DOI: 10.61782/fa.2023.0497

Adaptive CCOLSA: cognitive overload thresholds in a multi-talker speech test

Jan Heeren[1,2*] **Volker Hohmann**[2,3] **Michael Schulte**[1,2]
Kirsten C. Wagener[1,2]
[1] Hörzentrum Oldenburg gGmbH, Oldenburg, Germany
[2] Cluster of Excellence Hearing4all, Oldenburg, Germany
[3] Carl-von-Ossietzky-Universität Oldenburg, Oldenburg, Germany

ABSTRACT

The Concurrent OLSA (CCOLSA) is a multi-talker speech test that uses a temporal overlap between sentences from alternating talkers to achieve sensitivity at high signal-to-noise ratios (SNR). The paradigm is a dual-task design that combines call-sign detection and speech recognition for switching target talkers. This study proposes an adaptive CCOLSA approach, where the overlap parameter is adjusted adaptively to measure individual thresholds with 50 percent speech recognition at fixed SNR. For this, the adaptive procedure used in the corresponding standard matrix test is applied with a transformation of SNR steps to overlap time steps. As outcome measure, a consolidated result value is determined, which is called CCOLSA costs and includes the difference between the presented SNR and the individual Speech-Reception Threshold (SRT) as well as the difference of the individual overlap time compared to the median overlap time of normal-hearing subjects. The approach was tested for 31 elderly participants with various hearing profiles. Additionally, the Comprehensive Trail Making Test 2 (CTMT-2) was conducted. The measured individual CCOLSA costs at threshold amount 4.0-7.1 dB with a median test-retest deviation of 0.4 dB. Interactions between CCOLSA costs, SRTs and CTMT-2 results are discussed.

[*]***Corresponding author***: heeren@hz-ol.de.

Keywords: *speech recognition, cognitive load, multi-talker test, cocktail party*

1. INTRODUCTION

Heeren et al. [7] recently introduced the Concurrent OLSA Test (CCOLSA), which is a multi-talker speech test, where matrix sentences of the German Matrix Test [15] are presented alternately from three talkers. The sentences are overlapping across talkers so that the last word of the current sentence is presented simultaneously with the first word of the following sentence. During a continuous presentation of sentences, participants have to perform the task: „Repeat the last words of all sentences from the talker, who started a sentence with the name ‚Kerstin' the latest." Thus, the name „Kerstin" is used as a call sign, indicating target talker changes when occurring. Due to the simultaneous presentation of last and first words, call-sign detection and target word recognition are competing tasks within a dual-task paradigm. For more details on the method, please see [7].

In CCOLSA, the subjects' speech-recognition performance depends on the overlap time of the sentences. This overlap time corresponds to the overall speech rate, while the presentation speed of each sentence stays constant. Since cognitive load correlates with speech rate [9], the overlap time also corresponds to cognitive load.

This study proposes measuring thresholds with 50 percent correct responses based on an adaptive adjustment of the overlap time. For a basic exploration of this approach, following research questions were investigated:

1. Does the approach lead to reliable results?

2. What is the relationship between CCOLSA scores and classical single-talker SRTs?
3. Can the results be used as an individual cognitive performance measure?

Results might correlate with cognitive tests that correlate with speech-recognition measures. A previous study [10] showed that, compared with other cognitive tests, the Trail Making Test (TMT) [14] showed the highest correlation with CCOLSA results. Furthermore, TMT shows significant correlation with Speech-Reception Thresholds (SRT) [4]. Therefore, the CTMT-2 (Comprehensive Trail Making Test 2, extended version of TMT) [12] is included in the experiment to potentially support research question 3.

2. METHOD

2.1 Adaptive procedure

The original CCOLSA measurement software was supplemented with a stage that adaptively adjusts the overlap time of sentences for each new call sign. The adjustments are computed using a modified version of the adaptive procedure A1 by Brand and Kollmeier [1], which is also used in the matrix tests of the medical product „Oldenburg Measurement Applications“ (Hörzentrum Oldenburg gGmbH, 2022). The implementation for this study includes the following modifications:

2.1.1 Modification 1: intelligibility as a function of the overlap time instead of SNR

The original procedure of Brand and Kollmeier [1] adapts the intelligibility as a function of the SNR towards 50 percent speech recognition. In adaptive CCOLSA, the SNR term of the procedure was replaced by:

$$\Delta SNR = \frac{m_{CCOLSA}}{m_{SRT}} \Delta t_{Overlap} \tag{1}$$

The formula includes the slope of the SNR-based intelligibility function:

$$m_{SRT} = 12 \frac{\%}{dB}$$

and the slope of the intelligibility function in CCOLSA:

$$m_{CCOLSA} = 32.5 \frac{\%}{s}$$

The m_{SRT} value was defined based on a literature reference [3, noise condition MT]. m_{CCOLSA} was derived from the dual-task results (including missed-call-sign effects) of Heeren et al. [7].

2.1.2 Modification 2: last words of sentences used for word scoring

The original procedure adjusts the SNR based on word scoring for sentences of five words. In CCOLSA, only the last words of target sentences are repeated by the test subjects, but there is a sequence of 2-4 target sentences between two call signs. The correct responses ratio for such a sequence are used as word score.

2.1.3 Modification 3: SRTs are used for fixed SNR

Due to the target talker switches in CCOLSA, the intelligibility of the three talkers must be equal to achieve the same effect of overlap time adjustments across talkers. Thus, it is mandatory to measure SRTs for the three talkers in advance and use an SRT-based presentation level for each individual subject. In this study, talker levels of SRT + 5 dB were used. The original procedure starts at the fixed SNR value of 0 dB for all subjects and adapts towards the SRT.

2.1.4 Modification 4: minimal step size only

The step sizes of the original procedure are large in the beginning to account for large SNR variances in individual speech recognition. Later, smaller step sizes are used to increase the test resolution. As adaptive CCOLSA already uses an individual SNR value (SRT + 5 dB), large SNR variances are not expected. Consequently, large step sizes were not applied as they might lead to a decrease of the test resolution.

2.2 Participants

31 subjects with various hearing profiles participated in the experiment (13 male, 18 female). The ages ranged from 56-84 years with a mean of 71.9 years. Nine of the participants were normal-hearing (NH) showing hearing thresholds of <20 dB HL. The other 22 participants showed hearing losses with pure tone averages (PTA4; frequencies 500, 1k, 2k, 4k Hz) of 29-54 dB HL. Hearing-impaired (HI) and normal-hearing participants were matched in age, showing mean ages of 70.3 years (NH) and 72.6 years (HI), respectively. Nine of the HI subjects own and use hearing aids daily (HA users), whereas thirteen did not own hearing-aids (non-HA users). The mean ages for these groups are 72.4 years (HA users) and 72.8 years (non-HA

users). The experiment was approved by the ethics committee ("Kommission für Forschungsfolgen-Abschätzung und Ethik") of the Carl von Ossietzky University in Oldenburg, Germany.

2.3 Setup

Measurements were conducted in a free-field lab at the Hörzentrum Oldenburg, Germany. The room has dimensions of 5 m x 5.25 m x 2.50 m and a reverberation time of approximately 0.2 s. Participants were seated in the center of a horizontal loudspeaker array with 24 Genelec 8030B loudspeakers, that were set up on a circle in steps of 15 degree (starting from 0 degree). The circle had a radius of 2 m and the height of the tweeters was 1.25 m. Measurements were conducted using Matlab 2013. The implementation of the test was based on the version used in Heeren et al. [7], but included the additional adaptive stage described above. The speech signals were mapped to the loudspeaker outputs, directly, and the diffuse cafeteria noise, which was presented at 68 dB SPL, was panned in 5th order ambisonics using TASCAR [6].

2.4 Measurement Procedure

First, all participants completed the CTMT-2, followed by SRT measurements for the three CCOLSA talkers. The SRTs were used to determine the SNR for CCOLSA, which was set to the individual SRT + 5 dB. Third, a training for CCOLSA was performed. Afterwards, the main adaptive CCOLSA measurements were conducted. Retests were measured for a subgroup of the HI participants (N=9). All hearing-impaired subjects were measured in unaided conditions.
To allow for a comparison of the adaptive CCOLSA results with SRTs, the outcome measure of "CCOLSA costs" was defined as:

$$CCOLSA\ costs = (SNR_{CCOLSA} - SRT) + \frac{m_{CCOLSA}}{m_{SRT}} \left(t_{Overlap} - t_{Ov.\ median\ NH}\right) \quad (2)$$

3. RESULTS

In general, the observed CCOLSA costs range from 4.0-7.1 dB. The retest results (N=9) show a training effect compared with the test results of 0.4 dB in median CCOLSA costs (see Fig. 1), which is significant (t-test, p=0.017). In Fig. 2, CCOLSA costs for the three participant groups NH, non-HA users, and HA users are displayed. The groups performed significantly different (ANOVA, F(1, 30)=5.08, p=0.032). Posthoc, the difference between NH and HA users was significant (t-test, p=0.05, Bonferroni corrected).

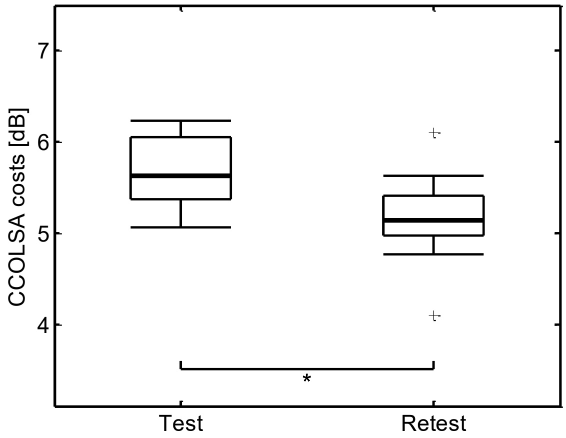

Figure 1. CCOLSA costs (medians and interquartile ranges) for test and retest; measured for a subgroup of hearing-impaired subjects (N=9); statistical significance is marked by asterisk (* p≤0.05, T-Test).

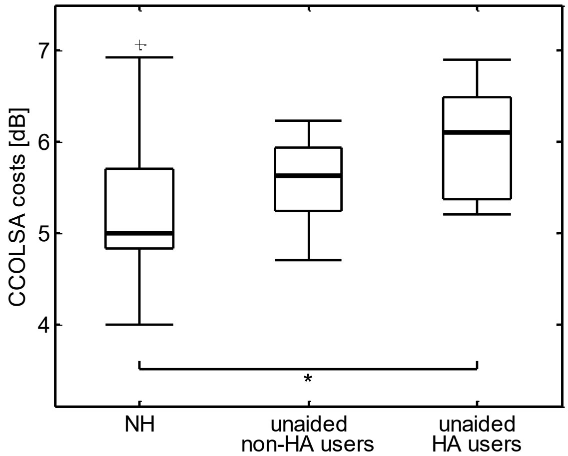

Figure 2. CCOLSA costs (medians and interquartile ranges) for normal-hearing subjects (NH, N=9) and hearing-impaired subjects grouped by hearing aid user status (non-HA users, N=13; HA users, N=9); results statistical significance is marked by asterisk (* p≤0.05, T-Test); all measurements were performed unaided.

The CCOLSA costs were analyzed for effects of age and hearing loss using linear regression models (see Fig. 3 and Fig. 4). The data significantly depends on both age ($p=0.014$, $F(1, 30)=6.75$, $R^2=0.184$) and hearing loss ($p=0.049$, $F(1, 30)=4.23$, $R^2=0.124$).

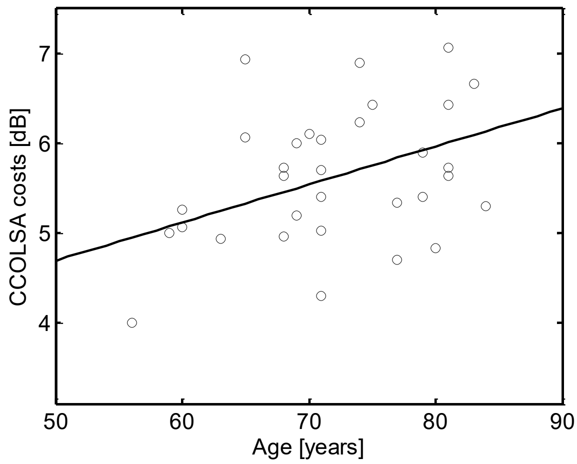

Figure 3. Individual CCOLSA costs and linear model as a function of age; the effect is statistically significant ($p<0.05$, F-Test).

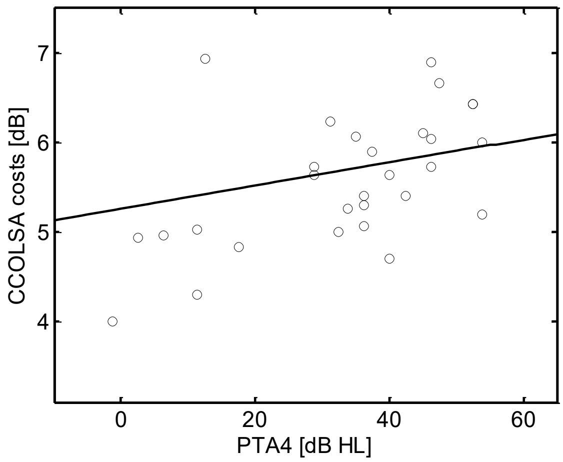

Figure 4. Individual CCOLSA costs and linear model as a function PTA4; the effect is statistically significant ($p<0.05$, F-Test).

CTMT-2 results (see Fig. 5) also show a significant age effect in response times for test part B ($p=0.027$, $F(1, 28)=5.46$, $R^2=0.163$). For CTMT-2 results and CCOLSA costs no correlation was found.

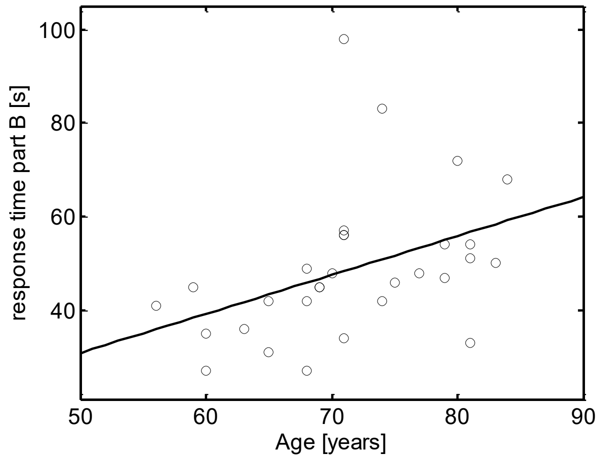

Figure 5. Individual CTMT-2 response times (part B) and linear model as a function of age; the effect is statistically significant ($p<0.05$, F-Test).

4. DISCUSSION

The adaptive overlap time procedure reliably converged towards the targeted threshold of 50 percent correct responses. Individual CCOLSA costs compared to SRTs are between 4 dB and 7 dB. Retest results show a decrease of 0.4 dB, which is comparable to the training effect in SRTs observed for the Oldenburg matrix corpus [13,15]. Furthermore, the variance of the individual CCOLSA costs does not exceed the standard deviation of the Matrix Test SRTs, which is approximately 0.6 dB [15].

Significant differences in CCOLSA costs were found between the groups NH and HA users (unaided measurements). Thus, the test is generally suitable to resolve group effects. As the groups were matched in age and all measurements were conducted without hearing aids, the main difference between the groups is hearing loss. However, the influence of peripheral hearing loss on speech recognition was equalized by using presentation levels based on individual SRTs, as well. There are several reasons for assuming that CCOLSA costs are a cognitive measure. First, the overlap time as the main variable of the test mainly affects the speech rate, which correlates with cognitive load [9]. According to Füllgrabe et al. [4], cognitive abilities and sensitivity to binaural cues can explain 68 percent of the differences in speech recognition for NH subjects. In this study, sensitivity to binaural cues is not assumed to have an effect, because talkers are spatially separated and have different voices (male 1 at -60 degree, female at 0 degree, male 2 at 60 degree). Due to the overlap construction, the number of concurrent talkers is two during the target word presentations for 90 percent of the cases (complete length of the word), while the other 10 percent still have a partial overlap of 150-300 ms. Moreover, binaural cues do not interfere with cognitive effects observed relative to individual SRTs [11, 16]. A plausible explanation for the poorer performance of the HA users (unaided measurements) would be that "increased demands due to hearing loss can result in changes in neural resource allocation, reducing available resources for cognitive function." [5]. The correlation of CCOLSA costs with age and with PTA4 also fit to this theory. Unfortunately, the CTMT-2 data does not show a correlation with CCOLSA costs. According to [4], TMT part B (included in CTMT-2) correlates with SRTs for NH subjects. Probably, the individually chosen SNRs in adaptive CCOLSA compensated the expected correlation. Still, there is some common ground, as the CTMT-2 data also shows an age effect. Likely, random effects included in both measures seem to be so high that this common effect is masked in their correlation. In addition, the CTMT-2 data variance in the test group was rather low and therefore also impede robust correlation analysis. For clarification, more research is necessary.

The approach has limitations regarding the individual ability to perform the task that is quite demanding. After participants are instructed, a training is performed, which on the one hand proves the general ability of the participant to perform the task and on the other hand reduces training effects in following measurements. So far, one person out of 173 (seven studies) had to be excluded because he could not perform the task. In this study, all participants successfully performed the training without prior testing or selection. However, the task is very demanding and it is recommended not to test more than 5-6 conditions per session to avoid mental fatigue effects. Further investigation is needed to assess which cognitive domains are reflected in CCOLSA costs. As the promising candidate TMT did not show any correlation, CCOLSA costs might show other dependencies of cognitive domains than SRTs (for an overview, see [2]). Another limitation is that the test is only validated for German language, yet. Since the Matrix test format is available for various languages [8], it could easily be translated to other languages.

5. ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungs-Gemeinschaft (DFG, German Research Foundation) – Project-ID 352015383 – SFB 1330 C4.

6. REFERENCES

[1] T. Brand and B. Kollmeier, "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *J. Acoust. Soc. Am.,* 111(6), pp. 2801–2810, 2002.

[2] A. Dryden, H.A. Allen, H. Henshaw, and A. Heinrich: "The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis," *Trends in hearing*, 21, pp. 1–21, 2017.

[3] T. Francart, A. Van Wieringen, and J. Wouters: "Comparison of fluctuating maskers for speech recognition tests," *Int. J. Audiol.*, 50(1), pp. 2–13, 2011.

[4] C. Füllgrabe, B. Moore, and M.A. Stone: "Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition," *Front. Aging Neurosci.*, 6 (347), pp. 1–25, 2015.

[5] S.E. Fulton, J.J. Lister, A.L.H. Bush, J.D. Edwards, and R. Andel: "Mechanisms of the hearing–cognition relationship," *Seminars in hearing*, 36(3), pp. 140–149, Thieme medical Publishers, 2015.

[6] G. Grimm, J. Luberadzka, and V. Hohmann: "A toolbox for rendering virtual acoustic environments in the context of audiology," *Act. Acust. United Acust.*, 105(3), pp. 566–578, 2019.

[7] J. Heeren, T. Nuesse, M. Latzel, I. Holube, V. Hohmann, K.C. Wagener, and M. Schulte: "The Concurrent OLSA test: A method for speech recognition in multi-talker situations at fixed SNR," *Trends in Hearing*, 26, pp. 1–12, 2022.

[8] B. Kollmeier, A. Warzybok, S. Hochmuth, M.A. Zokoll, V. Uslar, T. Brand, and K.C. Wagener: "The multilingual matrix test: Principles, applications, and comparison across languages: A review," *Int. J. Audiol.*, 54(sup2), 3-16, 2015.

[9] J.A. Müller, D. Wendt, B. Kollmeier, S. Debener, and T. Brand: "Effect of speech rate on neural tracking of speech," *Front. Psych.*, 10 (149), pp. 1–15, 2019.

[10] T. Nuesse, J. Heeren, M. Latzel, M. Schulte, and I. Holube: "Cognitive abilities affect speech recognition in multi-talker situations of young and elderly normal-hearing listeners using the Concurrent OLSA test," in prep.

[11] J. Rennies and G. Kidd Jr.: "Benefit of binaural listening as revealed by speech intelligibility and listening effort," *J. Acoust. Soc. Am.*, 144(4), pp. 2147–2159, 2018.

[12] C.R. Reynolds: "Comprehensive Trail Making Test-Second Edition (CTMT-2)," *PAR*, 2019.

[13] A. Schlüter, U. Lemke, B. Kollmeier, and I. Holube: "Normal and time-compressed speech: How does learning affect speech recognition thresholds in noise?," *Trends in Hearing*, 20, pp. 1–13, 2016.

[14] O. Spreen and E. Strauss: "A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary," *Oxford University Press*, New York, 1998.

[15] K.C. Wagener, T. Brand, and B. Kollmeier: "Entwicklung und evaluation eines Satztests in deutscher Sprache III: Evaluation des Oldenburger Satztests," *Z. Audiol.*, 38(3), pp. 86–95, 1999.

[16] A.A. Zekveld, M. Rudner, S.E. Kramer, J. Lyzenga, and J. Rönnberg: "Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech," *Front. Neurosci.*, 8(88), pp. 1–11, 2014.