

Comparison of the Oldenburg Sentence Test and the German Hearing in Noise Test

Vergleich von Oldenburger Satztest und deutschem Hearing in Noise Test

Abstract

The well-established German Oldenburg Sentence Test (OLSA) and the Hearing in Noise Test (HINT) are frequently used in audiological research; the latter of which was recently published in German (Joiko et al. 2021). Both sentence tests examine speech intelligibility in noise, but differ in sentence structure, calibration and scoring procedure. In this paper, the comparability of the measurement results of both tests, and differences in response time (an objective measure related to the listening effort) will be investigated. A total of ten test subjects with normal hearing took part in the study. The speech signal was presented from 0°. The background noise was presented from 180° at a level of 65 dB(A) for the HINT and 65 dB SPL for the OLSA. Following initial training, the signal-to-noise ratio (SNR) of the speech recognition threshold (SRT) at 50% was determined applying an adaptive measurement procedure with a starting SNR of 0 dB. The evaluation is done according to the standards of the OLSA and HINT, with word scoring and sentence scoring, respectively. Speech intelligibility was measured at various fixed SNRs (Harianawala 2019) and the measurement results were analyzed applying both sentence and word scoring. Psychometric functions were fitted to the measured values and 70% thresholds were estimated. The response time of the test participants was determined as the period between the end of the presented test sentence and the beginning of the spoken repetition of the test sentence. Results revealed a significantly lower SNR for the SRT 50% threshold for the OLSA than for the HINT. This can be explained by the evaluation differences between sentence and word scoring methods. When analyzing the SNR-dependent speech intelligibility and the 70% thresholds of the psychometric functions, it was shown that the OLSA and the HINT are not significantly different if the same scoring procedure is applied. We conclude that the results of the OLSA and the HINT could be comparable if measurement and scoring procedures are considered. When analyzing response time, it was found that an increasing SNR results in a decrease in response time. It was also determined that the response times of the OLSA are significantly lower than those of the HINT. The lower response times for the OLSA may be an indication of lower listening effort and lower cognition compared to the HINT.

Keywords: HINT, OLSA, German Hearing in Noise Test, German Matrix Test, response time

Zusammenfassung

Häufig in der audiologischen Forschung verwendete Satztests sind der etablierte Oldenburger Satztest (OLSA) und der kürzlich in Deutsch publizierte Hearing in Noise Test (HINT) (Joiko et al. 2021). Beide Satztests untersuchen die Sprachverständlichkeit im Störgeräusch, unterscheiden sich jedoch in Aufbau, vorgegebener Kalibrierung und Bewertung. In dieser Arbeit wurde die Vergleichbarkeit der beiden Tests untersucht.

Franziska Schweikert^{1,2}
Steffen Kreikemeier²
Lena Eipert¹

1 WS Audiology, Erlangen,
Germany

2 Hochschule Aalen, Germany

Zusätzlich wurden Unterschiede in der Antwortzeit, einer objektiven Messgröße zusammenhängend mit der Höranstrengung, untersucht. An der Studie nahmen insgesamt zehn normalhörende Testpersonen teil. Das Sprachsignal wurde aus 0° präsentiert. Das Störgeräusch wurde aus 180° mit einem Pegel von 65 dB(A) beim HINT beziehungsweise 65 dB SPL beim OLSA dargeboten. Einem anfänglichen Training folgend, wurde der SNR der Sprachverständlichkeitsschwelle 50% (SRT 50%) durch adaptive Messung mit einem Start-SNR von 0 dB bestimmt. Die Bewertung erfolgt standardisiert beim OLSA je Wort (word scoring), beim HINT je Satz (sentence scoring). Des Weiteren wurde die Sprachverständlichkeit bei verschiedenen festen SNR gemessen (Harianawala 2019) und die daraus resultierenden Messergebnisse sowohl mit sentence als auch mit word scoring analysiert. Eine psychometrische Funktion wurde angepasst und deren 70%-Schwelle geschätzt. Die Antwortzeit wurde jeweils zwischen dem Ende des präsentierten Testsatzes und dem Beginn der verbalen Wiederholung des Satzes ermittelt. Die Ergebnisse haben gezeigt, dass die SRT 50% des OLSAs bei signifikant niedrigeren SNR liegt als die des HINTs. Dies kann durch die Unterschiede durch sentence und word scoring erklärt werden. Bei der Analyse der SNR-abhängigen Sprachverständlichkeit und der 70%-Schwellen geschätzt aus den psychometrischen Funktionen hat sich gezeigt, dass der OLSA und der HINT bei gleichem Bewertungsverfahren nicht signifikant unterschiedlich sind. Bei der Analyse der Antwortzeit wurde festgestellt, dass ein steigender SNR zu einer Verringerung der Antwortzeit führt. Außerdem hat sich herausgestellt, dass die Antwortzeiten des OLSAs signifikant niedriger sind als des HINTs. Die niedrigeren Antwortzeiten beim OLSA könnten ein Hinweis auf eine niedrigere Höranstrengung und Kognition im Vergleich zum HINT sein.

Schlüsselwörter: HINT, OLSA, deutscher Hearing in Noise Test, Oldenburger Satztest, Antwortzeit

1 Introduction

Sentence tests have an important role in audiological research. A suitable sentence test is selected depending on the question or measurement requirement. In addition, different tests could be selected for similar questions depending on availability or standard procedures. It is often difficult to compare the results of different sentence tests due to differences in test sentences and evaluation criteria. One test frequently used in Germany is the well-established Oldenburg Sentence Test (OLSA), also known as the German Matrix Test. Internationally, the Hearing in Noise Test (HINT) is used more commonly, which was recently published in German [1], [2]. Both sentence tests can be conducted in quiet or in noise. The speech level or the signal-to-noise ratio (SNR) can be set at a fixed value or adaptively determined for the speech recognition threshold (SRT) at a given percentage. The OLSA stimuli are compiled from senseless matrix sentences (name-verb-number-word-adjective-object) [3]. In contrast, the HINT consists of meaningful everyday sentences with a length between three to seven words [1]. A major difference between the HINT and the OLSA is the method to evaluate correctness. The OLSA is evaluated with word scoring as standard [3]; correctness is calculated by dividing the number of correctly reproduced words by the total number of words in a sentence. The HINT is evalu-

ated using sentence scoring [1], [2], i.e., if a single word is not reproduced correctly, the entire sentence is considered incorrect. This difference is particularly significant when measuring SRT, as the speech level is adaptively adjusted depending on the correctness of the subject's response.

In literature pertaining to the HINT and the OLSA, tests are conducted via headphones with speech and noise from simulated 0° (from generic HRTFs). The SRT 50% of the HINT is at a higher SNR (-6.0 dB) than of the OLSA (-7.1 dB) [1], [4], [5]. Versfeld et al. [6] discovered, during the development of their own sentence test, that the SRT 50% with word scoring is at a lower SNR than with sentence scoring. Therefore, it can be predicted that the SRT 50% of the HINT measured in this study will be at a higher SNR than that of the OLSA. Zinner et al. [7] conducted comparative measurements of five German speech tests and determined that the measurement conditions, i.e., the setup configuration, and the different masking effects of the noise due to spectral imbalances were more critical to the speech intelligibility than the speech material itself. In this study measurement conditions were matched as closely as possible between both tests. Therefore, the assumption is that there are no significant differences in the speech intelligibility at fixed SNRs and the 70% thresholds of the psychometric function of the HINT and the OLSA with the same evaluation

procedure. Houben et al. [8] analyzed if the response time changes when measuring different SNRs. They concluded that the response time decreases with increasing SNR even if the speech intelligibility is already optimal. In addition to the differences in speech intelligibility, this study also investigates whether the response time differs between the HINT and the OLSA. Response time is an objective measure related to the listening effort, because it may show the processing time and may be an indicator of the cognitive effort that is needed when conducting a speech intelligibility test (e.g. [8], [9]).

2 Methods

In all measurements, the speech signal was presented from the front (0°) and the noise from the back (180°). This setup configuration was selected because it is a common setting for testing hearing aid features. The corresponding speech-shaped noise was presented continuously at 65 dB(A) for the German HINT and 65 dB SPL for the OLSA [2], [3]. It was randomized whether the HINT or the OLSA was conducted first. For both the HINT and the OLSA test lists including 20 sentences each were used. Before the measurements of each test, a training was conducted in quiet for the HINT with one test list and for the OLSA with two test lists [1], [5]. Then, the SRT 50% was determined adaptively with a starting SNR of 0 dB. The measurement was conducted using sentence scoring for the HINT and word scoring for the OLSA. To fit a psychometric function, speech intelligibility was determined at fixed SNRs. The SNRs chosen were -12 dB, -9 dB, -6 dB, -3 dB, 0 dB and 3 dB, based on a study by Hari-anawala et al. [10]. The order of the individual SNRs was randomized. Correctness was evaluated with word scoring for both the OLSA and the HINT, allowing the results of both sentence tests to be additionally evaluated with sentence scoring. A psychometric function for both word and sentence scoring was fitted to the speech intelligibility values of all measured SNRs for each test subject using a cumulative normal distribution function and 70% speech intelligibility thresholds were estimated. To calculate the response time between the end of the test sentence and the beginning of the spoken response, the speech signal and the test subjects' responses were recorded and analyzed. The envelope of the analytical speech signal was detected exceeding a defined threshold calculated from the median and the standard deviation of the noise floor. The response time was calculated as the time between the end of the test sentence and the beginning of the spoken response and averaged across all test sentences per SNR measurement.

3 Results

This study was conducted with a total of ten subjects (average age 29.6 years) with normal hearing.

3.1 SRT 50%

The results of the SRT 50% measurements of the OLSA and the HINT (Figure 1) show that the SRT 50% for the HINT with a mean of -8.15 dB is at a significantly higher SNR than for the OLSA with a mean of -10.20 dB (t-test for dependent samples, $p < 0.0005$).

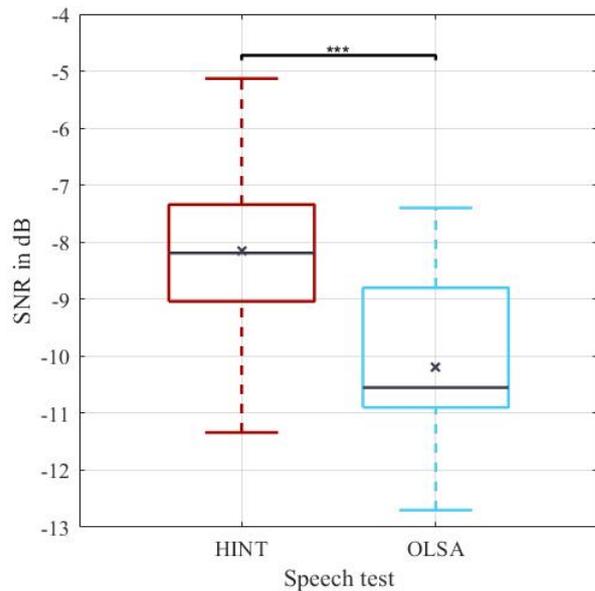


Figure 1: SRT 50% for the HINT and the OLSA
The boxplots show the mean (grey cross), median (grey line), upper and lower quartiles (box) and the values within 1.5 times the interquartile range (whiskers). The asterisks (***) indicate a statistically significant difference within $p \leq 0.0001$.

3.2 Speech intelligibility at fixed SNRs and estimated 70% threshold of the psychometric functions

Speech intelligibility at fixed SNRs, shown in Figure 2, revealed no significant differences between the HINT and the OLSA for the measured SNRs with the same evaluation procedure, i.e., either word or sentence scoring (Wilcoxon signed-rank test). For the estimated 70% thresholds, shown in Figure 3, a one-way ANOVA with repeated measures revealed significantly lower thresholds ($p < 0.0005$) with word scoring (mean of the HINT: -8.9 dB; mean of the OLSA: -9.0 dB) than with sentence scoring (mean of the HINT: -6.7 dB; mean of the OLSA: -6.8 dB). No significant difference was found for the 70% thresholds between the sentence tests.

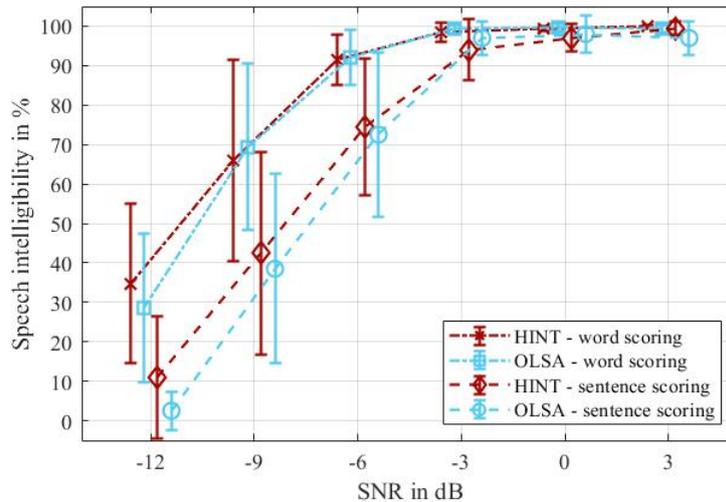


Figure 2: Speech intelligibility at fixed SNRs for the HINT and the OLSA. The error bars display the mean values and standard deviations.

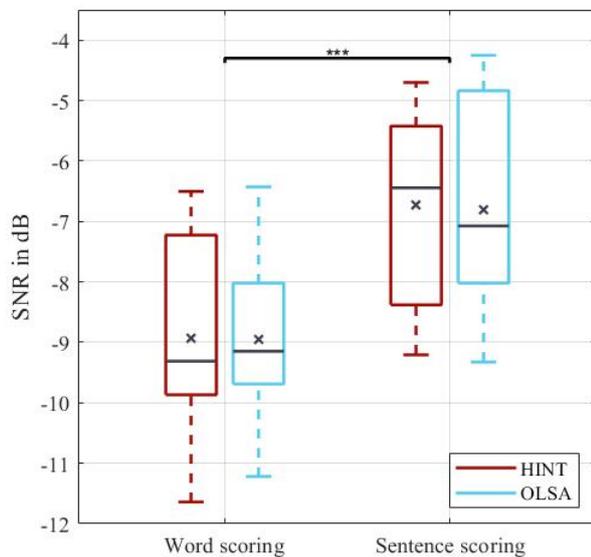


Figure 3: 70% thresholds of the psychometric function. The two boxplots on the left show the thresholds with word scoring and the two on the right with sentence scoring for the HINT and the OLSA.

3.3 Response times

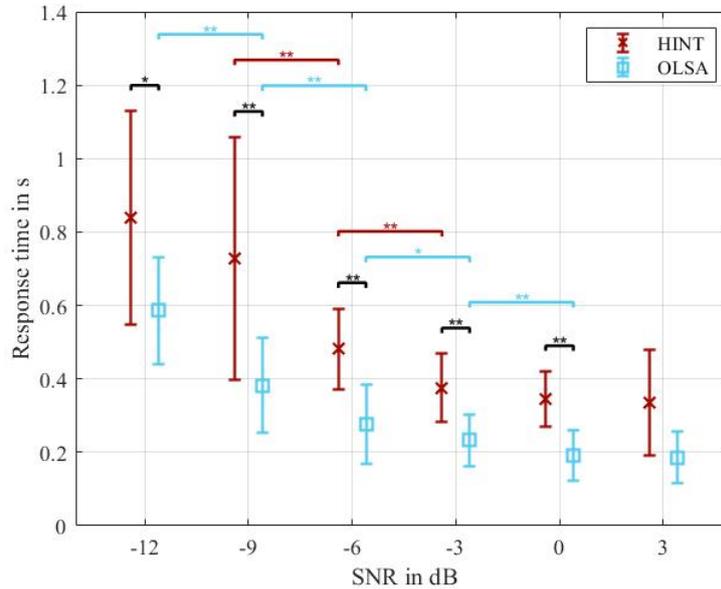
The response times for both the HINT and the OLSA significantly decreased with increasing SNRs (Friedman test: HINT: $p < 0.0005$; OLSA: $p < 0.0005$; Wilcoxon signed-rank test: results in Table 1). When comparing the response times between the OLSA and the HINT, the response times for the OLSA are significantly lower than those for the HINT, shown in Figure 4 (Friedman test for dependent samples, $p < 0.001$). Further pairwise comparisons indicated shorter response times for the OLSA than for the HINT (Wilcoxon sign-rank test) for the SNR -12 dB ($p = 0.017$), -9 dB ($p = 0.005$), -6 dB ($p = 0.005$), -3 dB ($p = 0.005$) and 0 dB ($p = 0.007$).

4 Discussion

The SRT 50% measure revealed higher SNR thresholds for the HINT with sentence scoring than for the OLSA with word scoring. These differences are in line with the previously reported differences in SNR thresholds between word and sentence scoring [6]. The SRT 50% of the HINT and the OLSA measured in this study revealed on average lower SNRs (3.10 dB for the OLSA and 2.15 dB for the HINT) than the published values of the two sentence tests [1], [4], [5]. The reasons for the deviations from the literature values are largely due to the different measurement conditions (e.g. noise from 180° in this study compared to 0° in previous work). Using the same evaluation the speech intelligibility at fixed SNRs and the 70% thresholds estimated from the psychometric functions of the HINT and the OLSA did not differ significantly. The differences in the mean 70% thresholds of the OLSA and the HINT with the same scoring were relatively low with 0.07 dB for sentence scoring and with 0.04 dB for word scoring. This leads to the conclusion that the measurement results of the HINT and the OLSA are comparable when evaluated using the same scoring procedure. The results of this study are consistent with the conclusions from Zinner et al. [7] that the setup configuration and the noise were more critical to the speech intelligibility than the speech material itself. The response times decreased with increasing SNR, as in Houben et al. [8], although not all differences were significant. Another notable aspect of the results is that the response times for the OLSA were significantly lower than for the HINT. An explanation for this could be that when measuring the OLSA, less thought is given to the sentences because the sentences always have the same structure than with the HINT, where an attempt is made to understand the meaning of the sentences. Thus, this could be an indication of lower listening effort and lower cognition when measuring the OLSA.

Table 1: P-values of the Wilcoxon sign-rank test of the response time

SNRs	-12 dB -9 dB	-9 dB -6 dB	-6 dB -3 dB	-3 dB 0 dB
HINT	n.s.	p=0.007	p=0.009	n.s.
OLSA	p=0.005	p=0.007	p=0.047	p=0.009

**Figure 4: Response times**

The response time is shown for the HINT and the OLSA for all measured SNRs. The stars show statistically significant differences (*: $p \leq 0.05$; **: $p \leq 0.001$).

Notes

Conference presentation

This contribution was presented at the 26th Annual Conference of the German Society of Audiology and published as an abstract [11].

Competing interests

The authors declare that they have no competing interests.

References

- Joiko J, Bohnert A, Strieth S, Soli SD, Rader T. The German hearing in noise test. *Int J Audiol.* 2021 Nov;60(11):927-33. DOI: 10.1080/14992027.2020.1837969
- Nilsson M, Soli SD, Sullivan JA. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am.* 1994 Feb;95(2):1085-99. DOI: 10.1121/1.408469
- Wagener K, Kühnel V, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Z Audiol.* 1999;38(1):4-15.
- Kollmeier B, Warzybok A, Hochmuth S, Zokoll MA, Uslar V, Brand T, Wagener KC. The multilingual matrix test: Principles, applications, and comparison across languages: A review. *Int J Audiol.* 2015;54(Suppl 2):3-16. DOI: 10.3109/14992027.2015.1020971
- Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests. *Z Audiol.* 1999;38(3):86-95.
- Versfeld NJ, Daalder L, Festen JM, Houtgast T. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am.* 2000 Mar;107(3):1671-84. DOI: 10.1121/1.428451
- Zinner C, Winkler A, Holube I. Vergleich von fünf Sprachtests im sprachsimulierenden Störgeräusch. *Z Audiol.* 2021 Jun;60(4):138-48. DOI: 10.3205/zaud000016
- Houben R, van Doorn-Bierman M, Dreschler WA. Using response time as a measure for listening effort. *Int J Audiol.* 2013 Sep;52(11):753-61. DOI: 10.3109/14992027.2013.832415
- Alhanbali S, Dawes P, Millman RE, Munro KJ. Measures of Listening Effort Are Multidimensional. *Ear Hear.* 2019 Sep/Oct;40(5):1084-97. DOI: 10.1097/AUD.0000000000000697
- Harianawala J, Galster J, Hornsby B. Psychometric Comparison of the Hearing in Noise Test and the American English Matrix Test. *J Am Acad Audiol.* 2019 Apr;30(4):315-26. DOI: 10.3766/jaaa.17112
- Schweikert F, Kreikemeier S, Eipert L. Vergleich von Oldenburger Satztest und Hearing in Noise Test. In: Deutsche Gesellschaft für Audiologie e.V., editor. 26. Jahrestagung der Deutschen Gesellschaft für Audiologie. Aalen, 06.-08.03.2024. Düsseldorf: German Medical Science GMS Publishing House; 2024. Doc183. DOI: 10.3205/24dga183

Corresponding author:

Franziska Schweikert
WS Audiology, Henri-Dunant-Str. 100, 91058 Erlangen,
Germany
franziska.schweikert@gmx.de

Please cite as

Schweikert F, Kreikemeier S, Eipert L. Comparison of the Oldenburg Sentence Test and the German Hearing in Noise Test. *GMS Z Audiol (Audiol Acoust)*. 2024;6:Doc20.
DOI: 10.3205/zaud000055, URN: urn:nbn:de:0183-zaud0000553

This article is freely available from
<https://doi.org/10.3205/zaud000055>

Published: 2024-11-27

Copyright

©2024 Schweikert et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.