

Technical University of Denmark



## How Hearing Impairment Affects Sentence Comprehension: Using Eye Fixations to Investigate the Duration of Speech Processing

**Wendt, Dorothea; Kollmeier, Birger; Brand, Thomas**

*Published in:*  
Trends in Hearing

*Link to article, DOI:*  
[10.1177/2331216515584149](https://doi.org/10.1177/2331216515584149)

*Publication date:*  
2015

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Wendt, D., Kollmeier, B., & Brand, T. (2015). How Hearing Impairment Affects Sentence Comprehension: Using Eye Fixations to Investigate the Duration of Speech Processing. *Trends in Hearing*, 19, 1-18. DOI: 10.1177/2331216515584149

**DTU Library**  
Technical Information Center of Denmark

---


### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# How Hearing Impairment Affects Sentence Comprehension: Using Eye Fixations to Investigate the Duration of Speech Processing

Trends in Hearing  
2015, Vol. 19: 1–18  
© The Author(s) 2015  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2331216515584149  
tia.sagepub.com  


Dorothea Wendt<sup>1,2</sup>, Birger Kollmeier<sup>1,3</sup>, and Thomas Brand<sup>1,3</sup>

## Abstract

The main objective of this study was to investigate the extent to which hearing impairment influences the duration of sentence processing. An eye-tracking paradigm is introduced that provides an online measure of how hearing impairment prolongs processing of linguistically complex sentences; this measure uses eye fixations recorded while the participant listens to a sentence. Eye fixations toward a target picture (which matches the aurally presented sentence) were measured in the presence of a competitor picture. Based on the recorded eye fixations, the single target detection amplitude, which reflects the tendency of the participant to fixate the target picture, was used as a metric to estimate the duration of sentence processing. The single target detection amplitude was calculated for sentence structures with different levels of linguistic complexity and for different listening conditions: in quiet and in two different noise conditions. Participants with hearing impairment spent more time processing sentences, even at high levels of speech intelligibility. In addition, the relationship between the proposed online measure and listener-specific factors, such as hearing aid use and cognitive abilities, was investigated. Longer processing durations were measured for participants with hearing impairment who were not accustomed to using a hearing aid. Moreover, significant correlations were found between sentence processing duration and individual cognitive abilities (such as working memory capacity or susceptibility to interference). These findings are discussed with respect to audiological applications.

## Keywords

hearing impairment, processing duration, eye-tracking, linguistic complexity

## Introduction

During speech understanding, 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 can be affected not only by processes in the auditory periphery but also by cognitive factors of the listener. In particular, elderly people often report increasing difficulties understanding speech in acoustically complex situations. Their difficulties may arise not only from deficits in impaired functioning of the inner ear or from age-related changes in cognitive factors (such as working memory capacity or susceptibility to interference) but also from an interaction between these two levels of processing (Pichora-Fuller, 2003; Schneider, Pichora-Fuller, & Daneman, 2010). The primary focus of the current study was to analyze the influence of hearing status on processing

duration at controlled speech intelligibility levels. For that purpose, an eye-tracking paradigm was applied that allows an *online* analysis of processing duration, that is, *during* the presentation of speech. For a systematic investigation of processing duration, the linguistic complexity of the speech material was varied, ranging from simple to more complex sentence structures. In addition, the effect of sensory demands on processing duration was examined for all sentence structures by

<sup>1</sup>Medizinische Physik, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

<sup>2</sup>Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark

<sup>3</sup>Cluster of Excellence Hearing4all, Oldenburg, Germany

## Corresponding author:

Dorothea Wendt, Hearing Systems, Ørstedts Plads, Building 352, Technical University of Denmark, Kgs. Lyngby 2800, Denmark.

Email: wendt@elektro.dtu.dk



Creative Commons CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (<http://www.creativecommons.org/licenses/by-nc/3.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<http://www.uk.sagepub.com/aboutus/openaccess.htm>).

measuring processing duration in quiet and in two different noise conditions. To account for possible age-related changes on the level of cognitive processing, processing duration was further examined with respect to individual cognitive abilities.

Previous studies have shown that hearing impairment can cause a decrease in comprehension accuracy or an increase in reaction time, even at high speech intelligibility levels. For instance, McCoy & Tun (2005) measured recall performance of spoken words presented at intensities at which words could be correctly recognized for participants with normal hearing and with mild-to-moderate hearing loss. The group of participants with hearing impairment showed poorer recall performance in a running memory task compared with the participants with normal hearing. Furthermore, Wingfield, McCoy, Peelle, Tun, & Cox (2006) reported 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: In particular, when the task became more difficult due to an increased level of linguistic complexity, hearing loss influenced comprehension accuracy. These results suggest that hearing loss may force participants 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, McCoy, & Wingfield, 2009; Wingfield, Tun, & McCoy, 2005). The fact that hearing loss can lead to a decrease in performance even at controlled speech intelligibility levels demonstrates possible individual processing difficulties that are usually ignored when using common audiological measures, such as speech reception thresholds (SRTs). So far, there are no standard measures in audiology that are designed to detect these difficulties in an effective and objective way.

Audiological research in the field of speech perception has long focused on the peripheral auditory domain. However, several studies have demonstrated a relationship between cognitive mechanisms and speech perception, in particular in adverse listening situations (Pichora-Fuller, 2008; Pichora-Fuller, Schneider, & Daneman, 1995; Schneider, Daneman, & Pichora-Fuller, 2002). It has been shown that sentence processing can be affected by cognitive abilities, such as working memory or susceptibility to interference (Akeroyd, 2008; Just & Carpenter, 1992; Rönnerberg, Rudner, Foo, & Lunner, 2008). Working memory capacity in particular is thought to be relevant for speech processing (e.g., Humes, Lee, & Coughlin, 2006; Zekveld, Kramer, & Festen, 2011). Humes et al. (2006), for instance, showed significant correlations between the performance of participants with hearing impairment in a digit-span test, where they had to repeat a chain of numbers, and

their performance in speech recognition tests. The output of the digit-span test was then interpreted as a measure of the individual's working memory capacity. Carroll & Ruigendijk (2013) used a word-monitor paradigm in order to investigate reaction times during processing sentences with different levels of syntactic complexity. In their study, reaction times were measured at different measuring points (across the sentence), and a three-way interaction of noise type, measuring point, and working memory (indicated by the reading span) was found. 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 & Ruigendijk (2013) observed this interaction only for syntactically critical measuring points (with assumedly higher processing effort), which suggested that participants were able to recover from this extra processing effort before the end of the sentence. These results further support the necessity of an online investigation of the process of speech understanding.

Eye movements have been used frequently in psycholinguistic research to analyze speech processing online. The temporal interplay between speech processing and eye movements was first shown in the pioneering study by Cooper (1974). The visual world paradigm (Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) was developed to reveal the interaction of vision and language by simultaneously presenting spoken language and visual scenes. Wendt, Brand, & Kollmeier (2014) showed that eye fixations can be used to detect changes in processing durations during sentence comprehension even at high audible levels (e.g., for clearly understandable sentences presented in quiet). This method works online: Processing duration is measured during the presentation of the sentence. This online measure enables the detection of temporary misinterpretations of the sentence's meaning and differences in processing duration. The eye-tracking paradigm applies sentences from the OLACS corpus (Uslar et al., 2013), which allows for a systematic variation of the type and level of linguistic complexity. Within the OLACS corpus, linguistic complexity is varied by changing linguistic parameters such as word order or ambiguity. Wendt, Brand, & Kollmeier (2014) reported increased sentence processing durations for linguistically complex sentence structures compared with less complex sentence structures, although speech intelligibility was high. The increased processing duration was interpreted as evidence for increased cognitive processing demands when processing complex sentence structures.

The main focus of the current study was to examine the contributions of sensorineural hearing loss on sentence processing duration using the eye-tracking

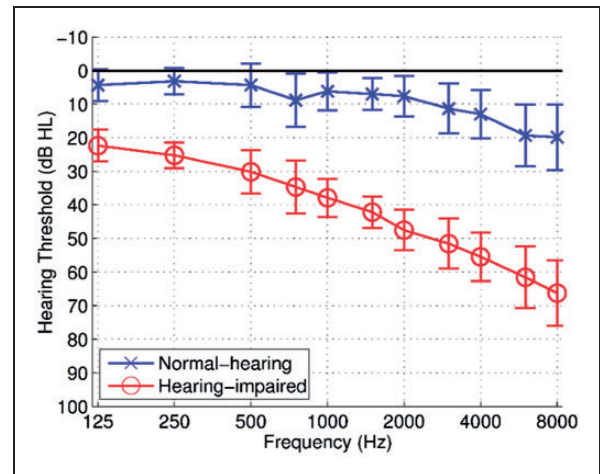
paradigm introduced by Wendt et al., (2014). The eye fixation analysis presented in Wendt, Brand, & Kollmeier (2014) has been extended in the present study to estimate the processing duration of individual participants in various listening conditions. These individual data are then averaged and discussed for two different groups of participants, that is, normally hearing participants and participants with hearing impairment, in order to investigate the interaction of linguistic complexity, background noise, and hearing loss on sentence processing duration at controlled speech intelligibility levels. In addition, the purpose of this study was to test whether the observed dependency of processing duration is specific for hearing impairment or whether it involves selected cognitive processing parameters (such as working memory capacity). For that reason, cognitive measures were used to estimate individual differences in cognitive abilities and to investigate the extent to which individual processing duration correlates with participants' cognitive abilities. Especially in more adverse listening conditions, the capacity for storing and remembering words and for manipulating the stored speech signal was expected to play an important role in speech processing (see, Rönnberg, 2003; Rönnberg et al., 2008; Rönnberg, Rudner, Lunner, & Zekveld, 2010). For that reason, a digit-span test and a word-span test (Tewes, 1991) were applied in the current study. Since noise may be generally viewed as a kind of interference, understanding speech in noise was expected to be affected by participants' susceptibility to interference and general attention (cf., Rönnberg et al., 2008, 2010). Therefore, the Stroop test, which is a selective attention task (May, Hasher, & Kane, 1999; Kim, Kim, & Chun, 2005), was used to investigate the participants' ability to ignore additional confounding visual information unrelated to the actual visual task. Both the span test and the Stroop test have already been shown to correlate with speech perception measures using the OLACS material (see Carroll & Ruigendijk, 2013; Uslar et al., 2013).

This study addressed the following hypotheses:

**Hypothesis 1:** Hearing impairment causes a substantial increase in processing duration even when participants with hearing impairment reach the same intelligibility as participants with normal hearing.

**Hypothesis 2:** The effect of hearing loss on processing duration is most pronounced in adverse listening conditions characterized by high cognitive and acoustical demands, such as when processing complex sentences in background noise.

**Hypothesis 3:** An increased processing duration is related to decreased cognitive performance in areas of speech perception.



**Figure 1.** Mean hearing threshold averaged across the left and right ears for a group of participants with normal hearing and a group of participants with hearing impairment. The error bars show the standard deviation.

## Material and Methods

### Participants

Seven male and 13 female participants with normal hearing (NH group) participated in the first 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 Hz and 4000 Hz, and hearing thresholds of 30 dB HL or better at 6000 Hz and 8000 Hz (see Figure 1). The pure-tone average thresholds across the frequencies ranging from 125 Hz to 8000 Hz was 6.9 dB HL ( $SD = 4.3$  dB HL).

Nine male and 13 female participants with hearing impairments (HI group) participated in the second experiment, with an average age of 65 years (ranging from 42 to 77 years). Participants belonging to this group had mild to moderate, sensorineural, postlingual hearing loss. The pure-tone average threshold across the frequencies ranging from 125 Hz to 4000 Hz was 39.3 dB HL ( $SD = 6.7$  dB HL), see Figure 1.

The HI group was further divided into two groups: participants in the HA group wore hearing aids in daily life for at least the past 6 months; participants in the noHA group did not use hearing aids (see Table 1).

### Material

**Speech material.** The OLACS corpus was developed and evaluated by Uslar et al. (2013) to systematically investigate the effect of linguistic complexity on speech processing and comprehension. OLACS contains seven different sentence structures that differ in their linguistic complexity. Different types of linguistic complexities are

**Table 1.** Participants with hearing impairment divided into two groups according to their hearing aid experience.

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

Note. Left panel: participants who are acclimatized with hearing aids (HA group) with their age and their PTA thresholds across the frequencies ranging from 125 Hz to 4000 Hz. Fourth column: duration of hearing aid use. Right panel: participants who do not use hearing aids in their everyday life (noHA group). HA = hearing aids; PTA = pure-tone average; HL = hearing level.

**Table 2.** Examples of the Three Different OLACS Sentence Structures (SVO, OVS, and ambOVS) Tested in This Study.

SVO structure	Der The <sub>nom</sub>	kleine little <sub>nom</sub>	Junge <sub>PTD</sub> boy	grüßt greet <sub>3sg</sub>	den the <sub>acc</sub>	lieben nice <sub>acc</sub>	Vater. father.
“The little boy greets the nice father.”							
OVS structure	Den The <sub>acc</sub>	lieben nice <sub>acc</sub>	Vater <sub>PTD</sub> father	grüßt greet <sub>3sg</sub>	der the <sub>nom</sub>	kleine little <sub>nom</sub>	Junge. boy.
“It is the nice father that the little boy is greeting.”							
ambOVS structure	Die The <sub>amb</sub>	liebe nice <sub>amb</sub>	Königin queen <sub>fem</sub>	grüßt greet <sub>3sg</sub>	der <sub>PTD</sub> the <sub>nom</sub>	kleine little <sub>nom</sub>	Junge. boy.
“It is the nice queen that the little boy is greeting.”							

Note. The disambiguating word—the word from which the target picture could theoretically first be recognized by the participant—is indicated by PTD. *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. PTD = point of target disambiguation; SVO = subject–verb–object; OVS = object–verb–subject; ambOVS = ambiguous object–verb–subject.

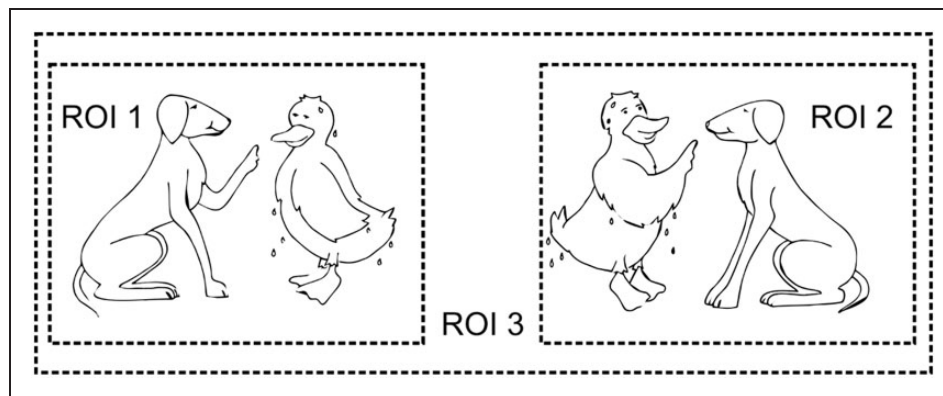
realized by varying word order and ambiguity (for detailed information about the different sentence structures, see Carroll & Ruigendijk, 2013; Uslar et al., 2013; Wendt, Brand, & Kollmeier, 2014). Of the seven structures, the three verb-second structures have been shown to be most appropriate for investigating processing duration in the eye-tracking paradigm applied here (Wendt, Brand, & Kollmeier, 2014) and are therefore used as auditory stimuli in this study (Table 2):

1. The subject–verb–object (SVO) structure represents the canonical word order in the German language with a transitive verb. Since the first article *Der* (“The<sub>nom</sub>,” 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.

2. The object–verb–subject (OVS) structure represents a noncanonical word order in the German language with a transitive verb. Since the first article *Den* (“The<sub>acc</sub>,” 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 its entities.

In general, the SVO structure is less complex compared with the OVS structure, since the subject-before-object sentence structure is preferred in the German language (Bader & Meng, 1999; Gorrell, 2000).



**Figure 2.** Example picture set for a sentence with the amBOVS structure: Die nasse Ente tadelt der treue Hund. (The wet duck [accusative] reprimands the loyal dog [nominative].) A picture set consists of two single pictures. The dashed lines indicate the three ROIs 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 (from Wendt, Brand, & Kollmeier, 2014). ROI = region of interest.

Thus, the SVO structure is expected to be processed more easily than the object-before-subject sentence structures.

3. The ambiguous object–verb–subject (amBOVS) structure also exhibits a noncanonical word order with a transitive verb. Since the first article *Die* (“The<sub>amb</sub>,” 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 SVO and OVS structures are unambiguous with respect to their meaning and to the grammatical role of the sentence components. For instance, the noun, *Junge*<sub>PTD</sub> “boy” in the SVO structure 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. The point of target disambiguation (denoted by PTD in Table 2) is defined as the start of the disambiguating word, which is the word that allows the correct assignment of the target picture. The same applies for the OVS structure, where the first noun *Vater*<sub>PTD</sub> “father” allows the role assignment of the object character in the target picture. However, for the amBOVS structure, the identification of subject and object function is not possible until the article *der*<sub>PTD</sub> (“the<sub>nom</sub>,” nominative) of the second noun phrase. Adding ambiguity further increases the level of complexity, since the thematic role assignment of agent and object can only be made at a late point in the sentence (see overview by Altmann, 1998). Participants might have to revise their initial hypotheses about sentence structure and reanalyze the sentence meaning after the PTD. This necessary reanalysis leads to an increase in processing cost.

**Graphical material.** Picture sets corresponding to the sentences of the OLACS corpus were used as visual stimuli (Wendt, Brand, & Kollmeier, 2014). 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 picture in Figure 2). In the competitor picture, the roles of the agent (the entity that carries out the action) and the object (the entity that is affected by the action) were interchanged (see right picture in Figure 2). Both pictures were of the same size, and within each picture, 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, but one was mirrored, so that 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 them did. The filler trials were incorporated to force the participants to analyze both pictures.

### Stimuli and Procedure

**Acoustic conditions.** Sentences were presented either in quiet 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 speech, created by overlapping 30 tracks, each consisting of the entire randomly overlapping speech material, as described by Kollmeier (1990) and Wagener, Josvassen, & Ardenkjaer (2003). 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, Verschuur, Ludvigsen, & Westermann, 2001, modified

according to Wagener, Brand, & Kollmeier, 2006, with a maximum pause length limited to 250 ms).

The sound level of the stimuli was 75 dB SPL but was adjusted if participants preferred a higher level. The spectrum of speech and noise was further adjusted according to the individual hearing loss using the National Acoustic Laboratories-Revised (NAL-R) formula (Byrne & Dillon, 1986) to ensure that each participant had roughly the same spectral information available. Based on the individual audiogram, the required gain was applied separately for different frequency bands using an eight-channel Butterworth filterbank with center frequencies at 0.25, 0.5, 0.75, 1, 2, 3, 4, and 6 kHz (filter order = 5). To ensure comparable speech intelligibility levels across all participants, every participant was measured at his or her individual SRT of 80% word understanding (SRT80) in noise. For that purpose, the individual SRT80 was measured for each sentence structure (SVO, OVS, and ambOVS) in stationary noise and in modulated noise.

**Procedure.** In total, 180 OLACS sentences (60 of each sentence structure) were 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. In total, 604 trials were conducted for each participant.

Each participant performed one training block consisting of 60 sentences 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 10 minutes. The complete measurement took about 3 hours per participant and was divided into two sessions, which were performed on different days within 1 week.

**Eye-tracking paradigm.** The visual stimulus containing the two alternative scenes was presented from 1,000 ms before the onset of the acoustic stimulus until the response of the participant. Participants were instructed to identify the picture that matched 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 button in the middle 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 9-point fixation stimulus.

**Apparatus.** An eye-tracker system (EyeLink 1000 desktop system, SR Research Ltd.) was used with a sampling rate of 1000 Hz to monitor participants' eye fixations. The pictures were presented on a 22-in. multiscan color computer screen with a resolution of 1,680 × 1,050 pixels. Participants were seated 60 cm from the computer screen and a chin rest was used to stabilize their 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 ISO 389-8: 2004(E). For the calibration of the speech signals, a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2 in. microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier were used. All experiments took place in a sound-insulated booth.

### Preparatory Measurements

**Speech recognition measurements.** To ensure that participants conducted the eye-tracking paradigm at the same speech intelligibility (independent of sentence structure and hearing status), speech recognition was measured for each participant beforehand. For that purpose, sentences from the OLACS corpus were presented in stationary noise and in modulated noise (the same noise types that were used for the eye-tracking paradigm) in a sound-insulated booth through headphones (Sennheiser HDA 200). Participants were asked to repeat all words of the presented sentence as accurately as possible. The correctly repeated words within one sentence were counted. An adaptive procedure was used to determine the SRT80, which is the signal-to-noise ratio at which 80% of the speech material was recognized correctly (see Uslar et al., 2013 for detailed information about the measurement procedure). Within the adaptive procedure, the speech level of each sentence was adjusted according to the number of correctly recognized words (see Brand & 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 (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 per test. Sentences with different sentence structures were presented in a random order within one list for each participant. The SRT80 was determined for each sentence structure and each participant. The averaged SRT80s for the NH and HI groups are listed in Table 3.

**Table 3.** Mean SRT ( $\pm$ SD) and cognitive measures ( $\pm$ SD) for participants with normal hearing (NH group) and hearing impairment (HI group).

	SVO structure	OVS structure	ambOVS structure	Cognitive tests		
				Stroop measure in ms	Digit-span score in %	Word-span score in %
NH	SRT80 in stationary noise			1,197 ( $\pm$ 225)	51 ( $\pm$ 1)	28.9 ( $\pm$ 9)
	-4.4 ( $\pm$ 1.3) dB	-3.6 ( $\pm$ 1.4) dB	-4.2 ( $\pm$ 0.7) dB			
	SRT80 in modulated noise					
	-9.8 ( $\pm$ 2.1) dB	-8.1 ( $\pm$ 2.7) dB	-7.8 ( $\pm$ 2.7) dB			
HI	SRT80 in stationary noise			1,295 ( $\pm$ 244)	48 ( $\pm$ 12)	27.9 ( $\pm$ 10)
	-1.5 ( $\pm$ 2.7) dB	-0.1 (2.6) dB	-0.5( $\pm$ 2.3) dB			
	SRT80 in modulated noise					
	0.1 ( $\pm$ 3.8) dB	2.3(3.1) dB	1.9 ( $\pm$ 3.0) dB			

Note. Left panel: Mean SRT measured for all three sentence structures in stationary and in modulated noise. Right panel: mean results of the cognitive tests: the Stroop measure, the digit-span, and the word-span score are shown for both group. NH = normal hearing; HI = hearing impaired; SRT = speech reception threshold; SVO = subject-verb-object; OVS = object-verb-subject; ambOVS = ambiguous object-verb-subject.

**Cognitive tests.** To reveal correlations between cognitive abilities and processing duration, 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, Kim, & Chun (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 10 trials, mean reaction times were measured for each participant.

Participants also 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 is presented aurally and participants are then asked to repeat the chain in reversed order. 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., Cheung & Kemper, 1992; Kemper, Kynette, Rash, & O'Brien, 1989). The word-span test is based on the same experimental design but uses semantically unrelated words (one and two syllables) instead of digits. The word-span test was conducted as a forward version: Participants were asked to repeat the chain of words in the original order. The word-span forward test is

expected to provide a measure for pure verbal memory (span) capacity. To calculate the span scores for both span tests, one point was awarded for each correctly repeated trial (according to the traditional scoring; see Tewes, 1991). The scores were presented in percentages: The number of points obtained was divided by the total points possible (see Table 3).

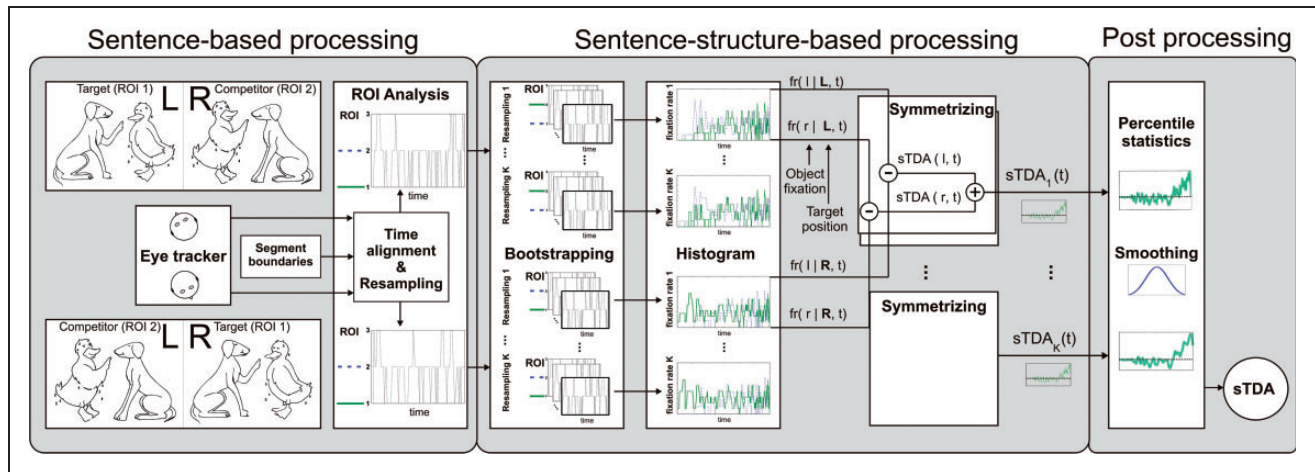
## Data Analysis

### Analysis of the Eye Fixation Data

The recorded eye fixation data were transformed into the single target detection amplitude (sTDA). The sTDA calculation was divided into three processing stages: the sentence-based processing stage, the sentence-structure-based processing stage, and the postprocessing stage (see Figure 3). In the sentence-based processing stage, the eye fixations toward different regions of interest (ROIs) on the display were determined. The target picture was assigned to ROI 1, the competitor picture was assigned to ROI 2, and the background was assigned to ROI 3. Moreover, 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 fixation time patterns that reflect the dynamics of the recognition process for correctly identified sentences.

The processing stage is introduced in order to calculate a mean sTDA for an individual participant in a particular condition. For each sentence structure, the sTDA is averaged across all trials within a certain acoustical condition. This 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 toward the target picture and a negative sTDA describes more fixations toward





**Figure 3.** Schematic visualization of the calculation of the sTDA. The calculation of the sTDA consists of three processing stages, namely, the sentence-based processing, the sentence-structure-based processing, and the postprocessing stage. ROI = region of interest; sTDA = single target detection amplitude.

**Table 4.** Time Segments Used for Time Alignment Across All Sentences.

	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
Segment length in samples	100	100	100	100	100	Maximum 200
	<i>No acoustic stimulus</i>	Der kleine The little	Junge boy	grüßt den greets the	lieben Vater. nice father	<i>Response time</i>
Mean segment borders in ms	0–1,000	(1,000–1,745) ±130	(1,745–2,340) ±135	(2,340–2,995) ±130	(2,995–4,140) ±151	(4,140–end) ±114

Note. The second row gives the length of each segment in number of 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 length of each segment calculated over all sentences after the resampling procedure is shown in milliseconds ( $\pm$ SD), (from Wendt, Brand, & Kollmeier, 2014).

the competitor picture. Note that the calculation of sTDA is similar to the calculation of TDA (introduced by Wendt, Brand, & Kollmeier, 2014) in most of the processing stages. In contrast to the TDA, the sTDA is derived independently for each participant whereas the TDA is based on average data for an ensemble of participants.

**Time alignment and resampling.** In the sentence-based processing stage, the eye fixations toward the target (ROI 1), the competitor (ROI 2), and the background (ROI 3) were recorded as a function of time for each sentence  $n = 1, \dots, N$  in that particular condition. Since the sentences differed in length, a time alignment of the recorded eye fixation data was employed, allowing comparisons across sentences. Therefore, each sentence was divided into several time segments, as depicted in Table 4. Segment 1 describes the time from the onset of the visual stimulus to the onset of the acoustic

stimulus, which had a fixed length of 1,000 ms. The spoken sentence was presented during Segments 2 through 5. The time from the end of the spoken sentence until the participant's response by pressing the response key was allotted to Segment 6. 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 depended on the mean reaction time of the participant, with a maximal length of 200 samples (see Table 4). This segment-based resampling led to a segment-dependent sampling rate due to the individual length of each segment. 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).

After this time alignment and resampling stage, the fixation  $F_i(n)$  toward one of the three ROIs, that is,

$F_t(n) \in \{\text{ROI1}, \text{ROI2}, \text{ROI3}\}$ , was available as a function of the time index  $t=1, \dots, T$  for all  $n=1, \dots, N$  sentences. Eye blinks were removed and not considered for the further calculation of the sTDA.

**Bootstrapping procedure and symmetrizing.** A bootstrapping resampling procedure (see Efron & Tibshirani, 1993; van Zandt, 2002) was applied in the sentence-structure-processing stage in order to transform a set of fixations into a set of fixation rates  $fr(t)$ , to be used for statistical analysis.<sup>1</sup> 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  $N \times K$  fixations  $F_t$  for each time index  $t$ :

$$M_t = \begin{bmatrix} F_t(1, 1) & \dots & F_t(n, 1) \\ \vdots & \ddots & \vdots \\ F_t(1, k) & \dots & F_t(n, k) \end{bmatrix}$$

Afterwards, the fixations  $F_t$  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  $K=1, \dots, 10,000$  time-dependent fixation rates  $\{fr_1(t), \dots, fr_k(t)\}$  for each participant.

Previous studies mentioned baseline effects, where one ROI is 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, Gann, & Pierce, 2011). Participants tend to fixate the left picture before the start of the trial, and shift their gaze to the right at the start of the trial. This left-to-right gazing strategy could stem from the common reading direction in the German language. So far, methods that account for such baseline effects have not been applied frequently in visual world studies. Moreover, there is no setup that safely prevents participants from following their initial gaze bias and looking initially at the left or the top area of the picture. Thus, the current study uses a simple method termed *symmetrizing* to account for any strategy effects used by the participants. This symmetrizing includes an experimental design that was randomized such that the visual location of the target image was counterbalanced. Moreover, the fixation rates toward a target picture in relation to the fixation rate toward a competitor picture are calculated, resulting in the sTDA. This is realized in the sentence-structure-based processing stage.

Thus, the calculation of the sTDA was split into different processing steps, which allowed analysis of the fixation rates for left and right targets separately. For symmetrizing, four different fixation rates  $fr_k(s|S, t)$

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. Note that the fixation rates of the background (ROI 3) were not considered in the calculation of the sTDA. Depending on the position of the target, the two fixation rates of the competitor pictures  $fr_k(r|L, t)$  and  $fr_k(l|R, t)$  were subtracted from the respective fixation rates of the target pictures  $fr_k(l|L, t)$  and  $fr_k(r|R, t)$ . This results in  $sTDA_{k,r}$  for the right side:

$$sTDA_{k,r}(r, t) = fr_k(r|R, t) - fr_k(r|L, t)$$

and  $sTDA_{k,l}$  for the left side:

$$sTDA_{k,l}(l, t) = fr_k(l|L, t) - fr_k(l|R, t)$$

Both resulting  $sTDA_{k,r}(r, t)$  and  $sTDA_{k,l}(l, t)$  were added to the  $sTDA_k$ .

**Postprocessing.** In the last processing stage, a percentile statistic was used for the calculation of the 95% confidence interval and the mean value, the sTDA. A bootstrapping resampling procedure was applied across all  $sTDA_k$  in that particular condition. Thus, the sTDA was calculated over all  $K$  realizations  $sTDA_k$  of an individual participant for each OLACS sentence structures. This was realized in quiet, in stationary noise, and in modulated noise. Finally, a Gaussian smoothing filter with a kernel size of 35 samples was applied in order to reduce random fluctuations. The final sTDA was then used in order to calculate the processing duration for a single participant (see Figure 4).

**Calculation of the decision moment and the disambiguation to decision delay.** To investigate the processing duration required for sentence comprehension, the decision moment (DM) and the disambiguation to decision delay (DDD) were determined for each individual sTDA. The DM is defined as the point in time for which the sTDA exceeds a threshold value of 15% for at least 200ms (see Figure 4).<sup>2</sup> The temporal distance between the 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 (see Table 2). The temporal delay between the PTD and the DM is interpreted as a measure for participant's processing duration and is termed *disambiguation to decision delay*. Note that the focus of the proposed analysis method is to estimate the point in time where the decision toward the correct picture can be detected based on the sTDA, that is, to determine the DDD. For that reason, the proposed data analysis methods was appropriate and preferred above more

**Table 5.** Mean Picture Recognition Rates ( $\pm SD$ ) and Mean Log Reaction Times ( $\pm SD$ ) Are Shown for the Three Sentence Structures Presented in Quiet, in Stationary Noise, and in Modulated Noise for Both Groups.

		Picture recognition rate in rau			Log reaction time		
		SVO	OVS	ambOVS	SVO	OVS	ambOVS
Quiet	HI	114.7 ( $\pm 7.5$ )	111.9 ( $\pm 8.1$ )	112.9 ( $\pm 8.5$ )	3.15 ( $\pm 0.22$ )	3.17 ( $\pm 0.20$ )	3.14 ( $\pm 0.22$ )
	NH	104.8 ( $\pm 8.9$ )	107.0 ( $\pm 9.9$ )	105.4 ( $\pm 9.1$ )	3.03 ( $\pm 0.22$ )	3.06 ( $\pm 0.13$ )	3.05 ( $\pm 0.14$ )
Stat. noise	HI	111.1 ( $\pm 6.6$ )	103.7 ( $\pm 11.8$ )	107.8 ( $\pm 9.3$ )	3.13 ( $\pm 0.25$ )	3.16 ( $\pm 0.20$ )	3.16 ( $\pm 0.23$ )
	NH	101.4 ( $\pm 10.1$ )	97.2 ( $\pm 11.0$ )	104.4 ( $\pm 9.3$ )	3.02 ( $\pm 0.18$ )	3.04 ( $\pm 0.16$ )	3.01 ( $\pm 0.18$ )
Mod. noise	HI	108.8 ( $\pm 8.9$ )	106.1 ( $\pm 14.6$ )	106.3 ( $\pm 4.8$ )	3.18 ( $\pm 0.24$ )	3.17 ( $\pm 0.22$ )	3.15 ( $\pm 0.22$ )
	NH	100.4 ( $\pm 10.2$ )	100.5 ( $\pm 11.6$ )	101.1 ( $\pm 10.0$ )	3.05 ( $\pm 0.17$ )	3.01 ( $\pm 0.17$ )	3.03 ( $\pm 0.17$ )

Note. SVO = subject-verb-object; OVS = object-verb-subject; ambOVS = ambiguous object-verb-subject; NH = normal hearing; HI = hearing impaired.

parametrical methods (such as the Grow Curve Analysis as proposed, e.g., by Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Mirman, Dixon, & Magnuson, 2008; Mirman, Yee, Blumstein, & Magnuson, 2011).

## Results

### Picture Recognition Rates and Reaction Times

In addition to eye fixations, the picture recognition rates (the percentage of correctly recognized pictures) and the reaction times (the time after the presentation of the spoken sentence until the participant's response) were measured for each participant. In general, participants' performance in identifying the target picture with the button press was very high, and therefore, picture recognition rates were transformed to rationalized arcsine transformed (rau) scores (see left panel in Table 5). 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 rather the participants' 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. Since reaction times were distributed logarithmically, they were logarithmically transformed (see right panel in Table 5).

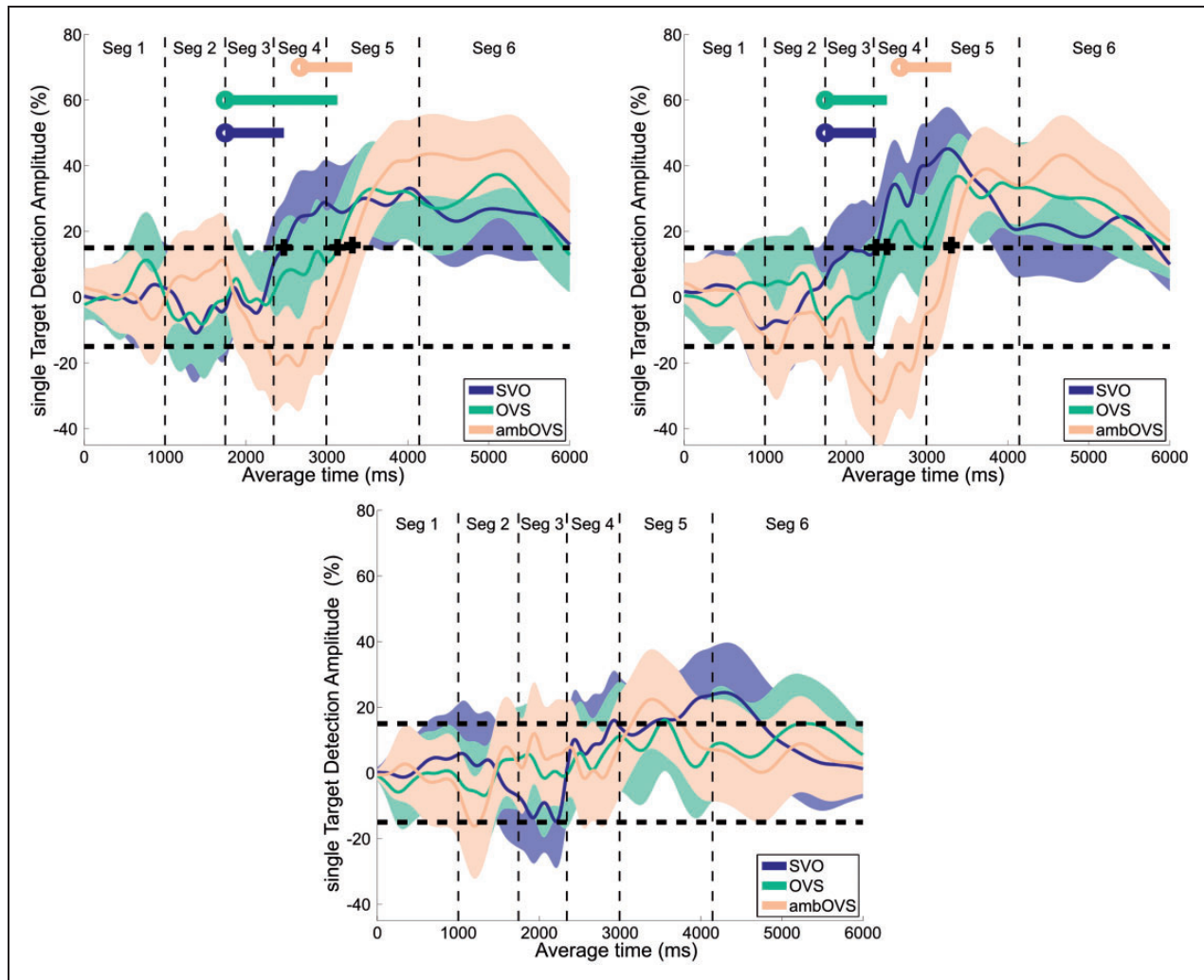
For statistical analysis, the picture recognition rate and the log reaction time were subjected to a repeated-measures analysis of variance (ANOVA) with sentence structure (SVO, OVS, and ambOVS) and acoustical condition (quiet, stationary noise, and modulated noise) as within-subject factors (i.e., across all participants) and hearing status as a between-subject factor (i.e., between NH group and HI group); picture recognition rate and log reaction time were used as the dependent measures. Bonferroni post hoc tests (level of significance set at

$p = .05$ ) were used to determine the sources of significant effects indicated by the ANOVA.

The ANOVA revealed an effect of acoustical condition on picture recognition rates ( $F(2, 74) = 15.3$ ,  $p < .001$ ), and post hoc tests showed significantly higher picture recognition rates in quiet compared with modulated noise for the SVO structure ( $p = .024$ ), the OVS structure ( $p = .029$ ), and the ambOVS structure ( $p = .004$ ). Moreover, post hoc tests showed significantly higher picture recognition rates in quiet compared with stationary noise for the OVS structure ( $p < .001$ ).

More importantly, the analysis revealed an effect of groups of participants (NH group and HI group) on the picture recognition rates ( $F(1, 37) = 15.1$ ,  $p < .001$ ), reflecting significant differences between both groups. For the NH group, paired comparisons showed significantly higher picture recognition rates for the SVO structure (in quiet:  $t(37) = 3.7$ ,  $p = .001$ ; in stationary noise:  $t(37) = 3.5$ ,  $p = .001$ ; and in modulated noise:  $t(37) = 2.7$ ;  $p = .01$ ) and for the ambOVS structure (in quiet:  $t(37) = 2.6$ ,  $p = .012$ ; and in modulated noise:  $t(37) = 2.1$ ,  $p = .045$ ). These results support the expectation that the NH group performed better than the HI group in some conditions, even though both groups were tested at levels selected to elicit equal intelligibility for all individuals.

The mean log reaction times for all experimental conditions are shown in Table 5. Again, an ANOVA was performed using log reaction time as the dependent measure. The analysis revealed no effect of sentence structure or acoustical condition, but a small but significant difference between the two groups ( $F(1, 37) = 4.2$ ,  $p = .05$ ). Paired comparisons showed significantly higher log reaction times for the HI group for the OVS structure (in stationary noise:  $t(37) = 2.0$ ;  $p = .048$ ; and in modulated noise:  $t(37) = 2.47$ ;  $p = .018$ ). Moreover, paired comparisons showed significantly higher reaction times for the HI group for the ambOVS structure in quiet:  $t(37) = 2.47$ ;  $p = .026$ .



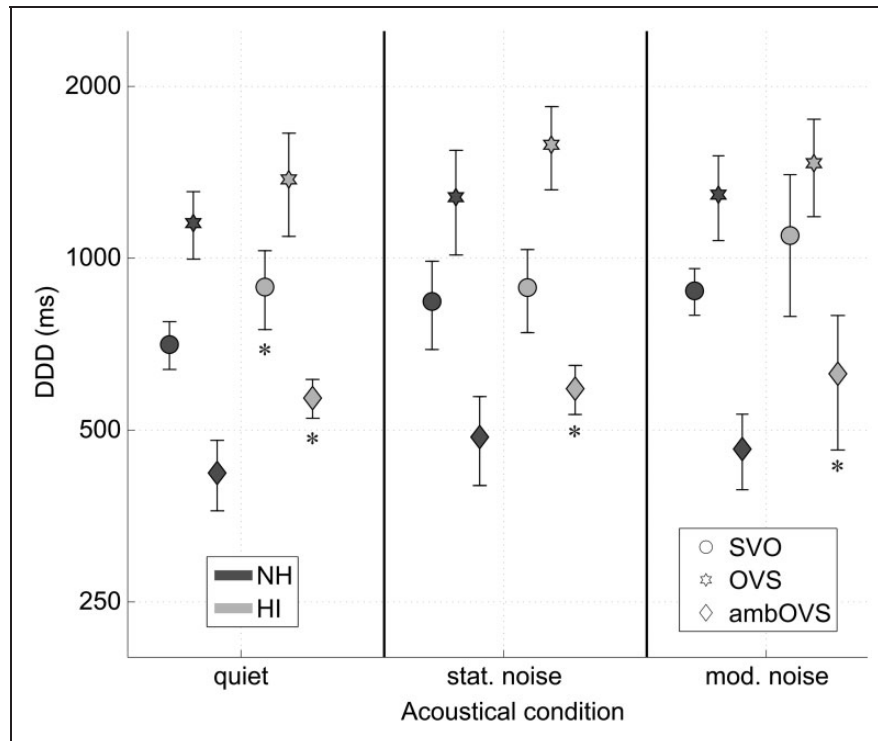
**Figure 4.** Examples of a sTDA for three individual participants. The shaded areas illustrate the 95% confidence interval for each individual curve. The circles denote the point of target disambiguation (PTD), which describes the onset of the word that allows the assignment of the spoken sentence to the target picture (see also Table 2). The plus signs denote the decision moments (DM), where the sTDA first exceeds the threshold (15% of the sTDA). The line starting from the circle denotes the disambiguation to decision delay (DDD), which is the temporal distance between the PTD and DM. SVO = subject–verb–object; OVS = object–verb–subject; ambOVS = ambiguous object–verb–subject.

### Eye-Tracking Data

*Individual data.* For a comparative investigation of the processing time, one DDD was calculated for each participant in each condition. A high DDD indicates a long processing duration for the corresponding sentence structure. Figure 4 depicts individual sTDAs with the corresponding DM and DDD for each sentence structure. These sTDAs were found to exemplify the different sTDA time courses that occurred for individual participants. The shaded areas illustrate the 95% confidence intervals of the sTDA for each individual participant. Most of the individual data showed similarities to one of these three sample participants, and the sTDAs and the corresponding DDDs of

the participants can be classified into the following groups:

- I Upper left panel in Figure 4: The sTDAs displayed here are characteristic for the majority of the participants. The DM was early for the unambiguous SVO structure and late for the OVS and ambOVS structures. The greatest DDD was observed for the OVS structure.
- II Upper right panel in Figure 4: For some participants, no differences were observed between the DM of the SVO and the OVS structures. The DMs were early for both unambiguous SVO and OVS structures. A late DM was observed only for the ambiguous sentence structure (ambOVS).

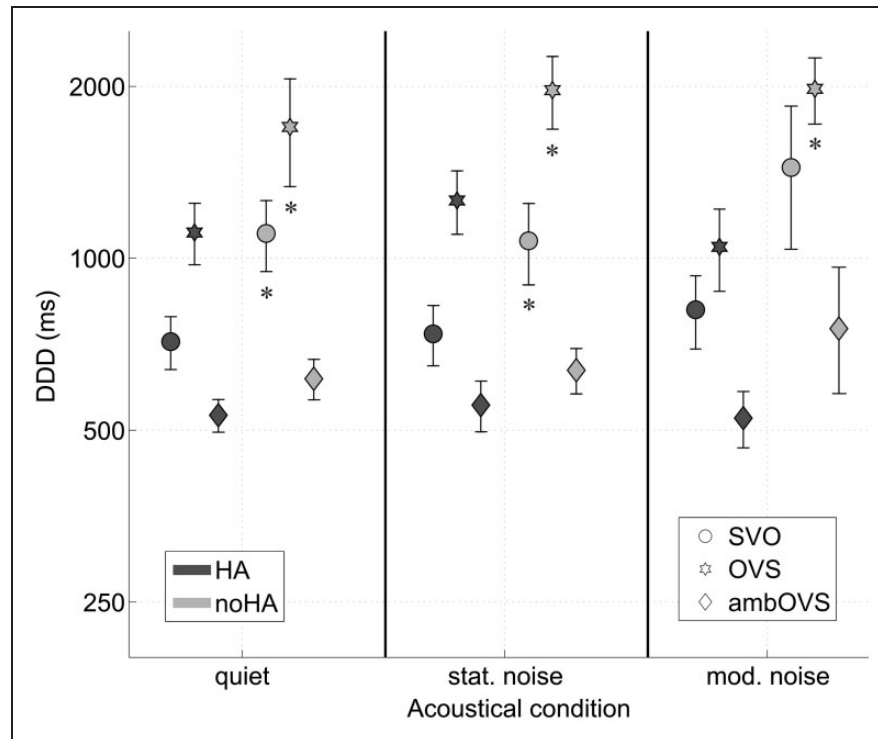


**Figure 5.** Mean DDD for participants with normal hearing (NH group; dark gray) and participants with hearing impairment (HI group; light gray) of three sentence structures (SVO, OVS, and ambOVS) in quiet, stationary noise, and modulated noise. \* denotes significant differences between both groups ( $p < .05$ ). Error bars show standard deviation. DDD = disambiguation to decision delay; NH = normal hearing; HI = hearing impaired; SVO = subject–verb–object; OVS = object–verb–subject; ambOVS = ambiguous object–verb–subject.

III Lower panel of Figure 4: For a small group of participants, the sTDA and the corresponding DDDs were not appropriate for obtaining a measure of processing duration, since the sTDAs of these participants exhibited only small amplitudes. This group included one participant from the NH group and two participants from the HI group. Since these participants failed to reach the threshold of 15%, their data were not considered for the statistical analysis.

To investigate the effect of hearing loss on processing duration, averaged DDDs were calculated across the NH group and across the HI group. Figure 5 depicts the averaged DDDs of the different sentence structures in quiet and in the two noise conditions. 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  $p = .05$ , adjusted for false discovery rate (FDR) correction) for comparison between different conditions (Nichols & Holmes, 2001). Unpaired tests were applied to investigate whether DDDs varied between both groups. To verify the statistical significance, the bootstrap  $p$ -values were determined.

*Effect of sentence structure and background noise.* Processing duration was affected by sentence structure. Significant differences in DDDs were observed between the three sentence structures in quiet as well as in background noise ( $p < .001$ ). The greatest temporal DDDs were measured for the OVS structure (about 1,400 ms averaged across both groups). In contrast, comparably small DDDs were measured for the ambOVS structure in all three acoustic conditions (about 500 in quiet and about 530 in stationary and in modulated noise). Negative sTDA values occurred for the ambOVS structure, indicating that participants fixated more frequently toward the competitor picture before the PTD was reached. This effect is interpreted as a temporary misinterpretation of the spoken sentence. Only after the PTD did the participants realize that they had identified the wrong picture as the target picture; they then had to correct 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 eye movement, which may indicate a different 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 (see also Wendt, Brand, & Kollmeier, 2014).



**Figure 6.** Mean DDD for hearing aid users (HA group; dark gray) and non-users (noHA group; light gray) of three sentence structures (SVO, OVS, and ambOVS) in quiet, stationary noise, and modulated noise. \* sign denotes significant differences between the two groups ( $p < .05$ ). Error bars show standard deviation. DDD = disambiguation to decision delay; HA = hearing aid; SVO = subject–verb–object; OVS = object–verb–subject; ambOVS = ambiguous object–verb–subject.

An increase in processing duration was expected for the more demanding listening conditions, that is, for complex sentence structures presented in background noise. To analyze the effect of acoustical conditions, paired tests were conducted for each sentence structure. The tests showed significant effects of noise for the SVO structure (in modulated noise:  $p = .001$ ), for the OVS structure (in stationary noise:  $p = .002$ ; and in modulated noise:  $p = .05$ ), and for the ambOVS structure (in stationary noise:  $p = .01$ ). Processing duration is therefore affected by background noise. However, sentence complexity did not cause any further increase in processing duration: Longer processing durations in noise were found not only for the complex sentence structures but also for the simple SVO structure.

**Effect of hearing impairment.** Unpaired tests were applied to investigate whether DDD varied depending on the hearing status, that is, between the participants with normal hearing and participants with hearing impairment. Significant differences between the two groups were found for the ambOVS structure in quiet ( $p = .001$ ), in stationary noise ( $p = .02$ ), and in modulated noise ( $p = .05$ ). Furthermore, a higher DDD for the HI group was measured for the SVO structure in quiet

( $p = .04$ ). These results support the idea that the HI group spent more time in sentence processing.

**Effect of hearing aid use.** Early results indicated that some participants from the HI group performed better than others, so the HI group was divided into two subgroups. One group consisted of 11 participants (average age of 65 years; see Table 1) who used hearing aids in their daily life (HA group) and the other group consisted of 9 participants (average age of 65 years) who did not use hearing aids in their daily life (noHA group). Figure 6 depicts the averaged DDDs of the different sentence structures in quiet and in the two noise conditions for both groups. Again, unpaired tests were applied to investigate whether DDDs varied between both groups. The test revealed significantly higher DDDs for the noHA group for the SVO structure (in quiet:  $p = .01$ ; and in stationary noise:  $p = .03$ ) and for the OVS structure in all three acoustical conditions (in quiet:  $p = .05$ ; in stationary noise:  $p = .01$ ; and in modulated noise:  $p = .001$ ), reflecting a higher processing duration for the noHA group. Note 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, and word-score) were found between the two groups.

**Table 6.** Correlation Coefficients Between the Cognitive Measures (i.e., Stroop Measure, Digit-Score, Word-Score) and the DDDs Are Shown for the Three Sentence Structures Presented in Quiet, in Stationary Noise, and in Modulated Noise for Both Groups.

		Quiet			Stationary noise			Modulated noise		
		SVO	OVS	ambOVS	SVO	OVS	ambOVS	SVO	OVS	ambOVS
NH	Word-Score	-0.19	-0.08	-0.12	-0.19	-0.15	-0.33	-0.33	-0.29	-0.18
	Digit-Score	-0.03	-0.03	-0.17	-0.17	-0.17	-0.23	-0.03	-0.44	-0.16
	Stroop Measure	0.51	0.46	0.59*	0.03	0.69*	0.48	0.40	0.61*	0.26
HI	Word-Score	-0.17	-0.30	-0.38	-0.16	-0.01	-0.59*	-0.21	-0.19	-0.66*
	Digit-Score	-0.25	-0.03	-0.02	-0.27	-0.30	-0.50	-0.56*	-0.24	-0.57*
	Stroop Measure	0.19	0.08	0.38	0.40	0.28	0.39	0.11	0.16	0.21

Note. DDD = disambiguation to decision delay; SVO = subject-verb-object; OVS = object-verb-subject; ambOVS = ambiguous object-verb-subject; NH = normal hearing; HI = hearing impaired.

\*indicates significant correlations with  $p \leq .01$ .

### Cognitive Measures

The cognitive abilities of the participants were assessed to account for their effects on sentence processing duration. Cognitive abilities were tested using the digit-span score, word-span score, and the Stroop measure. The results of the cognitive tests, averaged across all participants, are listed in Table 3. On average, lower values were measured for the word-span test than for the digit-span test. There were no significant differences between the HI and NH groups for all three tests (unpaired  $t$ -test;  $p < .05$ ). It is thus reasonable to conclude that the longer processing durations observed for the HI group did not result from differences in working memory and in participant's susceptibility to interference between the groups.

To determine whether individual differences in processing duration correlated with cognitive measures, Spearman rank correlations between the DDDs of the different sentence structures and the cognitive measures (Stroop measure, digit-score, and word-score) were calculated. Correlation coefficients were examined separately for the two groups in quiet and in the two noise conditions (see Table 6).

For the NH group, significant correlations were only observed between the Stroop measure and DDDs of the OVS and ambOVS structures. Significant correlations were observed in quiet (for the ambOVS structure:  $r = .59$ ,  $p = .01$ ) and in background noise (in stationary noise for the OVS structure  $r = .69$ ,  $p = .001$ ; in modulated noise for the OVS structure:  $r = .61$ ,  $p = .007$ ). In contrast to the NH group, the only correlation observed for the HI group was between DDDs of the ambOVS structure in background noise and the word-span score (in stationary noise:  $r = -.59$ ,  $p = .007$ ; and in modulated noise:  $r = -.66$ ,  $p = .002$ ). Moreover, the processing duration in background noise correlated with the digit-span

score backward (for the SVO structure in modulated noise:  $r = -.56$ ,  $p = .01$ ; for the ambOVS structure in modulated:  $r = -.57$ ,  $p = .01$ ).

### General Discussion

#### Effect of Hearing Impairment on Processing Duration

The main goal of the current study was to examine whether an increase in sentence processing duration can be measured for participants with hearing impairment. By recording eye fixations, processing duration was measured during sentence comprehension. Our results clearly show that sentence processing duration increased for the HI group, which supports Hypothesis 1. Even when demonstrating roughly equal performance for working memory or susceptibility in interference, the differences observed between the NH and the HI groups can be interpreted as a consequence of hearing loss. Increased processing durations suggest that the HI group required more time for sentence processing even though they were tested at the same intelligibility level as the NH group. Note that the NAL-R algorithm was applied for the HI group to ensure that the participants had roughly the same spectral information, but this did not restore the original speech signal as perceived by the NH group. In addition to the increased processing duration, the HI group also showed decreased performance in the audiovisual task; this was indicated by significantly reduced picture recognition rates and significantly higher reaction times compared with the NH group. Both the lower performance and longer processing durations may be indicators of processing difficulties due to hearing impairment, even at high intelligibility levels. The results of the current study are in line with several previous studies reporting that hearing loss affects not only speech intelligibility but also processing duration, accuracy in speech comprehension tasks, and rated effort (Larsby, Hällgren, Lyxell, & Arlinger,

2005; Tun, Benichov, & Wingfield, 2010; Wingfield et al., 2006; Zekveld et al., 2011). For instance, Larsby et al. (2005) analyzed subjectively rated effort in sentence recognition for elderly participants with normal hearing and hearing loss and reported higher subjectively rated effort due to hearing loss.

Moreover, it was hypothesized that the observed increase in processing duration caused by hearing impairment is strongest in adverse listening conditions, in this case, when processing complex sentences in background noise (Hypothesis 2). The current study found longer processing durations caused by background noise for all three sentence structures. However, a compounded effect of noise and complexity could not be measured, presumably due to the high picture recognition rates (above 90%), which indicate that participants performed very well in noisy conditions. Note that participants of both groups were tested at their individual 80% speech intelligibility threshold. The SRT80 was determined by using word scoring, which means 80% of the spoken words were recognized in noise (without pictures). However, not every word was essential to identifying the target picture (e.g., the adjectives) and the additional presentation of the visual stimulus presumably facilitated acoustical word recognition. In addition, the signal-to-noise ratio at which the processing durations were tested differed by as much as 10 dB between the two groups. Therefore, a part of the expected interactive effect of sentence structure, noise type, and hearing loss may already be covered by the differences in SRT80. Consequently, higher sensory demands caused by background noise at the individual SRT80s may be too small to detect when compounded with the effects of noise, sentence structure, and hearing loss.

### **Correlations Between Processing Duration and Cognitive Factors**

It was hypothesized that sentence processing duration is related to cognitive abilities linked to speech perception (Hypothesis 3). The results of the current study showed significant correlations between cognitive measures and processing duration, supporting this hypothesis. For the NH group, lower reaction times in the Stroop task correlated with lower processing durations of complex sentence structures. The Stroop measure, that is, the reaction time in the Stroop task, was interpreted as the participant's susceptibility to interference. However, no significant correlations between the span measures (digit-span and word-span) and processing duration were observed. In contrast to the NH group, the HI group showed significant correlations between the span measures and processing durations, indicating that reduced working memory capacity was linked to a higher processing duration in noise. In general, these results are in line with the

literature. Akeroyd (2008), for instance, 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).

In the current study, different cognitive abilities were found to correlate with processing duration for the NH than for the HI group in noise, which may indicate different processing strategies between the two groups. These findings suggest that sentence processing duration for the NH group was primarily affected by a measure of selective attention and by their ability to detect relevant speech information in the presence of noise. For the HI group, processing duration was more affected by their working memory capacity and their ability to store and manipulate the speech signal during processing.

A more theoretical framework for the ease of language understanding was proposed by Rönnberg et al. (2003). They suggested a model which describes the role of the working memory system in the context of speech understanding. The model assumes that for less adverse listening situations, no cognitive resources need to be mobilized beyond those required to understand and interpret speech anyway. In contrast, speech processing can become effortful when processing demands increase as a result of background noise or hearing impairment, which is termed *explicit processing* in the model (Rönnberg, 2003; Rönnberg et al., 2008, 2010). This requires seizing additional resources, for instance because missing speech information requires the activation of knowledge stored in long-term memory. As a result, the process of speech understanding is affected by the working memory capacity in effortful listening conditions. This has also been shown with behavioral measures, for instance in a dual-task paradigm (see, e.g., Sarampalis, Kalluri, Edwards, & Hafter, 2009). In the current study, the correlations between processing duration and working memory capacity, observed for the HI group, support this theory of explicit processing and the assumption that processing duration for complex sentence structures is a function of the working memory capacity. Moreover, the processing duration might provide an objective method for detecting the time-consuming aspects assumed by the ease of language understanding model. Higher cognitive processing demands (for sentences with higher linguistic complexity or due to hearing impairment) led to an increase in processing duration.

### **Effect of Hearing Aid Use**

Increased processing durations were measured for HI group in quiet and at controlled SRTs. This indicates that processing duration can be a sensitive measure for detecting processing difficulties in a situation where



speech intelligibility is high and speech intelligibility tests in audiology suffer from ceiling effects. Moreover, although the focus of the current study was not to study the effect of hearing aid algorithms on processing difficulties, processing duration was further investigated for participants according to their hearing aid experience. Hearing aid users did not differ significantly from non-users in their hearing thresholds, age, or cognitive measures. However, longer processing durations were observed for the non-users compared with the participants who had experience with hearing aid use.

A possible reason for the observed difference may have arisen from the preprocessing of the noisy speech signal. The signal was spectrally shaped and amplified according to the NAL-R formula for all participants of the HI group individually, to ensure similar spectral information compared with the NH group. However, the hearing aid users may have been more familiar with this kind of modified speech sound, which may have reduced the processing duration. The observed differences in processing duration suggest that the proposed paradigm may be suitable for detecting processing differences caused by acclimatization effects in the hearing aid users versus unfamiliar speech signals presented to the non-users. This idea is supported by recent studies testing aided speech recognition of hearing aid users with new or unfamiliar signal processing algorithms (Foo, Rudner, Rönnerberg, & Lunner, 2007; Rudner, Foo, Rönnerberg, & Lunner, 2009). For these unfamiliar processed speech signals, it becomes more effortful to reach a certain level of speech recognition performance. An alternative interpretation is that the non-users might have lost auditory processing capacity due to a lack of stimulation of part of their auditory system (see, e.g., Arlinger et al., 1996; Gatehouse, 1992). This interpretation is based on the hypothesis that their increased processing duration is independent of the level and spectral shaping of the speech material. 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. However, this is clearly beyond the scope of the current study.

In general, the results of the current study showed that the proposed paradigm provides an objective measure for investigating processing difficulties, indicated by a change in processing duration, under aided listening conditions. However, further studies are required to evaluate this measure for clinical application. First of all, the reliability of estimating the processing duration for individual participants needs to be evaluated. In addition, the next logical step would be to analyze whether the increase in processing duration observed for the participants without hearing aids experience can be related to an acclimatization effect and, hence, be compensated by

getting participants acquainted with preprocessed signals using the NAL-R formula. If this is the case, then the proposed paradigm could be used to control parameters of hearing aid signal processing algorithms for individual participants in order to reduce processing duration in aided listening situations. So far, the difference measured between both groups is not sufficient to distinguish between a long-term effect of hearing aid acclimatization (i.e., the mere exposure to a hearing aid) on the DDD and any short-term effect of adapting to a certain new hearing aid processing strategy. It is assumed that the mere exposure to a hearing aid has the largest effect and that any change in hearing aid processing strategies or refitting/readjusting of the hearing aid settings would probably have a much smaller effect. However, if future studies prove that a short-term effect of hearing aid processing strategy exists and is consistent, the paradigm developed here could be used to assess the effectiveness of different hearing aid processing strategies.

## Conclusions

The eye-tracking approach presented here appears to provide a useful tool for characterizing the individual processing duration involved in sentence comprehension in an objective way. The following conclusions can be drawn based on the comparison across (roughly) aged-matched participants with normal hearing and hearing loss, across acoustical conditions (quiet, stationary, and modulated noise), and across cognitive demands (three sentence structures with increasing linguistic complexity):

1. Although speech intelligibility was similar for participants with normal hearing and for participants with hearing impairment, the latter showed a significantly longer processing duration.
2. The longest processing duration was observed for participants with hearing impairment who had no hearing aid experience. It is unclear whether this is due to a lack of familiarity with the spectral shaping and amplification of the speech signals employed here or to a loss in auditory processing capacity resulting from a lack of auditory stimulation.
3. A longer processing duration was found in background noise for all sentence structures. In addition, longer processing durations for participants with hearing impairment were observed in particular for sentence structures with a higher level of linguistic complexity. However, no combined effect of background noise, complexity, and hearing loss on processing durations was detected.
4. Participants' processing durations for complex sentence structures in noise were linked to cognitive factors, such as working memory capacity and susceptibility to interference.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Deutsche Forschungsgemeinschaft (DFG; Grant No: KO 942/20-1 and BR 3668/1-2).

## Acknowledgments

The authors wish to thank Esther Ruigendijk for helpful suggestions and fruitful discussion of this study, Helga Sukowski and Jens Hjortkjær for help with the statistics, and Jana Müller for performing the measurements. The editor Andrew Oxenham and two anonymous reviewers are acknowledged for their very helpful comments.

## Notes

1. Wendt, Brand, & Kollmeier (2014) applied this resampling to the pooled data across participants.
2. On the one hand, the threshold value must be high enough to prevent small fluctuations in the sTDA at the beginning of a trial from leading to a detection of a DM. On the other hand, the value must be low enough to avoid a delayed DM. According to Wendt, Brand, & Kollmeier (2014), a threshold value of 15% was found to be a good compromise between these two requirements.

## References

- 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), 125–143.
- Allopenna, P. D., Magnuson, J. S., & 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), 419–439.
- Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences, 2*, 146–152.
- Arlinger, S., Gatehouse, S., Bentler, R. A., Byrne, D., Cox, R. M., Dirks, D. D., . . . , Willott, J. F. (1996). Report of the Eriksholm workshop on auditory deprivation and acclimatization. *Ear & Hearing, 17*(3), 87S–98S.
- Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic, 28*(2), 121–143.
- Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica, 137*(2), 201–207.
- Brand, T., & 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*, 2801–2810.
- Byrne, D., & Dillon, H. (1986). The national acoustic laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear & Hearing, 7*(4), 257–265.
- Carroll, R., & Ruigendijk, E. (2013). The effect of syntactic complexity on processing sentences in noise. *Journal of Psycholinguistic Research, 42*(2), 139–159.
- Cheung, H., & Kemper, S. (1992). Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics, 13*, 53–76.
- 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), 84–107.
- Dreschler, W. A., Verschuere, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *International Journal of Audiology, 40*, 148–157.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.
- Foo, C., Rudner, M., Rönnerberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology, 18*, 618–631.
- Gatehouse, S. (1992). The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids. *The Journal of the Acoustical Society of America, 92*(3), 1258–1266.
- Gorrell, P. (2000). The subject-before-object preference in German clauses. In B. Hemforth, & L. Konieczny (Eds), *German sentence processing* (pp. 25–63). Dordrecht, Netherlands: Kluwer Academic.
- Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *Journal of the Acoustical Society of America, 120*(5), 2926–2937.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.
- Kemper, S., Kynette, D., Rash, S., & O'Brien, K. (1989). Life-span changes to adults' language: Effects of memory and genre. *Applied Psycholinguistics, 10*, 49–66.
- Kim, S., Kim, M., & Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 102*, 16524–16529.
- Kollmeier, B. (1990). *Messmethodik, Modellierung, und Verbesserung der Verständlichkeit von Sprache [Methodology, modeling, and improvement of speech intelligibility measurements]* (Habilitation thesis). University of Göttingen, Göttingen, Germany.
- Larsby, B., Hällgren, M., Lyxell, B., & 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*, 131–143.
- Magnuson, J. S., Dixon, J., Tanenhaus, M., & Aslin, R. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 133–156.

- May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition*, 27, 759–767.
- McCoy, S. L., & Tun, P. A. (2005). Hearing loss and perceptual effort: Downstream effects on the older adults' memory for speech. *The Quarterly Journal of Experimental Psychology*, 58A(1), 22–33.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475–494.
- Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modelling. *Brain and Language*, 117, 53–68.
- Nichols, T. E., & Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1), 1–25.
- Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology*, 42(2), 26–32.
- Pichora-Fuller, M. K. (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), 144–154.
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97, 593–608.
- Rönnerberg, 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, 68–76.
- Rudner, M., Foo, C., Rönnerberg, J., & 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, 405–418.
- Rönnerberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, 47(2), 99–105.
- Rönnerberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, 12(49), 263–269.
- Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language and Hearing Research*, 52, 1230–1240.
- Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in aging adults: From discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology*, 56, 139–152.
- Schneider, B. A., Pichora-Fuller, M. K., & Daneman, M. (2010). Effects of senescent changes in audition and cognition on spoken language comprehension. In S. Gordon-Salant, R. D. Frisina, R. R. Fay, & A. N. Popper (Eds), *The aging auditory system* (pp. 167–210). New York, NY: Springer.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634.
- Tewes, U. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision 1991 (HAWIE-R)* [Hamburg-Wechsler-Intelligenztest for adults – revision 1991]. Bern, Germany: Huber.
- Tun, P. A., Benichov, J., & Wingfield, A. (2010). Response latencies in auditory sentence comprehension: Effects of linguistic versus perceptual challenge. *Psychology of Aging*, 25(3), 730–735.
- Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, 24(3), 761–766.
- Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., . . . , Kollmeier, B. (2013). Development and evaluation of a linguistically and audiological controlled sentence intelligibility test. *Journal of the Acoustical Society of America*, 124(4), 3039–3056.
- van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (pp. 461–516). New York, NY: Wiley.
- Wagener, K. C., Brand, T., & Kollmeier, B. (2006). The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology*, 45, 26–33.
- Wagener, K. C., Josvassen, J. L., & Ardenkjaer, R. (2003). Design, optimization, and evaluation of a Danish sentence test in noise. *International Journal of Audiology*, 42, 10–17.
- Wendt, D., Brand, T., & Kollmeier, B. (2014). An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities. *PLoS ONE*, 9(6), e100186.
- Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, L. C. (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, 487–497.
- Wingfield, A., Tun, P. A., & 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, 144–148.
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear & Hearing*, 32, 498–510.