

# Are Experienced Hearing Aid Users Faster at Grasping the Meaning of a Sentence Than Inexperienced Users? An Eye-Tracking Study

Trends in Hearing  
2016, Vol. 20: 1–13  
© The Author(s) 2016  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2331216516660966  
tia.sagepub.com  


Julia Habicht<sup>1</sup>, Birger Kollmeier<sup>1</sup>, and Tobias Neher<sup>1</sup>

## Abstract

This study assessed the effects of hearing aid (HA) experience on how quickly a participant can grasp the meaning of an acoustic sentence-in-noise stimulus presented together with two similar pictures that either correctly (*target*) or incorrectly (*competitor*) depict the meaning conveyed by the sentence. Using an eye tracker, the time taken by the participant to start fixating the target (the *processing time*) was measured for two levels of linguistic complexity (low vs. high) and three HA conditions: clinical linear amplification (National Acoustic Laboratories-Revised), single-microphone noise reduction with National Acoustic Laboratories-Revised, and linear amplification ensuring a sensation level of  $\geq 15$  dB up to at least 4 kHz for the speech material used here. Timed button presses to the target stimuli after the end of the sentences (*offline reaction times*) were also collected. Groups of experienced (eHA) and inexperienced (iHA) HA users matched in terms of age, hearing loss, and working memory capacity took part ( $N = 15$  each). For the offline reaction times, no effects were found. In contrast, processing times increased with linguistic complexity. Furthermore, for all HA conditions, processing times were longer (poorer) for the iHA group than for the eHA group, despite comparable speech recognition performance. Taken together, these results indicate that processing times are more sensitive to speech processing-related factors than offline reaction times. Furthermore, they support the idea that HA experience positively impacts the ability to process noisy speech quickly, irrespective of the precise gain characteristics.

## Keywords

hearing aids, speech comprehension, eye-tracking, acclimatization, linguistic complexity

Date received: 10 December 2015; revised: 1 July 2016; accepted: 1 July 2016

## Introduction

There is a large body of research documenting the effects of hearing aids (HAs) based on established measures of speech audiometry. Such measures are useful because they allow for quick and sensitive measurements and also for comparisons in different languages (e.g., Kollmeier et al., 2015). More recently, however, it has become increasingly apparent that these measures may be unsuitable for revealing HA-related effects on higher level (e.g., cognitive-linguistic) processes (cf. Edwards, 2007). In the last few years, audiological researchers have therefore explored alternative measures in terms of their ability to fill this gap. For example, Desjardins and Doherty (2014) used a dual-task paradigm to investigate the effects of noise reduction (NR) processing on

the performance of HA users on a speech intelligibility task and a simultaneous visual tracking task. Results showed that NR processing increased (improved) the tracking speed in some conditions, while speech intelligibility remained unchanged. Also using a dual-task paradigm, Neher, Grimm, and Hohmann (2014) measured speech intelligibility and reaction times to visual stimuli for a group of hearing-impaired listeners with different HA conditions. They found that whereas speech

<sup>1</sup>Medizinische Physik and Cluster of Excellence Hearing4all, Oldenburg University, Germany

### Corresponding author:

Julia Habicht, Department of Medical Physics and Acoustics, Carl-von-Ossietzky University, D-26111 Oldenburg, Germany.  
Email: julia.habicht@uni-oldenburg.de



intelligibility was generally unaffected, visual reaction times increased (worsened) with reduced high-frequency audibility (HFA).

The results summarized earlier illustrate that HA-related factors such as NR or HFA can affect performance on outcome dimensions beyond *basic* speech intelligibility. However, the measures reviewed so far do not address the arguably most important communication outcome: speech comprehension. Speech comprehension requires piecing together the information carried by more or less intelligible speech segments in order to determine the *meaning* that they convey. Recently, Fontan, Tardieu, Gaillard, Woisard, and Ruiz (2015) found that speech intelligibility in noise was only weakly correlated with performance on a task that required participants to infer the meaning of the presented speech signals, leading to the conclusion that intelligibility scores are poor predictors of *real* communication outcomes. In view of this finding and the relative shortage of related research in audiology, the current study examined the effects of a number of HA-related factors using a novel test paradigm tapping into speech comprehension abilities.

Over the last decades, considerable efforts have been devoted to investigating speech (or language) comprehension based on different experimental approaches. For example, researchers have used event-related potentials and neuroimaging methods to find out when, where, and how speech comprehension takes place in the brain (e.g., Erb & Obleser, 2013; Federmeier, 2007; Peelle, Troiani, Grossman, & Wingfield, 2011; Robertson et al., 2000). In the field of psycholinguistics, investigators have assessed the accuracy with which participants can answer comprehension questions or how long they take to respond to them under different experimental conditions, for example, with linguistically simple or complex stimuli (Tun, Benichov, & Wingfield, 2010; Wingfield, McCoy, Peelle, Tun, & Cox, 2006; Wingfield, Peelle, & Grossman, 2003; Wingfield & Tun, 2007) or with and without normal hearing (McCoy et al., 2005; Tun et al., 2010; Wingfield et al., 2006). Another well-known psycholinguistic test method is the so-called *visual world paradigm* (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Huettig, Rommers, & Meyer, 2011; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In this paradigm, a visual scene composed of a number of objects is presented to the participant together with an acoustic stimulus that somehow describes the scene. The participant first scans the scene and then spontaneously starts looking at the object that seems to match the acoustic stimulus. During the stimulus presentation, the participant's eye movements are recorded with an eye tracker. This makes it possible to investigate the interplay between acoustical and visual information processing *online*,

that is, *during* the comprehension process. This is in contrast to *offline* tasks that require the participant to respond *after* the stimulus presentation (e.g., speech recognition measurements) and that therefore cannot reveal when precisely comprehension must have occurred. As a consequence, online tasks may be able to provide more detailed information about higher level perceptual and cognitive functions than offline tasks.

To investigate this possibility for potential applications in audiology, Wendt, Brand, and Kollmeier (2014) recently combined the recording of eye movements with a special speech corpus featuring different levels of linguistic complexity. The resultant paradigm can determine how quickly a participant grasps the meaning of an acoustic sentence-in-noise stimulus presented together with two similar pictures that either correctly (*target*) or incorrectly (*competitor*) depict the meaning conveyed by the sentence. Using an eye tracker, the time taken by the participant to start fixating the target (the *processing time*) is measured. Additionally, offline reaction times to the target stimuli are gathered by asking the participant to press a button as quickly as possible after the end of each sentence. In a first study based on this paradigm conducted with normal-hearing listeners, Wendt et al. (2014) examined the effects of linguistic complexity and found that while processing times increased with higher linguistic complexity, offline reaction times did not. In a follow-up study, Wendt, Kollmeier, and Brand (2015) also investigated the influence of hearing loss by testing age-matched groups of normal-hearing and hearing-impaired listeners. Consistent with their previous findings, they found that processing times increased with higher linguistic complexity, whereas offline reaction times did not. Furthermore, the group of hearing-impaired listeners had longer processing times than the normal-hearing controls. Concerning the offline reaction times, no effects of linguistic complexity were observed. However, there was a small but significant group difference, with the hearing-impaired listeners reacting more slowly to linguistically complex sentences compared with the normal-hearing controls.

Interestingly, in an ad hoc follow-up analysis, Wendt et al. (2015) observed that when the hearing-impaired group was subdivided into participants with ( $N=11$ ) and without ( $N=9$ ) HA experience the former had shorter (better) processing times than the latter, despite comparable stimulus audibility and speech recognition performance. This would seem to suggest that HA use may enable faster speech comprehension in noise and thus potentially better communication abilities in complex environments. Nevertheless, given the somewhat coincidental nature of this finding and the fact that Wendt et al. (2015) did not analyze their offline reaction times for such a group difference, its robustness remains

unclear. Therefore, one aim of the current study was to follow up on this initial finding by trying to confirm the putative difference in processing times between hearing-impaired listeners with and without HA experience.

Another aim of the current study was to assess the effects of a number of HA conditions on processing times. These HA conditions were chosen such that they resulted in different types of spectral shaping. At a more general level, this enabled us to shed more light on their perceptual consequences. At a more specific level, it allowed us to check for a potential interaction between HA condition and HA experience. In the study of Wendt et al. (2015), the stimuli were amplified according to the National Acoustic Laboratories-Revised (NAL-R) prescription formula (Byrne & Dillon, 1986). Because this type of amplification is clinically established and thus relatively common, Wendt et al. (2015) speculated that their experienced HA users could have been more familiar with it, thereby putting them at an advantage compared with the inexperienced users. Thus, by including less common types of spectral shaping in our study we wanted to check if the group difference in processing times would persist for them. Apart from NAL-R, we included another type of amplification and one type of single-microphone NR. The other type of amplification was based on the approach of Humes (2007), intended to provide *adequate* audibility by raising the long-term speech spectrum several decibels above hearing threshold in the high frequencies. In contrast to NAL-R, this type of spectral shaping is presumably rather unfamiliar to experienced HA users because it results in substantial HFA that is atypical of clinically fitted HAs. The single-microphone NR condition, which we tested in combination with NAL-R, was based on a state-of-the-art algorithm (Gerkmann & Hendriks, 2012) that gave rise to yet another spectral envelope characteristic.

A third and final aim of the current study was to compare processing times and offline reaction times in terms of their ability to reveal effects related to speech processing. As pointed out earlier, Wendt et al. (2014, 2015) did not observe any effects of linguistic complexity in their offline reaction times and did not check these data for a possible effect of HA experience. In the current study, we analyzed both types of data with respect to both types of effects. In this manner, we wanted to find out if the eye-tracking measurements are indeed more sensitive to such influences than the offline data.

## Methods

Ethical approval for all experimental procedures was obtained from the ethics committee of the University of Oldenburg (reference number Drs. 2/2015). Prior to any data collection, written informed consent was

obtained from all participants. The participants were paid on an hourly basis for their participation.

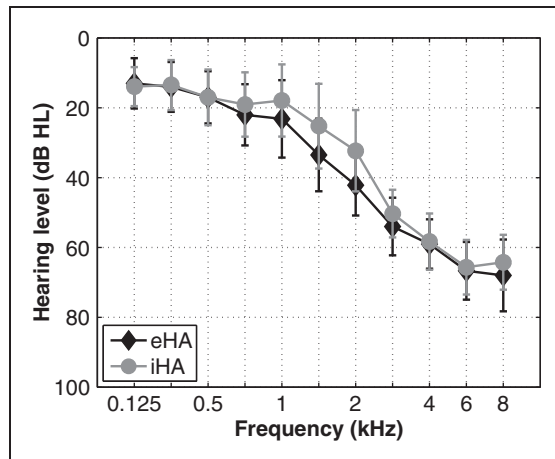
## Participants

A total of 30 participants were recruited from a large database of hearing-impaired listeners available at the Hörzentrum Oldenburg GmbH. Fifteen of the participants were habitual HA users with at least 1 year of HA experience (eHA group), whereas the other 15 participants had no previous HA experience (iHA group). None of the participants tested here had taken part in the study of Wendt et al. (2015). The two groups were matched closely in terms of age, pure-tone average hearing loss (PTA), and working memory capacity (see later). Inclusion criteria were as follows: (a) age from 60 to 80 years, (b) bilateral, sloping, sensorineural hearing loss in the range from 40 to 80 dB (HL) at 3 to 8 kHz, and (c) self-reported normal or corrected-to-normal vision. Because one participant from the iHA group could not perform the eye-tracking measurements reliably (see Preparatory Analyses section), he was excluded from the study, leaving 29 participants. Table 1 summarizes the main characteristics of the two groups of participants, while Figure 1 shows mean hearing threshold levels. On average, the eHA and iHA groups were 72 and 73 years of age. Average PTA as calculated across ears for the standard audiometric frequencies from 0.5 to 4 kHz ( $PTA_{0.5-4k}$ ) amounted to 35 and 31 dB HL for the eHA and iHA group, respectively. Corresponding values for the frequencies from 3 to 8 kHz ( $PTA_{3-8k}$ ) were 62 and 60 dB HL, respectively. Three independent  $t$  tests revealed no significant group differences in terms of age ( $t_{27}=0.7$ ,  $p>.5$ ),  $PTA_{0.5-4k}$  ( $t_{27}=-2.0$ ,  $p>.07$ ), and  $PTA_{3-8k}$  ( $t_{27}=-1.2$ ,  $p>.2$ ).

**Table 1.** Means (and Ranges) for the Age,  $PTA_{3-8k}$ ,  $PTA_{0.5-4k}$ , and Reading Span Data for the Two Groups of Participants. SRT80s for the Two Levels of Linguistic Complexity (Low, High) Are Also Shown.

	eHA	iHA
N	15	14
Age (year)	72 (60, 79)	73 (63, 78)
$PTA_{3-8k}$ (dB HL)	62 (57, 67)	60 (55, 65)
$PTA_{0.5-4k}$ (dB HL)	35 (31, 39)	31 (24, 38)
RS (%-correct)	44 (32, 50)	39 (26, 56)
SRT80,low (dB SNR)	-1.9 (-2.9, -0.2)	-1.3 (-3.2, -2.3)
SRT80,high (dB SNR)	-1.1 (-2.4, 0.4)	-0.6 (-3.0, 3.2)

PTA = pure-tone average hearing loss; RS = Reading Span; SNR = signal-to-noise ratio; eHA = group of experienced hearing aid users; iHA = group of inexperienced hearing aid users.



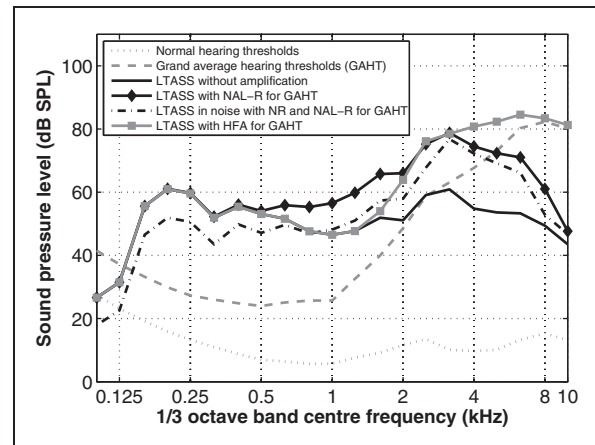
**Figure 1.** Mean hearing threshold levels averaged across both ears for the two groups of participants (eHA:  $N = 15$ ; iHA:  $N = 14$ ).

### Apparatus

All measurements took place in a soundproof booth. The (visual) working memory capacity measurements were performed with a laptop computer connected to an additional monitor. The speech recognition measurements were performed with a personal computer equipped with an Auritec (Hamburg, Germany) Earbox Highpower soundcard. For the offline reaction time and online processing time measurements, the setup of Wendt et al. (2014) was used (EyeLink 1000 desktop system, EyeLink CL high-speed camera, SR Research Ltd., Samsung 2253BW monitor). The visual stimuli were presented on a 22" multiscan color computer monitor with a resolution of  $1680 \times 1050$  pixels. Participants were seated such that their eyes were 60 cm in front of the monitor. Calibration of this setup was carried out at the start of each block of measurements (using a 9-point fixation stimulus procedure of the manufacturer of the eye tracker). Furthermore, before each stimulus presentation, participants had to fixate a point in the center of the screen for drift correction. The behavioral response (i.e., a button press) of the participant was collected using a hardware controller (Microsoft SideWinder Plug & Play). All acoustic stimuli were presented via free-field equalized Sennheiser (Wennebostel, Germany) HDA200 headphones. The acoustic stimuli were calibrated with a Brüel & Kjær (B&K; Nærum, Denmark) 4153 artificial ear, a B&K 4134  $\frac{1}{2}$ " microphone, a B&K 2669 preamplifier, and a B&K 2610 measurement amplifier.

### HA Conditions

The acoustic stimuli were processed using a simulated HA implemented on the Master Hearing Aid research



**Figure 2.** Illustration of the acoustic effects of the different hearing aid conditions. Gray dotted line: Hearing thresholds (in dB SPL) corresponding to normal hearing (i.e., 0 dB HL). Gray dashed line: Grand average hearing thresholds (in dB SPL) for the 30 participants of the current study. Black solid line without symbols: LTASS at the eardrum for an overall free-field level of 65 dB SPL ('baseline condition'). Black solid line with diamonds: Baseline condition with NAL-R amplification. Black dotted line without symbols: Long-term spectrum of speech in noise (each with a free-field level of 65 dB SPL) at the eardrum with NR and NAL-R amplification. Gray solid line with squares: Baseline condition with HFA. For the conditions including NAL-R or HFA, amplification was prescribed based on the grand average hearing threshold levels (gray dashed line).

platform (Grimm, Herzke, Berg, & Hohmann, 2006). All processing was carried out at a sampling rate of 16 kHz. Before presentation, the stimuli were resampled to 44.1 kHz.

**NAL-R.** To ensure comparability with the results of Wendt et al. (2015), we used NAL-R amplification as our reference condition. Figure 2 illustrates the use of NAL-R amplification for the sentence material used here. The gray dotted line corresponds to the hearing thresholds of a normal-hearing listener (i.e., thresholds of 0 dB HL) plotted in terms of 1/3-octave band sound pressure levels (SPLs). The gray dashed line corresponds to the grand average hearing thresholds (also in dB SPL) of the participants of the current study. The black solid line depicts the long-term average speech spectrum (LTASS) by adding the diffuse-field-to-eardrum transfer function specified in ANSI (1997) for an overall free-field level of 65 dB SPL (baseline condition). The black solid line with diamonds shows the baseline condition with NAL-R amplification prescribed for the grand average audiogram of our 30 participants.

**Single-microphone NR.** To assess the effects of NR on the offline reaction times and online processing times of our

participants, we applied a state-of-the-art single-microphone algorithm of Gerkmann and Hendriks (2012). This algorithm derives short-time Fourier transform-domain estimates of the noise power spectral density and the speech power by means of a speech presence probability estimator and temporal cepstrum smoothing (Breithaupt, Gerkmann, & Martin, 2008), respectively. The clean spectrum amplitude is then estimated from the speech power and the noise power estimates, after which the time-domain signal is resynthesized using a standard overlap-add procedure. In the current study, we set the maximum amount of attenuation to  $-9$  dB in order to minimize speech distortions. Furthermore, during stimulus generation we appended 2 s of noise to the beginning of each sentence-in-noise stimulus to allow the algorithm to adapt to the noise statistics and in this way to ensure a constant spectral shape afterwards. During stimulus presentation, we then discarded these 2-s noise segments. The black dotted line without symbols in Figure 2 shows the effects of the NR processing on the long-term spectrum of the speech signals used for the measurements (see later) mixed with stationary speech-shaped noise at signal-to-noise ratio (SNR) of 0 dB with NAL-R amplification prescribed for the grand average audiogram of our 30 participants at the eardrum. As shown, these manifested themselves primarily below about 3 kHz.

**High-frequency audibility.** As shown in Figure 2, NAL-R amplification provides only limited audibility at frequencies higher than about 2 kHz. As pointed out earlier, to ensure adequate audibility Humes (2007) proposed raising the LTASS 10–15 dB above hearing threshold from 0.25 through at least 4 kHz. In the current study, we followed this approach by making sure that the 1/3-octave band root mean square levels of our stimuli were at least 15 dB above threshold between 0.125 and 4 kHz. Because of hardware limitations coupled with the large dynamic range typical of human speech, we had to reduce our audibility criterion to at least 5 dB above hearing threshold at 6 kHz, and to at least 0 dB above hearing threshold at 8 kHz. The gray solid line with squares in Figure 2 shows the result of the HFA

condition on the LTASS for the grand average audiogram of our 30 participants at the eardrum.

### Working Memory Capacity

To characterize our participants in terms of cognitive function, we administered the reading span (RS) test of Carroll et al. (2015) to them. The RS test is a (visual) measure of working memory capacity that is rather widely used in hearing research (e.g., Lunner, 2003; Neher et al., 2014) and that includes three subtasks (Daneman & Carpenter, 1980). First, the participant has to read aloud sentence segments displayed successively on a screen. After the presentation of three such segments, there is a break of 1.75 s before the next presentation. The participant then has to answer “yes” if the three segments made up a meaningful sentence (e.g., “Das Mädchen–sang–ein Lied”; “The girl–sang–a song”) or “no” if the previous three segments did not make up a meaningful sentence (e.g., “Die Flasche–trank–Wasser”; “The bottle–drank–water”). After a block of three to six sentences, either *first words* or *final words* are displayed on the screen. The participant then has to repeat as many first or final words of the last block of sentences as possible in any order. Altogether, there were three training and 54 test sentences. As the performance measure, we used the percentage of correctly recalled target words across the 54 test sentences.

### Speech Recognition Measurements

Prior to the offline reaction time and processing time measurements, we assessed baseline speech recognition performance using the Oldenburg corpus of Linguistically and Audiologically Controlled Sentences (OLACS; Uslar et al., 2013). The OLACS consists of seven grammatically correct sentence structures with low semantic context that vary in linguistic complexity. For our measurements, we used two sentence structures (see Table 2): (a) subject–verb–object sentences with a canonical word order and therefore *low* linguistic complexity and (b) object–verb–subject sentences with a

**Table 2.** Example Sentences From the “Oldenburg Corpus of Linguistically and Audiologically Controlled Sentences” (Uslar et al., 2013) With Two Levels of Linguistic Complexity (Low, High).

Low	Der <sub>nom</sub>	müde <sub>nom</sub>	Drache	fesselt	den <sub>acc</sub>	großen <sub>acc</sub>	Panda.
	Meaning: “The tired dragon ties up the big panda.”						
High	Den <sub>acc</sub>	müden <sub>acc</sub>	Drachen	fesselt	der <sub>nom</sub>	große <sub>nom</sub>	Panda.
	Meaning: “The big panda ties up the tired dragon.”						

Note. In each case, the grammatically salient *word endings* and corresponding cases (Nom = Nominative; acc = Accusative) are indicated, as are the English meanings.

noncanonical word order and therefore *high* linguistic complexity. In each sentence, there are two characters (e.g., a dragon and a panda), one of which (the subject) performs a given action with the other (the object). In the German language, the linguistic complexity of these sentences is determined by relatively subtle grammatical or acoustic cues, for example, “*Der müde Drache fesselt den großen Panda*”; meaning: “The tired dragon ties up the big panda” (low linguistic complexity) versus “*Den müden Drachen fesselt der große Panda*”; meaning: “The big panda ties up the tired dragon” (high linguistic complexity).

The task of the participants was to repeat the words they had understood, which an experimenter then scored. Initially, a training measurement based on 40 sentences with low and high linguistic complexity in the reference (NAL-R) condition was performed to familiarize the participants with the sentences and the procedure. Using 40 additional sentences with low and high linguistic complexity, we then estimated the SNR corresponding to 80%-correct speech reception (SRT80) for the NAL-R condition. In each case, we presented stationary speech-shaped noise at a nominal level of 65 dB SPL. To assess the effects of NR and HFA on speech intelligibility, we also measured the SRT80s achievable with these two HA conditions using two subgroups of nine randomly chosen participants each.

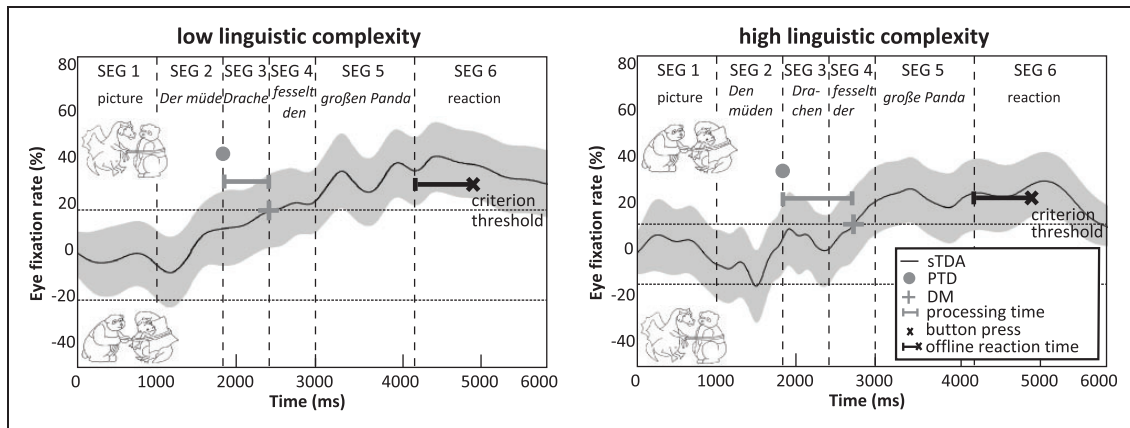
### Offline Reaction Time and Processing Time Measurements

**Acoustic stimuli.** Consistent with the speech recognition measurements, we used the sentences with low and high linguistic complexity of the OLACS as sentence material and presented them in stationary speech-shaped noise. In contrast to Wendt et al. (2015) who had tested their participants at individual SRT80s, we carried out the offline reaction time and processing time measurements at one (fixed) SNR. This had the advantage that we could test our HA conditions under acoustically equal conditions. Based on two similar sets of measurements carried out in our laboratory with comparable groups of participants we chose an SNR of 0 dB, as this corresponded to the grand average SRT80. Thus, for each participant, level of linguistic complexity, and HA condition we used a nominal presentation level of 65 dB SPL for both speech and noise. To estimate the offline reaction time and processing time achievable with a given HA condition, we presented a fixed set of 60 OLACS sentences (30 per level of linguistic complexity) in randomized order. The sentences ranged in length from 2.7 to 3.7 s. They were mixed with speech-shaped noise that started 200 ms before and ended 200 ms after each sentence.

**Visual stimuli.** As visual stimuli, we used the picture sets developed by Wendt et al. (2014) that complement the sentences of the OLACS. Each sentence has two corresponding picture sets, and each picture set consists of two similar pictures displayed next to each other. One of the pictures (the target) correctly depicts the situation conveyed by a given sentence. The other picture (the competitor) depicts the same situation but with interchanged roles of subject and object. In the first picture set, the left picture is the target; in the other picture set, the right picture is the target. A total of 60 picture sets, 30 per level of linguistic complexity, were used and presented with the corresponding sentence-in-noise stimuli.

**Combination of acoustic and visual stimuli and measurement of offline reaction times.** In the eye-tracking paradigm of Wendt et al. (2014), the acoustic and visual stimuli are combined as follows (see Figure 3). First, the visual stimulus is presented on its own for 1000 ms (Stimulus Segment 1). Next, the acoustic sentence-in-noise stimulus is added (Stimulus Segments 2–5). Following presentation of the acoustic sentence-in-noise stimulus, the participant has to identify the target picture by pressing one of three buttons on the hardware controller as quickly as possible (Stimulus Segment 6): a left button if the target picture appeared to be on the left-hand side, a right button if the target picture appeared to be on the right-hand side, or a key on top of the controller if neither picture appeared to match the spoken sentence. On each trial, the time taken to press a button relative to the end of the spoken sentence (the offline reaction time) was recorded (see Figure 3).

**Measurement of processing times.** Consistent with the method of Wendt et al. (2015), we proceeded as follows to estimate the processing times achievable with the three HA conditions and two levels of linguistic complexity. First, on the basis of the recorded eye-fixation data, we determined the so-called *single target detection amplitude* (sTDA) for each participant across all sentences of one level of linguistic complexity and one HA condition. The sTDA is a quantitative, normalized measure across time of the eye-fixation rate of a participant toward the target picture in relation to the eye-fixation rate toward the competitor picture and other regions on the screen; if at a given point in time the target picture is fixated more (across all sentences included in the sTDA estimation) than the competitor picture (or any other region on the screen), the sTDA is positive; if the competitor picture is fixated more the sTDA is negative. To illustrate, Figure 3(a) shows the sTDA of a given participant as a function of time for an example sentence with low linguistic complexity, while Figure 3(b) shows the sTDA for the corresponding sentence with high linguistic complexity. Since the sentence recordings differ in terms of their



**Figure 3.** Examples of single target detection amplitudes (quantitative tendency of eye fixations to the presented pictures; black line) for sentences with low linguistic complexity (left) and high linguistic complexity (right) over the course of the combined acoustic and visual stimulus presentation (segments 1–6). The shaded areas illustrate the 95% confidence interval. The gray dot denotes the point of target disambiguation (PTD), which defines the onset of the first word that allows matching the acoustic sentence to the target picture (upper picture). The gray + symbol denotes the DM, where the eye-fixation rate exceeds the criterion threshold. The horizontal gray bar defines the processing time, that is, the time difference between PTD and DM. The black × symbol denotes the time point of the behavioral response (i.e., button press), and the black bar defines the offline reaction time, that is, the time difference between the end of the sentence and the button press. Left: “Der müde Drache fesselt den großen Panda”; meaning: “The tired dragon ties up the big panda” (low linguistic complexity). Right: “Den müden Drachen fesselt der große Panda”; meaning: “The big panda ties up the tired dragon” (high linguistic complexity).

durations (see earlier), the eye fixations were time-aligned by segmenting them in a consistent fashion (Wendt et al., 2014). That is, Segments 1 to 5 were resampled to 100 samples each, while Segment 6 was resampled to 200 samples.

Second, on the basis of two specific sTDA values, we estimated the processing time. The first value is the so-called *point of target disambiguation* (PTD). The PTD corresponds to the onset of the first word that allows for disambiguation to occur, that is, the moment from which the target picture can in principle be identified by the participant. For the sentences used here, the PTD always corresponded to the start of stimulus Segment 3 (see Figure 3). The second specific sTDA value is the so-called *decision moment* (DM), which is defined as that point in time at which the sTDA exceeds a certain criterion threshold. In the study of Wendt et al. (2015), a fixed criterion threshold of 15% of the eye-fixation rate was used for all test conditions. In a recent follow-up analysis, however, it was found that the DM estimation can be made more reliable by using a relative criterion threshold corresponding to the 42%-point of the sTDA maximum of each test condition (Müller, Wendt, Kollmeier, & Brand, 2016). For our study, we therefore used this improved criterion threshold to estimate the DMs. The processing times were then derived by taking the difference (in milliseconds) between the PTDs and DMs. Using this approach, we estimated the processing time for each participant, level of linguistic complexity, and HA condition.

**Procedure.** The offline reaction time and processing time measurements were carried out at two visits (see Test Protocol section). Each visit started with a training block, consisting of 30 picture sets (i.e., one of the two picture sets per sentence) presented concurrently with the corresponding acoustic sentences in quiet, but with NAL-R amplification. After the training, eight test blocks consisting of 37 trials each were performed. Within a test block, the HA condition was kept constant. The level of linguistic complexity varied, however. Specifically, there were 30 trials based on 15 sentences with low and high linguistic complexity plus 7 catch trials. Two types of catch trials were used. First, either the target or competitor picture was depicted on both the left and the right side of the screen. Therefore, either both of the pictures matched the spoken sentence or neither did. These trials were included to force the participants to look at both pictures each time (and therefore to prevent them from developing alternative task-solving strategies). Second, two additional sentences with different structures (and thus levels of linguistic complexity) from the OLACS were included: subject-relative clauses (e.g., “Der Lehrer, der die Models bestiehlt, zittert”—“The teacher who is stealing from the models is shivering”) and object-relative clauses (e.g., “Der Maler, den die Vampire beschatten, gähnt”—“The painter whom the vampires are shadowing is yawning”). These types of sentences were included to prevent the participants from getting used to specific sentences structures, thereby forcing them to continuously attend

to them. The HA conditions were distributed across the second and third visit of each participant (see Test Protocol section). For each HA condition, four test blocks were carried out. These were presented in a row but in randomized order across participants. Across the two visits, each participant carried out 60 training trials and 592 test trials (37 trials  $\times$  4 test blocks  $\times$  [3 HA conditions + 1 retest]). One test block took about 7 min to complete. After three such blocks, participants had to take a break for at least 10 min.

### Test Protocol

Each participant attended three visits with a duration of about 1.5 h each. At the first visit, the SRT80s were measured, and the RS test was administered. At the second and third visit, the offline reaction times and processing times were measured. The NAL-R condition was tested twice per participant: once at the second visit (NAL-R<sub>test</sub>) and once at the third visit (NAL-R<sub>retest</sub>). The NR and HFA conditions were tested once. They were distributed across the two visits and participants in a balanced manner. Within a given visit, the different test conditions (i.e., levels of linguistic complexity and HA conditions) were presented in randomized order.

### Preparatory Analyses

Prior to any data analyses, we assessed the ability of our participants to identify the correct picture by determining the *picture* recognition rate for each level of linguistic complexity and HA condition (based on the offline reaction time data). Although the picture recognition rate cannot be compared directly with the SRT80, one would expect picture recognition rates of around 80% and higher due to the availability of both acoustic and visual information. All 29 participants achieved picture recognition rates of 80% and above (in contrast, the excluded participant had a grand average picture recognition rate of only 47%; see Participants section).

Furthermore, since the offline reaction times and processing times were logarithmically distributed, we applied a logarithmic transformation to them. We then used Shapiro–Wilk’s and Levene’s test to check the assumptions of normality and homogeneity of variance. In the case of the offline reaction times, both of these tests were significant ( $p < .05$ ) for some data subsets. Because no additional steps could remedy these issues, we analyzed these data using nonparametric statistics. In the case of the processing times, we identified six measurements (i.e.,  $2 \times$  NAL-R<sub>test</sub>,  $3 \times$  NAL-R<sub>retest</sub>, and  $1 \times$  HFA) stemming from five different participants that deviated by more than three times the interquartile range from the lower and upper quartiles of their respective datasets. Removal of these extreme values (Tukey,

1977) left us with complete datasets from 24 participants (eHA:  $N = 13$ ; iHA:  $N = 11$ ), which according to Shapiro–Wilk’s and Levene’s test fulfilled the requirements for normality and homogeneity of variance (all  $p > .05$ ). We therefore used parametric statistics to analyze these data further.

## Results

### Working Memory Capacity

The results of the RS test are summarized in Table 1. On average, the eHA and iHA groups could recall 44.2% and 39.4% of all target words, respectively. These results are in good agreement with other comparable studies (e.g., Arehart, Souza, Baca, & Kates, 2013; Neher et al., 2014). An independent  $t$  test revealed no significant difference in terms of RS between the two groups ( $t_{27} = 1.5$ ,  $p > .1$ ).

### Speech Recognition Measurements

Concerning the baseline speech recognition measurements (made with NAL-R), we obtained a grand average SRT80 of  $-1$  dB, thereby providing support for our chosen test SNR of 0 dB (see Acoustic Stimuli section). Furthermore, the two groups of participants achieved very similar SRT80s for sentences with both low ( $-1.9$  vs.  $-1.3$  dB) and high ( $-1.1$  vs.  $-0.6$  dB) linguistic complexity (see Table 1). Two independent  $t$  tests confirmed the lack of any significant differences among the eHA and iHA groups for the two sentence types (both  $t_{27} < 1.4$ ,  $p > .2$ ).

Concerning the effects of HFA and NR on speech intelligibility, we analyzed the SRT80s measured with the two subgroups of participants (recall that for each of these two HA conditions, nine randomly chosen participants were tested; see Speech Recognition Measurements section). For sentences with low linguistic complexity, the mean SRT80s for the HFA and NR conditions were  $-2$  and  $-1.8$  dB, respectively; for sentences with high linguistic complexity, they were both  $-1.3$  dB. Two paired  $t$  tests showed no differences in terms of the speech recognition performance achievable with HFA and NR relative to NAL-R (both  $t_8 < 1.3$ , both  $p > .05$ ).

Taken together, these measurements illustrate the absence of any differences in basic speech recognition performance between our two groups of participants or between our three HA conditions.

### Offline Reaction Times

Table 3 (left) shows median offline reaction times for the different levels of linguistic complexity, HA conditions, and listener groups. To analyze the effect of linguistic



**Table 3.** Median Offline Reaction Times (Left) and Median Processing Times (Right) With Lower and Upper Quartiles in Parentheses for the Different Levels of Linguistic Complexity (Low, High), HA Conditions (NAL-R<sub>test</sub>, NAL-R<sub>retest</sub>, NR, HFA), and Listener Groups (eHA, iHA).

	Offline reaction times (ms)		Processing times (ms)	
	Low	High	Low	High
NAL-R <sub>test</sub>				
eHA	880 (746, 1967)	1042 (702, 1559)	633 (570, 737)	866 (756, 1142)
iHA	1279 (805, 2011)	1261 (801, 2303)	776 (532, 990)	1443 (815, 2098)
NAL-R <sub>retest</sub>				
eHA	911 (865, 1486)	1096 (987, 1493)	633 (357, 763)	899 (570, 1179)
iHA	1057 (636, 2218)	1084 (640, 2392)	646 (413, 931)	1139 (763, 1569)
NR				
eHA	938 (660, 1294)	1073 (747, 1394)	704 (517, 1022)	821 (659, 1161)
iHA	1094 (786, 2004)	1221 (753, 2051)	782 (633, 1271)	1328 (828, 1868)
HFA				
eHA	1068 (762, 1481)	1026 (752, 1750)	665 (529, 730)	931 (646, 1126)
iHA	1141 (722, 2397)	1186 (719, 1962)	795 (704, 1061)	1297 (886, 1811)

NAL-R = National Acoustic Laboratories-Revised; NR = noise reduction; HFA = high-frequency audibility; eHA = group of experienced hearing aid users; iHA = group of inexperienced hearing aid users.

complexity (medians: 966 and 1068 ms for low and high complexity, respectively), we made use of the Wilcoxon signed-rank test, which was not significant ( $Z = -1.0$ ,  $p > .3$ ). To investigate the effect of HA condition (medians: 1039, 1089, 1005, and 1047 ms for NAL-R<sub>test</sub>, NAL-R<sub>retest</sub>, NR, and HFA, respectively), we conducted a Friedman's test, which was not significant either ( $\chi^2_{(54)} = 0.27$ ,  $p > .9$ ). Finally, to analyze the effect of listener group (medians: 962 and 1232 ms for eHA and iHA, respectively), we made use of the Mann-Whitney  $U$  test, which was also nonsignificant ( $U = -1.12$ ,  $p > .2$ ).

### Processing Times

Table 3 (right) shows median processing times for the different levels of linguistic complexity, HA conditions, and listener groups. Consistent with our expectations, both listener groups had longer (poorer) processing times for the sentences with high linguistic complexity than for the sentences with low linguistic complexity. Furthermore, the iHA group had longer (poorer) processing times than the eHA group.

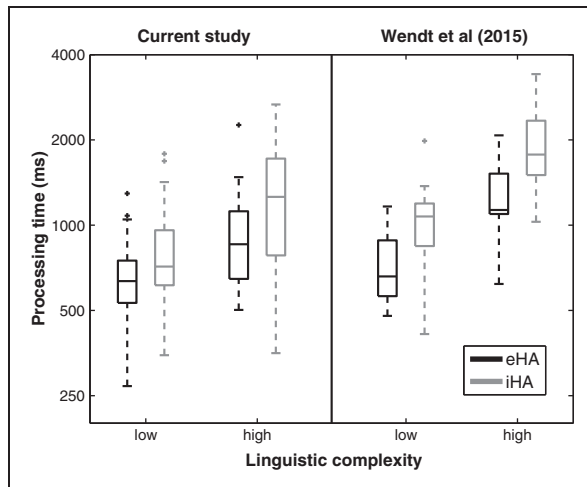
To analyze the processing times further, we performed a mixed-model analysis of variance on the log-transformed data with listener group (eHA, iHA) as a between-subject factor, and linguistic complexity (low, high) and HA condition (NAL-R<sub>test</sub>, NAL-R<sub>retest</sub>, NR, HFA) as within-subject factors. This analysis revealed significant effects of listener group ( $F(1, 22) = 6.2$ ,  $p < .02$ ,  $\eta_p^2 = 0.22$ ), linguistic complexity ( $F(1, 22) = 36.4$ ,  $p < .00001$ ,  $\eta_p^2 = 0.62$ ), and HA condition ( $F(1, 66) = 3.7$ ,  $p < .015$ ,  $\eta_p^2 = 0.15$ ). No interactions were

observed (all  $p > .5$ ). To follow up on the significant effect of HA condition, we carried out a post hoc analysis with Bonferroni correction. This revealed that NAL-R<sub>retest</sub> differed significantly from each of the other three HA conditions (all  $p < .05$ ). This finding is indicative of a training effect, corresponding to an improvement in processing time of  $\sim 70$  ms across the two sets of measurements (medians: 798 and 730 ms for NAL-R<sub>test</sub> and NAL-R<sub>retest</sub>, respectively).

Because removal of some extreme data points had reduced the size of our dataset to  $N = 24$  (see Preparatory Analyses section), we decided to follow the aforementioned analysis up with a confirmatory one. To that end, we averaged across the two NAL (test and retest) measurements of the 24 participants and combined the resultant data with those of the participants with a single remaining NAL (test or retest) measurement. In this manner, we could increase the size of our dataset to  $N = 2 \times 14$ . A mixed-model analysis of variance performed on these data confirmed the effects of listener group ( $F(1, 26) = 4.5$ ,  $p < .043$ ,  $\eta_p^2 = 0.15$ ) and linguistic complexity ( $F(1, 26) = 40.0$ ,  $p < .00001$ ,  $\eta_p^2 = 0.61$ ) found previously. No effect of HA condition was found, and neither were there any interactions (all  $p > .05$ ). Figure 4 (left) shows box plots of the processing times for the two levels of linguistic complexity and listener groups ( $N = 2 \times 14$ ), averaged across HA condition.

### Discussion

The main purpose of the current study was to investigate whether HA experience may offer an advantage in terms



**Figure 4.** Left: Box plots of the processing time measurements from the current study for the two levels of linguistic complexity (low, high) and participant groups (eHA:  $N = 14$ ; iHA:  $N = 14$ ), averaged across the different HA conditions (NAL-R<sub>mean</sub>, NR, HFA). Horizontal bars correspond to medians, boxes to lower and upper quartiles, and whiskers to one and a half times the interquartile range from the lower and upper quartiles. Right: Corresponding data from Wendt et al. (2015) obtained with NAL-R (eHA:  $N = 11$ ; iHA:  $N = 9$ ).

of grasping the meaning of acoustic sentence-in-noise stimuli presented together with two pictures that either correctly or incorrectly depict the meaning of the sentences. To that end, we measured processing times and offline reaction times using the eye-tracking paradigm of Wendt et al. (2014). Participants were closely matched groups of experienced and inexperienced HA users. Three HA conditions (NAL-R, NR, and HFA) resulting in rather different spectral envelope characteristics (see Figure 2) were tested. They were included here to check for a potential interaction with HA experience (or participant group) as well as to shed more light on their perceptual effects in general. Data analyses revealed no effects in the offline reaction times. In contrast, the expected effects of linguistic complexity and listener group were observed in the processing times (see Figure 4), whereas HA condition did not influence these results. In the following sections, we discuss our findings in more detail.

### Effects of HA Experience

The fact that we observed an effect of participant group (and linguistic complexity) on processing times replicates the earlier observation of Wendt et al. (2015) that HA experience seems to enable shorter processing times (and that higher linguistic complexity leads to longer processing times). Not only did we find this effect for the

NAL-R condition previously tested by Wendt et al. (2015), we also found it for the HFA and NR conditions included here. For comparison, Figure 4 (right) shows the processing time data from the study of Wendt et al. (2015), illustrating a good correspondence with our dataset. Nevertheless, the processing times that we measured were generally shorter than those of Wendt et al. A possible reason for this could be differences between the cohorts tested in the two studies (e.g., in terms of duration of auditory deprivation or HA use). Another reason could be that Wendt et al. (2015) included an additional sentence type with an ambiguous object-verb-subject structure (e.g., “*Die* liebe Königin grüßt *der* kleine Junge”; meaning: “The little boy greets the nice queen”). With this type of sentence, the first article “*Die*” can indicate either a subject or an object, and this ambiguity is first resolved by the article of the second noun (“*der*”). Therefore, disambiguation of the target occurs later for this sentence type than for the two sentence types included here. This could have forced the participants of Wendt et al. (2015) to be more cautious, thereby resulting in longer processing times compared with our study.

Altogether, these findings support the idea that HA experience positively influences speech comprehension abilities. That having been said, it remains unclear whether the observable across-group difference in processing times was due to acclimatization to HAs *per se* or due to other underlying factors. Until now, the research literature is divided with regard to the existence of acclimatization effects (see reviews by Munro, 2008; Palmer, Nelson, & Lindley, 1998; Turner & Bentler, 1998). That is, while some studies found acclimatization effects (e.g., Munro & Lutman, 2003) others did not (e.g., Humes & Wilson, 2003). At a more general level, some recent studies have provided indications that HA use can have a positive influence on higher level cognitive processes. For example, Dawes et al. (2015) investigated statistical relations between hearing loss, cognitive performance, and HA use using data from a large British database. They observed that HA use was associated with better cognition, even when known psychological side effects of hearing impairment such as social isolation and tendency to depression were partialled out. Recently, Amieva et al. (2015) also investigated statistical relations between cognition (as assessed via the Mini-Mental State Examination), self-reported hearing loss, and HA use over the course of 25 years. They reported an association between HA use and reduced cognitive decline. Given the indications from (and limitations of) these studies, it would be prudent to conduct a longitudinal study with the aim of confirming the putative link between auditory acclimatization to HAs and faster processing times as measured with our eye-tracking paradigm.

## Effects of HA Condition

With the exception that NAL- $R_{\text{retest}}$  led to significantly shorter processing times than NAL $_{\text{test}}$ , NR, and HFA, there were no differences among our chosen HA conditions—neither in the processing times nor in the offline reaction times. A number of recent studies have reported results in support of the idea that NR (e.g., Desjardins & Doherty, 2014; Houben, van Doorn-Bierman, & Dreschler, 2013; Sarampalis, Kalluri, Edwards, & Hafter, 2009) and HFA (McCreery & Stelmachowicz, 2013; Neher et al., 2014) can influence the speed with which different stimuli can be processed or responded to by a participant. One explanation for the lack of any such effects in our study could be the training effect that we observed (see earlier). This could have overshadowed the influence of NR and HFA. In future work, care should therefore be taken to train participants more on the eye-tracking paradigm. Another explanation for the divergent results could be the large number of methodological differences across studies, for example, in terms of experimental measures, HA conditions, and participants used. In this context, we consider it likely that our eye-tracking paradigm taps into another domain (i.e., that of online speech comprehension) compared with more indirect timed measures of speech processing used by others, for example, verbal processing times to speech stimuli (McCreery & Stelmachowicz, 2013) or reaction times to visual distractor stimuli unrelated to the acoustic speech signals (Neher et al., 2014; Sarampalis et al., 2009). To clarify if these measures indeed tap into different speech processing domains, future research would need to compare them directly using the same participants and HA conditions.

## Offline Reaction Times Versus Processing Times

In accordance with Wendt et al. (2014, 2015), we found no effects of linguistic complexity in our offline reaction times, whereas processing times were affected by it. The same was true for the effects of HA experience. As pointed out in the Introduction section, psycholinguists have previously found effects of linguistic complexity in timed responses to comprehension questions for listeners with and without hearing loss (e.g., Tun et al., 2010; Wingfield et al., 2003). Because these studies did not investigate potential effects of HA experience, it is currently unclear if timed responses to comprehension questions or other related measures would be sensitive to them and if they therefore could constitute viable experimental alternatives to our rather elaborate eye-tracking paradigm. Future work should ideally investigate this issue.

Taken together, our results indicate that processing times are more sensitive to acoustic and perceptual influences on speech processing than offline reaction times.

## Summary and Conclusions

In the current study, we examined the effects of HA experience on offline reaction times and processing times for complex sentence-in-noise stimuli using the eye-tracking paradigm of Wendt et al. (2014). We found that elderly hearing-impaired listeners without any HA experience had significantly longer (poorer) processing times than listeners matched in terms of age, high-frequency PTA, and working memory capacity who had at least 1 year of HA experience. This group difference was stable across three acoustically different HA conditions, which each gave rise to comparable processing times. No effects of either group or HA condition were observed in the offline reaction times.

Because our results were obtained using an across-group design, it is still unclear whether the group difference can indeed be traced back to a lack of auditory stimulation for the hearing-impaired listeners without prior HA experience. Follow-up research should therefore confirm the seemingly positive effect of HA experience on processing times using a longitudinal study design.

## Acknowledgments

The authors thank Jana Müller (Oldenburg University) and Dorothea Wendt (Technical University of Denmark) for their help with the eye-tracking paradigm.

## 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 funded by the DFG Cluster of Excellence EXC 1077/1 “Hearing4all.”

## References

- 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.
- Amieva, H., Ouvrard, C., Giulioli, C., Meillon, C., Rullier, L., Dartigues, J. F. (2015). Self-reported hearing loss, hearing aids, and cognitive decline in elderly adults: A 25-year study. *Journal of the American Geriatrics Society, 63*(10), 2099–2104.
- ANSI. (1997). S3. 5-1997, methods for the calculation of the speech intelligibility index. *New York: American National Standards Institute, 19*, 90–119.
- Arehart, K. H., Souza, P., Baca, R., & Kates, J. M. (2013). Working memory, age and hearing loss: Susceptibility to hearing aid distortion. *Ear and Hearing, 34*(3), 251.

- Breithaupt, C., Gerkmann, T., & Martin, R. (2008). *A novel a priori SNR estimation approach based on selective cepstral-temporal smoothing*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, IEEE.
- Byrne, D., & 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), 257–265.
- Carroll, R., Meis, M., Schulte, M., Vormann, M., Kießling, J., Meister, H. (2015). Development of a German reading span test with dual task design for application in cognitive hearing research. *International Journal of Audiology*, 54(2), 136–141.
- 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.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Dawes, P., Emsley, R., Cruickshanks, K. J., Moore, D. R., Fortnum, H., Edmondson-Jones, M., ... Munro, K. J. (2015). Hearing loss and cognition: The role of hearing aids, social isolation and depression. *PloS One*, 10(3), e0119616.
- Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and Hearing*, 35(6), 600–610.
- Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, 11(1), 31–46.
- Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience*, 7, 116. doi: 10.3389/fnsys.20.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505.
- Fontan, L., Tardieu, J., Gaillard, P., Woisard, V., & Ruiz, R. (2015). Relationship between speech intelligibility and speech comprehension in babble noise. *Journal of Speech, Language, and Hearing Research*, 58(3), 977–986.
- Gerkmann, T., & Hendriks, R. C. (2012). Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(4), 1383–1393.
- Grimm, G., Herzke, T., Berg, D., & Hohmann, V. (2006). The master hearing aid: A PC-based platform for algorithm development and evaluation. *Acta Acustica United With Acustica*, 92(4), 618–628.
- Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology*, 52(11), 753–761.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology*, 18(7), 590–603.
- Humes, L. E., & Wilson, D. L. (2003). An examination of changes in hearing-aid performance and benefit in the elderly over a 3-year period of hearing-aid use. *Journal of Speech, Language, and Hearing Research*, 46(1), 137–145.
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Usklar, V., Brand, T., ... Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology*, 54, 1–14.
- Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, 42, S49–S58.
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A*, 58(1), 22–33.
- McCreery, R. W., & Stelmachowicz, P. G. (2013). The effects of limited bandwidth and noise on verbal processing time and word recall in normal-hearing children. *Ear and Hearing*, 34(5), 585.
- Müller, J., Wendt, D., Kollmeier, B., & Brand, T. (2016). Comparing eye tracking with electrooculography for measuring individual sentence comprehension duration. *submitted to PloS one*.
- Munro, K. J. (2008). Reorganization of the adult auditory system: Perceptual and physiological evidence from monaural fitting of hearing aids. *Trends in Amplification*, 12(3), 254–271.
- Munro, K. J., & Lutman, M. E. (2003). The effect of speech presentation level on measurement of auditory acclimatization to amplified speech. *The Journal of the Acoustical Society of America*, 114(1), 484–495.
- Neher, T., Grimm, G., & Hohmann, V. (2014). Perceptual consequences of different signal changes due to binaural noise reduction: Do hearing loss and working memory capacity play a role? *Ear and Hearing*, 35(5), e213–e227.
- Palmer, C. V., Nelson, C. T., & Lindley, G. A. (1998). The functionally and physiologically plastic adult auditory system. *The Journal of the Acoustical Society of America*, 103(4), 1705–1721.
- Peelle, J. E., Troiani, V., Grossman, M., & Wingfield, A. (2011). Hearing loss in older adults affects neural systems supporting speech comprehension. *The Journal of Neuroscience*, 31(35), 12638–12643.
- Robertson, D. A., Gernsbacher, M. A., Guidotti, S. J., Robertson, R. R., Irwin, W., Mock, B. J., ... Campana, M. E. (2000). Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. *Psychological Science*, 11(3), 255–260.
- 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(5), 1230–1240.
- 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.

- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass: Addison-Wesley.
- Tun, P. A., Benichov, J., & Wingfield, A. (2010). Response latencies in auditory sentence comprehension: Effects of linguistic versus perceptual challenge. *Psychology and Aging, 25*(3), 730.
- Turner, C. W., & Bentler, R. A. (1998). Does hearing aid benefit increase over time? *The Journal of the Acoustical Society of America, 104*(6), 3673–3674.
- Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., ... Kollmeier, B. (2013). Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test. *The Journal of the Acoustical Society of America, 134*(4), 3039–3056.
- 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. doi:100110.101371/journal.pone.0100186.
- 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*. doi:10.1177/2331216515584149.
- Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & 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), 487–497.
- Wingfield, A., Peelle, J. E., & Grossman, M. (2003). Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *Aging, Neuropsychology, and Cognition, 10*(4), 310–322.
- Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology, 18*(7), 548–558.