1] (McClelland & Elman, 1986) is a model specifically developed to explain the processes involved in auditory comprehension. In this interactive activation model, several levels of processing are assumed: a feature, a phoneme and a word level. The feature level contains acoustic features, such as 'acuteness' or 'vocalic'. These are connected to the phoneme level, where single phonemes are represented. The third level holds complete words. The three levels are fully interconnected. Units within one level are connected via lateral inhibition. That means that units on the same level can

[1]The name TRACE refers to a specific function of the model: that of leaving traces of activations. Those traces are a delay window which store the activation patterns of any given moment before continuing the processing in order to account for the recovery of mispronunciations or context influences

inhibit each other. On the feature level, that means that activation of one value of a feature inhibits activation of a different value of the same feature. Across levels the connections are excitatory. This excitation applies both bottom-up and top-down. Every feature is connected to every phoneme and every phoneme to every word. However, the strength of activation differs. When the incoming speech sound is low on vocality (that means when it is a consonant), the feature 'vocalic' will be only weakly activated. The degree of this activation is then compared to the defaults specified for each phoneme. The matching phonemes with the most comparable value for 'vocalic' get the highest activation, while those who are furthest away will get the least activation. The input is temporally organized to assure the correct order of phonemes in a word. This model incorporates a delay window, called 'trace', to account for context and mispronunciation effects (where incorrectly pronounced words are still recognized). In the 'trace' all patterns of activation are stored that correspond to a stimulus that has not yet been identified. A schematic illustration of the model is depicted in Figure 1.2. A picture of the actual activation patterns of the computer model[2] is provided in appendix A.1.

In both the Logogen (e.g. Morton, 1969; Howard & Franklin, 1988) and the TRACE (McClelland & Elman, 1986) model, activation of entries at one level in course activates entries at the following level. In the Logogen model, this is accomplished by employing thresholds: whenever a certain entry reaches the critical threshold, it is chosen. In the TRACE model, the word with the highest activation after a certain time is selected. This is one of the small differences between the models. A major distinction is that the Logogen model aims to explain single word processing in all modalities, while the TRACE model concentrates on auditory processing. As TRACE is a connectionist model which is computationally implemented, many details about the activation patterns are given. All bottom-up, top-down and within-level connections are specified. In contrast to that, not many details are provided about the Logogen model, especially not concerning pre-lexical processing. Because of its highly interactive architecture, TRACE can be easily extended to include also the processing of visually perceived speech. Therefore, this model is particularly useful for the current studies.

[2] A reimplementation by Strauss, Harris, and Magnuson (2007), called JTRACE, has been used.

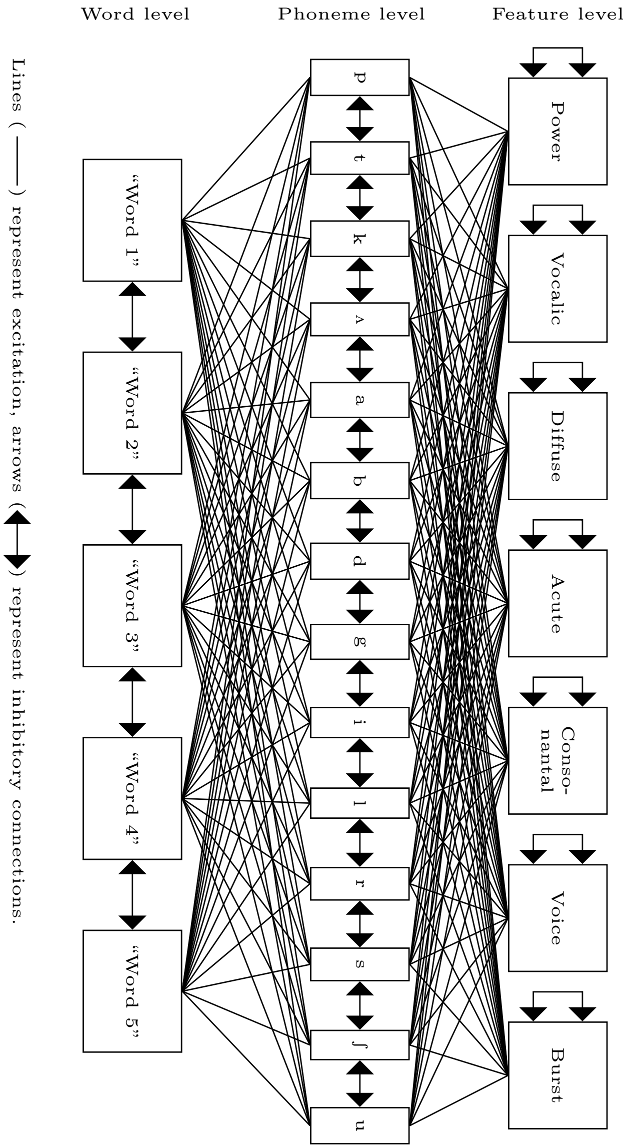


Figure 1.2: TRACE model of speech perception.

1.2 Audiovisual speech perception

The models described above account for the processing of auditory input. Language comprehension is, however, a multimodal process. Not only auditory but also visual information (seen speech) is employed in perception (Rosenblum, 2008). This has been demonstrated in different contexts, as discussed in the following section.

1.2.1 Evidence for multimodality

Evidence for the primacy of multimodal processing comes from different studies investigating auditory and audiovisual speech processing. These studies concern the comprehension of speech in noise (Sumby & Pollack, 1954), with demanding contents (Reisberg, McLean, & Goldfield, 1987) or of incongruent auditory and visual information (McGurk & MacDonald, 1976)

Speech perception in noise

Sumby and Pollack (1954) were the first to describe the influence of visual information on auditory speech perception in noise. In their experiment participants heard bisyllabic words which were presented with different levels of noise. The speech-to-noise ratio varied from 0 to -30 db. That means that the signal varied from speech and noise being evenly hard to the point where noise was 30 db louder than the speech. Participants were asked to select the heard word from a list. This task was used to define the intelligibility of the speech signal. Intelligibility scores decreased as the speech-to-noise ratio decreased for auditory only and audiovisual presentation. However the resistance to noise was much higher when speech was presented audiovisually. The difference between intelligibility scores in the auditory and audiovisual condition increased as the speech-to-noise ratio decreased: the worse the listening conditions (more noise) the bigger the difference between the auditory and the audiovisual condition. With these findings, Sumby and Pollack (1954) showed that the participants made use of the information presented visually, especially for speech difficult to comprehend solely based on auditory information (due to high levels of noise). The authors concluded that having access to visual information, i.e. seeing the speakers face, contributes to comprehension. This finding is highly relevant to ‘everyday speech perception’ since in normal listening situations, speech is usually accompanied by background noise.

Perception of demanding contents

While Sumby and Pollack's (1954) study focused on speech that was difficult to hear due to the added noise, Reisberg et al. (1987) took a different approach to analyze the influence of speechreading. In their study, they increased the cognitive load by presenting speech with demanding contents while the quality of the auditory signal was not influenced. The participants were asked to shadow (repeat as fast as possible) speech with and without the possibility to see the speaker. The text passages were taken from Smith's (1965) English translation of Kant's (1787) "Kritik der reinen Vernunft". The shadowing performance was scored by a judge blind to the condition of presentation (auditory or audiovisual). Shadowing performance was significantly better when speech was presented audiovisually rather than only auditorily. The authors concluded that perceiving intact auditory input can also be aided by speechreading. This indicates that visual information is not only taken into account when the auditory signal is degraded.

The McGurk effect

In the evidence for multimodality presented so far, auditory speech was difficult to understand, either because of noise or because of cognitive demands. But also clear auditory speech is influenced by visual information, as has been shown by the findings of McGurk and MacDonald (1976). In their study, participants watched dubbed videos with non-matching auditory and visual information and had to report what they perceived. Instead of answering with the auditory (/ba/) or the visual (/ga/) component of the video, they often reported a fusion of both: /da/. Information gained through speechreading was combined with the auditory information to form a percept, even though no necessity to depend on visual information (e.g. due to background noise) was given. This phenomenon, known as the 'McGurk effect' demonstrates that information from the seen face forms part of language processing, supporting the notion of primacy of multimodal processing: both auditory and visual information are processed under all circumstances rather than visual information being a mere fall-back mechanism that is only applied when needed.

Extending the described findings, Manuel, Repp, Studdert-Kennedy, and Liberman (1983) conducted a study in which the syllable /va/ (visual) was dubbed onto the syllable /ba/ (auditory). The majority of participants reported to have heard /va/, the visual input. This was interpreted as another indication for the influence of seen speech on heard speech. Later this combination of syllables was studied by other authors investigating the integration of speechreading (e.g. Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007) although it is not clear whether there really is an influence on the auditory stimulus or rather a reliance on the visual part. Despite the differences to the original McGurk effect this phenomenon is sometimes also referred to as 'McGurk effect'. In the current studies, however, the term 'McGurk effect' refers to the classical fusion of auditory and visual input into a new phoneme, different from both presented inputs.

Often, the audiovisual integration underlying the McGurk effect has been regarded as an automatic unconscious process (e.g. Colin et al., 2002; Soto-Faraco, Navarra, & Alsius, 2004). There is, however, convincing evidence that, next to experiencing the McGurk effect, also the unimodal information is consciously processed. Soto-Faraco and Alsius (2007, 2009) did a 'McGurk study', in which their participants were asked to judge the synchrony of the audiovisual stimuli as well as to report their perception. It was found that in a certain window of asynchrony, although this asynchrony was detected, a multisensory percept ('McGurk' response) was reported. Soto-Faraco and Alsius' (2007, 2009) data show that there is also conscious access to the unisensory inputs, which can explain the 'awkward' feeling reported by many participants.

The McGurk effect has been replicated investigating the influence of different factors, such as native language and age. The original study (McGurk & MacDonald, 1976) was conducted with English-speaking children and adults. Replications have been carried out in different (combinations of) languages: Sekiyama and Tohkura (1991) investigated the McGurk effect in Japanese. They report that the effect is less likely to occur than in English and is influenced by the intelligibility of the auditory signal. Also in Cantonese and Dutch, the effect occurs less often than in English (De Gelder, Bertelson, Vroomen, & Chen, 1995). There was no difference concerning the number of fusion responses for the Dutch and Cantonese speakers. The results for Dutch were supported by findings from Klitsch (2008). Aloufy, Lapidot, and Myslobodsky (1996) compared Hebrew and English speakers. English speakers were much more susceptible to the McGurk effect than speakers of Hebrew, who also recognized the incongruity more readily. Sekiyama (1997) tested Chinese native speakers with English and Japanese McGurk stimuli. When tested with English stimuli, the participants perceived more fusions than when tested with

Japanese stimuli. Generally, they showed only a modest McGurk effect, which was, also for the English stimuli, much lower than for English native speakers. Colin, Radeau, and Deltenre (1998) tested French native speakers with different sound intensities and found a generally low number of fusion responses, which increased with decreasing sound intensity. Grauwinkel and Fagel (2006) used synthetic speech to evaluate the McGurk effect in German. The number of McGurk responses was much lower than in English and was also dependent on the amount of white noise added. The more white noise was added, the more the participants shifted focus to the visual condition, resulting in more fusion, more visual and less auditory answers.

Another factor influencing the susceptibility to the McGurk effect is the age of the participant. The effect has been shown in infants of 4.5 months old (Burnham & Dodd, 2004) and increases with age. Even within the group of adults it has been shown, that older adults perceive fusions more readily than younger adults (Klitsch, 2008; Ohde & Abou-Khalil, 2001). Klitsch (2008) explained the increased occurrence of the McGurk effect in the group of older adults with a cognitive account: participants have to divide their attention to attend both modalities, as instructed. This is easy for the younger adults resulting in detection of the incongruity more often and therefore focusing on the dominant (auditory) modality. The group of older adults cannot divide the attention as easily which prevents them from isolating one modality resulting in more McGurk responses.

1.2.2 Models of audiovisual integration

To explain audiovisual integration, several models have been proposed. Robert-Ribes, Piquemal, Schwartz, and Escudier (1996) categorized these models in four distinct groups. The authors formulated three questions on the basis of which the categories can be established. Those questions are addressed subsequently, starting with the most general: is there a common representation of both modalities (auditory and visual)? If there is none the model is a 'direct integration model'. If there is a common representation, the next question is asked: is the integration of both types of information an early or late process? If it is a late process (occurring after the derivation of a code for each modality), the category of the model is a 'separate identification model'. If there is early integration, the last question needs to be asked: is there a dominant modality? Models that suggest a dominant modality (usually the auditory) are classified as ‘dominant recoding models’ while those with an amodal (possibly motor) representation are in the category of ‘motor space recoding models’. This classification taken from Robert-Ribes et al. (1996) is also depicted in the flowchart below (Figure 1.3).

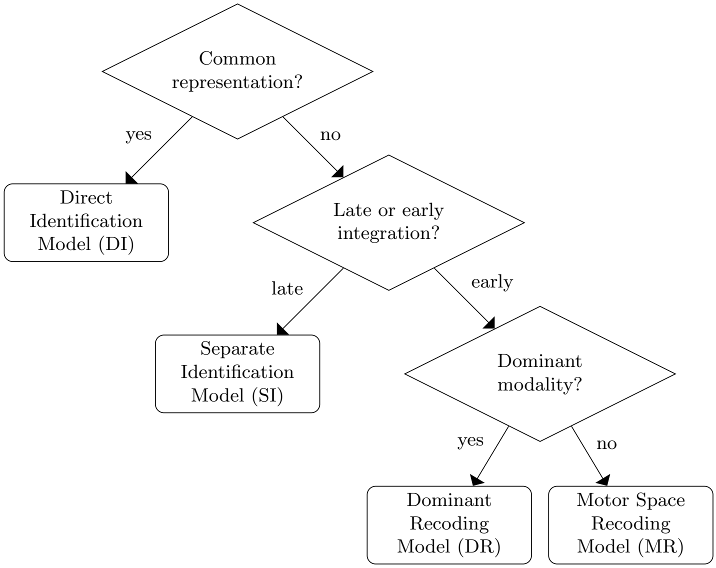


Figure 1.3: Flowchart of audiovisual speech recognition models (adapted from Robert-Ribes et al., 1996).

In the following sections the four types of models will be discussed. Whenever possible, examples of existing models will be given to illustrate the features of the category. Furthermore, possible disadvantages of the models will be highlighted.

Direct identification models

The first category of models is formed by those models in which the input signals (auditory as well as visual) are transmitted directly to a bimodal classifier, as illustrated in Figure 1.4 below. There is no access to the unimodal information of the input prior to integration. Robert-Ribes et al. (1996) call these types of models therefore ‘direct identification models’ (DI).

A model of this type was proposed by Summerfield (1987) by extending Klatt’s (1979) “Lexical Access From Spectra (LAFS)” model to incorporate

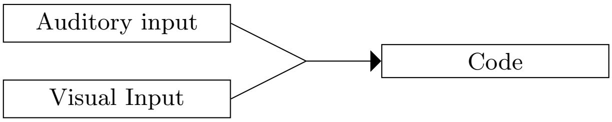


Figure 1.4: Sketch of DI model (adapted from Robert-Ribes et al., 1996).

speechreading. In the original model, words were identified from a finite, but large repertoire. The best match between a stored template and the spectral input was chosen. To account for coarticulation effects, the stored sequences were pairs of phonemes rather than single phonemes. Summerfield (1987) proposed an extension to include seen speech: next to spectra also images of the speaker's mouth were stored and compared.

Robert-Ribes et al. (1996) argued that this type of model should be rejected because a common representation seems necessary to explain experimental findings. It has been found, that participants can detect the audiovisual incongruity, but are still subject to an automatic fusion of both inputs (Summerfield & McGrath, 1984). This indicates that a common representation, where both modalities can be compared (before being fused), is required. Therefore, according to Robert-Ribes et al. (1996), direct integration models as described above have to be rejected.

Separate identification models

The second type of model is called 'separate identification model' (SI) by Robert-Ribes et al. (1996). A sketch of the general architecture of this kind of models is given in Figure 1.5. Two parallel recognition processes are assumed, one for each modality. The features from each modality are then fused into a single code. This fusion can be accomplished on logical or probabilistic values.

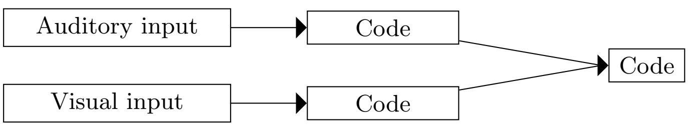


Figure 1.5: Sketch of SI model (adapted from Robert-Ribes et al., 1996).

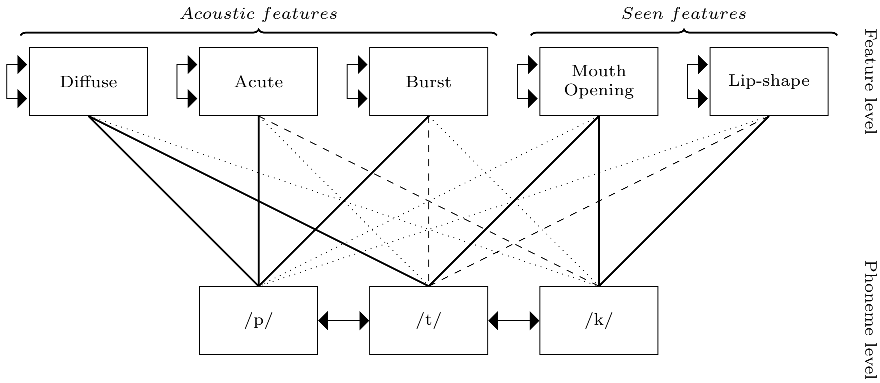
Examples of SI models assuming a fusion based on logical values are discussed by Summerfield (1987). This is for example the case in the 'Vision: Place, Audition: Manner' (VPAM) hypothesis: the percept relies on the visual input to gain information about the place of articulation and on the acoustic input for information about voicing and manner of articulation.[3] The fusion between the two codes based on the use of probabilistic (or fuzzy-logical) values has been assumed in the 'Fuzzy Logical Model of Perception' (Oden & Massaro, 1978; Massaro, 1987). In this model the percept is chosen from competing hypotheses on the basis of probability estimations. It operates in three steps: evaluation, integration and decision. Evaluation is performed for each modality separately based on sensory cues (phonetic features). Those are stored (depending on experience) in the long term memory as prototypes. The input is then compared to them and the probability of agreement with the prototypes is computed. In the integration stage the probabilistic values of both modalities are taken together to calculate an overall probability. In the last step, 'decision', the stimulus with the highest overall probability is chosen as percept.

Another model of this type has been proposed by Campbell (1988, 1990). Her model is an extension of the TRACE model of auditory speech perception (McClelland & Elman, 1986) described above, adding mechanisms accounting for speechreading. The visual features introduced by Campbell (1988, 1990) are 'mouth opening' and 'lip-shape'. Together with the acoustically perceived features, these seen features form the feature level, with interaction within this level and with the phoneme level. The features 'mouth opening' and 'lip-shape' mainly convey information necessary to decode the 'place of articulation'. This model therefore predicts that 'place of articulation' is influenced by speechreading, but it is not clear whether there is also influence on the dimensions 'manner of articulation' and 'voicing'.

Another advantage of Campbell's (1988, 1990) model is the fact that the McGurk effect can also be explained in terms of it: the relevant acoustic features to discriminate /p/, /t/, and /k/ are 'diffuse', 'acute', and 'burst'. Upon hearing a /p/, the phoneme /p/ naturally receives the highest activation. Because /t/ is still somewhat similar it also receives substantial activation (for instance /p/ and /t/ do not differ on 'diffuse'). The phoneme /k/, however, receives least activation, as it shares less similarities with /p/. Simultaneously to hearing /p/, the participants see the speaker articulating /k/. Therefore the phoneme /k/ receives the most activation from the two seen features. The

[3]Summerfield (1987) gives some evidence against a model as simple as that, but introduces other versions of rule-based models, which rely on more distinguished rules.

phoneme /t/ is identical to /k/ concerning ‘mouth opening’ and differs only slightly with regard to ‘lip-shape’. Therefore it also receives substantial activation from the two seen features. The phoneme /p/ however differs immensely from /k/ concerning both seen features. It is characterized not only by a closed mouth, but also by a different ‘lip-shape’, with wider spread lips. It, therefore, receives little to no activation from the seen features. As activation from both input types is cumulated, neither /p/ nor /k/, which served as inputs, but /t/ is finally selected because it received overall the highest activation. This is depicted in Figure 1.6.



Solid lines (——) represent strong excitation, dashed lines (- - -) weaker excitation and dotted lines (·······) a very weak or no excitation.

Arrows (◀—▶) represent inhibitory connections.

Figure 1.6: The adjusted TRACE model for McGurk items. Only the features that differ between /p/, /t/, and /k/ are taken into account. The strength of the connections represents the excitation based on the input (auditory /p/ dubbed on visual /k/). For example, a strong link between the feature ‘diffuse’ and the phonemes /p/ and /t/ is assumed because the ‘diffuse’ value of the input (/p/) strongly activates the phonemes /p/ and /t/. The phoneme /k/ differs substantially from the input with regards to this feature and, thus, the excitation is very weak, which is represented by a dotted line.

Robert-Ribes et al. (1996) give several arguments against late integration. They claim that the strongest argument is provided from studies investigating how speechreading can be aided by acoustic information (Rosen, Fourcin,

& Moore, 1981; Grant, Ardell, Kuhl, & Sparks, 1985; Breeuwer & Plomp, 1986). In those studies, (normal hearing) participants had to identify speech from speechreading, complemented with auditory information (e.g. the overall amplitude and/or fundamental frequency of the stimulus). While it was impossible to identify for example the feature 'voicing' with either type of information alone, participants succeeded with the combination of both input types. According to Robert-Ribes et al. (1996), this is incompatible with late integration, as neither decoding module provides any cue allowing a decision on the voicing feature. This indicates that fusion must occur at an early stage of processing and the separate identification models have to be rejected.

Yet, when considering the interactive architecture of the TRACE model, this criticism can be rejected: even though the model suggests that separate codes are generated, it does not imply that either of theses codes is sufficient to recognize a phonetic feature unambigously. It is rather the interaction of the different code types that leads to the correct result.

Dominant recoding models

The third type of model integrates both types of incoming information before a code is generated. Because there is a dominance of the auditory input, it is called 'dominant recoding model'(DR) (Robert-Ribes et al., 1996). The visual input is recoded into a representation of the auditory modality. This is done independently from the auditory input. Only after both types of input are processed, they are fused. A sketch can be seen in Figure 1.7. This architecture has not been used to build psychophysical models and was only a few times implemented in automatic speech recognition (e.g. Robert-Ribes et al., 1996).

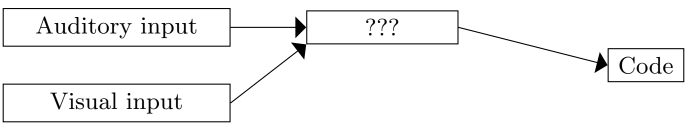


Figure 1.7: Sketch of DR model (adapted from Robert-Ribes et al., 1996).

Robert-Ribes et al. (1996) argue that in a dominant recoding model (with dominance of the auditory modality), the visual information can only influence the auditory percept if they are contradictory or if the auditory information is affected by noise. However, data from Lisker and Rossi (1992) show that visual information can bias a clear, congruent auditory stimulus. In their study participants had to judge the rounding category of a vowel, for example the

French /ɯ/ (the unrounded counterpart of /u/). When presented ‘visually only’, participants correctly rejected it as being rounded: /ɯ/ was considered rounded only 1 percent of the times. In an ‘auditory only’ condition, the same vowel was judged rounded 60 percent of the times. When presented audiovisually, the interaction of the two modalities became apparent, as the vowel was judged rounded 25 percent of the times. This is difficult to explain in a model, that assumes a dominance of the auditory modality as the visual information clearly influences the perception here. Therefore, also the dominant recoding model seems unlikely to Robert-Ribes et al. (1996).

Motor space recoding models

In the last type of model, the ‘Motor space recoding model’ (MR), both inputs are translated into an amodal description and then fused (Robert-Ribes et al., 1996). A sketch of this architecture is shown in Figure 1.8. A classical example of this type of model would be the ‘Motor Theory of Speech’ (Liberman & Mattingly, 1985). In this model, the amodal description is a motor description (vocal tract configurations or motor programs). Both input modalities establish their own vocal tract configurations, which are then fused and finally forwarded to a phonetic classifier. This process can take into account that some dimensions can hardly be seen (as the velum) and give less weight to input about that dimension gained through seen speech.

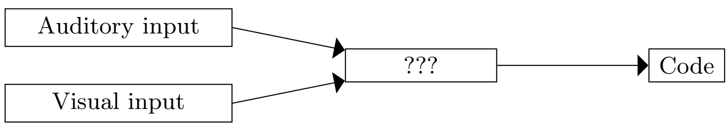


Figure 1.8: Sketch of MR model (adapted from Robert-Ribes et al., 1996).

The motor space recoding model can explain all the data given as counterarguments to the before described models. In fact, Robert-Ribes and colleagues state that they “do believe that the only model compatible with the whole set of experimental data in the field of AV perception is the MR model” (Robert-Ribes et al., 1996, p. 199).

This notion is controversial. Massaro and Chen (2008), for example argue against the motor theory of speech (Liberman & Mattingly, 1985; Galantucci, Fowler, & Turvey, 2006) and present data in favor of their own model (the fuzzy logical model of perception), which represents the group of separate identification models. One major problem of the motor theory of speech is that it cannot explain the top-down influence that lexical information has on phoneme perception, such as the biasing influence lexical constraints have on phoneme identification (Massaro & Oden, 1995). For example, the findings of Ganong (1980) showed this influence: a sound on the /d/-/t/ continuum was more likely to be judged /d/, when presented in the context 'dash-tash' (where only the first is an existing word). The same sound was however judged /t/, when in the context of 'dask-task', where 'dask' is a nonword. Another problematic issue for the motor theory of speech is the ability of infants to perceive speech sounds and speech sound distinctions they do not produce yet (MacNeilage, 1991). Massaro and Chen (2008) further argue that it is not clear how the McGurk effect can be explained within a framework based on motor representations. If the auditory input and the visual input are both transformed into a motor representation, there will be two conflicting motor representations. The motor theory does not provide any information on how this problem can be resolved and the observed fusion perception can be explained.

In the current studies, the focus therefore lies on Campbell's (1988, 1990) adoption of the TRACE model. This model can explain the beneficial influence of visual information on (aphasic) comprehension and especially the occurrence of the McGurk effect with interaction between the different kinds of features perceived auditorily and visually.

1.2.3 Audiovisual speech perception in aphasia

It has been found that seeing the speaker's face improves comprehension in aphasic individuals with word-sound deafness (Buchman et al., 1986). Shindo et al. (1991) investigated this advantage in more detail: one patient with word deafness and three patients with auditory agnosia[4] were tested (among other measures) with a speechreading test. Participants were asked to repeat or write down the perceived word or sentence. It was presented in three conditions: 'speechreading only', 'auditory', and 'audiovisual'. While a group of control participants scored at ceiling in the auditory and audiovisual conditions, all four patients showed difficulties in the auditory condition and much less so in the audiovisual condition. However, their 'speechreading only' results were lower than those of the control group. That means that they were not extraordinary

[4]Auditory agnosia is a more general deficit in auditory perception than word-sound deafness. Patients with auditory agnosia also suffer from difficulties in perceiving non-speech sounds (Brown, 1974).

speechreaders, but that combining both types of information in the audiovisual condition leads to the improvement that was found.

Next to the beneficial effect of speechreading reported above, there are also studies making use of the McGurk effect to evaluate audiovisual integration in aphasia. By investigating the McGurk effect it was hoped to find out whether audiovisual processing in aphasia differs from the processing in non-brain damaged populations. Several studies focusing on the McGurk effect in aphasia will be discussed in the following section.

McGurk effect in aphasia

In order to analyze the integration of speechreading, several authors studied the McGurk effect in aphasic populations (Campbell et al., 1990; Youse, Cienkowski, & Coelho, 2004; Klitsch, 2008). Campbell et al. (1990) conducted an experiment on the McGurk effect in four brain-damaged participants, two of whom suffered from left hemisphere brain damage. One of these participants presented with aphasia, the other with pure alexia and hemianopia. The aphasic participant was impaired in auditory processing, but was able to speechread. He exhibited a McGurk effect on words (e.g. auditory 'pick' dubbed on visual 'kick', McGurk perception 'tick') and consonants (e.g. auditory /aba/ dubbed on visual /aga/, perception /ada/), but reported the visual part for vowels (e.g. auditory 'erbee' dubbed on visual 'erboo', perception 'erboo'). The alexic participant had a good auditory performance, but was hardly able to speechread. Her responses were most often resembling the auditory part of the stimuli. The other two participants had right hemisphere lesions and suffered both from prosopagnosia (one developmental, one acquired). Both were able to speechread, but only the participant with acquired prosopagnosia showed fusion responses. Campbell (personal communication) tested another participant with global aphasia, which resolved to word deafness. He was poor in identifying auditorily presented syllables and did not improve when materials were presented audiovisually. For this participant, no McGurk effect was found.

Another description of the McGurk effect in aphasia was presented by Youse et al. (2004), who assessed a male patient with mixed aphasia and two non-brain-damaged participants. The aphasic participant exhibited problems in a syllable identification task for all three tested conditions: 'auditory only', 'visual only', and 'audiovisual' (congruent and McGurk). In each condition, syllables were presented and had to be identified by the participant. The syllables used in this experiment were /bi/, /di/, and /gi/ for all conditions and additionally McGurk stimuli for the audiovisual condition: auditory /bi/ dubbed onto visual /gi/ (so called 'fusion stimuli') and vice versa ('combination stimuli'). From a preliminary analysis, it seemed that the participant showed only McGurk responses (in this case /di/) on the incongruent stimuli. However, the results were seriously influenced by a response bias, as was shown by the congruent condition, where he also very often picked the answer /di/ (in 83% of the cases, while only 33% were correct). The reported McGurk responses could therefore have emerged as a result of this bias.

Klitsch (2008) compared the performance of a group of six Dutch aphasic participants to a group of age-matched non-brain-damaged control participants. The task consisted of watching a video and choosing between three answer options: the McGurk-type answer, the auditory component or the visual component of the video. The study was designed to examine the influence of different factors (e.g. the age of the participants and lexical status of the material) on the performance. The stimuli were therefore subdivided into sets of different lexicality. One set was formed by lexical auditory and visual forms, leading again to a lexical percept. A second set was made of lexical auditory and visual forms, however leading to a non-lexical outcome. Another set consisted of non-lexical auditory and visual forms, leading to a lexical percept and the last set was made of non-lexical inputs leading to a non-lexical response.

Aphasic and age-matched participants exhibited the same number of overall McGurk responses (43% and 45%, respectively). Also, the response pattern (most often McGurk type answers, followed by auditory and visual answers) did not differ significantly between the groups. Regarding the influence of lexical status on the McGurk effect, Klitsch (2008) reported that both groups showed the highest percentage of McGurk responses in the condition with non-lexical inputs leading to a lexical output, indicating a lexical influence on the McGurk effect. This difference is however only significant for the aphasic group. Results per group and lexical condition can be seen in Table 1.2.

The finding that lexical status can influence pre-lexical perception has also been reported by Ganong (1980) (see above). Based on his findings as well as her own results into account, Klitsch (2008) argues that the McGurk effect is a phonetic-phonological effect, which is influenced by lexical status. This lexical influence is stronger for aphasic than for non-brain-damaged listeners.

Table 1.2: Overview of McGurk responses in Klitsch's (2008) study; L = lexical, NL = non-lexical.

Auditory + Visual → McGurk	Aphasic Participants	Age-matched Controls	Younger Controls	All Participants
L + L → L	38.3%	39.5%	22.3%	33%
L + L → NL	27.5%	34.8%	16.6%	26%
NL + NL → L	66.7%	55.0%	28.0%	50%
NL + NL → NL	36.7%	44.8%	22.7%	36%
All Stimuli	43%	45%	22%	34.6%

In the studies described in this thesis, the audiovisual processing of speech is investigated in aphasic and non-brain-damaged listeners. Due to the lexical influences, as reported by Klitsch (2008), only nonwords will be used in the experiments in order to exclude a lexical bias. In chapter 2, the focus lies on identifying the phonetic dimensions that are particularly difficult for Dutch aphasic listeners and how these dimension are influenced by speechreading. Chapter 3 relates more specifically to the dimension 'place of articulation' and investigates the McGurk effect in Dutch aphasic listeners, combining offline scores with online reaction times, in order to gain more insight into the processes involved in audiovisual speech perception. Both studies have theoretical as well as practical aims: on the one hand they will contribute to the knowledge about audiovisual speech perception and on the other hand they will help to identify the damage to the processing system in aphasia in more detail. The latter is the basis for developing more specific treatment programs in the future. The aims and research questions of the individual studies are introduced in the respective chapters.

CHAPTER 2

Phonetic Dimensions in Aphasic Perception[1]

2.1 Introduction

This chapter reports a study which investigates the influence that different phonetic dimensions have on speech sound processing in Dutch aphasic participants and how speechreading can aid processing. First, we will, however, give a background on the phonetic dimensions, the influences of speechreading on comprehension and how that can be accounted for in a speech processing model. Deficits in speech sound discrimination will also be discussed.

2.1.1 Phonetic dimensions

Phonemes are considered to consist of phonetic features (Chomsky & Halle, 1968). These features can be categorized into the three (distinctive) phonetic dimensions 'place of articulation', 'manner of articulation', and 'voicing'. A combination of these dimensions uniquely identifies each phoneme. Changes in one phonetic dimension will lead to a different phoneme: changing, for example,

[1] This chapter is a slight adaption of Hessler, Jonkers, and Bastiaanse (2010). Terminology has been changed in order to be consistent throughout the thesis.

the ‘place of articulation’ from bilabial to alveolar would transform a /p/ into a /t/. However, also broader contrasts between two phonemes (changes in two or three of the phonetic dimensions) are possible: /p/ and /z/ for example, are distinguished by ‘voicing’, ‘place of articulation’, and ‘manner of articulation’. Therefore, it is not sufficient to classify two words or syllables as different in one phoneme, as this distinction can be formed by differences in one, two or all three distinctive dimensions. Rather the number and type of the phonetic dimensions differing should be mentioned as well.

The phonetic characteristics of a particular phonetic dimension are manifested differently across languages: The distinction ‘voiced’ vs. ‘voiceless’ is made by differences in voice onset time (VOT). According to Lisker and Abramson (1964) the voicing distinction in English is achieved by contrasting onset of voicing at the release of the lips (for /b/) with an onset of voicing up to 100ms later (/p/). Therefore the distinction in English is rather one between voiceless and voiceless-aspirated. For other languages such as Dutch and Hungarian, however, Lisker and Abramson (1964) found that the voiceless /p/ was produced by having the onset of voicing aligned with the release of the lips, while the voiced counterpart /b/ was produced with a voice onset at least 50ms prior to the lip-release.

A distinction in ‘voicing’ therefore refers to different phonetic distinctions in different languages, which makes a comparison of for example Dutch and English data difficult. In Dutch and Hungarian, however, the distinction between ‘voiced’ and ‘voiceless’ is accomplished by similar VOT patterns.

2.1.2 Speechreading

The extraction of phonetic information out of the speech stream is an early component of language comprehension. Language comprehension is however a multimodal process. Not only auditory but also visual information (seen speech) is employed in perception (Rosenblum, 2008). It has been demonstrated that seeing the speaker facilitates comprehension in a noisy environment (Sumby & Pollack, 1954) or with cognitively demanding contents under good listening conditions (Reisberg et al., 1987).

More evidence for the fact that speechreading is automatically integrated into speech perception is provided by experiments carried out by McGurk and MacDonald (1976). In their study, participants watched dubbed videos in which auditory and visual information did not match, and they were asked to report what they perceived. Instead of answering with the auditory (/ba/) or the visual (/ga/) component of the video, they usually reported a fusion of both (/da/). This even occurred when the participants were aware of the dubbing. This so-called 'McGurk' effect actually shows the influence of speechreading. It is also a demonstration of the integration of both modalities by producing a percept which is a fusion of seen and heard speech. This proves that the information a listener gains from the lip-movements of the speaker cannot be ignored and is automatically taken into account in generating a percept. Therefore, speechreading should not be understood as a substitute mechanism that only mediates when needed, but as one that supports auditory comprehension.

Campbell (1988, 1990) suggested a model explaining speech perception with multimodal (auditory and visual) input. Her model is based on the TRACE model of language processing (McClelland & Elman, 1986). In this interactive activation model of speech processing several levels of processing are assumed: a phonetic, a phonological, and an abstract phonemic level. The phonetic level consists of acoustic as well as lip-read features. These are connected to the phonological level, where phonemes are represented. The third level, consisting of abstract phonemic units, is necessary to explain why some people actually 'hear' phonemes which are not articulated, but only written, such as the /b/ in 'comb'. The three levels are fully interconnected. Units within one level are connected via lateral inhibition, that means that units can inhibit each other. Across levels the connections are excitatory. This excitation applies both bottom-up and top-down. The input is temporally organized to assure the correct order of phonemes in a word. This model incorporates a delay window, called TRACE, to account for context and mispronunciation effects (where incorrectly pronounced words are still recognized). In the TRACE all patterns of activation are stored that correspond to a stimulus that has not yet been identified.

To integrate the information from speechreading, Campbell (1988, 1990) introduced two new features at the level of phonetic units into the model: 'mouth opening' and 'lip-shape'. Together with the acoustically perceived features, these visual features form the phonetic units which interact with each other and with the phonological units. A schematic overview of the 1990 version of the model is shown in Figure 2.1.[2] The features 'mouth opening' and 'lip-shape' mainly convey information necessary to decode the 'place of artic-

[2] A more detailed description of the model can be found in Chapter 1.

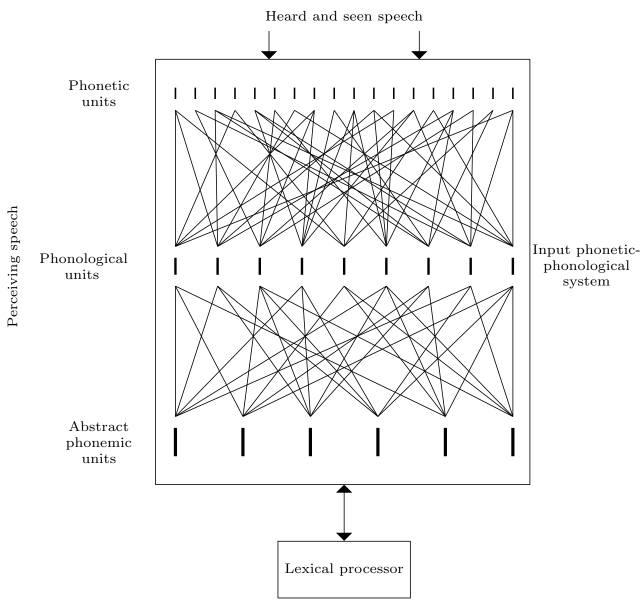


Figure 2.1: Schematic overview of a model of audiovisual processing, based on Campbell (1990).

ulation'. This model therefore predicts 'place of articulation' to be influenced by speechreading, but it is not clear whether there is also influence on the dimensions 'manner of articulation' and 'voicing'.

2.1.3 Impairments of processing

Brain damage can lead to an impairment in processing phonemes. This disorder was first described by Kussmaul (1877), who called it 'pure word deafness' because the patient he described did not suffer from other aphasic symptoms. Terminology was not consistent and now a common term is word-sound deafness (Franklin, 1989). Word-sound deafness can be diagnosed with auditory discrimination tasks, in which participants have to report whether two auditory (phonologically related) stimuli (words or nonwords) are the same or different (e.g. "house" and "mouse"). In word-sound deafness, the problems are restricted to linguistic material, while there are no problems in discriminating or identifying non-linguistic auditory stimuli.

Aphasic participants with a disorder in analyzing speech sounds will have problems in discrimination of items differing in fewer phonetic dimensions. The more dimensions are different, the easier the discrimination task becomes (Blumstein et al., 1977). Factors that have a beneficial influence are the use of context or slowed speech and, most importantly for the present study, the possibility to see the speaker, thus the possibility to gain speechreading information (Buchman et al., 1986; Shindo et al., 1991). The use of speechreading information has also been successfully utilized in treatment studies (Gielewski, 1989; Morris et al., 1996; Grayson et al., 1997; Hessler & Stadie, 2008).

Hessler and Stadie (2008) evaluated the effects of a systematic treatment of the auditory analysis of speech in a patient with aphasia. The treatment was based on the beneficial influence of speechreading. During treatment, six different tasks were carried out: auditory discrimination of syllables, auditory discrimination of phonemes, word-picture matching, word-picture verification, heard word-written word matching, and heard word-written word verification. In all tasks, the distractors were phonologically related to the targets. Treatment started with items with broad distinctions (three phonetic dimensions) and speechreading possible. After mastery of this condition, more difficult conditions were presented (less dimensions different, no speechreading possible). The efficiency of treatment was measured by comparing pre- and post-treatment performance in the treatment tasks on a set used during treatment and a matched, non-trained set of stimuli. Apart from showing general improvement in both the trained and the untrained set, the authors also analyzed the performance on individual phonetic dimensions. The aphasic patient improved in discrimination of 'place of articulation' contrasts as well as 'manner of articulation' contrasts (also for untrained stimuli). These results cannot be explained in terms of the model by Campbell (1988, 1990) described above. Influence on other dimensions than 'place of articulation' is not predicted by this model. The results of Hessler and Stadie (2008) indicate that speechreading can be beneficial for perceiving the other dimensions as well. It is therefore not clear which phonetic dimensions ('place of articulation', 'manner of articulation', and/or 'voicing') make use of the additional information from seen speech.

Processing of which phonetic dimensions is most impaired in aphasic comprehension disorders has been investigated previously: Blumstein et al. (1977) compared the processing of the dimensions 'place of articulation' and 'voicing' in English speaking aphasic participants and found that they have most problems with 'place of articulation'. They did not include the dimension 'manner of articulation'. Saffran et al. (1976) and Caplan and Aydelott-Utman (1994), however, found (also for English aphasic listeners) that 'voicing' actually is more difficult than 'place of articulation'. Similar results have been found for Hungarian by Csépe et al. (2001) for two aphasic participants with unilateral left-hemisphere lesions (opposed to the bilateral cases also investigated). Klitsch (2008) used two subtests of the Dutch version of the PALPA (Bastiaanse, Bosje, & Visch-Brink, 1995) to investigate whether aphasic listeners showed differences in detecting distinctions in the dimensions 'place of articulation', 'manner of articulation', and 'voicing'. In the two discrimination tasks carried out, words as well as nonwords were investigated. Generally the performance of the aphasic participants was better when word pairs had to be distinguished. The detection of differences in 'place of articulation' was worse than 'manner of articulation' but only for nonwords. The comparison of either 'place of articulation' or 'manner of articulation' with 'voicing' was more difficult, as 'voicing' distinctions were realized in initial positions, the other contrasts however in final (less salient) position or metathesis. For comparison with 'voicing', only the performance in metathesis distinctions in 'place of articulation' and 'manner of articulation' was taken into account. Compared like this, performance on 'voicing' distinctions was worse than on 'manner of articulation' distinctions, while there was no difference with 'place of articulation'. Klitsch (2008), therefore, came to the cautious conclusion that 'place of articulation' was affected most, but also noted that the dimension 'voicing' could not be compared reliably to the other dimensions, because they were not occurring in the same position within the stimuli.

In the current study, all three dimensions will be compared again, but manipulated in the same position (initially). In order to investigate the influence speechreading has on the discrimination performance, the task will be presented in three conditions: with only auditory input, with audiovisual input and with only visual input (a video of lip-movements, serving as a control condition).

Based on the former studies, we expect that aphasic participants with a disorder in speech sound processing benefit from information derived from speechreading in discriminating between similar phonemes. Therefore, their overall performance in the 'audiovisual' condition will be better than in the 'auditory only' condition. This beneficial influence of speechreading will be manifested in the phonetic dimension 'place of articulation'. Based on Campbell's (1988, 1990) model no beneficial influence is predicted on the dimensions 'manner of articulation' and 'voicing'.

It is also expected that the degree of difference has an influence on the performance: The more dimensions differ between two items, the easier discrimination becomes for the individuals with aphasia. They will have least difficulties with distinguishing stimuli when all three phonetic dimensions differ, while differences in only one phonetic dimension are expected to cause most problems.

Additionally, it will be investigated whether all three phonetic dimensions are equally difficult for individuals with aphasia or whether one of them is particularly difficult and, if so, which one. For the 'audiovisual' condition it is predicted that the dimension 'voicing' will be most difficult, as within this dimension it is not possible to make use of visual information. This prediction, however, does not hold for the 'auditor only' condition, as there is no visual information available.

2.2 Methods

2.2.1 Participants

Six participants with aphasia (three female) and 14 non-brain-damaged control participants (seven female) took part in this study. All participants were native speakers of Dutch, right-handed, and reported normal hearing. The hearing was also judged as within functional limits by their speech-therapists. Vision was normal or corrected to normal. The aphasic participants were between 47 – 64 years old (mean age: 52.33). The participants in the control group were matched with the aphasic participants for age (mean age 56.29; range 49 – 67), gender, and region of origin. They had never experienced neurological problems and had no (history of) language disorders.

All aphasic participants were at least 3 months post-onset. None of them had demonstrated any language disorders prior to the CVA. They did not suffer from any neuropsychological problems influencing the testing (such as severe attention disorders). The participants were selected on the basis of their results in the PALPA nonword discrimination task (Bastiaanse et al., 1995). In this task, participants hear two nonwords which are either the same or differ in one phonetic dimension from each other. They have to judge whether both heard stimuli are the same. A failure to do so has been attributed to an impairment of the auditory analysis of speech. This task was administered using a recording of the stimuli in order to maximize the comparability between participants. As the normative data of the PALPA (Bastiaanse et al., 1995) are collected with direct speech, Klitsch (2008) collected normative data for the recorded version. We compared the results of our participants to her normative data. Performance more than 2 SD below the mean of Klitsch's group were considered as impaired. Only aphasic participants with impaired performance on the discrimination task were included in this study. Thus, all aphasic participants in this study had a deficit in the auditory analysis of speech. Furthermore, all but one participant had been diagnosed with a standardized battery, the Akense Afasietest (AAT) (Graetz, De Bleser, & Willmes, 1992). The performance in two sub-parts, the Token Test and the comprehension part, gives an indication of the comprehension abilities of the participants. In the Token Test, participants have to follow commands such as "Touch the green rectangle" or "Put the red square under the red circle". These commands, of which 50 are presented, vary in length and complexity. The results are reported as error scores. A score of '0' would therefore mean 'no errors', while '50' represents the fact that no command at all could be executed correctly. The comprehension part of the AAT consists of word and sentence comprehension tasks. A word or sentence is presented and the participant is asked to choose between four pictures, one depicting the target and the other three distractors (one or two of which are related to the target). The maximum score that can be reached is 120. An overview of the personal data of the aphasic participants and their results on these two AAT tasks (Graetz et al., 1992) and the PALPA nonword discrimination task (Bastiaanse et al., 1995) are given in Table 2.1.

Table 2.1: Overview of the personal data of the aphasic participants.

Initials	Age	Gender	Type of Aphasia	Months post-onset	AAT Token Test	AAT Comprehension	PALPA Nonword Discrimination
WB	57	male	Wernicke	148	37	94	56/72
BB	64	male	Global	5	50	67	53/72
EK	48	male	Anomia	16	11	88	58/72
TB	47	female	Global	8	33	53	68/72
JH	51	female	Mixed	44	36	89	66/72
MB	47	female	Global	4	50	68	64/72

2.2.2 Materials

The materials consisted of one-syllable nonwords with CVC(C) structure. They were spoken by a male native speaker of Dutch, who was recorded in a quiet room with daylight. Additionally, a light diffuser was used to avoid shading on the recorded material in order to ensure optimal visual information. The recorded frame included the lower part of the speaker's face (from the bottom of the nose), the neck, and the upper chest. For the recording, a video camera and separate cardioid microphone were used. The video was then digitized into avi-files at a sampling rate of 48 kHz with 32-bit-stereo quantization. All stimuli were edited with Adobe Premiere to form video files with a duration of 3 seconds each. Recording was done with 25 frames per second (thus 40 ms per frame). Therefore each file consisted of 75 frames. The video showed the speaker in rest (with a closed mouth) for 12 frames (480 ms) at the end of each video. The resting phase in the beginning was varied slightly to ensure equal length of all videos. To provide equal length of rest, the last or first frame of the video was artificially prolonged, where necessary.

As the experiment was carried out in three conditions ('audiovisual', 'auditory only', and 'visual only' presentation), the audiovisual video files were then further edited to create the stimuli for the other conditions. For the 'auditory only' condition, the picture was deleted, leaving the sound and a blank screen. In the 'visual only' condition, the audio trace was removed resulting in a video without sound. Finally, the video files were converted into Windows Media files (.wmv), reducing file size in order to guarantee smooth running of the experiment without long delays for loading the files.

As said above, the material consisted of pairs of nonwords. These were presented either 'auditorily only', 'audiovisually' or 'visually only', depending on the sub-condition of the experiment. Distribution of the material can be seen in Figure 2.2. A complete overview of the used stimuli can be found in appendix B.1.

2.2.3 Procedure

Each participant was tested in three sessions: In the first session the PALPA nonword discrimination task (Bastiaanse et al., 1995) was carried out. The experimental task was administered in two further sessions. The task used in this study was a discrimination task, asking the participants to state whether

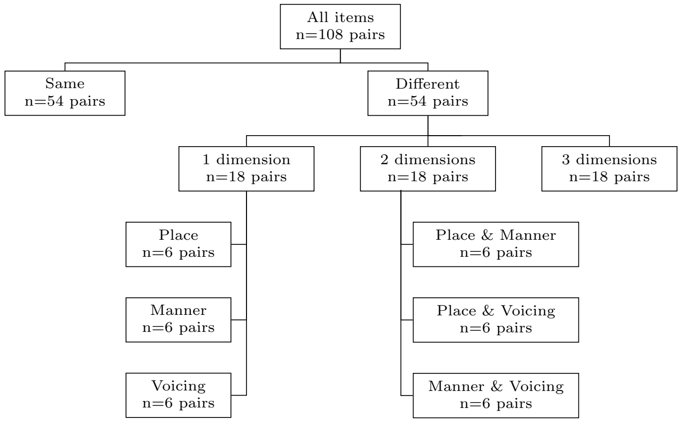


Figure 2.2: Overview of materials used.

two heard and/or seen syllables are the same. It was carried out in three different conditions: (1) 'auditory only' (2) 'audiovisual', and (3) 'visual only'. The last condition was introduced as a control condition, indicating that the presumed better performance in the 'audiovisual condition' was not solely due to the visual information. For all three conditions, the same stimuli were used. The items of each condition were split into two blocks, so that only half of the items were presented in one session. The order of presentation of the blocks was balanced between participants.

The materials were presented to the participant on a laptop equipped with headphones and a response box using E-Prime 2.0 (Psychology Software Tools). For each condition there were five practice trials before the experiment started. On those items feedback was provided. The practice trials were repeated if the participant requested it or if it seemed necessary to explain the procedure again. The experimental task was only started once the participants responded correctly to at least 80% of the trials in the 'auditory only' and 'audiovisual' conditions without help. Each condition was presented separately, not mixing the conditions. Items were randomized to prevent learning effects across conditions. The order in which the conditions were presented was varied between participants, so that a possible learning effect would not favor a certain condition.

A schematic outline of the procedure per trial is shown in Figure 2.3. Item presentation occurred self-paced. Prior to each pair of nonwords an asterisk was shown on the screen. The video only started when requested by the participant

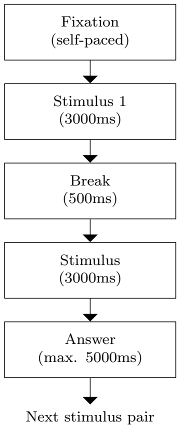


Figure 2.3: Overview of applied procedure.

by pushing a button. After both nonwords were presented, the participant had 5 seconds to respond with 'yes' or 'no'. This response was also given by pushing a button (a green button for 'yes' and a red one for 'no'). If no response had been recorded after 5 seconds, the nonword pair was presented again with another 5 seconds to decide. If there was no answer after the second presentation, the next stimulus pair was presented. Each condition of the experiment consisted of 108 stimulus pairs; however, as mentioned above, only half of those were presented within one session.

2.3 Results

2.3.1 Overall performance

The non-brain-damaged control participants scored at ceiling for the 'auditory only' and 'audiovisual' conditions. In the 'visual only' condition, they performed worse, failing mainly in contrasts involving only 'voicing' or 'manner of articulation' or the combination of both. The aphasic participants scored significantly lower than control participants on all conditions (2-tailed Mann-Whitney-U tests: 'auditory only': 99% - 87% correct, Z=-3.521, $p<0.001$; 'audiovisual': 99% - 90% correct, Z=-3.545, $p<0.001$; 'visual only': 83% - 63% correct, Z=-3.387, $p<0.001$). Because the non-brain-damaged control participants performed at ceiling the following analyses were only conducted within the group of aphasic participants.

2.3.2 Influence of speechreading

Based on previous studies it was expected that aphasic participants with speech sound processing disorders benefit from speechreading. It was investigated whether the performance of the aphasic participants also improved with speechreading. The results in the three conditions 'auditory only', 'audiovisual', and 'visual only' (control condition) differed significantly (Friedman Anova $\chi^2(2)=12$, $p<0.01$). Post-hoc Wilcoxon tests revealed that the 'audiovisual' condition was significantly easier than both the 'auditory only' ($Z=-2.207$, $p<0.05$) and the 'visual only' condition ($Z=-2.201$, $p<0.05$). There was also a significant advantage for the 'auditory only' over the 'visual only' condition ($Z=-2.207$, $p<0.05$). This also holds on an individual basis: The performance in the 'audiovisual' condition was better than in the 'auditory only' condition for five out of the six aphasic participants.

Analysis by phonetic dimension

According to Campbell's (1988, 1990) model, it was expected that performance concerning the dimension 'place of articulation' would improve with the addition of speechreading cues. No improvement was predicted for the other two dimensions. In order to investigate the influence of speechreading a comparison of the 'audiovisual' and the 'auditory only' conditions separately for each phonetic dimension was carried out. It revealed no significant differences for the dimensions 'place of articulation' (2-tailed Wilcoxon test: $Z=-0.816$, $p=0.414$) and 'voicing' (2-tailed Wilcoxon test: $Z=-.674$, $p=0.5$). For the dimension 'manner of articulation', a trend for better performance on 'audiovisual' stimuli could be found (2-tailed Wilcoxon test: $Z=-1.826$, $p=0.068$). The individual results in appendix B.2 show that this trend was not caused by single participants, but was found for four of the six aphasic participants, while the other two showed no difference between both conditions.

2.3.3 Number of distinguishing dimensions

Based on the results of Blumstein et al. (1977), it was predicted that the number of phonetic dimensions differing would influence the performance of the aphasic participants, such that the fewer phonetic dimensions differ the worse the performance becomes.

Analyses revealed that the number of dimensions differing within the pair played a role for aphasic participants in the 'auditory only' (Friedman Anova: $\chi^2(2)=8.667$, $p<0.05$) and 'audiovisual' (Friedman Anova: $\chi^2(2)=11.143$, $p<0.01$) conditions. In the auditory condition, it was found that differences in one distinctive dimension were significantly less likely to be detected than differences in two (2-tailed Wilcoxon test: Z=-2.023, $p<0.05$) or three (2-tailed Wilcoxon test: Z=-2.207, $p<0.05$) dimensions. There was however no significant difference between distinctions in two and three dimensions (2-tailed Wilcoxon test: Z=0.0, p=1.0). Similar results have been found for the audiovisual condition: distinctions in two and three dimensions were not significantly different from each other (2-tailed Wilcoxon test: Z=-1.604, p=0.109), while both were easier to perceive than distinctions in one dimension (2-tailed Wilcoxon test: Z=-2.201, $p<0.05$ for both comparisons). The individual data of the aphasic participants (appendix B.1) show that these findings were not caused by the performance of single participants, but hold for all aphasic participants in the 'auditory only' condition and all but one in the 'audiovisual' condition.

2.3.4 Type of distinguishing dimension

Previous studies have found contradictory results concerning the question which phonetic dimension is most impaired in aphasic perception. Therefore, no prediction was made for the 'auditory only' condition. For the 'audiovisual condition' it was expected that differences in the dimension 'place of articulation' would be the easiest to perceive, as, following Campbell's (1988, 1990) model, beneficial influence of speechreading is assumed for this dimension, but not for the other two dimensions.

An analysis of the influence of type of dimension ('place of articulation' vs. 'manner of articulation' vs. 'voicing') was carried out, showing significant results for the 'auditory only' condition (Friedman Anova: $\chi^2(2)=6.7$, $p<0.05$) and marginally significant results for the 'audiovisual' condition (Friedman Anova: $\chi^2(2)=4.727$, p=.094). For both conditions it appears that 'voicing' was the most difficult to distinguish, followed by 'place of articulation' and 'manner of articulation' (see Figure 2.4). In both the 'auditory only' and the 'audiovisual' condition, five out of six aphasic participants showed the same pattern as the group, with 'voicing' being most difficult (see appendix B.2). The group result, therefore, reflects a vast majority of the participants' performances, rather than being caused by extremes in the data distribution.

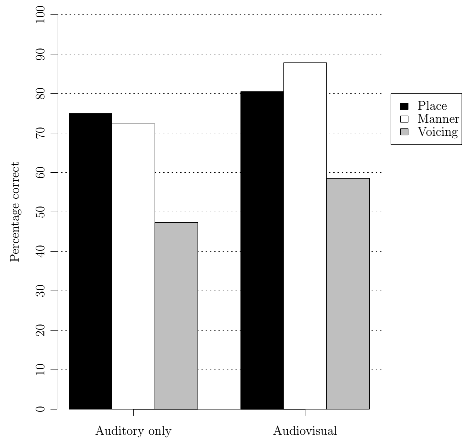


Figure 2.4: Percentage of correct aphasic responses to different dimensions in auditory only and audiovisual conditions.

2.3.5 Answer bias

Results of ‘yes-no-paradigms’ can be influenced by an answer bias of the participants. This can be corrected by using methods from signal-detection research. Within this paradigm the hit-rate and the false-alarm-rate are used to calculate a measure of discriminability, d-prime. The calculation of d-prime is a parametric procedure. As the current data do not fulfill the demands for parametric testing, d-prime could not be calculated. Instead a non-parametric variant, a-prime (A’), was calculated to correct for a response bias. A’-scores vary between ‘0’ (no discriminability) and ‘1’ (perfect discriminability), with ‘0.5’ being chance-level. In the current study we applied the algorithms from Snodgrass, Levy-Berger, and Haydon (1985) to calculate A’. All statistical analyses have been repeated using the bias-corrected A’-scores. Also using these scores (rather than the non-corrected ones) it becomes evident that the aphasic participants score significantly worse than the non-brain-damaged controls in all three conditions (‘audiovisual’, ‘auditory only’, and ‘visual only’). Regarding the analyses within the aphasic group, the results resembled those of the non-corrected scores, with two exceptions: The overall difference between the ‘audiovisual’ and the ‘auditory only’ condition does not yield significance, but forms a trend when based on A’-scores. The same is true for the difference between two and one dimension distinctions found in the ‘auditory only’ condition: Using the corrected A'-scores a trend, rather than a significant difference, can be found, indicating that two dimensions cause more difficulties than one. The individual A'-scores are mentioned in appendix B.2 and the results of the statistics using A'-scores are provided in appendix B.3.

2.4 Discussion

The aim of the current study was to investigate how perception of phonetic dimensions is impaired in speech processing by individuals with aphasia and how that processing is influenced by speechreading. A discrimination task was carried out in three conditions: 'auditory only', 'audiovisual', and 'visual only' stimulus presentation. A group of fourteen non-brain-damaged control participants and six aphasic participants took part in this study. The aphasic participants were diagnosed with different syndromes, but shared a deficit in processing speech sounds. The small number of aphasic participants does not allow for general conclusions, rather all conclusions drawn refer only to the group tested.

It was found (repeating numerous previous studies) that discriminating pairs of nonwords is more difficult for individuals with an aphasic disorder in speech sound processing than for non-brain-damaged control participants. When analyses use the bias-corrected A'-scores this observation also holds.

Generally, the aphasic participants showed a very homogeneous pattern. For all analyses reported in this paper, a broad majority of the aphasic participants showed performance in the same direction as the group. The group analyses were, therefore, based on a consistent pattern within the group rather than on extreme performances of single participants. The possibility that hearing problems influenced the results is ruled out by the fact that the aphasic participants had only slight problems with differences in three dimensions in the 'auditory only' condition. If the underlying problems were in hearing, this condition should have been affected as well.

Overall, there was a trend to better performance of individuals with aphasia in the 'audiovisual' condition than in the 'auditory only' condition, indicating that the additional visual information gained from speechreading facilitates their discrimination abilities. The performance in the control condition with 'visual only' stimulus presentation was worse than in both other conditions, indicating that the superiority of the 'audiovisual' condition is not due to pure visual information, but rather the combination of auditory and visual input.

For the non-brain-damaged control group no difference between the 'auditory only' and the 'audiovisual' conditions were found, as they performed at ceiling in both. No further analyses were carried out for the non-brain-damaged control group.

For the aphasic participants, it was further tested whether the general advantage of the 'audiovisual' over the 'auditory only' condition was due to improvement on one of the phonetic dimensions in particular. Therefore, the difference between the 'audiovisual' and the 'auditory only' conditions was analyzed individually for each of the three phonetic dimensions 'place of articulation', 'manner of articulation', and 'voicing'. According to the model of Campbell (1988, 1990) improvement, particularly in the dimension 'place of articulation' was expected when additional speechreading is possible. However, we did not find significant differences between 'audiovisual' and 'auditory only' presentation for any of the dimensions individually. It is, hence, not possible to say whether there was more improvement for one of the dimensions than for another. The general improvement is, therefore, not due to one dimension in particular, but rather to a summation of improvement on all of them.

These findings are not in line with Campbell's (1988, 1990) model, as only improvement for distinctions in 'place of articulation' was predicted. This prediction was previously questioned in the treatment study by Hessler and Stadie (2008). They found improvement for 'manner of articulation' after a treatment based on utilizing speechreading. Therefore, it seems that Campbell's (1988, 1990) model needs to be extended to account for influences from visual features other than 'mouth opening' and 'lip-shape'.

The difficulties individuals with aphasia experience when discriminating stimuli were more profound for smaller distinctions. As previously reported by Blumstein and Cooper (1972) and Blumstein et al. (1977) for English, we also found for Dutch that differences are less likely to be detected if the items within the pair differ in one phonetic dimension (rather than in two or three). This holds for the 'auditory only' as well as for the 'audiovisual' condition. Even though performance is generally better in the 'audiovisual' condition it is still impaired, especially regarding the small differences. Speechreading adds information that enhances speech sound processing for small as well as larger differences. Therefore, the distinction between one and two or three dimensions can also be found for the 'audiovisual' condition. In their explanation why smaller differences are more difficult to perceive, Blumstein and Cooper (1972) note that in a discrimination task it is not necessary to analyze the auditory information into its linguistic components. A mere comparison of the phonetic properties of the two stimuli is sufficient. Therefore, they argue, the worse performance for the small differences can be explained by the fact that they are perceptually closer together. This argument is supported by the fact that the dimensions are phonetically conveyed differently: 'voicing' is based on temporal cues while 'place of articulation' and 'manner of articulation' rely mainly on spectral cues. When 'voicing' and at least one of the spectral dimensions differ, both types of cues are involved, while in differences in one dimension only either temporal or spectral cues are altered. Therefore one type of cue is the same in the stimuli, making a distinction more difficult.

The question of which dimension is most difficult to perceive for individuals with aphasia has been addressed previously with ambiguous results. Blumstein et al. (1977), for English, and Klitsch (2008), for Dutch, found 'place of articulation' to cause most difficulties. Saffran et al. (1976) and Caplan and Aydelott-Utman (1994), for English, and Csépe et al. (2001), for Hungarian, however, found 'voicing' to be most impaired, as in the current study. As mentioned above, Dutch and English, though both Germanic languages, differ in their phonetic realization of 'voicing' of plosives: While Dutch contrasts voiced (voice onset before lip-release) and voiceless-unaspirated (voice onset during lip-release) sounds, English shows a differentiation between voiceless-unaspirated (voice onset during lip-release) and voiceless-aspirated (voice onset after lip-release) (Lisker & Abramson, 1964; Jansen, 2004). This difference cannot explain the ambiguous results within the English data. It, however, makes it difficult to compare the English and Dutch data. A comparison of the Dutch and Hungarian data, on the other hand, is possible as both languages have a similar phonetic realization of 'voicing' (Lisker & Abramson, 1964; Jansen, 2004).

The difference in performance between 'voicing' on the one hand and the two other dimensions on the other hand could, for the 'audiovisual' condition, be explained by the fact that 'voicing' cues are considered to be not visible. As we however found the same pattern in the 'auditory' condition, thus without visual information, we suggest a different analysis: The difference between the phonetic dimensions can be explained by the different phonetic cues encoding them. As explained above, 'voicing' is phonetically conveyed by temporal cues, while 'place of articulation' and 'manner of articulation' are based on spectral cues. As distinctions in 'voicing' are most difficult for the aphasic participants, they seem to have predominantly an impairment in processing the temporal cues necessary to perceive the difference between 'voiced' and 'voiceless'.

In conclusion, the current study shows that additional visual information (gained from speechreading) positively influences the discrimination abilities of aphasic participants with a speech sound processing disorder. To what degree the individual phonetic dimensions are influenced could not be conclusively answered. Contrasts between items are more easily detected if they result from wider distinctions (more differing phonetic dimensions). Furthermore, the type of dimension differentiating items is of importance, indicating that differences in 'voicing' are most difficult to perceive for Dutch individuals with an aphasic disorder of speech sound processing.

2.4.1 Clinical implications

In the current study, it was shown that different phonetic dimensions can be affected to different degrees in speech sound processing disorders. It is yet to be determined how this relates to more general comprehension tasks as lexical decision and word-picture matching or real-life comprehension. As lexical retrieval is dependent on the correct phonetic input, the auditory analysis of speech sounds is an important part of the comprehension process: It is the first step to accurate word processing. The actual influences of different phonetic dimensions on higher-level tasks and real-life comprehension, however, still need to be established in follow-up research. Only then can the next step, improving treatment, be made. However, Hessler and Stadie (2008) have shown that taking into account the phonetic structure of stimuli while utilizing speechreading is beneficial. The results of the current study give more information about the characteristics that need to be considered in developing treatment, such as the fact that distinctions in 'voicing' were more difficult to detect for the current group of aphasic participants than those in 'place of articulation' or 'manner of articulation'.

If these differences actually effect higher processes and real-life comprehension as well, it should be investigated for all patients prior to treatment, which dimensions are especially problematic for them. Treatment for patients as the ones described in the current study should then include a focus on the timing cues necessary to perceive distinctions in 'voicing'. The current study therefore provides not only the theoretical conclusions described above, but also preliminary clinical implications can be drawn.

CHAPTER 3

Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: A Study on the McGurk Effect

3.1 Introduction

Auditory speech perception has been described in terms of different models. One of these is the TRACE model of speech perception (McClelland & Elman, 1986). In this interactive activation model, several levels of processing are assumed: a feature level, a phoneme level, and a word level. The feature level consists of acoustic features, such as 'acuteness' and 'vocalic'. These are connected to the phoneme level, where single phonemes are represented. The third level consists of complete words. The three levels are fully interconnected. Units within one level are connected via lateral inhibition, so that units can inhibit each other. On the feature level, this means that activation of one value of a feature inhibits activation of a different value of the same feature. Across levels the connections are excitatory. This excitation applies both bottom-up and top-down. The input is temporally organized to assure the correct order of phonemes in a word. This model incorporates a delay window, called TRACE, to account for context and mispronunciation effects (where incorrectly pronounced words are still recognized). In this delay window all patterns of activation are stored that correspond to a stimulus that has not yet been identified.

Unlike as described by the TRACE model (McClelland & Elman, 1986), speech perception is based not only on auditory but also on visual information (seen speech). Therefore Campbell (1988, 1990) extended the model. The feature level consists of the same acoustic features as assumed in the original model. It has, however, been extended by also including visual features based on the input from seen speech. The visual features introduced by Campbell (1988, 1990) are 'mouth opening' and 'lip-shape'. While Campbell (1990) provided the values of the model for 'mouth opening', she did not implement or specify the feature 'lip-shape'. The feature 'lip-shape' is part of the model as it is discussed in the current study. Together with the acoustically perceived features, these visual features form the units on the feature level, which can inhibit each other. The feature level is connected by means of excitation to the phoneme level. The features 'mouth opening' and 'lip-shape' mainly convey information necessary to decode the 'place of articulation'. Based on this model, it is predicted that 'place of articulation' is influenced by speechreading. Other models of audiovisual processing have been proposed as well, for example the fuzzy-logical model of speech perception (Oden & Massaro, 1978; Massaro & Oden, 1995; Schwartz, 2010) and a recent neural network model by Loh, Schmid, Deco, and Ziegler (2010). However, the aim of the current study is not to discriminate between different models, therefore these models will not be discussed further. The extended TRACE model (Campbell, 1988, 1990) will be used to explain the McGurk effect in the following section and the results in the discussion.

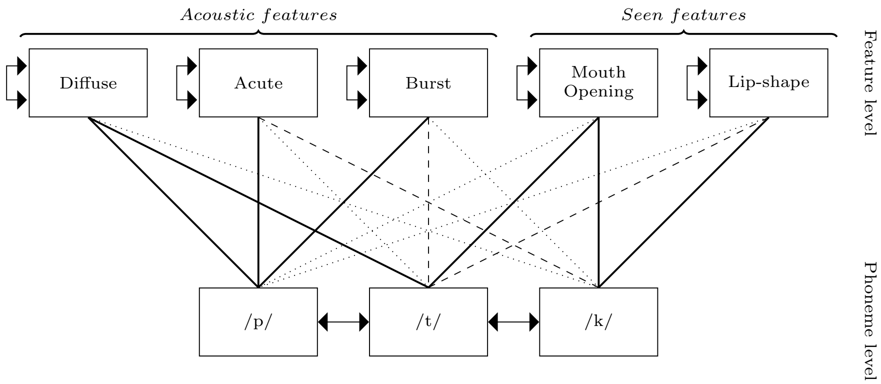
The notion that speech perception is audiovisual can be supported by the findings of McGurk and MacDonald (1976). In their study, participants watched dubbed videos with non-matching auditory and visual information and had to report what they perceived. Instead of answering with the auditory (/pa/) or the visual (/ka/) component of the video, they mostly reported a fusion of both: /ta/. This phenomenon, known as the McGurk effect, demonstrates that information from the seen face forms part of speech perception, supporting the notion of primacy of multimodal processing: both auditory and visual information are processed under all circumstances rather than visual information being a mere fall-back mechanism that is only applied when needed. Often the audiovisual integration underlying the McGurk effect has been regarded as an automatic unconscious process (e.g. Colin et al., 2002; Soto-Faraco et al., 2004). There is, however, convincing evidence that next to experiencing the McGurk effect, the unimodal information is also consciously processed. In studies of Soto-Faraco and Alsius (2007, 2009), participants were asked to judge the synchrony of the audiovisual stimuli as well as to report their perception. It was found that for some degrees of asynchrony, the asynchrony was detected, but also a multimodal percept ('McGurk' response) was reported. The first task required conscious access to unimodal information, while the latter is evidence in favor of audiovisual integration. Soto-Faraco and Alsius (2007, 2009) concluded that both unimodal and multimodal information were accessed during audiovisual integration.

The strength of the McGurk effect differs between languages. While the original findings on English (McGurk & MacDonald, 1976) showed a strong effect (98% for voiced stops, 81% for voiceless stops), much weaker effects were reported for other Germanic languages. De Gelder et al. (1995) reported an occurrence of 56% for Dutch participants. Klitsch (2008) observed a McGurk incidence of 43% for Dutch aphasic listeners, 45% for age-matched control participants, and 22% for young adults. For German, Grauwinkel and Fagel (2006) reported the rather low incidence of McGurk answers of 19.3% when stimuli were presented in clear hearing conditions. De Gelder et al. (1995) claim that language-specific speech-processing architectures, due to differences in the phonological system, influence the magnitude of the McGurk effect. It has been suggested that the size of the phoneme inventory plays an important role in explaining cross-linguistic differences (Massaro, Cohen, Gesi, Heredia, & Tsuzaki, 1993). The phoneme /ð/, for example, is only found in English and not in Dutch or German, but accounts for 10% of the responses reported by MacDonald and McGurk (1978).

The McGurk effect can be explained with the model introduced above: the relevant acoustic features to discriminate /p/, /t/, and /k/ are 'diffuse', 'acute', and 'burst'.[1] Upon hearing a /p/, the phoneme /p/ naturally receives the highest activation. Because /t/ is still somewhat similar it also receives substantial activation (for instance /p/ and /t/ do not differ on 'diffuse'). The phoneme

[1] The classification of phonemes is based on the original TRACE model (McClelland & Elman, 1986), in which the dimensions were taken from Jakobson, Fant, and Halle (1952), but treated as continua (cf. Oden & Massaro, 1978).

/k/, however, receives least activation, as it shares fewer similarities with /p/. Simultaneously to hearing /p/, the participants see the speaker articulating /k/. Therefore the phoneme /k/ receives the most activation from the two seen features. The phoneme /t/ is identical to /k/ concerning 'mouth opening' and differs only slightly with regard to 'lip-shape'. Therefore, it also receives substantial activation from the two seen features. The phoneme /p/ however differs immensely from /k/ concerning both seen features. It is characterized not only by a closed mouth, but also by a different 'lip-shape', with wider spread lips. It, therefore, receives little to no activation from the seen features. As activation from both input types is cumulated, neither /p/ nor /k/, which served as inputs, but /t/ is finally selected, because overall it received the highest activation. The outline of the model and how the McGurk effect can be explained in terms of it are depicted in Figure 3.1.



Solid lines (——) represent strong excitation, dashed lines (- - -) weaker excitation and dotted lines (·······) a very weak or no excitation.

Arrows (◀—▶) represent inhibitory connections.

Figure 3.1: The adjusted TRACE model for McGurk items. Only the features that differ between /p/, /t/, and /k/ are taken into account. The strength of the connections represents the excitation based on the input (auditory /p/ dubbed on visual /k/). For example, a strong link between the feature 'diffuse' and the phonemes /p/ and /t/ is assumed because the 'diffuse' value of the input (/p/) strongly activates the phonemes /p/ and /t/. The phoneme /k/ differs substantially from the input with regards to this feature and, thus, the excitation is very weak, which is represented by a dotted line.

Brain damage can lead to an impairment in speech perception, specifically affecting the processing of phonemes (Franklin, 1989). These problems of phonemic processing can occur independently of or in association with other aphasic symptoms. Blumstein et al. (1977) showed that a disorder in speech sound discrimination is more pronounced for small differences, thus for items that differ along fewer phonetic dimensions. These findings have been replicated for Dutch recently by Hessler et al. (2010).

Furthermore, Blumstein et al. (1977) also addressed the question of whether the dimension type also influenced the performance. They found differences in 'place of articulation' to be most impaired. Klitsch (2008) found similar results for Dutch. Contrary to those findings, some authors found the feature 'voicing' to cause most difficulties (Saffran et al., 1976; Caplan & Aydelott-Utman, 1994; Csépe et al., 2001; Hessler et al., 2010).

Participants with deficits in speech sound processing often benefit from seeing the speaker (e.g., Buchman et al., 1986; Shindo et al., 1991; Schmid, Thielmann, & Ziegler, 2009). Schmid and Ziegler (2006), however, described a group of aphasic participants who did not benefit from speechreading. They claimed that these participants suffered from a functional deficit affecting processing on a level after audiovisual integration occurred. Evidence in favor of a beneficial influence of speechreading also comes from treatment studies that utilized speechreading successfully (e.g., Morris et al., 1996; Hessler & Stadie, 2008). The beneficial effect of speechreading is not specific for one of the phonetic dimensions, but rather a summation of small improvements on all dimensions (Hessler et al., 2010). The fact that speechreading improves perception leads to the question of how participants with a deficit in phoneme processing handle the integration. This question can be addressed by investigating how these participants perceive McGurk-type stimuli.

Several authors looked at the McGurk effect in aphasic populations (Campbell, 1990; Youse et al., 2004; Klitsch, 2008). These studies confirm that participants with aphasia and an impairment in speech perception are subject to the McGurk effect. While the studies of Campbell (1990) and Youse et al. (2004) were case studies, Klitsch (2008) studied a group of six aphasic participants. Unfortunately, the participant in the Youse et al. (2004) study showed a strong answer bias, which leads to difficulties in interpreting the results. However, the studies of Campbell (1990) and Klitsch (2008) clearly showed that the outcome of audiovisual integration did not differ between aphasic and non-brain-damaged participants.

Klitsch (2008) investigated the effect in more detail and also had a look at the lexical influences on the McGurk effect. The stimuli were subdivided into sets of different lexicality. One set was formed by lexical auditory and visual forms, leading again to a lexical percept. A second set was made of lexical auditory and visual forms, however leading to a non-lexical outcome. Another set consisted of non-lexical auditory and visual forms, leading to a lexical percept and the last set was made of non-lexical inputs leading to a non-lexical response. Klitsch (2008) reported that both groups showed the highest percentage of McGurk responses in the condition with non-lexical inputs leading to a lexical output, indicating a lexical influence on the McGurk effect. Furthermore, the number of McGurk responses did not differ between aphasic and non-brain-damaged participants for either lexicality condition.

The previous research on the McGurk effect in aphasia, however, relied on offline measures and therefore provided limited information about processing itself. In the current study we will investigate audiovisual processing in aphasic and age-matched non-brain-damaged listeners by combining offline scores with online reaction times. The aim is to find out whether there are differences between healthy and aphasic processing, when looking closer into the processing itself, rather than only considering its result. In order to investigate this, a nonword identification task will be carried out in four conditions: 'auditory only' (participants can only hear the speaker), 'audiovisual' (the speaker can be heard and seen), 'McGurk' (auditory /p/ dubbed onto visual /k/, expected percept /t/), and 'visual only' (participants can only see the speaker).

We expect that the aphasic participants will be both slower and less accurate than the non-brain-damaged participants in the 'auditory only', the 'visual only', and the 'audiovisual' conditions. As the influence of speechreading on processing has proven beneficial (Buchman et al., 1986; Shindo et al., 1991; Schmid et al., 2009; Hessler et al., 2010), we furthermore expect higher accuracy of the aphasic participants and shorter reaction times for both groups in the 'audiovisual' compared to the 'auditory only' condition.

Finally, based on the results of Klitsch (2008), we assume that the number of McGurk answers will not differ between the aphasic and age-matched non-brain-damaged control participants, although it may vary in extent between individuals. The answer pattern is not expected to differ between both groups either.

The reaction times to the different answer types will be compared. Norrix, Plante, and Vance (2006) investigated reaction times in a McGurk experiment with healthy adults and adults with language learning disabilities (but no brain damage). They found that the reaction times of both groups were longer when a McGurk stimulus was presented than when congruent stimuli needed to be processed. However, Norrix et al. (2006) did not analyze the results related to the reported perception of the participants, but related to the presented stimulus. It is therefore not clear whether this slow-down for the McGurk stimuli was only present when participants experienced the McGurk effect. We assume that not the stimulus type, but the percept accounts for the reaction times, leading to longer reaction times when perceiving a fusion (thus giving a McGurk-type answer) than when perceiving either the auditory or the visual component of a McGurk-type stimulus, as only in the first case is audiovisual integration necessary. Therefore we will not compare the reaction times to different stimulus types, but to different percepts.

3.2 Methods

3.2.1 Participants

A total of 3 aphasic participants and a group of 14 control participants (7 female) took part in this study. All participants were native speakers of Dutch, right-handed, and reported normal hearing. Vision was normal or corrected to normal. The control participants were matched with the aphasic participants for age (mean age 56; range 49-67) and region of origin. They had never experienced neurological problems and had no (history of) language disorders.

The aphasic participants were tested by their speech and language therapists with the Aachen Aphasia Test (AAT) for Dutch (Graetz et al., 1992). The AAT consist of several subtests: (1) Token Test, (2) repetition, (3) written speech, (4) naming, (5) language comprehension, and (6) spontaneous speech.

Subsequently participants were tested with three subtests of the PALPA test battery (Bastiaanse et al., 1995): auditory discrimination of nonwords, auditory discrimination of words, and grapheme naming. The first two tasks served to assess auditory processing abilities. In both tasks pairs of nonwords or words were presented, and the participant was asked to state whether stimuli were the same or different. This test was presented from tape, not allowing for speechreading. The grapheme naming task was carried out in order to ensure that the aphasic participants were able to identify speech sounds from written letters, a capacity necessary in the design of the current study. Furthermore, all aphasic participants also took part in an earlier study on phoneme discrimination, where their abilities to discriminate between speech sounds differing in various dimensions were tested with auditory and audiovisual presentation (Hessler et al., 2010, see also chapter 2). The general description of the aphasic participants and the results on the described tests are presented below.

Participant WB

WB is a 57 year old male who worked as a sales director until he had a left hemisphere ischaemic CVA at age 45, 148 month prior to testing. He was diagnosed with Wernicke's aphasia. He had 37 errors on the Token Test and his scores for repetition, written speech, naming, and language comprehension were 66/150, 58/90, 100/120, and 94/120 respectively. His spontaneous speech was judged on a 0-5 scale as 3 (communicative ability), 5 (articulation & prosody), 4 (automatic speech), 3 (semantic structure), 3 (phonemic structure), and 3 (syntactic structure). WB had problems in both the auditory word and nonword discrimination tasks: he scored 56/72 correct on the nonword discrimination task and 65/72 on the word discrimination. In both tasks his main problems were in identifying differences, he hardly had any 'false alarm' responses. Problems occurred mainly when the distinction was in voicing (3/12 correct for nonwords), although he also exhibited problems in place contrasts (9/12 correct). Naming of graphemes did not cause any problems: both upper- and lowercase letters were correctly named in all cases.

In a previous study on nonword discrimination (Hessler et al., 2010), WB exhibited problems in nonword discrimination especially for small differences (in one phonetic dimension), which were most profound for voicing, but also affected distinctions in place of articulation. He was overall better when speechreading was possible than with auditory stimulus presentation.

Participant EK

EK is a 48 year old male who has worked as an interim manager at different businesses. EK presented with a deviance in the white matter of the brain which was discovered 16 months prior to testing. He suffered from memory loss, which affected the period prior to the onset. On the basis of the AAT the speech and language therapist diagnosed anomic aphasia. EK had 11 errors on the Token Test and his scores for repetition, written speech, naming, and language comprehension were 148/150, 85/90, 110/120, and 88/120 respectively. His spontaneous speech was judged on a 0-5 scale as 5 (communicative ability), 5 (articulation & prosody), 5 (automatic speech), 4 (semantic structure), 4 (phonemic structure), and 5 (syntactic structure).

Discrimination of words was hardly impaired (70/72), but EK showed more severe problems in discrimination of nonwords (58/72), where he had mainly difficulties with distinctions in voicing (6/12), but also with place (10/12) and manner (10/12) of articulation. He mastered grapheme naming with one error for both upper- and lowercase letters. This is well within the normal range.

EK also participated in the discrimination study. He had problems with distinctions in one and two phonetic dimensions, affecting all three dimensions (voicing, place, and manner of articulation). His scores differed not much between auditory and audiovisual presentation, indicating that he might not benefit from speechreading. This could be due to his poor performance in processing of seen speech: in a condition where only the articulatory movements were visible, he showed a discrimination performance at chance level.

Participant JH

JH is a 51 year old female who is a housewife and cleaning woman. She suffered from an ischaemic CVA in the left arteria cerebri media 44 months prior to testing. She was diagnosed with mixed aphasia based on the AAT results. She had 36 errors on the Token Test and her scores for repetition, written speech, naming, and language comprehension were 100/150, 60/90, 89/120, and 89/120 respectively. Her spontaneous speech was judged on a 0-5 scale as 2 (communicative ability), 5 (articulation & prosody), 5 (automatic speech), 3 (semantic structure), 4 (phonemic structure), and 2 (syntactic structure). She also suffered from apraxia of speech.

She scored 66/72 and 67/72 on the PALPA tasks on nonword and word discrimination, which is below the range of control participants who had been tested with the same audio recording (cf. Klitsch, 2008). In the nonword tasks she mainly made errors with distinctions in place of articulation. In the word task the pattern was more balanced between place of articulation and voicing distinctions. Grapheme naming was tested only for uppercase letters. She scored 20/26 correct, although she occasionally substituted the letter name with the sound it produces. The problems she experienced seemed to be caused by the apraxia of speech, rather than a problem in letter recognition.

In the discrimination study JH had problems with distinctions in one or two dimensions, which were most profound for voicing distinctions. She performed better when speechreading was possible than with auditory presentation only.

3.2.2 Materials

The testing materials consisted of 30 one-syllable nonwords with CVC(C) structure. They were spoken by a male native speaker of Dutch, who was video-recorded in a quiet room with daylight. Additionally, a light diffuser was used to avoid shading on the recorded material for optimal visual information. The recorded frame included the lower part of the speaker's face (from the lower part of the nose), the neck, and the upper shoulders. For recording, a video camera and separate cardioid microphone were used. The video was then digitized into avi-files at a sampling rate of 48 kHz with 32-bit-stereo quantization. All stimuli were edited with Adobe Premiere to form video files with a duration of 3 seconds each. As recording was done with 25 frames per second (i.e., duration of one frame is 40 ms), each file consists of 75 frames. The video showed the speaker in rest (with a closed mouth) for 12 frames (480 ms) in the end of each video. The resting phase in the beginning varied slightly to ensure equal length of all videos. To warrant equal length of rest, the last or first frame of the video was artificially prolonged, where necessary. The audio-visual congruent video files were then further edited to derive the stimuli for the other conditions. For the 'auditory only' condition the picture was taken away, leaving the sound and a blank screen. In the 'video only' condition the audio trace was deleted resulting in a video without sound. To establish the McGurk items, video and audio traces from different recordings were dubbed onto each other. Special attention was paid to the synchrony of picture and sound. Neither the auditory and visual input nor the expected McGurk answer comprised existing Dutch words.

As described above, the number of McGurk answers in Dutch was not expected to be high. This was not a major concern of the current study, as the focus was on comparing the reaction times for different answer types. Nonetheless we decided to evaluate the McGurk items in a pilot study in order to include the highest possible number of McGurk answers in the reaction time analysis. A total of 16 native Dutch speakers, who were not involved in the project, took part in the pilot study. McGurk items not provoking a 'McGurk effect' in any participant were re-edited (n=3). Finally the video files were converted into 'Windows Media' files (.wmv), reducing file size in order to guarantee smooth running of the experiment without long delays for loading the files.

The stimuli used in this experiment are given in the appendix (cf. Tables C.2.1 and C.2.2). Presentation of the stimuli occurred in four different conditions: 'auditory only' (participants could only hear the speaker), 'audiovisual' (congruent information from hearing and seeing the speaker), 'McGurk' (auditory /p/ dubbed onto visual /k/, expected percept /t/), and 'visual only' (participants could only see the speaker). The last served as a control condition to ensure that possible advantages in the 'audiovisual condition' were not solely due to the visual information but rather to a fusion of auditory and visual input. For each condition there were 30 stimuli. In the non-McGurk conditions 10 started with /p/, 10 with /t/, and 10 with /k/. The same rhymes were used in all four conditions, however the distribution of the initial phoneme differed, so that a syllable starting with /p/ in the auditory condition started with either /t/ or /k/ in the visual and with the remaining phoneme in the audiovisual condition. In the McGurk condition stems were presented with an auditory /p/ dubbed onto a visual /k/ as initial phoneme. Every rhyme was therefore used with each phoneme and also every rhyme occurred in each condition. Furthermore, there was the same number of /p/, /t/, and /k/ responses in each of the non-McGurk conditions. In order to avoid a disproportionately high number of /t/-responses (because of the McGurk condition) 60 audiovisual congruent fillers (30 beginning with /k/ and 30 beginning with /p/) were added.

Pilot study[2]

The McGurk stimuli were evaluated in a pilot study. They were presented to 16 native Dutch speakers (12 female) for evaluation. The mean age of the participants was 44.44 years (range 26-67). All participants, except one, were right-handed and none were involved in the project. Three of them had heard from the McGurk effect before, while the others were not familiar with it.

Video files were converted into flash format and displayed one by one on a website. Participants were instructed to pay attention to what they hear and what they see. They were required to write down what the speaker said for each video file. Also they were asked to state possible problems they encountered for

[2]This description of the pilot study was not part of the original paper (Hessler, Jonkers, & Bastiaanse, 2011) due to space restrictions, but is presented here, in order to provide additional information about the materials that were used.

each video file (such as noise or asynchrony). Overall, 50 stimuli were presented, consisting of 30 McGurk items, 9 other dubbed stimuli and 11 audiovisual congruent items to ensure variability in the answers. After the completion of this task participants were re-directed to another website where they were asked for some background information (age, gender, handedness, native language and familiarity with the McGurk effect).

Overall, participants responded to the items with 29.58% McGurk answers (range: 0% - 70%). The auditory part was provided in 24.17% of the responses (range: 3.33% - 60%) and 32.08% resembled the visual part of the videos (range: 3.33% - 60%). The remaining 14.17% were other responses, for example the syllable without the initial phoneme. When only considering the relevant age group (participants older than 45, n=8) the values slightly change. Results for all participants compared to only those older than 45 are shown in Figure 3.2. Values missing to form a hundred percent were other responses, mostly not giving any initial consonant as well as some completely unrelated answers. The individual results of the 16 participants and the results per item can be found in appendix C.

The results from an item-based analysis were used for the quality control of the material. Three items were re-analyzed and re-edited because none of the participants showed a McGurk effect on them. Furthermore, another five items were re-analyzed and/or re-edited because (one or two) participants reported noise or asynchrony of picture and sound. On none of the items, more than two participants stated problems in the video file.

3.2.3 Procedure

A nonword identification task was carried out. An identification task was chosen rather than a repetition task to be able to include participants with speech or language production difficulties. The participants watched videos of a speaker pronouncing a syllable. Then they had to choose which of three written syllables matched the video. It was ensured that all participants were able to read the nonwords correctly. The task was carried out in the four conditions described above: 'auditory only', 'audiovisual', 'McGurk', and 'visual only'. Items were presented on a laptop equipped with headphones and a response box using the program E-Prime (Psychology Software Tools). Items were split into two blocks, which were presented in separate sessions. Items of all four conditions were displayed within those blocks in randomized order. It was balanced between participants which block was displayed in which session.

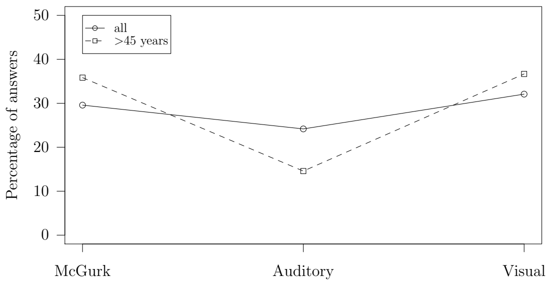


Figure 3.2: Percentage of occurrence per answer type for all participants and for only participants above 45 years.

Participants could determine the interstimulus interval by requesting a new item by pressing a button after seeing a fixation asterisk. The trial started with a short display of a picture (500 ms), indicating the modality of the following stimulus: 'auditory', 'visual' or 'audiovisual' (for 'audiovisual' and 'McGurk' condition). The indicator showed an ear for the 'auditory only' condition, an eye for the 'visual only' condition or both for the 'audiovisual' condition. This prepared the participants to direct their attention to either or both of the modalities. Subsequently the video file was played. Then, the participants saw a screen with three written syllables, from which they had to choose. The alternatives always had the same rhyme of the syllable, but started with /k/, /p/ or /t/. To avoid confusion for the aphasic participants these words were written in different colors, corresponding to colors on the response box. For reasons of consistency, /k/ was always red, appearing on top, /p/ was in the middle and green, and /t/ was on the bottom and blue. The position of the written syllable also related to the position of the matching button on the response box.[3] After the presentation of the video the participants had

[3] While this approach is helpful in avoiding confusion, it could become a problem if a participant has a response bias, preferring a certain position or color. This is not the case for the non-brain-damaged participants nor for two of the aphasic participants. WB, however, seemed to choose the phoneme /t/ more often than the others (56% of his answers started with /t/, 28% with /p/ and 16% with /k/). It is not clear whether this is due to a bias, however, because of circularity in the explanations: while the potential preference for /t/ can be explained by the fact that WB perceived syllables starting with /t/ most clearly and therefore made least errors for these syllables, it could also be that he made least errors because of an initial preference for /t/, the matching button or color. Therefore a bias cannot be excluded for WB, but also cannot be proven. For all remaining participants it was clear that there was no bias.

5 seconds to answer by pushing the appropriate button on the response box. Regardless of the type of answer (correct, incorrect, or no response) the next item was presented, but again a fixation asterisk was shown first, indicating that the participant could start the next trial. Before the actual experiment started, five practice items were presented. Two of them were in the auditory modality, one in the visual and two in audiovisual congruent. No practice item was presented in the McGurk condition in order to avoid biasing the participant for a certain answer. The participant got feedback on the practice items. If required by the participant or if it was necessary to clarify the procedure, the practice trials were repeated. During the experiment the answers and the reaction times were recorded, and no feedback was provided.

3.2.4 Analysis

In the current study the results will be analyzed with regard to the assumptions and questions introduced above. Different methods will be used to carry out these analyses. Due to the different profiles in the pre-testing, the three aphasic participants will not be grouped, but discussed independently. Accuracy and reaction times of each aphasic participant will be compared to the range of the control group to examine possible differences between the aphasic participants and the non-brain damaged control participants. A second analysis will be concerned with the difference between conditions. Because of handling data of individual participants, the nonparametric Wilcoxon test will be used to determine whether conditions differ significantly from each other. For the same reason, nonparametric tests (Friedman Anova and Wilcoxon test) will be used when comparing answer types in the McGurk condition. The analysis comparing the reaction times per answer type will be carried out with the nonparametric Kruskal-Wallis test and post-hoc Mann-Whitney U tests, where applicable. These are independent tests, but we will compare results produced by the same individuals. However, as we will analyze answer types rather than stimulus presentations, there is a different number of items in each category and reaction times cannot be compared on the basis of pairs either. In order to back up the findings from the Kruskal-Wallis tests, nonparametric correlations (Spearman rank correlations) will be carried out as well. All reaction times are included in the analysis, which means not only reaction times of correct answers. However, the overview of the results (Table 3.1) also provides the reaction times of correct answers as reference.

3.3 Results

The accuracy of the non-brain-damaged control participants was at ceiling in the 'auditory only' and 'audiovisual' condition. In the 'visual only' condition, their performance did not reach ceiling. A comparison of the reaction times of this group revealed that they reacted faster in the 'audiovisual' condition than with 'auditory only' stimulus presentation (Z = -2.152; p<.05). The results of the control group and the three individuals with aphasia can be found in Table 3.1. Individual results of the non-brain-damaged participants are provided in the appendix (Table C.3.1).

Table 3.1: Results and reaction times in the 'auditory only', 'audiovisual', and 'visual only' conditions, for each individual aphasic participant and the group of non-brain-damaged control participants. The reaction times for correct responses are given in parentheses. The mean of the control group was calculated by averaging the reaction times of all answers within the given condition, while the range reflects the minimum and maximum average reaction time of individual participants in the control group.

Initials	Auditory Only		Audiovisual		Visual	
	correct	RT	correct	RT	correct	RT
WB	53%	2176ms (2078ms)	73%	1674ms (1628ms)	52%	1899ms (1535ms)
EK	59%	2718ms (2536ms)	76%	2516ms (2395ms)	24%	3189ms (3198ms)
JH	55%	2755ms (2756ms)	89%	2353ms (2259ms)	47%	2938ms (2763ms)
controls:						
mean	99%	1462ms	100%	1422ms	78%	2177ms
range	90-100%	1085-1807ms	97-100%	1091-1786ms	67-93%	1674-2682ms
95% CI	98-100%	1347-1578ms	99-100%	1309-1534ms	74-82%	1985-2346ms

Within the McGurk condition there was a difference in prevalence of the three answer types for the non-brain-damaged participants (χ^2=30.964, df=2, p<.001). Post-hoc analyses revealed that there were more visual answers than auditory (Z=-2.548, p<.05) or McGurk-type answers (Z=-5.568, p<.001). Furthermore, auditory answers were more common than McGurk-type answers (Z=-3.079, p<.01). Reaction times differed between the three answer types (χ^2=27.405, df=2, p<.001). Participants reacted significantly more slowly when giving a McGurk-type answer than when giving an auditory (U=4166, p<.001) or visual (U=5601, p<.001) answer. There was no significant difference between the latter two (U=12965.5, p=.948). The additional analyses with the Spearman rank correlations showed comparable results: there is a significant correlation between answer type (McGurk, auditory or visual) and the reaction time for the control group (r=.201, p<.001). The results of the control group and the three aphasic participants in the McGurk condition are given in Table 3.2. The individual results of the non-brain-damaged participants can be found in the appendix (Table C.3.2).

Table 3.2: Overview of answer patterns in the McGurk condition for the three individuals with aphasia and the control group. The mean of the control group was calculated by averaging the reaction times of all answers with the given answer type, while the range reflects the minimum and maximum average reaction time of individual participants in the control group.

Initials	McGurk (/t/)		Auditory (/p/)		Visual (/k/)	
	Incidence	RT	Incidence	RT	Incidence	RT
WB	50%	1989ms	23%	2316ms	27%	2195ms
EK	18%	1912ms	46%	2061ms	36%	2297ms
JH	39%	2565ms	39%	2718ms	22%	2693ms
Controls:						
mean	22%	2021ms	33%	1650ms	44%	1644ms
range	0-50%	1136-3048ms	0-100%	1125-2636ms	0-93%	903-3617ms
95% CI	12-32%	1868-2174ms	15-52%	1537-1762ms	26-63%	1544-1743ms

The aphasic participant WB had problems in all three conditions. His accuracy was outside the normal range for auditory, audiovisual, and visual stimulus presentation. His reaction times were larger in the auditory only and in the visual only condition, but fell within the normal range in the audiovisual condition. WB responded significantly faster in the audiovisual condition than in the auditory only condition (Z=-2.293, p<.05). Furthermore he showed a trend to higher accuracy, which failed to reach significance (Z=-1.604, p=.109). In the McGurk condition, WB did not show a significant preference for either answer type (χ^2=3.800, df=2, p=.150). Moreover, the chosen answer type did not influence the reaction time (Kruskal-Wallis: χ^2=0.755, df=2, p=.686, Spearman rank correlation: r=-0.161, p=.395).

EK showed both a lower accuracy and higher reaction times than the control group in all three conditions. EK did not benefit from speechreading. Neither his accuracy (Z=-1.155, p=.248) nor his reaction times (Z=-1.375, p=.202) differed between auditory only and audiovisual stimulus presentation. In the McGurk condition, there was no difference in the prevalence of the different answer types (χ^2=3.500, df=2, p=.174). Also, the reaction times were not influenced by the given answer type (Kruskal-Wallis: χ^2=0.419, df=2, p=.811, Spearman rank correlation: r=-0.017, p=.930).

JH was both slower and less accurate than the non-brain-damaged control group in all three conditions. Her performance was improved by speechreading. She performed significantly more accurate (Z=-2.714, p<.01) and marginally faster (Z=-1.898, p=.058) in the audiovisual condition compared to the auditory only condition. JH did not show a preference for either answer type in the McGurk condition (χ^2=1.786, df=2, p=.409). Furthermore, the chosen answer type did not influence her reaction times (Kruskal-Wallis: χ^2=0.603, df=2, p=.740, Spearman rank correlation: r=-0.135, p=.494).

In summary, all three aphasic participants performed less accurate and slower than the control group in the three conditions 'auditory only', 'audiovisual', and 'visual only', except for WB, whose reaction time fell into the normal range for the 'audiovisual' condition. Additional speechreading speeded up the reaction times of the control group, WB, and JH, but not the reaction times of EK. JH also showed a higher accuracy with speechreading possible, while only a trend was found for WB. In the McGurk condition, the non-brain-damaged control participants produced mainly visual answers, followed by auditory answers and McGurk-type answers. None of the aphasic participants showed a preference for either answer type. The reaction times of the non-brain-damaged participants were influenced by the answer type they chose: reaction times were largest when a McGurk-type answer was given. There was no influence of answer type for either of the aphasic participants (see Figure 3.3).

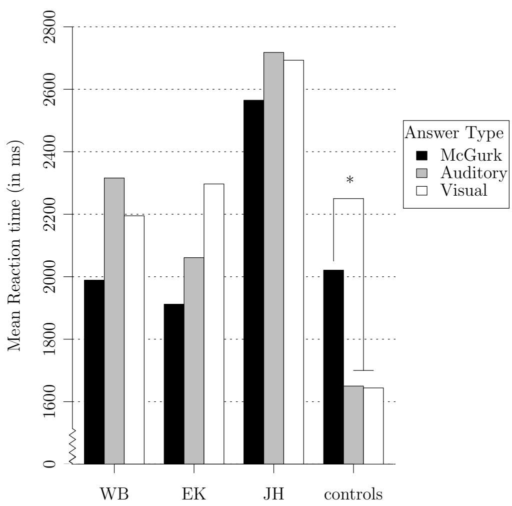


Figure 3.3: Reactiontime per answer type.
∗: Mann-Whitney-U test, p<.05.

3.4 Discussion

In the current study we investigated the processing of audiovisual speech in three participants with aphasia and a group of non-brain-damaged control participants. Several hypotheses were formulated and examined.

It was found that each of the aphasic participants performed worse in a nonword identification task than a group of non-brain-damaged control participants. This was true with 'auditory only' as well as 'audiovisual' stimulus presentation. This outcome was expected as the aphasic participants were selected on the basis of their accuracy on nonword discrimination, a task closely related to the current research task.

With regard to the reaction times, it was found that in the 'auditory only' condition all three aphasic participants answered slower than the non-brain-damaged control group. In the 'audiovisual' condition, however, the reaction times of one of the aphasic participants, WB, was within the range of the control group. Additional speechreading in this condition apparently provided sufficient information for him to increase his processing speed up to a normal level. The other two aphasic participants were, again, slower than the non-brain-damaged group in this condition. In the McGurk condition, there were three answer types possible. The patterns and reaction times for each chosen answer type will be discussed below.

One of the central questions of this study concerned the effect of speechreading on the perceptive abilities of the aphasic participants. In a previous study, it was shown that nonword discrimination improves when speechreading is possible (Hessler et al., 2010). Hence, we expected that the aphasic participants would benefit from additional visual information in an identification task as well. We compared their scores and reaction times in the 'auditory only' and 'audiovisual' condition. The later allowed for speechreading while the first only provided the auditory speech signal. WB and JH showed improved performance (higher accuracy for JH and faster processing for WB and JH) with speechreading than without, while no significant difference between the conditions was found for EK. Schmid and Ziegler (2006) also reported a group of aphasic participants that did not benefit from speechreading. They performed comparably in the auditory and the audiovisual condition of a discrimination task. While the authors assume that in those cases a stage of phonological processing was affected at which auditory and visual information have already been integrated, it seems more likely that EK had a problem in gaining information from speechreading: he was also particularly impaired in the 'visual only' condition, where he had only 24% correct answers. This is performance at chance-level. This explains why he did not perform better in the 'audiovisual' than in the 'auditory only' condition. In the 'visual only' condition, accuracy was lower and reaction times longer than in the 'auditory' condition for each aphasic participant and also for the non-brain-damaged participants. Therefore, the better performance in the 'audiovisual' condition did not solely rely on information from speechreading, but rather the combination of auditory and visual information. The difference between the 'auditory only' and the 'visual only' condition was larger for the non-brain-damaged participants than for either of the aphasic participants. This is not due to the fact that the aphasic participants do exceptionally well in the 'visual only' condition, but can be explained by their poor performance in the 'auditory only' condition.

The group of non-brain-damaged participants scored at ceiling in the 'auditory only' condition, so only reaction times were compared for this group. They showed faster processing of audiovisual than of mere auditory information. The decreased reaction times indicate that during processing all information provided was used automatically and speechreading and auditory information were integrated to speed-up processing. This supports previous reports that speechreading is not only beneficial when the auditory signal is disturbed but also in other contexts, such as cognitively demanding contents (Reisberg et al., 1987).

The 'McGurk' condition was analyzed with regard to the answer patterns of the participants. Within the group of non-brain-damaged control participants a difference between the prevalence of answer types was found. Visual answers were most common, followed by auditory answers that in turn were more common than McGurk-type answers. However, this is only true when looking at the non-brain-damaged participants as a group. The individual data given in Table C.3.2 in the appendix, show that different kinds of patterns exist: participants can have a preference for either answer type (see participants C4 and C8 for a strong preference for auditory or visual answers and participant C11 for a less extreme preference of McGurk answers) or a rather mixed pattern (see participant C7). The majority of the control participants, however, show a preference for a visual answer. Generally, a rather low number of McGurk type answers were given by the non-brain damaged control participants.

The strength of the McGurk effect differs between languages. Unlike the original findings for English (McGurk & MacDonald, 1976), the McGurk effect seems less strong for Dutch and German (De Gelder et al., 1995; Grauwinkel & Fagel, 2006; Klitsch, 2008). Another aspect potentially explaining the rather weak McGurk effect in the current study is that voiceless stops were used. McGurk and MacDonald (1976) found a weaker effect for voiceless (81%) than for voiced stops (98%). However, the Dutch phonological system does not include the sound /g/, which precludes using voiced stops when testing the McGurk effect in Dutch participants. While the rather low incidence of McGurk answers might be judged problematic, it does not influence the main concern of the current study, as this examined the reaction time patterns per answer type.

The aphasic participants show diverging patterns, with inter-individual variances. However, none of them preferred visual answers. Although this is opposed to the general pattern of the control group, the results of the aphasic participants resemble those of individual members of the control group. Differences in response patterns are quite common. The McGurk effect has been described as a phenomenon with large inter-individual differences (cf. Schwartz, 2010). Interestingly, WB and JH show only a small number of visual answers although they benefit from speechreading. Nonetheless they did not develop a bias for visual information.

The reaction times in the McGurk condition were analyzed in relation to the chosen answer type. For the non-brain-damaged participants, reaction times differed depending on what answer type was chosen. Their reaction time was the longest when they gave a McGurk type response. This could reflect additional resources necessary when a fusion of auditory and visual information is perceived. None of the aphasic participants showed a slow-down when giving a McGurk type answer. The performance of the aphasic participants is therefore not only quantitatively different (they reacted slower in all conditions, except for WB in the audiovisual condition) but also qualitatively different, which can be seen in the diverging patterns. In the following sections we will provide two different accounts for these findings, one based on the experimental findings of Soto-Faraco and Alsius (2007, 2009), the other based on the adapted TRACE model (Campbell, 1988, 1990).

Soto-Faraco and Alsius (2007, 2009) argued that their participants had conscious access to the unimodal information, and thus processed the auditory and visual information separately before integrating them to a McGurk percept.

This finding can explain the slow-down that the non-brain-damaged participants experienced: reaction times were slower due to processing two layers, the unimodal and the integrated input. We assume that the aphasic participants relied solely on multimodal processing, not gaining access to the underlying unimodal information, and therefore do not show a slow-down in processing compared to the other conditions. However, this hypothesis still needs to be tested, as we did not record whether participants had access to unimodal information in the current study. In a future study the research as carried out here could be combined with an additional task as described by Soto-Faraco and Alsius (2007, 2009), asking the participants to judge the synchrony of the stimuli. The results of this judgment may indicate whether aphasic participants indeed do not access the unimodal information of stimuli prior to fusion unlike non-brain-damaged control participants.

However, the diverging reaction time patterns can also be explained in terms of the adapted TRACE model (Campbell, 1988, 1990). If it is assumed that individuals can weigh activations from distinct input types differently, this accounts for the preference of different answer types. While some participants might value visual information more than auditory, others might have the opposite preference. As explained in the introduction, there is excitation (both bottom-up and top-down) between levels and inhibition within a level. This means that the auditory input matching /p/ activates features that in turn activate the /p/ on the phoneme level, which consecutively activates the values of other features matching /p/ as well. This inhibition is supportive in cases where the input is consistent, as it helps to reach threshold activation faster. However, in the McGurk stimuli the visual input does not match and therefore values of the seen features are activated that match /k/. The auditory input (via the phoneme level) and the visual input therefore activate different values of the same feature. Due to the inhibition principles, different values of one feature try to inhibit each other. This inhibition could play an important role with regard to the reaction times found. The mutual attempts to inhibition might delay the activation of the following processing level, the phoneme level. However, this would only be the case whenever there is no preference for either type of input. In the cases where subsequently a visual or auditory answer is given, there might have been an initial preference for this type of input, neglecting the other input type and therefore not causing a multimodal inhibition clash.

A reason why aphasic listeners did not suffer from a slow-down could then be that the inhibition (at least) on feature level is not working properly. A lack of inhibition on the lexical-semantic level of language comprehension has been reported for participants with aphasic speech perception deficits (Wiener, Connor, & Obler, 2004; Janse, 2006; Yee, Blumstein, & Sedivy, 2008). If a similar deficit applies to the feature level, activation is passed on to the next level without clashing inhibition. Therefore all answer types would be equally slow.

3.5 Conclusions

In the current study, we investigated audiovisual processing in aphasic and non-brain-damaged participants. It was found that the aphasic participants were generally slower and less accurate in identifying stimuli presented auditorily or audiovisually than non-brain-damaged participants. Two out of three aphasic participants reached higher accuracy and faster reaction times in the 'audiovisual' than in the 'auditory only' condition, indicating the beneficial influence of speechreading. Within the non-brain-damaged control group, faster reaction times were also found for the 'audiovisual' than the 'auditory only' condition. This proves the notion of primary multimodal processing. Regarding the 'McGurk' stimuli we found that, while the answer patterns differed slightly between the aphasic and the non-brain-damaged participants, we could find a difference especially in the reaction time patterns. Not only a quantitative, but also a qualitative difference was found. While non-brain-damaged participants reacted more slowly when they were subject to the McGurk illusion, this did not hold for the aphasic participants. We argued that while in undisturbed processing the listener has conscious access to the unimodal information, this might not be true for the aphasic participants. This extra processing layer might cause the slow-down in the group of non-brain-damaged participants.

These findings add to the understanding of multimodal speech processing in aphasia. While previous studies found no qualitative differences in processing (Klitsch, 2008), we refined the methodology in the current study and found different processing patterns. These differences need to be further analyzed to specify the exact steps of audiovisual processing in which the aphasic participants differ from the non-brain-damaged controls. Follow-up research, for example using event-related potentials, can answer this question. Once it is determined at what point exactly processing is disturbed, treatment can be developed to address the problems more specifically and more effectively.

CHAPTER 4

Event-Related Potentials

While the studies reported in the previous chapters employed offline behavioral techniques (combined with reaction times) only, chapters 5 and 6 report studies which make use of the neurophysiological online technique ERP (event-related potential). ERPs are calculated from recordings of the electroencephalogram (EEG). With the help of this technology, the brain processes involved in speech perception will be investigated. Therefore, not only the results of processing can be analyzed, leading to offline measures, but also information of the online processing is gained. This can provide insights over the actual processing steps that are impaired in aphasia. This chapter provides background of the technique and the components relevant in the current research.

4.1 From EEG to ERP

The variations of electrical activity of the brain can be measured with a technique called electroencephalogram (EEG). This has been reported for the first time for humans by Berger (1929). During an EEG recording, the changes in voltage between at least two electrodes are measured. These are amplified before they can be analyzed further. The output of the amplifier shows a pattern of variation in voltage over time. The amplitude can vary between -100 and +100 μV. The frequency ranges up to 40 Hz (Coles & Rugg, 1996). The EEG itself does not give information about cognitive processes, however it allows to judge the alertness of participants and serves for clinical diagnostic purposes, for instance diagnosing epilepsy. In order to investigate cognitive processes, such as language comprehension, the EEG needs to be linked to events, for example stimulus presentations. Then, the so called 'event-related potentials' (ERP) are measured. However, the brain produces random spontaneous background activity next to the relevant, stimulus-related activity. In order to isolate the stimulus-dependent activity, many stimuli of the same kind are presented and the activities linked to them are averaged. This averaging process removes the random activity and leaves the stimulus dependent activity, as the latter is constant during all trials of the same kind. A sketch of the recording of ERPs is given in Figure 4.1.

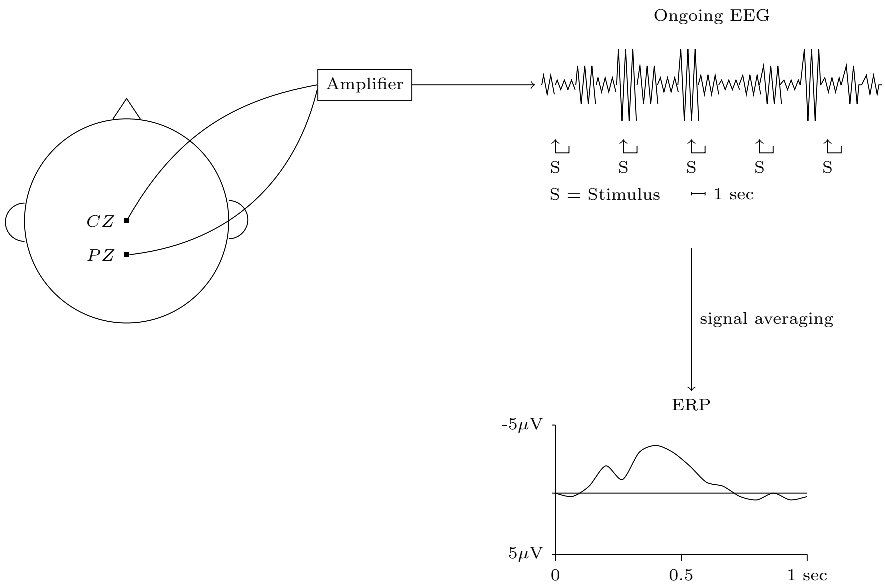


Figure 4.1: Sketch of ERP measurement, adapted from Hillyard and Kutas (1983): The activity is recorded from the scalp followed by amplification to yield the ongoing EEG with stimulus markers. The EEG is averaged over numerous repetitions based on the markers, resulting in the event-related potential.

The measured activation is generated by the post-synaptic (dendritic) activity of a substantial population of neurons, which must be synchronously active and in a certain configuration in order to yield measurable electrical fields. That means that the neurons must be aligned in a parallel orientation so that their individual fields add up to one dipolar field.[1] When recording ERPs, it must therefore be kept in mind that much of the neural activity is not measured because it occurs in configurations which are not captured. For example, the arrangement of neurons in the Thalamus does not lead to a dipolar field and therefore prevents Thalamic activity from being recorded. Therefore, it is certain that some functionally important neural processes cannot be detected using the ERP technique (Coles & Rugg, 1996). While measuring brain activity using ERPs reveals a considerable amount of information about the temporal domain (e.g. order of processing), it does not provide information about the location of this activity. Even although some activity can be specific to certain measurement areas, these do not correspond to the adjacent cortical regions, as activations generated in one area can be detected at distant locations. Source-localization techniques have been developed to infer the source of a recorded activity. These will, however, not be applied in the experiments reported here and are therefore not further discussed.

4.1.1 Electrode location

ERPs are obtained by recording the difference in voltage between two electrode sites, a measurement and a reference site. In the experiments discussed in this thesis, two linked reference electrodes are used (called 'common reference'), one of them placed on either mastoid bone. At these sites, no relevant electrical activity is recorded. The ERP per electrode is generated by first calculating the mean voltage of all electrodes in relation to the 'common reference' and then subtracting this mean voltage from each electrode's activity, resulting in recordings with respect to an 'average reference' (Lehmann, 1987). The placement of the electrodes on the scalp is described by the '10-20 system' (Jasper, 1958). An overview of this system can be seen in Figure 4.2.

The name '10-20 system' relates to the placement of the electrodes relative to the nasion, inion and pre-auricular points of the participant. The distance from nasion to inion is measured and divided in parts of 10, 20, 20, 20, 20 and 10 percent (from front to back). That means that the first electrode is placed after 10 percent of the complete length from nasion to inion. The next after

[1] A dipolar field consists of negative and positive charges between which a current flows.

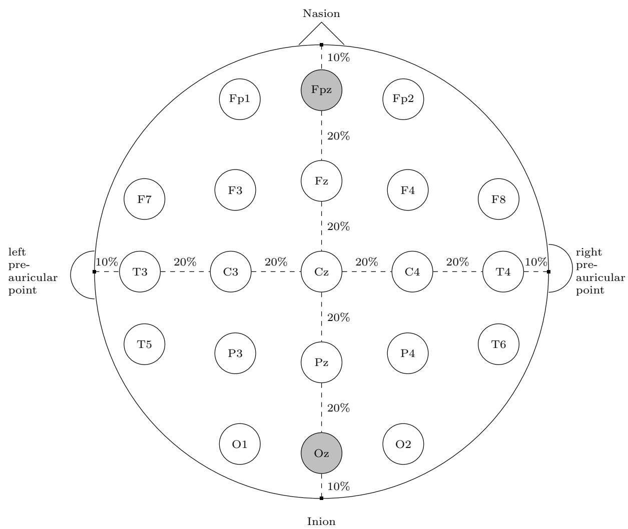


Figure 4.2: Schematic overview of the original 10-20 system and the electrode placement. Electrodes in gray were not part of the original description by Jasper (1958).

another 20 and so on, leaving the last one exactly 10 percent away from the inion. Thus, a total of five electrodes is placed on the central line from nasion to inion. The same applies to the left-right dimension, where the distance between both pre-auricular points is divided into segments of 10, 20, 20, 20, 20 and 10 percent. Again, the first electrode is placed after 10 percent of the complete distance from one pre-auricular point and the last one will be 10 percent away of the other pre-auricular point. The spaces in between are accordingly filled with electrodes, adding up to a total of 21 electrodes. The electrodes are built in caps, so that the location of most electrodes is automatically correct, when the border electrodes (front, back, left, right, central) are correctly placed. In modern electrode-caps the number of electrodes has been extended to 64 and even 128. The naming of the electrodes reflects their relative position: the electrode name starts with a letter, indicating the front- or backness of the electrode: FP for the most frontal (frontal pole), F for the second (frontal), C for the central, P for the parietal and O for the most posterior (occipital) electrode. Electrodes that are located closest to the pre-audicular points on the central and parietal rows are labeled T, as they lie above the temporal lobe. The second part of the electrode name is either a lowercase ‘z’ for electrodes on the midline, an even number for electrodes on the right scalp side or an uneven number for the left side. So, the midpoint of both the front-back line and the left-right line is marked 'Cz', while 'Fp1' is a frontal left electrode and 'O2' a posterior right electrode.

4.2 ERP components

The ERP outcomes are often described in terms of components. This raises the question 'What is an ERP component?'. ERP components have been defined in different terms. Donchin, Ritter, and McCallum (1978) pursue a functional approach to ERP definition. According to them, a component is defined by a combination of its polarity, latency, scalp distribution and sensitivity to characteristic experimental manipulations. The polarity of a component can be positive or negative. The name of a component is usually depending on the polarity (positivity or negativity) and its place in the ordering of components. A P3 is the third positivity in the course of ERP-measures, while an N2 is the second negativity.[2] In some cases, the name is defined by other characteristics, for example the name 'left anterior negativity (LAN)' is based on the scalp distribution and the polarity of the component. One common approach to isolate components is to subtract waveforms obtained in different experimental conditions. The difference between the two waveforms is the component, which is further attributed to the cognitive process present in one condition, but not in the other (Coles & Rugg, 1996). The components relevant to the current research will be discussed in the following sections.

4.2.1 Mismatch negativity (MMN)

The mismatch negativity (MMN) is an ERP-component specific to the auditory modality, which is elicited by the automatic recognition of a deviating sound. It was first reported by Näätänen, Gaillard, and Mäntysalo (1978) who used a dichotic stimulus presentation task, in which the participants had to attend to one ear and ignore the stimuli presented to the other. They were asked to push a button when they heard an occasional deviant stimulus in the sequence of identical tones in the attended ear. The deviant stimuli at both ears (regardless

[2] Terminology differs, however. While some authors prefer the notation described above, others combine the polarity with the latency in milliseconds, leading to components as P300 and N200.

if attended or ignored) led to a negativity in the ERP at 100 to 200 ms latency, which was not recorded for the standard stimuli. Näätänen et al. (1978) concluded that it was not necessary for the emergence of this negativity that the participants pay attention to the stimulus. They further argued, that the MMN can be best displayed in a difference wave, in which the wave of the standard sounds is subtracted from the wave of the deviants. According to Näätänen (1995) the mismatch negativity has several cortical generators: a bilateral auditory-cortex generator and a frontal cortex generator. He further assumes that deviant stimuli activate further auditory cortex sources and maybe also subcortical sources.

The MMN can be recorded as a response to any discriminable auditory change. It is not only found in experiments with tones but also with linguistically more complex material, like phonemes (Aaltonen, Niemi, Nyrke, & Tuhkanen, 1987; Sams, Aulanko, Aaltonen, & Näätänen, 1990), when presented auditorily. It has been reported (Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1987) that there is an influence of the inter stimulus interval (ISI) on the occurrence of the MMN. Shorter ISIs provoke larger MMNs. Böttcher-Gandor and Ullsperger (1992) however found an MMN even with ISIs from 6 and 10 seconds. It is often claimed that the MMN is specific for auditory input. However, there is evidence for a visual counterpart of the MMN: a negativity in the N2 time window with a posterior localization (Pazo-Alvarez, Cadaveira, & Amenedo, 2003).

The MMN is usually elicited in so-called oddball tasks in which a sequence of auditory stimuli is presented to the participants. Within this sequence one of the stimuli occurs frequently (e.g. 90%). This is called the 'standard'. The stimulus (or stimuli) occurring in the remaining instances are called 'deviant'. While the 90-10 distribution of stimuli is the standard distribution, also other distributions have been successfully applied. Deacon, Nousak, Pilotti, Ritter, and Yang (1998) explored different distributions. They showed that also a distribution with 70% standards and three deviants of 10% each did not result in a different MMN than those of the 90-10 distribution. They concluded that it is not the global probability of all deviants, but the probability of each individual deviant that is important for the generation of the MMN.

A distinction has been made between passive and active oddball tasks. In a passive oddball task, participants pay no attention to the stimuli, but read a book, watch a silent movie or do another non-auditory task. Auditory

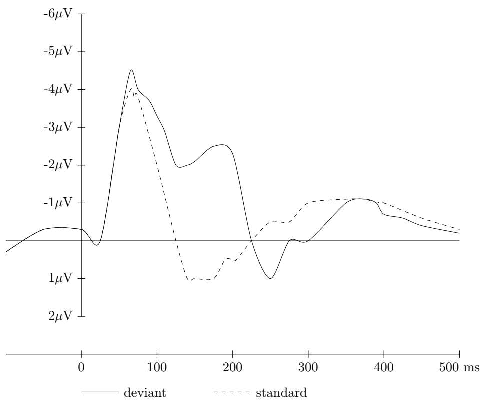


Figure 4.3: Prototypical MMN to tonal stimuli. By convention, negativity is plotted upward and positivity downward. Depicted are waveforms typically resulting from standard and deviant stimuli. The deviant stimuli provoke a more negative waveform than the standard stimuli, known as the MMN.

stimuli are presented to either both ears or only one ear. Sometimes, the participants are asked to attend one ear only. In passive oddball tasks only the mismatch negativity is elicited. Figure 4.3 displays a prototypical MMN as can be recorded to differences in pure tone pitch.

In active oddball tasks, participants are requested to attend to the stimuli. There are different tasks associated with the active oddball paradigm, such as counting the number of deviant stimuli or pushing a button when encountering one. In active oddball tasks, it is also possible to measure the MMN, but the attentional processes are reflected as well. Therefore, the difference between standard and deviant stimuli also elicits a N2b (negativity around 200 ms) and a P3 (positivity around 300 ms) (Näätänen, Simpson, & Loveless, 1982). The two attention-related components, N2b and P3, will be discussed in more detail in the following section.

It can be argued that motor activation influences the signal in those active oddball tasks that require pushing a button. It is the case that only for detected deviants a button is pushed and thus motor action is carried out. Sams, Paavilainen, Alho, and Näätänen (1985), however, showed that there is no influence of the additional motor action. They conducted a study in which two active oddball designs were compared: in the first set-up, deviant stimuli had to be counted (no motoric interference) and in the second, buttons were pushed upon encounter of a deviant stimulus. The MMN and the P3 did not differ between both conditions. The additional motor action did not alter the results. Therefore effects found between standards and deviants in the button-push set-up are not influenced by motor activity but rely solely on the frequency of occurrence.

4.2.2 Attention related ERP components

As mentioned above, active oddball tasks elicit more components than their passive counterparts. When processing an attended deviant stimulus the MMN is followed by another negativity, the N2b (Näätänen et al., 1982). This component is elicited by conscious discrimination tasks and shows an overlapping distribution with the MMN. Unlike the MMN, the N2b is not only sensitive to auditory, but also to visual deviants, if they are attended. The latency is somewhat later for the N2b than the MMN, generally peaking around 200-250ms. Novak, Ritter, and Vaughan Jr. (1992) report a MMN for pitch contrasts after 139ms, followed by the N2b 204ms after stimulus onset. The MMN and the N2b also differ slightly in their distribution: the MMN is largest in the frontal electrodes while N2b is largest centrally (Novak et al., 1992). The N2b is often followed by a large positive component, the P3 (Courchesne, Hillyard, & Galambos, 1975). Although this association is quite strong, the N2b can also occur without a P3, for example when discrimination was not successful (Sams et al., 1985).

The P3 follows the N2b and has a latency of 300 to 600ms. Novak et al. (1992) reported the P3 latency to be 342ms for the pitch contrast referred to above. The P3 has a broad distribution, but it is largest at the parietal electrodes. The P3, just like the N2b, is only elicited when the stimuli are attended and occurs with both visual and auditory deviants. The amplitude of the P3 is related to target probability: the smaller the probability to encounter a deviant stimulus, the larger the amplitude (Duncan-Johnson & Donchin, 1977). The amplitude is also dependent on the effort devoted to the task. Isreal, Chesney, Wickens, and Donchin (1980) therefore suggested that the P3 reflects the resources allocated to the task. However, the P3 amplitude is smaller when equivocation (uncertainty about perception) increases, meaning that the less sure the participants are about a deviant, the smaller the P3 becomes. According to Johnson (1984, 1986), the P3 amplitude is influenced by probability (P), equivocation (E), and resource allocation (R) in the following manner:

$$\text{P3 amplitude} = E \times (P + R)$$

The P3 is generated after stimulus categorization. Therefore the latency is increased by manipulations that postpone stimulus categorization, such as perceptual degradation of the stimuli (Luck, 2005). But the latency is not dependent on post-categorization processes. Once the stimulus has been categorized, the latency is not influenced anymore (Kutas, McCarthy, & Donchin, 1977). Figure 4.4 shows the waveforms resulting from an active oddball task with pitch difference, as described by Näätänen et al. (1978). A task like this results in the three components described above: the MMN, the N2b and the P3. Figure 4.4 also shows that the difference between the standards and the deviants is important, not the absolute positivity or negativity.[3]

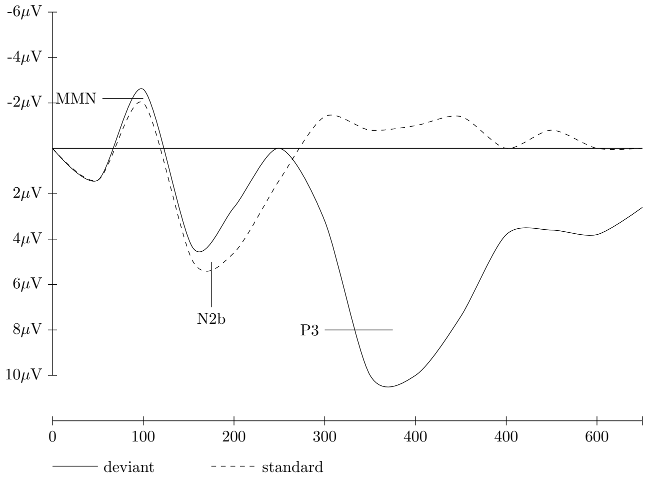


Figure 4.4: Prototypical sequence of MMN, N2b and P3 in an active oddball task (with pitch contrasts) as described by Näätänen et al. (1978). Depicted are waveforms resulting from standard and deviant stimuli. The components elicited by the deviant stimuli (MMN, N2b and P3) are marked in the figure.

[3]Note that the N2b has a positive absolute potential, however the deviant waveform is more negative than the standard one.

4.3 Phoneme processing in ERP studies

There are a number of aspects of phonological processing, which have been addressed with oddball studies. With an active oddball design, Lawson and Gaillard (1981) showed that the the peak latency and amplitude of mismatch responses were influenced by the number of phonetic features differing between standard and deviant. The N2 was found to be a good indication of phonetic distance: the larger the distance, the shorter the latency and the higher the amplitude of the N2.

Oddball designs have also been used to analyze processing differences of voice onset time and place of articulation. Several studies reported a relationship between the amplitude and/or latency of the MMN and the size of the difference, but no effect of categorical processing has been found. This means that the MMN changed continuously and did not show a larger effect at a category border (Sams et al., 1990; Kraus, McGee, Sharma, Carrell, & Nicol, 1992; Sharma, Kraus, McGee, Carrell, & Nicol, 1993; Maiste, Wiens, Hunt, Scherg, & Picton, 1995). Sharma and Dorman (1999) were, however, able to find a category effect for a voice onset time continuum.

The MMN has also been used to investigate effects of innateness of phonological contrasts: participants process native and nonnative contrasts differently (Dehaene-Lambertz, 1997; Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000; Tsui, 2000; Kirmse et al., 2008), even when the participants are advanced learners of a language, mastering the contrast in question (Peltola et al., 2003). More details can be found in an extensive review of studies using MMN paradigms to investigate phonological processing, but also non-speech auditory processing and processing on higher language levels by Näätänen, Paavilainen, Rinne, and Alho (2007)

4.3.1 Studies on audiovisual processing

Möttönen, Krause, Tiippana, and Sams (2002) conducted a magnetic encephalographic (MEG) study with audiovisual stimuli, which were either congruent or incongruent (where visual perception was supposed to influence auditory processing). The incongruent stimuli differed only in the visual part from the standard. In MEG, this kind of paradigm evokes a so-called 'mismatch field' (MMF), which is comparable to the MMN in ERP studies. Both congruent and incongruent deviants elicited an MMF. However, this MMF to the congruent deviants had a larger peak amplitude than the MMF to incongruent deviants. The MMF to the incongruent deviants, which differed only visually from the standards, had a shorter latency than that elicited in a purely visual condition without auditory input. This suggests that the interaction between auditory and visual information accelerates the detection of deviants.

The McGurk effect, which comprises a specific auditory incongruity, has also been studied with the ERP paradigm (e.g. Colin et al., 2002; Colin, Radeau, Soquet, & Deltenre, 2004; Saint-Amour et al., 2007). Colin et al. (2002) assessed the existence of an MMN evoked by McGurk percepts. The participants were eight adults without language problems. The experiment was carried out in three conditions of stimulus presentation: 'auditory only', 'visual only' (only seeing the speaker's articulatory movements) and 'audiovisual'. In the 'auditory only' and the 'visual only' conditions, the stimuli were formed by the syllables /bi/ and /gi/. One of these syllables formed the frequently occurring standard stimulus, while the other served as a deviant. Both syllables were used equally often in both functions. The audiovisual standard stimuli were syllables /bi/ or /gi/ with matching auditory and visual information. The deviant stimuli in this condition were formed out of the same auditory syllable as the standard but combined with different visual information (the other syllable). Thus, in the deviant condition, speech and lips did not match, eliciting a McGurk percept. Participants were not paying attention to the stimuli but had to perform a tactile discrimination task while listening. An MMN was found in the 'auditory only' condition and in the 'audiovisual' condition at the frontal midline electrode Fz. In the 'visual only' condition no MMN was found (neither at Fz nor at the occipital midline electrode Oz).

In a follow-up study, Colin et al. (2004) extended their earlier (2002) results to syllables with voiceless consonants. The design was similar. As in the previous study, MMNs at Fz were evoked in both the 'auditory only' and the 'audiovisual' (McGurk) condition. The authors concluded that there is an MMN for voiced as well as voiceless consonants in the 'auditory only' and the McGurk condition. As the visual mismatch within the McGurk condition was not expected to elicit a MMN, the authors concluded that the auditory perception of the participants was altered and therefore perceived as deviant although the auditory part of the stimulus was not different from the standard.

Saint-Amour et al. (2007) used an oddball design to investigate the lateralization of the McGurk-MMN. Eleven participants took part in this study.

The experiment was carried out in two conditions: 'visual only' and 'audiovisual'. All participants reported effects influenced by speechreading in the latter condition. The deviant stimuli were the auditory syllable /ba/ and the visual syllable /va/, which lead to the perception of /va/.[4] The standard stimuli were made of an audiovisual congruent /ba/. There was no MMN for the visual stimuli, but deviant and standard elicited different wave forms in occipital, parietal and central regions. The visual activity was subtracted from the audiovisual activity in order to avoid an overlay of it with the activity in the McGurk condition. After a further subtraction of the standard-wave form from the deviant-wave form, a typical MMN-like wave form was found at Fz. No MMN was found in the occipital regions.

4.4 MMN and Aphasia

Speech perception in aphasia has also been addressed using oddball designs. Aaltonen, Tuomainen, Laine, and Niemi (1993) investigated the abilities of four aphasic participants to discriminate vowels. Two of them suffered from posterior left hemisphere lesions and two from anterior left hemisphere lesions. The stimuli were made up of vowels (/i/ and /y/) and tonal sounds, presented in an oddball design. The participants were not attending the stimuli, but watching a silent movie. The two participants with anterior lesions showed MMNs both for the vowels and for the tonal stimuli. For the two participants with posterior lesions, a different picture was found. They also demonstrated MMNs for the tonal stimuli but not for the vowels. The authors give two possible explanations for this pattern: either speech-specific mechanisms were disturbed or the posterior region (where the patients had lesions) could be necessary for the generation of an MMN for all complex sound stimuli.

In a single case study, Sharma, Kraus, Carrell, and Thompson (1994) reported a similar pattern for an aphasic participant with a unilateral cortical lesion due to left hemisphere frontal-temporal-parietal craniotomy. He showed a normal MMN to pure tone pitch contrasts, but phonemic (place of articulation) contrasts did not elicit an MMN. These MMN results also reflected the behavioral performance in pitch and phoneme discrimination tasks.

[4] This is not the classical McGurk effect, but rather an effect of visual information dominating the auditory information. Nonetheless, it has been referred to as McGurk effect as well.

Auther, Wertz, Miller, and Kirshner (2000) related the MMN responses in aphasic adults to their comprehension abilities and site of lesion. In their study, 17 aphasic participants suffering from left hemisphere lesions took part. Their language comprehension level was evaluated by the Western Aphasia Battery (Kertesz, 1982). Groups were formed according to the results in the auditory subtest and the token test. Four of the participants in the group with good comprehension were classified as having Broca's aphasia, three anomic aphasia and two conduction aphasia. Two participants with good comprehension had suffered anterior lesions, one a posterior lesion and the remaining six both anterior and posterior lesions. In the group with poor comprehension, four participants were classified as having global, three Wernicke's and one Broca's aphasia. Three of them had suffered posterior lesions, one an anterior lesion and four of them a combination of posterior and anterior lesions. The experiment was conducted with an oddball design, involving the presentation of the syllables /ga/ and /da/ where /ga/ was the standard and presented 90% of the time. The authors found an MMN occurring in 89% of the participants with good auditory comprehension and only in 25% of the participants with poor auditory comprehension. Again, this was related to the site of the lesion. Good comprehension (and presence of MMN) was related to lesions that spared the temporal lobe. This correlation was, however, not perfect, as there was one participant with a temporal lobe lesion who showed an MMN. These results support the earlier findings of the same group (Wertz et al., 1998) that only 54% of the aphasic participants show an MMN to speech stimuli (compared to 100% of the non-brain-damaged control participants) and that the duration of an MMN is correlated to the comprehension abilities (as measured with behavioral tasks).

In an extensive study, Csépe et al. (2001) analyzed MMN data of four aphasic participants, two of these participants suffering from Broca's aphasia and two of Wernicke's aphasia. One participant with either syndrome had experienced a bilateral brain damage, the other an unilateral. Also, four age-, gender-, and education-matched non-brain-damaged participants took part in this study. The MMN was recorded for three conditions: contrasts in pure tone pitch, vowels and consonant-vowel-syllables. For each condition two deviant stimuli were created, which each occurred in 15% of all stimuli. Therefore the standard was presented in 70% of all stimuli. In the pure tone condition the standard was a tone of 1000 Hz, while the deviants were tones of 1050 respectively 1200 Hz. The standard vowel was an /e:/, the deviants were /i:/ and /ø:/, resulting in a difference in one phonetic dimension (height or rounding) between the standard and each deviant vowel. In the syllable condition, the standard /ba:/ was contrasted with the deviants /ga:/ and /pa:/, again exhibiting a difference in only one feature: place of articulation or voicing. The authors reported that the pure tone MMN did not distinguish between non-brain-damaged and aphasic participants. Furthermore, consonant contrasts appeared to be more vulnerable than vowel contrasts. For two of the aphasic participants (the Broca's patient with unilateral lesion and the Wernicke's patient with bilateral lesion), there was a noticeable difference between the MMNs to voicing and place of articulation contrasts: it was not present for voicing, while it was found, though abnormal, for place of articulation. The phoneme processing anomalies (as reflected by the MMN) correlated with results of a nonword discrimination task, but not with the results of a word discrimination task.

Jacobs and Schneider (2003) described a case of pure word deafness. The aim of their study was to determine whether their patient was suffering from pure word deafness or whether his problems had to be located at a pre-linguistic level. One of the measures used to assess the problems and abilities of their participant was an MMN study, as it is an objective tool providing information of the discrimination abilities of the participant. It was found that the participant showed an MMN for pitch difference (even on linguistic material), but he did not exhibit an MMN when the change was phonetic (/da/ vs. /ga/). In combination with the results of a wide range of other tests, the authors concluded that their participant suffered from pure word deafness.

As can be seen from the previously described studies, problems in language comprehension can be reflected by MMN, showing no or decreased (in duration) MMNs for phonetic differences while results did not differ from normal performance for non-linguistic (pure tone) changes.

4.5 ERP studies in the current thesis

None of the components discussed above is specific to language processing. Nonetheless, they all have been found to react to language related differences as well. Therefore, the active oddball technique is a very appropriate tool to assess phoneme processing in non-brain-damaged and aphasic participants.

As described above, the oddball design has been applied to aphasia studies. However, several questions remain. It is not yet clear whether the conscious and the automatic processing of phonemic differences show comparable impairments in aphasia. An active oddball design can reveal whether there is a difference between the automatic brain response (measured with the MMN) and the conscious processing (measured with button presses). Furthermore, the studies investigating the influence of speechreading concentrated on undisturbed processing. It is, however, interesting to address aphasic processing as well, as it has been shown that aphasic listeners benefit from speechreading (see chapters 1-3). These issues will be addressed in the following two chapters. The investigated topics will be introduced in detail in the respective chapters. Chapter 5 will report a study conducted with non-brain-damaged listeners. This will serve to verify and extend previous findings on auditory and audiovisual processing and the data will also function for comparison to the aphasic data, that are reported in chapter 6.

CHAPTER 5

Brain Correlates of Phonemic Processing and Audiovisual Integration in Non-Brain-Damaged Participants

5.1 Introduction

Language comprehension involves various processing steps, of which the analysis and identification of phonemes forms the first that is specific for language. Processing of phonemes, therefore, provides the basis of language comprehension. Often when investigating these early phonemic processes, only auditory processes are considered, while there is also an influence of visual information. The articulatory movements of the speaker, when visible, facilitate comprehension (see e.g. Sumby & Pollack, 1954; Reisberg et al., 1987). In the current study, we investigate both auditory and audiovisual processing. One of the central aims of this study is to examine the brain activity related to audiovisual processing of different phonemic contrasts. This will be done by using event-related potential (ERP) measures. We focus not only on the pre-attentive discrimination process, but also on activity related to conscious mismatch detection. Before going into more detail on the research questions and the applied methods, we provide a background on audiovisual speech processing and discuss studies using the ERP paradigm.

Language processing has been described in terms of different models. One of these is the TRACE model (McClelland & Elman, 1986). In this interactive activation model of speech perception, several levels of processing are assumed: a feature level, a phoneme level, and a word level. The three levels are fully interconnected. Units within one level are connected via lateral inhibition. Across levels, excitation takes places both bottom-up and top-down. In a model like this, distances between phonemes are defined on the basis of the feature values that differentiate them. The phoneme /p/ is, for example, closer to the phoneme /t/ than to the phoneme /k/ as the values on the features 'diffuse' and 'burst' are much more similar for the first two.

However, speech is not only perceived auditorily but also visually (speechreading). Evidence for the importance of multimodal processing comes e.g. from a study by Sumby and Pollack (1954), investigating speech perception in noise. Resistance to noise was much higher when speech was presented audiovisually rather than auditorily. More evidence in favor of multimodality in speech perception was added by the findings of McGurk and MacDonald (1976). Participants were presented with non-matching auditory and visual information and had to report their perception. Instead of answering with the auditory (/pa/) or the visual (/ka/) component of the stimulus, they frequently reported a fusion: /ta/. Information gained through speechreading was combined with the auditory information to form a percept, even though there was no necessity (e.g. due to background noise) or instruction to depend on visual information. This phenomenon is known as the 'McGurk effect'.

Campbell (1988, 1990) extended the TRACE model to incorporate audiovisual perception. The feature level has the same acoustic features as the original model, but was extended to include the visually perceived features 'mouth opening' and 'lip-shape'. All features can inhibit each other, regardless of their input modality. Thus, the activation pattern of 'lip-shape' can inhibit certain values of 'diffuse'.

5.1.1 Event-related potentials (ERPs)

Speech perception can be investigated online with event-related potentials (ERPs), studying neurophysiological activation patterns. Brain reactions to phonemic distinctions and differences between auditory and audiovisual processing can be investigated with an oddball design. In such a design, a sequence of stimuli is presented to the participants. Within this sequence, one of the stimuli, the 'standard', occurs frequently (e.g. 90%). The stimuli occurring in the remaining instances are called 'deviant'. While the 90-10 distribution of stimuli is the standard distribution, also other ratios have been successfully applied (Deacon et al., 1998). A response to the deviants shows that listeners perceive a difference and can be used to investigate what differences are perceptible during phonemic processing.

The mismatch negativity (MMN) is an ERP-component which is elicited by the automatic recognition of a deviating sound (Näätänen et al., 1978). It peaks between 100 and 200ms after the stimulus onset and is largest at frontal electrodes. It is not only found in experiments with tones, but also with linguistic materials, like phonemes (Aaltonen et al., 1987; Sams et al., 1990), when presented auditorily. It is often claimed that the MMN is specific for auditory input. But there is also evidence for a visual counterpart of the MMN: a negativity in the N2 time window. However, this negativity has a more posterior scalp distribution, suggesting that the effects are generated by different areas in the brain (Pazo-Alvarez et al., 2003).

In active oddball tasks, participants are requested to attend to the stimuli. When processing an attended deviant stimulus, the MMN is followed by another negativity, the N2b (Näätänen et al., 1982). This component is elicited by conscious discrimination tasks and shows an overlapping distribution with the MMN. The N2b is sensitive to auditory and visual deviants, if they are attended. The peak latency is around 200-250ms. The N2b also differs from the MMN in the distribution: the MMN is largest at the frontal electrodes while N2b is largest centrally (Novak et al., 1992). Often the MMN and the N2b are referred to together as the N2 (Luck, 2005).

The P3 follows the N2b with a latency of 300 to 600ms (Courchesne et al., 1975). It has a broad distribution, but is largest at the parietal electrodes. The P3, just like the N2b, is only elicited when the stimuli are attended and occurs with both visual and auditory deviants. The amplitude of the P3 is related to the stimulus probability (Duncan-Johnson & Donchin, 1977), the resources allocated to the task (Isreal et al., 1980) and the uncertainty (equivocation) of the participants (Johnson, 1984, 1986). The contribution of probability (P), equivocation (E), and resource allocation (R) to the overall P3 amplitude was formalized by Johnson (1984, 1986) in the following manner:

$$\text{P3 amplitude} = E \times (P + R)$$

None of the components discussed above is specific to language processing. Nonetheless, they all have been found to react to phonemic differences as well. Thus, the active oddball technique is an appropriate tool to assess phoneme processing.

Phoneme processing in ERP studies

Phonological processing has been addressed with oddball studies. Lawson and Gaillard (1981) showed with an active oddball design that the peak latency and amplitude of mismatch responses were influenced by the number of phonetic dimensions differing between standard and deviant. The N2 was found to be a good indication of phonetic distance: the larger the distance, the shorter the latency and the higher the amplitude of the N2. In this study, distinctions of different sizes (different number of dimensions) were investigated, but contrasts within one dimension were not looked at.

Processing differences of voice onset time and place of articulation have also been analyzed in oddball designs. Several studies reported a relationship between the amplitude and/or latency of the MMN and the magnitude of the difference between standard and deviant on a 'place of articulation' continuum, but no effect of categorical perception has been found. The MMN changed continuously between different exemplars of the same phoneme and did not show a larger effect at a phoneme border (Sams et al., 1990; Kraus et al., 1992; Sharma et al., 1993; Maiste et al., 1995). Sharma and Dorman (1999) were, however, able to find a category effect for a 'voice onset time' continuum.

Investigations of audiovisual processing have concentrated on the McGurk effect and were done with passive oddball designs. Möttönen et al. (2002) conducted a magnetic encephalographic (MEG) study with congruent and incongruent audiovisual stimuli. The incongruent stimuli differed only in the visual part from the standard. In MEG, this kind of paradigm evokes a so-called 'mismatch field' (MMF), which is comparable to the MMN in ERP studies. Both congruent and incongruent deviants elicited an MMF which had a larger amplitude for congruent than for incongruent deviants. The latency was shorter for the audiovisual incongruent deviants than for purely visual stimuli without auditory input. This suggests that the interaction between auditory and visual information accelerates the detection of deviants.

Studies investigating the McGurk effect with the ERP paradigm aimed to prove that the visual information alters auditory perception (e.g. Colin et al., 2002, 2004; Saint-Amour et al., 2007). Colin et al. (2002, 2004) found that an MMN was evoked for both auditory and McGurk stimuli, although the McGurk stimuli were auditorily identical to the standard. As the visual mismatch within the McGurk condition was not expected to elicit an MMN, the authors concluded that the auditory perception of the participants was altered and therefore perceived as deviant although the auditory part of the stimulus was not different from the standard.

Similar results were found by Saint-Amour et al. (2007) in an oddball study with visual and audiovisual stimuli. The deviant stimuli were the visual syllable /va/ and the incongruent audiovisual syllable /ba/[va] (auditory /ba/ dubbed on visual syllable /va/), which lead to the perception of /va/. The standard stimulus was the syllable /ba/ (either visual or audiovisual congruent). There was no mismatch response for the visual stimuli, but ongoing visual activity for standards and deviants. The visual activity was subtracted from the audiovisual activity in order to avoid an overlay of the response to the visual processing in the McGurk condition. A typical MMN was found at Fz. However, since there was no direct comparison with the auditory MMN, it is not clear whether the effects are in fact identical.

Most of the studies reported above were carried out with passive oddball tasks and focused on automatic processes. In the current study, we will extend the paradigm to an active oddball design, in order to investigate also conscious processing. This will provide valuable additional information, such as the response times and the accuracy rates per stimulus type. Furthermore, false alarms (for the standards) and misses (for the deviants) can be excluded from the analysis.

With the design adapted in the way described above, we aim to address the following issues:

(1) We will investigate whether the amplitude of any of the ERP responses to a phonemic deviant depends on the size of the mismatch, as has been reported for the processing of tones. The difference between standard and deviant will be within the dimension 'place of articulation', regarding the features 'diffuse', 'acute' and 'burst', as they were defined in the TRACE model (McClelland & Elman, 1986).

(2) Furthermore, we will study the integration of auditory and visual information. With the use of both congruent and McGurk type deviants, we will investigate whether the activity related to the integration of incongruent

audiovisual information is similar to the activity for audiovisual integration of congruent stimuli.

(3) Finally, we want to find out whether brain responses represent the integration of auditory and visual information by looking at audiovisual processing of congruent information. We will evaluate whether the response to audiovisual processing is more than a mere addition of the brain waves to auditory and visual processing.

5.2 Methods

In order to address the issues stated above, an active oddball experiment was carried out in four different variants: 'pure tones', 'auditory syllables', 'visual syllables' (videos of articulatory movements), and 'audiovisual syllables'. Participants were asked to identify infrequent deviant stimuli in a series of repeating standard stimuli.

5.2.1 Participants

Thirteen native speakers of Dutch (nine female) participated in this study after giving their informed consent. None of them reported neurological, language or hearing disorders. Vision was normal or corrected to normal in all individuals. All participants were right-handed. The mean age was 59 years (range 45-69).

5.2.2 Materials

In the first three sub-experiments a sequence of standard stimuli and two deviant stimuli was presented. There were 800 repetitions of the standard stimulus (80% of all stimuli) and 100 repetitions of each deviant (each 10% of all stimuli). In the audiovisual sub-experiment, a McGurk type deviant was added. Therefore, there were three deviants, each of which occurred 100 times (6.66% of all stimuli). The standard was repeated 1200 times, forming 80% of stimuli. This way the proportion of standards and deviants, as well as the number of items for each deviant type, were kept constant across sub-experiments.

The stimuli in the 'tones' sub-experiment were generated with the computer program Praat (Boersma & Weenink, 2009). The standard stimulus was a pure tone of 1000Hz, the deviant stimuli were pure tones of 1050Hz (near deviant) and 1200Hz (distant deviant). They were presented auditorily while displaying a white screen with a fixation cross in the middle to minimize eye movements during trials.

The stimuli in the remaining three sub-experiments were the standard /pa/ and the deviants /ta/ (near) and /ka/ (distant) as these syllables differ in only one dimension ('place of articulation') and are the ones involved in the McGurk effect. As described above, /pa/ and /ta/ are distinguished by the features 'acute' and 'burst', while the distinction between /pa/ and /ka/ is based on differences in 'acute', 'burst' and 'diffuse'. Overall, the distance between the latter two is larger. In the audiovisual sub-experiment, an audiovisual incongruent syllable, eliciting the McGurk effect (auditory /pa/ dubbed on visual /ka/) was added.

The syllables were spoken by a male native speaker of Dutch, who was video-recorded in a quiet room with daylight. Additionally, a light diffuser was used to avoid shading on the face for optimal visual information. The recorded image included the lower part of the speaker's face (from the lower part of the nose), the neck and the shoulders. For recording, a video camera and separate cardioid microphone were used. The video was then digitized into avi-files at a sampling rate of 48 kHz with 32-bit-stereo quantization. All stimuli were then edited with Adobe Premiere to form video files with a duration of 800ms each. As recording was done with 25 frames per second (i.e. the duration of one frame is 40ms), each file consisted of 20 frames. The video showed the speaker in rest (with a closed mouth) for 6 frames (240ms) in the beginning of each video (baseline for movement). The initial preparatory movements of the mouth lead the sound onset by 200ms, on average (range: 180-220ms).

In the 'auditory syllables' sub-experiment, the stimuli were presented with a white screen with a fixation cross replacing the speakers face. In the 'visual only' sub-experiment, the videos were played without sound, showing only the articulatory movements of the speaker. In the 'audiovisual' sub-experiment, both sound and articulatory movements were presented. The sound and articulatory movements were congruent for the standard and two of the deviants and were incongruent for the McGurk deviant.

5.2.3 Procedure

The experiment was set up as an active oddball task. Standard and deviant stimuli were presented in a semi-randomized order: each deviant was preceded by at least three and maximally five standards. Stimuli were presented with a stimulus onset asynchrony of 1500ms. Participants were instructed to pay close attention and press a button whenever they detected a deviating syllable. They were told that the first stimulus was a standard which would occur frequently. In a short practice trial, the procedure was illustrated. Response times and the number of correct detections were recorded. Furthermore, it was automatically recorded if the onset of the video was delayed, which occurred rarely, when the presentation program could not access the rather large video files in time.

All sub-experiments were split into four blocks to ensure that continuous recording time did not exceed 10 minutes, because it has been shown that the MMN decreases due to habituation after 10 minutes (McGee et al., 2001). Also, the participants needed to pay attention to the stimuli, which made regular breaks necessary.

Testing was carried out on two different days. Two blocks of each sub-experiment were presented per day. All participants started with the two blocks of the 'tones' sub-experiment. The order of the other sub-experiments was balanced between participants and recording days.

The volume of the stimuli was kept constant for all participants at 65dB. The screen refresh rate and the triggers sent to the EEG for segmentation were measured and compared with an oscilloscope. There was alignment between the refresh rate and the triggers, ensuring synchrony of the video onset with the trigger.

5.2.4 EEG recording and analysis

A 64-electrode elastic cap (Electro-Cap International) with tin electrodes was used to record the EEG. Reference electrodes were placed on both mastoid bones. A ground electrode, placed on the sternum, served as common reference. Bipolar horizontal and vertical electrooculograms (EOG) were recorded and used to correct for eye movements.

The mean impedance over all participants of the different electrodes varied from 2KΩ to 6.5KΩ. The impedance of individual electrodes per participant was kept below 10KΩ, with a few exceptions. The highest impedance of an electrode included in the analysis was 17KΩ, for one electrode of one participant. EEG and EOG signals were recorded with Brain Vision Recorder (Brain Products GmbH) and sampled at 250Hz.

The analysis of the EEG data was done with Brain Vision Analyzer 1.05 (Brain Products GmbH). In a first step, both recordings of one participant were combined to one dataset. The reference during recording was an average of all electrodes; the data were re-referenced off-line with the two mastoid electrodes as reference. The data were furthermore filtered with a low cut-off point of 0.1Hz and a high cut-off point of 50Hz. Ocular correction for blinks was carried out using the Gratton-Coles method. Semi-automatic artifact rejection was applied with the following automatic parameters: maximum voltage change on a single step was 50μV, maximum difference in values within 200ms was 200μV and the maximal amplitude was 200μV. All segments were inspected and additional irregularities were marked as artifacts. So as not to include more standards than deviants in the statistical comparison, only standards directly preceding a deviant were included in the analysis. These were also the standards which were most prototypical, as they are always preceded by at least two other standards. Trials with a delay in video presentation were excluded from the analysis, as were trials where the participants made errors (i.e. missed a deviant or responded to a standard). For all analyses, a baseline was set from 200ms before stimulus onset until the onset itself. Individual averages were calculated and served as a basis for the grand averages. An additional high cut-off filter of 10Hz was applied to the grand averages for use in figures only.

In the 'tones' and 'auditory syllables' sub-experiments, the onset was the beginning of the sound. The time windows investigated were from 120 to 160ms after sound onset for the MMN, between 200 and 240ms for the N2b and from 360 to 400ms. For the 'visual only' sub-experiments the same intervals were used, but taking the onset of visual differences as starting point. In the 'audiovisual' sub-experiment, the target onset (0) is the auditory onset which lags behind the visual onset by 200ms. The intervals were chosen with reference to the auditory onset and are equivalent to those reported above. With regard to the visual onset they are between 320-360, 400-440 and 560-600ms.

For all analyses, nine regions of interest (ROIs) were defined to evaluate for scalp distribution (frontality and laterality, each with three levels). Each

ROI consisted of two electrodes[1] for which sufficient data were available after removing stimuli with a video onset delay, incorrect answers and artifacts. Electrodes in the left and right ROIs were mirrored (see Figure 5.1).

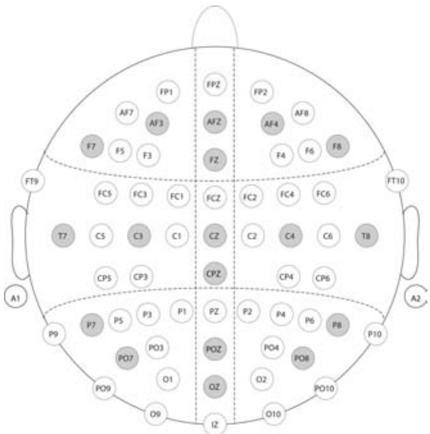


Figure 5.1: Position of the electrodes used in the current study. Electrodes marked in gray were used in the analyses. Regions of interest are indicated by dotted lines for the factors frontality (frontal, central, occipital) and laterality (left, midline, right).

For each time window, three-way repeated measures ANOVAs (3x3x3) were carried out with the factors frontality (frontal, central, occipital), laterality (left, midline, right) and stimulus type (standard, deviant1, deviant2). The scalp distribution factors were only of interest if they interact with the factor stimulus type, so significant effects limited to these two factors will not be reported. Whenever a main effect of stimulus type was found, pairwise comparisons were carried out to determine which of the three types led to the effect. While the test statistics and the significance value were calculated based on degrees of freedom corrected for sphericity (Greenhouse-Geisser correction), the uncorrected degrees are provided in this paper, for the sake of readability.

The behavioral results were analyzed with regard to the number of correct answers per sub-experiment and stimulus type and the response times when

[1]Two electrodes per region of interest were selected because in midline regions maximally three electrodes were available, with FPz being of rather poor quality, for a number of participants. Therefore, in the other regions also only two electrodes were chosen. These were the ones with best quality and, if possible, most 'prototypicality' for the region, for example the most central for the central ROI (C3 rather than FC3 or CP3).

detecting a deviant. Repeated measures ANOVAs with the factor 'stimulus type' (standard, deviant1, deviant2, McGurk) were carried out with regard to accuracy. Post-hoc pairwise comparisons were carried out when the ANOVA yielded significant results. For the response times, repeated measures ANOVAs with three levels of 'stimulus type' (deviant1, deviant2, McGurk) were carried out, which were followed-up by pairwise comparisons if significant.

5.3 Results

5.3.1 Behavioral results

Both the accuracy and the response times were recorded for the detection of the different deviants. For the standard, only accuracy can be reported, as the correct response was to not push a button. Participants showed hardly any false alarms, but missed some of the deviants. An overview of the behavioral results is given in Table 5.1.

Table 5.1: Percentage correct and mean reaction times for the different sub-experiments and stimuli. For McGurk stimuli, button pushes were counted as correct answers.

Condition	Standard (/pa/ or 1000Hz)	Deviant 1 (/ka/ or 1050Hz)		Deviant 2 (/ta/ or 1200Hz)		McGurk deviant (/pa/[ka])	
	% correct	% correct	RT	% correct	RT	% correct	RT
Tones	99.9%	94.3%	530ms	99.3%	450ms	-	-
Auditory Syll.	99.9%	74.7%	912ms	80.6%	909ms	-	-
Visual Syll.	99.9%	93.0%	701ms	96.9%	669ms	-	-
Audiovisual Syll.	99.9%	92.1%	784ms	96.1%	743ms	86.8%	782ms

The accuracy of the reactions of the participants differed per stimulus type in the auditory ($F(2,24)=15.312$, $p<0.001$), the visual ($F(2,24)=5.756$, $p<0.05$) and the audiovisual sub-experiments ($F(3,36)=11.998$, $p<0.01$), but not for the tones sub-experiment ($F(2,24)=1.605$, $p=0.229$). Pairwise comparisons revealed that there are more errors for both kinds of deviant than for the standards in the auditory sub-experiment ($p<0.01$, for both comparisons: standard – deviant /ka/ and standard – deviant /pa/). The accuracy did not differ significantly between deviant types ($p=0.074$). In the visual sub-experiment, the pairwise comparisons showed a different pattern: the accuracy for the standard is higher than for the deviant /ka/ ($p<0.01$), but does not differ significantly from the deviant /ta/ ($p=0.067$). The number of correct responses for the deviant /ka/ is also significantly lower than for the deviant /ta/ ($p<0.01$). In the audiovisual sub-experiment, all pairwise comparisons are significant on at least the 0.05 level. Accuracy is the highest for standards, followed by the deviant /ta/, then the deviant /ka/ and is lowest for the McGurk stimuli (when button pushes are counted as correct answer).

The deviants differed significantly from each other with regard to the response times in the tones sub-experiment ($F(1,12)=32.414$, $p<0.001$) and the visual sub-experiment ($F(1,12)=16.951$, $p<0.001$). The factor stimulus type also influenced the response time in the audiovisual sub-experiment ($F(2,24)=13.731$, $p<0.01$): reactions to deviant 2 (/ta/) were faster than to both deviant 1 (/ka/, $p<0.01$) and the McGurk stimulus ($p<0.01$). The latter two stimulus types did not differ from each other ($p=0.811$). In the auditory sub-experiment, there was no difference in the response times to the deviant stimuli ($F(1,12)=1.066$, $p=0.322$).

5.3.2 Pure tones

Figure 5.2 depicts the activity recorded in the 'tones' sub-experiment. In the first time window both deviants evoked a more negative response than the standard at the frontal and central electrodes. The effect is strongest at frontal electrodes (see left panel of Figure 5.2. In the second time window, both deviants elicited a more negative response than the standard at the central electrode (see middle panel). In the third time window, both deviants show a positivity at posterior electrodes, which is stronger for the more distant deviant.

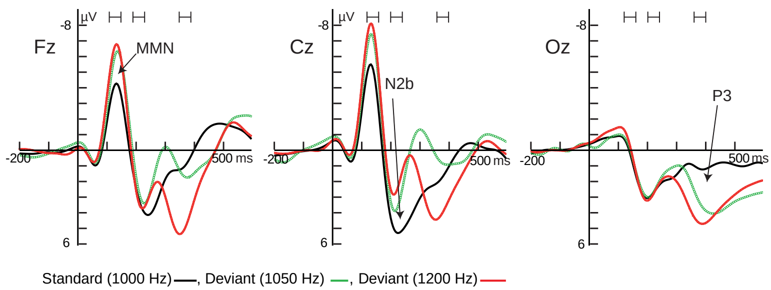


Figure 5.2: ERP of the three stimuli in the tones condition. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the figure.

MMN (120 to 160ms): A significant effect of stimulus type was found ($F(2,24)=25.674$, $p<0.001$). Pairwise analyses revealed that both deviants differed from the standard ($p<0.001$, for both comparisons), but not from each other ($p=0.126$). Furthermore an interaction of stimulus type and frontality was found ($F(4,48)=11.769$, $p<0.001$): the effect is the strongest for frontal electrodes.

N2b (200 to 240ms): In this time range, no main effect of stimulus type was found ($F(2,24)=0.833$, $p=0.445$). There was, however, an interaction between stimulus type and frontality ($F(4,48)=8.176$, $p<0.01$). This interaction implied an effect of stimulus type limited to central electrodes. When analyzing the central electrodes in a two-factor repeated measures analysis (stimulus type by laterality), a main effect for stimulus type was found ($F(2,24)=4.913$, $p<0.05$): both deviants differed from the standard ($p<0.05$, for both comparisons), but not from each other ($p=.242$).

P3 (360 to 400ms): A significant main effect of stimulus type ($F(2,24)=16.072$, $p<0.001$) was found: there were significant differences between the standard and the more distant deviant (1200Hz, $p<0.001$) and between both deviants ($p<0.01$). Furthermore, there was a three-way interaction between laterality, frontality and stimulus type ($F(8,96)=3.658$, $p<0.05$). While the effect for stimulus type appeared to be largest in the midline occipital region, there is clearly also an positive upswing visible at central electrodes which partially overlapped with the N2b negativity. In the occipital region, there was a significant main effect for stimulus type ($F(2,24)=16.097$, $p<0.001$). Also, all three pairwise comparisons were significant ($p<0.01$, for all three comparisons), indicating differences between the standard and both deviants and between the deviants.

5.3.3 Auditory syllables

Figure 5.3 depicts the ERP activity for the auditory syllables. In the first time window, no clear effect can be seen. In the second time window, there is a negativity for both deviants in the frontal and central electrodes (left and middle panels), which starts around 250ms for the frontal and around 200ms for the central electrode. In the third time window, a positivity for both deviants can be seen at all three locations, which appears larger for deviant /ta/ than deviant /ka/ at frontal and central electrodes.

MMN (120 to160ms): No main effect of stimulus type was found

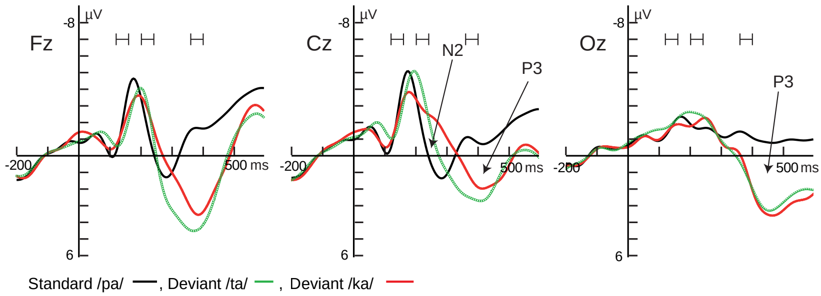


Figure 5.3: ERP of the three stimuli in the auditory syllables condition. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the figure.

(F(2,24)=0.730, p=0.474).

N2b (200 to 240ms): The brain activity depended on the stimulus type (F(2,24)=9.680, p<0.01). There was also an interaction of stimulus type and frontality (F(4,48)=4.124, p<0.05): the effect is the strongest for central electrodes. Post-hoc pairwise comparisons of the main effect revealed that both deviants differed from the standard (standard - /ka/: p<0.05; standard - /ta/: p<0.01), but not from each other (p=0.161).

P3 (360 to 400ms): We found a main effect of stimulus type (F(2,24)=34.716, p<0.001): both deviants differed from the standard (standard - /ka/: p<0.001; standard - /ta/: p<0.001). The effect was larger for the deviant /ta/ than the deviant /ka/ (p<0.01). Furthermore, we found an interaction of stimulus type and frontality (F(4,48)=4.313, p<0.05): the effect is the strongest in frontal electrodes.

5.3.4 Visual syllables

Figure 5.4 depicts the activity related to the visual syllables. Throughout the time windows, both deviants elicit far more positive waveforms than the standard. Furthermore, the deviants also differ from each other: in the second and third time window, the deviant /ta/ elicits a larger positivity than the deviant /ka/. The figure also illustrates how early the positivity starts and how long it lasts. This underlines the major effect that deviancy had in this sub-experiment. Note that unlike the previous sub-experiments, there is a clear event-related response during the baseline period related to visual input.

This is due to the fact that the measurement was time-locked to the onset of articulatory movement. The video, however, started earlier, showing the speaker in rest for 240ms before articulatory movements set in, evoking an event-related response to the visual input. Since the movement is part of the same visual event, it did not evoke such clear early components as the initial part of the event.

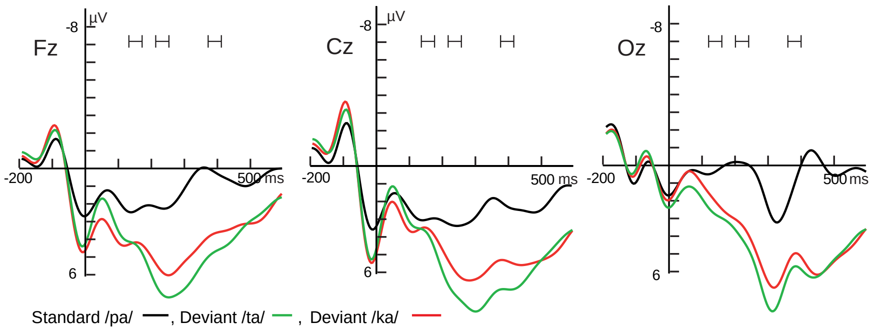


Figure 5.4: ERP of the three stimuli in the visual syllables condition. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the figure.

MMN (120 to 160ms): In the MMN-time window, a main effect of stimulus type was found ($F(2,24)=8.119$, $p<0.01$): both deviants differed from the standard (deviant /ka/: $p<0.01$, deviant /ta/: $p<0.05$). Furthermore, there was a three-way interaction between frontality, laterality and stimulus type ($F(8,96)=6.605$, $p<0.01$), indicating that this difference was largest at left frontal electrodes.

N2b (200 to 240ms): A main effect of stimulus type ($F(2,24)=27.43$, $p<0.001$) and an interaction of stimulus type with laterality ($F(4,48)=5.735$, $p<0.01$) as well as the three-way interaction between laterality, frontality and stimulus type ($F(8,96)=4.664$, $p<0.01$) were found. The post-hoc pairwise comparison of the main effect revealed that both deviants differed from the standard (standard vs. /ka/: $p<0.01$; standard vs. /ta/: $p<0.001$). They also differed from each other ($p<0.05$). The effect appeared largest at midline electrodes.

P3 (360 to 400ms): There was a significant main effect for stimulus type ($F(2,24)=53.293$, $p<0.001$). The interaction between stimulus type and laterality ($F(4,48)=12.533$, $p<0.001$) and the three-way interaction ($F(8,96)=6.111$, p<0.01) were also significant. They showed that the effect was strongest occipitally around the midline. The post-hoc pairwise comparison of stimulus type revealed that both deviants differed from the standard (p<0.001 for both comparisons). The two deviants also differed significantly from each other (p<0.01).

5.3.5 Audiovisual syllables

As can be seen from the visual sub-experiment, visual mismatches have an effect in an active oddball task. The visual input leads the auditory by 200ms. Therefore, in the time window where the visual deviance elicits a P3, the auditory deviance elicits a negativity. This had to be taken into account in the analysis of this sub-experiment. Therefore, the activity of the visual stimuli was subtracted from the audiovisual stimuli (hereafter called 'corrected audiovisual'), to remove the visual mismatch effect. For the McGurk deviant (visual /ka/, auditory /pa/), the activity of the visual deviant /ka/ was subtracted. The remaining activity should then be due to the auditory part /pa/ or additional audiovisual integration activity. The reported intervals are based on the onset of the auditory difference.

McGurk stimuli

For the evaluation of the McGurk effect, the comparisons were carried out with three-way repeated measures ANOVAs with the factors frontality (3 levels), laterality (3 levels), and stimulus type (4 levels: standard, deviant /ka/, deviant /ta/, and McGurk deviant).

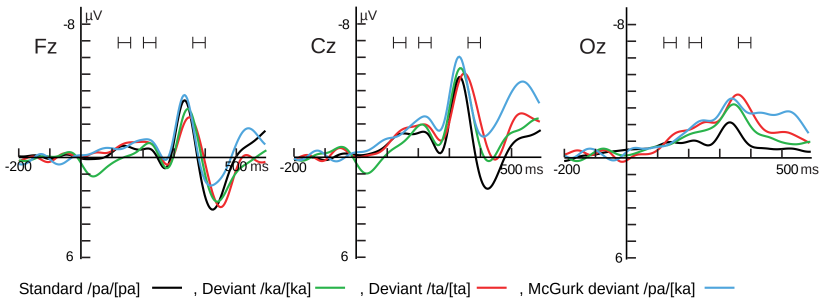


Figure 5.5: Difference waves (audiovisual - visual) for all four stimuli. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the figure.

In Figure 5.5, the waves of all four stimuli are depicted. The McGurk stimulus showed a more negative waveform than the standard at all three electrode locations, which started in the first and the second time window at occipital electrodes. In the third time window, this negativity was also found at the other electrode locations. Furthermore, the McGurk stimulus was then also more negative than the other two deviants, while they appeared quite equal in the earlier time windows.

MMN (120 to 160ms): There was no main effect of stimulus type ($F(3,36)=1.469$, $p=0.25$), but an interaction of stimulus type and frontality ($F(6,72)=4.677$, $p<0.05$). Separate two-way ANOVAs for each instance of frontality revealed a significant main effect of stimulus type at the occipital electrodes ($F(3,36)=4.191$, $p<0.05$), a trend at frontal electrodes ($F(3,36)=2.817$, $p=0.065$) and no significant difference at central electrodes ($F(3,36)=1.838$, $p=0.175$). Pairwise comparisons showed that the McGurk stimulus differed from the standard at the occipital electrodes ($p<0.01$), despite being auditorily identical. Moreover, the congruent deviant /ka/ also elicited a more negative response than the standard ($p<0.05$).

N2b (200 to 240ms): There was a main effect of stimulus type ($F(3,36)=6.908$, $p<0.01$): the response evoked by the McGurk stimulus differed from the responses evoked by the standard ($p<0.01$) or deviant /ta/ ($p<0.05$). Also, the deviant /ka/ showed a significantly more negative reaction than the standard ($p<0.01$). An interaction between stimulus type and frontality ($F(6,72)=4.303$, $p<0.05$) indicated that effect was strongest in the central and the occipital regions and less so in the frontal region. Also, an interaction between stimulus type and laterality was found ($F(6,72)=3.958$, $p<0.01$): the difference between McGurk and standard stimuli was largest around the midline.

P3 (360 to 400ms): Between 360 and 400ms, we found a main effect of stimulus type ($F(3,36)=5.878$, $p<0.01$): the McGurk stimulus differed from each of the other three stimuli: the standard ($p<0.01$), deviant /ka/ ($p<0.01$) and deviant /ta/ ($p<0.01$). The two deviants differed neither from the standard nor from each other (deviant /ka/ - standard: $p=0.536$, deviant /ta/ - standard: 0.592, /ka/ - /ta/: 0.782). A significant interaction between stimulus type and frontality ($F(6,72)=8.485$, $p<0.001$) indicated that the difference between the McGurk stimulus and the other stimuli is located in the central and occipital regions.

Comparison of auditory and audiovisual processing

The main interest in this sub-experiment was to investigate whether activation can be found that is additional to the activation from the separate auditory and visual inputs. Therefore, a further subtraction was applied: For both the auditory and the corrected audiovisual activity the difference between deviant and standard was calculated. These difference waves made it possible to compare the auditory part of the audiovisual deviance response to the pure auditory deviance response.

Figure 5.6 depicts the difference waves for both presentation modalities for each deviant. For the deviant /ka/, it can be seen that the negativity starts earlier for the 'corrected audiovisual' stimuli than for the auditory and that the positivity is larger for the auditory difference. For the deviant /ta/ too, the 'corrected audiovisual' negativity has an earlier onset than the auditory negativity and the positivity is larger in the auditory modality than in the 'corrected audiovisual' modality. This time-shift in the negativity cannot easily be directly caught when analyzing individual time windows, but only becomes apparent in visual inspection. The two presentation modalities were statistically compared in order to support the findings from visual inspection.

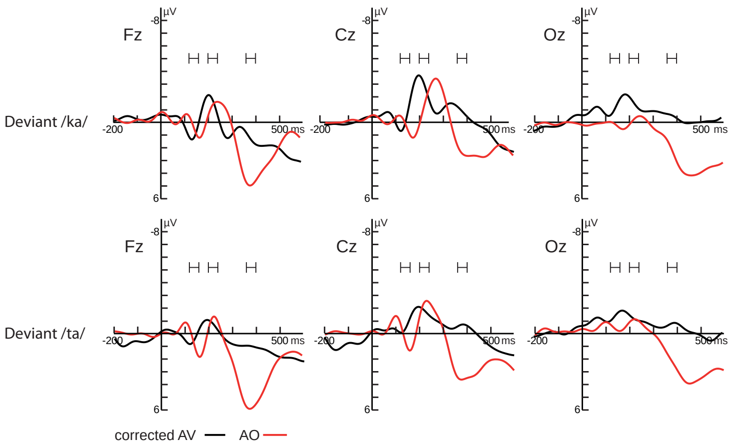


Figure 5.6: Difference waves (deviant-standard) for both deviants in the conditions with auditory presentation and the subtraction of visual presentation from audiovisual presentation. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the figure.

The time windows analyzed in this sub-experiment refer to the onset of the auditory difference, which is 200ms later than the visual difference. A four-way ANOVA with the factors presentation modality (2: corrected audiovisual and auditory), frontality (3), laterality (3), and deviant (2: deviant /ta/ and deviant /ka/) was carried out to investigate whether there were differences in the brain responses.

120 to 160ms: No main effect for modality ('corrected audiovisual' versus auditory) was found ($F(1,12)=0.428$, $p=0.525$). There was, however, an interaction of modality and frontality ($F(2,24)=12.229$, $p<0.001$). Separate three-way ANOVAs for each level of frontality revealed that there was no main effect of modality at either frontal ($F(1,12)=1.709$, $p=0.216$), central ($F(1,12)=1.027$, $p=0.331$) or occipital electrodes ($F(1,12)=3.812$, $p=0.075$).

200 to 240ms: No main effect of modality was found ($F(1,12)<1$). There were, however, interactions between modality and frontality ($F(2,24)=7.559$, $p<0.01$) and between modality and deviant ($F(1,12)=10.94$, $p<0.01$) When looking only at the deviant /ka/, no main effect of modality was found ($F(1,12)=1.63$, $p=0.226$), but again the interaction between modality and frontality emerged ($F(2,24)=7.458$, $p<0.01$). For the other deviant, /ta/, no main effect of modality ($F(1,12)=0.856$, $p=0.373$) and no significant interaction of modality with frontality were found ($F(2,24)=2.927$, $p=0.094$). Three-way ANOVAs for each level of frontality showed that no main effect of modality could be found for any level (frontal: $F(1,12)=3.123$, $p=0.103$; central: $F(1,21)=0.072$, $p=0.793$; occipital: $F(1,12)=2.507$, $p=0.139$).

360 to 400s: In this time window, there was a significant main effect of modality ($F(1,12)=35.38$, $p<0.001$). Furthermore, we found a significant interaction with deviant type ($F(1,12)=5.277$, $p<0.05$), indicating that the difference between the presentation modalities was larger for the deviant /ta/ than for the deviant /ka/.

5.3.6 Summary of results

Several analyses have been carried out to address the issues raised above. This summary provides the results with regard to each sub-experiment.

Tones: For the pure tones, the accuracy was equally high for the standard and both deviants. Participants reacted more quickly to the more distant deviant than to the less distant deviant. In the ERP, we found an MMN, which was strongest at frontal electrodes, an N2b (at central electrodes) and a

P3 (strongest at occipital electrodes) for both deviants.

Auditory syllables: For the auditory syllables, participants showed a higher accuracy for the standards than for either type of deviant. Responses to the deviants differed neither in accuracy nor in response time from each other. No ERP influence of stimulus type was found in the MMN time-window. In the period between 200 and 240ms, a significant negativity was found for the deviants (especially at central electrodes). Also, a P3 was found, which was larger for the less distant deviant /ta/ than for the deviant /ka/.

Visual syllables: In the visual sub-experiment, the accuracy was higher for the standards and deviant /ta/ than for the deviant /ka/. Also, the response time was shorter for deviant /ta/ than for deviant /ka/. The brain response to the deviant syllables was more positive than the response to the standard syllables throughout the different time-windows. The effect had a broad scalp distribution, but was largest occipitally. The deviant /ta/ evoked a more positive response than the deviant /ka/ in the time slots between 200 and 240 and between 360 and 400ms.

McGurk syllables: In the audiovisual sub-experiment, the accuracy was highest for the standards, followed by the deviant /ta/, the deviant /ka/ and was lowest for the McGurk deviant when the correct answer is regarded as pushing a button. Also, response times were shortest for deviant /ta/. Deviant /ka/ and the McGurk deviant evoked equally fast responses. Responses to the incongruent McGurk syllables were significantly more negative than to the congruent standards at occipital electrodes in all three time-windows. After correction for visual activity, the McGurk stimulus elicited also more negative responses than the deviants which were deviant in both the visual and auditory modalities (/ka/ and /ta/) between 360 and 400ms.

Audiovisual versus auditory syllables: The comparison of the brain responses between the auditory part of the audiovisual syllables (after correcting for the deviance response to visual syllables) and the pure auditory syllables revealed that there was a difference between the two presentation modalities, specifically in the latest time-window (360-400ms), with a larger positivity for the auditory only syllables. Also the scalp distribution of the positivity differs between both modalities. While the positivity is limited to frontal electrodes for the corrected audiovisual syllables, it can be found across the scalp for the auditory syllables.

5.4 Discussion

In this section, we will discuss the results reported above. First, we will discuss the results of the 'tones' sub-experiment, which served as a control for the interpretation of the other results. After that we will address three issues: (1) the degree of deviance in phonemic contrasts, (2) the brain correlates of the McGurk effect, and (3) the effect of audiovisual integration on phonemic processing.

5.4.1 Processing of pure tones

For the pure tones, the participants had no problems detecting both deviants. The more distant deviant (1200Hz) was however detected faster than the less distant one (1050Hz). This decrease in reaction time might reflect a higher certainty about the decision, caused by the larger physical difference.

The ERP findings in the 'tones' sub-experiment resemble the classical findings of for example Sams et al. (1985) for active oddball designs, which also determined the choice of time windows. The effects in the current study are not necessarily limited to the time windows used in the analysis, but these were considered most prototypical. In order to compare the measurements in the other sub-experiments to those taken for the 'tones', the choice of time windows was consistent throughout all sub-experiments. For the 'tones', the three components which were expected were found: the MMN, especially at frontal electrodes, the N2b at central electrodes and a large P3 at occipital electrodes, which was more pronounced for the more distant deviant. Participants showed faster responses indicating a higher certainty for the more distant deviant. The larger P3 amplitude for this deviant can thus be explained by the higher certainty (cf. Johnson, 1984, 1986). The factor probability, which Johnson also mentions as affecting the amplitude of the P3, is equal for both deviants in the current set-up. The third factor 'resource allocation' might play a role as well, predicting a difference in amplitude opposite to the recorded one.

This sub-experiment served as a control measure to test the setup of the experiment. Since we were able to find all expected components, we can conclude that the parameters for our recording and analysis are suitable for the planned analyses in the experimental sub-experiments. Moreover, the results of this sub-experiment indicate that the participants did not suffer from any auditory problems, understood the task and responded as expected.

5.4.2 Influence of degree of deviance in phoneme processing

The first speech-related sub-experiment was the 'auditory syllable' sub-experiment. Previous ERP research on phoneme processing concentrated on automatic, unconscious processing, using passive oddball designs (Sams et al., 1990; Kraus et al., 1992; Sharma et al., 1993; Sharma & Dorman, 1999) and does therefore not add to the understanding of attention-related processes. Lawson and Gaillard (1981) conducted an active oddball study, however they used rather large contrasts between standard and deviant. In the current study, we chose stimuli that differed in only one dimension ('place of articulation'), but to different degrees. The deviants, /ka/ and /ta/, were presented together with the standard, /pa/. Based on the TRACE model (McClelland & Elman, 1986), it was postulated that the difference between /pa/ and /ta/ is smaller than the difference between /pa/ and /ka/, when considering the overall difference of the three features 'acute', 'burst' and 'diffuse'. The phonemes /p/, /t/, and /k/ differ, for example, in their burst qualities: /p/ has a fairly faint burst that is scattered over a wide frequency range. The burst of /t/ is in a rather high frequency range and somewhat more intense than the one of /p/. The burst related to /k/ is in a middle frequency range and most intense. It is also longer than that of /t/, which in turn is longer than that of /p/ (Ladefoged, 2001). Therefore, /p/ and /t/ are closer to each other regarding the burst qualities than /k/ and /p/. This is not reflected in the behavioral results: there was no difference in accuracy nor in response times between the two deviants.

No MMN was found in this sub-experiment. In the designated time window, there was no difference between the deviants and the standard. However, there was a significant difference between 200 and 240ms (the N2b time window). Since this negativity started somewhat earlier than the time window we analyzed, it is most likely that it is a non-differentiated N2, consisting of both MMN and N2b influences. Sams et al. (1990) also found an N2 for phonemic contrasts in an active oddball design with no distinction into frontal early MMN and central later N2b. We found no difference in the responses to the two deviants in the first two time windows. Thus, in the phase of automatic processing, both contrasts are processed equivalently. When comparing differences in one and two dimensions, Lawson and Gaillard (1981), however, found an influence on the N2 amplitude: the amplitude was higher for a difference in two phonemic dimension than for a difference in one. In the current study the two contrasts were within the same phonetic dimension, ‘place of articulation’.

A large P3 caused by attention-related processes involved in difference detection was found. The amplitude was larger for the deviant /ta/ than for the deviant /ka/. As explained above, the distance between /pa/ and /ta/ is smaller than that between /pa/ and /ka/. This means that the smaller difference elicited the larger amplitude, unlike in the ‘tones’ sub-experiment.

As for the ‘tones’ sub-experiment, the probability of both deviants was equal. However, in this sub-experiment also the equivocation did not differ, as the response times were almost identical. Therefore, only the third factor mentioned in Johnson’s formula (1984, 1986), which was called ‘resource allocation’ distinguishes the two deviants. This means that the contrast needing most resources elicits the highest amplitude and that is in this case the deviant /ta/. ‘Resource allocation’ is a rather unspecific term that does not explain which neuropsychological process actually causes the influence on the amplitude. In this case, the ‘resource’ in question can probably best be described as the attention necessary to detect different types of deviants.

In the visual sub-experiment, the deviants elicited more positive responses than the standard in all three time windows with no sign of a mismatch negativity as in the auditory modality. The onset of this positivity is rather early, starting in what is considered to be the pre-attentive time window for auditory processing.

The ERP response in the P3 time window showed a positivity which was larger for /ta/ than for /ka/. Behaviorally, there is a difference between both deviants as well: the accuracy was higher and the reaction time shorter for the deviant /ta/ than for the deviant /ka/, indicating a higher certainty. This is comparable to what was found for the tones: the stimulus with the higher uncertainty evokes the larger component. It differs, however, from the results for auditory syllables, where no behavioral difference between the deviants was found.

5.4.3 Audiovisual processing

When studying audiovisual processing, not only brain responses to the auditory and the visual information are recorded but also activity related to the integration of both. These effects were studied in the audiovisual sub-experiment. In order to investigate the process of integration, we looked at McGurk stimuli and compared them to audiovisual congruent stimuli. Furthermore, we compared the correlates of audiovisual perception to those of auditory perception to determine whether the beneficial effect of audiovisual speech is represented in neural activity as well.

The behavioral results of the audiovisual sub-experiment show that /ta/ was detected as a deviant more easily than /ka/. The accuracy was lower and the response time was longer for the latter. This resembles the findings in the visual sub-experiment rather than the auditory results, in which both were equally detectable, and probably reflects the contribution of visual information.

The onset of the visual difference and the auditory difference were 200ms apart. Therefore, visual cues were picked-up earlier and the components evoked by the visual mismatch can overlay those related to the auditory mismatch detection. In an active design, due to the invested attention, a positivity related to the visual difference is expected. This is what we found and reported above for the visual syllables. Because the onset of the auditory difference is 200ms later than the onset of the visual difference, any auditory negativity might be covered by the large visual positivity. Therefore, we subtracted the visual activity from the audiovisual and then compared the remaining auditory response to the auditory syllables. Therefore, in the discussions below 'audiovisual' refers to the audiovisual activity after subtraction of the visual activity.

Brain correlates of the McGurk effect

The McGurk effect is a special case of audiovisual integration, as even though the auditory and the visual input do not match integration takes place. Earlier studies using the ERP paradigm concentrated on showing that the McGurk effect can elicit an MMN (Colin et al., 2002, 2004; Saint-Amour et al., 2007). As the MMN is regarded to react only to auditory mismatch detection, the authors assumed that the auditory perception is altered by the misleading visual information. However, these studies were all carried out with a passive oddball design, providing no information about the actual perception of the participants and limiting the analysis to components related to automatic processing.

In the current study we extended the paradigm and compared the components elicited by McGurk stimuli to those elicited by congruent audiovisual stimuli. Only the McGurk stimuli, which participants perceived as deviant (indicated by pushing a button), were included in the analysis. It is therefore clear that the participants did not perceive the auditory part of the McGurk syllable. They could either have an altered auditory perception or have reacted to the visual difference. As explained above, we subtracted the activity recorded for the visual syllables from the activity of the audiovisual syllables. For the McGurk stimuli, this means that the visual activity of the deviant /ka/ was subtracted from the measured audiovisual activity. Therefore, only the auditory part, which did not differ from the standard, was taken into the analysis.

The McGurk stimuli elicited a more negative wave than the standard stimuli in all analyzed time windows. In the latest time window, this activity was more negative than that related to congruent audiovisual deviants. The difference with the standards is noteworthy because standards and McGurk items did not differ auditorily. The only differences were in the visual part and in the result of audiovisual integration. As the visual activity had been subtracted, the standard and the McGurk items were actually physically the same. The response differed, however, substantially. As audiovisual processing was necessary for all stimulus types in this sub-experiment, the additional activation must be specifically caused by the integration of non-matching information. Möttönen et al. (2002) also found a difference between congruent and incongruent deviants in their (passive oddball) MEG study, which was limited to the right hemisphere. Processing of incongruent stimuli might be more complicated than processing of congruent stimuli: for congruent stimuli, both auditory and visual information contribute to the identification of the correct phoneme. For the incongruent stimuli, however, contradictory information is received, demanding more effort to select the matching phoneme and increase uncertainty about the given response.

Effects of audiovisual integration

In order to measure the effect of integration on processing, difference waves between the deviants and standards were calculated for the corrected audiovisual and for the auditory activity. As the audiovisual brain responses were corrected for the visual activity, any difference between the two presentation modalities should be due to the integration effects.

As shown in Figure 5.6, the observed negativity has an earlier onset for the 'corrected audiovisual' than the 'auditory' stimuli. Möttönen et al. (2002) found comparable results in an MEG study, in which they reported a shortened latency for audiovisual differences (compared to visual differences). This is another indication that audiovisual information facilitates processing. There is a faster response to the mismatch detection, when both auditory and visual information are present than when only the auditory information is provided. The direct influence of the visual information on the activity had been subtracted from the wave. Therefore, the remaining effects were due to audiovisual integration processes. The activity for the audiovisual stimuli is thus more than a mere addition of auditory and visual activation.

The positivity in the P3 time window was much larger for the purely auditory stimuli than for the 'corrected audiovisual' stimuli. This could reflect the difference in required 'resources' (such as attention): for the processing of audiovisual differences less attention is required. Therefore the amplitude of the P3 is smaller. This implies that the integration of audiovisual information does not come with a 'cost', but rather eases processing, resulting in a smaller P3 amplitude. Another possible explanation is based on the fact that the visual information leads the auditory by 200ms. Hence, the mismatch is first detected visually eliciting a P3. Once this has happened, there is no necessity to do any further mismatch processing based on the auditory input. Therefore, no P3 related to the auditory part of the input is recorded. Furthermore, the smaller P3 amplitude could be due to an overlay of a negativity related to integration. This would, however, imply that integration is a rather late process. This does not seem to be the case, as the effects seem to occur earlier in the audiovisual modality.

5.5 Conclusions

In this paper we addressed three main issues. We investigated whether the components representing automatic and conscious processing differed between two distinct contrasts (/pa/ vs. /ta/ and /pa/ vs. /ka/) as they do for tones, which we used as a control measure here. This was not the case for the components related to automatic processing, but it was true for the P3, representing conscious mismatch detection. We concluded that the smaller the difference is, the more attention is needed to detect the deviant and the larger the amplitude becomes. Since this was only tested with one contrasts, some caution needs to be paid in drawing conclusions. It is recommendable to investigate more contrasts to confirm these findings.

The second issue we addressed was whether the processing of McGurk stimuli differed from the processing of congruent audiovisual material. The

McGurk stimuli elicited a more negative waveform than congruent standards and deviants, which cannot be explained by physical differences. More difficult integration is, therefore, the most likely explanation for this additional activity.

Finally, we addressed the effect of audiovisual integration. We investigated whether audiovisual processing provoked responses differing from those of a summation of auditory and visual processing. A comparison of activity due to auditory stimuli and activity due to audiovisual activity (after subtracting the visual activity) showed activation patterns differed, with a diminished P3 and a shorter latency for the audiovisual stimuli. This indicates that audiovisual integration facilitates processing.

The current study emphasized the influence of audiovisual processing on comprehension. While it was known from behavioral studies that additional visual information facilitates (Sumby & Pollack, 1954; Reisberg et al., 1987) and influences (McGurk & MacDonald, 1976) comprehension, the present results from electrophysiological measures strengthen the claim that for audiovisual processing the whole is more than the sum of its parts.

CHAPTER 6

Using ERP measurements to investigate speech perception in aphasia — a reliable approach?

6.1 Introduction

The previous chapter focused on the brain reactions of healthy speakers when perceiving audiovisual speech. In the current chapter it is first investigated whether the same experiment as described in chapter 5 for non-brain-damaged participants can be successfully carried out with aphasic participants. If this is the case, their results will be compared to those of the non-brain-damaged participants.

Chapter 4 provided an overview of the different ERP components elicited in oddball tasks and discussed studies investigating the McGurk effect with this paradigm. In chapter 5, a more detailed background on research into phonemic processing and audiovisual integration with ERPs was provided. Oddball paradigms were also used to investigate aphasic phonemic processing (Aaltonen et al., 1993; Wertz et al., 1998; Sharma et al., 1994; Auther et al., 2000; Csépe et al., 2001; Jacobs & Schneider, 2003, see also chapter 4). All of these studies used passive oddball designs and concentrated on the mismatch negativity (MMN) component to investigate automatic, pre-conscious processing. It was found that the MMN was intact for processing of pure tones, while it was diminished or absent for phonemic stimuli (Aaltonen et al., 1993; Sharma et al., 1994; Csépe et al., 2001; Jacobs & Schneider, 2003). Wertz et al. (1998) and Auther et al. (2000) found a correlation between the amplitude of the MMN and the results of the comprehension parts of the Western Aphasia Battery (Kertesz, 1982). None of these studies, however, combined the ERP recordings with a task, in order to differentiate correct and incorrect reactions. Furthermore, the use of passive designs limited the analysis to the MMN, precluding from the evaluation of attention-related processing, as can be captured by the P3 component. So far, audiovisual integration has only been examined with the use of ERPs in non-brain-damaged speakers, but not for aphasic participants.

As an active oddball task has not been carried out with aphasic participants before, it will first be investigated whether the chosen design yields reliable results. In order to do so, a condition with 'pure tones' will be carried out as a pilot task. The participants have no deficits in the processing of pure tones, therefore their results should be comparable to those of the non-brain-damaged participants in chapter 5. If this baseline condition is successful, several issues will be addressed: it will be investigated whether the aphasic participants show brain responses comparable to those reported for the non-brain-damaged control participants when processing phonemic differences. Both pre-conscious (MMN) and attention-related (N2b and P3) components will be studied. Furthermore, audiovisual integration will be addressed. The current study will investigate whether there is also specific activity related to the integration of auditory and visual information, as has been found for non-brain-damaged participants. Another aim of this study is to analyze which levels of processing are affected by the brain damage. If the deficit leads to a diminished or absent MMN for the undetected deviants, processing is already deficient at an automatic level.

6.2 Methods

An active oddball design has been chosen to address these issues, as it provides information about the accuracy of the participants. Therefore, brain reactions in response to detected and undetected deviants can be compared. The procedure will be as described in chapter 5 for the non-brain-damaged participants.

A short summary of the design can be found below.

6.2.1 Participants

All three participants were native speakers of Dutch, right-handed, and reported normal hearing. Vision was normal or corrected to normal. The aphasic participants were tested by their speech and language therapists with the Aachen Aphasia Test (AAT) for Dutch (Graetz et al., 1992). Subsequently, they were tested with two subtests of the PALPA test battery (Bastiaanse et al., 1995): auditory discrimination of nonwords and auditory discrimination of words.[1] Furthermore, two of the aphasic participants (WB and TB) also took part in the discrimination experiment described in chapter 2. WB also participated in the identification task described in chapter 3. The general description of the aphasic participants and the results on the described tests are presented below.

WB

WB is a 59 year old male who worked as a sales director until he had a left hemisphere ischaemic CVA at age 45, 166 month prior to testing. He was diagnosed with Wernicke's aphasia. He had 37 errors on the Token Test and his scores for repetition, written speech, naming, and language comprehension were 66/150, 58/90, 100/120, and 94/120 respectively. WB had problems both in the auditory word and nonword discrimination tasks: he scored 56/72 correct on the nonword discrimination task and 65/72 on the word discrimination.

In the discrimination experiment, WB exhibited problems in nonword discrimination especially for small differences (in one phonetic dimension), which were most profound for voicing, but also affected distinctions in place of articulation. He was overall better when speechreading was possible than with pure auditory stimulus presentation. In the nonword identification task, he performed worse and answered slower when only auditory information was provided instead of audiovisual. In the McGurk trials he chose the McGurk answer in 50% of the trials.

TB

TB is a 48 year old female, who has done promotional work next to being a housewife. She had an ischaemic CVA in the area of the left arteria cerebri

[1] A short explanation of the subparts of the AAT and the PALPA was given in chapter 3, 57

media 26 months prior to testing. She suffered from a paresis of the right arm and was diagnosed with global aphasia on the basis of the AAT. She had 33 errors on the Token Test and her scores for repetition, written speech, naming, and language comprehension were 71/150, 0/90, 0/120, and 53/120, respectively. The speech and language therapist furthermore diagnosed apraxia of speech. TB made two errors (70/72 correct) in the discrimination of words and four errors (68/72 correct) in the discrimination of nonwords.

TB also participated in the discrimination study. She had problems with distinctions on one phonetic dimension, affecting especially the dimensions voicing and place of articulation. Her scores improved, when stimulus presentation was audiovisual rather than auditory only.

DM

DM is a 67 year old male who worked as a radio officer until his retirement. He had an ischaemic CVA in the area of the left arteria cerebri media eleven months prior to testing. He suffered from a right-sided facial paresis and a mild hemiparesis of the right arm and leg. Based on subparts of the AAT, the speech and language therapist diagnosed a moderate mixed aphasia. He had 41 errors on the Token Test and his scores for repetition, written speech, naming, and language comprehension were 81/150, 51/90, 66/120, and 83/120 respectively. Discrimination of words was not impaired (71/72), but DM had problems in discriminating nonwords (56/72).

As DM was recruited for this study only, he was not included in the analysis of the experiments discussed in the previous chapters. Nonetheless, both the discrimination and the identification experiment were carried out. However, no reaction time data are available for him. His performance on the discrimination task was impaired for contrasts in one dimension, showing errors only to differences in voicing. On the identification task he showed a rather low number of McGurk type answers (6.9%), with a clear preference to report the visual part of the stimulus. His performance for stimuli presented auditorily and audiovisually was equally impaired (70% correct). His accuracy for visual stimuli was slightly better (80% correct), which resembles the performance of the healthy listeners (cf. chapter 3). His visual performance and his preference for the visual information in incongruent trials did, however, not lead to good audiovisual integration, as he did not benefit from the additional information from audiovisual stimulus presentation.

The three aphasic participants were all in a chronic phase of aphasia.

They showed mild or moderate impairments in the discrimination and the identification of phonemes. While WB and TB benefited from speechreading, this was not the case for DM.

6.2.2 Materials and procedure

The materials of the current study were the same as described in chapter 5. They consisted of pure tones of 1000, 1050, and 1200Hz in the 'pure tones' condition, the syllables /pa/, /ka/, and /ta/ in the remaining three conditions, as well as an additional McGurk syllable (auditory /pa/ dubbed on visual /ka/) in the 'audiovisual condition'.

The procedure was also the same as in the experiment with the non-brain-damaged participants, described in chapter 5. An active oddball task was presented: participants had to detect deviant stimuli in a sequence of repeating 'standard' stimuli. This task was presented in four conditions: 'pure tones', 'auditory syllables', 'visual syllables', and 'audiovisual syllables'. The 'pure tones' condition was carried out as a pilot condition, in order to test the applicability of the design to an aphasic population.

6.2.3 Analysis

Due to the small number of participants and their diverging behavioral performance, the analysis was performed for each aphasic participant individually. The behavioral performance was analyzed with regard to the accuracy and the reaction times per stimulus type and condition and then compared to the results of the group of non-brain-damaged participants. Accuracy or reaction times that differed two standard deviations or more from the non-brain-damaged participants were considered deviant. The EEG was recorded and preprocessed as described in chapter 5. The statistic analysis of the ERP was carried out for each participant individually. Therefore, individual trials rather than grand averages formed the basis of the analyses. Repeated measures ANOVAs were carried out with the factors *stimulus type* (standard, deviant 1, deviant 2, and, where applicable, ignored deviant), *frontality* (frontal, central, occipital) and *laterality* (left, midline, right) for three time windows (MMN: 120-160ms, N2b: 200-240ms, and P3: 360-400ms). For the comparison of audiovisual and auditory stimuli, the same subtraction of visual activity as explained in chapter 5 was carried out. The results of interactions are only mentioned when they were significant. The test statistics and the significance level were determined based on the values corrected for sphericity (Greenhouse-Geisser correction). However, for readability, the degrees of freedom that are reported here, are the uncorrected ones.

For each participant, the number of included trials was chosen based on the availability of good quality trials. Conditions with fewer than 24 usable trials were excluded from the analysis. As the repeated measures design requires an equal number of trials for all conditions, the following procedure was applied to select trials: for each participant the lowest number of usable trials was established (yet at least 24). This was multiplied by two to determine the number of trials to include (thus, at least 48). In conditions with more usable trials, trials were selected on a random basis. If there were fewer than 24 trials, the condition was excluded. In conditions with more trials than the minimum, but fewer than twice the minimum, the missing values were replaced by the average of the condition. Finally, the order of all trials per condition was randomized before statistic comparison. An overview of the number of valid trials per stimulus type is given in the appendix (Table E.1.1).

6.3 Results

The results of the group of non-brain-damaged participants are given for comparison in the relevant sections. A more detailed description and discussion of these results can be found in chapter 5.

6.3.1 Results in the tones condition

The 'pure tones' condition served as a baseline condition in order to establish whether the paradigm is applicable to aphasic participants. The behavioral results of all three aphasic participants and the mean of the non-brain-damaged participants are provided in Table 6.1.

Behavioral results

WB's accuracy did not differ from the non-brain-damaged participants for the standard stimuli and the 1200Hz deviant. His performance for the 1050Hz deviant differed more than two standard deviations from the control group. The accuracy of neither TB nor DM differed from that of the non-brain-damaged

Table 6.1: Percentages correct and mean reaction times for the different stimuli in the tones condition for the three aphasic participants. For comparison, the mean and the standard deviation of the results of non-brain-damaged participants are also provided. Performances differing more than two standard deviations from the mean of the non-brain-damaged controls are marked with an asterisk.

Condition		Standard (1000Hz)	Deviant (1050Hz)		Deviant (1200Hz)	
		% correct	% correct	RT	% correct	RT
WB		98.1%	56%*	597ms	97%	472ms
TB		99.8%	100%	396ms	100%	345ms
DM		100%	99%	427ms	100%	369ms
Controls:	mean	99.9%	94.3%	530ms	99.3%	450ms
	SD	0.12%	14.4%	75ms	1.37%	76ms

controls. The reaction times of all three participants were inside the limit of two standard deviations from the mean of the control participants.

ERP results

For the non-brain-damaged participants, an MMN was found for both deviants, which was strongest at frontal electrodes. Furthermore, both deviants elicited an N2b (only at central electrodes) and a P3 (strongest at occipital electrodes).

For WB, the brain responses to the different stimulus types did not differ significantly from each other in the time windows of 120-160ms ($F(3,141)=1.534$, $p=0.214$) and 200-240ms ($F(3,141)=2.011$, $p=0.147$). In the third time window (360-400ms), there was a significant main effect of stimulus type ($F(3,141)=4.079$, $p<0.05$). 'Ignored deviants' elicited significantly more negative results than the standard ($p<0.01$) and the 1200Hz deviant ($p<0.001$). Interactions with laterality ($F(6,282)=2.568$, $p<0.05$) and frontality ($F(6,282)=2.759$, $p<0.05$) indicated that this effect was largest in midline and right hemisphere electrodes and in the frontal region (at least for the difference with the standard stimuli). In summary, WB did not show any expected effects of mismatch detection in either time window.

For TB, there was no significant main effect for stimulus type in the time window of 120-160ms ($F(2,242)=1.915$, $p=0.155$). However, there was an interaction of stimulus type and frontality ($F(4,484)=4.208$, $p<0.01$) and stimulus type and laterality ($F(4,484)=3.527$, $p<0.05$). Separate repeated measures ANOVAs for each instance of frontality revealed that, for occipital electrodes, there was a main effect of stimulus type ($F(2,242)=5.269$, $p<0.01$): the 1050Hz deviant elicited more negative responses than the standard ($p<0.01$) and the 1200Hz deviant ($p<0.05$). Furthermore there was a significant effect of stimulus type at midline electrodes ($F(2,242)=3.128$, $p=0.05$): the 1050Hz deviant evoked a more negative response than the standard ($p<0.05$) and the 1200Hz deviant ($p<0.05$). In the second time window (200-240ms), there was a main effect of stimulus type ($F(2,242)=7.606$, $p<0.01$): both the standard and the 1050Hz deviant elicited a more negative response than the 1200Hz deviant ($p<0.05$ and $p<0.001$, respectively). Between 360 and 400ms, there was no main effect of stimulus type ($F(2,242)=2.286$, $p=0.111$), but interactions between stimulus type and frontality ($F(4,484)=6.381$, $p<0.01$) and stimulus type and laterality ($F(4,484)=3.097$, $p<0.05$). Separate repeated measures ANOVAs were carried out for each instance of frontality and each instance of laterality. In the occipital region, there was a main effect of stimulus type ($F(2,242)=6.437$, $p<0.01$): the 1200Hz deviant elicited a more positive response than both the standard ($p<0.01$) and the 1050Hz deviant ($p<0.01$). For neither instance of laterality significant main effects of stimulus type were found. In summary, TB showed a negativity in the MMN time window, with a different distribution than the non-brain-damaged participants. She furthermore showed a positivity in the second time window (where a negativity was expected). In the P3 time window, she showed a P3 comparable to that of the non-brain-damaged participants for the more distant deviant, but not for the 1050Hz deviant.

For DM, there was no effect of stimulus type in the first time window (120-160ms: $F(2,338)=0.713$, $p=0.472$). Between 200 and 240ms, a main effect of stimulus type was found ($F(2,338)=4.948$, $p<0.01$): the 1200Hz deviant elicited a more positive reaction than the standard ($p<0.01$) and the 1050Hz deviant ($p<0.05$). Interactions with frontality ($F(4,676)=6.527$, $p<0.001$) and laterality ($F(4,676)=10.688$, $p<0.001$) indicated that this effect was strongest at frontal electrodes and less strong at right hemisphere electrodes. In the time window of 360-400ms, there was a main effect of stimulus type ($F(2,338)=12.010$, $p<0.001$). The 1050Hz deviant elicited less positive reactions than the standard ($p<0.01$) and the 1200Hz deviant ($p<0.001$). There was also an interaction with frontality ($F(4,676)=29.439$, $p<0.001$). The patterns differed for each instance of frontality. For the frontal electrodes, the same pattern as for the main effect was found ($F(2,338)=7.833$, $p<0.01$). At the central electrodes, a hierarchy of positivity was found: responses to the 1200Hz deviant

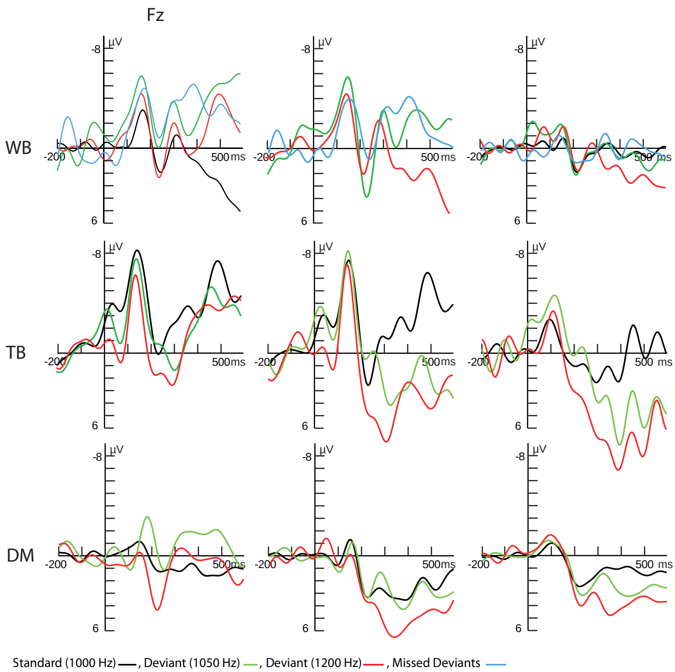


Figure 6.1: Brain activity of the three aphasic participants in the tones condition.

were most positive, followed by the standard and finally the 1050Hz deviant ($F(2,338)=17.295$, $p<0.001$; pairwise comparisons: standard versus deviant /ka/: $p<0.01$, standard versus deviant/ta/: $p<0.05$ and deviant /ka/ versus deviant /ta/: $p<0.001$). Also for the occipital electrodes, an effect of stimulus type was found ($F(2,338)=23.471$, $p<0.001$): the 1200Hz deviant evoked more positive responses than both the standard ($p<0.001$) and the 1050Hz deviant ($p<0.001$). In summary, DM did not show an effect in the MMN time window. In the second time window, a positivity was found, while a negativity was expected. In the P3 time window, he showed a P3 comparable to that of the non-brain-damaged participants, however limited to the more distant deviant.

As depicted in Figure 6.1, these results differed from those of the non-brain-damaged participants: in the MMN time window no effects were found for WB and DM. TB showed a negativity for the 1050Hz deviant only, which had, however, a different distribution than that of the non-brain-damaged

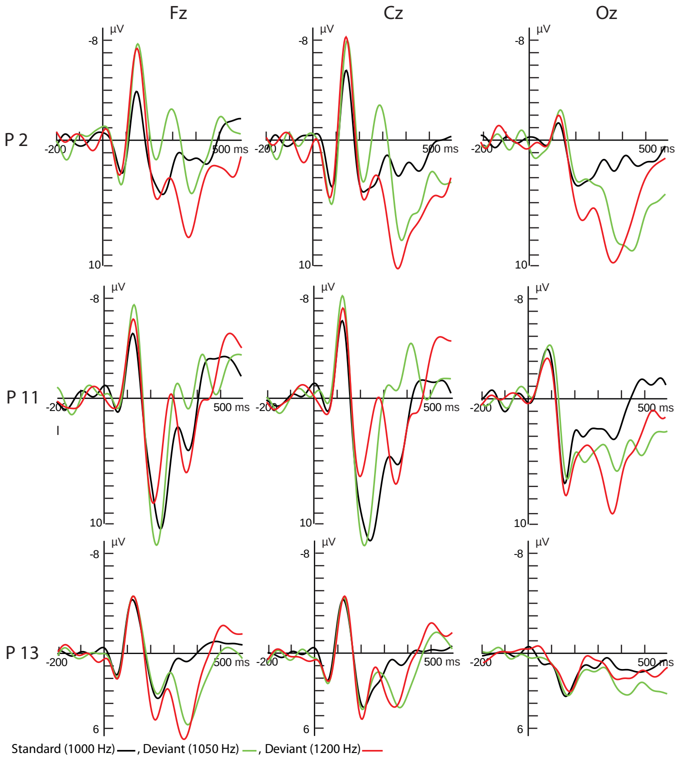


Figure 6.2: Brain activity of three randomly chosen non-brain-damaged participants in the tones condition.

controls. Between 200 and 240ms, WB did not show any effect, while TB and DM exhibited a positivity for the 1200Hz deviant, a pattern opposite to that of the control participants. In the P3 time window, TB and DM showed a positivity for the 1200Hz deviant, but not for the 1050Hz deviant. The non-brain-damaged participants had a similar effect although it was present for both deviants. For WB, no effect was found.

For illustration purposes, the activation patterns of the three aphasic participants and three randomly chosen[2] participants are given in Figures 6.1 and 6.2, above. It can be seen that the patterns of individual non-brain-damaged

[2]Per condition, three participants were randomly selected, using the L'Ecuyer portable combined random number generator (L'Ecuyer, 1988).

participants generally resemble the pattern of the group. This was not the case for either of the aphasic participants. For WB, it can be seen that there were no systematic differences between stimulus types. TB did show the P3, but in the MMN time window the standard is actually less negative than the deviants. For DM, also a small P3 was found, but no MMN: the standard actually lied in between the two deviants at the frontal electrodes in the relevant time window.

As the activity patterns in the 'pure tones' condition were not comparable to those of the control group, it was concluded that the current design does not allow for a statistic analysis of the results of the aphasic participants. For the remainder of this chapter, a more descriptive overview of the outcomes of the other conditions will be provided. The graphs will be presented for illustration purposes and visually inspected.[3]

6.3.2 Auditory syllables

For the auditory syllables, visual inspection of the graphs in Figure 6.3 revealed that none of the aphasic participants showed a pattern comparable to the control group. Only DM showed a positivity that resembled the P3 of the control group. However, he did not show any negativity in the earlier time windows. The individual results of three randomly chosen non-brain-damaged participants are more comparable to the group pattern. All three had a clear P3 and also the N2 pattern were found for participants 7 and 13. Participant 10 also showed a negativity, which was somewhat later than that of the group.

[3]For those who are interested in the statistical analysis, despite the outcome of the pilot 'tones' condition, the results are provided in appendix E.2.

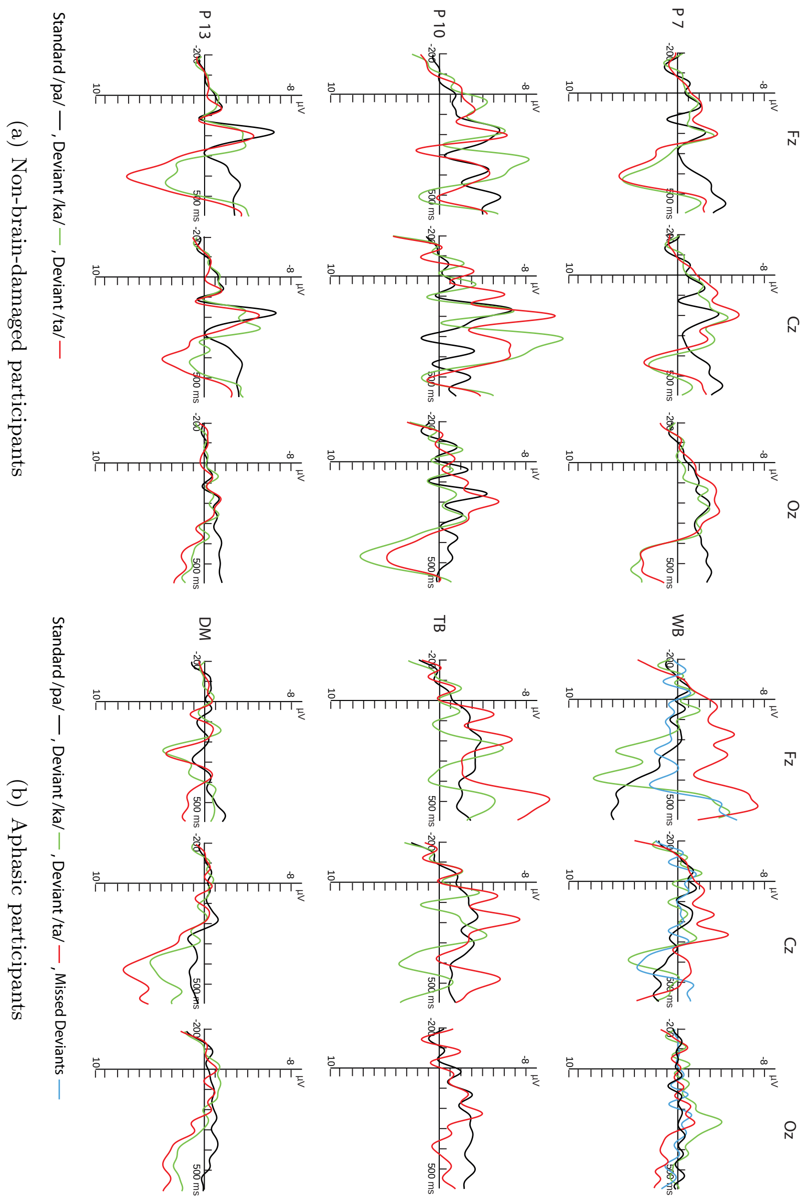
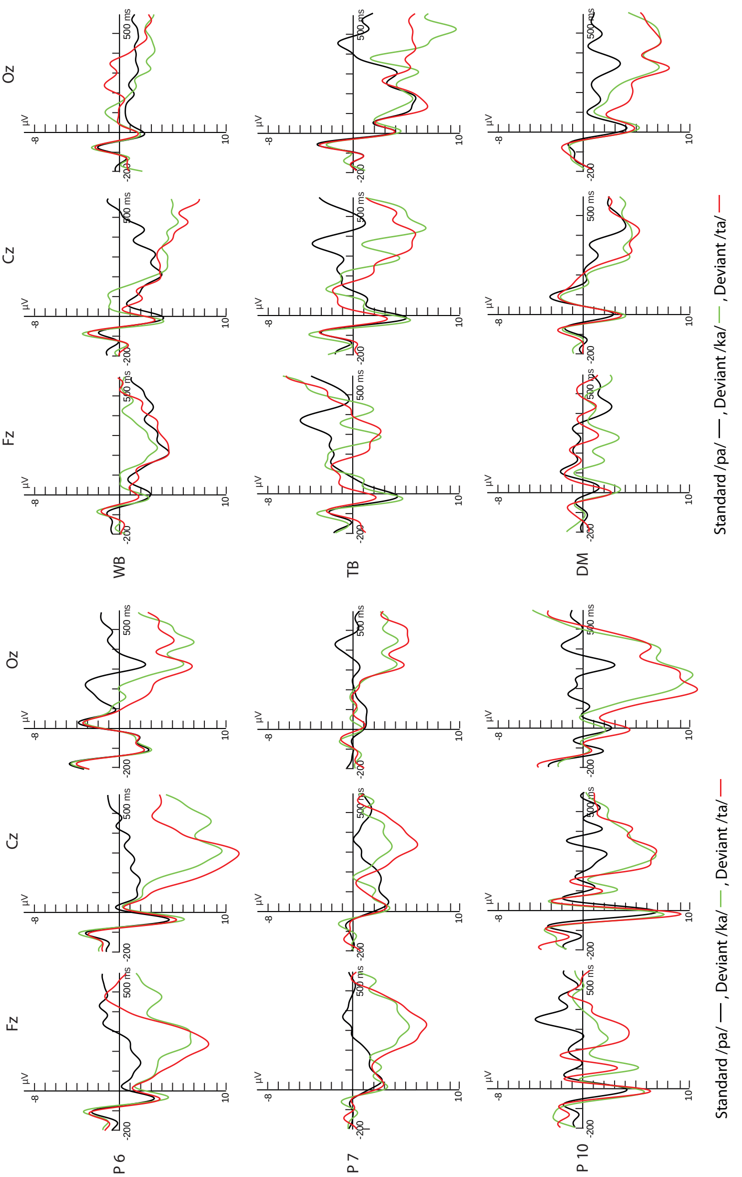


Figure 6.3: Brain activity for auditory syllables.



(a) Non-brain-damaged participants

(b) Aphasic participants

Figure 6.4: Brain activity for visual syllables.

6.3.3 Visual syllables

The three non-brain-damaged participants showed a large positivity across electrodes and time windows, as it was seen in the group results. Neither of the aphasic participants showed this pattern. They rather had a more traditional P3, restricted to central and occipital electrodes and beginning at 300ms (see Figure 6.4).

6.3.4 Comparison of audiovisual and auditory syllables

Visual inspection of the graphs presented in Figure 6.5 showed that, for the non-brain-damaged participants, the positivity was larger for auditory syllables and the negativity starts earlier for audiovisual syllables (for participant 1 and 3, after subtraction of the visual activity), resembling the group pattern. The larger positivity was also found for the aphasic participants TB and DM, but not for WB. The negativity was not identified for either of the aphasic participants.

6.3.5 McGurk stimuli and congruent audiovisual syllables

The group of non-brain-damaged participants showed more negative responses to the incongruent McGurk syllables than to the congruent standards at occipital electrodes in all three time-windows. The McGurk stimulus elicited also more negative responses than both congruent deviants (/ka/ and /ta/) between 360 and 400ms. In Figure 6.6 it can be seen that for the participants 7 and 12 the group pattern regarding the standard was found individually. The difference with the deviants was only evident for participant 12. WB also showed a large negativity for the McGurk stimulus throughout all time windows. The negativity seemed larger than for the standards and both congruent deviants. For DM, the McGurk stimulus elicited a more negative reaction than the standard, but no difference between the McGurk stimulus and the congruent deviants was found. No analysis was made for the aphasic participant TB, as not enough data was available for her.

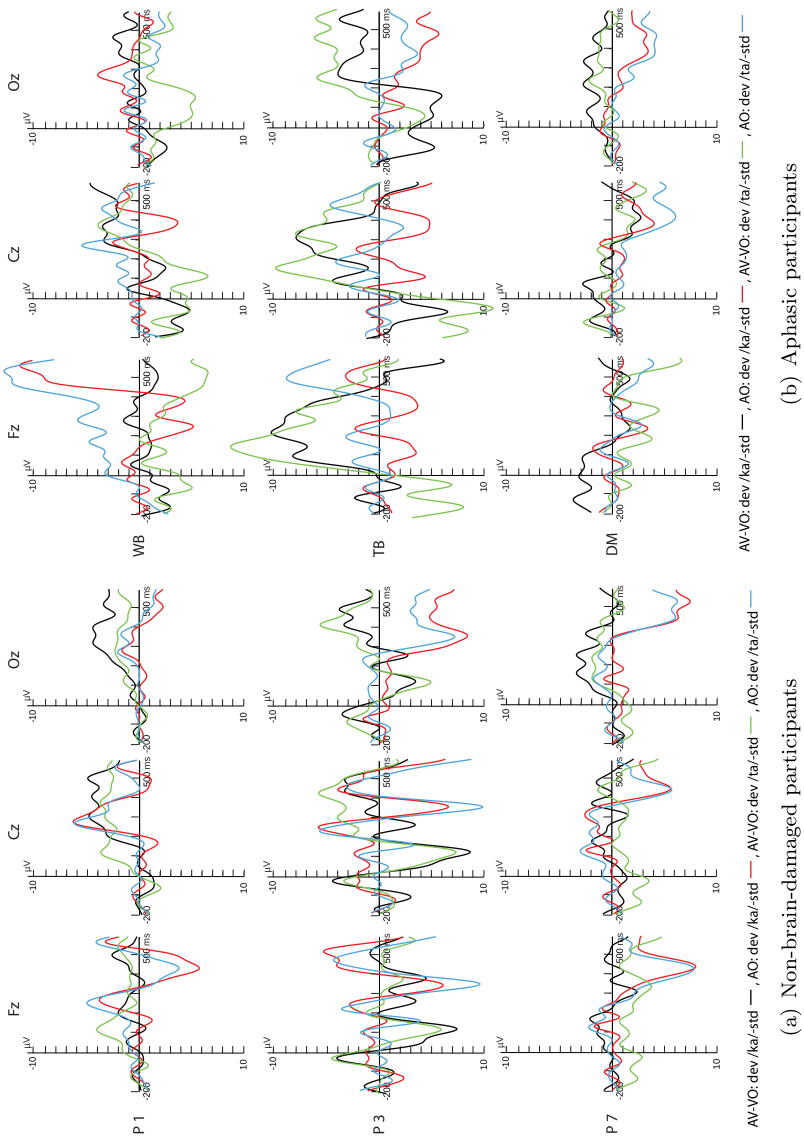


Figure 6.5: Comparison of brain activity for auditory and audiovisual syllables: presented are the difference waves (deviant minus standard) for both deviants in the auditory and the audiovisual condition (after subtraction of the visual activity).

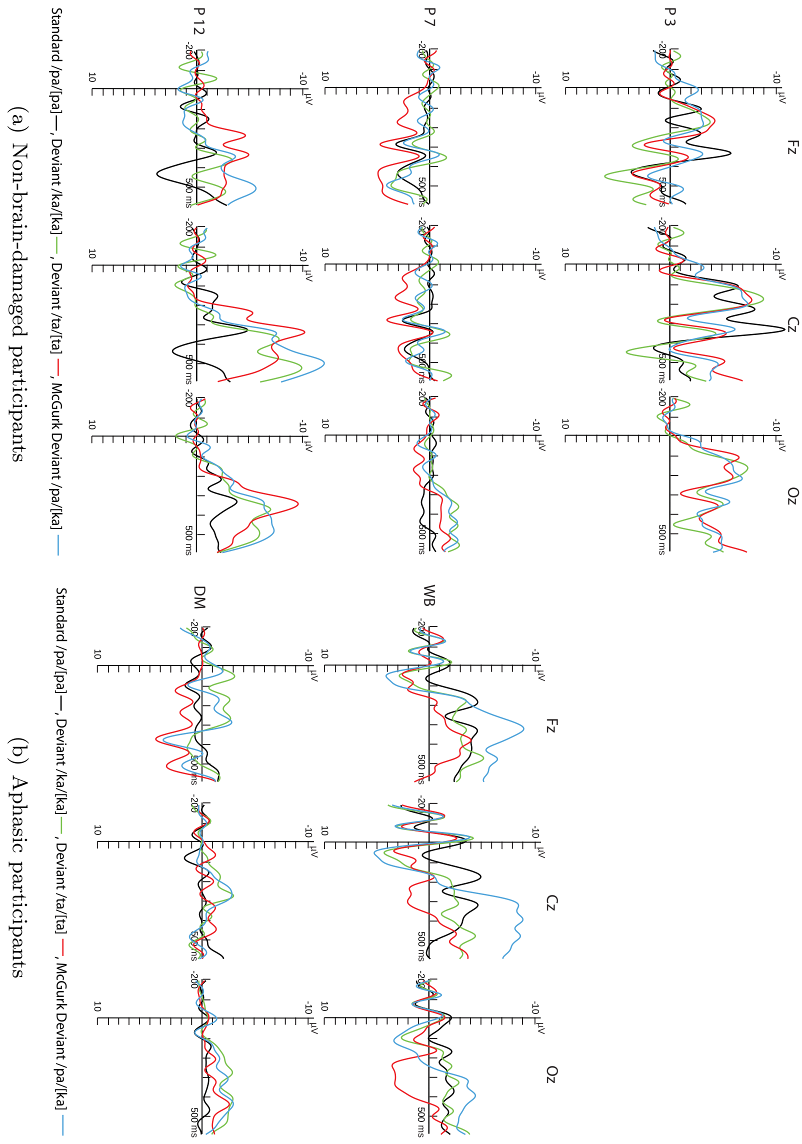


Figure 6.6: Brain activity for McGurk stimuli and audiovisual syllables. The waves depicted represent the audiovisual condition after subtraction of the visual activity. No graph is provided for participant TB, as there were not enough data available.

6.4 Discussion

This study aimed to investigate the processing of speech in the brains of aphasic participants. The setup was previously used in a study with non-brain-damaged participants, in which it proved its validity (see chapter 5). In order to determine the applicability of the paradigm to aphasic participants, the condition with 'pure tones' was analyzed as a baseline measure. Behaviorally, the aphasic participants did not exhibit problems in this condition, except for WB, who had problems in detecting the less distinct deviant.

However, the measured activity patterns of the ERP differed largely from that of the control group. These results indicate that the neural processing of participants with brain damage is difficult to compare with ERPs to those of participants without brain damage. The absence of a clear MMN, N2b, and, for WB, also the P3 in the 'tones' condition were not due to the fact that single cases were analyzed, as can be seen when inspecting the graphs for individual non-brain-damaged participants. The generation of the recorded brain activity may partly depend on lesioned areas although the lesion did not influence the behavioral performance. As the baseline condition showed that the paradigm did not elicit comparable results for non-brain-damaged and aphasic participants, the conditions related to auditory and audiovisual speech perception will not be further discussed. The graphs showing the brain activity have been presented above for illustration purposes.

The current study, therefore, was not able to to contribute knowledge to the issues related to speech perception. However, this study shows that carrying out measures of brain activity with patients with brain damage can be heavily influenced by their lesions. While other authors were more successful with a paradigm lacking a behavioral task, the addition of a task in the current study made the interpretation of the data almost impossible. It must, therefore, be concluded that the lesions of the participants alter the activity related to a task of which the performance itself is not influenced by the lesion. In future research, it should be carefully considered whether the use of the ERP methodology in an active oddball task is appropriate for aphasic participants.

CHAPTER 7

Discussion and Conclusion

7.1 Discussion

In the prolegomenon of this thesis, three main issues were raised which were addressed throughout the subsequent chapters. In the following sections, each of these issues is discussed, taking the results of the four experimental chapters into account.

The role of phonemes and phonetic features in processing

The first issue addressed regarded the perception of phonemes, the processing of contrasts between them and the role phonetic features play. The results of the discrimination study in chapter 2 and the ERP study in chapter 5 contribute to this issue.

Previous research has shown that the discrimination performance of aphasic listeners depends on the number of phonetic dimensions distinguishing two phonemes (Blumstein et al., 1977; Blumstein, 1994). It also is important in which dimension the phonemes differ (Saffran et al., 1976; Blumstein et al., 1977; Caplan & Aydelott-Utman, 1994; Csépe et al., 2001). These findings were confirmed for Dutch aphasic listeners in the current discrimination experiment, described in chapter 2. The closer two phonemes were to each other, the more difficulties the aphasic participants had in discriminating them. Specifically the dimension 'voicing' caused problems. This lead to the conclusion that the unit of processing must be smaller than phonemes. In the ERP study with non-brain-damaged participants, two contrasts in the same dimension (place of articulation) were further compared. Although the distinctions between /pa/ and /ka/ on the one hand and /pa/ and /ta/ on the other hand are both formed by one dimension (place of articulation), models of speech perception still distinguish between those contrasts by making use of non-binary features, which represent the degree of similarity (e.g. McClelland & Elman, 1986). In non-brain-damaged participants, both contrasts evoked a P3, an ERP component elicited by the recognition of a task-relevant deviance. The amplitude of the component differed, however, between both contrasts and was larger for the deviant /ta/, which is phonetically closer to the standard than the deviant /ka/ is. Therefore, even different contrasts within the same phonetic dimension are processed differently, implying that yet smaller entities play a role in speech perception. In the discussion related to the TRACE model, the issue of degrees of difference will be further discussed.This thesis, thus, showed that subphonemic entities play a role in speech perception. Even the differentiation into the three phonetic dimensions 'place of articulation', 'manner of articulation' and, 'voicing' is not sufficient to explain the neurophysiological evidence from the ERP study.

The influence of speechreading on perception

The second issue addressed was an investigation to the benefits from speechreading for both non-brain-damaged and aphasic listeners. Generally, in all experiments carried out, audiovisual input showed advantages compared to pure auditory input, confirming previous findings (Sumby & Pollack, 1954; Reisberg et al., 1987; Shindo et al., 1991). For the aphasic participants, higher accuracies were found in all three tasks: the discrimination task, the identification task and the oddball task in the ERP study. Also, their reaction times decreased in the identification task when materials were presented audiovisually rather than auditorily. In the discrimination study, it was investigated whether this advantage is based on a specific phonetic dimension. This was not the case. The summation of small effects for each dimension lead to the general improvement that was found. An effect of speechreading was also found for the non-brain-damaged participants: while their accuracy was at ceiling for both modalities, their reaction times were shorter for audiovisual than for auditory stimulus presentation in the identification task.

While both non-brain-damaged and aphasic participants benefited from speechreading, there were also differences between both groups: when perceiving a McGurk illusion, that means when integrating incongruent auditory and visual information, the reaction time of non-brain-damaged participants was longer than when not perceiving the illusion with the same kind of input. This is not the case for the aphasic participants. Nonetheless, they also integrate the two modalities. This was shown by the answer patterns, which were comparable to those of the non-brain-damaged participants. Two possible explanations were given for the influence of integration on the processing speed. The first is based on experiments by Soto-Faraco and Alsius (2007, 2009): participants in their studies reported awareness of the mismatch for the same stimuli for which they experienced the McGurk effect. That means that they access both unimodal and multimodal information, which is not the case for stimuli where no McGurk effect is experienced. If it is assumed that the aphasic participants do not have conscious access to the unimodal information before integration, that explains why they do not show increased reaction times for McGurk percepts. The other explanation is based on the TRACE model of speech perception and its extension for audiovisual material (McClelland & Elman, 1986; Campbell, 1988, 1990). In this model, it is assumed that different values for one feature inhibit each other. The incongruent McGurk stimuli activate different instances of features which leads to inhibition between them. This can cause the slow-down for the non-brain-damaged participants. It has been previously claimed that aphasic listeners have a deficit of inhibition at the lexical level (Wiener et al., 2004; Janse, 2006; Yee et al., 2008). If this deficit also exists at the feature level, no slow-down is expected.

The advantage of audiovisual processing over auditory processing was also shown by the results of the ERP study with non-brain-damaged participants. A comparison between the brain activities related to the auditory part of the audiovisual syllables (after subtraction of the visual activity) and the auditory syllables showed a difference in the amplitude of the P3 component. This indicates that the brain activity related to audiovisual processing is more than an addition of the auditory and the visual activity. The smaller amplitude of the P3 can be explained by the decreased amount of attention needed when both auditory and visual information are accessible. Thus, the difference between the two input modalities resembles the increased ease of processing when information is presented audiovisually.

The advantage of speechreading was, therefore, shown behaviorally for both aphasic and non-brain-damaged listeners. Furthermore, this advantage was also reflected in the neurophysiological data.

An evaluation of the TRACE model for the current data

The final point investigated in this thesis was the validity of the TRACE model of speech perception and its extension to audiovisual speech (McClelland & Elman, 1986; Campbell, 1988, 1990) for the current data. The model assumes seven phonetic features, such as ‘power’, ‘diffuse’, and ‘vocalic’. The feature level has been extended by two visual features (derived from seen speech): ‘mouth opening’, and ‘lip-shape’. While both features differentiate between different places of articulation, it is not clear whether different manners of articulation can also be distinguished by seen speech. The perception of ‘voicing’ can only be described on the basis of the acoustic features of the model.

In the discrimination study, however, it was found that not only differences in place of articulation were more easily perceived with speechreading possible, but that all contrasts contributed to the general improvement that was found. This cannot be explained in terms of the model as postulated by Campbell (1988, 1990). The two features for ‘seen speech’ cannot grasp the difference between phonemes with only a distinction in voicing. For this, the model needs to be extended by an additional feature for seen speech that is capable of detecting a distinction in voicing. A further detailed investigation is necessary to determine in which visual aspects voiced and voiceless phonemes differ. Speechreading was not limited to the lips or the face in the current experiments. Also, the neck and shoulders were visible, which is the reason why the term speechreading was chosen rather than lipreading. The voice-onset time is an important factor to discriminate between voiced and voiceless phonemes; it is likely that the movement of the larynx, which was visible, played an important role. This assumption needs to be confirmed by a study investigating the movement of the larynx in detail for different phonemes and phoneme combinations. This could be achieved by comparing the performance in a lipreading task and a speechreading task, where in the first only the mouth is shown, while in the second, the face and the neck are visible. In a lipreading task, the perception of ‘voicing’ might not differ from pure auditory perception.

In the identification task, an advantage of audiovisual perception over auditory perception has also been found. The aphasic participants showed both slower and less accurate reactions when only the auditory input was available than when speechreading was also possible. This can be explained in terms of the TRACE model as follows: for auditory stimulus presentation, only the seven acoustic features add activation for the identification of a phoneme. If the aphasic participants have trouble in extracting all features correctly from the incoming sound stream, the correct phoneme may not get sufficient activation to reach threshold, or at least it reaches threshold later. Additional visual information, which is extracted correctly, can provide the necessary activation to correctly identify a phoneme.

A second finding of this study was that, for McGurk stimuli, non-brain-damaged participants experienced a slow-down when they were actually subject to the McGurk effect, compared to the McGurk stimuli for which they perceived either the auditory or the visual part of the input. This slow-down was explained by a mutual inhibition of different values of a feature which could be diminished for the aphasic participants, as discussed above.

Predictions based on the TRACE model (McClelland & Elman, 1986) were also made for the outcome of the ERP study. While both the contrasts between /pa/ and /ta/ on the one hand and /pa/ and /ka/ on the other are generally defined as a difference in one phonetic dimension, that is place of articulation, the TRACE model provides more detail. The difference between /pa/, /ta/, and /ka/ is based on three of the acoustic features: ‘diffuse’, ‘burst’, and ‘acute’. While /p/ and /t/ have identical values for the feature ‘diffuse’, they differ slightly regarding ‘burst’ and substantially regarding ‘acute’. The difference between /p/ and /k/ is overall larger, as these two phonemes also differ with regard to ‘diffuse’ (see Table 7.1 for the exact values).

Table 7.1: Values for the three relevant features distinguishing /p/, /t/, and /k/ as used in the TRACE model (McClelland & Elman, 1986).

Phoneme	Diffuse	Acute	Burst
/p/	7	2	8
/t/	7	7	6
/k/	2	3	4

This theoretically assumed distinction in the degree of difference is represented in the measured ERP activity of the non-brain-damaged participants: the difference between /ka/ and /pa/ elicited a higher amplitude of the P3 component than the difference between /ta/ and /ka/. The amplitude of the P3 is influenced by three factors: equivocation, probability, and resource allocation. The probability between the deviants did not differ and the behavioral results showed that also the equivocation was equal. This higher amplitude must, therefore, have been elicited by a difference in resource allocation, that is the attention invested to detect the deviant. More attention is needed to detect smaller differences. The prediction based on the model was that the difference between /pa/ and /ta/ is smaller than that between /pa/ and /ka/. The deviant /ta/ also elicited a higher amplitude than the deviant /ka/, so that the theoretical predications are supported by the neurophysiological findings.

7.2 Conclusion

In this thesis, three main issues were discussed based on the results of four experimental studies. First, an investigation of the role of phonemes and phonetic features in speech perception was discussed. It was known previously that processing of phonemes depended on phonetic features. Blumstein et al. (1977); Blumstein (1994) showed that aphasic listeners perform better with distinctions in more phonetic dimensions. There were varying outcomes concerning the question which dimension is most difficult to perceive (Saffran et al., 1976; Blumstein et al., 1977; Caplan & Aydelott-Utman, 1994; Csépe et al., 2001). The discrimination experiment described in chapter 2 shows that, also for Dutch, the number of dimensions differing influences the performance. Furthermore, it was found that Dutch participants with a disorder in speech perception have the most problems in detecting a distinction in 'voicing'.

The ERP experiment with non-brain-damaged participants also added to the understanding of the role of phonetic features in speech perception. Even stimulus pairs differing in the same dimension ('place of articulation') elicit distinct activation patterns in the ERP. The amplitude of the P3, indicating the attention invested in the task, depended on the size of the difference. Therefore the current findings imply that entities even smaller than 'phonetic dimension' play a role in the neural processes related to speech perception.

The second issue that was addressed concerned the influence of speechreading on perception. While the beneficial effects of speechreading on perception have been described before (Sumby & Pollack, 1954; McGurk & MacDonald, 1976; Reisberg et al., 1987), the studies reported in this thesis add to the understanding of these influences. It has been assumed that speechreading is limited to the phonetic dimension 'place of articulation' (Campbell, 1988, 1990). The results of discrimination experiment show that the overall beneficial influence of speechreading cannot be attributed to a single phonetic dimension, but rather all three dimensions contribute to the improvement.

In the identification experiment, it was shown that also non-brain-damaged participants benefit from speechreading. While their accuracy was at ceiling for both auditory and audiovisual stimulus presentation, their reaction times decreased when speechreading was possible. An investigation of the reaction times for McGurk percepts showed, that the integration of non-matching auditory and visual information lead to a processing slow-down. This is, however, only true for the non-brain-damaged participants. For the aphasic participants, no slow-down was found. This lack of slow-down was explained in terms of an inhibition deficit on the feature level of the TRACE model. A second explanation was based on the findings by Soto-Faraco and Alsius (2007, 2009) that non-brain-damaged participants access both unimodal and multimodal information when processing incongruent audiovisual speech. If this is not the case for aphasic listeners, no slow-down is expected.

More evidence for the influence of speechreading was provided in the ERP study, reported in chapter 5. In this experiment with non-brain-damaged listeners, it was found that the activity evoked by audiovisual syllables is not a mere addition of the activities from auditory and visual processing. After a subtraction of the visual activation pattern from the audiovisual, there was still a difference to the purely auditory activation, especially regarding the amplitude of the P3. The P3 amplitude was smaller for audiovisual stimulus presentation, indicating that less attention is needed when both auditory and visual input are available than when the speaker can only be heard.

Finally, the predictions made on the basis of the TRACE model (McClelland & Elman, 1986) and its extension to audiovisual speech (Campbell, 1988, 1990) were evaluated with regard to the outcomes of the current studies. It was found in the discrimination study that the beneficial influence of speechreading is not limited to the dimension 'place of articulation'. The visual features of the model are, however, only capable of detecting visual information from that dimension.

An extension of the model to include other visual features, such as the larynx configuration, is, therefore, necessary.

The other findings of the studies reported in this thesis can be readily explained in the terms of the TRACE model. The influence of speechreading that was found in the two behavioral studies was attributed to the additional visual features that are extracted for audiovisual speech. With this additional information the threshold for the detection of the correct phoneme is reached more easily and faster, leading to an increase in accuracy for the aphasic participants and decreased reaction times for both non-brain-damaged and aphasic listeners.

Furthermore, the TRACE model was used to explain the slow-down of reaction times that non-brain-damaged participants showed in the identification experiment whenever they gave a McGurk answer. The perception of non-matching auditory and visual information was assumed to lead to a mutual inhibition of the activation levels of features, delaying phoneme recognition.

Finally, also the finding from the ERP study with non-brain-damaged participants that two different contrasts of the same phonetic dimension elicited different brain reactions was explained in terms of the TRACE model. Because the features in this model are not based on binary values, but on different degrees of activation, the distance between two phonemes could be predicted more accurately. The neurophysiological findings supported those predictions.

In summary, the findings presented in this thesis show that the perception of speech depends on entities smaller than phonemes and phonetic dimensions. Furthermore, it was shown that speechreading aids perception: aphasic participants improved in the discrimination and identification tasks and both aphasic and non-brain-damaged listeners showed decreased reaction times with speechreading possible. The outcome of the ERP experiment with non-brain-damaged participants show that also in neurophysiological terms audiovisual speech perception is more than an addition of auditory and visual perception.

The TRACE model of speech perception (McClelland & Elman, 1986) accounts for the finding that entities smaller than phonetic dimensions are processed by representing phonetic features by various activation levels, rather than by a binary value. The extension for audiovisual speech perception by Campbell (1988, 1990) also allows for an explanation of the beneficial effect of speechreading on perception. Furthermore, the processing slow-down in non-brain-damaged listeners and the lack of it in aphasic listeners when being subject to the McGurk effect, were explained in terms of the model. The model can, however, not account for the fact that the influence of speechreading in the discrimination task was not limited to the dimension ‘place of articulation’. Therefore, an extension of the visual features is necessary.

Next to addressing the three issues raised in the prolegomenon and discussed in the previous sections, the findings from this thesis also lead to clinical implications. These are presented in the following section.

7.3 Clinical Implications

The clinical implications of the research described in this thesis are based on the findings from the two behavioral experiments reported in chapters 2 and 3. In the discrimination study, it was found that the problems of aphasic participants in speech perception are more severe when small distinctions need to be differentiated. The phonetic dimension ‘voicing’ causes the most problems.

The PALPA diagnostics battery (Kay et al., 1992) and its Dutch translation (Bastiaanse et al., 1995) diagnose a disorder in speech sound discrimination with both a word and a nonword discrimination task. The pairs are differentiated by one phonetic dimension only. No information is gained about the processing abilities concerning larger differences. The findings from the discrimination experiment in this thesis, however, indicate that the deficit may be limited to the detection of differences in one dimension, but can also can affect larger differences. As this is essential when determining the appropriate treatment level, Morris et al. (1996) introduced a ‘maximal pair’ screening: a discrimination task, which test differences in one, two or three phonetic dimensions. An adoption for German was developed by Hessler (2007).

While these screenings tell more about the severity of the deficit, they do not tell which phonetic dimension is affected. The outcomes from the discrimination experiment show that the deficit can be specific to one phonetic dimension. It, therefore, needs to be tested for each individual which dimensions are affected by the impairment in order to address those specifically in treatment. The Dutch version of the PALPA (Bastiaanse et al., 1995) compares all three dimensions. It does not, however, compare all three dimensions in the same syllable position. The dimensions ‘place of articulation’ and ‘manner of articulation’ are tested in final position or metathesis and the dimension ‘voicing’ in initial, thus more salient position. Therefore, it is difficult to compare the results of all three dimensions. Therefore, new screenings need to be developed which can also detect problems with 'voicing' reliably.

Another outcome of the discrimination and the identification experiments was that speechreading aids perception in most aphasic listeners. This influence is not tested with the diagnostic batteries mentioned above, but should be assessed as well. The utilization of speechreading has been a successful treatment method (Gielewski, 1989; Morris et al., 1996; Grayson et al., 1997; Hessler & Stadie, 2008). The finding that the beneficial influence is not limited to the dimension 'place of articulation' (see chapter 2) stresses the importance of utilizing speechreading in treatment. In the discrimination and the identification task it became, however, evident that not all participants benefited from speechreading. For these participants, a treatment based on speechreading is not appropriate unless their speechreading and audiovisual integration skills are trained first.

Morris et al. (1996) and Hessler and Stadie (2008) reported successful treatment studies based on the utilization of speechreading. In both studies, several tasks were applied, such as phoneme-grapheme matching, syllable discrimination and word-picture matching. Depending on the task, words or nonwords were used. During treatment, the difference within pairs or between target and distractor was gradually decreased from three dimensions to one dimension. Materials were initially presented allowing for speechreading. This information was no longer available when the participant mastered the task. In the tasks using nonwords, such as the syllable discrimination task, distinctions in all three dimensions were equally common. This was the case to ensure that treatment for all dimensions was balanced. No attention was paid to the question which dimensions were actually affected by the impairment. Treatment could have been even more efficient if only problematic dimensions would have been addressed.

In summary, the clinical implications of the research studies reported in this thesis are that an individual profile needs to be established when diagnosing the deficit. It needs to be known how severe the deficit is (i.e. whether also larger differences are affected), whether the deficit is specific to a certain dimension, and whether the participant benefits from speechreading. Once this is known, treatment can address the specific deficit of an individual, increasing the effect of a treatment based on methods as reported in Morris et al. (1996) and Hessler and Stadie (2008).

Summary

In this thesis, the auditory and audiovisual processing of phonemes in non-brain-damaged and aphasic listeners was investigated. In the prolegomenon, three main issues were raised, which were addressed in the four experimental chapters: (1) the role of phonetic features in processing, (2) the influence of speechreading on speech perception, and (3) an evaluation of the validity of the TRACE model of speech perception (McClelland & Elman, 1986) and its extension for audiovisual perception (Campbell, 1988, 1990).

Chapter 1 provided a theoretical background of auditory and audiovisual speech perception. Two models of auditory processing were introduced: the Logogen model (e.g. Morton, 1969; Howard & Franklin, 1988) and the TRACE model (McClelland & Elman, 1986). While the first served to identify the deficits of the aphasic listeners, the second formed the theoretical basis of the research studies. Furthermore, different models of audiovisual processing have been discussed, amongst which an extension of the TRACE model to audiovisual processing (Campbell, 1988, 1990). Literature was reviewed, which shows that aphasic listeners have deficits in speech sound discrimination that depend on the size of the difference: the more phonetic dimensions differ, the fewer errors they make. Studies differ in their findings regarding which dimension is most difficult to perceive: 'place of articulation' or 'voicing'. There is, however, agreement on the fact that speechreading influences comprehension. Not only do aphasic listeners benefit from speechreading, but also non-brain-damaged

listeners benefit under adverse conditions, such as noise (Sumby & Pollack, 1954) or demanding contents (Reisberg et al., 1987). Even when speech is clear and easy to understand, visual information influences the perception: McGurk and MacDonald (1976) found out that presentation of non-matching auditory and visual information for a speech sound (for example auditory /p/, visual /k/) often leads to the perception of an intermediate speech sound (for example /t/), which unites features of both input types. This is called the 'McGurk effect'.

Chapter 2 was the first experimental chapter. It discussed a study on auditory and audiovisual perception of different phonetic dimensions by non-brain-damaged and aphasic listeners. A nonword discrimination experiment was carried out in different presentation conditions (auditory versus audiovisual) and for different stimulus types. It was found that the performance of the aphasic participants improved with the magnitude of the difference between stimuli: the more features differed, the higher the accuracy became. The aphasic participants had most problems detecting differences in the dimension 'voicing'. This was true for both auditory and audiovisual stimulus presentation. Overall, the accuracy of the aphasic listeners was higher in the audiovisual than in the auditory condition. This beneficial influence of speechreading was not caused by any specific phonetic dimension, rather all three dimensions contributed to the improved accuracy.

In chapter 3, a study on audiovisual perception and the McGurk effect in non-brain-damaged and aphasic participants was presented. A nonword identification task was carried out and reaction times were recorded. It was found that the aphasic participants were generally slower and less accurate than non-brain-damaged participants in identifying stimuli presented auditorily or audiovisually. Two of the three aphasic participants clearly benefited from speechreading, showing a higher accuracy and faster reaction times with audiovisual stimulus presentation than with only auditory information. The group of non-brain-damaged participants also showed decreased reaction times for the audiovisual condition compared to the auditory condition. For the McGurk stimuli, various answer patterns were found amongst the non-brain-damaged and the aphasic participants. The reaction time patterns, however, showed a clear difference in processing: while the non-brain-damaged participants reacted slower when they were subject to the McGurk illusion, this did not hold for either of the aphasic participants. Therefore, not only a quantitative, but also a qualitative difference of audiovisual processing was found between the two groups. This difference could be explained in two ways. The first explanation was based on results from Soto-Faraco and Alsius (2007, 2009), who claim that McGurk stimuli are processed unimodally (separately for the auditory and the visual modality) before integration takes place, which would account for a slow-down in reaction times. If the aphasic participants do not have access to the unimodal information first, that would explain the absence of the slow-down. The second explanation was based on the inhibition mechanism assumed in the TRACE model (McClelland & Elman, 1986; Campbell, 1988, 1990): due to contradictory information from auditory and visual speech, different instances of a feature are activated and inhibit each other. Therefore reaching the threshold for selecting the correct phoneme is delayed. The lack of the slow-down for the aphasic participants can be explained if they lack inhibition at feature level. A diminished inhibition was reported in studies focusing on the lexical level (Wiener et al., 2004; Janse, 2006; Yee et al., 2008).

Chapter 4 formed an introduction to the event-related potential (ERP) methodology. In this chapter an overview of the recording of ERPs was provided and previous findings relevant to the current research studies were discussed. Three components, which were of relevance to the planned studies, were introduced: the mismatch negativity (MMN), the N2b, and the P3. The MMN is a component that is elicited by an auditory deviant stimulus in a sequence of repeating 'standard' stimuli. The MMN represents automatic processing and is recorded even when the participants do not pay attention to the stimuli. Two other components have been discussed, the N2b and the P3, which both react to the conscious, task-relevant detection of a mismatch. The amplitude of the P3 is influenced by the probability of the deviant stimulus, the certainty of the participant and the resources that were allocated to detect it, such as attention (Johnson, 1984, 1986).

In chapter 5, an ERP study with non-brain-damaged participants was described. An active oddball task was carried out in four presentation modalities (tones, auditory syllables, visual syllables, and audiovisual syllables). For the pure tones, the participants showed the expected pattern of MMN, N2b, and P3 for two different deviants. The fact that the baseline condition was in line with the literature was taken as evidence for the validity of the current setup. In the auditory sub-experiment, the two deviants /ta/ and /ka/ were inserted in the sequence of standard stimuli (/pa/). Neither of the deviants elicited a clear

MMN or N2b, but rather an N2 was found, which was difficult to differentiate into its parts. A clear P3 was found for both deviants. The amplitude was larger for the for the deviant /ta/, which is theoretically assumed to be less distant to /pa/ than /ka/ is. This increased P3 has been attributed to the larger amount of attention needed to detect the difference. Furthermore, audiovisual and auditory processing have been compared. After a subtraction of the visual activity from the audiovisual activity, any differences found were assumed to be due to the integration of auditory and visual information. In the P3 time window, the auditory syllables evoked a larger P3 amplitude than the auditory part of the audiovisual syllables. This reflects the fact, that for audiovisual stimuli, processing is eased, requiring less resources to detect a difference between stimuli. For the McGurk stimuli, a difference to both congruent deviants and standards was found. As the visual activity was subtracted from the stimuli, there was no physical difference between the standard and the McGurk stimuli. Therefore, it was concluded that the difference in activation patterns is due to the more difficult integration of incongruent stimuli.

Chapter 6 described the investigation of the possibility to carry out the same ERP study, as described before, with aphasic participants. Unfortunately, only three aphasic participants could be included in this study. Therefore, their data were not analyzed as a group, but rather investigated in a single case approach. No clear results could be obtained. In the tones condition, which should not differ between aphasic and non-brain-damaged participants, large differences were found. This lead to the conclusion that the generation of the MMN, N2b, and P3 was corrupted by the brain damage although the behavioral performance was not influenced. Therefore, the absence of components in the experimental conditions could not be interpreted as related to their perception deficits. Thus, the data of the sub-experiments with auditory, visual, and audiovisual syllables were not analyzed and discussed with regard to the issues raised in the prolegomenon. Also, the visual inspection of the patterns of the aphasic participants lead to outcomes different from those of three randomly chosen non-brain-damaged participants. While the latter show patterns generally consistent with those of the group, this was not the case for the aphasic participants. The conclusion of this study was, therefore, that carrying out ERP research with participants with brain damage is difficult, as their anatomical lesion might influence the generation of components elicited by tasks of which the performance is not influenced by the deficit.

The final chapter consisted of a discussion of the three issues raised in the prolegomenon, a general conclusion and a short description of the clinical impact of the current findings. With regard to the main issues addressed in this thesis, it was found that the perception of speech depends on entities smaller than phonemes and phonetic dimensions. Secondly, it was concluded on the basis of the two behavioral studies and the ERP experiment with non-brain-damaged participants that speechreading improves the performance of the participants and that the influence of speechreading on perception is also represented in neurophysiological measures. The evaluation of the TRACE model of speech perception (McClelland & Elman, 1986; Campbell, 1988, 1990) revealed that almost all findings could be explained in terms of the model. The influence of speechreading on other dimensions than 'place of articulation' can, however, not be accounted for by the model in its current form. An extension of the visual features is necessary. Furthermore, the findings reported in this thesis show that a deficit in speech sound discrimination can affect the phonetic dimensions differently. This implies that it is important to establish an individual profile of an aphasic listener during diagnostics, which serves as a basis for choosing an appropriate treatment.

Nederlandse Samenvatting

Dit proefschrift beschrijft de auditieve en audiovisuele verwerking van fonemen door mensen met en zonder hersenbeschadiging. In het 'prolegomenon' worden drie hoofdpunten geïntroduceerd, die in de vier experimentele hoofdstukken onderzocht worden: (1) de rol van fonetische kenmerken bij de verwerking, (2) de invloed van liplezen[1] op taalbegrip en (3) een evaluatie van de validiteit van het TRACE-model (McClelland & Elman, 1986) en de uitbreiding ervan naar audiovisuele verwerking (Campbell, 1988, 1990).

Hoofdstuk 1 geeft een theoretische achtergrond van auditieve en audiovisuele spraakwaarneming. Twee taalverwerkingsmodellen worden voorgesteld: het multimodale Logogen model (e.g. Morton, 1969; Howard & Franklin, 1988) en het TRACE-model (McClelland & Elman, 1986), dat zich beperkt tot spraakwaarneming. Het eerste model wordt gebruikt om de taal- en spraakproblemen van proefpersonen met verworven hersenletsel (afasie) die aan dit onderzoek deelnamen in kaart te brengen. Het tweede model vormt de theoretische basis voor de wetenschappelijke studies. Verder worden verschillende modellen van audiovisuele verwerking bediscussieerd, waaronder onder meer een uitbreiding van het TRACE-model (Campbell, 1988, 1990). Uit

[1] In dit proefschrift wordt de Engelse term 'speechreading' gebruikt, om duidelijk te maken dat niet alleen de bewegingen van de lip een rol spelen, maar ook andere aspecten van taalproductie zichtbaar zijn, bijvoorbeeld de beweging van de larynx. In de Nederlandse samenvatting wordt echter de term 'liplezen' gebruikt, omdat dit de meest gangbare term is.

voorgaande studies is gebleken dat de mate waarin afatische luisteraars moeite hebben met spraakklankdiscriminate, afhangt van de grootte van het verschil tussen klanken: hoe meer fonetische dimensies verschillen, des te minder maken de proefpersonen met afasie fouten (Blumstein et al., 1977; Blumstein, 1994). Tot op heden bestaat er nog geen consensus over de kwestie welke dimensie het moeilijkst waarneembaar is: plaats van articulatie of stemhebbendheid. Men is het er echter over eens dat liplezen taalbegrip begunstigt. Niet alleen afatische luisteraars profiteren van liplezen, ook luisteraars zonder hersenbeschadiging hebben er profijt van indien omstandigheden niet optimaal zijn, zoals bij achtergrondgeluid (Sumby & Pollack, 1954) of moeilijke inhoud (Reisberg et al., 1987). Zelfs als spraak helder en goed te begrijpen is, beïnvloedt de visuele informatie de waarneming. McGurk and MacDonald (1976) kwamen erachter dat als klanken worden aangeboden die auditief en visueel niet overeenkomen (bijvoorbeeld /p/ auditief en /k/ visueel), er regelmatig een klank wordt waargenomen (bijvoorbeeld /t/) die tussen de gehoorde en geziene klank in ligt en kenmerken van beide verenigt. Dit wordt het 'McGurk-effect' genoemd.

Hoofdstuk 2 vormt het eerste experimentele hoofdstuk. Het beschrijft een onderzoek naar auditieve en audiovisuele waarneming van verschillende fonetische dimensies door luisteraars met en zonder hersenbeschadiging. De taak was lettergrepen te discrimineren die auditief dan wel audiovisueel gepresenteerd werden. De lettergrepen onderscheidden zich van elkaar in verschillende kenmerken. De prestatie van de proefpersonen met afasie hing samen met de grootte van het verschil tussen stimuli: hoe meer kenmerken verschilden, des te hoger was het aantal correcte reacties. De proefpersonen met afasie hadden de meeste problemen met het herkennen van verschillen in de dimensie 'stemhebbendheid'. Dit was het geval bij zowel auditieve als audiovisuele stimuluspresentatie. In het algemeen was de prestatie van de proefpersonen met afasie beter voor de audiovisuele dan voor de auditieve stimuli. Uit de resultaten van dit experiment blijkt dat alle drie fonetische dimensies bijdragen aan de verbeterde prestaties van de afatische proefpersonen bij audiovisuele stimuluspresentatie.

In hoofdstuk 3 wordt een onderzoek naar audiovisuele waarneming en het McGurk-effect bij proefpersonen met en zonder hersenbeschadiging gepresenteerd. Een nonwoord-identificatietaak werd uitgevoerd, waarbij ook de reactietijden gemeten werden. De deelnemers met afasie antwoordden langzamer en minder correct dan de deelnemers zonder hersenbeschadiging bij het identificeren van stimuli die auditief of audiovisueel gepresenteerd werden. Twee van de drie deelnemers met afasie hadden duidelijk baat bij het liplezen. Ze lieten snellere en correctere antwoorden zien in de audiovisuele conditie dan in de auditieve. Ook de deelnemers zonder hersenbeschadiging profiteerden van liplezen. Hun reactietijden daalden eveneens bij audiovisuele presentatie van de stimuli. Voor de McGurk-stimuli werden uiteenlopende antwoordpatronen gevonden bij zowel de proefpersonen met als zonder afasie. De reactietijdpatronen lieten daarentegen een duidelijk verwerkingsverschil tussen de groepen zien: terwijl de deelnemers zonder hersenbeschadiging een vertraagde reactie lieten zien wanneer er sprake was van een McGurk-effect, was dit niet het geval voor de proefpersonen met afasie. Dit betekent dat er niet alleen een kwantitatief, maar ook een kwalitatief verschil in de audiovisuele verwerking tussen beide groepen werd gevonden. Dit verschil kan op twee manieren verklaard worden. De eerste verklaring is gebaseerd op resultaten van Soto-Faraco and Alsius (2007, 2009), die postuleren dat McGurk-stimuli eerst unimodaal (afzonderlijk auditief en visueel) verwerkt worden voordat er integratie plaatsvindt, wat een vertraging kan opleveren. Als de mensen met afasie geen toegang tot de unimodale informatie hebben, zou dat verklaren waarom bij hen geen vertraging plaats vindt. De tweede verklaring is gebaseerd op het inhibitiemechanisme, dat in het TRACE-model aangenomen wordt (McClelland & Elman, 1986; Campbell, 1988, 1990): door de tegenstrijdige informatie van auditieve en visuele input, worden verschillende waardes van een kenmerk van een klank geactiveerd, die elkaar wederzijds inhiberen. Daardoor wordt de drempel voor het herkennen van het correcte foneem vertraagd bereikt. Het uitblijven van deze vertraging bij de deelnemers met afasie zou door een gebrek aan inhibitie op het niveau van kenmerken verklaard kunnen worden. Een verminderde inhibitie werd eerder vastgesteld in studies waarin de rol van inhibitie op het lexicale niveau werd onderzocht (Wiener et al., 2004; Janse, 2006; Yee et al., 2008).

Hoofdstuk 4 geeft een inleiding in het gebruik van event-related potentials (ERPs) tijdens de taal- en spraakverwerking. In dit hoofdstuk wordt beschreven hoe ERPs worden gemeten en worden resultaten van voorgaande studies besproken die relevant zijn voor het onderhavige onderzoek. Verder worden drie componenten geïntroduceerd die ook in de huidige studies worden verwacht: de mismatch negativity (MMN), de N2b en de P3. De MMN is een component die optreedt als reactie op een auditieve afwijkende stimulus in een reeks van herhaalde 'standaardstimuli'. De MMN representeert automatische verwerking en wordt ook gevonden als proefpersonen geen aandacht aan de stimuli besteden. Daarnaast worden de N2b en de P3 besproken die beide gerelateerd zijn aan bewuste, taakrelevante herkenning van een mismatch. De amplitude van de P3 wordt beïnvloed door de waarschijnlijkheid van de afwijkende stimulus, de zekerheid van de proefpersoon en de benodigde verwerkingscapaciteit om de mismatch te herkennen, zoals aandacht (Johnson, 1984, 1986).

In hoofdstuk 5 wordt een ERP-onderzoek bij deelnemers zonder hersenbeschadiging beschreven. Er werd een zogenaamde 'active oddball task' uitgevoerd. Dat wil zeggen dat de proefpersonen een taak kregen waarbij een reeks stimuli gepresenteerd werd, waarin één stimulus vaak en een ander zelden voorkomt. De proefpersonen werden gevraagd op de uitzonderlijke stimulus reageren. De stimuli behoorden bij deze taak tot vier verschillende types: tonen, auditieve lettergrepen, visuele lettergrepen en audiovisuele lettergrepen. In de conditie met de tonen werden alle verwachte componenten gevonden: zowel de MMN, N2b als P3 kwamen naar voren bij de twee uitzonderlijke stimuli. Omdat de resultaten in deze controleconditie overeenkwamen met resultaten zoals die eerder voor tonen werden gevonden in de literatuur, werd de opzet van de studie als valide beschouwd. In het experiment met auditieve lettergrepen werden de twee uitzonderlijke lettergrepen /ta/ en /ka/ ingevoegd in een reeks van de standaard lettergreep /pa/. Geen van deze uitzonderlijke stimuli veroorzaakte een duidelijke MMN of N2b. In plaats daarvan werd een N2 gevonden die moeilijk op te splitsen was in de subcomponenten MMN en N2b. Voor beide uitzonderlijke stimuli werd een eenduidige P3 gevonden. De amplitude was groter voor de stimulus /ta/, waarvan theoretisch geclaimd wordt dat deze minder dan /ka/ van de standaard stimulus /pa/ verschilt. Deze vergrote P3 werd toegekend aan de grotere hoeveelheid aandacht die nodig is om het verschil te detecteren. Verder werden audiovisuele en auditieve verwerking vergeleken. Nadat de visuele activiteit afgetrokken was van de audiovisuele, werden alle verschillen tussen de condities toegekend aan de integratie van informatie. In het tijdsinterval van de P3 werd gevonden dat de auditieve lettergrepen een grotere amplitude veroorzaakten dan het auditieve gedeelte van de audiovisuele lettergrepen. Dit weerspiegelt het feit dat de verwerking van audiovisuele input eenvoudiger is en minder capaciteit vergt, om een verschil tussen twee lettergrepen te herkennen dan de verwerking van auditieve input. De activiteit gerelateerd aan de McGurk-stimuli verschilde van zowel congruente uitzonderlijke lettergrepen alsook van de standaardlettergreep. Omdat voor de visuele activiteit gecorrigeerd werd, was er geen fysiek verschil tussen de standaard en de McGurk-stimuli. Derhalve werd geconcludeerd dat het verschil in activiteit door de moeizamere integratie van incongruente stimuli veroorzaakt wordt.

Hoofdstuk 6 beschrijft een onderzoek naar de mogelijkheid om het zojuist beschreven onderzoek ook met proefpersonen met afasie uit te voeren. Helaas konden slechts drie proefpersonen met afasie geïncludeerd worden in dit onderzoek. Om deze reden werden hun data niet als die van een groep geanalyseerd, maar als gevalsbeschrijvingen benaderd. Uit de analyse kwamen geen eenduidige resultaten naar voren. In de conditie met tonen, waarin geen verschillen met de controle groep werden verwacht, werden alsnog grote afwijkingen van het ERP-patroon gevonden. Dit leidde tot de conclusie dat de generatie van de MMN, N2b en P3 componenten verstoord werd door het hersenletsel, hoewel het aantal correcte reacties en de reactietijden niet afweken van de controlegroep. Daardoor kon het ontbreken van ERP-componenten in de experimentele condities niet toegeschreven worden aan de problemen met spraakherkenning. De data van de deelexperimenten met auditieve, visuele en audiovisuele lettergrepen werden daarom niet geanalyseerd met betrekking tot de punten die in het prolegomenon genoemd werden. Ook uit een visuele inspectie van de golven bleek dat de activiteit van de proefpersonen met afasie afwijkt van die van drie willekeurig geselecteerde proefpersonen zonder hersenbeschadiging. De individuele proefpersonen zonder hersenbeschadiging lieten patronen zien die grotendeels overeen kwamen met het groepspatroon, terwijl dit niet het geval was voor de individuele deelnemers met afasie. De conclusie van deze studie is dan ook dat het uitvoeren van ERP-onderzoek met hersenbeschadigde deelnemers niet mogelijk is, omdat hun letsel invloed heeft op het genereren van ERP-componenten die gerelateerd zijn aan een taak waarvan de prestatie niet beïnvloed wordt.

Het laatste hoofdstuk bestaat uit een discussie van de drie hoofdpunten die in het prolegomenon opgeworpen werden, een algemene conclusie en een korte beschrijving van de klinische relevantie van de beschreven resultaten. Wat betreft de hoofdvragen kan geconcludeerd worden dat de waarneming van spraak afhangt van eenheden die kleiner zijn dan fonemen en fonologische dimensies. Verder werd aan de hand van de twee gedragsmatige studies en het ERP-onderzoek met proefpersonen zonder hersenbeschadiging geconcludeerd dat liplezen de prestatie bevordert. Verder kan de invloed van liplezen op de waarneming ook in neurale metingen vastgesteld worden. De evaluatie van het TRACE-model van spraakwaarneming (McClelland & Elman, 1986; Campbell, 1988, 1990) liet zien dat bijna alle resultaten op basis van dit model verklaard kunnen worden. De invloed van liplezen op dimensies anders dan 'plaats van articulatie' kan echter niet aan de hand van het model in zijn huidige staat verklaard worden. Hiervoor is een uitbreiding van de visuele kenmerken in het model nodig. Ten slotte blijkt uit de resultaten van dit proefschrift dat een probleem in de herkenning van spraakklanken de fonetische dimensies verschillend kan aantasten. Dit impliceert dat het belangrijk is om tijdens de diagnostiek een individueel profiel van elke afatische luisteraar te verwerven, dat de basis vormt voor het kiezen van een geschikte behandeling.

Appendix

A Appendix to chapter 1: Auditory and Audiovisual Speech Perception

A.1 Illustrations on models

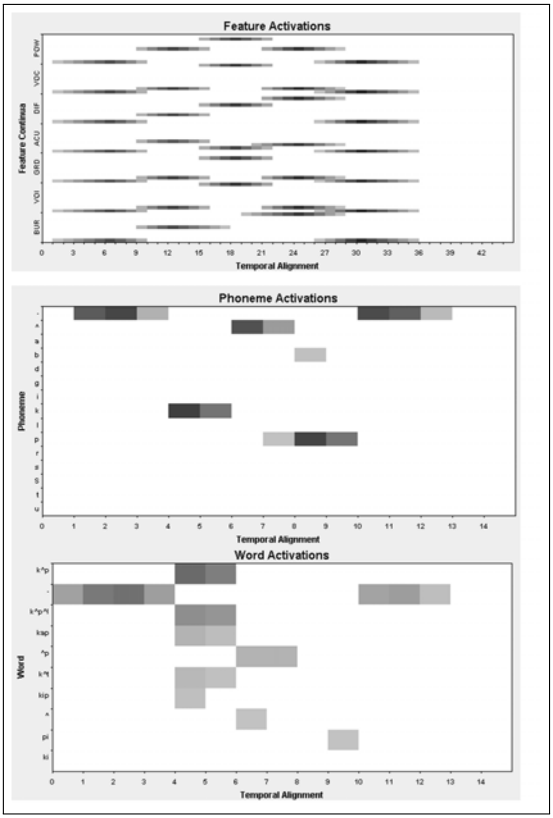


Figure A.1.1: Example of activation pattern for the word "cup"(/kʌp/) generated with a computational implementation of TRACE, JTrace (Strauss et al., 2007). In the top panel the activation of each feature at different time points is shown. The degree of activation is color-coded: a darker color represents a higher level of activation. The middle panel shows the activation of each phoneme relative to the time: while in the beginning and end no phoneme is activated, it can be seen that activation goes from /k/ via /ʌ/ to /p/, with a small co-activation for /b/. The word-level activation is shown at the bottom panel. Upon the identification of /k/, several words starting with /k/ are co-activated, but the overall activation is strongest for /kʌp/ (indicated by the darkest color).

B Appendix to chapter 2: Phonetic Dimensions in Aphasic Perception

B.1 Stimuli

Table B.1.1: Pairs with identical stimuli.

Stimuli			
/ba:f/ - /ba:f/	/bo:f/ - /bo:f/	/bœ͡yp/ - /bœ͡yp/	/du:p/ - /du:p/
/dɛ͡if/ - /dɛ͡if/	/dɔ͡uk/ - /dɔ͡uk/	/dɔ͡um/ - /dɔ͡um/	/dœ͡ys/ - /dœ͡ys/
/fe:t/ - /fe:t/	/fu:p/ - /fu:p/	/fɛ͡ip/ - /fɛ͡ip/	/fø:l/ - /fø:l/
/fø:p/ - /fø:p/	/fœ͡yp/ - /fœ͡yp/	/ke:m/ - /ke:m/	/kɔ͡uk/ - /kɔ͡uk/
/ky:m/ - /ky:m/	/la:p/ - /la:p/	/ly:m/ - /ly:m/	/lø:l/ - /lø:l/
/lø:p/ - /lø:p/	/ma:f/ - /ma:f/	/me:m/ - /me:m/	/my:k/ - /my:k/
/mø:l/ - /mø:l/	/mœ͡yp/ - /mœ͡yp/	/ni:x/ - /ni:x/	/no:k/ - /no:k/
/pa:f/ - /pa:f/	/pi:x/ - /pi:x/	/py:k/ - /py:k/	/pɛ͡if/ - /pɛ͡if/
/pøm/ - /pøm/	/pœ͡yp/ - /pœ͡yp/	/sa:f/ - /sa:f/	/si:x/ -/si:x/
/sy:n/ - /sy:n/	/sø:m/ - /sø:m/	/sø:p/ - /sø:p/	/sɛ͡if/ - /sɛ͡if/
/ta:f/ - /ta:f/	/ti:x/ - /ti:x/	/tɔ͡uf/ - /tɔ͡uf/	/tɔ͡un/ - /tɔ͡un/
/vø͡ys/ - /vø͡ys/	/vɔ͡ul/ - /vɔ͡ul/	/xi:m/ - /xi:m/	/xø:p/ - /xø:p/
/xœ͡ys/ - /xœ͡ys/	/xɔ͡uk/ - /xɔ͡uk/	/za:f/ - /za:f/	/za:p/ - /za:p/
/zi:m/ - /zi:m/	/zi:x/ - /zi:x/		

Table B.1.2: Pairs with different stimuli.

Condition (difference in)		Stimuli	
		/fe:t/ - /se:t/	/fø:p/ - /sø:p/
	Place	/pɛ͡if/ - /tɛ͡if/	/py:m/ - /ty:m/
		/si:x/ - /fi:x/	/ti:x/ - /pi:x/
		/du:p/ - /nu:p/	/kø:p/ - /xø:p/
1 Dimension	Manner	/kɔ͡uk/ - /xɔ͡uk/	/sa:f/ - /ta:f/
		/tɛ͡if/ - /sɛ͡if/	/ti:x/ - /si:x/
		/ba:f/ - /pa:f/	/bø:m/ - /pø:m/
	Voicing	/dɛ͡if/ - /tɛ͡if/	/dɔ͡uf/ - /tɔ͡uf/
		/pœ͡yp/ - /bœ͡yp/	/tɔ͡um/ - /dɔ͡um/
		/fø:l/ - /tø:l/	/kɛ͡ip/ - /fɛ͡ip/
	Place & Manner	/pø:m/ - /sø:m/	/tɔ͡uf/ - /xɔ͡uf/
		/vœ͡ys/ - /dœ͡ys/	/xø:p/ - /tø:p/
		/bo:f/ - /to:f/	/dɔ͡um/ - /pɔ͡um/
2 Dimensions	Place & Voicing	/fi:x/ - /zi:x/	/pɛ͡if/ - /dɛ͡if/
		/za:p/ - /fa:p/	/zi:m/ - /xi:m/
		/di:x/ - /si:x/	/dy:n/ - /sy:n/
	Manner & Voicing	/mœ͡yp/ - /pœ͡yp/	/py:k/ - my:k/
		/sɛ͡if/ - /dɛ͡if/	/zi:x/ - /ti:x/
		/ba:f/ - /sa:f/	/dœ͡ys/ - /xœ͡ys/
		/dɔ͡uk/ - /xɔ͡uk/	/fœ͡yp/ - /dœ͡yp/
		/fo:k/ - /no:k/	/fœ:l/ - /lœ:l/
		/ke:m/ - /me:m/	/kɔ͡ul/ - /vɔ͡ul/
3 Dimensions		/ky:m/ - /ly:m/	/la:p/ -/fa:p/
		/lø:p/ - /xø:p/	/ni:x/ - /fi:x/
		/nu:p/ - /fu:p/	/pi:x/ - zi:x/
		/sø:m/ - /bø:m/	/ta:f/ - /ma:f/
		/tø:l/ - /mø:l/	/za:f/ - /pa:f/

B.2 Individual data

Table B.2.1: Individual results of the aphasic listeners for the differences in 1, 2, and 3 dimensions.[1]

Initials	Audiovisual 1 dim.	Audiovisual 2 dim.	Audiovisual 3 dim.	Auditory only 1 dim.	Auditory only 2 dim.	Auditory only 3 dim.	Visual only 1 dim.	Visual only 2 dim.	Visual only 3 dim.
WB (W)	83%	100%	100%	83%	94%	94%	61%	72%	44%
(A')	(95%)	(99%)	(99%)	(94%)	(97%)	(98%)	(79%)	(84%)	(71%)
BB (G)	47%	63%	71%	39%	72%	61%	22%	44%	44%
(A')	(84%)	(89%)	(91%)	(77%)	(89%)	(85%)	(63%)	(67%)	(50%)
EK (A)	67%	94%	100%	72%	78%	94%	33%	44%	56%
(A')	(88%)	(97%)	(98%)	(92%)	(97%)	(98%)	(60%)	(68%)	(75%)
TB (G)	94%	100%	100%	72%	100%	100%	50%	50%	56%
(A')	(97%)	(99%)	(99%)	(88%)	(88%)	(97%)	(70%)	(70%)	(73%)
JH (M)	89%	94%	100%	83%	83%	100%	50%	0%	86%
(A')	(97%)	(99%)	(100%)	(96%)	(96%)	(100%)	(84%)	(24%)	(95%)
MB (G)	72%	78%	78%	39%	83%	78%	33%	67%	33%
(A')	(87%)	(89%)	(89%)	(76%)	(92%)	(90%)	(52%)	(76%)	(52%)
MEAN	75%	88%	91%	65%	87%	88%	42%	46%	53%
(A')	(91%)	(95%)	(96%)	(87%)	(95%)	(95%)	(68%)	(67%)	(74%)

[1] W = Wernicke's Aphasia, G = Global Aphasia, A = Anomia, M = Mixed Aphasia
A'= A'-Scores calculated according to Snodgrass et al. (1985)

Table B.2.2: Individual results of the aphasic listeners for the different dimensions.[1]

Initials	Audiovisual			Auditory only			Visual only		
	Place	Manner	Voicing	Place	Manner	Voicing	Place	Manner	Voicing
WB (W)	83%	100%	67%	100%	100%	50%	67%	67%	50%
(A')	(95%)	(99%)	(90%)	(99%)	(99%)	(85%)	(82%)	(82%)	(74%)
BB (G)	67%	60%	17%	50%	50%	17%	33%	17%	17%
(A')	(90%)	(88%)	(72%)	(82%)	(82%)	(63%)	(86%)	(57%)	(57%)
EK (A)	83%	100%	17%	83%	67%	67%	33%	17%	50%
(A')	(93%)	(98%)	(66%)	(95%)	(91%)	(91%)	(60%)	(36%)	(71%)
TB (G)	100%	100%	83%	67%	100%	50%	50%	67%	33%
(A')	(99%)	(99%)	(94%)	(86%)	(97%)	(80%)	(70%)	(79%)	(58%)
JH (M)	100%	100%	67%	100%	67%	83%	100%	0%	33%
(A')	(100%)	(100%)	(92%)	(100%)	(92%)	(96%)	(99%)	(24%)	(78%)
MB (G)	50%	67%	100%	50%	50%	17%	50%	17%	33%
(A')	(78%)	(85%)	(96%)	(80%)	(80%)	(60%)	(66%)	(22%)	(52%)
MEAN	81%	88%	58%	75%	72%	47%	56%	31%	36%
(A')	(92%)	(95%)	(85%)	(90%)	(90%)	(79%)	(75%)	(50%)	(65%)

[1] W = Wernicke's Aphasia, G = Global Aphasia, A = Anomia, M = Mixed Aphasia
A'= A'-Scores calculated according to Snodgrass et al.

B.3 Statistics with A'-scores

Table B.3.1: Comparison of conditions for the aphasic group using A'- Scores.

Comparison	Test used	Teststatistics	df	p
AV vs AO vs VO	Friedman Anova	χ^2=10.333	2	p=0.006
AV vs AO	Wilcoxon Test	Z=1.725		p=0.084
AV vs VO	Wilcoxon Test	Z=2.201		p=0.028
AO vs VO	Wilcoxon Test	Z=2.201		p=0.028
AV vs AO (place)	Wilcoxon Test	Z=0.406		p=0.684
AV vs AO (manner)	Wilcoxon Test	Z=2.023		p=0.043
AV vs AO (voicing)	Wilcoxon Test	Z=0.943		p=0.345

Table B.3.2: Comparison of dimensions for the aphasic group using A'- Scores.

Comparison	Test used	Teststatistics	df	p
	Auditory only			
1 vs 2 vs 3 dimensions	Friedman Anova	χ^2=8.667	2	p=0.013
2 vs 1 dimensions	Wilcoxon Test	Z=1.826		p=0.068
3 vs 1 dimensions	Wilcoxon Test	Z=2.201		p=0.028
3 vs 2 dimensions	Wilcoxon Test	Z=1.089		p=0.276
place vs manner vs voicing	Friedman Anova	χ^2=6.700	2	p=0.035
manner vs place	Wilcoxon Test	Z=0.000		p=1.000
voicing vs place	Wilcoxon Test	Z=2.207		p=0.027
voicing vs manner	Wilcoxon Test	Z=1.753		p=0.080
	Audiovisual			
1 vs 2 vs 3 dimensions	Friedman Anova	χ^2=11.143	2	p=0.004
2 vs 1 dimensions	Wilcoxon Test	Z=2.226		p=0.026
3 vs 1 dimension	Wilcoxon Test	Z=2.207		p=0.027
3 vs 2 dimensions	Wilcoxon Test	Z=1.633		p=0.102
place vs manner vs voicing	Friedman Anova	χ^2=4.727	2	p=0.094

C Appendix to chapter 3: Audiovisual Processing: A McGurk Study

C.1 Pilot study

Table C.1.1: Results of the pilot study per participant. Values missing from 100% were other responses, such as leaving out the initial phoneme.

Participant	Age	Gender	Answertype		
			McGurk	Auditory	Visual
1	26	female	13%	62%	15%
2	27	female	62%	10%	26%
3	28	female	23%	31%	38%
4	29	female	0%	59%	28%
5	32	female	26%	56%	8%
6	32	male	36%	31%	23%
7	33	female	0%	41%	28%
8	43	male	10%	69%	10%
9	49	female	13%	8%	54%
10	50	female	46%	21%	26%
11	53	female	3%	31%	31%
12	58	female	41%	36%	23%
13	59	male	54%	3%	27%
14	62	male	23%	8%	59%
15	63	female	5%	41%	51%
16	67	female	61%	36%	3%
average: all	44		26%	34%	28%
average: >45 years	58		31%	23%	34%
average: <45 years	31		21%	45%	22%

Table C.1.2: Results of the pilot study per item. Values missing from 100% were other responses, such as leaving out the initial phoneme.

Item	Answertype		
	McGurk	Auditory	Visual
/tɪlm/	25%	31%	38%
/tɔ͡um/	50%	19%	31%
/tiːlp/	13%	38%	50%
/tɛnf/	50%	19%	19%
/tɑlp/	31%	25%	0%
/tøːp/	25%	13%	44%
/tøːlf/	25%	44%	19%
/tɪlx/	25%	25%	44%
/tɛms/	38%	38%	0%
/tɑrp/	6%	25%	19%
/taːrm/	38%	13%	6%
/tɪrk/	25%	56%	13%
/tyːst/	0%	31%	69%
/tyːl/	31%	6%	63%
/tiːx/	50%	44%	0%
/taːmst/	31%	44%	0%
/tɛŋ/	69%	13%	0%
/tɪrn/	38%	6%	19%
/tɛ͡ilf/	50%	38%	6%
/tɔ͡uf/	13%	38%	38%
/tuːm/	0%	6%	81%
/tulk/	38%	0%	63%
/tɑnf/	56%	19%	0%
/tøːf/	6%	6%	50%
/tyːm/	0%	6%	94%
/tɔ͡ux/	31%	19%	44%
/tɪlf/	25%	31%	31%
/tɔ͡un/	44%	31%	25%
/tuːsp/	6%	44%	50%
/tœ͡ynk/	50%	0%	50%

C.2 Stimuli

Table C.2.1: Overview of stimuli used: given are the rhyme and its respective onsets per condition together with the written answer choices that were provided.

Rhyme	Onset per condition AO	AV	VO	Written answer choice presented on top	middle	bottom
/ɪlm/	/t/	/k/	/p/	kilm	pilm	tilm
/ɔ͡um/	/p/	/t/	/k/	koum	poum	toum
/iːlp/	/t/	/k/	/p/	kielp	pielp	tielp
/ɛnf/	/k/	/p/	/t/	kenf	penf	tenf
/ɑlp/	/t/	/k/	/p/	kalp	palp	talp
/øːp/	/t/	/p/	/k/	keup	peup	teup
/øːlf/	/k/	/p/	/t/	keulf	peulf	teulf
/ɪlx/	/t/	/k/	/p/	kilg	pilg	tilg
/ɛms/	/p/	/k/	/t/	kems	pems	tems
/ɑrp/	/k/	/p/	/t/	karp	parp	tarp
/aːrm/	/p/	/t/	/k/	kaarm	paarm	taarm
/ɪrk/	/k/	/t/	/p/	kirk	pirk	tirk
/yːst/	/t/	/p/	/k/	kuust	puust	tuust
/yːl/	/t/	/p/	/k/	kuul	puul	tuul
/tiːx/	/p/	/t/	/k/	kieg	pieg	tieg
/aːmst/	/k/	/p/	/t/	kaamst	paamst	taamst
/ɛŋ/	/p/	/k/	/t/	keng	peng	teng
/ɪrn/	/p/	/t/	/k/	kirn	pirn	tirn
/ɛ͡ilf/	/p/	/t/	/k/	kijlf	pijlf	tijlf
/ɔ͡uf/	/t/	/p/	/k/	kouf	pouf	touf
/uːm/	/k/	/t/	/p/	koem	poem	toem
/ulk/	/k/	/t/	/p/	koelk	poelk	toelk
/ɑnf/	/p/	/k/	/t/	kanf	panf	tanf
/øːf/	/k/	/t/	/p/	keuf	peuf	teuf
/yːm/	/p/	/k/	/t/	kuum	puum	tuum
/ɔ͡ux/	/t/	/k/	/p/	koug	poug	toug
/ɪlf/	/k/	/p/	/t/	kilf	pilf	tilf
/ɔ͡un/	/p/	/k/	/t/	koun	poun	toun
/uːsp/	/t/	/p/	/k/	koesp	poesp	toesp
/œ͡ynk/	/k/	/t/	/p/	kuink	puink	tuink

Table C.2.2: Fillers used to compensate for too many /t/- responses in the McGurk condition.

Stimuli					
/kœ͡yx/	/kɛrts/	/køːx/	/pœ͡yp/	/pɛrts/	/pøːrf/
/kɑls/	/keːm/	/kɔ͡uk/	/pœ͡yx/	/peːf/	/pøːst/
/kɑlx/	/keːx/	/kɔ͡ul /	/pɑlx/	/peːx/	/pøːx/
/kɑnk/	/kɪrf/	/kɔ͡up/	/paːf/	/pɪrf/	/pɔ͡up/
/kɛ͡ip/	/koŋ/	/kurf/	/pɛ͡if/	/polt/	/purf/
/kɛ͡is/	/køːlp/	/kuːx/	/pɛkst/	/poŋ/	/pʏŋ/
/kɛlm/	/køːm/	/kʏlm/	/pɛlm/	/pøːlp/	/pʏrf/
/kɛlx/	/koːm/	/kʏŋ/	/pɛlp/	/poːlt/	/pyːf/
/kɛnk/	/køːrn/	/kyːf/	/pɛlx/	/pøːm/	/pyːk/
/kɛrp/	/køːst/	/kyːs/	/pɛrp/	/poːm/	/pyːs/

C.3 Results

Table C.3.1: Demographics and individual results of the non-brain-damaged control participants per condition.

Participant	Age	Gender	Auditory Only		Audiovisual		Visual Only	
			correct	RT	correct	RT	correct	RT
C1	52	f	100%	1396ms	100%	1322ms	93%	1944ms
C2	52	m	100%	1085ms	100%	1091ms	69%	1773ms
C3	54	m	100%	1807ms	100%	1786ms	83%	2304ms
C4	52	m	100%	1789ms	100%	1657ms	70%	2682ms
C5	49	f	100%	1411ms	100%	1342ms	82%	2350ms
C6	55	f	100%	1489ms	100%	1627ms	76%	2354ms
C7	59	m	100%	1643ms	100%	1530ms	67%	2418ms
C8	64	f	100%	1398ms	100%	1315ms	77%	1920ms
C9	64	f	97%	1169ms	100%	1096ms	90%	1674ms
C10	50	m	97%	1357ms	100%	1421ms	76%	1937ms
C11	65	m	100%	1534ms	100%	1449ms	83%	2155ms
C12	49	f	100%	1392ms	100%	1374ms	80%	2142ms
C13	61	f	100%	1511ms	100%	1422ms	71%	2674ms
C14	62	m	90%	1514ms	97%	1470ms	76%	1985ms

Table C.3.2: Individual answer patterns in the McGurk condition for the non-brain-damaged control participants.

Participant	McGurk (/t/)		Auditory (/p/)		Visual (/k/)	
	Incidence	RT	Incidence	RT	Incidence	RT
C1	31%	2806ms	14%	2260ms	55%	1823ms
C2	7%	1468ms	83%	1153ms	10%	1394ms
C3	47%	1903ms	6%	2522ms	47%	1955ms
C4	0%	—	100%	1624ms	0%	—
C5	10%	2553ms	10%	1626ms	80%	1722ms
C6	13%	3048ms	74%	1887ms	13%	2487ms
C7	46%	1922ms	27%	2636ms	27%	2087ms
C8	7%	1341ms	0%	—	93%	1818ms
C9	13%	1136ms	30%	1125ms	57%	903ms
C10	7%	2776ms	3%	2370ms	90%	1237ms
C11	50%	1885ms	20%	2262ms	30%	2058ms
C12	43%	1564ms	54%	1406ms	3%	3617ms
C13	17%	2615ms	33%	1622ms	50%	1789ms
C14	20%	1987ms	13%	1773ms	67%	1423ms

D Appendix to Chapter 5: Brain Correlates of Phonemic Processing

D.1 Individual results

Table D.1.1: Individual results in the tones condition for each stimulus type. Next to the participant number, age and gender (m=male, f=female) of the participants are provided. Participant 4 showed a rather low recognition rate of the less distant deviant. This could be an indication of a mild hearing disorder. He, however, showed normal performance for the syllable tasks and was, therefore not excluded. Also, the ERP is not influenced by his lower accuracy, as only trials with correct answers were included in the analysis.

Participant	standard 1000Hz	deviant 1050Hz		deviant 1200Hz	
	% correct	% correct	RT	% correct	RT
1 (66,m)	99.9 %	100 %	533ms	100 %	460ms
2 (67,m)	100 %	100 %	428ms	100 %	363ms
3 (59,f)	100 %	99 %	547ms	99 %	474ms
4 (69,m)	99.6 %	45 %	522ms	100 %	429ms
5 (53,f)	100 %	100 %	426ms	100 %	348ms
6 (46,f)	100 %	100 %	578ms	100 %	549ms
7 (65,f)	100 %	100 %	543ms	100 %	491ms
8 (56,f)	100 %	94.5 %	616ms	100 %	457ms
9 (45,f)	100 %	94 %	617ms	98 %	549ms
10 (60,m)	99.8 %	100 %	383ms	100 %	288ms
11 (63,f)	100 %	97.6 %	560ms	95 %	543ms
12 (57,f)	99.9 %	100 %	501ms	100 %	465ms
13 (61,m)	100 %	96.4 %	635ms	98.8 %	436ms

Table D.1.2: Individual results in the auditory syllables condition for each stimulus type. Next to the participant number, age and gender (m=male, f=female) of the participants are provided.

Participant	standard /pa/	deviant /ka/		deviant /ta/	
	% correct	% correct	RT	% correct	RT
1 (66,m)	100 %	73 %	961ms	73 %	960ms
2 (67,m)	99.9 %	93 %	867ms	88 %	865ms
3 (59,f)	100 %	88 %	913ms	91 %	924ms
4 (69,m)	99.9 %	93 %	868ms	94 %	858ms
5 (53,f)	100 %	85 %	878ms	79 %	855ms
6 (46,f)	100 %	76 %	922ms	86 %	931ms
7 (65,f)	99.9 %	57 %	966ms	62 %	961ms
8 (56,f)	99.9 %	94 %	907ms	95 %	900ms
9 (45,f)	99.9 %	26 %	970ms	30 %	957ms
10 (60,m)	99.8 %	99 %	811ms	97 %	817ms
11 (63,f)	100 %	49 %	926ms	83 %	939ms
12 (57,f)	100 %	67 %	922ms	85 %	916ms
13 (61,m)	100 %	71 %	949ms	85 %	939ms

Table D.1.3: Individual results in the visual syllables condition for each stimulus type. Next to the participant number, age and gender (m=male, f=female) of the participants are provided.

Participant	standard /pa/	deviant /ka/		deviant /ta/	
	% correct	% correct	RT	% correct	RT
1 (66,m)	99.9 %	94 %	721ms	100 %	655ms
2 (67,m)	99.9 %	82 %	772ms	98 %	750ms
3 (59,f)	100 %	98 %	743ms	100 %	676ms
4 (69,m)	99.9 %	96 %	614ms	99 %	628ms
5 (53,f)	100 %	96 %	621ms	100 %	560ms
6 (46,f)	100 %	100 %	678ms	100 %	666ms
7 (65,f)	99.8 %	95.2 %	773ms	100 %	733ms
8 (56,f)	100 %	100 %	577ms	100 %	568ms
9 (45,f)	99.9 %	79 %	860ms	91 %	840ms
10 (60,m)	99.5 %	97 %	526ms	100 %	494ms
11 (63,f)	100 %	75 %	885ms	72 %	866ms
12 (57,f)	100 %	97 %	660ms	100 %	582ms
13 (61,m)	99.9 %	100 %	684ms	100 %	679ms

Table D.1.4: Individual results in the audiovisual syllables condition for each stimulus type. Next to the participant number, age and gender (m=male, f=female) of the participants are provided.

Participant	standard /pa/	deviant /ka/		deviant /ta/		deviant McGurk	
	% correct	% correct	RT	% correct	RT	% correct	RT
1 (66,m)	99.9 %	98 %	780ms	98 %	711ms	92 %	754ms
2 (67,m)	99.8 %	74 %	900ms	84 %	872ms	51 %	918ms
3 (59,f)	99.8 %	84 %	891ms	94 %	872ms	85 %	901ms
4 (69,m)	99.8 %	100 %	671ms	98 %	679ms	95 %	680ms
5 (53,f)	99.9 %	100 %	701ms	100 %	639ms	93.6 %	725ms
6 (46,f)	100 %	92 %	811ms	95 %	812ms	92 %	797ms
7 (65,f)	99.8 %	85 %	901ms	97 %	850ms	75 %	894ms
8 (56,f)	99.9 %	97 %	703ms	97 %	658ms	96 %	692ms
9 (45,f)	99.9 %	85.9 %	827ms	95 %	815ms	88.3 %	845ms
10 (60,m)	99.8 %	100 %	576ms	100 %	511ms	97 %	560ms
11 (63,f)	99.9 %	88 %	885ms	92 %	860ms	83 %	886ms
12 (57,f)	100 %	98 %	808ms	99 %	660ms	98 %	770ms
13 (61,m)	99.9 %	96 %	731ms	100 %	715ms	83 %	747ms

E Appendix to chapter 6: Reliability of ERP measures in aphasia

E.1 Number of valid trials

Table E.1.1: Number of valid (correct, artifact-free) trials. Values marked with an asterisk represent conditions which were not included in the analysis due to too little data. In the 'tones' condition the standard was a 1000Hz pure tone, deviant 1 a 1050Hz tone and deviant 2 a 1200Hz tone. In the remaining three conditions, the standard was /pa/, deviant 1 /ka/, and deviant 2 /ta/. In the audiovisual condition a third deviant, auditory /pa/ dubbed on a visual /ka/ was added to elicit the McGurk illusion.

Participant	Condition	standard	deviant1	deviant2	missed deviants	McGurk	Used trials
WB	Tones	133	47	70	33	—	48
	Auditory Syllables	93	25	24	53	—	
	Visual Syllables	142	58	62	23*	—	
	Audiovisual Syllables	203	55	65	8*	34	
TB	Tones	134	72	70	—	—	122
	Auditory Syllables	190	81	92	22*	—	
	Visual Syllables	146	61	71	11*	—	
	Audiovisual Syllables	289	92	90	12*	18*	
DM	Tones	186	85	94	1*	—	170
	Auditory Syllables	191	96	98	3*	—	
	Visual Syllables	196	88	95	9*	—	
	Audiovisual Syllables	297	100	97	2*	95	

E.2 Results of statistic analyses

WB

Auditory syllables

120–160ms: no main effect of stimulus type ($F(3,141)=1.548$, $p=0.219$)
200–240ms: no main effect of stimulus type ($F(3,141)=1.131$, $p=0.335$)
360–400ms: no main effect of stimulus type ($F(3,141)=0.379$, $p=0.728$)

Visual syllables

120–160ms: no main effect of stimulus type ($F(2,94)=0.304$, $p=0.714$)
200–240ms: no main effect of stimulus type ($F(2,94)=0.162$, $p=0.847$)
360–400ms: no main effect of stimulus type ($F(2,94)=0.694$, $p=0.476$)

Audiovisual syllables versus auditory syllables

120–160ms: no main effect of condition ($F(1,47)=0.003$, $p=0.959$)
200–240ms: no main effect of condition ($F(1,47)=0.696$, $p=0.408$)
360–400ms: no main effect of condition ($F(1,47)=0.089$, $p=0.767$)

McGurk stimuli

120–160ms: no main effect of stimulus type ($F(3,141)=0.460$, $p=0.676$)
200–240ms: no main effect of stimulus type ($F(3,141)=0.263$, $p=0.833$)
360–400ms: no main effect of stimulus type ($F(3,141)=2.278$, $p=0.102$)

TB

Auditory syllables

120–160ms: no main effect of stimulus type ($F(2,242)=0.693$, $p=0.496$)
200–240ms:

- no main effect of stimulus type ($F(2,242)=1.182$, $p=0.306$)
- interaction of stimulus type and frontality ($F(4,484)=3.647$, $p<0.05$): trend to main effect of stimulus type occipitally ($F(2,242)=2.678$, $p=0.074$): /ta/ more negative than /ka/ ($p<0.05$)

360–400ms:

- main effect of stimulus type ($F(2,242)=5.236$, $p<0.05$): standard more positive than deviant /ta/ ($p<0.01$) and deviant /ka/ ($p<0.05$)
- interaction of stimulus type and frontality ($F(4,484)=4.711$, $p<0.01$): effect strongest at frontal electrodes

Visual syllables

120–160ms:

- no main effect of stimulus type ($F(1.785)=1.902$, $p=0.151$)
- interaction between stimulus type and frontality ($F(2.349)=3.153$, $p<0.05$): main effect of stimulus type occipitally ($F(1.802)=7.093$, $p<0.01$): standard and deviant /ka/ more positive than deviant /ta/ ($p<0.05$, $p<0.001$)
- interaction between stimulus type and laterality ($F(2.826)=15.192$, $p<0.001$): main effect of stimulus type at left hemisphere electrodes ($F(1.746)=7.838$, $p<0.01$): deviant /ka/ more positive than /ta/ ($p<0.001$)

200–240ms:
- main effect of stimulus type (F(1.869)=4.084, $p<0.05$): deviant /ta/ more positive than deviant /ka/ ($p<0.05$).
- interaction of stimulus type and frontality (F(2.397)=5.508, $p<0.01$): effect strongest for frontal electrodes
- interaction of stimulus type and laterality (F(2.762)=11.199, $p<0.001$): effect smallest on right side electrodes

360–400ms:
- main effect of stimulus type (F(1.842)=35.736, $p<0.001$): deviant /ta/ more positive than deviant /ka/ ($p<0.001$), deviant /ka/ more positive than standard ($p<0.001$)
- interaction of stimulus type and laterality (F(3.014)=19.556, $p<0.001$): effect smallest at the right hemisphere electrodes

Audiovisual syllables versus auditory syllables

120–160ms: no main effect of condition (F(1,121)=0.554, p=0.458)
200–240ms: main effect of condition (F(1,121)=5.767, $p<0.05$):
corrected AV more negative than AO
360–400ms: no effect of condition (F(1,121)=2.062, p=0.154)

McGurk stimuli

There were too few correct, artifact-free responses for an evaluation.

DM

Auditory syllables

120–160ms:
- no effect of stimulus type (F(2,338)=0.270, p=0.746)
- interaction of stimulus type and frontality (F(4,676)=2.764, $p<0.05$): no main effects for stimulus type in either instance of frontality (frontal: F(2,338)=0.176, p=0.827; central: F(2,338)=0.021, p=0.973; occipital: F(2,338)=1.515, p=0.223)
- interaction of stimulus type and laterality (F(4,676)=3.695, $p<0.01$): no main effects for stimulus type in either instance of laterality (left: F(2,338)=0.385, p=0.666; midline: F(2,338)=0.679, p=0.494; right: F(2,338)=0.242, p=0.769)

200–240ms: main effect of stimulus type (F(2,338)=7.122, $p<0.01$):
standard more negative than deviants /ka/ ($p<0.05$) and /ta/ ($p<0.01$)

360–400ms:

- main effect stimulus type ($F(2,338)=5.873$, $p<0.01$): both deviants more positive than standard ($p<0.01$, for both comparisons)
- interaction of stimulus type and frontality ($F(4,676)=29.37$, $p<0.001$): effect strongest at occipital electrodes
- interaction between stimulus type and laterality ($F(4,676)=5.349$, $p<0.01$): effect least strong at left hemisphere electrodes

Visual syllables

120–160ms:

- main effect of stimulus type ($F(2,338)=5.107$, $p<0.01$): standard elicited more negative responses than either deviant (/ka/: $p<0.01$, /ta/: $p<0.05$)
- interaction of stimulus type with frontality ($F(4,676)=13.087$, $p<0.001$): effect strongest occipitally
- interaction of stimulus type and laterality ($F(4,676)=6.879$, $p<0.001$): effect strongest at left and midline electrodes

200–240ms:

- main effect of stimulus type ($F(2,338)=4.139$, $p<0.05$): standard more negative than deviant /ka/ ($p<0.05$)
- interaction of stimulus type and laterality ($F(4,676)=4.237$, $p<0.01$): effect strongest at midline electrodes

360–400ms:

- main effect of stimulus type ($F(2,338)=11.464$, $p<0.001$): both deviants more positive than standard (/ka/: $p<0.01$, /ta/: $p<0.001$)
- interaction of stimulus type and laterality ($F(4,676)=4.135$, $p<0.01$): effect largest at midline electrodes
- interaction of stimulus type and frontality ($F(4,676)=50.424$, $p=<0.001$): effect increases with posteriority of electrodes

Audiovisual syllables versus auditory syllables

120–160ms:

- no main effect of condition ($F(1,169)=1.112$, $p=0.293$)
- interaction of condition and frontality ($F(2,338)=12.931$, $p<0.001$): separate ANOVAs per instance of frontality did not yield significant results

200–240ms:

- trend to main effect ($F(1,169)=3.720$, $p=0.055$): AV-VO more negative AO
- interaction of condition and frontality: corrected AV more negative AO at central ($F(1,169)=4.654$, $p<0.05$) and occipital ($F(1,169)=10.418$, $p<0.01$) electrodes

360–400ms:

- no main effect of condition ($F(1,169)=0.622$, $p=0.431$)
- interaction of condition and frontality ($F(2,338)=32.006$, $p<0.001$): AO more positive than AV at occipital electrodes ($F(1,169)=6.653$, $p<0.05$)

McGurk stimuli

120–160ms:

- no main effect of stimulus type ($F(3,507)=0.968$, $p=0.4$)
- interaction of stimulus type and frontality ($F(6,1014)=7.553$, $p<0.001$): separate ANOVAs per instance of frontality did not yield significant results

200–240ms:

- no main effect of stimulus type ($F(3,507)=0.328$, $p=0.778$)
- interaction of stimulus type and frontality ($F(6,1014)=15.292$, $p<0.001$): separate ANOVAs per instance of frontality did not yield significant results

360–400ms:

- no main effect of stimulus type ($F(3,507)=1.96$, $p=0.131$)
- interaction of stimulus type and laterality ($F(6,1014)=2.627$, $p<0.05$): main effect of stimulus type at left hemisphere electrodes ($F(3,507)=3.328$, $p<0.05$): McGurk stimuli more positive than standard stimuli ($p<0.05$);
- interaction between stimulus type and frontality ($F(6,1014)=2.999$, $p<0.05$): separate ANOVAs per instance of frontality did not yield significant results

References

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

Aaltonen, O., Tuomainen, J., Laine, M., & Niemi, P. (1993). Cortical differences in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN). *Brain and Language*, *44*, 139–152.

Aloufy, S., Lapidot, M., & Myslobodsky, M. (1996). Differences in susceptibility to the "blending illusion" among native Hebrew and English speakers. *Brain and Language*, *53*, 51–57.

Auther, L. L., Wertz, R. T., Miller, T. A., & Kirshner, H. S. (2000). Relationships among the mismatch negativity (MMN) response, auditory comprehension, and site of lesion in aphasic adults. *Aphasiology*, *14*, 461–470.

Bastiaanse, R., Bosje, M., & Visch-Brink, E. (1995). *Psycholinguistic assessment of language processing in aphasia: The Dutch version.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Berger, H. (1929). Über das Elektroenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*, 527–570.

Blumstein, S. E. (1994). Impairments of speech production and speech perception in aphasia. *Philosophical Transactions: Biological Sciences*, *346*, 29–36.

Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, *15*, 19–30.

Blumstein, S. E., & Cooper, W. E. (1972). Identification versus discrimination of distinctive features in speech perception. *Quarterly Journal of Experimental Psychology*, *24*, 207–214.

Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer [computer progam].* Version 5.2.17, retrieved 6 February 2009 from http://www.praat.org/.

Böttcher-Gandor, C., & Ullsperger, P. (1992). Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval. *Psychophysiology*, *29*, 546–550.

Breeuwer, M., & Plomp, R. (1986). Speechreading supplemented with auditorily presented speech parameters. *The Journal of the Acoustical Society of America*, *79*, 481–499.

Brown, J. (1974). *Aphasia, apraxia and agnosia: Clinical and theoretical aspects.* Springfield, Illinois: Charles C. Thomas.

Buchman, A., Garron, D. C., Trost-Cardamone, J. E., Wichter, M. D., & Schwartz, M. (1986). Word deafness: One hundred years later. *Journal of Neurology, Neurosurgery and Psychiatry*, *49*, 489–499.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*, 204–220.

Campbell, R. (1988). Tracing lip movements: Making speech visible. *Visible Language*, *22*, 32–57.

Campbell, R. (1990). Lipreading, neuropsychology, and immediate memory. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 268–286). Cambridge: Cambridge University Press.

Campbell, R., Dodd, B., & Burnham, D. (1998). Introduction. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye II – Advances in the psychology of speechreading and auditory-visual speech* (pp. ix–xiv). Hove: Psychology Press.

Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., & Regard, M. (1990). Neuropsychological studies of auditory-visual fusion illusions. Four case studies and their implications. *Neuropsychologia*, *28*, 787–802.

Caplan, D., & Aydelott-Utman, J. (1994). Selective acoustic phonetic impairment and lexical access in an aphasic patient. *The Journal of the Acoustical Society of America*, *95*, 512–517.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Coles, M. G. H., & Rugg, M. D. (1996). Event-related brain potentials: An introduction. In M. G. H. Coles & M. D. Rugg (Eds.), *Electrophysiology* *of mind, event-ralated brain potentials and cognition* (pp. 1–27). Oxford: Oxford University Press.

Colin, C., Radeau, M., & Deltenre, P. (1998). Intermodal interactions in speech: A French study. In *AVSP'98 International Conference on Auditory-Visual Speech Processing.*

Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clinical Neurophysiology*, *115*, 1989–2000.

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. *Clinical Neurophysiology*, *113*, 495–506.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*, 131–143.

Csépe, V., Osman-Sági, J., Molnár, M., & Gósy, M. (2001). Impaired speech perception in aphasic patients: Event-related potential and neuropsychological assessment. *Neuropsychologia*, *39*, 1194–1208.

Deacon, D., Nousak, J. M., Pilotti, M., Ritter, W., & Yang, C.-M. (1998). Automatic change detection: Does the auditory system use representations of individual stimulus features or gestalts? *Psychophysiology*, *35*, 413–419.

De Gelder, B., Bertelson, P., Vroomen, J., & Chen, H. C. (1995). Inter-language differences in the McGurk effects for Dutch and Cantonese listeners. In *Proceedings of the Fourth European Conference on Speech Communication and Technology* (pp. 1699–1702).

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*, 919–924.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-related brain potentials in man* (pp. 349–411). New York: Academic Press.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456–467.

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology.* Hove: Lawrence Erlbaum.

Franklin, S. (1989). Dissociations in auditory word comprehension: Evidence from nine fluent aphasic patients. *Aphasiology*, *3*, 189–207.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, *13*, 361–377.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.

Gielewski, E. J. (1989). Acoustic analysis and auditory retraining in the remediation of sensory aphasia. In C. Code & D. J. Muller (Eds.), *Aphasia therapy* (pp. 138–145). London: Whurr.

Graetz, P., De Bleser, R., & Willmes, K. (1992). *Akense afasietest (AAT).* Amsterdam: Hogrefe.

Grant, K. W., Ardell, L. H., Kuhl, P. K., & Sparks, D. W. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *The Journal of the Acoustical Society of America*, *77*, 671–677.

Grauwinkel, K., & Fagel, S. (2006). Crossmodal integration and McGurk-effect in synthetic audiovisual speech. In *Proceedings of the International Conference on Speech and Computer.*

Grayson, E., Hilton, R., & Franklin, S. (1997). Early intervention in a case of jargon aphasia: Efficacy of language comprehension therapy. *European Journal of Disorders of Communication*, *32*, 257–276.

Hessler, D. (2007). *War die störungsspezifische Therapie der auditiven Analyse effektiv? Eine Einzelfallstudie bei Aphasie.* Unpublished Diploma thesis, University of Potsdam.

Hessler, D., Jonkers, R., & Bastiaanse, R. (2010). The influence of phonetic dimensions on aphasic speech perception. *Clinical Linguistics and Phonetics*, *24*, 980–996.

Hessler, D., Jonkers, R., & Bastiaanse, R. (2011). Processing of audiovisual stimuli in aphasic and non-brain-damaged listeners. *Aphasiology*. Advance online publication. doi:10.1080/02687038.2011.608840

Hessler, D., & Stadie, N. (2008). War die störungsspezifische Therapie der auditiven Analyse effektiv? Eine Einzelfallstudie bei Aphasie. In M. Wahl, J. Heide, & S. Hanne (Eds.), *Spektrum Patholinguistik 1* (pp. 131–134). Potsdam: Universitätsverlag Potsdam.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology*, *34*, 33–61.

Howard, D., & Franklin, S. (1988). *Missing the meaning? A cognitive neuropsychological study of the processing of words by an aphasic patient.* Cambridge, MA: MIT Press.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*, 259–273.

Jacobs, B., & Schneider, S. (2003). Analysis of lexical-semantic processing and extensive neurological, electrophysiological, speech perception, and language evaluation following a unilateral left hemisphere lesion: Pure word deafness? *Aphasiology*, *17*, 123–141.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis.* Cambridge: MIT Press.

Janse, E. (2006). Lexical competition effects in aphasia: Deactivation of lexical candidates in spoken word processing. *Brain and Language*, *97*, 1–11.

Jansen, W. (2004). Laryngeal contrast and phonetic voicing (Doctoral dissertation, University of Groningen). *Groningen Dissertations in Linguistics (GRODIL)*, *47*.

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical neurophysiology*, *10*, 371–375.

Johnson, R. (1984). P300: A model of the variables controlling its amplitude. *Annals of the New York Academy of Sciences*, *425*, 223–229.

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*, 367–384.

Kant, I. (1787). *Kritik der reinen Vernunft.* Riga: Hartknoch.

Kay, J., Lesser, R., & Coltheart, M. (1992). *PALPA: Psycholinguistic assesment of language processing in aphasia.* Lawrence Erlbaum.

Kertesz, A. (1982). *Western aphasia battery.* New York: Grune & Stratton.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., & Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology*, *67*, 131–143.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*, 279–312.

Klitsch, J. (2008). Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia (Doctoral dissertation, University of Groningen). *Groningen Dissertations in Linguistics (GRODIL)*, *67*.

Kraus, N., McGee, T., Sharma, A. M. A., Carrell, T., & Nicol, T. B. S. (1992). Mismatch negativity event-related potential elicited by speech stimuli. *Ear and Hearing*, *13*, 158–164. Using Smart Source Parsing

Kussmaul, A. (1877). *Die Störungen der Sprache.* Leipzig: Vogel.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time.

Science, *197*, 792–795.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages.* Oxford: Blackwell Publishers Ltd.

Lawson, E. A., & Gaillard, A. W. K. (1981). Mismatch negativity in a phonetic discrimination task. *Biological Psychology*, *13*, 281–288.

L'Ecuyer, P. (1988). Efficient and portable combined random number generators. *Communications of the ACM*, *31*, 742-749, 774.

Lehmann, D. (1987). Principles of spatial analysis. In A. S. Gevins & A. Remond (Eds.), *Handbook of electroencephalography and clinical neurophysiology* (pp. 309–354). Amsterdam: Elsevier.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Lisker, L., & Rossi, M. (1992). Auditory and visual cueing of the [± rounded] feature of vowels. *Language and Speech*, *35*, 391–417.

Loh, M., Schmid, G., Deco, G., & Ziegler, W. (2010). Audiovisual matching in speech and nonspeech sounds: A neurodynamical model. *Journal of Cognitive Neuroscience*, *22*, 240–247.

Luck, S. J. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: MIT Press.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics*, *24*, 253–257.

MacNeilage, P. F. (1991). Comment: The gesture as a unit in speech perception theories. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception: Proceedings of a conference to honor alvin m. liberman* (pp. 61–68). Hillsdale, N.J.: Lawrence Erlbaum.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, *16*, 68–89. Article Using Smart Source Parsing

Manuel, S. Y., Repp, B. H., Studdert-Kennedy, M., & Liberman, A. M. (1983). Exploring the "McGurk effect". *Journal of the Acoustical Society of America*, *74*, 66.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin Review*, *15*, 453–457.

Massaro, D. W., Cohen, M. M., Gesi, A., Heredia, R., & Tsuzaki, M. (1993).

Bimodal speech perception: An examination across languages. *Journal of Phonetics*, *21*, 445–478.

Massaro, D. W., & Oden, G. C. (1995). Independence of lexical context and phonological information in speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1053–1064.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McGee, T. J., King, C., Tremblay, K., Nicol, T. G., Cunningham, J., & Kraus, N. (2001). Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology*, *38*, 653–658.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Morris, J., Franklin, S., Ellis, A. W., Turner, J., & Bailey, P. J. (1996). Remediating a speech perception deficit in an aphasic patient. *Aphasiology*, *10*, 137–158.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76,*, 165–178.

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research*, *13*, 417–425.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*, 6–18.

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

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1987). Inter-stimulus interval and the mismatch negativity. In C. Barber & T. Blum (Eds.), *Evoked potentials iii* (pp. 392–397). London: Butterworths.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590.

Näätänen, R., Simpson, M., & Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology*, *14*, 53–98.

Norrix, L. W., Plante, E., & Vance, R. (2006). Auditory-visual speech integration by adults with and without language-learning disabilities. *Journal of Communication Disorders*, *39*, 22–36.

Novak, G., Ritter, W., & Vaughan Jr., H. G. (1992). Mismatch detection and the latency of temporal judgments. *Psychophysiology*, *29*, 398–411.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*, 172–191.

Ohde, R. N., & Abou-Khalil, R. (2001). Age differences for stop-consonant and vowel perception in adults. *Journal of the Acoustical Society of America*, *110*, 2156–2166.

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: A review. *Biological Psychology*, *63*, 199–236.

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*, 25–28.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 97–114). London: Lawrence Erlbaum.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native english speakers. *Behavioural Brain Research*, *111*, 13–23.

Robert-Ribes, J., Piquemal, M., Schwartz, J.-L., & Escudier, P. (1996). Exploiting sensor fusion architectures and stimuli complementarity in av speech recognition. In D. Storck & M. Hennecke (Eds.), *Speechreading by humans and machines* (pp. 193–210). Berlin: Springer.

Rosen, S. M., Fourcin, A. J., & Moore, B. C. J. (1981). Voice pitch as an aid to lipreading. *Nature*, *291*, 150–152.

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405–409.

Saffran, E., Marin, O., & Yeni-Komshian, G. (1976). An analysis of speech perception in word deafness. *Brain and Language*, *3*, 209–28.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*, 587–597.

Sams, M., Aulanko, R., Aaltonen, O., & Näätänen, R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *Journal of Cognitive Neuroscience*, *2*, 344–357.

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

Schmid, G., Thielmann, A., & Ziegler, W. (2009). The influence of visual and auditory information on the perception of speech and non-speech oral movements in patients with left hemisphere lesions. *Clinical Linguistics and Phonetics*, *23*, 208–221.

Schmid, G., & Ziegler, W. (2006). Audio-visual matching of speech and non-speech oral gestures in patients with aphasia and apraxia of speech. *Neuropsychologia*, *44*, 546–555.

Schwartz, J.-L. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. *The Journal of the Acoustical Society of America*, *127*, 1584–1594.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception and Psychophysics*, *59*, 73–80.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, *90*, 1797–1805.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*, 1078–1083.

Sharma, A., Kraus, N., Carrell, T., & Thompson, C. (1994). Neurophysiologic bases of pitch and place of articulation perception: A case study. *The Journal of the Acoustical Society of America*, *95*, 3011.

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*, 64–71.

Shindo, M., Kaga, K., & Tanaka, Y. (1991). Speech discrimination and lipreading in patients with word deafness or auditory agnosia. *Brain and Language*, *40*, 153–161.

Smith, N. K. (1965). *Critique of pure reason.* New York: St. Martins Press.

Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). *Human experimental psychology.* New York: Oxford University Press.

Soto-Faraco, S., & Alsius, A. (2007). Conscious access to the unisensory components of a cross-modal illusion. *Neuroreport*, *18*, 347–350.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *Journal of Experimental Psychology: Human Perception Performance*, *35*, 580–587.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, *92*, B13–B23.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE:

A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods*, *39*, 19–30.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*, 212–215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–51). London: Lawrence Erlbaum.

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *36*, 51–74.

Tsui, V. (2000). *Mismatch negativity cortical event-related potential.* Unpublished master's thesis, University of British Columbia.

Wertz, R. T., Auther, L. L., Burch-Sims, G. P., Abou-Khalil, R., Kirshner, H. S., & Duncan, G. W. (1998). A comparison of the mismatch negativity (MMN) event-related potential to tone and speech stimuli in normal and aphasic adults. *Aphasiology*, *12*, 499–507.

Wiener, D., Connor, L., & Obler, L. (2004). Inhibition and auditory comprehension in Wernicke's aphasia. *Aphasiology*, *18*, 599–609.

Yee, E., Blumstein, S., & Sedivy, J. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Journal of cognitive neuroscience*, *20*, 592–612.

Youse, K. M., Cienkowski, K. M., & Coelho, C. A. (2004). Auditory-visual speech perception in an adult with aphasia. *Brain Injury*, *18*, 825–834.

Groningen Dissertations in Linguistics (GRODIL)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach.*
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure.*
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation.*
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation.*
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar.*
6. Peter Blok (1993). *The Interpretation of Focus: an epistemic approach to pragmatics.*
7. Roelien Bastiaanse (1993). *Studies in Aphasia.*
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist.*
9. Wim Kosmeijer (1993). *Barriers and Licensing.*
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach.*
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity.*
12. Ton van der Wouden (1994). *Negative Contexts.*
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar.*
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar.*
15. Maarten de Wind (1995). *Inversion in French.*
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance.*
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition.*
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items.*
19. Karen Lattewitz (1997). *Adjacency in Dutch and German.*
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch.*
21. Henny Klein (1997). *Adverbs of Degree in Dutch.*

22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The Case of French 'des'/'du'-NPs.*
23. Rita Landeweerd (1998). *Discourse Semantics of Perspective and Temporal Structure.*
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program.*
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in Aphasic Speakers.*
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics.*
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses.*
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*
29. Hae-Kyung Wee (1999). *Definite Focus.*
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean Tense and Aspect in Discourse.*
31. Ivilin Stoianov (2001). *Connectionist Lexical Processing.*
32. Klarien van der Linde (2001). *Sonority Substitutions.*
33. Monique Lamers (2001). *Sentence Processing: Using Syntactic, Semantic, and Thematic Information.*
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement.*
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding.*
36. Esther Ruigendijk (2002). *Case Assignment in Agrammatism: a Cross-linguistic Study.*
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and Segments in Level-specific Deficits.*
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*
43. Hein van Schie (2003). *Visual Semantics.*
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures.*
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.*
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology Approach to English, Hungarian and Dutch.*
48. Judith Rispens (2004). *Syntactic and Phonological Processing in Developmental Dyslexia.*
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: les cas de la planification familiale, du sida et de l'excision.*

50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin.*
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis.*
54. Leonoor van der Beek (2005). *Topics in Corpus-Based Dutch Syntax.*
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency.*
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality.*
59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*
61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*
63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*
66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*
69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering.*
70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition.*
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*

77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment.*
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval.*
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg.*
81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships.*
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.*
84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning.*
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development.*
86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Students.*
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students.*
88. Jelena Prokić (2010). *Families and Resemblances.*
89. Radek Šimík (2011). *Modal existential wh-constructions.*
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease.*
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching.*
92. Aletta Kwant (2011). *Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters.*
93. Marlies Kluck (2011). *Sentence amalgamation.*
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish.*
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation.*
96. Barbara Plank (2011). *Domain Adaptation for Parsing.*
97. Çağrı Çöltekin (2011). *Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts.*

Grodil
Secretary of the Department of General Linguistics
Postbus 716
9700 AS Groningen
The Netherlands