Characterizing sensory and cognitive factors of human speech processing through eye movements

Dorothea Christine Wendt

Characterizing sensory and cognitive factors of human speech processing through eye movements

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

von Frau
Dorothea Christine Wendt
geboren am 6. Dezember 1982
in Celle

Gutachter: Prof. Dr. Dr. B. Kollmeier
Zweitgutachter: Prof. Dr. Steven van de Par
Tag der Disputation: 11. Juli 2013

Glossary

List of acronyms

ambSR	Ambiguous subject-relative
ambOR	Ambiguous object-relative
ambOVS	Ambiguous object-verb-subject
ANOVA	Analysis of variance
DDD	Disambiguation to decision delay
DM	Decision moment
ELU	Ease of language understanding
EOG	Electrooculographic
FDR	Finite detection rate
FIR	Finite impulse response
HI	Hearing impaired
HL	Hearing level
NL	Normal hearing
OLACS	Oldenburg linguistically and audiologically controlled sentences
OR	Object-relative
OVS	Object-verb-subject
PTA	Pure tone average
PTD	Point of target disambiguation
ROI	Region of interest

SNR	Signal-to-noise ratio
SPL	Sound pressure level
SR	Subject-relative
SRT	Speech reception threshold
SRT80	Speech reception threshold at 80 % word recognition
sTDA	Singe target detection amplitude
SVO	Subject-verb-object
TDA	Target detection amplitude

Contents

List of Figures

List of Tables

1

Introduction

Imagine an everyday situation where you are sitting on a bus. Some people are talking to each other, while others get on the bus and a voice announces the next bus stop. Now try to understand your partner sitting next to you and telling you about the working day. You will soon find it almost impossible having a conversation, which is due to several factors related to the surroundings. Some of them can have acoustical origin but also non-acoustical factors can influence speech understanding. Acoustical factors, like the presence of competing speakers or the engine noise of the driving bus, often have a disturbing effect on speech understanding. In particular in situations when the acoustic information lets you down but you are not entirely lost in the conversation, an integration of multi-sensory information, i.e. not only to hear the voice of the speaker but also to see her or his face can improve speech communication. Hence, non-acoustical environmental factors can lead to a better speech understanding. Moreover, individual abilities of the listener can affect speech understanding. These individual factors are partly of a physiological nature, for instance hearing impairment and cognitive abilities can negatively affect speech communication. Thus, difficulties in speech communication can occur due to disturbing factors related to the environment (external factors) and/or related to specific individual parameters (internal factors). In standard speech audiometry, speech *recognition* performance is often investigated while listeners repeat the heard items, i.e. words or sentences without necessarily extracting the meaning of the speech signal. But speech *understanding* - especially in dialogs such as in the aforementioned bus situation - is required to extract the meaning of the speech signal, i.e. to comprehend and to interpret what your partner says. Hence, speech understanding involves higher level operations and cognitive processes which are not necessarily included in speech recognition. If one considers that in such a dialog the speech rate can reach about 140-180 words per minute (Wingfield and Tun, 2007), it seems reasonable that time-dependent effects in speech processing, i.e. when is a

sentence understood, play an important role.
In this thesis, both external (acoustical and non-acoustical) and internal (listener-specific) factors are considered. To that end, an eye-tracking paradigm is designed to gain further insights into the effect of these factors on the process of speech understanding. The primary aim is to develop a new method by the use of eye fixations in order to gain an *online* (i.e. *during* the presentation of the speech stimulus) investigation of the process of speech understanding. By using the novel method, it is possible to analyze the effect of external and internal factors on the speed of processing sentences, and therefore, to obtain a measure which - in contrast to standard audiological measures - allows a more detailed analysis of communication problems. In the following sections, previous research concerned with the effect of several factors on speech *recognition* is reviewed to provide a basis for the speech *understanding* studies carried out within the framework of this thesis.

1.1 Influence of external factors

Speech communication difficulties often arise from the acoustical environment, which may contain disturbing background noise or competing speakers. The characteristics of the noise signal play an important role for the difficulties in speech communication. Going back to the situation in the bus, the background noise level fluctuates over time, i.e. the noise is non-stationary, which is characteristic for most listening situations in daily life. It is well known that noise modulations can provide a listening benefit in comparison to non-modulated or stationary noise (Fastl, 1982, Duquestnoy, 1983, Sotscheck, 1985, Fastl and Zwicker, 2007), which is often referred to as *release from masking*. An early study conducted by Festen and Plomp (1990) investigated the speech recognition in modulated noise. For normally hearing listeners, they reported a benefit in the speech-reception-threshold (SRT, the signal-to-noise-ratio (SNR) required for 50 % correct speech recognition) of about 4-6 dB SNR, depending on the modulation characteristics of the noise masker. A release from masking indicates that listeners are able to exploit the dips in the envelope of the modulated masker to improve their speech recognition performance, which is known as "listening in the dips" (see Bronkhorst,2000 for a review). In contrast, hearing impairment seems to cause difficulties using these gaps, since a similar release from masking has not been found for hearing-impaired listeners (Duquestnoy, 1983, Gustafsson and Arlinger, 1994, Holube *et al.*, 1997, Wagener and Brand, 2005).
Moreover, listeners often use non-acoustical information to improve speech recognition. Visual information (e.g. lip-reading) in particular can help to overcome processing problems in difficult listening situations. Middelweerd and Plomp (1987) and MacLeod and Summerflied (1990) reported that audio-visual cues can lead to a decrease in the SRT compared to using auditory cues only. They investigated the improvement in the SRT due to lip-reading and observed a visual benefit of about 4-6 dB SNR using short sentences (five words on average).
Another non-acoustical factor influencing speech communication is related to the speech material

itself. Several studies have demonstrated that speech recognition performance can vary depending on the characteristics of the speech material, such as linguistic complexity or context. For example, recognition performance for words in a meaningful sentence increases compared to recognition performance of words presented in isolation (Miller *et al.*, 1951, Boothroyd and Nittrouer, 1988). Hence, the linguistic context can help to overcome speech recognition problems resulting from missing speech information in background noise. That is, knowing that your partner is telling you about his or her working day in the bus situation might help you to follow the conversation. A mathematical treatment of the effect of context on phoneme and word recognition was proposed by Boothroyd and Nittrouer by calculating the so-called k- and j-factors (Boothroyd and Nittrouer, 1988). The k-factor considers the error probabilities in phoneme or word recognition with and without context, and therefore, describes the contextual constraints on the recognition of linguistic units. The j-factor in turn is interpreted as the number of statistically independent parts within a whole, i.e. a measure of the perceiver's tendency to chunk the parts into larger perceptual units. Speech recognition performance is further affected by the level of linguistic complexity of the speech material. For instance, Uslar *et al.* (2011) could demonstrate that when controlling for linguistic complexity of sentences included in a standard speech recognition test called Göttinger sentence test (Kollmeier and Wesselkamp, 1997), speech intelligibility decreased as a function of complexity (Uslar *et al.*, 2011).

1.2 Influence of internal factors

Besides the external factors, individual abilities of the listener can affect speech communication. It has been argued that under everyday (unaided) listening conditions hearing loss is the most important cause of speech communication difficulties (see e.g. Kollmeier,1990, van Rooij and Plomp,1992, Humes,1991, 1996, 2002). Hearing impairment can lead to a less audible and degraded speech signal, which in turn can cause difficulties for the listener to understand incoming speech information and hence cause errors in speech recognition (Plomp and Mimpen, 1979). However, the audiogram, which is typically used to assess the degree of hearing impairment, it is not able to fully predict speech recognition performance by psychoacoustical or audiological models (see e.g. Rhebergen and Versfeld, 2005). In fact, speech communication involves not only sensory or signal-driven processes, for instance to detect and encode the speech signal. Moreover, understanding speech also requires cognitive, knowledge-driven processes to interpret and comprehend the encoded acoustical information. Besides impaired functioning of the inner ear, such as a loss of sensitivity as a function of frequency (typically measured with the audiogram) or a reduction or loss of the compressive functioning of the cochlea (recruitment phenomenon), cognitive abilities such as working memory capacity can influence speech comprehension in noise. Recently, a considerable interest has been raised in the scientific community in the relationship between individual cognitive abilities and speech recognition performance. Akeroyd (2008) presented an overview of experimental studies that reported a correlation between speech recognition

performance and measures of working memory capacity. Cognitive abilities were further tested in aided listening situations to investigate the predictive effect of cognitive factors on hearing-aid benefit (e.g. Gatehouse *et al.*, 2003; Lunner *et al.*, 2003, 2007, 2009). For that purpose, the relationship between cognitive measures, like individual differences in working memory capacity, and the performance in sentence processing under hearing aid signal processing was tested. In general, there is a link between cognitive performance and speech reception (in aided and unaided situations), indicating that a decrease in cognitive abilities can impair effective communication (Schneider *et al.*, 2010).

It is important to stress that age is related to both, i.e. sensory and cognitive changes. Thus, in particular elderly people complain about difficulties to follow conversation situations. However, age per se is not a disturbing factor. In fact, there is some evidence that in some listening situations older adults appear to derive a greater benefit from supportive context that younger adults (Kalikow *et al.*, 1977; Pichora-Fuller *et al.*, 1995, 2008).

1.3 Speech processing and eye movements

As demonstrated by the aforementioned studies, in audiological research speech perception is commonly investigated using speech audiometry, e.g., SRTs or recognition rates. However, several studies have demonstrated that measuring speech recognition performance using standard speech audiometry does not capture the whole picture of speech perception. For instance, it has been shown that even when the speech recognition rate is high, the perceived effort required to obtain successful recognition can be high. Furthermore, the effort can vary although the speech recognition performance is constant (Pichora-Fuller *et al.*, 1995, Surprenant, 1999, Fraser *et al.*, 2010). This already suggests that additional measures are required to better reveal communication difficulties, which may not be caught by standard audiometry.

Moreover, it was already mentioned above that cognitive or linguistic operations are needed to understand and interpret a speech signal. But speech recognition, which is typically tested with standard speech audiometry, does not necessarily require these higher level operations. That is, in contrast to standard speech audiometry, speech understanding requires the listener to comprehend the meaning of a sentence - a feat that is arguably very important in everyday conversations. As a consequence, to get a better insight into communication difficulties, novel objective measures of speech perception are needed to supplement traditional measures of speech recognition performance. Within the context of this thesis, the focus is on the development of a measure that allows studying the process of speech understanding. In particular, the novel measure should allow an *online* (in the sense of *during* the presentation of the spoken sentence) analysis of the time course of the speech understanding process to enable us to detect any time-dependent effects. Since an increase in processing demands due to disturbing factors such as background noise or linguistic complexity can lead to a slowing down of the system (Tun *et al.*, 2010a), it may prove worthwhile to obtain a measure of speed of sentence comprehension. A listener who is

slow at sentence processing may miss speech information late in the sentence because he/she is still processing a "backlog" of past speech information. In the long run this slowing down may prevent listeners from participating in a conversation. So far, there is no appropriate method in audiological research that allows for such an online analysis of speech understanding.
In order to get an online analysis of speech processing, eye-movements have frequently been used in psycholinguistic research. The highly temporal interplay between speech processing and eye-movements was first shown by the pioneering study by Cooper (1974) and was then confirmed in more recent studies. The *visual world paradigm* (Tanenhaus *et al.*, 1995, Allopenna *et al.*, 1998) was developed to reveal the interaction of vision and language by simultaneously presenting spoken language and visual scenes. In the basic setup of this paradigm, listeners hear a spoken utterance, such as a sentence or a single word, and look at an experimental display, which can be a visual scene depicted on a computer screen. The experimental display contains (visual) objects that are matched with the spoken utterance, or at least with parts of the utterance, and listeners spontaneously fixate on this object while their eye fixations are recorded with an eye-tracking device. Several subsequent studies have investigated how and when the acoustical and the visual information are integrated. Some of these studies examined the effect of the linguistic information on eye movements, and whether the interpretation of the linguistic information evoked anticipation of the corresponding visual information (Altmann and Kamide,1999, 2007, Kamide *et al.*, 2003). Furthermore, several studies analyzed how the depicted visual information constrained the inherent meaning of the acoustical input (see e.g.Trueswell *et al.*,1994, Chambers *et al.*,2004). Thus, the visual world paradigm provides an appropriate framework for an online investigation of the interplay between acoustical and visual information processing during speech understanding with a high temporal resolution (see also Huettig *et al.*,2011 for a review). However, there is no suitable approach so far, including the experimental design and the analysis of the eye fixation data, which is appropriate for audiological applications. That is, previous eye-tracking paradigms do not allow for an analysis of processing speed during speech understanding as a function of hearing status and/or individual cognitive abilities.

1.4 Aims and outline of this thesis

In order to shed light on the aforementioned difficulties in speech communication, which so far have typically been investigated using speech recognition measures and standard audiometry, this thesis focuses on the process of speech understanding with the following objectives:

1. To develop a new eye-tracking paradigm including appropriate data analysis that allows for an online investigation of the process of speech understanding. This paradigm should be sensitive to changes in sentence processing even at a high and constant level of speech intelligibility. For that purpose, sentences with different levels of linguistic complexity are used to gain insights into the processes underlying sentence understanding while manipulating

cognitive demands during processing.

2. To analyze how external factors, such as background noise and linguistic complexity, affect the process of speech understanding using the proposed paradigm. By systematically varying the sensory (via two different noise types) and cognitive (via changing the level of linguistic complexity) processing load, it is investigated if sensory and cognitive factors influencing the process of speech understanding are independent of one another or if an interaction of two factors occurs.

3. To analyze individual differences in the process of speech understanding due to intrinsic factors like hearing loss or cognitive abilities. With a view to a more audiological application of the eye-tracking paradigm, the focus in the third part of this thesis is on the investigation of the effect of hearing impairment on the process underlying sentence understanding even under conditions of equal intelligibility. Moreover, it is investigated if individual differences in this process can be explained by individual cognitive abilities.

The thesis is structured as follows:

Chapter 2 introduces an eye-tracking approach that allows for an online investigation of the process of speech understanding. The *target detection amplitude* (TDA), which is calculated from the eye fixation data, is introduced as a measure to analyze the time course of this process. By varying the level of linguistic complexity under conditions of good audibility, the process of speech understanding is tested with respect to processing speed as a function of cognitive processing load. For different sentence structures, the *disambiguation to decision delay* (DDD) is calculated from the TDA as measures of processing speed during speech understanding. The DDD is defined as the time interval that passes between the earliest possible point in the sentence where understanding would have been possible and the actual understanding of the sentence at the *decision moment* (DM) indicated by eye fixations. Furthermore, it is investigated if temporarily increasing difficulties in the process of speech understanding, such as a temporarily occurring misinterpretation of the sentence, can be detected with the proposed paradigm.

In **Chapter 3**, the effect of acoustical background noise and linguistic complexity on visual (picture) recognition performance and processing speed is investigated. For sentence structures with different levels of linguistic complexity the effect of the acoustical condition, i.e. quiet vs. stationary or modulated noise, on processing speed is tested. Sensory demands are manipulated during speech understanding by presenting the sentences in different noise conditions with comparable speech intelligibility level. Moreover, it is investigated if recognition performance in the proposed paradigm is able to reveal difficulties in speech processing caused by background noise and linguistic complexity, or if processing speed can provide a more sensitive measure for revealing processing difficulties.

To gain a better insight into how hearing impairment affects processing speed, a third study is conducted with elderly listeners, with normal hearing and with hearing impaired listeners (**Chapter 4**). Moreover, the analysis of the eye fixation data (introduced in Chapter 2) is modified in this chapter to gain a measure of processing speed for individual listeners. It was tested if hearing impairment can lead to an increase in processing speed even at a constant level of speech intelligibility. To analyze if and how individual differences in speech processing can be explained by individual cognitive abilities, correlations between processing speed and cognitive parameters are calculated.

Overall, this thesis introduces a novel audio-visual eye-tracking paradigm which was developed as an online analysis of the process of speech understanding with possible applications in the field of audiology. A wide range of factors, i.e., external and internal (listener-specific) factors, influencing speech processing were tested using the proposed paradigm to obtain a better understanding of the normal and impaired human auditory system. The new paradigm was found to be able to provide information about the online processing of speech understanding, such as at which point in time a sentence is understood, and as such constitutes a significant contribution to the field of hearing research.

Parts of this chapter are published as:

- Wendt *et al.* (2014): "An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities," PLoS ONE 9(6): e100186.

2

An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities

An eye-tracking paradigm was developed for use in audiology in order to enable online analysis of the speech comprehension process. This paradigm should be useful in assessing impediments in speech processing. In this paradigm, two scenes, a target picture and a competitor picture, were presented simultaneously with an aurally presented sentence that corresponded to the target picture. At the same time, eye fixations were recorded using an eye-tracking device. The effect of linguistic complexity on language processing time was assessed from eye fixation information by systematically varying linguistic complexity. This was achieved with a sentence corpus containing seven German sentence structures. A novel data analysis method computed the average tendency to fixate the target picture as a function of time during sentence processing. This allowed identification of the point in time at which the participant understood the sentence, referred to as the decision moment. Systematic differences in processing time were observed as a function of linguistic complexity. These differences in processing time may be used to assess the efficiency of cognitive processes involved in resolving linguistic complexity. Thus, the proposed method enables a temporal analysis of the speech comprehension process and has potential applications in speech audiology and psychoacoustics.

2.1 Introduction

Speech intelligibility tests are an indispensable tool in clinical audiology. They can evaluate the consequence of sensory hearing loss (characterized by a frequency dependent hearing impairment) for the patient's communication abilities (Laroche *et al.*, 2003, Ozimek *et al.*, 2010, Haumann *et al.*, 2012, Zokoll *et al.*, 2013). Beyond diagnostic applications, speech intelligibility tests are also often used to quantify the benefit of hearing aids or cochlear implants for individual patients. Typically, speech intelligibility tests measure the proportion of correctly repeated speech items, usually single words or single sentences (Plomp and Mimpen, 1979, Hagerman, 1982, Kollmeier and Wesselkamp, 1997, Nilsson *et al.*, 1994). However, research has shown that additional performance information about the ease of speech comprehension or cognitive effort during speech processing can complement traditional speech intelligibility measures. Increased cognitive effort is indicated by poorer task performance and processing time and can be measured in terms of recognition accuracy or reaction time, for instance (Wingfield *et al.*, 2006, Tun *et al.*, 2010a). The current study focuses on developing a method for assessing the speech comprehension process and processing speed as indicators of the cognitive effort required at levels of high intelligibility. The proposed method is characterized by two main aspects: Firstly, a special speech corpus is applied that is optimized for both speech intelligibility measurements and controlled variation of linguistic complexity. Secondly, eye movements are tracked to provide an online assessment of speech processing during sentence comprehension. This study aims to determine whether this combination of speech intelligibility testing and eye tracking can detect a systematic deceleration in speech processing due to an increase in cognitive processing effort that is sufficiently large and robust to be used in audiology. A further question is whether the deceleration effect is detected by either recognition scores or reaction times alone.

2.1.1 Speech intelligibility and linguistic complexity

Several studies reported that speech intelligibility is influenced by linguistic aspects of the speech material, such as context information, sentence structure, or level of complexity (Kalikow *et al.*, 1977, Boothroyd and Nittrouer, 1988, Uslar *et al.*, 2011). However, the role of linguistic aspects in speech comprehension, in particular in connection with hearing loss, has been largely neglected in standard audiological testing. In addition, speech intelligibility measurements provide little information about linguistic aspects in language comprehension, such as processing costs arising from different levels of cognitive load and/or linguistic complexity (Uslar *et al.*, 2011). Recently, Uslar et al. (Uslar *et al.*, 2013a) developed the Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) material to differentiate between acoustical and linguistic factors and their respective contributions to speech intelligibility measurement. Using the OLACS corpus, Uslar *et al.* (2013a) measured spech reception thresholds (SRT) and reported a small effect of complexity on speech intelligibility (about 1-2 dB). However, studies in which participants

were asked a comprehension question following sentence presentation revealed a stronger effect of linguistic complexity on sentence processing. For instance, Tun and colleagues (Tun *et al.*, 2010a) measured reaction times for sentences with different sentence structures presented at a clearly audible level. They observed reduced speech processing speeds for structures with higher linguistic complexity. It was argued that the reduced comprehension speed was caused by the increased cognitive processing demands of the more complex sentence structures. Hence, sentence complexity can lead to slower sentence processing. This suggests that sentence processing speed may be a more sensitive measure for detecting difficulties during sentence understanding than standard methods used in audiology, such as speech intelligibility tests. Reaction time, as reported by Tun et al. (Tun *et al.*, 2010a), and speech intelligibility measures are taken after the speech is presented. These offline measures do not provide any time-resolved information about the process of sentence comprehension, but instead reflect the end point of this process. On the other hand, an online analysis of processing time occurring *during* the presentation of the sentence is expected to provide a more direct measure of any temporal changes in speech processing that are not reflected by offline measures.
Another advantage of using response measures based on eye movements is their relative robustness against age effects (Pratt *et al.*, 2006); latency and reaction times using a button press exhibit age-related differences (Cerella and Hale, 1994). This is an important issue when testing listeners with hearing impairment because hearing loss typically increases with age. For this reason, this study recorded both eye fixation and reaction time derived from pressing buttons.

2.1.2 Analysis of eye movements with respect to speech processing

Eye movements are frequently used in psycholinguistic research in order to better understand how people process spoken sentences and to investigate linguistic aspects during sentence processing. A temporal relationship between speech processing and eye movements was shown in the pioneering study by Cooper (Cooper, 1974), and confirmed in more recent studies (see Huettig *et al.* (2011) for a review). The visual world paradigm (Tanenhaus *et al.*, 1995, Eberhard and Spivey-Knowlton, 1995, Allopenna *et al.*, 1998) was introduced in psycholinguistics to reveal the interaction between language and vision. In that paradigm, eye movements were recorded while simultaneously presenting spoken language and a visual scene that typically included the objects mentioned in the presented speech. Participants spontaneously fixated on the object that corresponded to the acoustical input. Several subsequent studies have investigated how and when the linguistic and visual information are integrated (Altmann and Kamide (1999), Snedeker and Trueswell (2004), Altmann and Kamide (2007), Kamide *et al.* (2003), Knoeferle *et al.* (2005), Knoeferle and Crocker (2006), Knoeferle (2007), Knoeferle and Crocker (2007)). These recorded data were often used to investigate how linguistic processes determine the participants' sentence processing and understanding. The method of analyzing the recorded eye-tracking data in the visual world paradigm, however, depends on the research question (Huettig *et al.*, 2011) and has not been

adapted for use in audiology or made available to answer the research questions of the current study. For these reasons, an approach was adapted which combined several techniques from other (visual world) studies. The new approach was designed to meet the following requirements: a) the eye-tracking data must have a high temporal resolution; b) the test design must be symmetric, averaging out any systematic eye movement strategies, such as a preference for analyzing the pictures from left to right; c) the eye-tracking data analysis should shed light on speech comprehension and the decision process. Since the combination of these processing techniques is novel, the motivation behind each step is outlined in the following.

To investigate the effect of linguistic aspects on the comprehension process, the speech stimuli (words or sentences) were subdivided into separate time windows, as in previous studies (e.g. Altmann and Kamide (1999), Knoeferle (2007)). Due to the nature of speech, these segments varied slightly in duration. For this reason, a time alignment was applied. This allowed temporal averaging across segments and a high temporal resolution on a sub-segment basis.

As in previous visual world studies, the visual stimulus was subdivided into regions of interest (ROIs): one for the target picture and one for the competitor picture. Previous studies have analyzed whether these ROIs differ in their likelihoods of being fixated during each time segment (Chambers *et al.*, 2002, Huettig and McQueen, 2007), or whether a ROI is looked at earlier in an experimental condition than in a control condition (Altmann and Kamide, 1999, Snedeker and Trueswell, 2004). Accordingly, the current study analyzed fixation rate as a function of time for different ROIs. Previous studies found that one region of interest was more likely to be fixated even before stimulus presentation, and emphasized that these baseline effects should be taken into account when analyzing the eye-tracking data (Barr *et al.*, 2011). However, methods that account for baseline effects have not often been applied in visual world studies. Therefore, the current study proposes a method that calculates the rates of fixations towards a target picture (in the current study a picture that matches the spoken sentence) in relation to the rate of fixation towards a competitor picture. As this is done both for target pictures on the left and on the right side, any systematic eye movement strategy that the participant uses, such as gazing preferably from left to right, is averaged out from the data. This is referred to as *symmetrizing* in the following. The applicability of assessing differences between fixations towards a target and a competitor was previously shown by other studies (Arnold *et al.*, 2003, Kaiser and Trueswell, 2008). A post-processing step is proposed that includes a bootstrap method to calculate the 95 % confidence interval of the estimated probability that the participant fixates the target picture. Bootstrapping is an appropriate method for analyzing measurement statistics in situations where observed values violate normality or are unknown (Efron and Tibshirani, 1993, van Zandt, 2002). In order to obtain a defined measure of processing speed and to detect the point in time when the target is recognized by the participant, a fixed threshold criterion is used, as described by McMurray and colleagues (McMurray *et al.*, 2008, Toscano and McMurray, 2012).

The underlying hypothesis of this study is that the proposed eye-tracking paradigm can detect significant and robust reductions in sentence processing speed for sentence structures with increased linguistic complexity. This would qualify the proposed method for use in audiology. An increase

in processing time, indicated by eye fixations as well as by reaction times, is then interpreted as evidence for a greater cognitive processing effort during sentence comprehension. This study had four main goals:

- Introduction of an eye-tracking paradigm that is adapted to the OLACS speech intelligibility test and enables online analysis of the time course of the sentence comprehension process for use in audiology.
- Introduction of a time-resolved statistical analysis technique for eye-tracking data that derives the decision moment (DM), defined as the point in time when the target is recognized by the participant. The analysis should take into account any systematic eye movement strategy employed by the participants.
- Evaluation of this paradigm and provision of normative data testing listeners with normal hearing in quiet.
- Identification of those sentence structures that show the most significant effects of linguistic complexity. As a prerequisite for a time-efficient clinical application, a reduced subset of test sentences will be needed for testing speech processing in listeners with hearing impairment in quiet and in noise.

2.2 Material and methods

2.2.1 Participants

Seventeen volunteer participants (ten male and seven female) with normal hearing took part in the experiment. Hearing thresholds were measured at octave frequencies from 125 Hz to 8000 Hz. All participants had otologically normal hearing, defined here as having a pure tone audiometry hearing threshold of 15 dB hearing level (HL) or better at the measured frequencies. All participants were native German speakers between 18 and 30 years of age (average age: 26 years) and either had uncorrected vision or wore corrective eyewear (glasses or contact lenses) when necessary.

2.2.2 Stimuli

Speech material

A total of 148 sentences from the OLACS corpus were used (Uslar *et al.*, 2013a). Each sentence corresponded to one of seven different syntactic structures; there were approximately 21 sentences

of each structure. The seven syntactic sentence structures fall into two major groups: verb-second structures and relative-clause structures (Table 2.1). Both groups contain sentences with canonical (subject-before-object) and non-canonical (object-before-subject) word orders.

The group of verb-second structures includes three sentence structures: subject-verb-object (SVO), object-verb-subject (OVS), and ambiguous object-verb-subject (ambOVS). The SVO structure has the canonical word order for simple main clauses in German and is considered syntactically simple and easy to process (Bader and Bayer, 2006). The OVS structure is more complex because of its non-canonical word order. The SVO and OVS structures are unambiguous with respect to their meaning and to the grammatical role of the sentence components (see Table 2.1). For example, the grammatical function of the first noun phrase is clearly marked for both the SVO structure (*Der kleine Junge*$_{PTD}$, 'The little$_{nom}$ boy$_{nom}$' *nom* indicates the nominative case marking) and the OVS structure (*Den lieben Vater*$_{PTD}$, 'The nice$_{acc}$ father' *acc* indicates the accusative case marking). In both of these sentence structures, the disambiguating word, which is the word that clarifies the agent/object role assignment, is the first noun. For instance, the noun, *Junge*$_{PTD}$ 'boy' in the SVO sentence disambiguates the sentence in such a way that participants are theoretically able to relate the spoken sentence to the target picture as soon as the noun is spoken. In all cases, the onset of the word that disambiguates subject and object is termed the point of target disambiguation (PTD). Thus, the PTD was defined as the onset of the word that first enabled correct recognition of the target picture. Note that we chose the onset of the word even though in some sentence structures the recognition of the target was only made possible by the suffix of the disambiguating word. This was necessary because it was not possible to determine the exact point in time at which the disambiguation occurs during the spoken word.

The third verb-second structure, ambOVS, has an object-before-subject structure with a later point of disambiguation. In these sentences, the first article is ambiguously marked for case: the first article, *Die* ('The$_{amb}$' *amb* indicates the ambiguous case marking; see Table 2.1) could indicate either subject or object function (and subsequently agent or object role) and only the article of the second noun, *der*$_{PTD}$ ('the$_{nom}$' *nom* indicates the nominative case marking; see Table 2.1) is unambiguously case-marked.

The second group of sentence structures includes four different structures of embedded relative clauses (Table 2.1): subject-relative (SR) clauses and object-relative (OR) clauses, with a PTD at the first relative pronoun *der*$_{PTD}$ ('who$_{nom,\ sg}$') or *den*$_{PTD}$ ('who$_{acc,\ sg}$' *sg* indicates singular form; see Table 2.1); and ambiguous subject-relative clauses (ambSR) and ambiguous object-relative clauses (ambOR) with a late PTD. The ambSR and ambOR sentence structures are disambiguated by the verb, *fangen*$_{PTD}$ ('catch$_{3pl}$' *3pl* indicates the third person plural form) or *fängt*$_{PTD}$ ('catches$_{3sg}$'), of the embedded clause (Table 2.1).

The speech material provides different levels of linguistic complexity by varying three different structural factors of the sentence material: word order, embedding, and ambiguity. The preferred,

Table 2.1: Example sentences for the seven sentence structures of OLACS. The disambiguating word from which the target picture could theoretically first be recognized is indicated with PTD (point of target disambiguation). *Nom* (nominative), *acc* (accusative), and *amb* (ambiguous case) indicate the relevant case markings. *Sg* indicates singular forms and *pl* indicates plural forms. Verbs are either in their third person singular (*3sg*) or third person plural (*3pl*) form. *fem* indicates feminine nouns. SVO, OVS, and ambOVS structure belong to the verb-second structures since they have either a subject-verb-object or an object-verb-subject sentence structure. SR, OR, ambSR, and ambOR structures belong to the relative-clause structures. The meaning of the example sentence is given by the sentence in quotation marks.

SVO	Der	kleine	$\text{Junge}_{\text{PTD}}$	grüsst	den	lieben	Vater.
	The_{nom}	little_{nom}	boy	greets_{3sg}	the_{acc}	nice	father.
	"The little boy greets the nice father."						
OVS	Den	lieben	$\text{Vater}_{\text{PTD}}$	grüsst	der	kleine	Junge.
	The_{acc}	nice_{acc}	father	greets_{3sg}	the_{nom}	little_{nom}	boy.
	"It is the nice father that the little boy is greeting."						
ambOVS	Die	liebe	Königin	grüsst	der_{PTD}	kleine	Junge.
	The_{amb}	nice_{amb}	queen_{fem}	greets_{3sg}	the_{nom}	little_{nom}	boy.
	"It is the nice queen that the little boy is greeting."						
SR	Der	Bauer,	der_{PTD}	die	Ärztinnen	fängt,	lacht.
	The_{sg}	farmer_{sg}	$\text{who}_{nom,sg}$	the	$\text{doctors}_{amb,pl,fem}$	catches_{3sg}	laughs_{3sg}.
	"The farmer who is catching the doctors is laughing."						
OR	Der	Bauer,	den_{PTD}	die	Ärztinnen	fangen,	lacht.
	The_{sg}	farmer_{sg}	$\text{who}_{acc,sg}$	the	$\text{doctors}_{amb,pl,fem}$	catch_{3pl}	laughs_{3sg}.
	"The farmer whom the doctors are catching is laughing."						
ambSR	Die	Bäuerinnen,	die	die	Ärztin	$\text{fangen}_{\text{PTD}}$,	lachen.
	The_{amb}	$\text{farmers}_{amb,pl,fem}$	who_{amb}	the_{amb}	$\text{doctor}_{sg,fem}$	catch_{3pl}	laugh_{3pl}.
	"The farmers who are catching the doctor are laughing."						
OR	Die	Bäuerinnen,	die	die	Ärztin	$\text{fängt}_{\text{PTD}}$,	lachen.
	The_{amb}	$\text{farmers}_{amb,pl,fem}$	who_{amb}	the_{amb}	$\text{doctor}_{amb,sg,fem}$	catches_{3sg}	laugh_{3pl}.
	"The farmers whom the doctor is catching are laughing."						

canonical word order in German, like many other languages, is subject-before-object (Bader and Meng, 1999, Gorrell, 2000). The non-canonical object-before-subject word order is considered syntactically more complex (Fanselow *et al.*, 2008) and has been shown to increase processing costs in the form of reduced accuracy and longer reaction times (Tun *et al.*, 2010a, Wingfield *et al.*, 2006, Gibson, 2000). Another factor leading to increased processing costs is embedded relative-clauses (Gordon *et al.*, 2002, Carroll and Ruigendijk, 2013). Within the relative-clause structures, processing costs can be further increased by word order (Bader and Meng, 1999, Carroll and Ruigendijk, 2013) (SR and OR structures in Table 2.1). The OLACS corpus further includes temporally ambiguous sentence structures, in which disambiguation occurs later. The ambiguity of these sentence structures (ambOVS, ambSR, ambOR) can lead to temporary uncertainty with regard to the grammatical role of the sentence components (Carroll and Ruigendijk, 2013, Altmann, 1998). Because of this ambiguity, the participant has to reanalyze the initial subject after the point of disambiguation. Hence, the ambiguity can lead to both increased processing cost and temporary misinterpretation of the sentence.

Visual stimuli

In total, picture sets for 150 sentences of the OLACS corpus were created. Each picture set consisted of two pictures (Figure 2.1). One of the two pictures, the target picture, illustrated the situation described by the sentence. In the competitor picture, the roles of agent (the active character) and object (the passive character) were interchanged. In each picture, the agent was always shown on the left side in order to facilitate fast comprehension of the depicted scene. Presenting both pictures at the same time ensured that participants did not assign agent and object roles using only visual information. All of the figures illustrated in the picture sets had the same size in order to avoid effects of contrast between the figures. Care was taken in selecting actions, agents, and objects that were non-stereotypical, such that the action was not characteristic for the agent (for example, baking is a typical action of a baker). This constraint was employed to make sure that participants did not make premature role assignments based on any anticipation of an agent's characteristic action. The picture set was divided into three regions of interest (ROI): ROI1 defined the target picture, ROI2 the competitor picture, and ROI3 defined the background. The target picture was shown randomly either on the left or right side of the computer screen. Consequently, the positions of ROI1 and ROI2 were not fixed, but changed randomly from trial to trial.

Validation of the visual stimuli

To ensure that both pictures in a particular picture set could be parsed and interpreted equally well, a subset of the graphical material was tested by measuring the reaction times of 20 participants. For 106 picture sets, the reaction time for each picture was measured (212 single pictures). For

that purpose, each sentence was presented visually in written form on a computer screen for 1500 ms. Afterward one picture, either the target or the competitor picture, was shown on the computer screen, and the participants had to decide whether the presented picture matched the previously displayed sentence. Participants were instructed to respond as quickly as possible and reaction times were measured. Note that the sentences were simplified for the validation of the visual stimuli: the modified sentences all had a subject-verb-object structure, and the adjectives of the verb-second structures and the matrix verbs of the relative-clause structures were omitted in this pre-test. For instance, Figure 2.1 shows the picture corresponding to the example sentence, "The dog reprimands the duck." By modifying the sentences to have the same syntactical structure, any effects of linguistic complexity on reaction times were avoided. The statistical significance of the differences in reaction times for the two pictures of one set was calculated for all participants using a paired t-test with a 5 % significance level. If a significant difference was found, the picture set was excluded from the eye-tracking study. Of the 106 picture sets tested, two sets were excluded. Because so few picture sets had to be excluded, no formal reaction time validation was performed for the additional 44 picture sets that were produced later and added to the experimental set. Thus, in total, 148 different picture sets were used for the eye-tracking experiment.

2.2.3 Procedure

For the experiments, an OLACS picture set was presented visually on a computer screen while the recorded sentence was presented via headphones. First, the participants performed one training block, which contained all 148 picture sets. After training, six test blocks, containing 110 sentences each, were performed. In total, each participant listened to 660 sentences. 148 sentences were

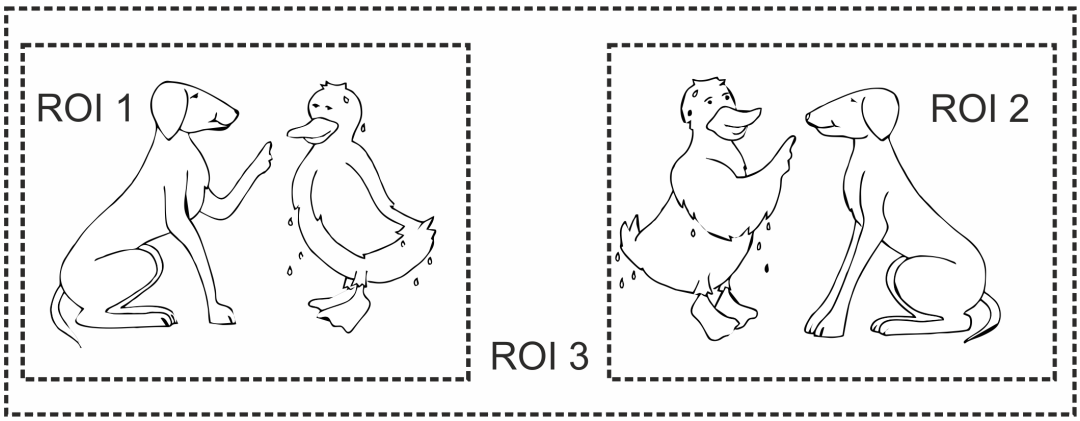

Figure 2.1: Example picture set for a sentence of the ambOVS sentence structure: *Die nasse Ente tadelt der treue Hund*. (The wet duck (acc.) reprimands the loyal dog (nom.), which means, "It is the wet duck that is reprimanded by the loyal dog"). A picture set consists of two single pictures. The dashed lines indicate the three regions of interest (ROI) and are not visible for the participants. ROI1 is the target picture and can be located on the left or right side of the picture set. ROI2 is the competitor picture. ROI3 is the background.

presented in quiet at a level of 65 dB SPL. Two conditions with different background noises were employed for a different study: 444 sentences were presented in different noise conditions. These 592 sentences were randomly distributed across the six test blocks. In order to avoid retrieval strategies, 68 filler trials were presented across all test blocks (11-12 filler trials per test block). During a filler trial, either the target or the competitor picture was depicted on both sides of the screen, with the positions of the agent and object reversed in one of the two pictures. Therefore, either both of the pictures matched the spoken sentence or neither did. These trials forced the participants to fixate on both pictures.
The visual stimulus was presented 1000 ms before the onset of the acoustic stimulus. Participants were instructed to identify the picture that matched the acoustic stimulus by pressing one of three keys as quickly as possible: The "A" indicated that participants assigned the target to the left picture, and "L" indicated assignment to the right picture; participants were instructed to press the space bar if they were not able to clearly assign one target picture to the spoken sentence. The position of the selected keys enabled the participants to leave their hands on the keyboard during the experiment so they did not have to look at the keyboard to search for the right key. After each trial, participants were asked to look at a marker at the center of the screen so that a drift correction could be performed. At the beginning of each test block a calibration was done using a nine-point fixation stimulus. The completion of one test block of trials took about 20 min. After each block, participants had a ten-minute break. The entire measurement took about three hours per participant, which was divided into two sessions.

2.2.4 Apparatus

An eye-tracker system (EyeLink 1000 desktop system including the EyeLink CL high-speed camera, SR Research Ltd.) was used with a sampling rate of 1000 Hz to monitor participants' eye movements. The pictures were presented on a 22 inches multi-scan color computer screen with a resolution of 1680 x 1050 pixels. Participants were seated 60 cm from the computer screen. A chin rest was used to stabilize the participant's head. Although, viewing was binocular, the eye-tracker sampled only from the dominant eye. Auditory signals were presented via closed headphones (Sennheiser HDA 200) that were free-field compensated according to DIN EN389-8 (2004). For the calibration of the speech signals a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2 inch microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier were used. All experiments took place in a sound-insulated booth.

2.2.5 Data analysis

Time alignment

Since the sentences differed in length, a time alignment was employed to allow comparisons across sentences. This was realized by dividing each trial into six segments, as shown in Table 2.2. Note that the choice of segment borders and the evaluation of eye-tracking data during these segments were selected to best fit the employed OLACS speech material. Knoeferle and colleagues (Knoeferle, 2007) showed that for German sentences with an initially ambiguous structure, sentence interpretation happens immediately after the point in time at which the combination of visual and linguistic information disambiguates the sentence. Therefore, segment borders were defined according to the word that first enabled correct recognition of the target picture. Segment 1 corresponds to the time from the onset of the visual stimulus until the onset of the acoustical stimulus. The spoken sentence was presented during segments 2 through 5. The time from the end of the spoken sentence until the participant responded by pressing the response key was denoted as segment 6. The segment borders and the corresponding points in time (in ms) during the eye-tracking recordings were determined for each sentence and averaged over all sentences of a single sentence structure (see Table 2.2).

Calculation of the target detection amplitude (TDA)

The eye-tracking data were used to calculate the target detection amplitude (TDA). The TDA quantifies the tendency of the participant to fixate on the target picture in the presence of the competitor picture. The data analysis for the TDA was divided into three stages (Figures 2.2 and 2.3).

In the first stage, the calculation was sentence based (left panel in Figure 2.2). The recorded eye-tracking data were analyzed and the fixations on the target (ROI1), the competitor (ROI2), and the background (ROI3) were calculated as functions of time. Trials in which the target was presented on the left side were considered separately from those in which the target was on the right. A time alignment and a resampling stage were employed to associate the observed fixations of the ROIs with the appropriate sentence segment (see Table 2.2). To synchronize the segment borders across sentences, the first five segments were individually rescaled to a fixed length of 100 samples using an interpolation algorithm. The length of segment 6 depended on the mean reaction time of the participant, with a maximal length of 200 samples (see Table 2.2). For instance, if the reaction time was 1500 ms, the last segment was rescaled to a length of 150 samples. For reaction times longer than 2000 ms, the signal was cut to a length of 200 samples. This was done because 1000 ms after the offset of the sentence, on average, participants fixated less frequently on the target picture (as can be seen in segment 6 in Figure 4 and Figure 5). This may have been because no more information could be gained after this time. The segment-based

Table 2.2: Time segments for the verb-second and relative-clause structures used for time alignment across sentences. The first row gives the borders of each segment in time samples. Segment 1 describes the time from the onset of the measurement until the onset of the acoustical stimulus. The spoken sentence was presented during segments 2 through 5. Segment 6 corresponds to the time between the end of the spoken sentence and the participant's response. An example sentence is given for each group. The mean segment borders (in milliseconds) were calculated over all sentences in the group after the resampling procedure (± standard deviation).

	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6
Segment borders/samples	0-100	100-200	200-300	300-400	400-500	500-end
Verb-second structure	no acoustic stimulus	Der kleine	Junge	grüsst den	lieben Vater.	response
		The little	boy	greets the	nice father.	
Mean segment border/ ms	0-1000	1000-1745 (±130)	1745-2340 (±135)	2340-2995 (±130)	2995-4140 (±151)	4140-end (±114)
Relative-clause structure	no acoustic stimulus	Der Bauer,	der die Ärztinnen	fängt,	lacht.	response
		The farmer	who the doctors	catches	smiles.	
Mean segment borders/ ms	0-1000	1000-1885 (±200)	1885-2755 (±136)	2755-3430 (±131)	3430-4450 (±143)	4450-end (±238)

resampling used a fixed number of samples per segment (except for the last segment), which resulted in a segment-dependent sampling rate depending on the individual length of each segment. This resampling not only allowed comparison across sentences of one structure, but also across different sentence structures.

The second stage of the TDA calculation was sentence-structure-based (Figure 2.2). For a given (interpolated) time sample, the fixated ROIs were averaged across all sentences of a given sentence structure, resulting in an average fixation rate (right panel in Figure 2.2). Note that the fixation rates of the background (ROI3) were not considered in the calculation of the TDA[1]. Thus, the fixation rates of target (ROI1) and competitor (ROI2) did not add up to 100 %. Only trials in which the participants selected the correct picture were used for further analysis. This selection was done in order to analyze time patterns of eye fixations that reflected the dynamics of the

1 Further analysis of the data showed that the fixation rates for ROI3 did not differ significantly between sentence structures. Since this study examines the differences in the time courses of the TDAs for different sentence structures, the fixation rates of the background were not considered for further analysis.

recognition process for correctly identified sentences only.

Symmetrizing

In general, participants tended to fixate more frequently on the left picture. This effect was independent of the position of the target picture and was most noticeable in segment 1, before the acoustical stimulus was presented. This tendency towards the left picture probably arose from the usual reading direction. This behavior was exploited in the paradigm by always presenting the agent of each scene on the left side of each picture (except in filler trials). This agent-left convention supported the participant in systematic and fast analysis of each picture as uncertainties about the agent's and the object's roles within each picture were reduced. The agent-left convention may have supported the listeners' left-to-right strategy. To correct for this, the test design was symmetrized: in random order, the target picture was presented equally often on the left and right sides. Subsequently, the fixation rate was averaged across all trials, averaging out any left-to-right picture reading strategy. One half was subtracted from the resulting averaged target fixation rate (which ranges between 0 and 1) in order to center it around 0. The result was then multiplied by 2. This resulted in the TDA, which assumed the value -1 for sole fixations of the competitor, 0 for random fixation, and 1 for sole fixations of the target. The calculation of the TDA was split into different processing steps, which allowed analysis of the fixation rates for left and right targets separately. Four different fixation rates FR(s|S, t) were considered, with s denoting the position of the fixated picture (with l for left side and r for right side), S denoting the position of the target picture (with L for left side and R for right side), and t denoting the time. Depending on the position of the target, the two fixation rates of the competitor pictures FR(r|L, t) and

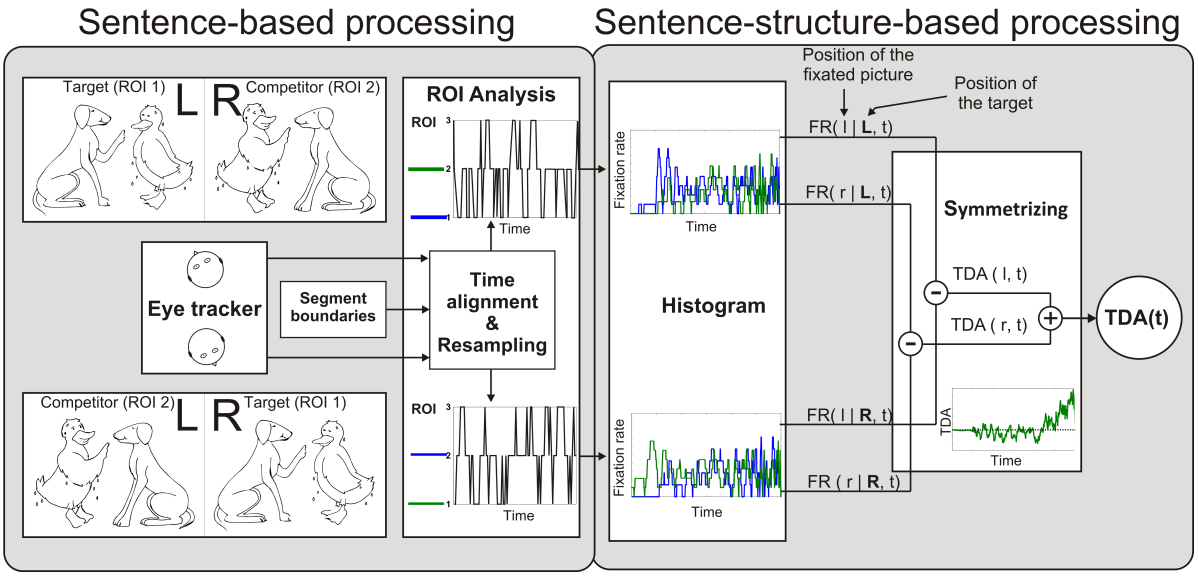

Figure 2.2: Schematic diagram for the first two stages of the calculation of the target detection amplitude (TDA), namely the sentence-based processing and the sentence-structure-based processing stage.

FR(l|R, t) were subtracted from the respective fixation rates of the target pictures FR(l|L, t) and FR(r|R, t). This gave the TDA for the left picture:

$$TDA(l,t) = FR(l|L,t) - FR(l|R,t) \tag{2.1}$$

and for the right picture:

$$TDA(r,t) = FR(r|R,t) - FR(r|L,t). \tag{2.2}$$

The position-independent total TDA was expressed using the sum of the two side-dependent TDA(s,t):

$$TDA(t) = TDA(l,t) + TDA(r,t). \tag{2.3}$$

The total $TDA(t)$ was a function of time and quantified the tendency to fixate on the target picture within the arrangement of alternative pictures. Positive values indicated more fixations on the target picture and negative values indicated more fixations on the competitor picture. A value near zero reflected the inability to differentiate between the two pictures at a given point in time. The $TDA(t)$ was computed for all 17 participants, resulting in a set M_{TDA} of 17 values for each sentence structure at a given point in time t:

$$M_{TDA} = TDA_1(t), ..., TDA_{17}(t). \tag{2.4}$$

Post processing

To compute the time-smoothed mean value and estimate the 95 % confidence interval of the TDA, this set was input to a post-processing stage, as depicted in Figure 2.3. A bootstrapping resampling procedure was applied (Efron and Tibshirani, 1993, van Zandt, 2002) to estimate the mean value and 95 % confidence interval of the average TDA across participants for the different OLACS sentence structures without assuming any underlying distribution. This type of bootstrapping procedure has been successfully applied before to analyze eye-tracking data (Ben-David *et al.*, 2011). This bootstrapping was necessary because the underlying distribution of the mean value across the set M_{TDA} at a given point in time was unknown and could vary across different sentence structures. For each time point, a sample from M_{TDA} was randomly selected with replacement 17 times and averaged to provide a random estimate of the mean value $\langle TDA(t) \rangle$ across participants. This process was repeated 10,000 times, resulting in a resampled data set containing 10,000 values that approximated the estimated distribution of $\langle TDA(t) \rangle$. From this distribution, the 95 % confidence intervals and the mean value $\langle TDA(t) \rangle$ were obtained. Finally, a Gaussian smoothing filter with a kernel size of 25 samples was applied in order to reduce the random fluctuations of the $\langle TDA(t) \rangle$. The resulting signal was called TDA (see Figure 2.4).

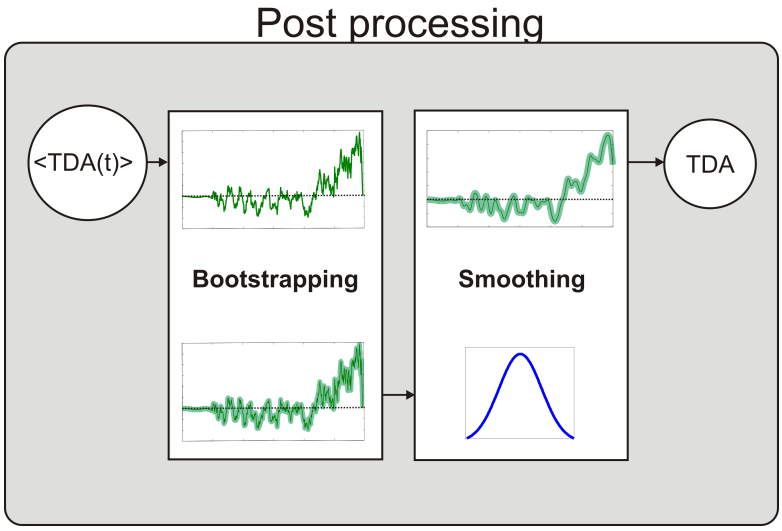

Figure 2.3: Post processing of the TDA, including bootstrapping resampling procedure and Gaussian smoothing

Calculation of the decision moment (DM) and the disambiguation to decision delay (DDD)

The decision moment (DM) was defined as the point in time from which the mean TDA exceeded the 15 % threshold for at least 200 ms. The threshold was chosen as 15 % TDA because small fluctuations in the TDA are not relevant for the investigation of speech processing. The time between the PTD and the DM was calculated for each sentence structure and defined as disambiguation to decision delay (DDD). This DDD is interpreted as a measure of processing time: The greater the DDD, the longer the processing time and the slower the speed of sentence processing.

2.3 Results and discussion

2.3.1 Picture recognition rates

The picture recognition rates-the percentage of correctly identified target pictures (by pushing the correct button)-for each sentence structure (see Table 2.3) were averaged across all participants. Before conducting further analysis, picture recognition rates were transformed to rationalized arcsine units (rau) according to Sherbecoe and Studebaker (2004).

To investigate the effect of sentence structure on picture recognition, a one-way repeated measures ANOVA was conducted for both groups of sentence structures. The factor sentence structure was significant for both groups of sentence structures (verb-second: $F(2; 32) = 36.2$, $p < 0.001$; relative-clause: $F(3; 48) = 7.4$, $p < 0.001$). Multiple pairwise comparisons with Bonferroni correction revealed differences in picture recognition rates between the SVO and

ambOVS structures ($p < 0.001$), reflecting lower picture recognition rates for the ambOVS structure. The picture recognition rate for ambOVS sentences was lower than that for OVS sentences ($p < 0.001$). For the relative-clause structures, the pairwise comparisons revealed significant differences between SR and OR structures ($p = 0.001$) and between OR and ambSR structures ($p = 0.002$). In general, significantly lower picture recognition rates, in particular for the object-first sentence structures (ambOVS and OR structures) suggest that linguistic complexity affects picture recognition performance. This is not self-evident: all of the sentences were presented in quiet at a constant sound pressure level of 65 dB and were acoustically controlled for equal intelligibility (for detailed information, see Uslar *et al.* (2013a)), so they should all have been equally understandable. For that reason, the differences in picture recognition rates found here are evidence that linguistic factors influence the processing of syntactically complex structures in combination with the visual stimuli.

2.3.2 Reaction time

The reaction times were measured offline: participants were asked to press the response button after the end of the sentence. To investigate the effect of sentence structure on reaction time, a one-way repeated-measures ANOVA was conducted for both groups of sentence structures. The factor sentence structure was not significant for either group, indicating that sentence complexity did not affect reaction time within this paradigm. Note that the offline measures, recognition rate and reaction time, did not follow the same pattern across sentence structures, suggesting different response strategies and criteria. However, this effect was not considered further because the online measures used in this paper took place markedly before the (offline) button press. In addition, only correct trials were considered for the online analysis.

2.3.3 Eye fixation data

The target detection amplitude (TDA) functions for the verb-second and relative-clause structures are depicted in Figures 2.4 and 2.5, respectively. The dashed vertical lines reflect the averaged segment borders. The time points corresponding to these segment borders are shown for both groups of sentence structures in Table 2.2. The dashed horizontal lines in Figures 2.4 and 2.5 indicate the thresholds of $\pm 15\,\%$ TDA. The decision moment (DM) is the point in time at which the TDA exceeded the threshold for at least 200 ms; it is indicated with a plus sign for each sentence structure. The DM was interpreted as the moment at which participants recognized the target, since they fixated the target picture significantly more frequently than the competitor. The circles indicate the PTD corresponding with the words denoted in Table 2.1. The horizontal lines starting at the PTDs depict the disambiguation to decision delay (DDD).

Table 2.3: Picture recognition rates and reaction times obtained from the keyboard responses, and the calculated decision moments (DM) for each sentence structure. The mean picture recognition rates in rationalized arcsine units (rau), DMs (ms), and reaction time (ms) were calculated over all participants for both verb-second and relative-clause structures of the OLACS corpus. The calculated DMs are listed for each sentence structure with the corresponding width Δt (in milliseconds) at the 15 % threshold along the timeline.

Sentence structure		Picture recognition rate / rau	DM / ms	Reaction time / ms
Verb-second structures	SVO	97.6 ± 5.0	2045 (Δt =645)	2057 ± 477
	OVS	105.8 ± 8.1	2715 (Δt =1380)	1956 ± 421
	ambOVS	81.0 ± 4.3	3315 (Δt =275)	1944 ± 300
Relative-clause structures	SR	101.4 ± 8.8	2615 (Δt =1515)	2029 ± 411
	OR	91.6 ± 9.9	2625 (Δt =335)	1965 ± 477
	ambSR	100.9 ± 8.7	3600 (Δt =895)	2084 ± 643
	ambOR	96.2 ± 4.4	3510 (Δt =340)	1898 ± 367

Verb-second structures

Figure 2.4 shows the TDAs for the three sentence structures with verb-second structures. The TDAs fluctuated between the thresholds (±15 % TDA) around zero during the first two segments for all three sentence structures: neither target nor competitor picture was fixated preferably. Since the PTDs for the two unambiguous sentence structures (SVO, OVS) did not occur until the beginning of segment 3, the DM was not expected before the beginning of segment 3. The fact that the TDA fluctuated around zero during the first segments indicated the success of the symmetrizing method in averaging out any systematic strategy of the participants. If the tendency of fixating the left picture first would not have been compensated for, the TDA would have differed significantly from zero.

The early case marking of the first noun phrase *Der kleine Junge*$_{PTD}$ ('The$_{nom}$ little$_{nom}$ boy'; Table 2.1) in the SVO structure allowed an early thematic role assignment, so participants were able to identify the noun phrase referent (*Junge*) as the agent and to recognize the target even before the end of the spoken noun. This was indicated by an early DM during segment 3, with a DDD of 300 ms, for the SVO structure. Considering that the oculomotor delay is approximately 200 ms, the DM for the SVO structure occurred quite quickly after the PTD. The first noun phrase, *Den lieben Vater*$_{PTD}$ ('The$_{acc}$ nice$_{acc}$ father'; Table 2.1), of the unambiguous OVS structure also provided role information at the very beginning of the spoken sentence. But despite the early PTD, the DM of the OVS structure was observed during segment 4, one segment after the first noun (*Vater*) was spoken. Thus, the DDD for the OVS structure was about 970 ms. So although the ±95 % confidence intervals of the SVO and OVS structures overlapped slightly at the DMs, their DDDs differed by more than 600 ms.

Object-first sentences with a late PTD, as in the ambOVS structure, had a markedly different TDA time course. The DM of the ambOVS structure occurred during segment 5, after the onset of the second article, *der*$_{amb}$ ('the$_{nom}$'; Table 2.1), which disambiguated the sentence in

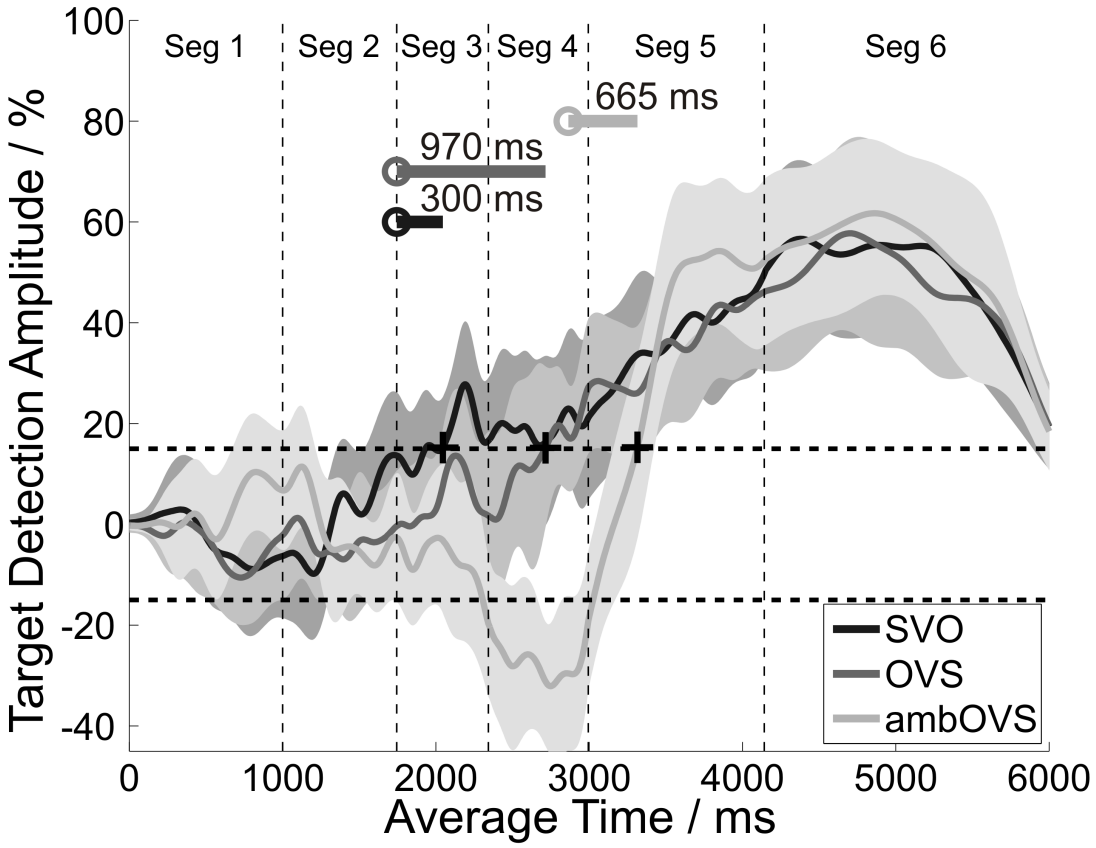

Figure 2.4: Mean target detection amplitude (TDA) averaged over all subjects for the verb-second structures, i.e., the subject-verb-object (SVO), object-verb-subject (OVS), and the ambiguous object-verb-subject (ambOVS) structures. The shaded areas illustrate the 95 % confidence intervals for each individual curve. The + signs at 2045 ms, 2715 ms, and 3315 ms denote the DMs where the TDA first exceeded the threshold (15 % of the TDA). The circles denote the point of target disambiguation (PTD): at 1745 ms for the SVO and OVS sentences and at 2650 ms for the ambOVS sentences. The horizontal lines denote the disambiguation to decision delay (DDD), i.e. the distance between the PTD and the DM.

segment 4. This resulted in a DDD of about 665 ms. Note that the DDD for the ambOVS structure was about 300 ms shorter than that of the unambiguous sentence structure, OVS. In addition, a strongly negative TDA was observed for the ambOVS structure at the end of segment 3, indicating that participants were preferentially fixating on the competitor picture. The negative TDAs were interpreted as a temporary misinterpretation arising out of listeners' preferences for subject-before-object word order. German shows a general preference of subject-before-object word order (Bader and Meng, 1999, Gorrell, 2000). So listeners expected a subject-before-object sentence structure and tended to interpret the first noun phase, *Die liebe Königin* ('The$_{amb}$ nice$_{amb}$ queen$_{fem}$' see Table 2.1), as the subject of the sentence. As a result, the competitor was fixated more frequently at the beginning of the sentence. This temporary misinterpretation only occurred before the sentence had been disambiguated by the article of the second noun phrase, der$_{PTD}$ ('the$_{nom}$').

Relative-clause structures

The left panel of Figure 2.5 shows the average TDAs of the unambiguous relative-clause structures (SR and OR structures). For both structures, the TDAs fluctuated around zero during the first two segments, indicating that the target was not recognized. For both sentence structures, the case-marking relative pronoun, der_{PTD} ('who$_{nom}$') or den_{PTD} ('who$_{acc}$'; Table 2.1), of the embedded phrase disambiguated the sentence; this is indicated by the PTD at the very beginning of segment 3. The DMs of both sentence structures occurred at the end of segment 3 and the DDDs varied between 730 ms and 740 ms. The right panel of Figure 2.5 shows the TDAs of the two ambiguous relative-clause structures (ambSR and ambOR). It is clear that the embedded verbs, *fangen*$_{PTD}$ ('catch$_{3pl}$') and *fängt*$_{PTD}$ ('catches$_{3sg}$'; Table 2.1) resolved the roles of agent and object: the PTD was located at the beginning of segment 4. The DMs were observed in segment 5, with a DDD of 755 ms for the ambOR structure and 845 ms for the ambSR structure. Note that for the unambiguous structures, the first article of the embedded sentence, *der*$_{PTD}$ ('who$_{nom}$') or *den*$_{PTD}$ ('who$_{acc}$'; Table 2.1), which had an average length of about 135 ms, disambiguated the spoken sentence. In contrast, the disambiguating word for the ambiguous sentence structure was the embedded verb (*fangen*$_{PTD}$ 'catch$_{3pl}$' and *fängt*$_{PTD}$ 'catches$_{3sg}$' see Table 2.1), with an average length of about 575 ms. For most of these embedded verbs the disambiguating information about the agent/object role assignment was not given until the suffix. Since the PTD was defined as the onset of the disambiguating word, the different word lengths (135 ms vs. 575 ms) had to be accounted for when comparing the DDDs of the different relative-clause structures. After subtracting the length of the disambiguating word, the remaining DDD was much smaller for the ambiguous structures than for the unambiguous structures.

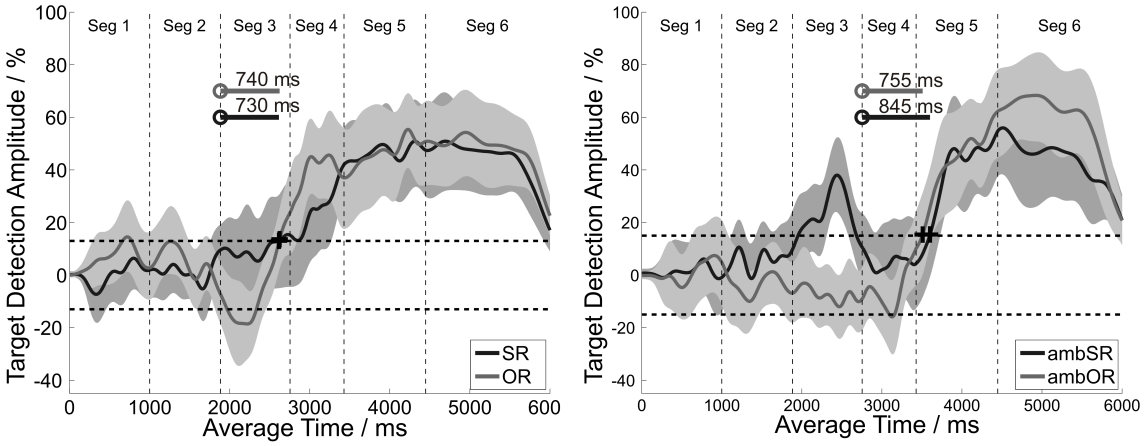

Figure 2.5: Mean TDA averaged over all participants for the relative-clause structures of the OLACS. The shaded areas illustrate the 95 % confidence intervals for each curve. *Left panel*: unambiguous subject-relative clause (SR structure) vs. unambiguous object-relative clause (OR structure); DMs (+) at 2615 ms and 2625 ms, respectively. *Right panel*: ambiguous subject-relative clause (ambSR structure) vs. ambiguous object-relative clause (ambOR structure); DMs (+) at 3600 ms and 3510 ms, respectively. The horizontal lines denote the DDD.

Participants were not expected to discriminate between the two pictures before the PTD, so the TDAs of the two sentence structures should not differ markedly before the PTD. Surprisingly, a significant positive TDA was observed for the ambSR structure shortly after the relative pronoun *die* ('the$_{amb}$' see Table 2.1) in segment 3. If this unexpected early increase in the TDA had been caused by the participants' subject-first preference, then it should have also been reflected in the time course of the ambOR structure. For instance, if the plural form of the noun used for the ambiguous subject-relative and object-relative clauses had helped the participants to recognize the target earlier, this should have been indicated in the TDA of both sentence structures. It would have appeared as an early increase in the TDA for the ambSR structure and a decrease in the TDA for the ambOR structure. However, this was not the case: no significant decrease in the TDA was observed in segment 3 or at any later point in time. There is some evidence that this unexpected effect was caused by the presence of more acoustical cues in the ambSR sentences. Carroll and Ruigendijk (2013) pointed out that there was a small but significant difference in the speech rate between the words in segment 2 in the ambSR and ambOR structures. The participants may have used the slower speech of the ambSR sentences to differentiate between the two sentence structures even before the PTD was reached. However, further investigations are needed to identify the reason for the early increase. With the rationale of this study and an audiological application in view, the ambSR structure is not recommended for further studies using the eye-tracking paradigm.

Precision of the estimated DM

In order to define the temporal precision of the DM, the temporal width Δt of the confidence interval of the TDA was determined at the DM (Table 2.3). That is, the width Δt of the confidence interval was calculated at the point in time at which the TDA began to exceed the $\pm 15\,\%$ threshold for a period that lasted at least 200 ms. The width Δt varied from about 275 ms to 1515 ms across the seven different sentence structures. Sentence structures with a steep slope at the DM exhibited a small Δt. The steepest slopes were measured for the ambiguous sentence structures. While Δt was the smallest for the object-first sentences with ambiguous structures (ambOVS and ambOR; $\Delta t < 500\,ms$), for unambiguous subject-first sentence structures (SVO and SR) Δt showed high variability, due to the flat slope of the TDA at the DM. Possible differences in the process of recognizing the target between unambiguous and ambiguous sentence structures may have influenced the time course of the TDA and caused a smaller Δt for the ambiguous structures. Different decision-making processes are discussed in the following section.

2.4 General discussion

An eye-tracking paradigm was introduced with a time-resolved statistical data analysis technique that enabled online analysis of the time course of the sentence comprehension process. The main objective of this study was to evaluate the paradigm for a group of listeners with normal hearing using a speech intelligibility test that was audiologically controlled with respect to speech intelligibility and linguistic complexity. The novel data analysis technique was designed to detect time-dependent effects in speech comprehension at various levels of linguistic complexity even at high speech intelligibility levels. The technique was designed with a potential application in audiological research in mind. An increase in processing time could indicate that people have trouble in everyday communication situations, since the speech rate can be about 140-180 words per minute in ordinary conversations (Wingfield and Tun, 2007). A person who is slow at sentence processing may miss speech information later in the conversation because he/she is still processing a "backlog" of past sentences or words. This slower sentence processing is interpreted as an indicator of increased cognitive processing demands even at high speech intelligibility levels. Speech intelligibility tests, in which speech recognition performance is recorded sentence by sentence, failed to detect these increased processing demands at high intelligibility. In the long run, however, this slowing down and an increased processing effort may prevent people from participating in a conversation. So far, there is no established method in audiological research that allows this kind of online analysis of speech comprehension. The results reported in this study highlighted another important advantage of the online measure: misinterpretations could be detected while the speech was presented; offline measures of processing time may be insensitive to these difficulties in sentence comprehension since participants can overcome them before the sentence is completed.

2.4.1 Effect of sentence structure on TDA and processing time

In general, processing time was expected to be increased for sentences with a higher level of linguistic complexity. Different levels of linguistic complexity were achieved using the OLACS material by altering word order, embedding relative clauses, and introducing ambiguity. In general, the results indicated that the DDD, which was interpreted as a measure of processing time, greatly depended on the sentence structure. Word order had a strong effect on sentence processing time for the verb-second structures. Longer processing times were found for the non-canonical compared to the canonical sentence structure, indicated by an increase in the DDD of almost 600 ms. An increase in processing time indicated additional cognitive processing costs, which were expected to arise from the non-canonical word order. This canonicity effect has been reported in many other psycholinguistics studies (Bader and Bayer, 2006, Gorrell, 2000, Gibson, 2000). As expected, sentence processing was slower for embedded structures: the DDD was 300 ms for the SVO structure and 730 ms for the SR structure. Interestingly, no increase in processing time was

observed for the object-relative (OR) structure compared to the subject-relative (SR) structure. It is possible that the additional processing cost of the embedded sentence structure covered any smaller differences in processing time caused by changes in word order.

Several earlier studies already reported that sentence structure complexity caused processing difficulties, increasing the cognitive processing load during speech comprehension. This was revealed using different measures, such as reaction times, recognition scores, and pupil size (Wingfield *et al.*, 2006, Tun *et al.*, 2010a, Piquado *et al.*, 2010).Tun *et al.* (2010a) presented different sentences structures and examined participants' reaction times when answering comprehension questions. They reported an increase in reaction time for complex sentence structures, indicating an imposed cognitive processing effort due to linguistic complexity even at a high intelligibility level. Piquado *et al.* (2010) reported that pupil size increased significantly during storing and processing of complex object-before-subject sentence structures compared to syntactically less complex subject-before-object sentence structures. They interpreted the pupillary enlargement as an indicator of the engagement of cognitive effort during the processing of the complex sentences. However, a significant effect of sentence structure on pupil size could only be measured after the verbal presentation of the sentence.

The results of the current study supported most of these findings, underscoring the validity of this paradigm. The DDD greatly depended on sentence structure: syntax-related difficulties during sentence processing were observed by measuring processing time. In contrast to measures such as reaction times or pupil size, used in the previously mentioned studies, the proposed eye-tracking paradigm taps into sentence processing while the sentence is being spoken. This is in line with early literature about the visual world paradigm reported that participants had difficulties during speech comprehension either on the sentence or the word processing level (Tanenhaus *et al.*, 1995, Allopenna *et al.*, 1998, Knoeferle, 2007). The fact that sentence structure had no significant effect on offline reaction times (measured by participants' button press) in this paradigm strengthens the assertion that the proposed online measure of processing speed is more sensitive for detecting processing difficulties.

Processing was expected to be slower for ambiguous sentence structures than for unambiguous structures. Interestingly, this was not the case; instead, sentence processing time was actually smaller for the ambiguous sentence structures than for their unambiguous counterparts. This was particularly evident for the ambOVS structure. Furthermore, negative TDAs indicated more fixations towards the competitor picture and were interpreted as a temporary misinterpretation of the agent and object roles. Temporal processing difficulties have been reported by Knoeferle and colleagues using the visual world paradigm (Knoeferle, 2007). They assessed online the participants' processing difficulties that arose from their expectations of thematic roles in German SVO and OVS sentence structures. The negative TDA values in the current study indicate that the eye movements and the time curve of the TDA was influenced not only the speech signal but also by the listeners' preferences and expectations. Only after the PTD did the participants realize that they had identified the wrong picture as the target picture; they then had to adjust their decision and choose the other picture; this decision is indicated by a steep increase in the TDA. This

temporary misinterpretation of the sentence led to a sudden acceleration in the decision-making process: the participant just had to choose the other picture. This may make processing faster than for unambiguous sentence structures, and is reflected in the smaller DDDs.

2.4.2 Audiological application and further research

As discussed in the previous section, our results are largely consistent with other studies, especially in psycholinguistic research investigating linguistic aspects in sentence processing. Those studies did not address audiological aspects. Moreover, (psycho-) linguistic aspects of the speech material have been considered to a lesser extent in the audiological research field to date. The data presented demonstrate the value of the paradigm for assessing aspects of cognitive processing in a speech comprehension task. The paradigm presented here was developed as a combination of methods from both research fields: recording eye-fixation data during sentence processing, which is typically used in psycholinguistic studies, and using a sentence corpus that was developed for speech intelligibility measurements. This combination may provide a useful tool for diagnostic purposes in audiology.

Research concerning the benefit of hearing aid signal processing traditionally focused on the effects on speech recognition scores or intelligibility measures (such as the SRT). However, speech reception measures often showed no sensitivity when testing, for instance, the benefit of hearing aid algorithms. One reason is that SRTs for standard speech intelligibly tests are typically at negative signal-to-noise ratios (SNRs). However some hearing aid algorithms, such as noise reduction algorithms often require positive SNRs (Marzinzik, 2000, Fredelake *et al.*, 2012) for optimum performance, i.e. a situation where speech intelligibility is high and speech intelligibility tests in audiology suffer from ceiling effects. In addition, several studies propose to focus less on improving speech intelligibility measures but rather on the effort during speech processing (Brons *et al.*, 2013, Sarampalis *et al.*, 2009). It has been shown that effort may change between conditions for which speech intelligibility remains constant and, moreover, that different hearing aids and hearing aid algorithms might affect the required effort in different ways. For instance, Brons *et al.* (2013) investigated subjectively rated effort of participants for different hearing aids and the effect of their noise-reduction outputs on the effort. They reported that the rated effort varied between different hearing aids and their noise-reduction systems even though the intelligibility was roughly the same. Hence, minimizing listening effort might be a desirable goal for fitting and adjusting hearing devices which should be supported by an effective and objective way of testing processing effort in audiology, e.g., during sentence processing as proposed here. So far, there is no effective and objective way of testing sentence processing and processing effort with standard measures and methods in audiology. The fact that the eye-tracking method introduced here was able to detect differences in processing time depending on sentence structure could also be relevant for diagnostic purposes in order to differentiate between peripheral, sensorineural-hearing loss-associated individual deficits in speech comprehension and more cognition-related, centrally

located deficits.
However, this study is only the first step towards the application of this paradigm in audiology. Note that the scope of this manuscript includes presenting the proposed method and evaluating it with the OLACS sentence corpus. A systematic study of the influence of bottom-up vs. top-down processing in background noise or hearing impairment is beyond the scope of this study, and several issues need to be clarified before the method can be broadly applied:

1. Further studies are needed to examine the interaction of sensory factors, such as hearing loss and masking noise, with the linguistic factors investigated in this study. By applying different noise types, the effect of energetic, modulation, and informational masking on speech processing and the required effort at controlled speech intelligibly levels should be investigated systematically. In addition, it has been shown that speech intelligibility can also be influenced by the rate of speech (Schlueter *et al.*, 2014), so the sensitivity of the proposed paradigm to changes in speech rates is a relevant aspect that should be addressed in future studies.

2. To gain better insight into how individual factors, such as hearing loss, might affect processing speed, it is important to assess speech processing in individual participants. The results of the current study indicate that the TDA varied widely across participants. The confidence intervals shown here include both inter-individual and intra-individual test-retest variance. A more precise TDA time course and DM could be estimated for a single participant by increasing the number of sentences per sentence structure.

3. For clinical studies, it is important to have a relatively small number of trial repetitions, so the number of sentence structures tested should be reduced for this purpose. In general, the set of verb-second structures showed strong effects on processing speed in response to changes in word order. In contrast, the expected word order effects were not seen for the relative-clause sentence structures. Consequently, of the seven different sentence structures from the OLACS corpus, the verb-second structures are the most promising for analyzing processing time and are likely to be sufficient for audiological applications.

2.5 Conclusions

This study developed and evaluated an eye-tracking paradigm that provides a time-resolved, online measure of sentence processing, revealing the influence of linguistic complexity. Experimental data from 17 participants with normal hearing tested in quiet showed that the proposed method was able to detect syntax-related delays during sentence processing using speech material that was optimized for use in audiology. As the results were in line with findings of other psycholinguistic studies, it can be concluded that the method proposed here is valid. Moreover, the experimental data showed that the proposed methods can be relevant with regard to audiological research:

1. The target detection amplitude (TDA) provides a statistically supported, time-resolved measure that directly reflects the participants' comprehension of the sentence. This measure can even be negative, which indicates a temporary misinterpretation of the presented sentence. This underlines the advantage of an online measure that provides information about the time course of speech processing.

2. The eye-tracking paradigm reveals effects of linguistic complexity on processing time that were not found in offline measures of processing speed, such as reaction time, assessed by pressing a button. Processing time was influenced by sentence structures in a systematic way, even though all measurements were performed at the same high level of intelligibility. This indicates that the proposed measure provides information about cognitive processes in speech understanding that go beyond classical speech intelligibility measures.

3. The highest contrast in processing time was observed for the SVO, OVS, and ambOVS sentence structures. Thus the verb-second structures provide a reasonable subset for practical applications, for example in audiology.

In conclusion, the paradigm presented here has a strong potential for use in audiology, where measures revealing differences in speech processing at high levels of intelligibility are highly desired.

3

Investigating sensory and cognitive effects on sentence processing speed using eye fixations

This study examined the effect of sensory factors on the speed of sentence processing. In an eye-tracking paradigm participants were asked to choose the correct pictorial representation of a sentence that was presented aurally while eye fixations were recorded. A group of 19 listeners with normal hearing aged from 20 to 31 years (mean age: 25 years) participated voluntarily. Picture recognition rates -the percentage of correctly identified pictorial representations- were obtained. Analysis of the eye fixations enabled an online measurement of processing speed during processing of sentences of varying complexity in different acoustic situations (quiet vs. stationary and modulated noise). Results indicated that even though the picture recognition rate was constant, processing was slower in the presence of noise, indicating a complex interaction between sentence structure and noise. In conclusion, using eye-tracking to assess processing speed provides a sensitive measure for detecting processing impediments caused by sensory and cognitive factors. While the most complex sentence structure employed here was unaffected by background noise, due to compensation by an appropriate processing strategy, the superadditive effect of background noise and sentence complexity indicates that sensory and cognitive effects cannot be considered as independent variables.

3.1 Introduction

Speech perception is not only studied in speech audiometry and hearing research; it is also the subject of psycholinguistic research, which considers the linguistic aspects of speech processing. While speech audiometry primarily aims at the sensory aspects of speech perception, i.e. signal-driven or bottom-up processes, psycholinguistics investigates more knowledge-driven (top-down) and cognitive processes in speech perception that include linguistic operations. However, speech comprehension includes an interaction of both signal-driven and knowledge-driven processes. Hence, a disturbed speech signal leads to a reduction of speech information and, subsequently, to a larger demand on knowledge-driven processes to recover the lost speech information in order to achieve speech comprehension (Pichora-Fuller, 2003, Zekveld *et al.*, 2010, Rönnberg *et al.*, 2008). As a consequence, an interdisciplinary approach investigating both sensory (acoustic) and cognitive (linguistic) aspects might provide a better understanding of the complex process underlying speech comprehension. The current study applies a recently developed eye-tracking paradigm by measuring eye fixations during sentence processing, which is commonly used in psycholinguistic research. Analyzing eye fixations enables an online investigation of speech processing (online in this context refers to what happens during the presentation of the speech) to analyze the speed of processing sentences. By varying the noise masker on the one hand and the level of linguistic complexity on the other hand, this study aims at clarifying the respective influence and mutual interaction of these two types of challenges on processing speed of sentence comprehension.

3.1.1 Effect of linguistic complexity on speech processing

Although linguistic complexity can influence speech intelligibility, sentence recognition tests usually do not yet take linguistic aspects into account. For instance, Uslar *et al.* (2011) analyzed the effect of linguistic complexity on speech intelligibility using short and meaningful sentences from the Göttinger sentence test (Kollmeier and Wesselkamp, 1997), which is frequently used in German audiological practice. Linguistic complexity had a small but significant effect on speech intelligibility for a group of young listeners with normal hearing. Uslar *et al.* (2010) then developed the Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) for a more systematic investigation of the effect of linguistic complexity on speech intelligibility and speech understanding (Carroll and Ruigendijk, 2013). Different levels and types of linguistic complexity were realized by manipulating different linguistic parameters, such as word order, sentence embedding, and ambiguity. Although they used appropriate test material with parametric control of linguistic complexity, the effect of linguistic complexity was slight (but still significant: 1 to 2 dB across sentence structures; Uslar *et al.*, 2013a). This observation suggests that the effect of linguistic complexity on sentence recognition is rather small. However, by using comprehension tasks and testing response times, several studies focused on speech comprehension and demonstrated that processing problems caused by linguistic complexity can be revealed even

at a high level of speech intelligibility (see Wingfield *et al.*, 2003, 2006, Tun *et al.*, 2010a). In particular, in acoustically challenging conditions (caused by, for example, background noise, decreased speech level or increased speech rate) the effect of syntactic complexity on sentence processing is enhanced. For instance, Wingfield *et al.* (2003) reported a multiplicative effect of complexity and speech rate, finding that syntactic complexity was amplified in response times when speech rates became more rapid.

3.1.2 Relationship between processing speed and processing effort

Since noise can lead to disturbances in the speech signal, knowledge-driven processes are required to recover the lost speech information needed for speech understanding. When processing demands increase, e.g. due to noise, the cognitive resources allocated for speech processing, commonly termed listening effort, also increase (Fraser *et al.*, 2010, Hicks and Tharpe, 2002). Several studies demonstrated that, even though speech recognition is unaffected by background noise, listening effort can increase (Rabbitt, 1968, Pichora-Fuller *et al.*, 1995). Thus, in order to obtain a more complete description of speech processing in noise, it is important not only to examine speech intelligibility (i.e. speech recognition performance), but also to analyze the demands of the processing underlying speech understanding. Several studies investigated subjective and objective measures of listening efforts and reported that listening effort can be sensitive to the signal-to-noise ratio (SNR; Zekveld *et al.*, 2010, 2011) and to noise type (Hällgren *et al.*, 2005, Rudner *et al.*, 2012). For instance, Rudner *et al.* (2012) reported that better performance in speech recognition (or sentence comprehension) does not necessarily result in less effortful listening when using Hagerman sentences (Hagerman, 1982). Although participants performed better in modulated noise than in stationary noise, listening in modulated noise was rated more effortful by the listeners.

Rönnberg (2003, 2008) proposed a conceptual framework for understanding effort by introducing the ease of language understanding (ELU) model. This model assumes that under optimal conditions, the linguistic information extracted from the speech signal matches the phonological representation in long-term memory, and language understanding can be successful and effortless. Under less advantageous conditions, e.g. a speech signal degraded by noise, the probability of a phonological mismatch between the extracted speech information and long-term memory increases. In this mismatch situation, the process of language understanding becomes more effortful, which the model terms explicit processing. The model assumes that working memory resources are activated when mismatch occurs, which is expected to be time-consuming.
In Chapter 2, an objective method was proposed to detect the time-consuming aspects assumed by the ELU model. An eye-tracking paradigm was used to analyze differences in the speed of processing sentences with varying linguistic complexity. The paradigm allows an online investigation of the sentence processing speed as a function of cognitive load even at a high speech intelligibility

level. A higher cognitive load (for sentences with higher linguistic complexity) led to a decrease in speech or sentence processing speed (see Chapter 2). This audio-visual paradigm required participants to combine audio and visual information, so it is more suitable to talk about processing effort rather than listening effort. The current study therefore investigated whether this audio-visual paradigm detects the increase in processing demands expected to result from background noise, which the ELU model predicts to be explicit and time-consuming.

3.1.3 Purpose of this study

The purpose of this study was to examine the contributions of sensory and cognitive factors to sentence processing in the presence of background noise using a developed audio-visual paradigm introduced in Chapter 2. For that purpose, sentences from the OLACS corpus, which vary in linguistic complexity, were presented in quiet, in stationary, and in modulated noise conditions. Both picture recognition rate and processing speed were analyzed for three different sentence structures. Processing speed is thought to provide a measure of processing effort, and therefore enables a sensitive detection of processing difficulties which cannot be revealed by picture recognition rates alone. This study addressed four hypotheses:

- Hypothesis 1: Sentence complexity reduces processing speed in all three acoustic conditions in a way that cannot be predicted from picture recognition rates. Based on the results of Chapter 2 that analyzed processing speed in quiet, the strongest decrease in sentence processing speed is hypothesized for object-verb-subject sentence structure.

- Hypothesis 2: Background noise reduces sentence processing speed. A poorer acoustical speech signal in the presence of background noise leads to a higher demand on linguistic processing and thus on cognitive processing. The second hypothesis predicts that this extra effort of sentence processing in noise can be detected by applying the proposed paradigm.

- Hypothesis 3: The noise type has an influence on processing speed. Since there is some evidence that speech processing in modulated noise is more effortful than in stationary noise (Rudner *et al.*, 2012, Larsby *et al.*, 2005), sentence processing in modulated noise is expected to be slower than in stationary noise.

- Hypothesis 4: Processing is substantially slower for linguistically complex sentences in noise than for sentences with a simple linguistic structure. As shown in the aforementioned studies, the role of cognitive processing is expected to increase with increasing sentence complexity and in noise. The interaction between the two variables is hypothesized to be superadditive : the combined effect of complexity and noise is larger than the sum of the two effects in isolation.

3.2 Material and methods

3.2.1 Participants

Nineteen voluntary participants (ten male and nine female) with normal hearing participated in the experiment. Each participant's hearing was confirmed using pure-tone audiometry. Hearing thresholds were measured at octave frequencies from 125 to 8000 Hz. Only participants with hearing thresholds better than 10 dB hearing level for all frequencies participated in this study. All participants were native German speakers between 20 and 31 years of age, with an average age of 25 years, and either had uncorrected vision or wore corrective eyewear (glasses or contact lenses) when necessary.

3.2.2 Material

Speech material

Three different German sentence structures from the OLACS corpus were used as auditory stimuli (see Table 3.1). The OLACS corpus was developed and evaluated by Uslar and colleagues (Uslar *et al.*, 2013a) to systematically investigate the effect of linguistic complexity on speech processing and comprehension (Carroll and Ruigendijk, 2013). OLACS contains seven different sentence structures that are acoustically controlled and differ in their level of linguistic complexity. Note that all seven sentence structures were presented in the current study. However, the results from only three sentence structures, namely the verb-second sentences (see Table 3.1), are discussed here and the other five sentence structures, the relative-clause structures, were used as filler sentences.

Table 3.1: Examples of the three OLACS structures used in this study. The disambiguating word from which the target picture could theoretically first be recognized by the listener is indicated by PTD (point of target disambiguation). *Nom* (nominative), *acc* (accusative), and *amb* (ambiguous case, here nominative or accusative) indicate the relevant case markings. *3Sg* indicates third person singular forms; *fem* indicates feminine gender. The meaning of the example sentence is given in quotation marks.

SVO	Der	kleine	$\text{Junge}_{\text{PTD}}$	grüsst	den	lieben	Vater.
	The_{nom}	little_{nom}	boy	greets_{3sg}	the_{acc}	nice	father.
	"The little boy greets the nice father."						
OVS	Den	lieben	$\text{Vater}_{\text{PTD}}$	grüsst	der	kleine	Junge.
	The_{acc}	nice_{acc}	father	greets_{3sg}	the_{nom}	little_{nom}	boy.
	"It is the nice father that the little boy is greeting."						
ambOVS	Die	liebe	Königin	grüsst	der_{PTD}	kleine	Junge.
	The_{amb}	nice_{amb}	queen_{fem}	greets_{3sg}	the_{nom}	little_{nom}	boy.
	"It is the nice queen that the little boy is greeting."						

The following three sentence structures from the OLACS corpus were used for the current study:

1. The subject-verb-object structure (SVO structure) represents a canonical word order with a transitive verb. Since the first article *Der* ('The'; nominative) clearly marks the subject function, the SVO structure is unambiguous right from the start of the sentence with respect to its meaning as well as to the grammatical role of each of its entities.

2. The object-verb-subject structure (OVS structure) represents a non-canonical word order with a transitive verb. Since the first article *Den* ('The'; accusative) clearly marks the object function, the OVS structure is unambiguous right from the start of the sentence with respect to its meaning as well as to the grammatical role of each of its entities.

3. The ambiguous object-verb-subject structure (ambOVS structure) exhibits a non-canonical word order with a transitive verb. Since the first article *Die* ('The'; ambiguous) could indicate either subject or object function (and subsequently agent or object role), the ambOVS structure is temporarily ambiguous with respect to its meaning as well as to the grammatical role of its entities. The identification of subject and object is not possible until the point of target disambiguation (PTD in 3.1), as the article der_{PTD} ('the'; nominative) of the second noun phrase is the first word that allows correct assignment of the subject role.

Different levels of linguistic complexity are realized by varying word order and ambiguity. In general, in the German language the subject-before-object structure is preferred to the object-before-subject structure (e.g., Bader and Meng, 1999, Gorrell,2000), which is less frequently used and argued to be derived from the canonical word order (Haegeman, 1995, Radford, 1997). Thus the subject-before-object structure is expected to be processed more easily. Adding ambiguity increases the level of complexity, since the thematic role assignment of subject and object can only be made late in the sentence (see overview by Altmann, 1998).

Graphical material

The graphical material consisted of 148 picture sets. Each picture set consisted of a target picture that illustrated the situation described by the sentence, and a competitor picture in which the agent (the entity that carries out the action) and the object (the entity that is affected by the action) roles were interchanged. The target picture was shown randomly either on the left or right side of the computer screen; the competitor picture was shown on the other side of the screen (see 3.1). The agent was always presented on the left side of each picture. In addition, filler displays were used, which contained two pictures depicting the same situation; that is, either the target or the competitor picture was depicted on both sides of the screen. However, one of the two pictures was a mirror image: the agent was presented on the right and the object on the left. Hence, in these filler trials, either both pictures matched the spoken sentence or neither did.

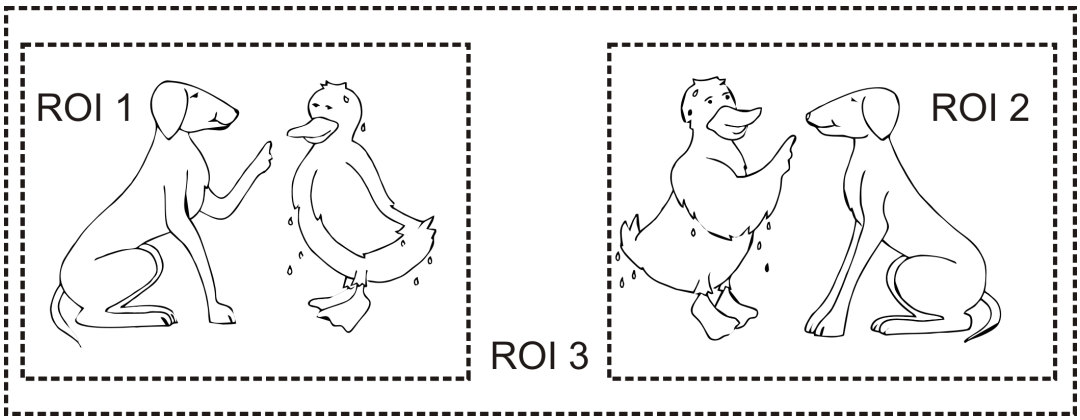

Figure 3.1: Example picture set for a sentence with the ambOVS structure: *Die nasse Ente tadelt der treue Hund*. (The wet duck (*acc.*) reprimands the loyal dog (*nom.*)). A picture set consists of two single pictures. The dashed lines indicate the three regions of interest (ROI) and are not visible for the participants. ROI 1 is the target picture and can be located on the left or right side of the picture set depending on the acoustical stimulus. ROI 2 is the competitor picture. ROI 3 is the background.

3.2.3 Stimuli and procedure

Acoustical conditions

Sentences were presented either in quiet at a level of 65 dB SPL or with one of two different noise maskers. The first noise masker was a stationary speech-shaped noise with the long-term frequency spectrum of the speaker, created by overlapping 30 tracks, each consisting of the entire randomly overlapping speech material. The second noise was the modulated ICRA4-250 noise, which is a speech-shaped noise with a female frequency spectrum and fluctuations of a single talker and originates from an English text spoken by a female speaker (original ICRA4 noise by Dreschler *et al.* (2001) modified according to Wagener *et al.* (2006), with a maximum pause length limited to 250 ms).
Participants were tested at -7 dB SNR in stationary noise and at -16 dB SNR in modulated noise. The different SNR conditions correspond to the SRT80 (the speech reception threshold at which 80 % of the words were repeated correctly) averaged over all sentence structures of OLACS, which were measured by Uslar *et al.* (2013a). Note that Uslar and colleagues found small differences in the SRT depending on the sentence structure; e.g. differences in the SRT ($< 2dB$) were found between OVS and ambOVS structure. However they also found that the speech material has nearly identical intelligibility when the sentence fragments are presented without the sentence context. Hence, differences in the SRT arise from context and not from acoustic aspects of the speech material. In the current study, this effect of context was not taken into account, and a fixed SNR, corresponding to the averaged STR80 across all sentence structures, was chosen for both noise conditions. This was realized in order to have a comparable sensory load, or the

same noise level, across all sentence structures, independent of the context effect. Moreover, the measurements of the current study differed from the SRT measurements of Uslar and colleagues, since the proposed paradigm used additional visual information to test sentence processing during speech understanding. Furthermore, no differences in intelligibility between stationary noise and modulated noise were expected at the chosen SNRs [1]. The condition in quiet, which corresponds to 100 % intelligibility, was chosen as a reference condition with minimal sensory load compared to the two noise conditions.
Eighteen sentences of each sentence structure (SVO, OVS, and ambOVS) were presented in the three acoustical conditions, resulting in 162 tested sentences. Additionally, 216 sentences with relative-clause structures functioned as filler sentences and were intermingled with the experimental sentences. Additionally, 70 filler displays were used in order to force the participants to fixate on both pictures and to avoid retrieval strategies. These filler displays were used as distractors. Taken together, 448 sentences (162 tested sentences, 216 filler sentences, and 70 distractors) were presented in five blocks in random order.

Eye-tracking paradigm

An eye-tracking paradigm was chosen in which an OLACS picture set was presented visually on a computer screen while a spoken sentence was presented via headphones. Participants were instructed to identify the target picture by pressing one of three buttons on the computer keyboard: the button "A" if the target was detected on the left side and the button "L" if it was detected on the right side of the screen. If participants were not able to clearly assign a target picture to the spoken sentence or if a filler display was presented, they were instructed to press the space button. The different buttons for the response were chosen such that participants were able to leave their hands on the keyboard during the measurement. In this way, participants did not have to look at the keyboard to search for the right button.
The visual stimulus was displayed starting 1000 ms before the onset of the acoustic stimulus until the participant responded. After each trial, participants were asked to look at a marker centered on the screen in order to perform a drift correction of the eye-tracking device. After a single training block that contained all of the picture sets, participants performed the five test blocks. At the beginning of each test block a calibration of the eye-tracking device was done using a nine-point fixation stimulus. The completion of one block of trials took about 20 minutes. After each block, participants had a break of ten minutes.

1 No significant differences were measured between the intelligibility at -7 dB SNR in stationary noise and at -16 dB SNR in modulated noise in the speech intelligibility measurements from Uslar and colleagues (Uslar *et al.*, 2013a). Note that SRT80, averaged across only the three sentence structures used in this study (SVO, OVS, ambOVS) was -6.4 dB SNR in stationary and -15.2 dB SNR in modulated noise (Uslar *et al.*, 2013a). But these higher SRT80 values lie within the test-retest variability observed for the Oldenburg sentence test (Wagener *et al.*, 2006) in stationary and modulated background noise.

3.2.4 Apparatus

An eye-tracker system (EyeLink 1000 desktop system including the EyeLink CL high-speed camera, SR Research Ltd.) was used with a sampling rate of 1000 Hz to monitor the participants' eye movements. The pictures were presented on a 22 inches multi-scan color computer screen with a resolution of 1680 x 1050 pixels. Participants were seated 60 cm from the computer screen. A chin rest was used to stabilize the participant's head. Auditory signals were presented via HDA200 Sennheiser headphones that were free-field equalized using an finite impulse response (FIR) filter with 801 coefficients according to DIN EN389-8 (2004). All experiments took place in a sound-insulated booth and the technical equipment was situated outside the booth, except for the headphones, the computer screen and the camera. For the calibration of the speech signals a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2 inch microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier were used.

3.2.5 Data analysis: Picture recognition rates

The picture recognition rate was calculated as the percentage of correctly identified pictures for each sentence structure (SVO, OVS, ambOVS) in each acoustical condition (quiet, stationary noise, and modulated noise) for all participants. Note that these picture recognition rates are not identical to speech intelligibility as the graphical display contains information that is not available in a noisy acoustic presentation. A two-way repeated measures analysis of variance (ANOVA) was applied on the picture recognition rates with sentence structure and acoustical condition as within-subjects factors. Significant effects were followed up with pairwise comparisons using post-hoc tests (applying a Bonferroni correction).

3.2.6 Data analysis: Eye fixations

The target detection amplitude (TDA) was calculated from the eye fixation data. The TDA quantifies the tendency of the participant to fixate on the target picture in the presence of the competitor picture: a positive TDA describes more fixations towards the target picture and a negative TDA describes more fixations towards the competitor picture at a given point in time. For that purpose, the eye fixations towards different regions of interest (ROI) in the display were analyzed: the target picture was defined as ROI 1 and the competitor picture as ROI 2 (Figure 3.1). In general, the TDA analysis was divided into three stages, which were introduced in Chapter 2: the first two stages consisted of the calculation of the TDA, including the time alignment and the symmetrizing. The last stage was a post-processing procedure of the TDA, including the bootstrapping procedure.

Time alignment

Since the sentences differed in length, time alignment and resampling of the recorded eye fixations were employed to associate the ROI fixations with the appropriate sentence segments (see Chapter 2 for the detailed information). For that purpose, each trial was divided into six segments. Table 3.2 shows the segment borders and the corresponding points in time (in ms), which were determined for each sentence and averaged over all sentences of all sentence structures. To synchronize the segment borders across sentences, the first five segments were individually rescaled to a fixed length of 100 samples using an interpolation algorithm. The length of segment 6 depended on the mean reaction time of the participant, with a maximal length of 200 samples. Note that only those trials in which the target was recognized correctly were used for calculation of the TDA. This resampling resulted in a segment-dependent sampling rate depending on the individual length of each segment allowing for a comparison across sentences of one structure and also between different sentence structures.

Symmetrizing and post-processing procedure

In general, participants tended to fixate more frequently on the left-hand picture. This effect was independent of the position of the target picture and was most noticeable in segment 1, i.e., before the acoustical stimulus was presented. This tendency fixating more frequently towards the left picture probably stemmed from the reading direction of the participants and the fact that the agent was always presented on the left side of the picture (except for in some filler trials). In order to eliminate the influence of the target picture position on the fixation rate, a position-dependent symmetrizing was applied in the second stage of the TDA calculation that considered four different fixation rates (see Chapter 2 for detailed calculation of the symmetrizing).
After symmetrizing, the TDA was calculated for all participants and the 95 % confidence interval was calculated by using a bootstrapping resampling procedure (Efron and Tibshirani, 1993, van Zandt, 2002). The bootstrapping was necessary because the underlying distribution of the mean value across the set of the TDA values at given point in time was unknown and could vary across sentence structures. Figure 3.3 on page 49 shows the mean TDAs (with the 95 % confidence interval) for all sentence structures and all acoustical conditions.

3.2.7 Measure of processing speed

The decision moment (DM) was calculated for each sentence structure. The DM was defined as the point in time from which the mean TDA exceeded the 15 % threshold for at least 200 ms. The threshold was chosen at 15 % TDA because smaller fluctuations in the TDA are not relevant for the investigation of speech processing (see Chapter 2). The 200 ms time requirement was set to

Table 3.2: Time segments used for time alignment across all sentences for the calculation of the TDA. The first row gives the segment borders in number of time samples. Segment 1 describes the time from the onset of the measurement until the onset of the acoustical stimulus. The spoken sentence was presented during segments 2 through 5. Segment 6 corresponds to the time between the end of the spoken sentence and the participant's response. The mean borders of each segment in ms was calculated across all sentences after the resampling procedure (with standard deviations across all sentence of the sentence structure; third row).

	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6
Segment border/ samples	0-100	100-200	200-300	300-400	400-500	500-end
Sentence structure	no acoustic stimulus	Der kleine	Junge	grüsst den	lieben Vater.	response
		The little	boy	greets the	nice father.	
Mean segment border/ms	0-1000	1000-1745 (±130)	1745-2340 (±135)	2340-2995 (±130)	2995-4140 (±151)	4140-end (±114)

account for oculomotor delay: it takes approximately 200 ms to plan and launch an eye movement (McMurray *et al.*, 2008). In order to define the temporal accuracy of the DM, the width Δt of the confidence interval at the DM was calculated along the time axis. The temporal delay between the point of target disambiguation (PTD) and the DM was calculated for each sentence structure. The PTD was defined as the onset of the word which first enabled correct recognition of the target picture (the PTD for each sentence structure is marked in Table 3.1). The time interval between the PTD and the DM was termed disambiguation to decision delay (DDD) and is interpreted as the processing time during sentence understanding within this audio-visual paradigm.

3.3 Results and discussion

3.3.1 Picture recognition rates

Figure 3.2 shows the picture recognition rates in quiet and in the two different noise conditions. The highest picture recognition rates, of about 96 %, were found for the SVO and OVS structures in quiet and in modulated noise. In all acoustical conditions, picture recognition rates were about $6-7$ % smaller for the ambOVS structure than for the SVO structure. In stationary noise, however, the highest picture recognition rate, of about 85 %, was found for the SVO structure and the lowest picture recognition rate, of about 70 %, was observed for the OVS structure. A two-way repeated measures analysis of variance (ANOVA) was applied on the picture recognition rates with

sentence structure (SVO, OVS, and ambOVS) and acoustical condition (quiet, stationary noise, and modulated noise) as within-subjects factors. The effects of sentence structure, acoustical condition and the interaction of both factors were significant ($F(2, 36) = 8.4$, $p = 0.001$; $F(2, 36) = 40.4$, $p < 0.001$; $F(2, 36) = 8.5$, $p < 0.001$, respectively). To further analyze the influence of acoustical condition and sentence structure on the picture recognition rates, one-way repeated measures ANOVAs were conducted for both factors separately.

Effect of acoustical condition

For each sentence structure, significant main effects were found for the factor acoustical condition. Post-hoc tests revealed significant differences in picture recognition rates in quiet and stationary noise for each sentence structure (SVO structure: $p < 0.001$; OVS structure: $p < 0.001$; ambOVS structure: $p = 0.03$). The picture recognition rate was expected to decrease in background noise, since speech information necessary for correct identification of the target picture could be masked by the noise; this was confirmed here: intelligibility measured in quiet was much higher (100 %) than in stationary noise (80 %). Although visual information may aid in identifying the correct target picture, stationary background noise degraded the speech signal and led to a reduction in picture recognition performance.
Surprisingly, no significant differences were found between picture recognition rates in quiet and in modulated noise. Picture recognition performance in quiet was comparable with performance in modulated noise, although speech intelligibility was lower in modulated noise (80 %). Moreover, significantly higher recognition rates were detected in modulated noise than in stationary noise for the OVS structure (p=0.001) and for the SVO structure ($p = 0.002$), although the level of speech intelligibility was comparable between the two noise conditions. This indicates that picture recognition performance, i.e. recognizing the required (linguistic) information from the speech signal in order to detect the correct target picture, was easier in modulated noise than in stationary noise for these structures, possibly resulting from the exploitation of pauses in the noise to obtain hints about the correct target picture.

Effect of sentence structure

To investigate the effect of sentence structure, a one-way repeated measures ANOVA, using sentence structure as the within-subjects factor, was conducted separately for each acoustical condition. The factor sentence structure was significant for each acoustical conditions (quiet: $F(2, 36) = 6.727$, $p = 0.003$; stationary noise: $F(2, 36) = 9.258$, $p = 0.001$; modulated noise: $F(2, 36) = 7.657$, $p = 0.002$). In quiet, pairwise comparisons applying a Bonferroni correction revealed significant differences in picture recognition rates between the SVO and ambOVS structures ($p = 0.01$); the picture recognition rate was slightly lower for the ambOVS than the OVS structure ($p = 0.05$). In general, lower recognition rates for the ambOVS structure

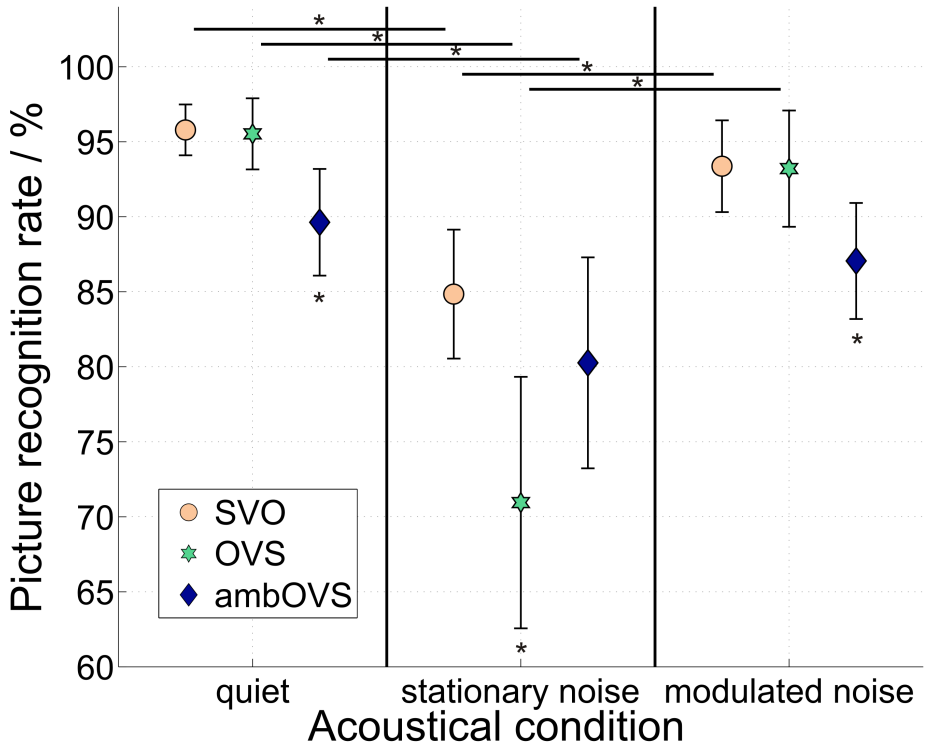

Figure 3.2: Mean picture recognition rates averaged across all participants in quiet, in stationary noise, and in modulated noise. Error bars represent interindividual standard deviations. $*$ indicates significant differences in picture recognition rates between sentence structures and between acoustical conditions (black horizontal lines).

in quiet suggest that linguistic complexity affects picture recognition performance even though the sentences were presented at an audible level, and therefore in a less demanding acoustical situation. Hence, a decrease in the recognition rate is interpreted as an initial indicator of a higher demand on cognitive resources during processing caused by an increased level of linguistic complexity (in this case due to the ambiguity of the sentence structure).

A similar trend was detected in modulated noise. Bonferroni-corrected pairwise comparisons also showed significant differences between picture recognition rates for the ambOVS and SVO structures ($p = 0.01$) and between the ambOVS and OVS structures ($p = 0.008$); recognition rates were lowest for the ambOVS structure. Although speech intelligibility was lower in modulated noise (80 %) compared to quiet (100 %), the picture recognition rate was comparable to the performance in quiet across all sentence structures and, here too, a significant reduction in recognition performance was only observed for the ambiguous structure.

In contrast to quiet and modulated noise, in stationary noise, picture recognition rates differed significantly between the OVS and SVO structures ($p = 0.004$) and between the OVS and ambOVS structures ($p = 0.02$). The considerably lower picture recognition rate for the OVS structure in stationary noise than in quiet (71 %) indicates an interaction between noise and complexity. If the information required for more linguistic operations (like role assignment of agent

and object) are masked by noise, sentence processing is expected to be more knowledge-driven. In this situation the probability of a misinterpretation increases even more for complex sentence structures. Listeners may tend to interpret the sentence as a subject-before-object structure, which is preferred to the object-before-subject structure (Gorrell, 2000, Uslar *et al.*, 2013a, Carroll and Ruigendijk, 2013). This misinterpretation is expected to lead to a stronger effect of noise on picture recognition rate for the object-before-subject structure. This was confirmed in the current study by the finding of a significantly reduced picture recognition rate for the OVS structure in stationary noise.

3.3.2 Eye fixation data

To investigate whether eye fixations can provide additional information about sentence processing or whether they just confirm the results of the picture recognition performance, TDAs were calculated to determine processing speed during sentence recognition. Figure 3.3 shows the mean TDA for all three sentence structures in quiet (panel a), stationary noise (panel b), and modulated noise (panel c). Note that the TDA for the OVS structure in stationary noise exceeded the threshold before the PTD. However, participants were not expected to discriminate between the two pictures before the PTD, and therefore this early increase in TDA was not rated as a DM.
The DM and its 95 % confidence interval and the corresponding DDDs were calculated from the TDA (see Figure 3.4). Differences between DM values (or between DDD values) were considered significant if the 95 % confidence intervals did not overlap.
In quiet, significant differences were measured between the DDDs of all three sentence structures (see lower panel in Figure 3.4). The earliest DM occurred for the simplest sentence structure, SVO; the latest DM was observed for the ambOVS structure (see upper panel in Figure 3.4). Note that the PTD of the ambOVS structure occurred later in the sentence (at the beginning of the article of the second noun phrase) compared with the other structures (see Table 3.1) resulting in a smaller DDD for the ambOVS structure. In fact, significant lower DDDs, corresponding to the highest processing speed, were found for the ambOVS structure (in all three acoustic conditions). Moreover, differences in processing speed between the SVO and OVS structures were measured even though the two sentence structures did not differ in picture recognition rates. This indicates that processing speed can provide additional information about difficulties in sentence processing that cannot be resolved by the picture recognition rates. The dependence of DDD on sentence structure matched well the expectations. In Chapter 2, differences in processing speed were measured between different sentence structures although the sentences were presented in quiet at a comfortable level, which correlates with near-to-perfect intelligibility. Differences in processing speed probably resulted from increased cognitive effort during processing of more linguistically complex sentences.
When comparing DMs between modulated noise and quiet, a significant temporal shift of the DM, causing a significantly higher DDD, was only observed for the SVO structure. This indicates

that modulated noise caused a reduction in processing speed even for the simplest sentence structure used in this study. Although no significant increase in DDD was found for the OVS and ambOVS structures, a tendency towards larger DDDs was observed, indicating a negative effect on processing speed in modulated noise compared to quiet.

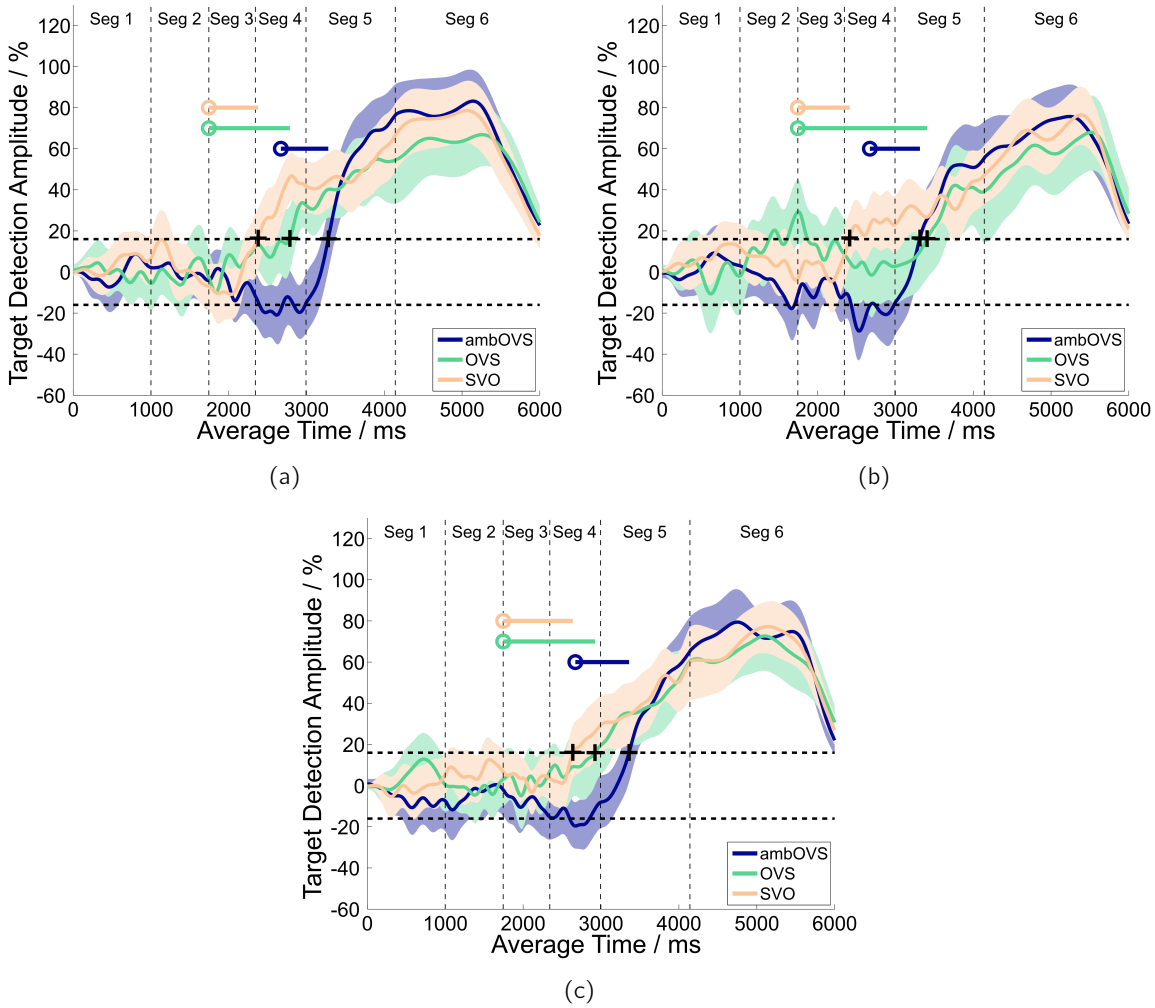

Figure 3.3: Mean TDA averaged over all participants for all three sentence structures (subject-verb-object: SVO; object-verb-subject: OVS; and ambiguous object-verb-subject: ambOVS), presented in quiet (panel a), stationary noise (panel b), and modulated noise (panel c). The x-axis of each subplot describes an averaged time scale resulting from a time alignment and resampling of the recorded eye fixations, which were applied to calculate the TDA. The vertical dashed lines mark segment borders (see Table 3.2). The horizontal dashed lines indicate the thresholds at ±15 % of the TDA. The shaded areas illustrate the 95 % confidence intervals. The + signs denote the DM where the TDA first exceeded the threshold for more than 200 ms. The circles denote the PTD, which describes the onset of the word that allows an assignment of the spoken sentence to the target picture (see Table 3.1). The horizontal lines denote the distance between the PTD and the DM, which is the DDD.

In contrast to quiet and modulated noise, in stationary noise the DM of the OVS structure occurred as late as the DM of the ambOVS structure, leading to a significantly higher DDD for the OVS structure in stationary noise compared to quiet (Figure 3.4). No significant differences were measured in DM, and therefore in DDD, between quiet and stationary noise for the SVO and ambOVS structures. Hence, this strong increase in DDD for the OVS structure confirmed the interaction between sentence complexity and stationary noise, which was indicated by the recognition rates.

3.3.3 Individual differences in the target detection amplitude (TDA) and the corresponding decision moment (DM)

This study determined the TDA time courses and corresponding DM for each sentence structure for a group of adults with normal hearing. The 95 % confidence intervals were calculated by the

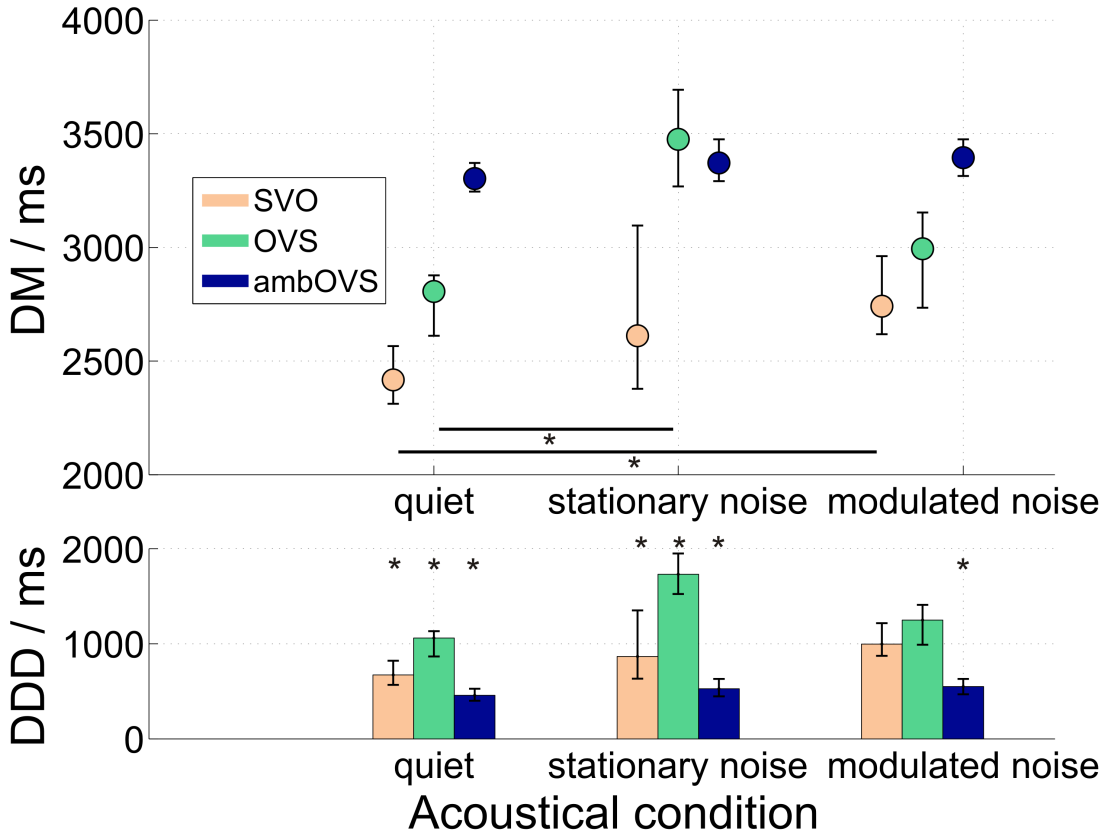

Figure 3.4: The upper panel depicts the DM and the corresponding 95 % confidence intervals along the timeline for the different sentence structures extracted from the TDA in quiet, in stationary noise, and in modulated noise. The DDD is depicted in the lower panel for each sentence structure and for each acoustical condition. The error bars indicate the 95 % confidence interval, which is identical to the confidence interval of the DM, since no error was assumed for the PTD. Horizontal lines and $*$ indicate significant differences between sentence structures and acoustical conditions (confidence intervals do not overlap).

bootstrapping procedure (shaded areas in Figure 3.3). In order to define the temporal precision of the DM and measure the individual differences between subjects, the temporal width Δt of the confidence interval was determined at the DM (Figure 3.4): that is, the width Δt of the confidence interval was calculated at the point in time at which the TDA began to exceed the 15 % threshold for a period of at least 200 ms. Note that Δt for the DDD is similar to the Δt for the DM, since the PTD is a constant value for each sentence structure, and therefore has no additional error. The width Δt varied from about 120 ms (for the ambOVS structure in quiet) to 630 ms (for the SVO structure in stationary noise) across the different sentence structures and acoustic conditions. The smallest Δt values in all three acoustic conditions were measured for the ambiguous sentence structures ($\Delta t < 300$ ms in all three acoustic conditions). Small values of Δt are connected to a steep increase in TDA. For unambiguous structures (SVO and OVS) Δt showed high variability due to the flat slope of the TDA around the DM (see Figure 3.3).

3.4 General discussion

In general, significant differences in processing were observed as a function of sentence complexity and background noise, even when no significant difference in picture recognition performance was detected. This confirms our hypothesis 1, which predicted that testing the speed of sentence processing within the proposed audio-visual paradigm can reveal processing difficulties which cannot be detected by the picture recognition rate alone. The reduction in sentence processing speed is interpreted as an indicator for increased processing effort, even when picture recognition performance is constant.

The ELU model proposed by Rönnberg (2003, 2008), describing a mathematical framework for the ease of language understanding, assumes that under unfavorable listening conditions, such as for speech processing in noise, the probability that the perceived speech or language signal does not match the stored (phonological) long-term representation increases. Therefore, speech processing becomes effortful, or explicit in ELU terminology. The model assumes that this effect results in an increase in cognitive effort to resolve the mismatch between incoming speech signal and the representation in the listener's long-term memory. Hence, within this model, explicit processing is expected to be time-consuming. We interpreted the slowing down in sentence processing as an indicator for this time-consuming explicit processing predicted by the ELU model in challenging listening conditions. That is, when processing demand increases, for instance due to linguistic complexity or background noise, a reduction in processing speed can be measured with the proposed eye-tracking paradigm.

3.4.1 Sentence complexity reduces processing speed

Slower processing of complex structures was clearly observed in quiet and in noise, further supporting hypothesis 1. At high speech intelligibility in quiet, differences in processing speed as a function of complexity was expected, since this effect was measured in Chapter 2 using the same paradigm. This effect of a non-canonical object-before-subject word order on sentence processing is well known in psycholinguistic research (see Gibson (1998) for an overview). Confirming hypothesis 1, the effect of linguistic complexity in background noise was indicated by higher DDDs for the OVS structure than for the SVO structure. A slowing down in processing is in agreement with previous studies that tested reaction times in comprehension tasks using subject-relative structures vs. object-relative structures under increased sensory demands (Wingfield *et al.*, 2003, Tun *et al.*, 2010a).

Contrary to hypothesis 1, complexity had no effect on processing speed for the ambOVS structure, even though the complexity of this structure was expected to produce a high DDD. Instead, small DDDs (between 640 ms and 740 ms, and thus smaller than those of the SVO structure) reflected a small reduction in sentence processing speed in all three acoustic conditions. There are two possible reasons for this: first, the late PTD of the ambiguous structure leads to a longer time period that can be used to visually analyze the two pictures before the correct picture has to be chosen by the participants. Hence, the DDD of the ambiguous structure might not be comparable to the DDDs of the unambiguous structures. Second, the negative threshold of the ambOVS TDA was exceeded, indicating a misinterpretation of the sentence. After participants realized that they had chosen the wrong picture, they simply had to adjust their decision and choose the other picture. A steep increase in the TDA after the DM (Figure 3.3) indicates that this processing strategy is faster than the processing strategy used for the unambiguous structures. This processing strategy used for the ambiguous structure resulted in a stable DM across all three acoustic conditions. Moreover, the stable DMs came along with a small Δt, suggesting only small individual differences between participants for the ambiguous structure. As a consequence, the ambiguous structure might not be appropriate for investigating the effect of background noise on processing speed using the proposed audio-visual paradigm. Hence, hypothesis 1 was supported for the OVS structure but not for the ambOVS structure.

3.4.2 Effect of background noise

Processing speed was measured in different noise conditions to test hypotheses 2 and 3, which predicted that background noise affects processing speed and that the type of noise is important. In both noise conditions speech intelligibility was lower than in quiet, which was expected to be reflected in a reduced picture recognition performance in noise. Surprisingly, a significant reduction in recognition performance was only measured in stationary noise. Note that only trials in which participants answered correctly were used for the calculation of the TDA, since the current study

addressed whether and how a higher sensory load would affect the speed of successful sentence processing.

Background noise reduces processing speed

Stationary noise significantly reduced picture recognition performance for all three sentence structures. In order to understand the sentence, subjects had to extract relevant speech information from the degraded speech signal, which was expected to be more demanding than understanding in quiet. However, stationary noise only had a strong impact on processing speed for the OVS structure; even though picture recognition performance was reduced in stationary noise for all three sentence structures, processing speed of the simpler SVO structure was less affected by stationary noise. Thus, stationary noise does not necessarily result in a general slowing down in sentence processing, but rather its effects are compounded by the complexity of the sentence structure. These results partly confirm hypothesis 2 (at least for the OVS structure) and are in agreement with previous studies reporting a deceleration in processing complex sentences under increased sensory demands resulting from background noise or increased speech rate (Carroll and Ruigendijk, 2013, Wingfield *et al.*, 2003). This reduced processing speed for the linguistically complex sentences suggests that the higher induced sensory load is initially detrimental - at least at the 80 % intelligibly level - when cognitive effort is higher because of increased linguistic complexity. In modulated noise, a significant decrease in sentence processing speed was only measured for the SVO structure; the reduction in processing speed for the OVS structure was not significant. Hence, in modulated noise, hypothesis 2 was only supported for the SVO structure. In general, hypothesis 2 was confirmed for the unambiguous sentence structures but not for the ambiguous structure.

Effect of noise type

To test hypothesis 3, which predicted that the noise type influences processing speed, picture recognition performance and processing speed were measured for modulated and stationary noise. Noise type influenced picture recognition performance: recognition rates were higher in modulated noise (89.7 %) than in stationary noise (78.8 %). Speech intelligibility, however, was comparable between the two noise conditions, so the difference in recognition rate cannot be a pure sensory effect, such as a simple release from masking which may arise from listening in the gaps (Bronkhorst, 2000, Wagener *et al.*, 2006). Instead, modulated noise may have enabled a better accumulation of the required acoustical information in the short dips in the noise, which could then be combined with the visual information for increased correct target recognition.
Since modulated noise did not significantly reduce recognition performance, it is also reasonable to predict that it would not cause much reduction in processing speed. However, modulated noise did reduce processing speed for the least complex structure, SVO. Thus, hypothesis 3

was supported for the simple SVO structure: processing speed was even slower in modulated than in stationary noise, reflecting more effortful processing. This is in line with previous studies investigating the perceived effort during speech processing in noise (Rudner *et al.*, 2012, Larsby *et al.*, 2005). For instance, Rudner *et al.* (2012) reported that better performance in speech recognition (or sentence comprehension) does not necessarily result in less effortful listening when using Hagerman sentences (Hagerman, 1982). Although performance was better in modulated noise than in stationary noise, participants rated listening in modulated noise as more effortful. The SVO structure of the present study, for which processing speed was significantly reduced in modulated noise, is comparable with the sentence structure of the Hagermann sentences (at least in their SRT80, see Müller, 2013). Hence, the current study confirms the observations of previous studies, in which subjective ratings demonstrated increased effort for processing in modulated noise, using an objective measure of processing speed.
However, hypothesis 3 was not supported for the more for complex sentence structures: processing was not more effortful in modulated noise for OVS and ambOVS structures than in stationary noise. The effect of noise on ambOVS processing was minimal in both modulated and stationary noise. In general, hypothesis 3 was only supported for the simple SVO sentence structure.

3.4.3 Compound effect of complexity and noise

Hypothesis 4 predicted that processing speed would be substantially slower for linguistically complex sentences than for simple sentences in noise. Linguistic complexity and noise had a compounded effect on processing speed. Most notable is the effect of stationary noise, which had a strong impact on processing speed for the complex OVS structure. The interaction of stationary noise and complexity resulted in a superadditive effect: the observed effect is greater than the sum of the effects of the two factors in isolation. For instance, linguistic complexity led (in quiet) to an increase in the DDD of about 400 ms. Furthermore, stationary noise caused an increase in DDD of about 245 ms for the simple SVO structure.[2] Thus, an additive effect of the two factors would lead to an increase of about 645 ms, yielding a DDD of 1290 ms. However, the measured interactive effect of stationary noise and complexity yielded a DDD in the range of about 1700 ms. Thus, the measured effect clearly exceeded the expected value of a simple additivity of effect size. This interactive effect of noise and complexity was not observed for the ambiguous ambOVS structure. In summary, hypothesis 4 is only supported for the OVS structure but not for the ambiguous structure.

2 Linguistic complexity can lead to an increase in DDD from 645 ms for the SVO structure to 1045 ms for OVS structure in quiet condition. Stationary noise causes an increase in DDD of SVO structure from about 645 ms to almost 890 ms.

Wingfield *et al.* (2003) reported a multiplicative effect of complexity and speech rate by measuring response times, so it made sense to test whether the observed superadditive effect could also explained by multiplication of two factors. A multiplicative interaction would have increased the DDD by a factor of 2.22 (corresponding to a DDD of 1440 ms) [3]. However, the effect observed in the current study was even stronger than a multiplicative effect: DDD for the OVS structure in stationary noise was 1700 ms (and not 1440 ms). The strong reduction in sentence processing speed observed in this study may be explained by the experimental design. Whereas Wingfield *et al.* (2003) measured processing effort after the sentence was spoken, the current study measured processing speed online. Thus, processing difficulties can be detected that occur during processing and may be overcome by the time the sentence is completed. Moreover, to increase the induced sensory load, we measured speech in noise instead of increasing the speech rate (as done by Wingfield *et al.*, 2003). Hence, our partial support of hypothesis 4 with a superadditive effect only observed for the OVS structure in stationary noise does not contradict the findings of Wingfield *et al.* (2003), because the two studies employed different experimental designs.

3.5 Conclusions

Applying the eye-tracking paradigm, four hypotheses were tested and partially confirmed:

1. A reduction in processing speed was measured for OVS vs. SVO structures in quiet and in noise which was not revealed by picture recognition rates alone.

    - Hypothesis 1 was partly confirmed: sentence complexity reduced sentence processing speed in quiet and in noise. Interestingly, complexity did not affect processing speed for the ambOVS structure, suggesting a different processing strategy within this paradigm for the ambiguous sentence structure.

2. In stationary noise, the reduction in sentence processing speed was only significant for the OVS structure, while a weak effect was observed for the SVO structure. In modulated noise, a significant deceleration in processing was observed for the SVO structure, whereas the effect for the OVS structure was weak (but not significant). Again, noise had no effect for the ambOVS structure, contrary to our expectation.

    - Hypothesis 2 was partly confirmed: background noise slowed down sentence processing

3 The DDD for the SVO structure in quiet is about 645 ms. The relative effect due to complexity (DDD for OVS/DDD for SVO) amounts to an increase in the DDD by a factor of 1.62. The relative effect due to stationary noise (DDD for SVO in noise/DDD for SVO in quiet) leads to an increase in DDD by a factor of 1.38. Thus, the expected relative delay due to multiplying the two factors amounts to a factor of $1.38 * 1.62 = 2.23$. Hence, a multiplicative effect of complexity and noise would lead to an expected delay of $645\ ms * 2.23 = 1440\ ms$ for OVS structure in stationary noise.

for SVO and OVS structures. However, noise had no effect for the ambOVS structure.

3. The noise type influenced the reduction in processing speed.

    - Hypothesis 3 was partly confirmed: the noise type influenced the reduction in processing speed. However, modulated noise only resulted in more effortful processing for the simple sentence structure.

4. The compounded effect of noise and complexity observed for the OVS structure was found to be superadditive: the combined effects of noise and complexity were stronger than the additive effect of each factor in isolation.

    - Hypothesis 4 was partly supported: a superadditive effect was observed for the OVS structure but not for the ambOVS structure.

In general, the results demonstrate that cognitive and sensory factors interact in sentence processing; a clear separation of the two factors with regard to sentence processing speed is difficult. However, the results indicate that both complexity and background noise have a strong impact on sentence processing speed as an indicator of processing effort. This underlines the need to consider both sensory and cognitive effects when analyzing sentence processing speed. Moreover, the proposed paradigm allows online detection of processing difficulties. However, the unexpected absence of a reduction in processing speed for the complex ambiguous sentence structure (ambOVS) suggests that listeners may overcome the subtle traps laid out by the experimenter by adopting an appropriate decision strategy.

Parts of this chapter are published as:

- Brand *et al.* (2012): "Recognition rates and linguistic processing: Do we need new measures of speech perception," in proceedings of ISAAR 2011: *Speech perception and auditory disorders*, 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, T. Poulsen and J. Cristensen-Dalsgaard. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, pp. 45–56.
- Uslar *et al.* (2013b): "Warum die Ente der Hund tadelt: Mögliche neue Wege in der Audiologie mit den Oldenburger Linguistisch und Audiologisch Kontrollierten Sätzen," *Zeitschrift für Audiologie* 51(1), pp. 6–15.

4

How hearing impairment affects understanding: Using eye fixations to test speed of sentence comprehension

This study investigated whether hearing impairment causes a specific change in processing speed during sentence comprehension and whether this change can be detected using eye fixations. An eye-tracking paradigm is used that records eye fixations towards a target picture that matches the aurally presented sentence in comparison to a simultaneously presented competitor picture. The single target detection amplitude (sTDA) is derived as an online measure of fixating the target picture during sentence processing. The decision moment (DM) is indicated by significant elevation of the sTDA above zero. This measure of processing speed during sentence understanding was compared across normally hearing and hearing impaired participants as across sentences that differed in linguistic complexity both in quiet and in two different noise conditions. A specific deceleration of sentence processing was found for hearing impaired listeners indicating an extra effort in sentence processing due to hearing impairment even at high levels of intelligibility. Hearing impaired listeners without acclimatization to a hearing aid exhibited the highest deceleration in sentence processing, suggesting an increased effort for this group of listeners when listening to amplified and filtered speech. Moreover, the comparison across normally hearing and hearing impaired listeners indicate significant correlations between speed of sentence processing and individual cognitive measures (such as working memory capacity).

4.1 Introduction

During speech comprehension, signal-driven (bottom-up) processes in the auditory system interact with knowledge-driven (top-down) processes and cognitive mechanisms of stimulus interpretations. Hence, the complex process of understanding speech is not only affected by the peripheral auditory level but is further influenced by more central auditory functions on a cognitive level. Elderly people in particular often report increasing problems in understanding speech in acoustically complex situations. The listeners' difficulties during sentence comprehension might arise from hearing impairment or deficits in cognitive factors, but also from an interaction between these two levels of processing (Pichora-Fuller, 2003, Schneider *et al.*, 2010). The current study aims at gaining insight into the changes of processing speed during speech understanding due to hearing impairment. The speed of processing sentences is investigated as an indicator of the corresponding processing effort of individual participants with and without hearing impairment.
The primary focus of the current study is to analyze the influence of hearing status on processing speed by applying an audio-visual paradigm. This paradigm was introduced in Chapter 2 and allows for an online analysis of processing speed by recording eye fixations during sentence understanding. Processing speed is evaluated by means of participants' eye fixations in several listening conditions with systematic variation of the required processing effort. For a systematic investigation of processing speed as a function of the required cognitive processing effort, the type and level of linguistic complexity of the presented sentences is changed, ranging from simple to more complex sentence structures. In addition, the effect of sensory demands on processing speed is examined for all sentence structures by measuring processing speed in different acoustic conditions, i.e. in quiet and in two different noise conditions. Moreover, to account for possible age-related changes on the cognitive processing level, processing speed is further analyzed with respect to individual cognitive abilities.

4.1.1 Processing effort during speech comprehension

Degradation of speech due to hearing loss not only causes listeners to miss parts of the speech signal, but further induces an increase in listening effort and processing costs (e.g. McCoy *et al.*, 2005, Wingfield *et al.*, 2005, Zekveld *et al.*, 2011). Listening effort is often used to describe the increase in cognitive resources allocated for speech processing and understanding. To investigate whether and how an increased effort caused by hearing impairment affects speech understanding, McCoy *et al.* (2005) measured recall performance of spoken words presented at intensity levels at which words could be correctly identified for listeners with normal hearing and with mild-to-moderate hearing loss. The group with hearing impairment showed poorer recall abilities for words heard in a running memory task than the group with normal hearing. Furthermore, Wingfield *et al.* (2006) showed that even a relatively mild hearing loss can increase the detrimental effects of rapid speech rates and syntactic complexity on the comprehension accuracy of spoken

sentences. Especially when the task becomes more difficult due to an increased level of linguistic complexity, hearing loss influenced the comprehension accuracy of participants. In general, the aforementioned studies showed that even at high speech intelligibility level hearing impairment can cause performance to deteriorate and thus can be indicated by a decrease in comprehension accuracy or an increase in reaction time. The results imply that hearing loss may force listeners to invest extra effort into sentence processing at the cost of other resources that would otherwise be available for the comprehension process (Pichora-Fuller, 2003, Tun *et al.*, 2010b, Wingfield *et al.*, 2005). In particular when listening becomes more effortful due to increased sensory difficulty, i.e. by a change in speech level or speech rate, and/or by increased cognitive difficulty resulting from higher linguistic complexity, the performance of the hearing impaired listeners decreased compared to that of normally hearing listeners. These results highlight the difficulties of capturing differences in processing effort using common audiological measures, such as speech reception thresholds (SRTs). At the same time they underline the need for new measures in audiological application to detect changes in processing effort due to hearing impairment.
Moreover, to gain more insight into processing difficulties during speech comprehension, an online measure for processing effort, i.e. a measure that is able to detect changes in processing effort during spoken sentence presentation, would be more appropriate. An online measure would enable detection of a temporal increase in effort during speech or sentence processing even if listeners are able to recover from this extra processing effort before the end of the spoken sentence. In Chapter 2 an online measure of processing speed was proposed and showed that eye fixations enable an investigation of processing speed during sentence processing even at good audible levels. By applying sentences from the Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) corpus (Uslar *et al.*, 2013a) a parametric variation of the type and level of linguistic complexity was reached. Within the OLACS corpus linguistic complexity is varied by changing linguistic parameters such as word order or ambiguity. As demonstrated in Chapter 2, sentence processing was slower for a group of normally hearing listeners when listening to more complex sentence structures, although speech intelligibility was high. In addition, a slowing down in sentence processing in background noise (at a fixed signal-to-noise-ratio) was detected by applying the eye-tracking paradigm, as reported in Chapter 3. It was found that in particular speed of processing complex sentences structures decreased in background noise whereby a superadditive effect of stationary noise and complexity was revealed. The reduced processing speed was interpreted as evidence for increased cognitive processing effort. The current study extends this eye-tracking approach (which so far only allowed the investigation of the processing speed of groups of participants) to individual participants.

4.1.2 The role of cognitive factors on speech processing

Audiological research in the field of speech perception has long focused on the peripheral auditory domain. However, cognitive mechanisms are also known to be important, in particular in adverse

listening situations (Pichora-Fuller *et al.*, 1995, 2008, Schneider *et al.*, 2002). A further aim of this study is to account for cognitive factors affecting speech processing. Akeroyd (2008) recently presented an overview of experimental studies that reported a relationship between participants' performance in speech recognition and their cognitive abilities. Working memory capacity in particular has been argued to be relevant for speech processing (e.g. Zekveld *et al.*, 2011, Humes *et al.*, 2006). Humes *et al.* (2006), for instance, showed significant correlations between the performance in a digit span test and performance in speech recognition tests for participants with hearing impairment. In the digit span test participants had to repeat a chain of numbers and the output of this test was interpreted as a measure of the individual's memory.

Carroll and Ruigendijk (2013) used a word-monitor paradigm to investigate reaction times for OLACS sentences. They measured reaction times at different measuring points (across the sentence) during processing of different OLACS sentences and reported a three-way interaction of noise type, measuring point, and reading span (as a measure of working memory). At certain measuring points they observed increased reaction times during sentence processing in noise, which were interpreted as an increase in local processing cost. Note that Carroll and Ruigendijk observed this interaction only for syntactically critical measuring points (with assumedly higher processing load), which suggests that listeners are able to recover from this extra processing load before the end of the sentence. These results further support the necessity of an online investigation of processing load.

Currently, it is not entirely clear how cognitive abilities influence listening effort (Zekveld *et al.*, 2011). In the current study cognitive measures are used to estimate individual differences in cognitive abilities and to ensure that the cognitive abilities of the normally hearing listeners and the hearing impaired listeners do not differ significantly. Moreover, whether and to what extent individual processing speed correlates with individual cognitive factors is also investigated. For this purpose, cognitive tests are applied to obtain a measure of working memory capacity for storage and processing. In challenging or adverse listening conditions especially, the capacity for storing and remembering words and for manipulating the stored speech signal is expected to play an important role in speech processing and understanding (see also Rönnberg, 2003, 2008, 2010). Thus, a digit span test and a word span test are applied (Tewes, 1991). In addition, understanding speech in noise is expected to be affected by the susceptibility to interference and further general attention, since noise generally may be viewed as a kind of interference (Rönnberg *et al.*, 2008, 2010). Therefore, the Stroop test (Kim *et al.*, 2005), which is a selective attention task and has been argued to be independent of span measures (May *et al.*, 1999), is used to investigate the participants' ability to ignore additional confounding visual information unrelated to the actual visual task. The span tests and the Stroop test have already shown correlations with speech perception measures using the OLACS material (see Carroll, 2012, Uslar *et al.*, 2013a).

4.1.3 The current study

This study tests if an extra deceleration of speech processing can be observed which is specifically due to sensorineural hearing loss. To that end an audio-visual paradigm is applied to obtain an online measure of processing speed as an indicator of individual processing effort. The experimental design of this paradigm has been introduced and applied in Chapter 2 and Chapter 3. The current chapter introduces a modified version of the statistical analysis of the eye fixation data, and - in contrast to the previous chapters - provides a measure reflecting the individual differences in processing speed in various listening conditions. Processing speed is analyzed by a systematical variation of external parameters such as linguistic complexity and background noise. That is, on the one hand cognitive processing demands during sentence understanding are systematically varied by changing the level and type of linguistic complexity of the speech material. On the other hand, the paradigm is applied in different acoustic conditions to account for the effect of increased sensory load on processing speed caused by background noise. The main focus is to test if the observed dependency of processing time is specific for hearing impairment rather than to selected cognitive processing parameters (such as working memory capacity). Based on the aforementioned studies, it is hypothesized that:

- Hypothesis 1: Even at a given speech intelligibility level hearing impairment causes an extra effort in speech processing in comparison to normal listeners, and therefore produces a substantial decrease in processing speed.
- Hypothesis 2: The extra effort caused by hearing impairment is highest in adverse listening conditions characterized by high cognitive demands, i.e. processing complex sentences in background noise.
- Hypothesis 3: The decrease in processing speed is related to a decreased performance in cognitive abilities that are linked to speech perception. This may be revealed by correlations between processing speed obtained with the audio-visual paradigm and individual cognitive abilities.

4.2 Material and methods

4.2.1 Participants

Seven male and thirteen female participants with normal hearing (NH group) participated in the experiment, with an average age of 59 years (ranging from 41 to 71 years). Participants had pure tone hearing thresholds of 20 dB hearing level (HL) or better at the standard audiometric frequencies in the range between 125 and 4000 Hz, and hearing thresholds of 30 dB HL or better

at 6000 and 8000 Hz (see Figure 4.1). The pure tone average (PTA) thresholds across the frequencies ranging from 125 Hz to 8000 Hz was 6.9 dB HL (standard deviation was 4.3 dB HL). In addition, nine male and thirteen female participants with hearing impairments (HI group) participated, with an average age of 65 years (ranging from 42 to 77 years). Participants belonging to this group had a mild to moderate, sensorineural, post-lingual hearing loss. The pure PTA thresholds across the frequencies ranging from 125 Hz to 4000 Hz was 39.3 dB HL (standard deviation was 6.7 dB HL).

The HI group was further divided into a group of participants acclimatized with hearing aids (i.e. hearing aids usage in daily life for more than 6 months, HA group) and a group of listeners which do not use hearing aids in their daily life (noHA group). The HA group consisted of eleven participants[1] (mean age 65 years ranging from 42 to 77 years; see Table 4.5) and the noHA group consisted of nine participants (mean age 64 years ranging from 59 to 69 years).

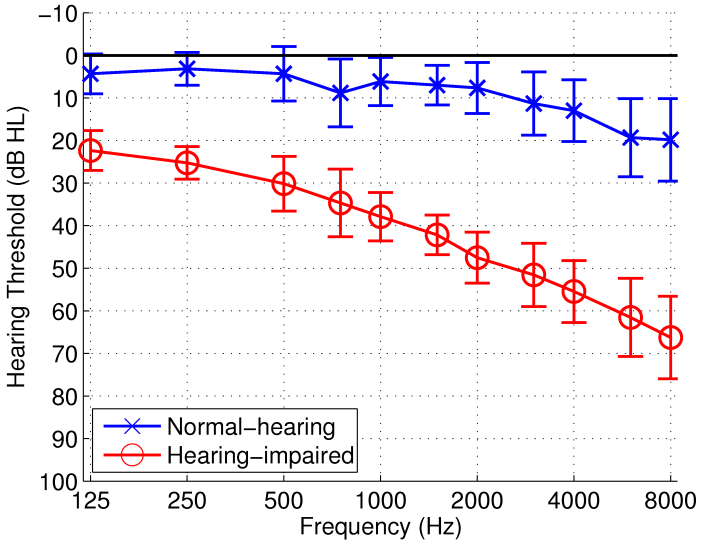

Figure 4.1: Mean hearing threshold averaged across the left and right ears for the normally hearing group and the hearing impaired group (error bars represent standard deviations across participants of the group).

4.2.2 Material

Speech material

Three different German sentence structures from the Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) corpus were used as auditory stimuli (see Table 4.1). The OLACS

1 Note that two participants were not considered for the statistical analysis, since their sTDA and the corresponding DDDs did not represent valid measures of processing speed (see Section 4.5.2).

Table 4.1: Examples of the three different OLACS sentence structures (SVO, OVS, and ambOVS) used in the current study. The disambiguating word from which the target picture could theoretically first be recognized by the listener is indicated by PTD (point of target disambiguation). *Nom* (nominative), *acc* (accusative), and *amb* (ambiguous case, here nominative or accusative) indicate the relevant case markings. *3sg* indicates third person singular forms; *fem* indicates feminine gender. The meaning of the example sentence is given by the sentence in quotation marks.

SVO	Der	kleine	Junge$_{PTD}$	grüsst	den	lieben	Vater.
	The$_{nom}$	little$_{nom}$	boy	greets$_{3sg}$	the$_{acc}$	nice	father.
	"The little boy greets the nice father."						
OVS	Den	lieben	Vater$_{PTD}$	grüsst	der	kleine	Junge.
	The$_{acc}$	nice$_{acc}$	father	greets$_{3sg}$	the$_{nom}$	little$_{nom}$	boy.
	"It is the nice father that the little boy is greeting."						
ambOVS	Die	liebe	Königin	grüsst	der$_{PTD}$	kleine	Junge.
	The$_{amb}$	nice$_{amb}$	queen$_{fem}$	greets$_{3sg}$	the$_{nom}$	little$_{nom}$	boy.
	"It is the nice queen that the little boy is greeting."						

corpus was developed and evaluated by Uslar and colleagues (Uslar *et al.*, 2013a) to systematically investigate the effect of linguistic complexity on speech processing and comprehension (see Carroll and Ruigendijk, 2013). OLACS contains seven different sentence structures that are acoustically controlled and differ in their type of linguistic complexity. Three out of the seven sentence structures, namely the verb-second sentences (see Table 4.1), were presented in the current study:

- The subject-verb-object (SVO) structure represents a canonical word order with a transitive verb. Since the first article *Der* ('The', nominative) clearly denotes the subject function, the SVO structure is unambiguous right from the start of the sentence with respect to its meaning as well as to the grammatical role of each of its entities.

- The object-verb-subject (OVS) structure represents a non-canonical word order with a transitive verb. Since the first article *Den* ('The', accusative) clearly marks the object function, the OVS structure is unambiguous right from the start of the sentence with respect to its meaning as well as to the grammatical role of each of its entities.

- The ambiguous object-verb-subject (ambOVS) structure also exhibits a non-canonical word order with a transitive verb. Since the first article *Die* ('The', ambiguous) could indicate either subject or object function (and subsequently agent or object role), the ambOVS structure is temporarily ambiguous with respect to its meaning as well as to the grammatical role of its entities. The identification of subject and object function is not possible until the point of target disambiguation (denoted by PTD in Table 4.1), as only the article *der* ('the', nominative) of the second noun phrase allows a correct assignment of the subject role.

Different types of linguistic complexities are realized by varying word order and ambiguity. In general, the SVO structure is preferred in the German language compared to the OVS and ambOVS structures, which are less frequently used, and are argued to be derived from the canonical word

order (Gorrell, 2000, Bader and Meng, 1999). Thus the subject-before-object structure is expected to be processed more easily. Adding ambiguity further increases the level of complexity, since the thematic role assignment of agent and patient can only be made at a late point in the sentence (see overview by Altmann,1998).

Visual stimuli

In order to present the visual stimuli, picture sets corresponding to the sentences of the OLACS corpus were used. For each spoken sentence, a set of two pictures was shown, consisting of one target and one competitor picture. The target picture illustrated the situation described by the spoken sentence (see left panel in Figure 4.2). In the competitor picture, the roles of the agent (i.e. the entity that carries out the action) and the object (i.e. the entity that is affected by the action carried out) were interchanged (see right panel in 4.2). Both pictures were of the same size and the agent was always presented on the left side. In addition, filler displays were used, in which either the target or the competitor picture was depicted on both sides of the screen, where one was mirrored, i.e. the agent was presented on the right and the object on the left side of the picture. Hence, when filler displays were presented, either both of the pictures matched the spoken sentence or neither of the pictures corresponded to the acoustical stimulus.

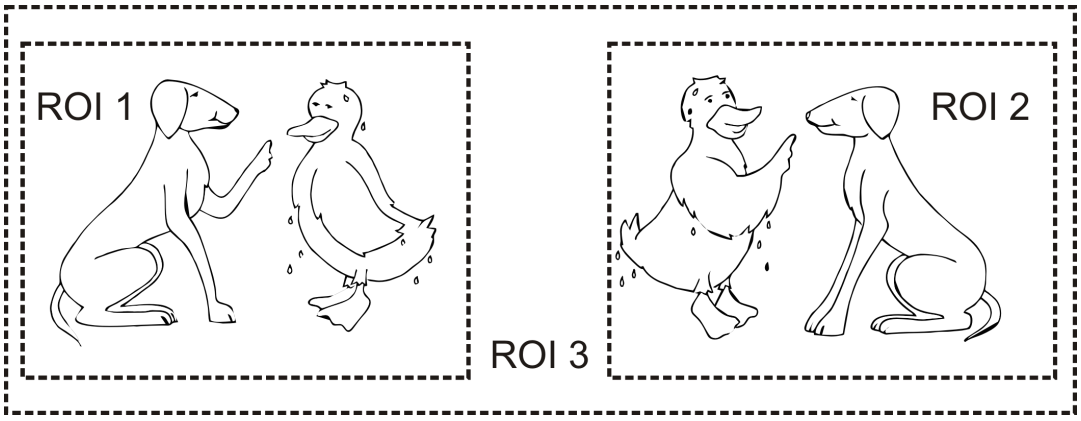

Figure 4.2: Example picture set for a sentence with the ambOVS structure: *Die nasse Ente tadelt der treue Hund*. (The wet duck (*acc.*) reprimands the loyal dog (*nom.*)). A picture set consists of two single pictures. The dashed lines indicate the three regions of interest (ROI) and are not visible for the participants. ROI 1 is the target picture and can be located on the left or right side of the picture set. ROI 2 is the competitor picture. ROI 3 is the background.

4.2.3 Stimuli and procedure

Acoustical conditions

Sentences were presented either in quiet or with one of two different noises. The first noise masker was a stationary speech-shaped noise with the long-term frequency spectrum of the speaker, created by overlapping 30 tracks, each consisting of the entire randomly overlapping speech material. The second noise was the modulated ICRA4-250 noise, which is a speech-shaped noise with a female frequency spectrum and fluctuations of a single talker and originates from an English text spoken by a female speaker (original ICRA4 noise by Dreschler *et al.*, 2001, modified according to Wagener *et al.*, 2006, with a maximum pause length limited to 250 ms).
The sound level of the stimuli was 75 dB SPL, but was adjusted if listeners preferred a higher level. The spectrum of speech and noise was additionally adjusted according to the individual hearing loss using the NAL-R formula (Byrne and Dillon, 1986) to ensure that each listener had roughly the same spectral information available. Based on the individual audiogram, the required gain was applied separately for different frequency bands using a multichannel filter bank.
To ensure comparable speech intelligibility levels across all participants, every participant was measured at his or her individual speech reception threshold of 80 % word understanding (SRT80). For that purpose, the individual SRT80 was measured for each sentence structure (SVO, OVS, and ambOVS structure) in stationary noise and in modulated noise (see Section 4.3).

Procedure

In total, 180 OLACS sentences (60 of each sentence structure) were presented in all three acoustic conditions. That is, each sentence was presented in quiet (at 100 % speech intelligibility) and in two noise conditions (at 80 % speech intelligibility in stationary and in modulated noise) in a randomized order. In addition, 64 filler displays with the corresponding sentences from OLACS were presented in filler trials (see Section 4.2.2). The filler trials were incorporated to force the participants to analyze both pictures. In total, 604 trials were conducted for each participant. One training block consisting of 60 sentences was performed by each participant at the beginning of each session to become familiar with the material (especially with the visual stimuli). After the training block, participants performed 14 test blocks. One test block took about 8 minutes. After three blocks, participants had a break of ten minutes. The complete measurement took about three hours per participant and was divided into two sessions, which were performed on different days within one week.

Eye-tracking paradigm

The visual stimulus containing the two alternative scenes was presented from 1000 ms before the onset of the acoustic stimulus until the response of the participant. Participants were instructed to identify the picture that matches the acoustic stimulus by pressing a button as soon as possible after the presentation of the spoken sentence. To identify the position of the target picture, which could be located either on the left or the right side of the display, a box with three buttons was used. Participants were asked to push the left button, if the target was presented on the left side and the right button, if they identified the target on the right side of the screen. If participants were not able to clearly assign one target picture to the spoken sentence, they were instructed to press the middle button of the box. After each trial, participants were asked to look at a marker, which was centered on the screen in order to perform a drift correction. At the beginning of each test block a calibration was done using a nine-point fixation stimulus.

4.2.4 Apparatus

An eye-tracker system (EyeLink 1000 desktop system including the EyeLink CL high-speed camera, SR Research Ltd.) was used with a sampling rate of 1000 Hz to monitor participants' eye movements. The pictures were presented on a 22 inches multi-scan color computer screen with a resolution of 1680 x 1050 pixels. Participants were seated 60 cm from the computer screen. A chin rest was used to stabilize the participant's head. The eye tracker sampled only from one eye. Auditory signals were presented via closed headphones (Sennheiser HDA 200) that were free-field equalized according to DIN EN389-8 (2004). For the calibration of the speech signals a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2 inch microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier were used. All experiments took place in a sound-insulated booth.

4.3 Preparatory measurements

4.3.1 Speech recognition measurements

In order to ensure that participants conducted the eye-tracking experiment at the same speech intelligibility (independent of sentence structure and/or hearing status), speech recognition was measured for each participant before the eye-tracking experiment started. For that purpose, sentences from the OLACS corpus were presented in stationary noise or in modulated noise (the same noise types that were used for the eye-tracking paradigm, see Section 4.2.3) in a sound-insulated booth over headphones (Sennheiser HDA 200). Participants were asked to repeat all words of the presented sentence as accurately as possible. The correctly repeated words within

Table 4.2: Mean SRT80 (with standard deviations) averaged across the participants with normal hearing (NH group) and hearing impairment (HI group) for all three sentence structures. In addition, the mean results (with standard deviations) of the cognitive tests, i.e. the Stroop test, the digit span test, and the word span test, are shown for the NH and HI groups.

	Sentence structure			**Cognitive tests**		
	SVO	OVS	ambOVS	Stroop	Digit-score	Word-score
NH group	SRT80 in stationary noise			1197 ms (±225 ms)	51 % (±17 %)	28.9 % (±9 %)
	-4.4 dB (±1.3 dB)	-3.6 dB (±1.4 dB)	-4.2 dB (±0.7 dB)			
	SRT80 in modulated noise					
	-9.8 dB (±2.1 dB)	-8.1 dB (±2.7 dB)	-7.8 dB (±2.7 dB)			
HI group	SRT80 in stationary noise			1295 ms (±244 ms)	48 % (±12 %)	27.9 % (±10 %)
	-1.5 dB (±2.7 dB)	-0.1 dB (±2.6 dB)	-0.5 dB (±2.3 dB)			
	SRT80 in modulated noise					
	-0.1 dB (±3.8 dB)	2.3 dB (±3.1 dB)	1.9 dB (±3.0 dB)			

one sentence were counted. An adaptive procedure was used to determine the SRT80, i.e., the SNR at which 80 % of the speech material was recognized correctly (see Uslar *et al.* (2013a) for detailed information about the measurement procedure). Within the adaptation procedure, the speech level of each sentence was adjusted according to the number of correctly recognized words (see Brand and Kollmeier (2002) for details). In order for participants to become familiarized with the test material, they first performed one training list. After training, two test lists were presented, one for each noise condition (i.e. stationary noise and modulated noise). The training list and the test lists each contained 20 sentences of each sentence structure, resulting in 60 sentences in total. Sentences with different sentence structures were presented in a random order within one list for each listener. The SRT80 was determined for each sentence structure and each participant. The averaged SRT80s for both groups (NH and HI group) are listed in Table 4.2.

4.3.2 Cognitive tests

In order to reveal correlations between cognitive abilities and processing speed, all participants performed three cognitive tests: a Stroop test, a digit span (backward) test and a word span (forward) test, in random order, as described below.

The Stroop test was employed to obtain a measure of the selective attention of the participant and a measure of his or her susceptibility to interference. The paradigm of the Stroop test followed that used by Kim and colleagues (Kim *et al.*, 2005). A colored rectangle and a written word were presented simultaneously on a computer screen. Participants were asked to decide whether the meaning of the word matched the color of the rectangle. Since the color of the written word and

the color of the rectangle could differ, the color of the written word (not the meaning of the word) was to be ignored while performing the task as quickly as possible. After a training block of ten trials, mean reaction times were measured for each participant.
Furthermore, participants performed two different span tests, including the digit span test, which is part of the verbal HAWIE-R intelligence test (the revised German version of the Wechsler Adult Intelligence scale; Tewes, 1991). In the backwards version of this test, a chain of digits was presented aurally and participants were then asked to repeat the chain in reversed order. To calculate the span scores for the span test, one point was awarded for every correctly repeated trial (according to the traditional scoring; see Tewes, 1991).The digit-scores were presented in percentages, i.e. the reached point were divided by the possible points (see Table 4.2). The digit span backwards test is expected to measure the storage and processing capacity of the working memory system and the ability to manipulate the content of working memory (e.g. Kemper *et al.*, 1989, Cheung and Kemper, 1992).
The word span test was based on the same experimental design but used semantically unrelated words (one and two syllables) instead of digits. The word span test was conducted as a forward version, i.e. participants were asked to repeat the chain of words in the original order. The word-scores also were presented in percentages, i.e. the reached points were divided by the possible points. The word span forward test is expected to provide a measure for pure verbal memory (span) capacity.

4.4 Data analysis

4.4.1 Analysis of the eye fixation data

The recorded eye fixation data were transformed into the single target detection amplitude (sTDA). For that purpose, the eye fixations towards different regions of interest (ROI) on the display were determined: the target picture (ROI 1) and the competitor (ROI 2) picture (see Figure 4.3). The sTDA quantifies the tendency of a single participant to fixate on the target picture in the presence of the competitor picture. Thus, a positive sTDA describes more fixations towards the target picture and a negative sTDA describes more fixations towards the competitor picture. The calculation of the sTDA was divided into three processing stages: the sentence-based processing stage, the sentence-structure-based processing stage, and the post-processing stage. Location of the target picture was taken into account in the calculation of the sTDAs (see left panel in Figure 4.3); this distinction was made necessary since participants tended to fixate more frequently on the left picture regardless of whether this was the target picture or not, which was assumed to be caused by the German reading direction (see Chapter 2). Further, only trials in which the participants selected the correct picture were considered for data analysis. This was done in order to analyze only those eye movement time patterns that reflect the dynamics of the recognition

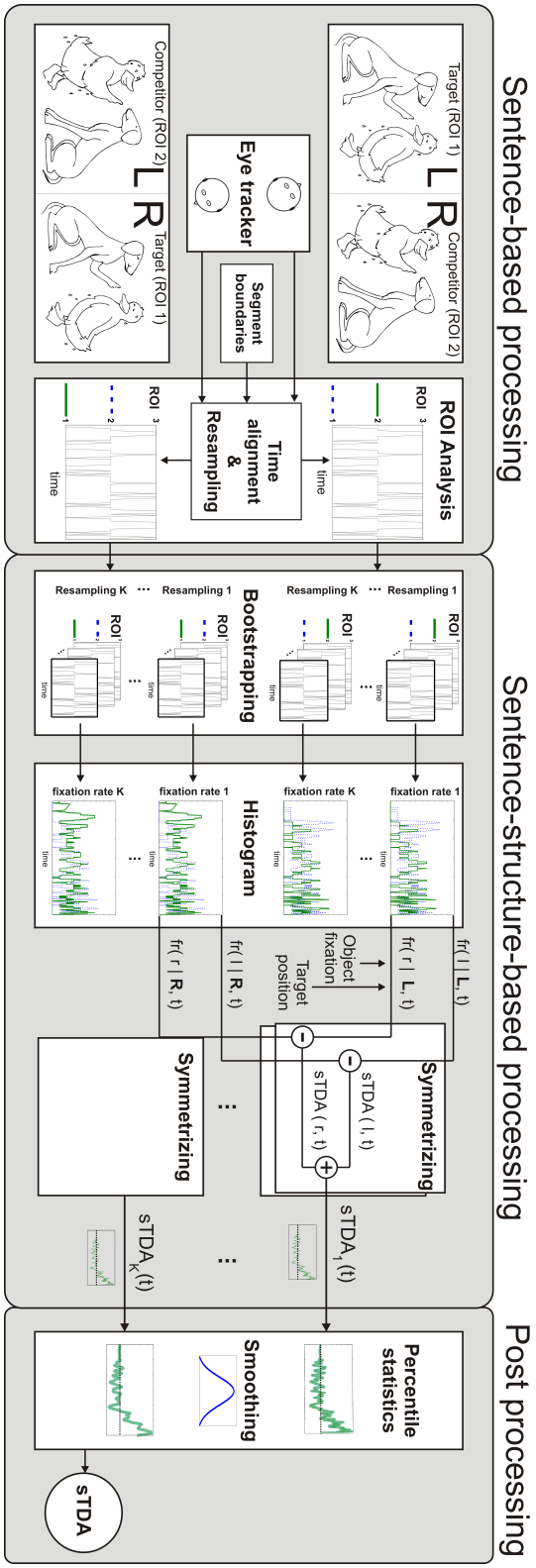

Figure 4.3: Schematic visualization of the calculation of the single target detection amplitude (sTDA). The calculation of the sTDA consists of three processing stages, namely the sentence-based processing, the sentence-structure-based processing, and the post processing stage.

process for correctly identified sentences. Note that the calculation of sTDA is based on the calculation of TDA (introduced in Chapter 2) in most of the processing stages, with only a few differences. Moreover, the sTDAs of the different sentence structures were derived independently for each participant. The calculation of sTDA is introduced in the following.

Time alignment and resampling

In the first sentence-based processing stage, the eye fixations towards the target (ROI 1), the competitor (ROI 2), and the background (ROI 3) were calculated as a function of time for each sentence $n = 1, ..., N$ in that particular condition. Since the sentences differed in length, a time alignment of the recorded eye fixation data was employed to allow comparisons across sentences. Therefore, each sentence was divided into 4 segments, as depicted in Table 4.3. Segment 1 describes the time from the onset of the visual stimulus until the onset of the acoustical stimulus, which had a fixed length of 1000 ms. The spoken sentence was presented during segments 2 through 5. The time from the end of the spoken sentence until the listener's response by pressing the response key was allotted to segment 6. The segment borders and the corresponding points in time (in ms) during the eye-tracking recordings were determined for each sentence and averaged over all sentences of a single sentence structure (see Table 4.3). The time alignment and resampling were performed for each sentence n to allow a comparison across all sentences. To synchronize the segment borders across different sentences, the first five segments were individually rescaled to a fixed length of 100 samples using linear interpolation. The length of segment 6

Table 4.3: Time segments used for time alignment across all sentences for the calculation of the TDA. The first row gives the segment borders in number of time samples. Segment 1 describes the time from the onset of the measurement until the onset of the acoustical stimulus. The spoken sentence was presented during segments 2 through 5. Segment 6 corresponds to the time between the end of the spoken sentence and the participant's response. The mean borders of each segment in ms was calculated across all sentences after the resampling procedure (with standard deviations across all sentence of the sentence structure; third row).

	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6
Segment border/ samples	0-100	100-200	200-300	300-400	400-500	500-end
Sentence structure	no acoustic stimulus	Der kleine / The little	Junge / boy	grüsst den / greets the	lieben Vater. / nice father.	response
Mean segment border/ms	0-1000	1000-1745 (±130)	1745-2340 (±135)	2340-2995 (±130)	2995-4140 (±151)	4140-end (±114)

depended on the mean reaction time of the participant with a maximal length of 200 samples (see Table 4.3). If the reaction time, for instance, was 1500 ms, the last segment was rescaled to a length of 150 samples. For reaction times above 2000 ms, the signal was restricted to a maximal length of 200 samples. This segment-based resampling led to a segment-dependent sampling rate due to the individual length of each segment.
After this time alignment and resampling stage, the fixation $F_t(n)$ towards one of the three ROIs, that is $F_t(n) \in ROI1, ROI2, ROI3$, was available as a function of the time index $t = 1, ..., T$ for all $n = 1, ..., N$ sentences.

Bootstrapping procedure and symmetrizing

In the second stage, a bootstrapping resampling procedure (Efron and Tibshirani, 1993, van Zandt, 2002) was applied in order to transform a set of fixations into a set of fixation rates, allowing for statistical analysis. Therefore, for a given time index t, data resampling was performed by randomly selecting a set of N fixations with replacement. This resampling was repeated K times ($K = 10,000$), resulting in a matrix consisting of NxK fixations for each time index t:

$$M_t = \begin{pmatrix} F_t(1,1) & \dots & F_t(N,1) \\ \vdots & \ddots & \vdots \\ F_t(1,K) & \dots & F_t(N,K) \end{pmatrix} \tag{4.1}$$

Afterwards, the fixations for all bootstrapping realizations K were transformed into a set of fixation rates by computing a histogram over N fixations for each of the K realizations at a given time index t. This histogram analysis was realized for every time index, resulting in a time-dependent fixation rate $fr(1,t), ..., fr(K,t)$.
In order to compensate for the participants' tendency to fixate more frequently on the left picture, the computation of the fixation rate was performed independently for the target picture (ROI 1) and the competitor picture (ROI 2). Further, the computation was performed separately for target presentation on the left or on the right side of the computer screen, resulting in the following four average fixation rates:

- $fr(r|L,t)$: fixation towards the competitor picture, while target was present of the left side
- $fr(l|R,t)$: fixation towards the competitor picture, while target was present of the right side
- $fr(l|L,t)$: fixation towards the target picture, while target was present of the left side
- $fr(r|R,t)$: fixation towards the target picture, while target was present of the right side

Given these four fixation rates, the symmetrizing was performed as follows:

$$\begin{aligned} sTDA(r,t) &= fr(r|R,t) - fr(r|L,t) \\ sTDA(l,t) &= fr(l|L,t) - fr(l|R,t), \end{aligned} \tag{4.2}$$

where the sTDAs for the right side ($sTDA(r,t)$) and the left side ($sTDA(l,t)$) were added to $sTDA(t)$. Note that the fixation rates of the background (ROI 3) were not considered in the calculation of the $sTDA(t)$.

Post-processing

Finally, a percentile statistic was used to determine the mean $sTDA(t)$ and the 95 % confidence interval over all K realizations, ranging from $sTDA_1(t)$ to $sTDA_K(t)$ (cf. Figure 4.3). A Gaussian smoothing filter with a kernel size of 25 samples was applied in order to reduce the random fluctuations. The resulting signal was termed sTDA and was calculated for each participant and for each sentence structure. In general, the sTDA quantifies the tendency of an individual participant to fixate on the target picture during sentence processing. Thus, a positive sTDA indicates more fixations towards the target picture at a certain point in time, whereas a negative sTDA describes more fixations towards the competitor picture, i.e. towards the wrong picture.

Calculation of the decision moment (DM) and the disambiguation to decision delay (DDD)

In order to investigate processing speed during sentence comprehension, the decision moment (DM) and the disambiguation to decision delay (DDD) were determined, as described in Chapter 2 on the basis of the TDA. The DM is defined as the point in time for which the mean sTDA exceeds a threshold value of 15 % for at least 200 ms. For the following discussion of the temporal position of the DM, an oculomotor delay was taken into account, since it was assumed that the time to plan and perform an eye movement takes about 200 ms (see, e.g. McMurray *et al.*, 2008). Thus, time points at which the threshold was exceeded for fewer than 200 ms were not considered for the determination of the DM. The distance between the point of target disambiguation (PTD) and the DM was calculated for each sentence structure. The PTD is defined as the onset of the word from which the recognition of the target picture is theoretically possible (the PTD for each sentence structure is marked in Table 4.1). The temporal delay between the PTD and the DM is interpreted as a measure for processing time and is termed disambiguation to decision delay (DDD).

4.4.2 Statistical analysis

For statistical analysis the picture recognition rate (i.e. the percentage of correctly recognized pictures) and the reaction time (i.e. the time after the presentation of the spoken sentence until the participant's response) were subjected to a repeated-measures analysis of variance (ANOVA) with sentence structure (SVO, OVS, and ambOVS) and acoustic condition (quiet, stationary noise, and modulated noise) as within-subject factors (i.e. across all participants) and hearing status as a between-subject factor (i.e. between NH and HI group); picture recognition rate and reaction time were used as the dependent measures. Bonferroni post-hoc tests (level of significance set at $p = 0.05$) were used to determine the sources of significant effects indicated by the ANOVA.

4.5 Results and discussion

4.5.1 Picture recognition rates and reaction times

The mean picture recognition rates, i.e. the percentage of correctly answered trials, for all conditions are depicted in Table 4.4. In general, picture recognition rates were very high (95.5 % averaged over all conditions). Note that the picture recognition rates are not comparable with speech intelligibility, which is at 100 % in quiet and 80 % in both noise conditions, since the graphical display contains additional visual information. Hence, the picture recognition rate describes more the listeners' ability to combine the acoustical and the visual information and to extract important speech information out of the noise signal during sentence processing in noise.

The ANOVA revealed an effect of sentence structure on picture recognition rates ($F(2, 74) = 6.49$, $p = 0.003$), and post-hoc tests showed significantly higher picture recognition rates for the SVO structure compared to the OVS structure ($p = 0.018$) and for the ambOVS structure compared to the OVS structure ($p = 0.032$). High picture recognition rates for the ambOVS structure are rather striking, since the ambOVS structure was expected to exhibit a high level of linguistic complexity. However, Uslar *et al.* (2013a) showed that speech intelligibility is higher for the ambOVS structure compared to the OVS structure. In the proposed paradigm, the depicted scenes provide additional information, and therefore may lead to higher recognition performance for the ambOVS structure. In contrast to the unambiguous structures, the ambOVS structure contains either the plural or the female form of the object (see Table 4.1 'Die liebe Königin'), so subject and object roles might be better distinguished with the help of the additional visual information. The effect of sentence structure suggests that recognition rates were affected by linguistic complexity, although all sentence structures were presented at a comparable level of speech intelligibility in all three acoustic conditions.

Table 4.4: Mean picture recognition rates (with standard deviations across participants) for the NH and the HI group and mean reaction times (with standard deviations across participants) are shown for the three sentence structures presented in quiet, in stationary noise, and in modulated noise for both groups.

		Picture recognition rate / %			**Reaction time / ms**		
		SVO	OVS	ambOVS	SVO	OVS	ambOVS
quiet	NH	98.8 (±2.1)	97.8 (±1.8)	98.1 (±1.9)	1144 (±484)	1194 (±437)	1178 (±476)
	HI	95.4 (±3.6)	96.1 (±3.5)	95.7 (±3.9)	1628 (±1062)	1655 (±1049)	1606 (±1051)
stat. noise	NH	97.9 (±1.4)	94.4 (±5.2)	96.5 (±2.8)	1139 (±481)	1161 (±438)	1107 (±483)
	HI	93.7 (±4.9)	91.4 (±6.4)	95.2 (±4.1)	1639 (±1271)	1621 (±1049)	1685 (±1151)
mod. noise	NH	96.8 (±2.8)	94.8 (±9.5)	96.6 (±2.3)	1227 (±536)	1106 (±423)	1159 (±461)
	HI	93.3 (±7.1)	93.1 (±7.3)	93.7 (±6.7)	1765 (±1250)	1691 (±1115)	1617 (±1219)

To further investigate the effect of sentence structure within the different acoustical conditions, one-way repeated ANOVAs with sentence structure (SVO, OVS, and ambOVS) as within-subject factors were conducted separately for all three acoustic conditions (quiet, stationary noise, and modulated noise). Sentence structure had no effect in quiet or in modulated noise. However, in stationary noise, an effect of sentence structure on picture recognition rate was detected ($F(2, 76) = 9.894$, $p < 0.001$). The post-hoc tests revealed significantly lower rates for the OVS structure compared to the SVO structure ($p = 0.002$) and compared to the ambOVS structure ($p = 0.004$) in stationary noise.

More importantly, the analysis revealed between-subject effect of the picture recognition rates ($F(1, 37) = 4.329$, $p = 0.044$), reflecting significant differences between the NH and HI group. Paired comparisons showed significantly higher picture recognition rates for the NH group (for the SVO structure in quiet: $t(37) = 2.8$; $p = 0.009$ and for the SVO structure in stationary noise: $t(37) = 3.218$; $p = 0.003$). These results support the expectation that the NH group performed better than the HI group in some conditions, although both groups were measured at the same level of speech intelligibility.

The mean reaction times for all experimental conditions are shown in Table 4.4. In general, the HI group tended to have larger reaction times across all conditions compared to the NH group. Again, an ANOVA was performed using sentence reaction time as the dependent measure. The analysis revealed no effect of sentence structure or noise, and no between-subject effect, i.e. no significant differences between both groups. Thus, although the listeners with hearing impairment tended to have longer reaction times, this effect was not significantly different between the two groups.

4.5.2 Eye fixation data

For a comparative investigation of the processing speed for all three sentence structures, the disambiguation to decision delay (DDD) was calculated for each listener. A high DDD indicates a strong deceleration in processing speed for the corresponding sentence structure. Figure 4.4 depicts individual single target detection amplitudes (sTDAs) of four listeners with the corresponding decision moment (DM) and DDD for each sentence structure. These sTDAs are shown to exemplify the different sTDA time courses that can occur for individual participants. Since most of the individual data show similarities to either one of these four examples, the sTDAs and the corresponding DMs of the participants can be classified into the following groups:

1. Panel (a) in Figure 4.4: The sTDAs displayed here are characteristic for the majority of the participants. That is, early DMs occurred for the unambiguous SVO structure, and late DM for the OVS and the ambOVS structures. The greatest DDD was observed for the OVS structure.

2. Panel (b) in Figure 4.4: No differences in the DM of the SVO and the OVS structures were observed for some participants. That is, early DMs occurred for both unambiguous SVO and OVS structures independent of the complexity of the sentence structure. A late DM was only observed for the ambiguous sentence structure (ambOVS), but no differences in the corresponding DDD were observed between the three sentence structures.

3. Panel (c) in Figure 4.4: Late DMs occurred after the presentation of the spoken sentence for some participants, independent of the sentence structure. Again, hardly any differences between the DM and the corresponding DDD of the SVO, OVS, and ambOVS structures were observed for this participant.

4. Panel (d) in Figure 4.4: The sTDA and the corresponding DDD of participants were not appropriate for obtaining a measure of processing speed (number of participants: 3). The sTDAs of these participants exhibited only small amplitudes, which barely exceeded the 15 % threshold. One participant out of twenty in the NH group and two of the HI group are classified to belong to this group as they showed a flat time course of the sTDAs. As a result, the corresponding DDDs did not represent valid measures of processing speed, and therefore, the data from these participants were not considered for the statistical analysis.

To investigate the effect of hearing loss on the processing time, averaged DDDs were calculated for both groups (NH and HI). Figure 4.5 depicts the averaged DDDs of the different sentence structures in quiet and in the two noise conditions for both groups. To exclude possible effects of non-normal distribution, small samples or unequal variance, bootstrap tests for paired samples (two-tailed) based on 10,000 bootstrap samples were applied (alpha level of significance set to $p = 0.05$, adjusted for FDR correction) for comparison between different sentence structures and

noise conditions (Nichols and Holmes, 2001). Unpaired tests were applied to investigate whether DDDs varied between the two groups (NH and HI group). To verify the statistical significance the bootstrap p values were determined.

Effect of sentence structure

Significant differences in processing speed were observed between all three sentence structures in quiet as well as in background noise ($p < 0.001$). The greatest temporal DDDs were measured for the OVS structure in quiet and in background noise (about 1400 ms averaged across both

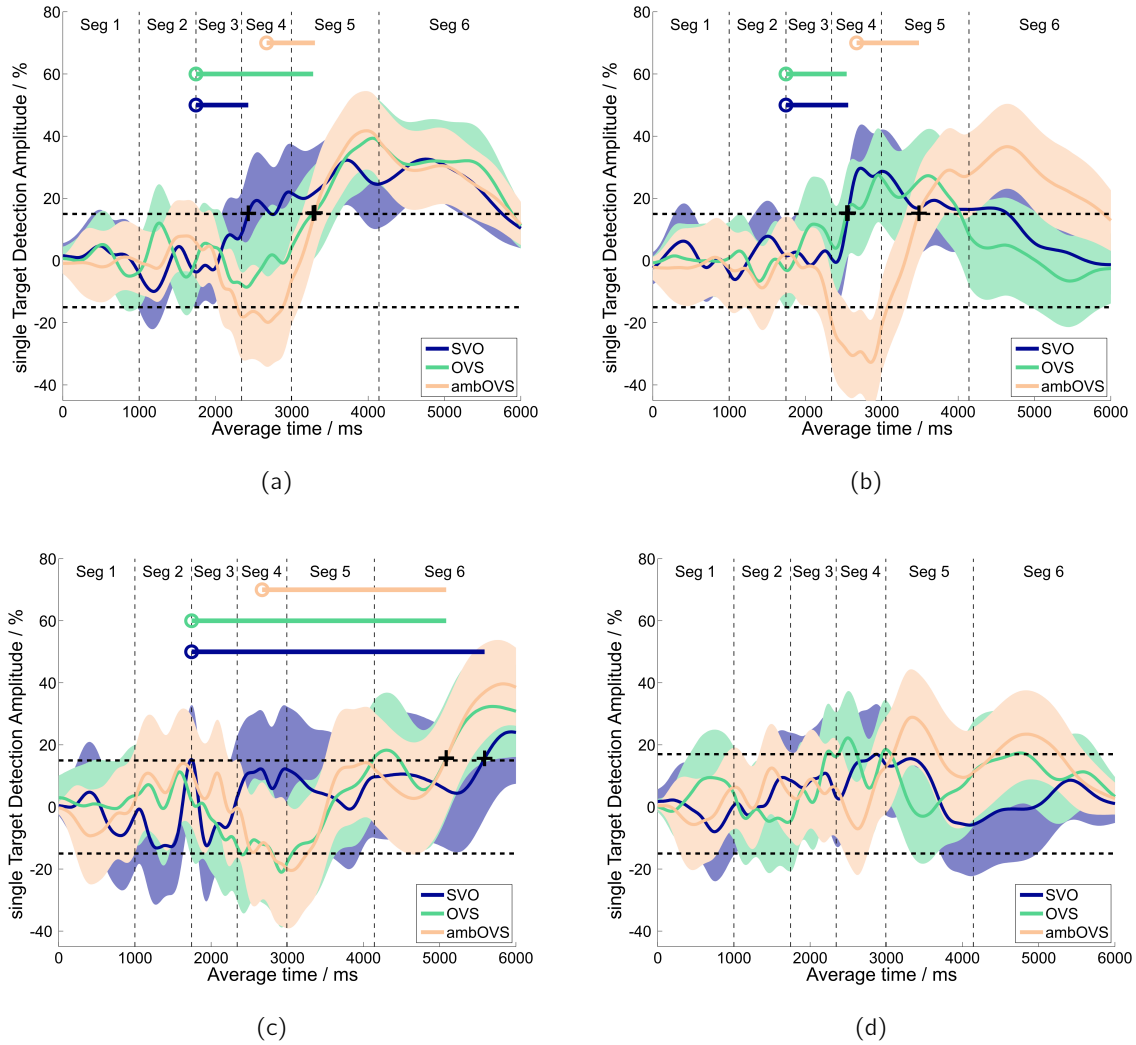

Figure 4.4: Examples of sTDAs of different participants (panels (a)-(d)) for the three sentence structures. The shaded areas illustrate the 95 % confidence interval for each individual curve. The circles denote the PTD, which describes the onset of the word that allows an assignment of the spoken sentence to the target picture (see also Table 4.1). The plus signs denote the DM where the sTDA first exceeds the threshold (15 % of the sTDA). The line starting from the circle denotes the DDD, i.e. the temporal distance between the PTD and DM.

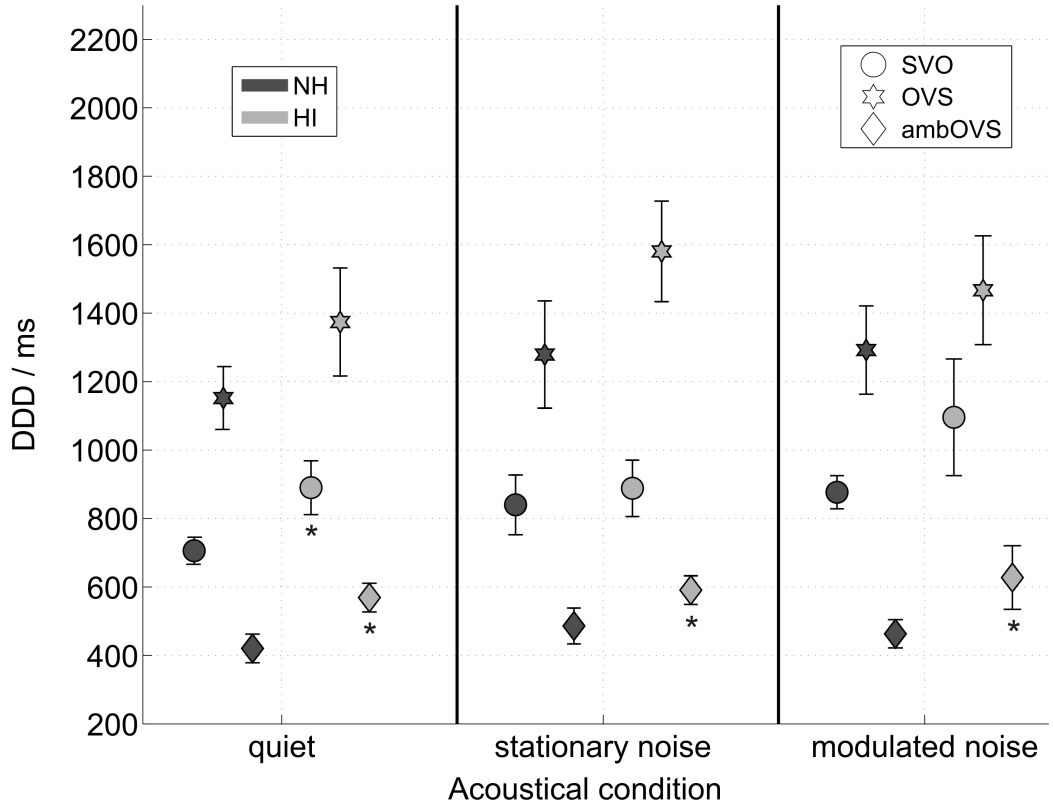

Figure 4.5: Mean DDD (with standard error across participants) for the normally hearing group (dark grey) and the hearing impaired group (light grey) of three sentence structures (SVO, OVS, and ambOVS) in quiet, stationary noise and modulated noise. $*$ denotes significant differences in DDD between both groups ($p < 0.05$) for the sentence structure in the acoustical conditions.

groups). Since in Chapter 2 a decrease in processing speed for the OVS structure in quiet was already observed and commented on a DDD of almost 1000 ms for a group of younger listeners with normal hearing, the results of the current study are in line with this expected decrease in processing speed due to a more complex sentence structure.

As already reported in Chapter 2 and 3 a comparably small DDD was observed for the ambOVS structure. It is argued that the decision process for the ambOVS structure differed from that of the unambiguous structures. Negative sTDA values occurred for the ambOVS structure, indicating that participants fixated more frequently towards the competitor picture before the PTD was reached. This effect was interpreted as a temporal misinterpretation of the spoken sentence, which was further assumed to cause a different decision process after the PTD; participants already had chosen the wrong pictures and they just had to revise their first decision by fixating the other picture. In contrast, no premature decision was made for the unambiguous structures, so participants had to choose between two possible alternative pictures, resulting in a greater DDD. Having a closer look at the individual data in Figure 4.4, negative sTDAs, indicating a temporal misinterpretation, can be observed. Thus, the comparably small DDD for the ambOVS structure in the current study may be explained by this effect.

Effect of background noise

According to the findings of Chapter 3, a decrease in processing speed was expected for the more demanding listening conditions, i.e. for complex sentence structures presented in background noise. It was commented on a decrease in processing speed at a fixed signal-to-noise level (at averaged 80 % speech intelligibility) for a group of normally hearing listeners in Chapter 3. In the current chapter, paired tests were conducted for each sentence structure to analyze the effect of acoustical condition. The tests showed significant effect of acoustical condition for the SVO structure indicating higher DDD in noise (in modulated noise: $p = 0.001$), for the OVS structure (in stationary noise: $p = 0.002$ and modulated noise: $p = 0.05$), and for the ambOVS structure (in stationary noise: $p = 0.01$).

Effect of hearing impairment

Unpaired tests were applied to investigate whether DDDs varied between the two groups (NH and HI group). Significant differences between the two groups were found for the ambOVS structure in quiet ($p = 0.001$), in stationary noise ($p = 0.02$), and in modulated noise ($p = 0.05$). Furthermore, a higher DDD for the HI group was measured for the SVO structure in quiet ($p = 0.04$). These results support the idea that the HI group has a stronger decrease in processing speed, particularly when listening becomes more demanding (i.e. for more complex sentence structures presented in quiet and in background noise). No significant differences in the cognitive measures were detected between the two groups; it is thus reasonable to conclude that the greater decrease in processing speed observed for the HI group did not result from differences in cognitive abilities between the groups.

Effect for hearing aid use

Since the results indicated that some listeners from the HI group did not perform as well as others, the HI group was divided into two subgroups. One group consisted of eleven participants (average age of 66 years; see Table 4.5) who used hearing aids in their daily life (HA group) and the other group consisted of nine participants (average age of 64 years) who did not use hearing aids (noHA group) in their daily life. Figure 4.6 depicts the averaged disambiguation to decision delay (DDD) of the different sentence structures in quiet and in the two noise conditions for both groups.
Again, unpaired tests (alpha level of significance set at $p = 0.05$, adjusted for FDR correction) were applied using the resamped data to analyze whether parameters varied between the two groups (HA and noHA). The test revealed significantly higher DDDs for the noHA group for the SVO structure (in quiet: $p = 0.01$ and in stationary noise: $p = 0.03$) and for the OVS structure (in quiet: $p = 0.05$, in stationary noise: $p = 0.01$, and in modulated noise: $p = 0.001$), reflecting a stronger decrease in processing speed for the noHA group compared to the HA group. Note

Table 4.5: Participants of the HI group with their age (second column), the pure tone average (PTA) thresholds across the frequencies ranging from 125 Hz to 4000 Hz (third column). Participants, which do not used hearing aids in their daily live are highlighted with the grey lines. For the other participants it is shown how long they do already used hearing aids (fourth column). Note that two participant were not considered for the statistical analysis due to a small single target detection amplitude (sTDA).

Participant	Age / years	PTA / HL	use of HA / years
HI_1	42	49	9
HI_2	72	43	7
HI_3	68	41	4
HI_4	57	41	13
HI_5	69	41	3
HI_6	59	33	2
HI_7	71	42	2
HI_8	74	47	14
HI_9	70	46	11
HI_10	68	33	1
HI_11	77	30	7
HI_12	59	40	-
HI_13	61	40	-
HI_14	62	35	-
HI_15	62	33	-
HI_16	65	33	-
HI_17	66	51	-
HI_18	67	47	-
HI_19	69	31	-
HI_20	69	35	-

that no significant differences in the hearing thresholds (tested for each audiometric standard frequency between 125 Hz and 8000 Hz) or in cognitive measures (stroop measure, digit-score, word-score) between the two groups were found. Hence, the smaller decrease in processing speed observed for the HA group indicates a smaller processing effort during sentence recognition for the group that uses hearing aids in daily life.

4.5.3 Cognitive measures

The cognitive abilities of the participants were assessed to account for any effects of individual cognitive abilities on sentence processing speed. The results of the cognitive tests are listed in Table 4.2, i.e. the digit-span score, word-span score and the results of the Stroop test (stroop measure) averaged across all participants. On average, lower values were measured for the word span than for the digit span test. Statistical analysis (paired t-test; $p < 0.05$) showed that there were no significant differences between the HI and NH groups for all three tests. No significant

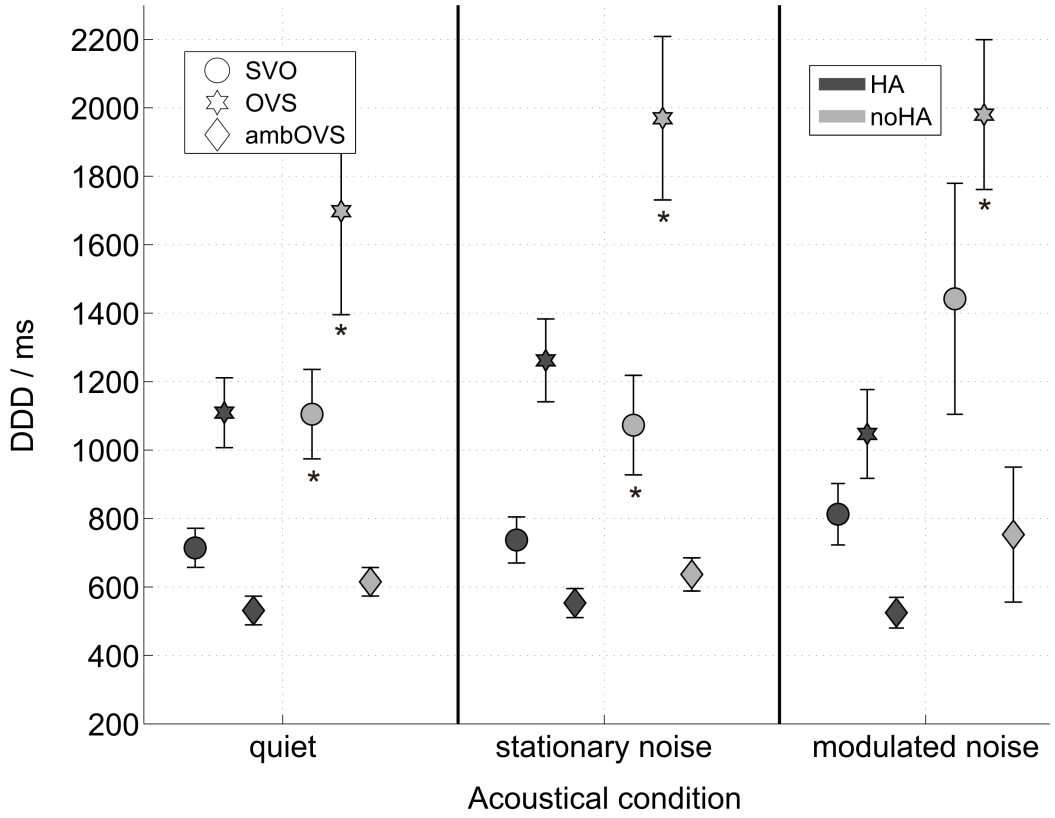

Figure 4.6: Mean DDD (with standard errors across participants) for hearing aid users (HA) group (dark grey) and non-users (noHA) group (light grey) of three sentence structures (SVO, OVS, and ambOVS) in quiet, stationary noise and modulated noise. $*$ denotes significant differences in DDD between both groups ($p < 0.05$) for the sentence structure.

differences in the cognitive tests suggest that the two groups did not differ significantly in their cognitive processing, at least in the tested abilities. Note that also no significant differences between the HA and the noHA group occurred.

To determine whether individual differences in processing speed correlated with cognitive measures, rank correlations between the disambiguation to decision delays (DDDs) of the different sentence structures and the cognitive measures (stroop measure, digit-score, word-score) were calculated according to Spearman. Correlation coefficients were examined separately for the two groups in quiet and in the two noise conditions.

In quiet, the averaged processing speed of the NH group showed correlations with the stroop measure for the complex sentence structure ambOVS ($r = 0.58$, $p = 0.01$).

In stationary noise, correlations were measured between the processing speed of the NH group and the stroop measure (OVS structure: $r = 0.69$, $p = 0.001$). In contrast, processing speed of the HI group correlated with word-span (OVS structure: $r = -0.53$, $p = 0.01$).

In modulated noise, correlation coefficients showed significant correlations between the stroop measure and the DDDs of the NH group (for the OVS structure: $r = 0.61$, $p = 0.007$). For the HI group, DDDs correlated with the digit-span (SVO structure: $r = -0.56$, $p = 0.01$, ambOVS

structure: $r = -0.47$, $p = 0.04$) and with the word-span (ambOVS structure: $r = -0.47$, $p = 0.006$).
Significant correlations between the DDDs and cognitive measures suggest that processing speed was influenced by individual cognitive abilities. The interpretations of the reported correlations between processing speed and cognitive functions observed for the two groups are further discussed in the following section.

4.6 General discussion

The current study focuses on an audiological viewpoint when investigating the deceleration in sentence processing as an indicator of the required extra processing effort caused by hearing loss. It was tested whether a deceleration in sentence processing can be measured for normally hearing and hearing impaired listeners on the individual level by applying an audio-visual paradigm.
Our results are clear in showing that sentence processing can decelerate due to hearing loss, and therefore support our hypothesis of a more effortful processing due to hearing impairment (hypothesis 1). A reduction in processing speed caused by hearing impairment was indicated by significantly smaller disambiguation to decision delays (DDDs) in some listening conditions even though both groups were measured at the same level of speech intelligibility. Notable differences in processing speed between the NH group and the HI group were mainly found for sentences with a higher level of linguistic complexity. Since the two groups did not differ significantly in their cognitive abilities, the observed averaged differences between the two groups are interpreted as the impact of the hearing impairment. The decreased processing speed suggests that the HI group required more processing effort for sentence understanding due to hearing impairment, even when controlling for speech intelligibility. Note that the NAL-R algorithm was applied to ensure that the participants had roughly the same spectral information, but this did not restore the original speech signal as perceived for normally hearing listeners.
The results of the current study are in line with several former studies that reported that hearing loss affects not only speech intelligibility, but also processing speed, accuracy in speech comprehension tasks, and rated effort (Wingfield *et al.*, 2006, Tun *et al.*, 2010a, Zekveld *et al.*, 2011). For instance, Larsby *et al.* (2005) analyzed subjectively rated effort in sentence recognition for elderly normally hearing and hearing impaired listeners and reported higher rated effort due to hearing impairment at a fixed SNR. In contrast, the current study provided a more objective measure of processing effort and demonstrated that even when the SNR was adapted to the individual SRT80 to take into account individual differences in intelligibility, longer speech processing times (indicating higher processing effort) were measured for hearing impaired listeners. Hence, the results of the current study support this hypothesis of a more effortful listening due to hearing impairment even at a constant level of speech intelligibility (hypothesis 1). In addition to the eye fixation data, the reduced picture recognition rates and the tendency of the HI group to show longer response times in all conditions (although these differences in comparison to the NH group

were not statistically significant) might give a further indication for a more effortful processing for the HI group.

It was hypothesized that extra effort caused by hearing impairment is strongest in adverse listening conditions based cognitive demands, i.e. processing complex sentences in background noise (hypothesis 2). This hypothesis is based on results of Chapter 3, where a compounded effect of linguistic complexity and noise on the processing speed depending on the noise type was reported. The TDA was only calculated for a group of younger listeners with normal hearing and speed of processing was measured for sentence presented at fixed signal to noise rations. An effect of noise on processing time was observed in the current study, since higher processing times were measured in noise compared to quiet. However, the expected compounded effect of noise and complexity could not be found, since higher processing times were measured for all three sentence structures in noise. The rates of correctly identified target pictures already indicate that participants performed well in noisy conditions (picture recognition rate was above 90 %), although they were tested at their individual 80 % speech intelligibility threshold (at the individual SRT80 without pictures). The SRT80 was determined by using word scoring, which means 80 % of the spoken words were recognized in noise. However, for the assignment of the target picture, chance level is already at 50 % and the recognition of each word is not necessary. For instance, the adjectives are not necessarily required for correct identification of the target picture. Furthermore, the additional presentation of the visual stimulus presumably facilitates acoustical word recognition. As a consequence, sensory demands caused by background noise at the individual intelligibility level at 80 % are too small to investigate the compounded effect of linguistic complexity and noise on speed of sentence processing within the audio-visual paradigm.

4.6.1 Correlations between processing effort and cognitive factors

Individual differences that cannot be explained by hearing impairment may be caused by individual differences in cognitive abilities. It was hypothesized that a deceleration in sentence processing is related to cognitive abilities that are linked to speech perception (hypothesis 3). Significant correlations between cognitive measures and processing speed support this assumption. The findings of the current study suggest that for the NH group, processing speed of complex sentence structures in noise correlates with the measured reaction time in the Stroop test. The stroop measure is interpreted as the susceptibility of the listeners to interferences. However, for the HI group correlations between the span measures (digit-span and word-span) and the speed of sentence processing were found in both noise conditions, reflecting that processing speed of this sentence structure was affected by the working memory capacity and the listeners' ability to store and manipulate the speech signal during processing. This is in line with Akeroyd (2008), which concluded that attention and working memory can explain at least parts of the variance in speech intelligibility measurements. More specifically, the digit span test was argued to be related to the

cognitive resources involved in processing complex sentences (e.g. Humes *et al.*, 2006).
A more theoretical framework for the ease of language understanding was proposed by Rönnberg (2003). They suggested a model for predicting the working memory system in the context of speech understanding. The model assumes that the probability of a mismatch between the incoming speech and the corresponding memory representation increases in challenging listening conditions; thus processing becomes more demanding, which is termed explicit processing within the model (Rönnberg, 2003, 2008, 2010). As a result, the process of speech understanding is affected by the working memory capacity in effortful listening conditions. Moreover, the model assumes that this effortful processing can be time-consuming. In the current study, the observed correlations between processing speed and working memory capacity support this theory of explicit processing and the assumption that speed of processing complex sentence structures is a function of the working memory capacity for the HI group. Note that, although significant correlations between the span measure and processing speed were found, the HI group showed somewhat smaller correlation coefficient between cognitive measured and the DDDs than for the NH group. However, regarding the variation in hearing thresholds for listeners with hearing loss, it seems reasonable to assume that the reduced correlation coefficients is due to variations in hearing status.
In general, the results are in line with the hypothesized correlations between speed of sentence processing and individual cognitive abilities (hypothesis 3). In particular in noise, listeners' susceptibility to interferences partly explained interindividual differences in processing speed for the NH group, which is indicated by high correlation coefficients between the DDD measures and the stroop measure (correlation coefficient up to 0.69). In contrast to the NH group, processing speed in modulated noise correlates with a measure for working memory capacity and the listeners' ability to store and manipulate the speech signal during processing for the HI group. The observed differences between both groups might indicate different processing strategies in noise. For instance, whereas normally hearing listeners' speed of sentence processing might be affected by a measure of selective attention and by listeners' capability to filter relevant speech information out of the modulated noise within the proposed paradigm, hearing impaired listeners' speed of sentence processing is more affected by their working memory capacity. However, this is more a speculated interpretation of the observed results, which needs further investigations.

4.6.2 Does hearing aid use ameliorate the specific deceleration effect of hearing impairment?

When the HI group was divided into hearing aid users (HA group) and non-users (noHA group), processing differences between the two groups were found with the current method (see Figure 4.6). A stronger reduction in processing speed was observed for the noHA group, although the two groups did not differ significantly in their hearing thresholds (verified by the pure tone audiogram), in their age or in their respective cognitive measures. Since the speech signal was filtered and

amplified according to the NAL-R formula to ensure the same speech information transmission independent from the individual hearing status, hearing aid users may have been more familiar with this modified and processed speech sound. As a consequence, less effort may have been needed for the processing of the adjusted speech material. In contrast, the reduced processing speed found in non-users may be due to an increased processing effort. Hence, the differences in processing speed between these two groups underline Rönnberg's theory of explicit processing. These results might go in the same direction of recent studies testing aided speech recognition of hearing aid users with new or unfamiliar signal processing algorithms (Foo *et al.*, 2007, Rudner *et al.*, 2009). It was reported that in particular for these unfamiliar processed speech signals, processing becomes more effortful to reach a certain speech recognition performance. An alternative interpretation is that hearing impaired non-users might have lost auditory processing capacity due to a lack of stimulation of parts of their auditory system. This interpretation bases on the hypothesis, that their reduced processing speed is independent from the level and spectral shaping of the speech material employed. To test this hypothesis, a comparison across (simulated) aided and unaided acoustical conditions with hearing-aid users of different degrees of hearing aid acclimatization would be desirable which is clearly beyond the scope of the current study.
In any case, the possibility of individually assessing the processing speed of hearing impaired listeners as demonstrated in the current study might be a valuable tool for the individualization of hearing aid fitting based on the individual parameters of cognitive abilities and processing effort of the patients. So far, subjectively rated efforts were used to test individual processing effort regarding speech perception with several hearing aid settings. For instance, a recent study showed that hearing aid compression settings influenced subjectively rated effort involved in listening to speech in noise (Brons *et al.*, 2013). The proposed objective measure of processing speed may be used for the design, selection and fitting of hearing devices to the individual listener that are adapted to the individual processing speed and/or the individual processing effort in perceiving speech in acoustical *difficult* situations.

4.7 Conclusions

The eye-tracking approach presented here appears to provide a useful tool for characterizing individual cognitive processing effort during sentence processing in a reliable and objective way. From a comparison across (roughly) aged-matched normally hearing (average age of 59 years) and hearing impaired listeners (average age of 65 years), across listening conditions (quiet, stationary, and modulated noise) and across cognitive demand (3 sentence structures with increasing linguistic complexity), the following conclusions can be drawn:

1. Although speech intelligibility was similar for age-matched hearing impaired and normally hearing listeners, hearing impairment can lead to a specific and significant reduction in processing speed (supports hypothesis 1).

2. The eye-tracking data of the hearing impaired group indicates that those listeners who did not use hearing aids in their daily life showed the strongest deceleration in processing speed. It is unclear if this is due to a missing acclimatization of the non-users to the spectral shaping and amplification of the speech signals employed here or due to a loss in auditory processing capacity resulting from a lack of auditory stimulation.

3. A deceleration in processing speed caused by hearing impairment was measured in particular for sentence structures with a higher level of linguistic complexity (supports hypothesis 2). However, no compounded effect of background noise and linguistic complexity on processing speed was detected.

4. Inter-individual variance of processing speed for complex sentence structures in noise appears to be associated with cognitive factors, such as working memory capacity and participant's susceptibility to interference (supports hypothesis 3).

In general, the eye-tracking approach presented here has been demonstrated to detect differences in processing effort caused by hearing impairment on an individual basis. This opens a wide range of applications in audiology both for diagnostical purposes (sensorineural vs. central factors in hearing impairment) and for the design, selection and fitting of hearing devices to the individual listener that are adapted to the individual processing speed.

5

Summary and concluding remarks

5.1 Summary

The primary focus of this work was to investigate the effect of external and internal factors on the process of speech understanding using a novel audio-visual approach. For that purpose, an audio-visual paradigm was developed that realizes an online analysis of the speech understanding process. More specifically, recorded eye fixations were transformed into the target detection amplitude (TDA), which can be used to visualize and analyze the time course of the complex process of speech understanding. With the help of this paradigm, the speed of sentence processing was examined in order to reveal difficulties during speech understanding. The most important findings of this thesis can be summarized as follows:

1. The decision moment (DM) and the corresponding disambiguation to decision delay (DDD) were introduced. Both the DM and DDD were shown to be effective measures of processing speed in audio-visual speech understanding (Chapter 2). The DM was defined as that point in time at which eye fixations towards the target picture exceed a critical threshold, and therefore denotes the moment at which the participant correctly recognizes the target picture. The DDD was defined as the temporal distance between the DM and the first point in time at which the target picture can theoretically be recognized given the grammar of the spoken sentence. The DM can be measured while a sentence is presented, and therefore allows an online analysis of the speed of sentence processing.

2. The audio-visual paradigm proved able to reveal a temporary misinterpretation of the spoken sentence. Negative TDA values, which were measured for ambiguous sentence structures,

indicate more fixations towards the competitor picture (i.e., the wrong picture). Thus, negative TDAs during sentence processing suggest temporary misunderstanding of the meaning of the sentence (Chapter 2). These results demonstrate the advantage of an online measure: misinterpretations can be detected during the presentation of the stimulus while offline measures might overlook difficulties in sentence understanding since listeners can overcome them before the end of the spoken sentence.

3. Systematically changing the sentence structure demonstrated that processing speed is highly dependent on linguistic complexity. In particular, speech processing is slower when the cognitive demands during sentence processing are higher, even under conditions of high speech intelligibility (Chapter 2). Moreover, this audio-visual paradigm detected different processing strategies for different sentence structures (ambiguous vs. unambiguous sentence structures).

4. Processing speed is also affected by changing the demands at the sensory level. Background noise leads to a reduction in processing speed during sentence understanding. The type of noise masker plays an important role in processing speed, even under conditions of equal speech intelligibility. In stationary noise, a joint, superadditive effect of noise and complexity was indicated by a strong reduction in the speed of processing complex sentence structures compared to quiet conditions. In contrast, in modulated noise a reduction in processing speed was observed even for simple sentence structures. This slower processing in modulated noise is due to characteristics of the background noise rather than to the linguistic complexity of the sentences (Chapter 3). Contrary to expectations, neither stationary noise nor modulated noise had any effect on processing speed for ambiguous sentence structures. This may reflect different processing strategies used for the unambiguous and ambiguous sentence structures within this paradigm.

5. Differences in processing speed were detected even when picture recognition rates were constant (Chapter 3). The picture recognition rate describes the number of correctly recognized target pictures. This rate was constant across several conditions of linguistic complexity: for the subject-verb-object and object-verb subject structures in quiet at 100 % intelligibility and for both structures in modulated noise at 80 % intelligibility. At the same time, TDAs across these conditions varied, indicating that processing speed provides a sensitive measure for detecting difficulties during sentence processing, which cannot be revealed by testing only (picture) recognition performance.

6. Individual difficulties in sentence processing were investigated by calculating the single target detection amplitude (sTDA). The sTDA is a modification of the TDA that allows the analysis of the target recognition process of a single listener. Comparisons between normally hearing and hearing impaired participants revealed a reduced processing speed for the hearing impaired, indicating a specific increase in processing effort due to hearing

impairment (Chapter 4). In addition, hearing impaired listeners that did not use hearing aids in daily life exhibited the highest reduction in sentence processing speed. This indicates an increased effort for this group of listeners when listening to speech that was amplified and filtered in a way which should have compensated for the sensory component of the hearing loss. It was therefore expected that, for hearing aid users who were better acclimatized to this kind of speech processing, the effect on processing speed would not be as great.

7. Correlations were detected between processing speed and cognitive abilities in individual listeners. In particular, for hearing impaired listeners, significant correlations with working memory capacity were found for complex sentence structures in background noise (Chapter 4). These findings indicate the importance of individual cognitive capacities in processing speed during sentence understanding.

5.2 Interpretation of the experimental results

To further understand the reduction in sentence processing speed, the ease of language understanding (ELU) model proposed by Rönnberg and colleagues was used (Rönnberg, 2003, 2008). This conceptual model can be seen as a framework which describes the working memory system during the processing and understanding of speech. The model assumes that for less adverse listening situations, speech processing is *implicit* (as termed by the model authors): no cognitive resources need to be mobilized beyond those required to understand and interpret speech anyway. Moreover, when processing demands increase as a result of background noise or hearing impairment, speech processing can become effortful, which is termed *explicit* processing in the ELU model. This requires seizing additional resources, for instance because missing speech information requires activation of knowledge stored in long-term memory. Thus, the model describes listening effort as an increase in cognitive resources required for speech understanding. Since mapping the speech information to and from long-term memory takes time, this explicit processing is expected to be time consuming. The proposed paradigm is able to uncover this time-consuming process in both cognitively and sensorily demanding situations. Calculating the sTDA enables a direct view into this processing: difficulties in speech understanding caused by time-consuming processes can be detected.

One objective of this thesis was to analyze the effect of external factors, such as background noise and linguistic complexity, on the speed of sentence processing. The influence of sensory (via two different noise types) and cognitive (via changing the level of linguistic complexity) factors on the process of speech understanding was demonstrated. The influence of linguistic complexity can be clearly detected at 100 % speech intelligibility (Chapter 2), because complex sentence structures can cause a significant reduction in sentence processing speed. In contrast, the effect of background noise on complexity is strongly influenced by the noise type (Chapter 3). In addition,

the compounded effect of noise and complexity clearly highlights the difficulties in isolating and determining the impact of background noise.

The second objective of this thesis was to investigate the impact of intrinsic factors, such as hearing loss or cognitive abilities, on the process of speech understanding. Significant correlations between processing speed and cognitive abilities suggested a link to the memory and processing capacity (and the ability to manipulate the content of the working memory system) of the test subjects (Chapter 4). These correlations support the model assumption that this process is limited by cognitive factors, such as the general capacity of processing and storing information (Rönnberg, 2003). Nevertheless, cognitive abilities, such as working memory capacity, can only explain a certain amount of this effect. The results of the current study clearly show that sensory factors like hearing impairment have a major impact on sentence processing speed, indicated by a stronger speed reduction for hearing impaired listeners: if, due to hearing impairment, listeners did not extract the required acoustical cue - the (linguistic) speech information required to identify the target picture - then they had to wait for the next cue. This takes time and leads to a reduction in sentence processing speed, indicated by an increase in the DDD. In addition, Chapter 4 demonstrated that participants with hearing impairment who do not use hearing aids in their daily life, and are therefore not accustomed to amplified and filtered speech signals, process sentences more slowly than hearing impaired listeners who are familiar with processed speech signals. Differences between the two groups were hypothesized to indicate the auditory deprivation effect. Arlinger *et al.* (1996) defined this effect as a systematic decrease over time in auditory performance associated with the reduced availability of acoustic information. That is, this reduction in speed of sentence processing may be explained not only by missing acclimatization in the cognitive domain (in the sense of the ELU theory) but considerably more by reduced auditory processing capacity arising from missing stimulation of parts of the auditory system. In this case, a reduced processing speed would be independent of the level and the spectral adjustment.

In general, the present work demonstrates the advantage of online methods beyond standard speech audiometry, namely their ability to characterize individual speech processing capabilities that relate to sensory factors as well as to the required (cognitive) processing effort. The new method provides a measure of processing speed that seems to be coupled to the effort required during speech processing, even at a high level of speech intelligibility, where standard methods typically fail. This yields the possibility of studying the relative contribution of sensory and cognitive factors and their combined effect on speech processing for individual subjects.

5.3 Suggestions for future research and possible applications of the methodology

The logical next step is to acquire a substantial body of normative data in order to characterize the *standard* responses and to be able to detect, consequently, participants' deviant behavior. So

far, sTDA and the associated measure of processing speed (DDD) were only measured for two groups of roughly age-matched elderly adults with and without hearing impairment; no data were collected from younger participants with normal hearing. Even though the current study design avoids the usual concatenation of age with hearing loss, it is not yet clear whether any effects of age can be measured using the audio-visual paradigm employed here.

On the one hand, there is some evidence that reaction times are identical for younger and elderly subjects with normal hearing. Tun *et al.* (2010a) tested speech comprehension tasks for sentences that differed in terms of linguistic complexity. Using offline measures of reaction time, they found no differences between younger and elderly participants with normal hearing even when the sensory load was expected to increase (due to low sound levels). However, this does not necessarily mean that there were no differences during sentence processing: there may have been processing difficulties that were overcome by the end of the spoken sentence. Wingfield *et al.* (2003), for example, revealed age-related differences in response times at higher levels of linguistic complexity for normally hearing listeners. However, the effect of age was only measured at higher levels of sensory load, realized with increased speech rates. Thus, it is not clear how age affects the speed of sentence processing, and therefore it is important to address this question with the proposed paradigm in future studies.

Using response times in comprehension tasks to study speech processing can reveal differences in processing speed, but it is not clear at what level of processing these differences occur. Caplan and Waters (1999) differentiated between the *interpretative processing level* (the process of extracting the meaning of the spoken sentence) and a *post-interpretative level* (the processes for performing the task such as answering a comprehension question). Waters and Caplan (2001) and Kemper *et al.* (1989), for instance, hypothesized that age-related differences in response to syntactic complexity are due to an age-related disadvantage in post-interpretative processes. The paradigm proposed here may be more appropriate for distinguishing between these two levels than the reaction time studies employed so far: changes in the DDD could potentially reveal effects at the interpretative level, whereas changes in reaction times (see Chapter 4) would reflect effects at the post-interpretative level.

5.3.1 Acclimatization effects and application in hearing aid testing

Chapter 4 revealed new insights into acclimatization in hearing aid use. The reduction in sentence processing speed for the listeners that did not regularly use hearing aids was hypothesized to result from a reduced auditory processing capacity arising from the missing stimulation of parts of the auditory system. To test this hypothesis, further studies are required that focus on testing the auditory acclimatization of this group of listeners, addressing the change in acoustic information available to the listener that leads to a change in auditory performance over time. This acclimatization effect is expected to lead to an increase in sentence processing speed over time.

In addition, the differences observed between the hearing aid users and the non-hearing aid users point towards the fact that the proposed paradigm may be suitable for measuring processing differences caused by unfamiliar speech signals. Some research has already focused on the question of how speech recognition using new hearing aid settings, to which the listener is not accustomed, interacts with cognitive aspects of hearing Foo *et al.* (2007), Rudner *et al.* (2009). Modern hearing aids offer several signal processing technologies for adapting to different environments, depending on the type of hearing impairment. These technologies include, for instance, dynamic range compression and noise reduction. So far, little is known about the effect of these technologies on the effort required for speech understanding. For instance, Brons *et al.* (2013) compared listening effort in speech recognition when testing noise reduction algorithms of different hearing aids. Participants were asked to rate the effort that they perceived was necessary for speech recognition with different hearing aid algorithms: subjective ratings of effort differed depending on the noise reduction system.
In order to test the effectiveness of hearing aids, it is important to have reliable measures of the (cognitive) processing effort in aided listening situations. This requires a measure of the (extra) effort during speech processing while using signal processing algorithm that can be applied easily. Previously, there was no objective way to detect an increase in processing effort caused by a certain signal processing algorithm. The results reported in this thesis suggest that the proposed paradigm may provide an objective measure of processing effort under aided listening conditions. The logical next step would be to analyze whether processing difficulties can be explained as an effect of signal processing, and hence be relieved by getting test subjects acquainted with their hearing aids. In the long term, the proposed paradigm could be used to control parameters of hearing aid signal processing algorithms for individual listeners in order to reduce processing effort in aided listening situations.

5.3.2 Link to further methodologies

Recently, Müller (2013) demonstrated that the proposed analysis of eye fixations can also be applied to electrooculographic (EOG) data. In that study pairs of electrodes were positioned next to the left and the right eyes to measure horizontal eye movements. Differences in the electrical potential between the two electrodes when the eyes moved from the center position towards one electrode enabled detection of eye movements and eye fixations. Müller compared sTDAs calculated from the eye-tracking data with sTDAs calculated from the EOG data and reported high correlations between these two measures. This indicates that the calculation of the sTDA is independent of the system employed for analyzing gaze. Having two equivalent methodologies may facilitate clinical application, since not all medical facilities have access to both techniques. In contrast to this thesis, which tested three acoustic conditions, Müller (2013) detected differences in processing speed using only two acoustic conditions: the proposed paradigm can therefore successfully analyze processing speed for different user groups using far fewer trials,

thereby requiring far less time.

5.4 Conclusions

Overall, this thesis introduces a novel audio-visual eye-tracking paradigm which was developed as an online analysis of the process of speech understanding, with possible applications in the field of audiology. On the one hand, the new paradigm can detect differences in processing speed that cannot be detected by standard speech audiometry. On the other hand, the proposed paradigm was used to test a wide range of factors in this thesis, including external and internal (listener-specific) factors influencing speech processing, in order to obtain a better understanding of the normal and impaired human auditory system. The new paradigm can provide information about the online processing of speech understanding, such as the point in time at which a sentence is understood. This novel opportunity of time-resolved monitoring of sentence understanding and the impact of acoustic conditions and individual prerequisites on this task is therefore certain to contribute to the field of hearing research in a significant way.

Summary

Understanding speech provides the basis for human communication. However, speech understanding is influenced by a variety of external and internal factors. External factors can be attributed to background noise or linguistic complexity of the speech signal while internal factors such as hearing loss or cognitive abilities are specific to the individual listener. The primary goal of this thesis is to gain a better insight into any impediments in speech processing that occur due to external and internal factors. Consequently, a new experimental design is developed here which allows for an online analysis of the speech understanding process with possible applications in the field of audiology. In particular, the aim is to separate sensory and cognitive aspects of speech processing and their respective influence on processing speed. A reduction in processing speed might indicate that listeners are missing sensory information and compensate by deciphering past information which in turn increases the required cognitive resources. Speech comprehension can be made more demanding on the sensory side by adding background noise. Moreover, the cognitive processing effort can be varied by changing the complexity of the sentence structure.

In the first part of this thesis, an experimental setup is developed with an eye-tracking device in order to realize an online measure of processing speed. The appropriate analysis of the recorded eye fixation data includes the computation of the target detection amplitude (TDA), which is further employed to assess the time course of the sentence understanding process. The decision moment (DM) and the corresponding disambiguation to decision delay (DDD) are calculated from the TDA to obtain a measure for processing speed. In the first study, the effect of linguistic complexity on processing speed is examined by using the proposed eye-tracking paradigm. For a systematical variation of syntactic complexity, the Oldenburg Linguistically and Audiologically Controlled Sentences (OLACS) are used. OLACS consists of seven different sentence structures that differ in their linguistic complexity. The experimental results indicate that the proposed eye-tracking paradigm permits an assessment of sentence processing speed on a cognitive level. A reduction in processing speed is observed for a more complex sentence structure, even at a high speech intelligibility level.

In the second part of this thesis, the interaction of external factors, such as sentence complexity and background noise, is investigated. In order to vary the sensory demands, sentences are presented in quiet and with different noise maskers (stationary speech shaped noise and modulated noise).

The results indicate that noise influences sentence processing, whereas the speed of processing strongly depends on the noise type. A compounded effect of noise and sentence complexity is observed in stationary noise, which can be superadditive.

In the third part, the influence of listener-specific (internal) factors is examined. The interaction of processing speed and hearing loss is investigated by measuring processing speed for a group of hearing impaired people and a control group of people with normal hearing roughly matched in age. The idea of the third part is to separate the cognitive aspects from more sensory aspects of speech processing for both groups by measuring the ability to understand speech at a constant level of speech intelligibility. Therefore, the individual SRT80 (the signal-to-noise ratio at which 80 % of the speech signal is correctly recognized by the listener) is determined for every sentence structure and every single listener. Furthermore, the single target detection amplitude (sTDA) is introduced, which allows for the investigation of speech processing for individual listeners. Although measuring at a constant level of speech intelligibility, differences in processing speed between both groups occur due to extra effort in processing for the hearing impaired listeners. In addition, correlation between processing speed and individual parameters provide evidence that the slowing down of processing speech is related to individual cognitive abilities. In particular, hearing impaired listeners that are experienced hearing aid users exhibit a smaller deceleration effect in processing speed than those without acclimatization to a hearing aid - an effect that has to be pursued in further studies.

In summary, the approach presented in this thesis provides new, time-resolved information about the processing of speech understanding and the impact of external and internal factors on this complex process. The methods and results might be useful both in basic research and in audiology.

Zusammenfassung

Das Verstehen von Sprache bietet die Basis für die menschliche Kommunikation und wird von vielen Faktoren beeinflusst. Dazu zählen externe Faktoren, wie z.B. Hintergrundrauschen als auch die linguistische Komplexität des Sprachsignals. Zusätzlich wird unser Sprachverstehen von hörerspezifischen Faktoren (interne Faktoren) beeinflusst, wie z.B. Hörverlust oder kognitive Fähigkeiten. Das wesentliche Ziel dieser Arbeit ist es, auftretende Probleme beim Sprachverstehen, die durch die oben genannten Faktoren entstehen können, genauer zu analysieren. Zu diesem Zweck wird in dieser Arbeit eine Methodik vorgestellt, welche eine zeitaufgelöste Analyse des Sprachverstehensprozesses ermöglicht und somit einen wichtigen Beitrag zur audiologischen Diagnostik bieten könnte. Das wesentliche Ziel bei der Anwendung dieser Methodik ist es, den Einfluss sensorischer und kognitiver Aspekte auf die Verarbeitungsgeschwindigkeit beim Verstehen von Sätzen zu untersuchen. Eine Reduzierung der Verarbeitungsgeschwindigkeit könnte dazu führen, dass der Zuhörer wichtige sensorische Informationen verpasst, da noch vorhergehende Sprachinformationen verarbeitet werden müssen. Das Verpassen von wichtigen Informationen wiederum führt zu einer erhöhten kognitiven Belastung.

Im ersten Teil dieser Arbeit wird eine Methodik vorgestellt, die auf der Messung der Augenbewegung während des Sprachverstehensprozesses beruht. Mithilfe einer statistischen Analyse der aufgenommenen Daten, welche aus der Berechnung der *target detection amplitude* (TDA) besteht, ist eine zeitaufgelöste Analyse der Verarbeitungsgeschwindigkeit beim Verstehen von Sprache möglich. Mithilfe der TDA kann der *decision moment* (DM) und *disambiguation to decision delay* (DDD), als Maß für die Verarbeitungsgeschwindigkeit, bestimmt werden. In einer ersten Studie wird mithilfe dieser Methodik der Einfluss von linguistischer Komplexität auf die Geschwindigkeit beim Verstehen von Sätzen systematisch überprüft. Um eine systematische Untersuchung zu ermöglichen, wurden die Oldenburger Linguistisch und Audiologisch Kontrollierten Sätze (OLAKS) verwendet. Dieses Sprachmaterial setzt sich aus insgesamt sieben verschiedenen Satztypen zusammen, welche sich in ihrer syntaktischen Komplexität unterscheiden. Zu diesen Sätzen zählen z.B. eingebettete Relativsätze und nicht eingebettet Hauptsätze. Durch die Analyse der Augenbewegung kann, trotz hoher Sprachverständlichkeit, eine Reduzierung der Verarbeitungsgeschwindigkeit beim Verstehen des Satzes nachgewiesen werden.

Im zweiten Teil dieser Arbeit werden der Einfluss von Störgeräusch und die Interaktion von syntak-

tischer Komplexität und Störgeräusch auf den Verstehensprozess untersucht. Die vorgestellten Ergebnisse zeigen, dass das Störgeräusch zu einer Reduzierung der Verarbeitungsgeschwindigkeit führt. Zusätzlich ist im stationären Rauschen eine Wechselwirkung von Störgeräusch und Komplexität zu beobachten, die sich in einer Superadditivtät von Rauschen und Komplexität in der Verarbeitungsgeschwindigkeit äußert.

Im dritten Teil dieser Arbeit wird der Einfluss von internen Faktoren analysiert. Zum einen wird der Einfluss von Schwerhörigkeit genauer untersucht, indem für eine Gruppe von Normalhörenden und eine Gruppe von Schwerhörenden die Verarbeitungsgeschwindigkeiten mithilfe der vorgestellten Methodik für verschiedene OLAKS Satztypen in Ruhe und im Störgeräusch gemessen und verglichen wird. Die Idee der dritten Studie ist es, kognitive von sensorischen Aspekten beim Sprachverstehen zu trennen. Daher wird jeder Proband bei gleicher Sprachverständlichkeit gemessen. Zu diesem Zweck wird die im ersten Teil beschriebene Methodik modifiziert um eine individuelle Bestimmung der Verarbeitungsgeschwindigkeit jedes Probanden beim Verstehen von Sätzen zu ermöglichen. Die Ergebnisse zeigen, dass bei gleicher Sprachverständlichkeit eine Reduzierung der Geschwindigkeit beim Satzverstehen für Schwerhörende gemessen werden kann. Ein besonders großer Effekt zeigt sich dabei für Schwerhörende, die bisher kein Hörgerät im Alltag tragen. Um den Einfluss individueller kognitiver Fähigkeiten, als weiteren internen Faktor, auf die Verarbeitungsgeschwindigkeiten zu untersuchen, werden zusätzlich kognitive Maße erhoben. Es kann gezeigt werden, dass das entwickelte Maß zur Bestimmung der Verarbeitungsgeschwindigkeit mit individuellen kognitiven Maßen korreliert.

Zusammenfassend bietet die vorgestellte Methodik neue, zeitaufgelöste Informationen über den Sprachverstehensprozess und zusätzliche Informationen über die Auswirkungen von externen und internen Faktoren auf den komplexen Prozess des Sprachverstehens. Diese Methodik und die neuen Erkenntnisse können daher sowohl für die Grundlagenforschung als auch für die Audiologie nützlich sein.

Bibliography

Akeroyd, M. A. (**2008**), "Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults," *International Journal of Audiology* **47**(2), pp. 53–71. (Cited on pages 3, 60, and 82)

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (**1998**), "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models," *Journal of Memory and Language* **38**(4), pp. 419–439. (Cited on pages 5, 11, and 30)

Altmann, G. T. M. (**1998**), "Ambiguity in sentence processing," *Trends in Cognitive Science* **2**, pp. 146–152. (Cited on pages 16, 40, and 64)

Altmann, G. T. M. and Kamide, Y. (**1999**), "Incremental interpretation at verbs: Restricting the domain of subsequent reference," *Cognition* **73**(3), pp. 247–264. (Cited on pages 5, 11, and 12)

Altmann, G. T. M. and Kamide, Y. (**2007**), "The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing," *Journal of Memory and Language* **57**(4), pp. 502–518. (Cited on pages 5 and 11)

Arlinger, S., Gatehouse, S., Bentler, R. A., Byrne, D., Cox, R. M., Dirks, D. D., Humes, L. E., Neuman, A., Ponton, C., Robinson, K., Silman, S., Summerflied, A. Q., Turner, C. W., Tyler, R. S., and Willott, J. F. (**1996**), "Report of the Eriksholm workshop on auditory deprivation and acclimatization," *Ear and Hearing* **17**(3), pp. 87S–98S. (Cited on page 90)

Arnold, J. E., Fagnano, M., and Tanenhaus, M. K. (**2003**), "Disfluencies signal theee, um, new information," *Journal of Psycholinguistic Research* **32**, pp. 25–36. (Cited on page 12)

Bader, M. and Bayer, J. (**2006**), *Case and linking in language comprehension: Evidence from German*, Springer. (Cited on pages 14 and 29)

Bader, M. and Meng, M. (**1999**), "Subject-object ambiguities in German embedded clauses: An across-the-board comparison," *Journal of Psycholinguistic Research* **28**(2), pp. 121–143. (Cited on pages 16, 26, 40, and 64)

Barr, D. J., Gann, T. M., and Pierce, R. S. (**2011**), "Anticipatory baseline effects and information integration in visual world studies," *Acta psychologica* **137**(2), pp. 201–207. (Cited on page 12)

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, K. M., Reingold, E. M., and Schneider, B. A. (**2011**), "Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements," *Journal of Speech, Language and Hearing Research* **54**(1), pp. 243–262. (Cited on page 22)

Boothroyd, A. and Nittrouer, S. (**1988**), "Mathematical treatment of context effects in phoneme and word recognition," *The Journal of the Acoustical Society of America* **84**(1), pp. 101–114. (Cited on pages 3 and 10)

Brand, T. and Kollmeier, B. (**2002**), "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *Journal of the Acoustical Society of America* **111**(6), pp. 2801–2810. (Cited on page 67)

Brand, T., Uslar, V. N., Wendt, D., and Kollmeier, B. (**2012**), "Recognition rates and linguistic processing: Do we need new measures of speech perception?" in *Proceedings of ISAAR: Speech perception and auditory disorders. 3rd International Symposium on Auditory and audiological Research*, edited by T. Dau, M. L. Jepsen, T. Poulsen, and J. Christensen-Dalsgaard, The Danavox Jubilee Foundation, vol. 34, pp. 45–56. (Cited on page 57)

Bronkhorst, A. W. (**2000**), "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica* **86**, pp. 117–128. (Cited on pages 2 and 53)

Brons, I., Houben, R., and Dreschler, W. A. (**2013**), "Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort," *Ear and Hearing* **34**(1), pp. 29–41. (Cited on pages 31, 84, and 92)

Byrne, D. and Dillon, H. (**1986**), "The national acoustic laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid," *Ear and Hearing* **7**(4), pp. 257–265. (Cited on page 65)

Caplan, D. and Waters, G. S. (**1999**), "Verbal working memory and sentence comprehension," *The Behavioral and Brain Sciences* **22**(1), pp. 77–126. (Cited on page 91)

Carroll, R. (**2012**), "Effects of syntactic complexity and prosody on sentence processing in noise," Ph.D. thesis, Carl-von-Ossietzky University, Oldenburg, Germany. (Cited on page 60)

Carroll, R. and Ruigendijk, E. (**2013**), "The effects of syntactic complexity on processing sentences in noise." *Journal of psycholinguistic research* **42**(2), pp. 139–159. (Cited on pages 16, 28, 36, 39, 48, 53, 60, and 63)

Cerella, J. and Hale, S. (**1994**), "The rise and fall in information-processing rates over the life

span," *Acta psychologica* **86**, pp. 109–197. (Cited on page 11)

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., and Carlson, G. N. (**2002**), "Circumscribing referential domains during real-time language comprehension," *Journal of Memory and Language* **47**(1), pp. 30–49. (Cited on page 12)

Chambers, C. G., Tanenhaus, M. K., and Magnuson, J. S. (**2004**), "Actions and affordances in syntactic ambiguity resolution," *Journal of Experimental Psychology. Learning, Memory, and Cognition* **30**(3), pp. 687–696. (Cited on page 5)

Cheung, H. and Kemper, S. (**1992**), "Competing complexity metrics and adults' production of complex sentences," *Applied Psycholinguistics* **13**, pp. 53–76. (Cited on page 68)

Cooper, R. M. (**1974**), "The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing," *Cognitive Psychology* **6**(1), pp. 84–107. (Cited on pages 5 and 11)

DIN EN389-8 (**2004**), "Akustik-Standard-Bezugspegel für die Kalibrierung audiometrischer Geräte-Teil 8: Äquivalente Bezugs-Schwellenschalldruckpegel für reine Töne und circumaurale Kopfhörer (ISO 389-8:2004) [Acoustics - Reference zero for the calibration of audiometric equipment - Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (ISO 389-8:2004), German version]," *European Committee for Standardization, DIN Deutsches Institut für Normung e.V., Berlin: Beuth* . (Cited on pages 18, 43, and 66)

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**), "ICRA noises: Artificial noise signals wiht speech-like spectral and temporal properties for hearing instrument assessment," *Audiology* **40**, pp. 148–157. (Cited on pages 41 and 65)

Duquestnoy, A. J. (**1983**), "The intelligibility of sentences in quiet and in noise in aged listeners," *Journal of the Acoustical Society of America* **68**, pp. 537–544. (Cited on page 2)

Eberhard, K. and Spivey-Knowlton, M. (**1995**), "Eye movements as a window into real-time spoken language comprehension in natural contexts," *Journal of* **24**(6), pp. 409–436. (Cited on page 11)

Efron, B. and Tibshirani, R. (**1993**), *An introduction to the bootstrap*, Chapman and Hall, New York, NY. (Cited on pages 12, 22, 44, and 71)

Fanselow, G., Lenertovà, D., and Weskott, T. (**2008**), "Studies on the acceptability of object movement to Spec, CP," in *Language, Context & Cognition: The discourse potential of underspecified structures*, edited by A. Steube, De Gruyter, New York, pp. 413–438. (Cited on page 16)

Fastl, H. (**1982**), "Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise," *Hearing Research* **8**, pp. 59–69. (Cited on page 2)

Fastl, H. and Zwicker, E. (**2007**), *Psychoacoustics: Facts and models*, Springer, Berlin Heidelberg. (Cited on page 2)

Festen, J. M. and Plomp, R. (**1990**), "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing." *The Journal of the Acoustical Society of America* **88**(4), pp. 1725–1736. (Cited on page 2)

Foo, C., Rudner, M., Rönnberg, J., and Lunner, T. (**2007**), "Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity," *Journal of American Academy of Audiology* **18**, pp. 618–631. (Cited on pages 84 and 92)

Fraser, S., Gagné, J.-P., Alepins, M., and Dubois, P. (**2010**), "Evaluating the effort expended to understand speech in noise using a dual-task paradigm : The effects of providing visual speech cues," *Journal of Speech, Language and Hearing Research* **53**, pp. 18–33. (Cited on pages 4 and 37)

Fredelake, S., Holube, I., Schlueter, A., and Hansen, M. (**2012**), "Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms." *International Journal of Audiology* **51**(4), pp. 299–308. (Cited on page 31)

Gatehouse, S., Naylor, G., and Elberling, C. (**2003**), "Benefits of hearing aids in relation to the interaction between the user and the enviroment," *International Journal of Audiology* **42**, pp. 77–85. (Cited on page 4)

Gibson, E. (**1998**), "Linguistic complexity: Locality of syntactic dependencies." *Cognition* **68**(1), pp. 1–76. (Cited on page 52)

Gibson, E. (**2000**), "The dependency locality theory: A distance-based theory of linguistic complexity," in *Image, Language, Brain*, edited by A. Marantz, Y. Miyashita, and W. O'Neil, Massachusetts Institute of Technology Press, pp. 95–126. (Cited on pages 16 and 29)

Gordon, P., Hendrick, R., and Levine (**2002**), "Memory load interference in syntactic processing," *Psychological Science* **13**(5), pp. 425–430. (Cited on page 16)

Gorrell, P. (**2000**), "The subject-before-object preference in German clauses," in *German Sentence Processing*, edited by B. Hemforth and L. Konieczny, Kluwer Academic, Dordrecht, pp. 25–63. (Cited on pages 16, 26, 29, 40, 48, and 64)

Gustafsson, H. a. and Arlinger, S. D. (**1994**), "Masking of speech by amplitude-modulated noise." *The Journal of the Acoustical Society of America* **95**(1), pp. 518–529. (Cited on page 2)

Haegeman, L. (**1995**), *The Syntax of Negation*, Cambridge University Press. (Cited on page 40)

Hagerman, B. (**1982**), "Sentences for testing speech intelligibility in noise," *Scandinavian Audiology*

11(2), pp. 79–87. (Cited on pages 10, 37, and 54)

Hällgren, M., Larsby, B., Lyxell, B., and Alinger, S. (**2005**), "Speech understanding in quiet and noise, with and without hearing aids," *International Journal of Audiology* **44**, pp. 574–583. (Cited on page 37)

Haumann, S., Hohmann, V., Meis, M., Herzke, T., Lenarz, T., and Büchner, A. (**2012**), "Indication criteria for cochlear implants and hearing aids: Impact of audiological and non-audiological findings," *Audiology Research* **2**(1), pp. 55–64. (Cited on page 10)

Hicks, C. B. and Tharpe, A. M. (**2002**), "Listening effort and fatigue in school-age children with and without hearing loss," *Journal of Speech, Language, and Hearing Research* **45**, pp. 573–584. (Cited on page 37)

Holube, I., Wesselkamp, M., Dreschler, W. A., and Kollmeier, B. (**1997**), "Speech intelligiblity prediction in hearing-impaired listeners for steady and fluctuating noise," in *Modeling sensorineural hearing loss*, edited by W. Jesteadt, Lawrence Erlbaum Associates, Inc., Dordrecht, pp. 447–459. (Cited on page 2)

Huettig, F. and McQueen, J. M. (**2007**), "The tug of war between phonological, semantic and shape information in language-mediated visual search," *Journal of Memory and Language* **57**(4), pp. 460–482. (Cited on page 12)

Huettig, F., Rommers, J., and Meyer, A. S. (**2011**), "Using the visual world paradigm to study language processing: A review and critical evaluation." *Acta Psychologica* **137**(2), pp. 151–171. (Cited on pages 5 and 11)

Humes, L. E. (**1991**), "Understanding the speech-understanding problems of the hearing impaired," *Journal of the American Academy of Audiology* **2**, pp. 59–70. (Cited on page 3)

Humes, L. E. (**1996**), "Speech understanding in the elderly," *Journal of the American Academy of Audiology* **7**, pp. 161–167. (Cited on page 3)

Humes, L. E. (**2002**), "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," *The Journal of the Acoustical Society of America* **112**, pp. 1112–1132. (Cited on page 3)

Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**), "Auditory measures of selective and divided attention in young and older adults using single-talker competition," *The Journal of the Acoustical Society of America* **120**(5), pp. 2926–2937. (Cited on pages 60 and 83)

Kaiser, E. and Trueswell, J. C. (**2008**), "Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution," *Language and Cognitive Processes* **23**(5), pp. 709–748. (Cited on page 12)

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**), "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *Journal of the Acoustical Society of America* **61**(5), pp. 1337–1351. (Cited on pages 4 and 10)

Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (**2003**), "The time-course of prediction in incremental sentence processing : Evidence from anticipatory eye movements," *Journal of Memory and Language* **49**, pp. 133–156. (Cited on pages 5 and 11)

Kemper, S., Kynette, D., Rash, S., and O'Brien, K. (**1989**), "Life-span changes to adults language: Effects of memory and genre," *Applied Psycholinguistics* **10**, pp. 49–66. (Cited on pages 68 and 91)

Kim, S.-Y., Kim, M.-S., and Chun, M. M. (**2005**), "Concurrent working memory load can reduce distraction," *Proceedings of the National Academy of Sciences of the United States of America* **102**(45), pp. 16524–16529. (Cited on pages 60 and 67)

Knoeferle, P. (**2007**), "Comparing the time-course of processing initially ambiguous and unambiguous German SVO/OVS sentences in depicted events," in *Eye movements: A window on mind and brain*, edited by R. P. G. Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill, Elsevier, Oxford, chap. 4, pp. 517–533. (Cited on pages 11, 12, 19, and 30)

Knoeferle, P. and Crocker, M. W. (**2006**), "The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking," *Cognitive Science* **30**(3), pp. 481–529. (Cited on page 11)

Knoeferle, P. and Crocker, M. W. (**2007**), "The influence of recent scene events on spoken comprehension: Evidence from eye movements," *Journal of Memory and Language* **57**(4), pp. 519–543. (Cited on page 11)

Knoeferle, P., Crocker, M. W., Scheepers, C., and Pickering, M. J. (**2005**), "The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events." *Cognition* **95**(1), pp. 95–127. (Cited on page 11)

Kollmeier, B. (**1990**), "Messmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache," Ph.D. thesis, Fachbereich Physik der Georg-August-Universität Göttingen. (Cited on page 3)

Kollmeier, B. and Wesselkamp, M. (**1997**), "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," *Journal of the Acoustical Society of America* **102**(4), pp. 2412–2421. (Cited on pages 3, 10, and 36)

Laroche, C., Soli, S., Giguere, C., Lagace, J., Vaillancourt, V., and Fortin, M. (**2003**), "An approach to the development of hearing standards for hearing-critical jobs," *Noise and Health* **6**, pp. 17–37. (Cited on page 10)

Larsby, B., Hällgren, M., Lyxell, B., and Arlinger, S. (**2005**), "Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects," *International Journal of Audiology* **44**(3), pp. 131–143. (Cited on pages 38, 54, and 81)

Lunner, T. (**2003**), "Cognitive function in relation to hearing aid use," *International Journal of Audiology* **42**(1), pp. 49–58. (Cited on page 4)

Lunner, T., Rudner, M., and Rönnberg, J. (**2009**), "Cognition and hearing aids," *Scandinavian Journal of Psychology* **50**(5), pp. 395–403. (Cited on page 4)

Lunner, T. and Sundewall-Thorén, E. (**2007**), "Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid," *Journal of the American Academy of Audiology* **18**, pp. 539–552. (Cited on page 4)

MacLeod, A. and Summerflied, Q. (**1990**), "A procedure for measuring audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," *British Journal of Audiology* **24**(1), pp. 29–43. (Cited on page 2)

Marzinzik, M. (**2000**), "Noise reduction schemes for digital hearing aids and their use for the hearing impaired," Ph.D. thesis, Carl-von-Ossietzky University, Oldenburg, Germany. (Cited on page 31)

May, C. P., Hasher, L., and Kane, M. J. (**1999**), "The role of interference in memory span," *Memory and Cognition* **27**, pp. 759–767. (Cited on page 60)

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (**2005**), "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," *The Quarterly Journal of Experimental Psychology* **58**, pp. 37–41. (Cited on page 58)

McMurray, B., Clayards, M., Tanenhaus, M. K., and Aslin, R. N. (**2008**), "Tracking the time course of phonetic cue integration during spoken word recognition." *Psychonomic bulletin & review* **15**(6), pp. 1064–1071. (Cited on pages 12, 45, and 72)

Middelweerd, M. J. and Plomp, R. (**1987**), "The effect of speechreading on the speech reception threshold of sentences in noise," *Journal of the Acoustical Society of America* **82**(5), pp. 2145–2147. (Cited on page 2)

Miller, G. A., Heise, G. A., and Lighten, W. (**1951**), "The intelligibility of speech as a function of the context of the test materials," *Journal of Experimental Psychology* **41**(5), pp. 329–335. (Cited on page 3)

Müller, J. (**2013**), "Analyse des Sprachverstehen in fluktuierendem Strögeräusch mithilfe von Blickbewegungsmessungen durch Eye-Tracking und Elektrookulografie," Master's thesis, University of Oldenurg. (Cited on pages 54 and 92)

Nichols, T. E. and Holmes, A. P. (**2001**), "Nonparametric permutation tests for functional neuroimaging: A primer with examples," *Human Brain Mapping* **15**(1), pp. 1–25. (Cited on page 76)

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**), "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *The Journal of the Acoustical Society of America* **95**(2), pp. 1085–1099. (Cited on page 10)

Ozimek, E., Warzybok, A., and Kutzner, D. (**2010**), "Polish sentence matrix test for speech intelligibility measurement in noise." *International Journal of Audiology* **49**(6), pp. 444–454. (Cited on page 10)

Pichora-Fuller, K. M. (**2003**), "Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension," *International Journal of Audiology* **42**, pp. S59–S67. (Cited on pages 36, 58, and 59)

Pichora-Fuller, K. M. (**2008**), "Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing," *International Journal of Audiology* **47**(2), pp. 72–82. (Cited on pages 4 and 60)

Pichora-Fuller, K. M., Schneider, B. A., and Daneman, M. (**1995**), "How young and old adults listen to and remember speech in noise," *The Journal of the Acoustical Society of America* **97**(1), pp. 593–608. (Cited on pages 4, 37, and 60)

Piquado, T., Isaacowitz, D., and Wingfield, A. (**2010**), "Pupillometry as a measure of cognitive effort in younger and older adults," *Psychophysiology* **47**(3), pp. 560–569. (Cited on page 30)

Plomp, R. and Mimpen, A. M. (**1979**), "Improving the reliability of testing the speech reception threshold for sentences," *Audiology* **18**, pp. 43–52. (Cited on pages 3 and 10)

Pratt, J., Dodd, M., and Welsh, T. (**2006**), "Growing older does not always mean moving slower: Examining aging and the saccadic motor system." *Journal of Motor Behavior* **38**(5), pp. 373–82. (Cited on page 11)

Rabbitt, P. M. A. (**1968**), "Channel-capacity, intelligibility and immediate memory," *Quarterly Journal of Experimental Psychology* **20**, pp. 241–248. (Cited on page 37)

Radford, A. (**1997**), *Syntactic theory and the structure of English*, Cambridge University Press. (Cited on page 40)

Rhebergen, K. S. and Versfeld, N. J. (**2005**), "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *Journal of the Acoustical Society of America* **117**(4), pp. 2181–2192. (Cited on page 3)

Rönnberg, J. (**2003**), "Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model," *International Journal of Audiology* **42**(1), pp. 68–76. (Cited on pages 37, 51, 60, 83, 89, and 90)

Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (**2008**), "Cognition counts: A working memory system for ease of language understanding (ELU)," *International Journal of Audiology* **47**(2), pp. 99–105. (Cited on pages 36, 37, 51, 60, 83, and 89)

Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. A. (**2010**), "When cognition kicks in: Working memory and speech understanding in noise," *Noise and Health* **12**(49), pp. 263–269. (Cited on pages 60 and 83)

Rudner, M., Foo, C., Rönnberg, J., and Lunner, T. (**2009**), "Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids," *Scandinavian Journal of Psychology* **50**(5), pp. 405–418. (Cited on pages 84 and 92)

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., and Rönnberg, J. (**2012**), "Working memory capacity may influence perceived effort during aided speech recognition in noise," *Journal of the American Academy of Audiology* **23**(8), pp. 577–589. (Cited on pages 37, 38, and 54)

Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (**2009**), "Objective measures of listening effort: Effects of background noise and noise reduction," *Journal of Speech, Language and Hearing Research* **52**, pp. 1230–1240. (Cited on page 31)

Schlueter, A., Lemke, U., Kollmeier, B., and Holube, I. (**2014**), "Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms." *The Journal of the Acoustical Society of America* **135**(3), pp. 1541–1555. (Cited on page 32)

Schneider, B. A., Daneman, M., and Pichora-Fuller, K. M. (**2002**), "Listening in aging adults: From discourse comprehension to psychoacoustics," *Canadian Journal of Experimental Psychology* **56**(3), pp. 139–152. (Cited on page 60)

Schneider, B. A., Pichora-Fuller, K. M., and Daneman, M. (**2010**), "Effects of senescent changes in audition and cognition on spoken language comprehension," in *The Aging Auditory System*, edited by S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay, Springer, New York, NY, vol. 34 of *Springer Handbook of Auditory Research*, pp. 39–74. (Cited on pages 4 and 58)

Sherbecoe, R. L. and Studebaker, G. A. (**2004**), "Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units," *International Journal of Audiology* , pp. 442–448. (Cited on page 23)

Snedeker, J. and Trueswell, J. C. (**2004**), "The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing." *Cognitive*

Psychology **49**(3), pp. 238–299. (Cited on pages 11 and 12)

Sotscheck, J. (**1985**), "Sprachverständlichkeit bei additiven Störungen," *Acustica* **57**, pp. 258–267. (Cited on page 2)

Surprenant, A. M. (**1999**), "The effect of noise on memory for spoken syllables," *International Journal of Psychology* **34**(5-6), pp. 328–333. (Cited on page 4)

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (**1995**), "Integration of visual and linguistic information in spoken language comprehension," *Science* **268**(5217), pp. 1632–1634. (Cited on pages 5, 11, and 30)

Tewes, U. (**1991**), *Hamburg-Wechsler-Intelligenztest für Erwachsene - Revision 1991*, Bern, Stuttgart, Toronto: Huber. (Cited on pages 60 and 68)

Toscano, J. C. and McMurray, B. (**2012**), "Cue-integration and context effects in speech: evidence against speaking-rate normalization." *Attention, Perception & Psychophysics* **74**(6), pp. 1284–1301. (Cited on page 12)

Trueswell, J. C., Tanenhaus, M. K., and Garnsey, S. M. (**1994**), "Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution," *Journal of Memory and Language* **33**, pp. 285–318. (Cited on page 5)

Tun, P. A., Benichov, J., and Wingfield, A. (**2010**a), "Response latencies in auditory sentence comprehension: Effects of linguistic versus perceptual challenge," *Psychology and Aging* **25**(3), pp. 730–735. (Cited on pages 4, 10, 11, 16, 30, 37, 52, 81, and 91)

Tun, P. A., McCoy, S., and Wingfield, A. (**2010**b), "Aging, hearing acuity, and the attentional costs of effortful listening," *Psychology and Aging* **24**(3), pp. 761–766. (Cited on page 59)

Uslar, V. N., Brand, T., Hanke, M., Carroll, R., Ruigendijk, E., Hamann, C., and Kollmeier, B. (**2010**), "Does sentence complexity interfere with intelligibility in noise? Evaluation of the oldenburg linguistically and audiologically controlled sentence test, (OLACS)." in *Proceedings of Interspeech, Makuhari, Chiba, Japan*. (Cited on page 36)

Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., and Kollmeier, B. (**2013**a), "Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test," *The Journal of the Acoustical Society of America* **134**(4), pp. 3039–3056. (Cited on pages 10, 13, 24, 36, 39, 41, 42, 48, 59, 60, 63, 67, and 73)

Uslar, V. N., Carroll, R., Wendt, D., Ruigendijk, E., and Brand, T. (**2013**b), "Warum die Ente der Hund tadelt: Mögliche neue Wege in der Audiologie mit den Oldenburger Linguistisch und Audiologisch Kontrollierten Sätzen," *Zeitschrift für Audiologie* **51**(1), pp. 6–15. (Cited on page 57)

Uslar, V. N., Ruigendijk, E., Hamann, C., Brand, T., and Kollmeier, B. (**2011**), "How does linguistic complexity influence intelligibility in a German audiometric sentence intelligibility test?" *International Journal of Audiology* **50**(9), pp. 621–631. (Cited on pages 3, 10, and 36)

van Rooij, J. C. G. M. and Plomp, R. (**1992**), "Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion," *Journal of the Acoustical Society of America* **91**(2), pp. 1028–1033. (Cited on page 3)

van Zandt, T. (**2002**), "Analysis of response time distributions," in *Stevens handbook of experimental psychology: Vol.4. Methodology in experimental psychology*, edited by J. Wixted, Wiley, New York, pp. 461–516. (Cited on pages 12, 22, 44, and 71)

Wagener, K. C. and Brand, T. (**2005**), "Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters," *International Journal of Audiology* **44**(3), pp. 144–156. (Cited on page 2)

Wagener, K. C., Brand, T., and Kollmeier, B. (**2006**), "The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners," *International Journal of Audiology* **45**(1), pp. 26–33. (Cited on pages 41, 42, 53, and 65)

Waters, G. S. and Caplan, D. (**2001**), "Age, working memory, and on-line syntactic processing in sentence comprehension," *Psychological Aging* **16**, pp. 128–144. (Cited on page 91)

Wendt, D., Brand, T., and Kollmeier, B. (**2014**), "An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities," *PLoS ONE* **9**(6), pp. e100186. (Cited on page 9)

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., and Cox, C. L. (**2006**), "Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity," *Journal of the American Academy of Audiology* **17**(7), pp. 487–497. (Cited on pages 10, 16, 30, 37, 58, and 81)

Wingfield, A., Peelle, J. E., and Grossman, M. (**2003**), "Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults," *Aging Neuropsychology and Cognition* **10**, pp. 310–322. (Cited on pages 37, 52, 53, 54, 55, and 91)

Wingfield, A. and Tun, P. A. (**2007**), "Cognitive supports and cognitive constraints on comprehension of spoken language," *Journal of the American Academy of Audiology* **18**, pp. 548–558. (Cited on pages 1 and 29)

Wingfield, A., Tun, P. A., and McCoy, S. L. (**2005**), "Hearing loss in older adulthood: What it is and how it interacts with cognitive performance," *Current Directions in Psychological Science* **14**(3), pp. 144–148. (Cited on pages 58 and 59)

Zekveld, A. A., Kramer, S. E., and Festen, J. M. (**2010**), "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," *Ear and Hearing* **31**(4), pp. 480–490. (Cited on pages 36 and 37)

Zekveld, A. A., Kramer, S. E., and Festen, J. M. (**2011**), "Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response," *Ear and Hearing* **32**(4), pp. 498–510. (Cited on pages 37, 58, 60, and 81)

Zokoll, M. A., Hochmuth, S., Warzybok, A., Wagener, K. C., Buschermoehle, M., and Kollmeier, B. (**2013**), "Speech-in-noise tests for multilingual hearing screening and diagnostics," *American Journal of Audiology* **22**, pp. 175–179. (Cited on page 10)

Danksagung

Es ist ein lobenswerter Brauch: Wer was Gutes bekommt, der bedankt sich auch.
(W. Busch, 1832 – 1908)

An erster Stelle geht mein Dank an Birger Kollmeier. Bedanken möchte ich mich für seine konstruktiven Diskussionen und entscheidenden Denkanstöße, die meine Arbeit überhaupt entstehen lassen konnten und wesentlich zum Gelingen der Arbeit beigetragen haben. Über die fachlich exzellente Betreuung meiner Arbeit hinaus, hat Birger mich auch in vielen weiteren Bereichen unterstützt. Er hat mich motiviert über den Tellerrand der Doktorarbeit hinauszuschauen und stets mein Engagement in Bereichen wie z.B. Öffentlichkeitsarbeit, Gleichstellungsarbeit oder Studentenbetreuung gefördert. Vielen Dank Birger für diese großartige Unterstützung!

Bei Steven van de Par bedanke ich mich ganz herzlich für die Übernahme des Korreferats. Trotz rappel-vollem Terminkalender hat er sich spontan bereit erklärt die Aufgaben des Zweitgutachters für dieser Arbeit zu übernehmen.

Bei Thomas Brand möchte ich mich bedanken für die anregenden Diskussionen und Kommentare. Seine Nachfragen haben dazu geführt, sich immer wieder kritisch mit der Arbeit auseinanderzusetzten.

Esther Ruigendijk danke ich für ihre hilfreichen Impulse und Diskussionen, die mich dazu gebracht haben über die physikalischen/audiologischen Aspekte meiner Arbeit hinauszudenken. An dieser Stelle sei auch dem AULIN-Team gedankt. Regelmäßige Projekttreffen, Workshops und die mit ihnen verbunden Diskussionen haben meine Arbeit erwähnenswert vorangetrieben. Zusätzlich hat sich meine Kenntnis über Produktnamen eines großen schwedischen Möbelhauses erheblich gesteigert. Bedanken möchte ich mich auch bei der SprAud-Gruppe für den fachlichen Austausch, die inhaltlichen Vorträge und Diskussionen, die immer in einer sehr aufgelockerten Atmosphäre statt gefunden haben.

Natürlich hätte ich meine Arbeit nicht fertig stellen können ohne die freundlichen und fleißigen Helfer im Hintergrund. Dies soll nun nicht den Verdacht einer plagiativen Arbeit erwecken, sondern vielmehr möchte ich hier den Menschen danken, die mich durch inhaltliche und mentale Höhen und Tiefen meiner Arbeit begleitet haben.

Großer Dank gilt Jörg-Hendrik Bach, Sabine Hochmuth, Tobias Neher und Darrin Reed für das fleißige Korrekturlesen besonders gegen Ende meiner Arbeit. Trotz zeitlicher Engpässe konnten sie mir stets hilfreiche und zügige Kritik und Verbesserungsvorschläge geben. Helga Sukowski danke ich für ihren geduldigen Einsatz in Sachen Statistik-Notruf. Danke für euer phantastisches Engagement.

Ein ganz allgemeiner Dank geht an die Arbeitsgruppe der Medizische Physik. Meine Kollegen und Freunde haben zum einen dazu beigetragen, dass ich mich inhaltlich gut aufgehoben gefühlt habe. Zum anderen haben sie für ein extrem hohes Wohlfühlklima gesorgt. Unter Ihnen gibt es einige, denen besonderer Dank gebührt. Sabine, Helga, Jörg-Hendrik, Niklas, Bernd und Dirk, danke für eure sensationelle Unterstützung in vielen Dingen. Ihr habt mich auch in angespannter und gestresster Laune wunderbar ertragen.

Ein ganz besonderer Applaus gebührt Anita, Frank, (G&G Transporte), Ingrid, Katja und Niklas G., ohne die ich sicherlich mehrfach verzweifelt wäre. Besonders bedanken möchte ich mich bei Anita (bekannt als Miss Medi) für ihre fleißige, geduldige und spontane Unterstützung bei der Durchführung der Messungen. Danke Frank und Niklas für euren Rat und eure Taten, wenn es darum ging Beziehungskrisen mit meinem Computer zu bewältigen. Danke Katja und Ingrid für eure Hilfe in allen organisatorischen Belangen oder entscheidende Hinweise wie z.B. zum Aufenthaltsort von Birger.

Ein riesengroßes Dankeschön geht an meine Freunde, auf die ich mich immer verlassen konnte und die notwendig sind um eine Promotion zu überstehen. Sie haben es immer geschafft mich in den richtigen Momenten auf andere Gedanken zu bringen. Danke Almut und Jutta, sie waren immer da, wenn es emotional wurde. Danke auch an die Mädels für die kleinen Highlights in der Mitte der Woche und den (unwissenschaftlichen) Austausch über das sinushafte Verhalten einer Promotion.

Ein besonderer Dank geht an meine Eltern und an meine Geschwister Katharina und Rainer, die mir während der gesamten Zeit starken familiären Rückhalt geboten haben. Die Zeiten in denen ich nur im Büro zu erreichen bin, sind nun hoffentlich vorbei!

Zuallermeist danke ich Tobias für seine unschätzbare Hilfe. Ohne ihn würde meine Arbeit nicht in dieser Form vorliegen. Es gibt vieles für das ich dankbar bin, aber das Beste bist du!

Curriculum Vitae

Personal:	
name	Dorothea Christine Wendt
date of birth	December 6th, 1982
place of birth	Celle, Germany
nationality	German
Education:	
September 2008 - July 2013	**Ph.D. candidate** University of Oldenburg, Germany
September 2008 - August 2013	**Research assistant** in the DFG project AULIN (part I and II), University of Oldenburg, Germany
April 2007 - April 2008	**Diploma thesis** in Physics: *Untersuchung der zeitlich und spektral kodiereten Tonhöhe von harmonischen Tonkomplexen mit Magnetresonanztomographie* Supervisors: Prof. Dr. Dr. Birger Kollmeier and Prof. Dr. Jesko Verhey, Medical Physics Section, University of Oldenburg, Germany
October 2002 - April 2007	**Graduate student of physics** University of Oldenburg, Germany
2002	**Abitur** Gymnasium Hankensbüttel, Germany

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst habe und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Oldenburg, den 16. Mai 2013

.....
(Dorothea Christine Wendt)