Revealing Perceptual and Cognitive Mechanisms in Static and Dynamic Cocktail Party Listening by Means of Error Analyses

Moritz Wächtler[1] (iD), Josef Kessler[2], Martin Walger[1,3] and Hartmut Meister[1]

Abstract

In cocktail party situations multiple talkers speak simultaneously, which causes listening to be perceptually and cognitively challenging. Such situations can either be static (fixed target talker) or dynamic, meaning the target talker switches occasionally and in a potentially unpredictable way. To shed light on the perceptual and cognitive mechanisms in static and dynamic cocktail party situations, we conducted an analysis of error types that occur during a multi-talker speech recognition test. The error analysis distinguished between misunderstood or omitted words (random errors) and target-masker confusions. To investigate the effects of aging and hearing impairment, we compared data from three listener groups, comprised of young as well as older adults with and without hearing loss. In the static condition, error rates were generally very low, except for the older hearing-impaired listeners. Consistent with the assumption of decreased audibility, they showed a notable amount of random errors. In the dynamic condition, errors increased compared to the static condition, especially immediately following a target talker switch. Those increases were similar for random and confusion errors. The older hearing-impaired listeners showed greater difficulties than the younger adults in trials not preceded by a switch. These results suggest that the load associated with dynamic cocktail party listening affects the ability to focus attention on the talker of interest and the retrieval of words from short-term memory, as indicated by the increased amount of confusion and random errors. This was most pronounced in the older hearing-impaired listeners proposing an interplay of perceptual and cognitive mechanisms.

Introduction

In everyday life, people often listen to a talker while others speak at the same time. Based on the seminal work by Cherry (1953), this is commonly referred to as a "cocktail party situation". Competing speech can not only reduce audibility by energetic masking (EM), but may also cause "informational masking" (IM). IM refers to the fact that competing speech might contain irrelevant and potentially distracting information which has to be ignored. The degree of IM depends on the similarity of the target and masker as well as stimulus uncertainty (cf. Kidd et al., 2008) and is closely related to attentional processes (Shinn-Cunningham, 2008). As a consequence, cocktail party listening poses both auditory and cognitive challenges, and may thus be especially detrimental to older individuals or listeners with hearing impairment (e.g., Helfer & Freyman, 2008).

In a cocktail party situation, listeners are exposed to a multitude of simultaneous sounds with overlapping frequency spectra. According to the model of auditory scene analysis (Bregman, 1990), at any given point in time, listeners have to group the individual spectral components of these sounds based on the source from which they originate. As described in the review by Bronkhorst (2015), a "primitive" grouping of speech sounds is enabled by different cues, some of which are characteristics of the voice itself such as the

[1]Jean-Uhrmacher-Institute for Clinical ENT-Research, University of Cologne, Cologne, Germany
[2]Department of Neurology, University Hospital of Cologne, Cologne, Germany
[3]Clinic of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Cologne, Germany

Corresponding Author:
Moritz Wächtler, Jean-Uhrmacher-Institute for Clinical ENT-Research, University of Cologne, Geibelstr. 29-31, D-50931 Cologne, Germany.
Email: moritz.waechtler@uhrmacherinstitut.de

fundamental frequency and vocal tract length. Other cues like interaural time and level differences (ITDs, ILDs) arise when speech sounds arrive at the listener's head from different directions. Bronkhorst (2015) views this primitive grouping as the preattentive segregation and streaming of sounds. Subsequently, attention comes into play and allows the selection of parts of the incoming auditory information (Shinn-Cunningham, 2008). In line with Bregman's concept of scene analysis, this mechanism enables the listener to focus their attention on the talker of interest and to suppress competing speech (cf. Alain & Arnott, 2000).

In daily communication situations the talker of interest often changes in a possibly unpredictable manner. Thus, one may distinguish between "static" and "dynamic" cocktail party listening (Brungart & Simpson, 2007; Lin & Carlile, 2015; Meister et al., 2020). Here, dynamic refers to the stimulus presentation but does not assume movements of the listener or the sound sources. Whereas static listening entails focusing attention on a target talker known in advance, dynamic situations require to monitor multiple sound sources due to the ubiquitous stimulus uncertainty, and to occasionally switch attention from one talker to another. For this reason, dynamic cocktail party situations are associated with an additional cognitive load (Lin & Carlile, 2015), which typically results in decreased speech recognition performance compared to static situations (Brungart & Simpson, 2007).

Several studies have investigated dynamic listening situations in the context of aging. Singh et al. (2013) presented sentences of the coordinate response measure (CRM, Bolia et al., 2000) with three competing talkers at different locations. They demonstrated similar speech recognition performance in young and older normal-hearing listeners, when the target was cued using a call-sign. In a second experiment, the task was slightly changed, requiring listeners to repeat back the words from the talker opposite the location indicated by the call-sign. In comparison to the first experiment, the findings revealed significantly decreased performance in older—but not in younger—listeners, indicating that difficulties due to temporarily misdirecting attention may partly explain age-related problems in dynamic cocktail party situations.

Getzmann et al. (2015) compared younger and older adults with normal hearing using a setup with four competing talkers, which required the listeners to switch attention based on a predefined call-sign. While behavioral results only showed slightly different performance in the young and older listeners, analysis of event-related potentials suggested deficits that are potentially based on less flexible inhibitory control and increased distraction in older adults.

This suggestion is in line with a study by Oberem et al. (2017) who asked younger and older subjects with normal hearing to recognize spoken digits from occasionally changing locations indicated by a visual cue in the presence of a distractor talker. While higher reaction times and error rates were observed in the older listeners, the influence of target location changes was similar for both age groups. The authors concluded that switching auditory attention is not affected by aging, but ignoring competing speech is more challenging for older listeners.

In order to shed more light on whether the ability to switch attention or to monitor multiple sources of information is subject to age-related changes in dynamic cocktail party listening, Meister et al. (2020) analyzed "general" and "specific" switch costs. Specific costs, which reflect the performance decrease associated with a target talker change, did not show significant differences between groups of normal-hearing young and older listeners. In contrast, a significant difference was found for general costs that reflect the load associated with monitoring multiple talkers when no a priori target information is given.

Finally, Wächtler et al. (2021) compared older listeners with and without hearing impairment in static and dynamic cocktail party scenarios using matrix sentences presented from three different locations. The dynamic scenarios considered either a variable or a fixed target voice as well as probabilities of 20% or 100% target switches across a test list. In order to characterize problems in dynamic listening, the authors calculated costs by comparing performance in the static and dynamic condition. Notably, a significant difference in costs between the listeners with and without hearing impairment could only be shown for the condition with a low switching probability of a variable target voice. In contrast, costs were similar in both groups when the target voice was kept constant or the target talker changed continually.

Taken together, these findings give some evidence that dynamic cocktail party listening is associated with greater difficulties for older than for younger listeners. In comparison to static situations in which focusing attention on a pre-known target is important, dynamic situations cause a higher cognitive load due to the need to monitor different talkers and to switch attention from one source to another. The studies presented above are not fully conclusive regarding the extent to which inhibitory control, switching attention and monitoring multiple information sources underlie age-related changes. There is, however, some evidence that monitoring multiple sources, which is vital in situations of stimulus uncertainty, shows some decline with increasing age. In contrast, switching attention seems to be unaffected by aging, unless the task poses high demands on cognitive processes (e.g., when attention is momentarily misdirected and must be deliberately refocused, Singh et al., 2013).

Since the mechanisms behind static and dynamic cocktail party listening are still unclear, a more detailed analysis of the outcome, such as determining different error patterns that occur in multi-talker listening, may be beneficial. This could also be helpful for distinguishing peripheral effects (such as availability of grouping cues, hearing loss) from cognitive aspects (such as attention).

Basically, two different types of errors can be distinguished: First, the target may be confused with one of the competing talkers. In terms of the "primitive" grouping mechanism described by Bronkhorst (2015), confusion errors might reflect the ability to assign the words to the correct auditory stream based on the auditory cues provided. For instance, if spatial cues are not very salient, such as with small ITDs or ILDs, the resulting target-masker similarity will increase IM and thus the likelihood to confuse target and masker. On an attentional level, confusions might reflect the ability to focus on the target and to suppress the competing information. Alternatively, it is conceivable that words are assigned to the correct auditory streams during perception, but are confused during the "post-stimulus decision phase" (Lin & Carlile, 2015), e.g. when they are recalled from short-term memory.

Second, unsystematic errors ("random" errors) might occur, including misunderstood or omitted words. Random errors might reflect EM or effects of hearing loss. For instance, restricted audibility will increase the likelihood of misperception and omissions. Furthermore, random errors may also reflect the inability to retrieve the information from short-term memory—possibly as a result of cognitive load (Lin & Carlile, 2015). Consistent with limited capacity theory (Kahneman, 1973), it has been shown that cognitive load affects recall (Barrouillet et al., 2007; Camos & Portrat, 2015), potentially increasing omissions or misperceived words (Mattys & Palmer, 2015).

Several studies have carried out such error analyses in the framework of dynamic cocktail party listening. Using the CRM Test (Bolia et al., 2000), which provides a call-sign as well as a color and number target word, Kidd et al. (2005) presented young normal-hearing listeners with three competing talkers at different positions. Presentations varied in terms of uncertainty regarding the position of the target talker who was indicated by the call-sign. As expected, errors grew with increasing uncertainty. The most frequently observed error was a confusion of both the number and color word from one and the same competing talker, indicating that attention was misdirected on the wrong source.

Using the CRM Test with two competing talkers and an error analysis similar to that of Kidd et al. (2005), Ihlefeld and Shinn-Cunningham (2008) investigated the reasons that underlie the speech recognition benefits that can be achieved by spatially separating competing talkers. They found that when the location of the target talker is cued, spatial separation reduced "masker errors" (all reported words from masker) and "mix errors" (mixing words from target and masker). This suggests that in this case spatial separation helps both focusing attention on the target (object selection) as well as linking words of a talker over time and thus forming a stream (object formation). When only the voice characteristic of the target but not its location is indicated, spatial separation only reduced the "mix errors" which suggests that in this case only the formation of auditory objects can be improved.

The study by Lin and Carlile (2015) examined errors during switch and non-switch trials using three simultaneous matrix sentences recorded from the same talker that were presented from different locations. The sentence of interest was indicated by the first word (name), and switches comprised changes in the spatial position of the target. The authors contrasted confusions and "passes", the latter specifying that single words were not repeated back. Confusions were taken as an indication of IM, whereas passes were attributed to failures in word recall, due to the high cognitive load caused by target talker changes. They found that both the probability of confusion errors and passes were significantly increased after switches of the target when compared to non-switch trials. Thus, both misdirecting attention as well as failures in recall appear to play a role in dynamic multi-talker listening.

The mechanisms behind misdirected attention were further examined in the follow-up study by Lin and Carlile (2019) using a setup with two co-located competing talkers with different voices. Two types of switches were used: either the previous target voice became the masker in the next trial ("old voice switch"), or it was replaced with a new voice ("new voice switch"). Relative to the no-switch conditions, they found a significantly higher amount of confusions for old voice switches, but not for new voice switches. This finding suggests that not only refocusing attention to a new target talker, but also disengaging attention from the preceding target can be challenging, as listeners may be biased by the voice of the previous talker (cf. Best et al., 2008).

These error analyses give valuable information on mechanisms of dynamic cocktail party listening. However, to our knowledge, such analyses only included data from young normal-hearing listeners and are thus not informative regarding possible effects associated with age and hearing loss. Additionally including older and hearing impaired listeners would be of particular interest due the possible interplay between sensory and cognitive factors. For example, the ELU-model (Rönnberg et al., 2013) and FUEL-model (Pichora-Fuller et al., 2016) suggest that speech degraded by an impaired auditory periphery causes perceptual load. As a consequence, top-down compensation processes are needed, resulting in an additional cognitive load. Given the idea that cognitive capacity is limited (Kahneman, 1973), hearing loss and aging could therefore reduce the amount of cognitive resources available for other tasks like switching attention, possibly making dynamic situations especially problematic and effortful. Thus, the aim of the present examination is to extend error analyses in multi-talker situations to different listener groups and to shed light on age-related mechanisms. Therefore, groups of young and older listeners with normal hearing as well as older listeners with typical mild-to-moderate hearing loss were examined in static and dynamic cocktail party situations with three competing talkers (Meister et al., 2020; Wächtler et al., 2021). In the

static conditions, a priori knowledge about the target talker was given, minimizing stimulus uncertainty and cognitive load. In the dynamic conditions, the target talker was unpredictable and could change from trial to trial constituting high cognitive load. Confusion and random errors were determined in both static and dynamic conditions. Based on the distribution of the errors, an attempt was made to discuss different mechanisms considering the cognitive load applied.

We hypothesized that older listeners would show more confusions than the young individuals due to problems with inhibitory control, especially in the demanding dynamic situation. Moreover, we expected hearing loss to increase random errors as a consequence of audibility in the static situation and due to the combined perceptual and cognitive load in the dynamic situation. In addition, we anticipated age-related differences in error rates to be relatively small for trials following a switch, since the ability to refocus attention seems to be minimally affected by aging. By contrast, we expected non-switch trials, in which target uncertainty and thus monitoring multiple talkers poses the main difficulty, to reveal greater effects of age. Lastly, due to potential problems with attention disengagement from the voice of the previous target, we hypothesized that most of the confused words after a target talker switch originate from the previous target talker.

Methods

Participants

Data from 47 listeners (33 female, 14 male), all of which passed a dementia screening test (Kalbe et al., 2004), were included in the study. Participants were grouped based on their age and hearing status. In accordance with the World Health Organization (Mathers et al., 2000), hearing status was classified based on the average of the pure-tone thresholds at 500, 1000, 2000, and 4000 Hz in the better ear (better ear hearing loss, BEHL), whereby a BEHL value of 25 dB HL or less is defined as normal hearing.

The 14 participants in the young normal-hearing group (young NH) had a mean age of 24.6 years (range: 18–33 years) and BEHL values of 2.2 ± 2.5 dB HL (mean $\pm$ SD). Eighteen older participants with normal hearing (old NH group) had a mean age of 69.7 years (62–78 years) and BEHL values of 13.3 ± 4.9 dB HL. Fifteen older listeners (71.1 years, range: 65–83 years) had mild-to-moderate symmetrical hearing impairments (BEHL of 32.9 ± 6.1 dB HL, old HI group).

Hearing thresholds were predominantly symmetrical, with the mean of the thresholds at 500, 1000, 2000, and 4000 Hz differing between both ears by an average of 3.5 dB (maximum: 13.8 dB). None of the subjects were fitted with hearing aids.

Figure 1 compares the audiometric pure-tone thresholds of the three groups.
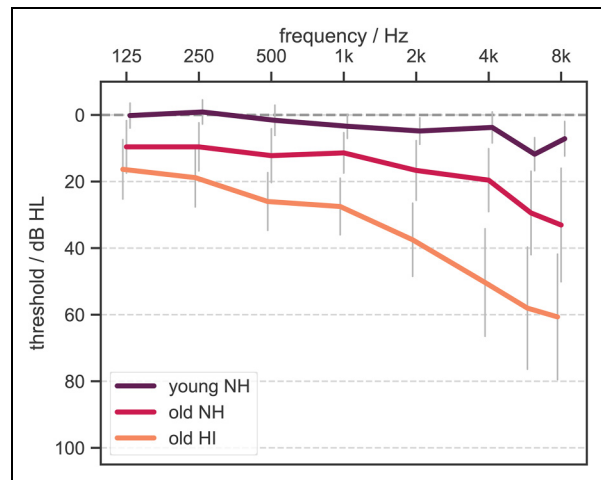
Figure 1. Mean audiometric pure-tone thresholds of the three listener groups. Thresholds were averaged across left and right ears. Error bars indicate standard deviations. Note that data points were horizontally shifted by a small amount to prevent the error bars from overlapping.

Setup and Stimuli

Experiments were conducted in a sound-treated and sound-proof booth to reduce room acoustic effects and to eliminate external noise. Participants sat in front of three loudspeakers (active near-field monitor KRK Rokit 5, providing a flat frequency response across the speech spectrum) located at azimuth angles of −60 (left), 0 (center) and 60 degrees (right). Each of the loudspeakers had a distance of 1.2 m from the listener's head. Through each of the loudspeakers, a sentence was played back at 65 dB SPL.

The experimental setup including the voices, locations and speech materials used are visualized in the upper part of Figure 2. The three sentences were presented simultaneously and were uttered by talkers with different voice characteristics: a male voice (fundamental frequency 122 Hz) as well as a deep and a high female voice (143 Hz and 202 Hz). The talkers remained at fixed positions; for example, the high female voice was always presented through the right loudspeaker. This resembles a typical real-world listening situation with people located at fixed positions, e.g. around a table. Speech stimuli were derived from the German version of the Oldenburg sentence test (Wagener et al., 1999), a matrix sentence test containing five-word sentences such as "Stefan kauft acht nasse Autos" (Stefan buys eight wet cars) or "Nina sieht neun weiße Sessel" (Nina sees nine white armchairs). The duration of each sentence was 2.5 s. Three different sentences with no words in common were chosen for each trial. Participants had the task to verbally repeat back the words uttered by the target talker, which was indicated by the cue word "Stefan".

Using this setup, errors during speech recognition in static as well as in dynamic cocktail party situations were

investigated. In the static conditions, the talker uttering the target sentence remained constant for the duration of a test list (10 trials). Moreover, the experimenter verbally provided participants with a priori information about the voice and the location of the target talker before the start of a test list. There was one static test list for each of the three target talkers.

The upper part of Figure 2 shows an example for a target talker switch, whereas the lower part illustrates the composition of the test list for the dynamic condition. The target talker switched in 20% of the trials, resulting in 6 switch trials and 24 non-switch trials across a test list of 30 trials. Switches occurred at pseudo-random intervals of 2 to 5 trials. Participants were not given any information about the position and voice of the target talker. This was done to create stimulus uncertainty, requiring listeners to monitor all three speakers and to listen for the keyword in order to identify the target talker.

Procedures

The study was approved by the local ethics committee. After the participants had signed the informed consent forms, the pure tone-threshold measurement and the dementia screening test were carried out. Subsequently, the multi-talker listening test was conducted. The test began by introducing participants to the test materials and procedures by first presenting them with multiple sentences from the Oldenburg sentence test uttered by the three different voices. Next, 30 practice trials that represented the actual test were presented so that participants could get used to the multi-talker situation and rehearse the task at hand. The actual test included the aforementioned static and dynamic conditions. Listeners were instructed to face towards the loudspeaker at the center position throughout the presentation of a test list. The test had a total duration of about 1 h.

Data Analysis

Error Analyses in Static and Dynamic Conditions

We distinguished between confusion and random errors. In general, errors were quantified as error rates—that is the number of omitted or incorrectly reported words divided by the total number of target words presented. The first word of each sentence was ignored for the calculation of error rates since it solely served as a cue word, meaning there were four target words in each trial.

For the static condition, a total of 120 target words were considered (3 voices * 10 trials * 4 target words/trial) per participant. For the dynamic condition, only switch trials and trials directly preceding a switch were analyzed. The rationale here is that error rates increase directly after a switch and then gradually decrease over the next few trials until they are back to their initial state (e.g., Brungart & Simpson, 2007). Therefore, it can be assumed that switch trials (T0) are those most affected by a

switch, which is why only these trials were analyzed to examine the effect of switches. In contrast, for trials directly preceding a switch (T-1), the effect of the previous switch is assumed to be minimal, and they were thus used to investigate effects of monitoring different targets due to stimulus uncertainty. This results in a total of 24 target words (6 trials * 4 target words/trial) being considered for T-1 and T0 trials, respectively.

Random and confusion errors for the dynamic condition were calculated as the increase in error rates relative to an individual static baseline. The rationale here is that we wanted to focus on the additional errors that were made in the dynamic condition compared to static situations, making error rates comparable across listeners with different levels of performance. Static baselines represented the individual error rates in the static condition and also reflected the presentation schemes (i.e., the different voices and locations) in the dynamic condition. This was necessary since the different target voices did not occur equally often in T-1 and T0 trials. To obtain the baseline, static error rates were averaged across target voices, using the frequency of occurrence of the respective target voice in the corresponding trial type (T-1, T0) as a weighting factor. For example, the six switch trials comprised three presentations of the target from the left position uttered by the deep female voice, two from the center by the male talker, and one from the right by the high female voice. The baseline for the switch trials was therefore $(3 \cdot E_{static, \, left, \, femalelow} + 2 \cdot E_{static, \, center, \, male} + 1 \cdot E_{static, \, right, \, femalehigh}) \, / \, 6$, with $E_{static, \, x}$ being the error rate for the respective talker and position in the static conditions.

Statistical Analysis

Data were analyzed by means of linear mixed models (LMMs) with participants as random factor. Three separate LMMs were fitted—namely, for the analysis of errors in the static and dynamic conditions as well as for the comparison of different types of confusion errors (Figure 5).

For each analysis, different fixed effect structures were tested: We first fitted the full models including all fixed factors and every possible interaction between them. Next, non-significant factors and interactions were successively removed from the models until either a model with only significant factors or with just one remaining factor (i.e. no further reduction possible) was reached. Since the *p*-values between the full models and the reduced models only differed slightly, we will only report the results of the full models in the following.

Regarding the random effects, fitting the models was first attempted using the maximal effects structure (by-subject random slopes and intercepts for all within-subject factors). In case of singular fits or errors due to an insufficient number of observations, the random effects structure was simplified by removing the random slope of the corresponding factor.

The formulas and coefficients for all fitted models as well as outcomes of the statistical tests (*p*-values) can be found in the supplementary materials.
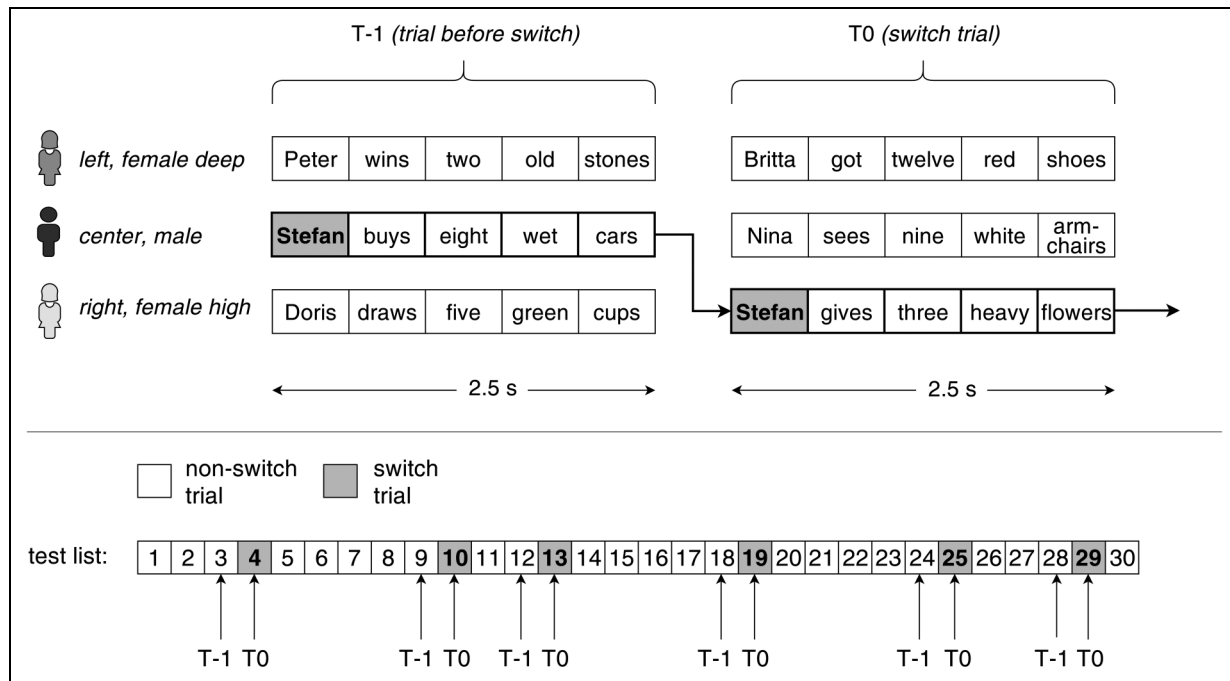
Figure 2. Upper part: example of a T-1 and a T0 trial illustrating the concurrent sentences as well as the switch of the target talker which is indicated by the cue word "Stefan". Lower part: Graphic representation of the dynamic test list consisting of 30 trials. Switch trials and trials immediately before a switch are labeled T0 and T-1, respectively.
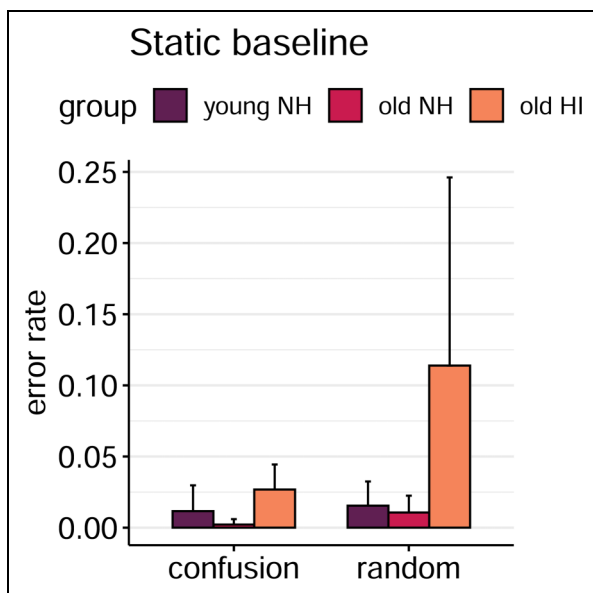
Figure 3. Mean error rates of the static baseline broken down by error type and listener group. Error bars indicate standard deviations. The plot shows the average of the baselines for T-1 and T0 trials.

LMMs were calculated in R (version 3.6.1; R Core Team, 2019) using version 1.1–26 of the lme4 package (Bates et al., 2015). The car library (version 3.0–3; Fox & Weisberg, 2019) was used to obtain p-values from the models.

Post-hoc analyses were carried out by comparing estimated marginal means using version 1.4.3.01 of the R package emmeans (Lenth, 2019). *p*-values were Bonferroni corrected in case of multiple comparisons. The significance level was set to 0.05.

Results

Errors in the Static Condition

Figure 3 shows the random and confusion error rates for the static condition. In general, error rates appear to be low, with the exception of the random error rate of the old HI group. The factors error type ($x^2(1) = 9.1$, $p < 0.01$), group ($x^2(2) = 20.2$, $p < 0.001$) as well as their interaction ($x^2(2) = 12.4$, $p < 0.01$) had a significant effect on the error rates. Pairwise comparisons between groups conducted separately for each error type confirm that confusion errors as well as random errors are significantly higher for old HI than for both other groups (all $p < 0.02$). However, both the rate of confusion errors as well as the group differences for this error type are small and seem to be affected by floor effects.

Errors in the Dynamic Condition

Figure 4 depicts confusion and random error rates in the dynamic condition relative to the static baseline for both
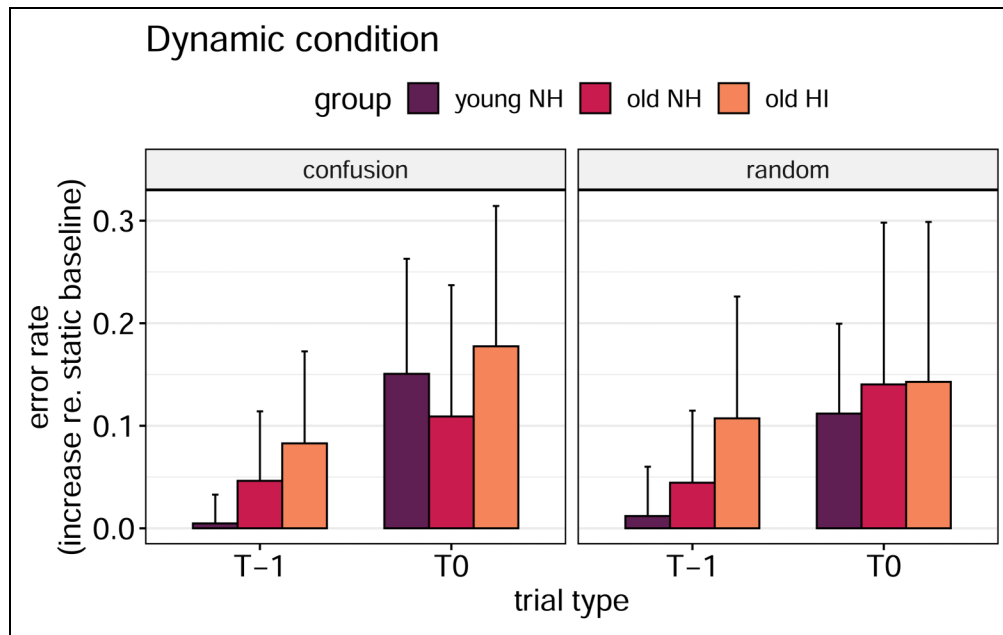
Figure 4. Confusion and random error rates for the different trial types in the dynamic condition. T-1 and T0 denote trials directly before and after a switch, respectively. Error rates are given as the increase relative to the static baseline. Means and standard deviations are shown. A version of this figure depicting the absolute error rates (not relative to the baseline) is shown in the supplementary materials.
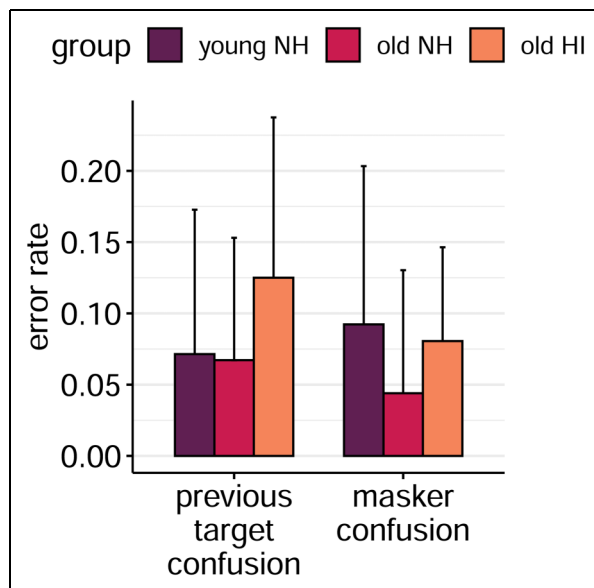
Figure 5. Means and standard deviations of the confusion error rates broken down by previous target confusions and masker confusions. Note that, in contrast to the data in Figure 4, error rates are not given relative to a static baseline.

trial types (T-1, T0). Confusion and random errors yield roughly similar patterns. It can be seen that the increases in error rates were higher in T0 compared to T-1 trials for all listener groups. Older HI listeners appear to have the highest increase of all groups. The LMM reveals significant differences between trial types ($x^2(1) = 40.9$, $p < 0.001$) and groups ($x^2(2) = 9.0$, $p = 0.011$), but not between confusion and random errors ($x^2(1) = 0.001$, $p = 0.97$). None of the interactions between these factors were significant (all $p \geq 0.23$). Motivated by our hypothesis regarding the age effect in non-switch trials, a post-hoc analysis of the group effect was conducted for both trial types separately: In T-1 trials, old HI had a significantly larger increase in error rates than the young NH ($p = 0.0049$), but not than the old NH ($p = 0.156$). No significant group differences could be observed for T0 trials.

Figure 5 shows results of an analysis inspired by Lin and Carlile (2015), which further distinguishes between confusions with the previous target and with the other masker. Note that this figure depicts absolute error rates, i.e., in contrast to the previous analysis, error rates are not given relative to the static baseline. The differences between both confusion types and the effect of listener group seem to be small. The LMM revealed no significant effects for the factors listener group ($x^2(2) = 4.15$, $p = 0.126$), confusion type ($x^2(1) = 0.75$, $p = 0.39$) or the confusion type × group interaction ($x^2(2) = 1.80$, $p = 0.41$).

Discussion

This study examined different error types in static and dynamic cocktail party listening. Two error types were distinguished—namely, confusions of target and maskers and random errors. In order to uncover potential age and

hearing loss effects, three different listener groups were enrolled.

Static Condition

In the static situation, cognitive load was low: listeners had a priori information about the voice and the position of the target talker and could accordingly focus their attention on the talker of interest (Kidd et al., 2005). In this situation, we assumed that random errors reflect problems with identifying words correctly and confusions reflect problems with assigning words to the correct auditory stream. We expected hearing loss to increase random errors as a consequence of limited audibility.

Analysis of the errors in the static situation revealed a significant listener group by error type interaction. Confusions were generally rare, suggesting that the voice and spatial cues provided were salient and allowed robust streaming. Still, a significantly higher proportion of confusions was found in the HI listeners. This would suggest that the use of the talker characteristics is restricted in this group, in accordance with findings based on F0 cues (Grimault et al., 2001) as well as spatial cues (Glyde et al., 2013). However, despite statistical significance, the amount of confusions was generally very low (i.e., around 1–3%).

In line with our hypothesis, the old HI listeners showed a clearly higher rate of random errors than both of the normal hearing groups. This supports the assumption that hearing loss affected audibility of the speech signal, consequently yielding more omissions and words misunderstood. Nevertheless, as random errors amounted on average to about 10%, performance in the static situation was still very high in this group despite the effects of their (mild-to-moderate) hearing impairment. Thus, it can be concluded that the static situation presented the listeners with robust voice and spatial cues, allowing them to focus attention on a given talker.

Dynamic Condition

In the dynamic condition, cognitive load was substantially higher, since the listeners had no a priori information about the target talker. Thus, they were always required to monitor different potential targets. In addition, they had to switch attention to a new position and voice when the target changed. In order to focus on the effects of the increased cognitive load in the dynamic situation, the errors were calculated relative to the static situation. This considers the aforementioned group differences as a baseline condition, thus allowing a more direct view on the effects of cognitive load. In the dynamic situation, we assumed that an increase in confusions reflects a failure to focus attention on the target talker and to inhibit the competing information as a result of stimulus uncertainty and target variability. Increase in random errors was interpreted as a load effect

on the ability to retrieve the information from short term memory beyond the effects of audibility. For instance, Mattys and Palmer (2015) have shown that cognitive load decreases perceptual accuracy, potentially resulting in incorrectly perceived words. Moreover, consistent with limited capacity theory (Kahneman, 1973), it is expected that cognitive load affects recall (Barrouillet et al., 2007; Camos & Portrat, 2015), possibly increasing omissions.

We hypothesized that both error types would increase in the dynamic situation and that this would be especially pronounced in the older HI listeners. We further expected that older listeners show problems with monitoring different talkers and inhibiting irrelevant information, reflected in a growing amount of confusions.

The analysis revealed that both confusion and random errors increased by a similar amount in the dynamic relative to the static situations. In line with the argumentation above, both focusing attention on the target as well as retrieving the information from memory seem to be compromised in the dynamic situation. In order to distinguish between the load due to monitoring different talkers as well as additionally switching attention, errors in the trial before (T-1) and after target changes (T0) were contrasted. Compared to the static baseline condition, T0 revealed a significantly higher increase in confusion and random errors than T-1. This appears to reflect the additional load of switching attention to a new target, apart from the need to monitor different talkers. Moreover, the analysis revealed a significantly higher increase in errors in the old HI listeners compared to the young NH subjects, whereas the difference between old HI and old NH did not reach statistical significance. This gives evidence that the load associated with the dynamic situation is especially detrimental in older listeners with hearing impairment. Note that this is unlikely a simple hearing loss effect since the increase in errors relative to the static situation was considered here. In line with our interpretation of the error types, this would suggest that hearing impairment affects the ability to focus attention as well as the retrieval from memory beyond pure peripheral effects. We propose that this is due to the combination of perceptual and cognitive mechanisms: hearing impairment compromises the speech input and causes a perceptual load. Consistent with the idea that this calls for top-down compensation (e.g., ELU-model, Rönnberg et al., 2013) this results in a cognitive load. In addition, the task-specific load of the dynamic situation becomes effective, possibly bringing cognitive capacity near or to its limit.

We further expected to detect group differences, especially in the non-switch trials. This was motivated by the fact that several studies did not find age-related differences for switching attention (Oberem et al., 2017; Singh et al., 2013), but rather showed that the stimulus uncertainty in the non-switch trials revealed age effects (Getzmann et al., 2017; Meister et al., 2020). Despite the fact that our analysis did not show a significant trial (T-1, T0) by group interaction,

the pattern in Figure 4 suggests such an effect. While the young NH subjects revealed virtually no increase in confusion or random errors for T-1 relative to the static baseline, both error types increased in the older listeners, particularly in the old HI group. Indeed, an analysis of the errors in T-1 showed significantly higher increases in error rates for the old HI than for the young NH, but not than the old NH.

Regarding the confusion errors found in the switch trial (T0), Lin and Carlile (2019) suggested that two different mechanisms are involved when the target voice changes—namely, re-engaging attention to the new target voice, but also disengaging attention from the previous target voice. This seems plausible since listeners may develop an attentional bias towards one talker, which may hinder switching attention to a new talker of interest. For instance, Brungart and Simpson (2007) found a negative priming effect when examining talker switches in a dynamic situation. Repeatedly presenting the target sentence from one talker builds up an expectation that the target talker will remain the same in subsequent trials, which makes the listeners less able to switch attention. Based on these assumptions, we hypothesized that confusions with the previous target talker would occur more frequently than with the other talker. However, our analysis revealed that the proportion of confusions originating from the two non-targets did not differ significantly from each other. This result does not give evidence that such a negative priming effect emerged and that disengaging attention played an important role. This might be explained with the relatively short periods (2 to 5 trials) of the target talker remaining the same in our study, in contrast to Brungart and Simpson (2007), who presented up to about 50 trials. In this regard, our paradigm more resembles a typical dynamic conversation situation in which the buildup of expectation may be rather weak. In line with the outcomes of our study, an analysis of confusion errors by Lin and Carlile (2015) also did not reveal problems with attention disengagement from the previous target. However, in this study, all three competing sentences were uttered by the same voice and switches were only spatial, which may explain the discrepancies with their follow-up study (Lin & Carlile, 2019), which only included switches of the target voice. Comparing the findings of the two studies by Lin and Carlile and the present study, one could suggest that either the absence of spatial cues (Lin & Carlile, 2019) facilitates bias effects or that listeners become more easily biased towards a voice than a location. The latter hypothesis begs the question why no problems with attention disengagement could be observed in the present study, even though talkers with distinct voices were used. It is conceivable that methodological differences explain this: In the experiments by Lin and Carlile, switches occurred between pairs of target sentences that were presented in quick succession (temporal gap of 300–400 ms), whereas in the present study, target sentences were separated by a gap of several seconds, during which participants had to

repeat back the words they recognized. Those prolonged pauses between stimulus presentations could have mitigated the biasing effect.

Summary and Conclusions

In the present study, we analyzed speech recognition errors in static and dynamic cocktail party situations in young as well as older listeners with and without hearing loss. The analyses distinguished between random and confusion errors.

In the static condition, in which cognitive load was assumed to be low, performance was generally high. As expected, the older HI listeners showed a higher rate of random errors than the two other groups, which was probably due to reduced audibility. Confusion errors were rare in all listener groups.

In the dynamic condition, which is assumed to be more cognitively demanding, both random and confusion errors increased by a similar amount relative to the static condition. This suggests increased problems with focusing attention and memory recall associated with higher cognitive demand. Increases in errors were higher in switch trials than in trials immediately before a switch, which can be linked to the additional cognitive load caused by the need to switch attention to a new target. In addition, increases in errors differed significantly between listener groups.

Contrary to our expectations, there was no significant interaction between trial type (T-1, T0) and group. However, a post-hoc analysis revealed that old HI listeners showed a significantly higher increase in error rates than the young NH group (but not than the old NH) for trials before switches (T-1), whereas there were no group differences for switch trials. This might suggest that hearing loss in combination with the task specific demand of the dynamic condition challenges cognitive resources. Specifically, this mainly affected trials before switches, in which stimulus uncertainty is assumed to be the main challenge. By contrast, the adverse effect of target talker switches on errors does not seem to depend on age or hearing loss.

We further hypothesized that due to potential problems with attention disengagement, listeners would tend to keep their attention focused on the previous target talker after a switch. However, an analysis of switch trials could not confirm this: confused words originated from the previous target talker and the other masker with a similar frequency, suggesting there was no attentional bias towards the previous target.

Taken together, the analysis of confusion and random errors proposed that the load associated with dynamic cocktail party listening affects the ability to focus attention on the talker of interest as well as the retrieval of words from short-term memory. The finding that older listeners with hearing loss showed the highest increase in both error types in dynamic cocktail party listening is in line with an interplay of perceptual and cognitive mechanisms.

ORCID iD

Moritz Wächtler [ID] https://orcid.org/0000-0003-0116-6419

Supplemental Material

Supplemental material for this article is available online.

References

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience*, *5*(3), 202–212. https://doi.org/10.2741/Alain

Barrouillet P., Bernardin S., Portrat S., Vergauwe E., & Camos V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

Bates D., Mächler M., Bolker B., & Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Best V., Ozmeral E. J., Kopco N., & Shinn-Cunningham B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences*, *105*(35), 13174–13178. https://doi.org/10.1073/pnas.0803718105

Bolia R. S., Nelson W. T., Ericson M. A., & Simpson B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066. https://doi.org/10.1121/1.428288

Bregman A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.

Bronkhorst A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, *77*(5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9

Brungart D. S., & Simpson B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception & Psychophysics*, *69*(1), 79–91. https://doi.org/10.3758/BF03194455

Camos V., & Portrat S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review*, *22*(4), 1029–1034. https://doi.org/10.3758/s13423-014-0772-5

Cherry E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Fox J., & Weisberg S. (2019). *An R companion to applied regression (third)*. Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Getzmann S., Hanenberg C., Lewald J., Falkenstein M., & Wascher E. (2015). Effects of age on electrophysiological correlates of speech processing in a dynamic "cocktail-party" situation. *Frontiers in Neuroscience*, *9*, 1–17. https://doi.org/10.3389/fnins.2015.00341

Getzmann S., Jasny J., & Falkenstein M. (2017). Switching of auditory attention in "cocktail-party" listening: ERP evidence of cueing effects in younger and older adults. *Brain and Cognition*, *111*, 1–12. https://doi.org/10.1016/j.bandc.2016.09.006

Glyde H., Cameron S., Dillon H., Hickson L., & Seeto M. (2013). The effects of hearing impairment and aging on spatial processing. *Ear & Hearing*, *34*(1), 15–28. https://doi.org/10.1097/AUD.0b013e3182617f94

Grimault N., Micheyl C., Carlyon R. P., Arthaud P., & Collet L. (2001). Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. *British Journal of Audiology*, *35*(3), 173–182. https://doi.org/10.1080/00305364.2001.11745235

Helfer K. S., & Freyman R. L. (2008). Aging and speech-on-speech masking. *Ear & Hearing*, *29*(1), 87–98. https://doi.org/10.1097/AUD.0b013e31815d638b

Ihlefeld A., & Shinn-Cunningham B. (2008). Disentangling the effects of spatial cues on selection and formation of auditory objects. *The Journal of the Acoustical Society of America*, *124*(4), 2224–2235. https://doi.org/10.1121/1.2973185

Kahneman D. (1973). *Attention and effort*. Prentice-Hall.

Kalbe E., Kessler J., Calabrese P., Smith R., Passmore A. P., Brand M., & Bullock R. (2004). Demtect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International Journal of Geriatric Psychiatry*, *19*(2), 136–143. https://doi.org/10.1002/gps.1042

Kidd G., Arbogast T. L., Mason C. R., & Gallun F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*, *118*(6), 3804–3815. https://doi.org/10.1121/1.2109187

Kidd G., Mason C. R., Richards V. M., Gallun F. J., & Durlach N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143–189). Springer US. https://doi.org/10.1007/978-0-387-71305-2_6

Lenth R. (2019). emmeans: Estimated marginal means, aka least-squares means (R package version 1.4.3.01). https://CRAN.R-project.org/package=emmeans

Lin G., & Carlile S. (2015). Costs of switching auditory spatial attention in following conversational turn-taking. *Frontiers in Neuroscience*, *9*, 1–11. https://doi.org/10.3389/fnins.2015.00124

Lin G., & Carlile S. (2019). The effects of switching non-spatial attention during conversational turn taking. *Scientific Reports*, *9*(1), 8057. https://doi.org/10.1038/s41598-019-44560-1

Mathers C., Smith A., & Concha M. (2000). Global burden of hearing loss in the year 2000. *Global Burden of Disease*, *18*(4), 1–30.

Mattys S. L., & Palmer S. D. (2015). Divided attention disrupts perceptual encoding during speech recognition. *The Journal of the Acoustical Society of America*, *137*(3), 1464–1472. https://doi.org/10.1121/1.4913507

Meister H., Wenzel F., Gehlen A. K., Kessler J., & Walger M. (2020). Static and dynamic cocktail party listening in younger and older adults. *Hearing Research*, *395*, 108020. https://doi.org/10.1016/j.heares.2020.108020

Oberem J., Koch I., & Fels J. (2017). Intentional switching in auditory selective attention: Exploring age-related effects in a spatial setup requiring speech perception. *Acta Psychologica*, *177*, 36–43. https://doi.org/10.1016/j.actpsy.2017.04.008

Pichora-Fuller M. K., Kramer S. E., Eckert M. A., Edwards B., Hornsby B. W. Y., Humes L. E., Lemke U., Lunner T., Matthen M., Mackersie C. L., Naylor G., Phillips N. A., Richter M., Rudner M., Sommers M. S., Tremblay K. L., & Wingfield A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear & Hearing*, *37*(1), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, https://www.R-project.org/

Rönnberg J., Lunner T., Zekveld A., Sörqvist P., Danielsson H., Lyxell B., Dahlström O., Signoret C., Stenfelt S., Pichora-Fuller M. K., & Rudner M. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Shinn-Cunningham B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Singh G., Pichora-Fuller M. K., & Schneider B. A. (2013). Time course and cost of misdirecting auditory spatial attention in younger and older adults. *Ear & Hearing*, *34*(6), 711–721. https://doi.org/10.1097/AUD.0b013e31829bf6ec

Wächtler M., Kessler J., Walger M., & Meister H. (2021). Costs of dynamic cocktail party listening: Investigating the effects of cognitive abilities and hearing impairment. *JASA Express Letters*, *1*(7), 075201. https://doi.org/10.1121/10.0005511

Wagener K., Brand T., & Kollmeier B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I-III: Design, Optimierung und Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test for the German language. I-III: Design, optimization and evaluation of the Oldenburg sentence test). *Zeitschrift Für Audiologie (Audiological Acoustics)*, *38*, 4–15.