

Using the Repeat-Recall Test to Examine Factors Affecting Context Use

Francis Kuk, PhD¹ Christopher Slugocki, PhD¹ Petri Korhonen, MSc¹

¹WS Audiology, Widex Office of Research in Clinical Amplification (ORCA-USA), Lisle, Illinois

Address for correspondence Francis Kuk, PhD, francis.kuk@wsa.com

J Am Acad Audiol 2020;31:771–780.

Abstract

Background The effect of context on speech processing has been studied using different speech materials and response criteria. The Repeat-Recall Test (RRT) evaluates listener performance using high context (HC) and low context (LC) sentences; this may offer another platform for studying context use (CU).

Objective This article aims to evaluate if the RRT may be used to study how different signal-to-noise ratios (SNRs), hearing aid technologies (directional microphone and noise reduction), and listener working memory capacities (WMCs) interact to affect CU on the different measures of the RRT.

Design Double-blind, within-subject repeated measures design.

Study Sample Nineteen listeners with a mild-to-moderately severe hearing loss.

Data Collection The RRT was administered with participants wearing the study hearing aids under two microphone (omnidirectional vs. directional) by two noise reduction (on vs. off) conditions. Speech was presented from 0 degree at 75 dB sound pressure level and a continuous speech-shaped noise from 180 degrees at SNRs of 0, 5, 10, and 15 dB. The order of SNR and hearing aid conditions was counterbalanced across listeners. Each test condition was completed twice in two 2-hour sessions separated by 1 month.

Results CU was calculated as the difference between HC and LC sentence scores for each outcome measure (i.e., repeat, recall, listening effort, and tolerable time). For all outcome measures, repeated measures analyses of variance revealed that CU was significantly affected by the SNR of the test conditions. For repeat, recall, and listening effort measures, these effects were qualified by significant two-way interactions between SNR and microphone mode. In addition, the WMC group significantly affected CU during recall and rating of listening effort, the latter of which was qualified by an interaction between the WMC group and SNR. Listener WMC affected CU on estimates of tolerable time as qualified by significant two-way interactions between SNR and microphone mode.

Conclusion The study supports use of the RRT as a tool for measuring how listeners use sentence context to aid in speech processing. The degree to which context influenced scores on each outcome measure of the RRT was found to depend on complex interactions between the SNR of the listening environment, hearing aid features, and the WMC of the listeners.

Keywords

- ▶ Repeat-Recall Test
- ▶ semantic context use
- ▶ directional microphone
- ▶ realistic signal-to-noise ratios

received
January 13, 2020
accepted after revision
April 17, 2020

© 2021. American Academy of Audiology. All rights reserved. Thieme Medical Publishers, Inc., 333 Seventh Avenue, 18th Floor, New York, NY 10001, USA

DOI <https://doi.org/10.1055/s-0040-1719136>.
ISSN 1050-0545.

In a previous paper, we reported on the use of the Repeat-Recall Test (RRT) as an integrative tool to examine the efficacy of a directional microphone (DIRM) and a noise reduction (NR) algorithm.¹ We examined how the signal-to-noise ratios (SNRs) of the environment and the working memory capacities (WMCs) of the listeners affected the efficacy of these two features on outcome measures of speech intelligibility (repeat), word/sentence retention (recall), and ratings of listening effort and tolerable time. We showed that the noted efficacy interacted with SNR, WMC, and passage context. Specifically, while all participants benefited from the use of the DIRM on the repeat task, participants in the good WMC group received more DIRM benefit at the poorer SNRs and no benefit at the SNR of 15 dB while those in the poorer WMC group showed slightly less benefit at the poorer SNR but such benefit continued to a SNR of 15 dB for the low context (LC) materials. Furthermore, those in the poorer WMC group benefited from NR on rating of listening effort. Space limitations did not permit us to explore how the use of context was affected by the study parameters. We report on how context use (CU) is affected in this article.

Speech comprehension involves both bottom-up and top-down processes. Bottom-up processes include factors that affect stimulus audibility such as room acoustics, SNRs, and hearing losses. Top-down processes include factors that facilitate stimulus comprehension. They could include the cognitive capacity of the listener,²⁻⁴ knowledge of the language, or the listeners' ability to use context, among others.

Contextual cues refer to any and all social, physical, visual, tactile, linguistic, and/or semantic information that a listener might use to gain communication success. Contextual cues might increase the speed and/or accuracy of speech identification and free up cognitive resources for storage and processing of the intended communication. In turn, this may decrease the perceived effort associated with communication. Although some have suggested that context contribution increases with the difficulty of the listening situation,^{5,6} it is not immediately clear how SNRs representative of real-world conditions affect CU on different tasks.

Early studies on the effects of semantic context^{5,7,8} used the Speech Perception in Noise (SPIN) test.⁹⁻¹¹ The SPIN test quantifies context effects by comparing intelligibility scores for sentence-final words between sentences where said words are either predictable (e.g., He is sleeping on the *bed*) or unpredictable (e.g., He is going to buy the *bed*) based on the sentence context. Other methods and materials have also been used to study context effects. For example, Boothroyd and Nittrouer¹² created their own high and low probability sentences. Helfer and Freyman¹³ reported that providing knowledge of sentence topic improved the perception in noise. Zekveld et al⁴ asked subjects to generate text cues and examined their effects on the intelligibility of natural sentences in noise. Guediche et al¹⁴ manipulated voice onset time to produce ambiguous and unambiguous target word stimuli (goat and coat) and investigated the effects of prior sentence context on phonetic perception of the target words. Indeed, many manipulations that may affect the top-down processing of target stimuli could serve as context.

Study of semantic context effect is not limited to speech intelligibility task. Many researchers have shown that semantic context also improved sentence retention and recall.^{3,4,15-17} Holmes et al¹⁸ reported that listeners rated semantically congruent sentences on the Connected Speech Test¹⁹ as less effortful compared with semantically incongruent sentences. Similarly, Winn²⁰ reported that semantic context reduced listening effort as evaluated by pupillometry.

Together, these studies support the benefits that context adds to speech understanding, recall, and listening effort. However, speech understanding tasks have different functional and cognitive requirements than recall tasks which are yet different than those required for rating of listening effort. Hence, the SNR(s) at which CU is maximal may differ across the different evaluative criteria or outcome measures. Furthermore, differences in the cognitive requirements of each measure suggest that the WMC of listeners may also modulate CU. For example, a hearing aid (HA) feature that provides slight SNR improvements might improve speech-in-noise performance but no improvement in listening effort. In such a scenario, wearers may still be dissatisfied with the performance of the HAs. A study that systematically examines how CU changes with SNR for different evaluative criteria may provide a better understanding of factors affecting CU in listeners with different cognitive capacities and offer guidance for the future design and selection of HAs as well as patient counseling.

For the average listener, the overall sound level and the SNRs of the listening environment determine the listening difficulty. For a hearing-impaired listener, the degree of hearing loss and HA status could also affect the difficulty of the listening situation. For example, a hearing-impaired listener would hopefully (but not always) find the listening situation less difficult when aided than unaided. The type of technology within the HA, including the use of NR and DIRM, could further affect the listener's difficulty in the listening situation. This could, in turn, affect the degree to which context might alleviate listening difficulty. For example, DIRMs reportedly improved the SNRs of the listening environment by 1 to 6 dB.²¹ While there is limited evidence to support that NR algorithms improve SNR, they have been shown to reduce listening effort.^{1,17} Thus, the use of processing features on a HA could change CU across realistic SNRs.

When studying context effects, it is important that the chosen speech materials minimize variables that may bias the observed effect. For example, Zekveld et al⁴ criticized the SPIN in that the SNR of the sentence context was the same as the target word. Thus, the audibility of the context cues was not assured at SNRs where the audibility of the target words was questionable. Other lexical factors such as word frequency and familiarity, phonological similarity, or age of word acquisition could affect the use of context.²² Moulin and Richard²³ reported that spondees that occurred more frequently provided more contextual information than spondees that occurred less frequently. Thus, to best reflect true context effects, high (HC) and LC speech materials should match their frequency of occurrence and word difficulty. Furthermore, the syntactic structure of the materials with and without context should also be similar (if not

identical) to minimize bias. These issues complicated the use of the SPIN to study context effects because high and low probability SPIN sentences are not identical in syntactical structure or word familiarity.

We developed the RRT as an integrated speech test that allows for study of context effects across several outcome measures.²⁴ This includes the listener's ability to (1) repeat sentences in quiet and in noise, (2) retain and recall those sentences, (3) rate the perceived listening effort required of the test conditions, and (4) judge their willingness to stay engaged in conversation (referred to as "tolerable time"). The test uses high and LC sentences presented under SNRs of 0, 5, 10, 15 dB, and quiet that are representative of real-world communication conditions.^{25,26} HC sentences are short meaningful 6 to 8 word sentences each with 3 to 4 target words. These are grouped into 6-sentence lists, where sentences within a list are related to a theme (e.g., food). LC sentences are then made by rearranging target words within a HC list such that the resulting list of 6 sentences are still syntactically similar or identical to the original HC sentences, but semantically meaningless. Thus, the meaningfulness of the sentences is used as a context to help listeners identify the target words. CU is calculated as a difference in target word scores between the HC and LC sentences. An example of a list of complementary HC and LC sentences is shown in **Appendix A**.

This manner of defining context has several advantages. First, the same words are used in both HC and LC materials; this minimizes any issues with word familiarity, frequency of occurrence, and/or difficulty. Second, because both versions use the same words, the long-term spectra of complementary HC and LC materials are similar. This helps to control possible confounds related to word audibility. Third, the same sentence structure is used for both HC and LC materials, which minimizes syntactical biases. On the other hand, because the target words and the rest of the sentence are presented at the same SNR, the audibility of the contextual cues is necessarily tied to the audibility of the target words.

In this study, we wanted to use the different outcome measures (i.e., repeat, recall, listening effort, and tolerable time) on the RRT to study changes in CU across a range of realistic SNRs. In addition, we wanted to examine if CU depends on HA features such as DIRMs and NR and/or on the WMC of the listener. Answers to these questions will allow one to know (1) if the RRT can be used to study context effects, (2) how CU changes for each RRT outcome measure, (3) how HA technology influences CU, and (4) if WMC affects how much context is used.

Methods

The readers are referred to Kuk et al.¹ for a detailed description of the Methods. A brief summary of the study details is reported here.

Participants

Nineteen hearing-impaired adults (average age of 73.6 years) with a bilaterally symmetrical mild-to-moderately severe

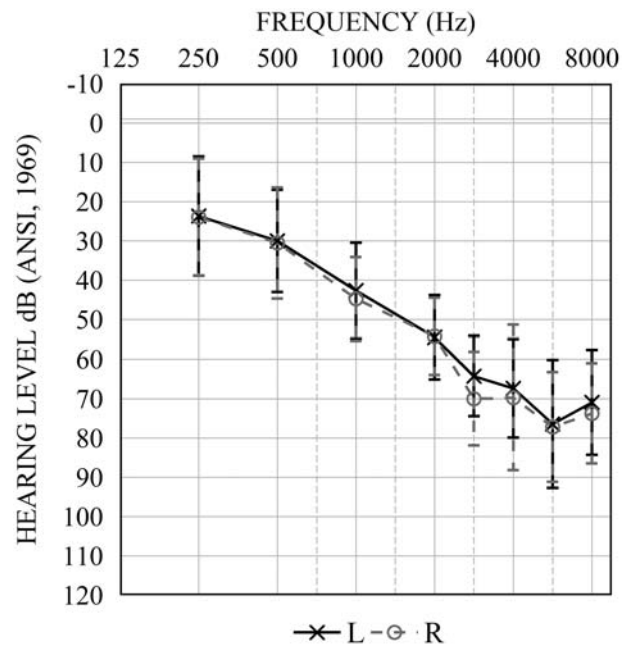


Fig. 1 Average pure-tone thresholds for the left (black exes) and right (gray circles) ears of 19 hearing-impaired listeners. Error bars represent 1 standard deviation.

sensorineural hearing loss and normal cognition participated (**Fig. 1**). Informed consent was obtained from all participants in accordance with protocols approved by an external institutional review board.

Hearing Aid Conditions

Participants completed all testing in the aided mode with bilaterally fitted receiver-in-canal HAs coupled to fully occluding "double-dome" instant-fit ear-tips. The NAL-NL2²⁷ fitting target was used and all fittings were verified for adequate audibility using the SoundTracker feature of the fitting software.²⁸ The fully adaptive beamformer was set to a fixed hypercardioid mode during testing. When activated, the modulation-based NR algorithm reshaped the frequency response to optimize the speech intelligibility index with a maximum gain reduction of 12 dB and maximum gain increase of 4 dB in the mid frequencies. Four combinations of microphone and NR conditions were evaluated: omni-DIRM with NR enabled (OMNI.NR.ON); omni-DIRM with NR disabled (OMNI.NR.OFF); DIRM with NR enabled (DIRM.NR.ON); and DIRM with NR disabled (DIRM.NR.OFF).

Test Materials and Procedure

The study followed a double-blind within-subjects design. Subject performance on the RRT was evaluated using a different list for each HA condition. Listeners first repeated the sentence that they heard. After all 6 sentences within a list were repeated, listeners recalled as many of the sentences (or target words) as they could recall. Afterwards, listeners rated the amount of perceived listening effort using a 1- to 10-point scale with "1" being "not effortful" and "10" being "extremely effortful." Listeners then estimated the amount of time (in minutes with a minimum of less than 1 minute and a

maximum of 2 hours) that they were willing to spend listening under the specific SNR condition. A practice trial at a SNR of 10 dB was completed. The LC sentences were always presented prior to the HC sentences.

Speech stimuli were delivered in the free-field at a fixed peak level of 75 dB sound pressure level 1 m from the front. A spectrally matched, continuous speech-shaped noise was presented 1 m directly behind the listener so both the DIRM and NR algorithm may be activated. Background noise was presented at fixed levels to produce SNRs of 0, 5, 10, and 15 dB in a random order.

Results

The recall score for HC sentences presented at a SNR of 15 dB was used to group listeners into good and poor WMC categories. This test condition was used because the repeat scores were $\geq 95\%$ in all participants to assure audibility. Because two peaks (at 35 and 50%) were noted in the distribution of the recall scores, listeners with recall performance $\geq 43\%$ were placed into the “good” WMC group and those with recall performance $< 43\%$ were placed into the “poor” WMC group. There were 10 participants in the good WMC group and 9 in the poor WMC group. Participants in both groups (good vs. poor) were similar in their ages (73 vs. 74 years), pure-tone averages (47 vs. 51 dB hearing level), and Montreal Cognitive Assessment scores²⁹ (27 vs. 26). The absolute scores for different test conditions were reported in the previous report¹ and detailed in the “Discussion” section. In this report, CU was the dependent variable and it was calculated as the difference between HC and LC sentence scores with the restriction that CU is not smaller than 0.

Repeated measures analyses of variance (ANOVAs) were conducted to assess the within-subjects factors of *Microphone* (2 levels, DIRM and OMNI), *NR* (2 levels, NR.ON and NR.OFF), *SNR* (4 levels, 0, 5, 10, and 15 dB), and *WMC group* (2 levels, good WMC and poor WMC) on CU separately for repeat, recall, listening effort, and tolerable time. Analyses assessed all interactions of these factors. Degrees of freedom were adjusted using Greenhouse–Geisser correction wherever the assumption of sphericity was violated. All ANOVAs were calculated using Type III sums of squares. The value of η^2 is reported to allow judgment of effect size. It has been suggested that η^2 values of 0.01, 0.09, and 0.25 may reflect small, medium, and large effect sizes, respectively.³⁰

Repeat Performance

For repeat performance, CU was significantly affected by the main effect of *SNR* ($F_{(3,51)} = 11.93, p < 0.001, \eta^2 = 0.12$). This main effect was qualified by a two-way interaction between *SNR* and *Microphone* ($F_{(3,51)} = 24.07, p < 0.001, \eta^2 = 0.17$). These were medium size effects. ▶ **Fig. 2** compares CU between WMC groups for each *Microphone* condition. With OMNI processing, maximum CU was observed at SNR = 10 dB. With DIRM processing, maximum CU was observed at SNR = 5 dB, decreasing as SNR increased beyond that level. CU was also higher in the OMNI mode than in the DIRM mode at SNR ≥ 10 dB. There was no significant effect of the WMC group or NR.

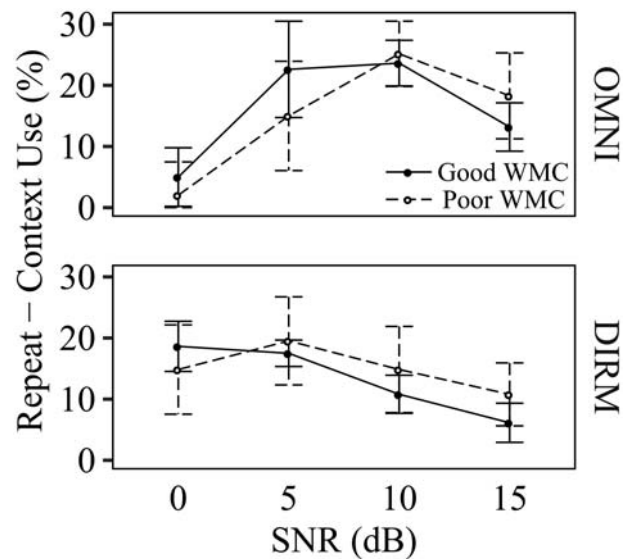


Fig. 2 Context use (CU) for repeat performance across signal-to-noise ratios (SNRs) for people with good (solid line) and poor (dotted line) working memory capacities (WMCs) in the omnidirectional (OMNI, top) and directional mode (DIRM, bottom) mode.

The amount of CU exceeded 30% in some test conditions. When collapsed across all SNR and microphone conditions, CU was estimated to be approximately 4.5 dB at a speech reception threshold criterion of 75%. This is less than the 6.5 dB improvement offered by the use of the DIRM.¹ It is difficult to compare the magnitude of the context effect measured in this study with those of others¹⁸ because of the differences in test materials.

Recall Performance

CU during recall was significantly affected by the main effects of *WMC group* ($F_{(1,17)} = 24.39, p = 0.001, \eta^2 = 0.14$), *Microphone* ($F_{(1,17)} = 6.03, p = 0.025, \eta^2 = 0.02$), and *SNR* ($F_{(3,51)} = 16.70, p < 0.001, \eta^2 = 0.17$). The effect size of *WMC group* and *SNR* was medium while that of *Microphone* was small. The effect of *NR* was not significant. These main effects were qualified by a significant *Microphone* \times *SNR* interaction ($F_{(3,51)} = 22.00, p < 0.001, \eta^2 = 0.14$) with a medium effect size. ▶ **Fig. 3** compares CU for recall between participants in each microphone mode. The good WMC group benefitted more from context than the poor WMC group at all SNRs for both microphone modes. Listeners made more use of context in the DIRM versus the OMNI mode at poorer SNRs (0 and 5 dB). This pattern reversed at SNR = 10 dB. In addition, CU was relatively stable across SNRs in the DIRM mode but increased as SNR increased in the OMNI mode. These results support previous observations that context improves recall⁴ and that people with better WMC show more CU than those with a poorer WMC.³

Ratings of Listening Effort

CU in rating of perceived listening effort was affected by *WMC group* ($F_{(1,17)} = 9.64, p = 0.006, \eta^2 = 0.11$) and *SNR* ($F_{(3,51)} = 22.65, p < 0.001, \eta^2 = 0.14$). These effect sizes were medium. The effect of *NR* was not significant. These

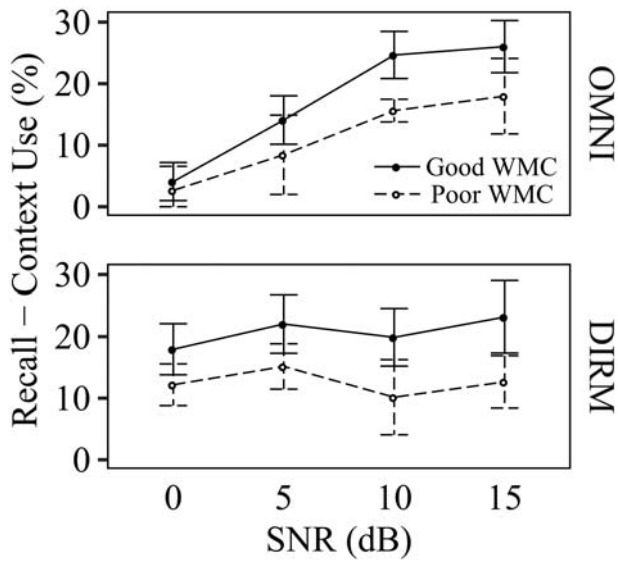


Fig. 3 Context use (CU) for recall performance across signal-to-noise ratios (SNRs) for people with good (solid line) and poor (dotted line) working memory capacities (WMCs) in the omnidirectional (OMNI, top) and directional mode (DIRM, bottom) mode.

effects were further qualified by significant *WMC group* × *SNR* ($F_{(3,51)} = 4.02, p < 0.012, \eta^2 = 0.03$) and *Microphone* × *SNR* ($F_{(3,51)} = 9.78, p < 0.001, \eta^2 = 0.05$) interactions with a small effect size. ▶**Fig. 4** compares CU between participants in each microphone mode. In general, good WMC listeners reported a greater reduction in perceived listening effort when processing HC versus LC sentences than did poor WMC listeners, except at SNR = 0 dB where the two groups did not differ. In addition, all listeners benefited more from context in the DIRM than in the OMNI mode at SNRs of 0 and 5 dB, but not at SNRs of 10 and 15 dB where CU was similar between the two microphone modes. For the DIRM,

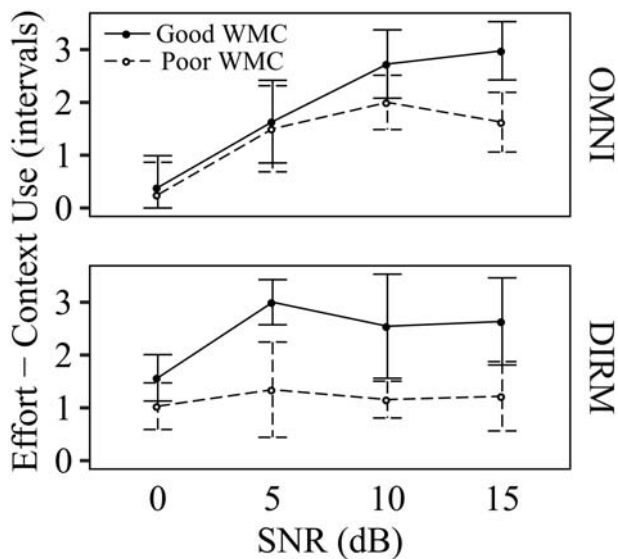


Fig. 4 Context use (CU) for perceived listening effort ratings across signal-to-noise ratios (SNRs) for people with good (solid line) and poor (dotted line) working memory capacities (WMCs) for the omnidirectional (OMNI, top) and directional mode (DIRM, bottom) mode.

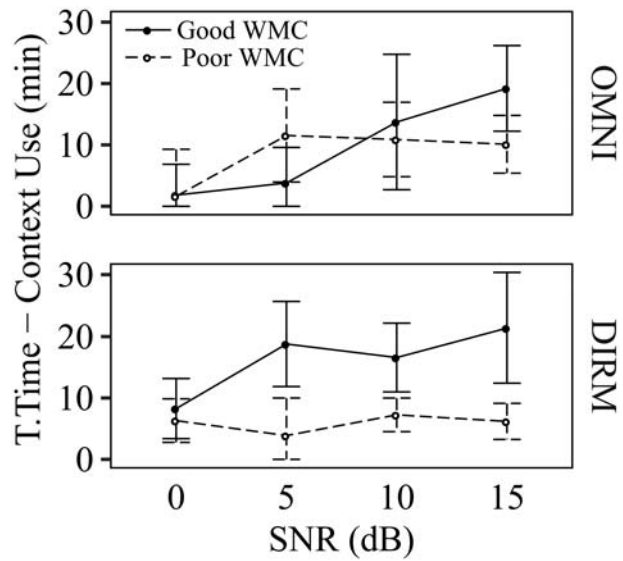


Fig. 5 Context use (CU) for tolerable time ratings across signal-to-noise ratios (SNRs) for people with good (solid line) and poor (dotted line) working memory capacities (WMCs) for the omnidirectional (OMNI, top) and directional (DIRM, bottom) mode.

CU was relatively constant across SNRs, whereas for the OMNI, CU increased as SNR increased.

Estimates of Tolerable Time

CU for tolerable time was affected by *SNR* ($F_{(3,51)} = 8.81, p < 0.001, \eta^2 = 0.06$) with significant two-way interactions between *SNR* and *WMC group* ($F_{(3,51)} = 2.95, p = 0.041, \eta^2 = 0.02$) and between *WMC group* and *Microphone* ($F_{(1,17)} = 5.61, p = 0.030, \eta^2 = 0.02$). These effect sizes were small. The effect of *NR* was not significant. ▶**Fig. 5** compares CU between participants in each microphone mode. Context improved estimates of tolerable time as SNR increased; however, this effect was stronger in listeners with good WMC than those with poor WMC. Listeners with good WMC reported longer tolerable time from context in the DIRM (vs. OMNI) mode, whereas CU was unaffected by microphone mode in listeners with poor WMC.

Discussion

The current study shows that the degree to which listeners use context depends on the interaction between the SNR of the environment, availability of a DIRM on the HA, and the WMC of the listeners. In addition, the pattern of CU may be different among the four outcome measures used on the RRT. Medium effect size was observed in most of the comparisons.

Starting at a poor SNR (i.e., 0 dB), CU increases with SNR until it reaches a maximum and then it either levels off (as observed for recall, listening effort, and tolerable time) or decreases (as observed of repeat) as SNR increases. This suggests that at SNRs of 0 or 5 dB, inaudibility of the speech signal limits the usability of any semantic cues. As SNR improves, some of the semantic cues become audible and contribute to improving performance for target words. Beyond a particular SNR, the audibility of the speech material is

sufficient for target word identification even in the absence of semantic cues. Thus, CU decreases for the repeat task when audibility is the determining factor. On the other hand, CU remains the same when task performance is not solely dependent on audibility as observed for recall, listening effort, and tolerable time measures when the HAs were in the DIRM mode.

HA technology influenced the SNR at which CU was maximal. Use of a DIRM improved audibility and thus usability of the semantic cues even at a SNR of 0 dB. Conversely, CU was not observed in the OMNI condition until a SNR of 5 dB. Maximum CU occurred at SNR of 5 dB in the DIRM mode and 10 dB in the OMNI mode. A DIRM alters the effective SNR at the listener's ears, which increases the availability of usable context cues to the listener. The NR algorithm used in the current study, while improving listening effort,¹ did not influence CU on any of the RRT measures.

Knowing the lowest SNR where CU occurred may provide an estimate of the minimum internal SNR required for optimal performance. For a repeat task, this was the SNR favorable enough to make the available semantic cues audible and maximally usable. When the HA was in the OMNI mode, maximum CU was observed at SNR = 10 dB, suggesting this SNR may meet the minimal internal SNR requirement. The observation of maximum CU at SNR = 5 dB in the DIRM mode reinforced this speculation. This is because the 5 dB input SNR in the DIRM mode, when added to the 6.5 dB benefit from DIRM,¹ was equal to an effective input SNR of 11.5 dB (5 + 6.5 dB). Thus, the use of a DIRM is mandatory if listeners are to benefit from semantic context at input SNRs \leq 5 dB. Otherwise, the input SNR must be $>$ 10 dB to fully utilize context. Interestingly, Smeds et al²⁵ and Wu et al²⁶ observed that the realistic SNRs of listeners with a mild-to-moderate hearing loss peaked around 10 dB. The results of the current study raise the possibility that these listeners might have chosen environments where they can fully utilize semantic contextual cues.

Patterns of CU varied depending on the outcome measure. On the repeat measure, CU reached a maximum and then decreased as SNR increased. For the other tasks (recall, listening effort, and tolerable time), CU stayed at similar levels in the DIRM mode and increased in the OMNI mode as SNR increased. The WMC of the listeners did not affect CU on the repeat measure, whereas listeners with better WMC were able to use more context on measures including recall, listening effort, and tolerable time. This difference in CU patterns across outcome measures may have implications for the test conditions under which we examine context effects in aided hearing-impaired listeners. Maximal context effects were noted on the repeat measure at SNRs between 5 and 10 dB. However, on the recall, listening effort, and tolerable time measures, similar CU was seen across SNRs \geq 5 dB in the DIRM mode and at SNRs \geq 10 dB in the OMNI mode. This suggests that the SNR where an aided hearing-impaired listener makes the most use of context depends on the outcome measure. If speech intelligibility is used to examine CU, then SNRs should be $<$ 10 dB. On the other hand, if the listener's tasks involve recall, rating of effort, or willingness to stay in noise, the

required SNR would be higher. If used with a DIRM, an SNR between 5 and 10 dB may be adequate. However, if it were an OMNI mic, then an SNR between 10 and 15 dB may be required to observe any effects.

Previous studies have suggested that CU depends on the listeners' WMCs.⁴ In this study, we observed CU during the repeat task to be similar between the good and poor WMC groups. However, listeners in the good WMC group showed more CU on recall, listening effort, and tolerable time measures. One possible explanation is that the cognitive capacities of all listeners in our sample were good enough to make use of contextual cues during the repeat task, but those in the good WMC group had additional cognitive spare capacity that could be directed at using contextual cues for encoding strategies and later retrieval. Spare capacity might also explain why listeners in the good WMC group found HC materials to be less effortful and more tolerable than LC materials at certain SNRs.

A Source of Difference in Context Use between WMC Groups

Because CU is a difference score between HC and LC sentences, a review of the absolute scores may provide additional insights into how listeners with good and poor WMC differed across measures. **Fig. 6** summarizes the absolute scores for each measure reported in Kuk et al,¹ averaged across SNRs and participants in each WMC group. For repeat, the good WMC listeners scored higher for both the HC and LC sentences than the poor WMC listeners; however, both groups were equally effective in utilizing context to help in speech understanding. Thus, CU was not different between WMC groups on the repeat task.

The good WMC listeners again scored higher than the poor WMC listeners on both the HC and LC sentences on the recall measure. However, the difference between WMC groups was less with the LC sentences than the HC sentences. Thus, the

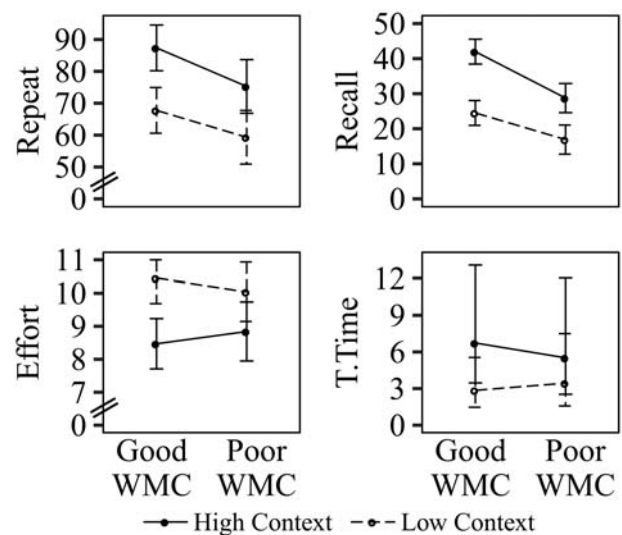


Fig. 6 High context (HC, solid line) and low context (LC, dotted line) sentence scores for good working memory capacities (WMC) and poor WMC listeners for the four outcome measures (repeat, recall, listening effort, and tolerable time) used on the Repeat-Recall Test (RRT).

good WMC listeners benefitted more from context than the poor WMC listeners to facilitate recall.

A different pattern emerged on the listening effort measure. Listeners in the good WMC group rated the HC sentences less effortful (8.5 for the good WMC vs. 9 for the poor WMC) and LC sentences more effortful (10.5 for the good WMC vs. 10 for the poor WMC) than the poor WMC listeners. Thus, a smaller difference in effort ratings between HC and LC materials was seen in the poor WMC listeners than the good WMC listeners. This resulted in greater CU in the good WMC group than the poor WMC group (2 vs. 1, a medium effect size). Intuitively, one would expect that the poor WMC listeners to rate the test conditions as more effortful than the good WMC listeners. This was indeed true for the HC sentences but not for the LC sentences.

Observations on the tolerable time (willingness to stay in noise) trended similarly as the listening effort ratings. Listeners in the good WMC group were willing to stay longer than the poor WMC listeners when HC materials were used (7 vs. 5.5 minutes) but less when LC materials were used (3 vs. 3.2 minutes). This resulted in greater CU in the good WMC group than the poor WMC group (4 vs. 2.3 minutes, a small effect size). This means that the meaningfulness of the message could increase the willingness of good WMC listeners to stay in a noisy situation but less so for those with poor WMC.

This finding may be related to the motivation of the listeners.³¹ In a challenging condition, some listeners may have given up on the task and rated their effort for all test conditions similarly. Thus, the subjective ratings sampled at these conditions did not solely reflect the true difficulty of the task but were biased by the motivation, or lack thereof, of the listeners. Listeners in the poor WMC group may have perceived greater difficulties with the task and became more easily demotivated under some of the same test conditions than their good WMC peers. If so, this would suggest that listeners with poor WMC may be at a higher risk (than listeners with better WMC) of giving up in a communication task when it becomes difficult. The narrower range of effort ratings between HC and LC materials (i.e., CU) in the poor WMC listeners may suggest that these listeners have a smaller range of listening conditions where they may remain motivated. Kochkin³² reported that HA wearers' satisfaction for their HAs correlated with the number of listening situations in which they were successful. Thus, it is not unreasonable to speculate that listeners with a poorer WMC are more likely to be dissatisfied with their HAs. For these listeners, it is important that they are provided HA technology that can expand the range of listening situations they are engaged in. Technologies such as DIRMs (which improve SNR and effort rating), adaptive sound classifiers (which adapt HA processing automatically based on acoustic analysis), or multiple programs (fixed set of different frequency gain characteristics) may be beneficial.

In summary, CU was similar between WMC groups on the repeat task, but smaller for the poor WMC group on the recall, listening effort, and tolerable time tasks. On the recall task, the smaller CU (in the poor WMC group vs. the good WMC group) was the result of a lower score on both the LC

and HC sentences. On the other hand, the smaller CU in the poor WMC group on the listening effort and tolerable time tasks was the result of an “inflated” score on the LC sentences and a “deflated” score on the HC sentences (compare with the good WMC).

Conclusion

The current study supports the use of the RRT to evaluate CU. The study demonstrated that the amount of CU was a result of the interaction between the SNR of the test environment, the processing features on the HA, and the WMC of the listeners. In addition, the interaction of these factors on CU depends on the outcome measures used.

Conflict of Interest

All the authors are employees of WS Audiology.

References

- 1 Kuk F, Slugocki C, Korhonen P. (published ahead of print). An integrative evaluation of the efficacy of a directional microphone and noise reduction algorithm under realistic signal-to-noise ratios. *J Am Acad Audiol* 2020;31(04):262–270
- 2 Daneman M, Carpenter P. Individual differences in working memory and reading. *J Verbal Learn Verbal Behav* 1980;19(04):450–466
- 3 McCoy SL, Tun PA, Cox LC, Colangelo M, Stewart RA, Wingfield A. Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. *Q J Exp Psychol A* 2005;58(01):22–33
- 4 Zekveld AA, Rudner M, Johnsrude IS, Festen JM, van Beek JH, Rönnerberg J. The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear Hear* 2011;32(06):e16–e25
- 5 Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. *J Acoust Soc Am* 1995;97(01):593–608
- 6 Rönnerberg J, Rudner M, Foo C, Lunner T. Cognition counts: a working memory system for ease of language understanding (ELU). *Int J Audiol* 2008;47(Suppl 2):S99–S105
- 7 Dubno JR, Ahlstrom JB, Horwitz AR. Use of context by young and aged adults with normal hearing. *J Acoust Soc Am* 2000;107(01):538–546
- 8 Obleser J, Wise RJ, Dresner MA, Scott SK. Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci* 2007;27(09):2283–2289
- 9 Bilger RC, Nuetzel JM, Rabinowitz WM, Rzezczkowski C. Standardization of a test of speech perception in noise. *J Speech Hear Res* 1984;27(01):32–48
- 10 Elliott LL. Verbal auditory closure and the Speech Perception in Noise (SPIN) test. *J Speech Hear Res* 1995;38(06):1363–1376
- 11 Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 1977;61(05):1337–1351
- 12 Boothroyd A, Nitttrouer S. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am* 1988;84(01):101–114
- 13 Helfer KS, Freyman RL. Aging and speech-on-speech masking. *Ear Hear* 2008;29(01):87–98
- 14 Guediche S, Salvata C, Blumstein SE. Temporal cortex reflects effects of sentence context on phonetic processing. *J Cogn Neurosci* 2013;25(05):706–718
- 15 Bäckman L, Mäntylä T, Erngrund K. Optimal recall in early and late adulthood. *Scand J Psychol* 1984;25:306–314

- 16 Mäntylä T, Nilsson L. Are my cues better than your cues? Uniqueness and reconstruction as prerequisites for optimal recall of verbal materials. *Scand J Psychol* 1983;24:303–312
- 17 Sarampalis A, Kalluri S, Edwards B, Hafter E. Objective measures of listening effort: effects of background noise and noise reduction. *J Speech Lang Hear Res* 2009;52(05):1230–1240
- 18 Holmes E, Folkeard P, Johnsrude IS, Scollie S. Semantic context improves speech intelligibility and reduces listening effort for listeners with hearing impairment. *Int J Audiol* 2018;57(07):483–492
- 19 Cox RM, Alexander GC, Gilmore C. Development of the Connected Speech Test (CST). *Ear Hear* 1987;8(5, Suppl):119S–126S
- 20 Winn MB. Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends Hear* 2016;20:2331216516669723
- 21 Ricketts TA. Directional hearing aids. *Trends Amplif* 2001;5(04):139–176
- 22 Goldinger S. Auditory lexical decision. *Lang Cogn Process* 1996;11:559–568
- 23 Moulin A, Richard C. Lexical influences on spoken spondaic word recognition in hearing-impaired patients. *Front Neurosci* 2015;9:476
- 24 Slugocki C, Kuk F, Korhonen P. Development and clinical applications of the ORCA Repeat and Recall Test (RRT). *Hear Rev* 2018;25(12):22–26
- 25 Smeds K, Wolters F, Rung M. Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol* 2015;26(02):183–196
- 26 Wu YH, Stangl E, Chipara O, Hasan SS, Welhaven A, Oleson J. Characteristics of real world signal-to-noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear Hear* 2018;39(02):293–304
- 27 Keidser G, Dillon H, Flax M, Ching T, Brewer S. The NAL-NL2 prescription procedure. *Audiology Res* 2011;1(01):e24
- 28 Oeding K, Valente M. Differences in sensation level between the Widex Soundtracker and two real-ear analyzers. *J Am Acad Audiol* 2013;24(08):660–670
- 29 Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;53(04):695–699
- 30 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates Publisher; 1988
- 31 Matthen M. Effort and displeasure in people who are hard of hearing. *Ear Hear* 2016;37(Suppl 1):28S–34S
- 32 Kochkin S. Customer satisfaction with hearing instruments in the digital age. *Hear J* 2005;58(09):30–39

Appendix A Example of a list of high context (HC) and low context (LC) sentences

| |
|---|
| High context (HC) sentences |
| Keep the ice cream in the freezer |
| The chef cooks food in a restaurant |
| The barbecue grill used hickory wood |
| Wash the fruit in the sink . |
| The tart pie had too much lemon |
| He tried new foods in different countries |
| Low context (LC) sentences |
| Keep the ice foods in the lemon |
| The cream cooks food in a country |
| The barbecue chef used hickory freezer |
| Wash the grill in the restaurant |
| The tart fruit had too much wood |
| He tried new pie in different sinks |

Note: The bolded words are the target words.

