MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Speech perception of children and adults in a dual-task paradigm under noisy conditions

Loh, Karin[1] and Fels, Janina[2]
Teaching and Research Area of Medical Acoustics, Institute of Technical Acoustics, RWTH Aachen University
Kopernikusstraße 5, 52074 Aachen, Germany

ABSTRACT

Human hearing benefits from several binaural and monaural cues. Therefore, it is possible to hear and understand speech in complex acoustic scenes. Children spend a considerable amount of time in educational buildings, where mostly a highly complex acoustic scene is existing. Previous studies have already shown, that children have to expend more effort to recognize speech in these situations, especially in noisy situations. However, these studies were conducted using adult target speech and multi-talker babble based on adults' voices as background noise, but a realistic scene consists of adults as well as children voices and relevant and irrelevant noises. This work presents an experiment using a dual-task paradigm to assess children speech perception and listening effort under different noise conditions considering aurally-accurate reproduction methods. The same experiment has been carried out with adults in a further step as a comparison with the initial children run.

Keywords: Noise, Educational buildings, Listening effort
I-INCE Classification of Subject Number: 61, 63

1. INTRODUCTION

Noise in educational institutions is a well-known problem. Different studies on noise exposition in educational institutions revealed high noise levels [13, 15], which also affect learning abilities and the development of children. An overview of effects caused by noise can be found in the work by Shield and Dockrell [14] and by Stansfeld and Lercher [16]. However, more and more children are joining educational institutions at an early age and therefore, a considerable part of children's education and development takes place in highly noisy acoustic situations.

A paradigm to directly analyze effects caused by noise is the dual-task paradigm, which combines two important factors for hearing in noise: the speech recognition and

[1]karin.loh@akustik.rwth-aachen.de
[2]janina.fels@akustik.rwth-aachen.de

the cognitive capacity taken together as listening effort. The dual-task paradigm was developed based on the assumption that the cognitive system of a human has limited capacity of resources available [7]. An overview on different dual-task experiments is given in the work by Gosselin and Gagné [5]. The primary task is mostly a speech recognition task. The secondary task is either a memory task or a reaction time task. Taking the cognitive load theory [9] into account and bringing it into relation with complex acoustic scenes, the assertion can be made that the more complex an acoustic scene is, the more cognitive effort is needed to manage both tasks. This means that the performance of the secondary task will decreases when both tasks are performed together.

To reach a sufficient assessment of noise exposition of adults and children, the acoustic scenes for experimental research needs to be as life-like as possible. Using virtual acoustics and binaural reproduction methods, this requirement can be met. Previous research by Feuerstein [4] as well as Rennies and Kidd [11] showed the benefit of binaural listening in dual-task experiments with adults. Furthermore, Fels et al. [2, 3] indicated differences between children and adults' hearing due to the differences in anthropometric data of head, torso and pinna.

The goal of this study is to develop and validate a child-suited dual-task paradigm to analyze the impact of a stepwise change in complexity of an acoustic scenes. Starting with the basic acoustic setting of a realistic classroom, a virtual acoustic environment has been created including a target adult female speaker, representing the teacher, and background noise with the spectrum of children speech. Furthermore, binaural reproduction methods has been taken into account in order to guarantee an aurally-accurate presentation of stimuli that is sufficient not only for adults but also for children. Moreover, individualization processes has been considered. Finally, this study provides direct insight into the differences of listening effort between adults and children.

2. METHODS

The present study obtained ethical approval by the Medical Ethics Committee at the RWTH Aachen University (EK 217/18). Informed consent was obtained from all participating persons and families prior testing. Children gave verbal consent, with the possibility to revoke the consent and cancel at any time during the experiment session.

2.1. Subjects

One group consisting of 36 children, from six and ten years old (median age: 8 years), were recruited from a primary school in Aachen. Inclusion criteria were normal hearing abilities, German speaking, no behavioral syndromes and no color blindness. Every child received a voucher (5 Euro) for participating in the experiment. Another group consisted of 36 normal hearing and German speaking adults with the age range between 19 and 33 years old (median age: 23 years). For participating in the experiment, every adult received a financial compensation (8 Euro). All listener never participated in a similar experiment before and could therefore be considered as non-expert listeners.

2.2. Task and experimental design

For this study, a dual-task paradigm was chosen and adapted for children. The primary task and the secondary task were performed consecutively. The primary task consisted of a closed word recognition task: the participant had to choose the correct picture considering the spoken target word out of four distractor pictures using the colored input device (cf. Figure 1). The secondary task consisted of a reaction time task in which the participant had to push the button in the center of the input device as soon as a stimulus picture appeared after a random time period (250 ms - 1500 ms) following the primary task. Hence, response time and error rate considering the primary task (RT_{T1} and PE_{T1}, respectively) and the reaction time considering the secondary task (RT_{T2}) were the dependent variables.

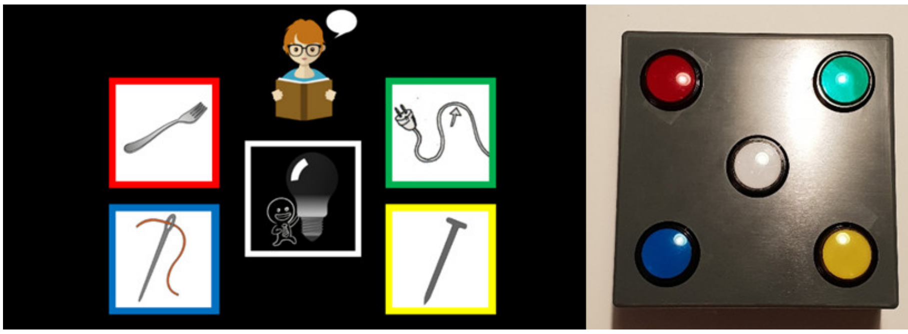

Figure 1: Input device for the experimental task according the presented picture.

The independent variables were noise type, noise position and the signal-to-noise ratio (SNR) in relation to the speech reception threshold (SRT).
The noise type (N_{type}) referred to the tested background noise types. For this study, a speech-shaped noise (SSN), an irrelevant multi-talker babble (IMTB), in another language than the participant's mother tongue, and a relevant multi-talker babble (RMTB), in which the background noise contained understandable distracting information, was chosen.
The noise position (N_{pos}) referred to the target and distractor position combination. The target sound source (S) was always positioned at 0 ° azimuth in the horizontal plane and the two distractor noise sources (N) were either positioned together with the target (S0N0) or spatially separated from the target at ±90 ° azimuth in the horizontal plane (S0N90).
The SNR in relation to the SRT ($N_{\Delta \text{SRT}}$) referred to the test SNR at which the participants were tested. It was either at the determined SRT (p0) or at the SRT + 3 dB (p3).
Since the experiment needed to be feasible for children and the duration of one session was limited to a maximum of 90 minutes. A mixed-subject-design was chosen. Every participant was tested in all combinations of N_{pos} and $N_{\Delta \text{SRT}}$, counterbalanced using a Latin Square. However, each participant was tested in two out of three conditions of N_{type}, which were also counterbalanced over all participants.

2.3. Stimulus material

Word lists adapted from the NORAH study [8] were selected as target stimuli material. Each word list consisted of four-word-combinations with similar phonetics: one target word and three distractor words. The whole set consisted of three training word lists (18 word combinations) and six experimental word lists (28 word combinations). The target words were spoken by a trained, native German-speaking female.
For the relevant multi-talker babble, four native German speaking girls (aged eight to nine years old) were recorded as competing speakers. Each of the girls was recorded individually in an anechoic chamber while reading a short fairy tale. The irrelevant multi-talker babble was spoken by four Swedish girls reading out stories in a competing speech

manner [17]. The speech-shaped noise was filtered white noise with a spectrum that is almost identical to the long-term spectrum of both, the German and the Swedish, multi-talker babble.
All recordings and speech materials were normalized and loudness-adjusted following the EBU-R128 standard. The levels of all background noises were overall calibrated to $L_{\mathrm{eq},30} = 65\mathrm{dB[SPL]}$.

2.4. Experiment environment and sound reproduction

The listening experiment took place in the mobile hearing laboratory "MobiLab" developed by the Teaching and Research Area of Medical Acoustics, RWTH Aachen Universtiy (cf. Figure 2). The MobiLab is a modified trailer including an acoustically optimized hearing booth, which can be easily set up close to the institutions where participants are available (e.g. schools or offices) to allow on-site listening experiments. The hearing booth ($l \times w \times h = 1.86\,\mathrm{m} \times 2.40\,\mathrm{m} \times 1.77\,\mathrm{m}$) ensured a quiet environment during the listening experiment.

Figure 2: Mobile hearing laboratory "MobiLab" on primary school playground.

The acoustic virtual environment for the listening experiment was implemented using the Virtual Acoustics (VA) integration for Matlab [6]. To ensure an aurally-accurate perception of the virtual sound sources, an individualized set of head-related transfer functions (HRTF) was calculated based on the participant's individual head dimensions, i.e. head width, depth and height, by modifying the ITD cues following Bomhard et al. [1] and the HRTF set of the ITA artificial head [12]. A static binaural reproduction via open headphones was chosen for this study using a robust headphone equalization following Masiero and Fels [10]. The virtual sound sources were set up in a free field condition and in a distance of two meters to the participant.

2.5. Experimental procedure

Every participant was tested individually seated in front of a monitor screen with the input device placed in front of them (cf. Figure 3).

The experiment started with a pure tone audiometry to ensure normal hearing (limit: 20 dB[HL]) as well as the individualization and headphone equalization process. Following a recorded, spoken introduction was given in which the task and the feedback was explained step by step accompanied by visual explanations. Training sessions were

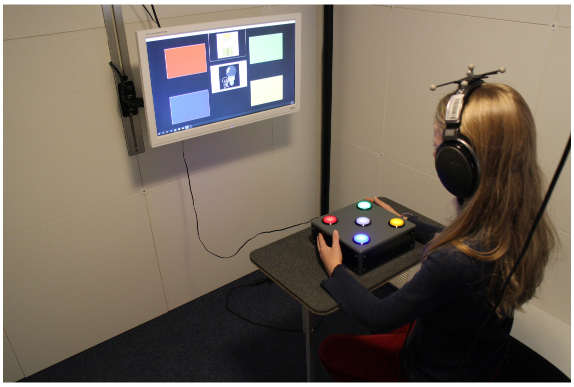

Figure 3: Child participant in experimental trial.

integrated into the introduction. In order to build up the task for children appropriately step by step, primary task and secondary task were trained separately in a first step. In a second step, both tasks were combined.
The whole experimental part was divided in two sections with one N_{type} each. Every section consisted of six blocks each in which the word lists were presented randomly under different conditions. After every block, the participant was forced to have a short break. After the first section, a longer break was recommended to the participant. The first block was noise free so that a baseline measurement could be done. The second block contained an adaptive procedure to determine the SRT. For this purpose, the target level was changed according to the presented SNR while the noise level remained constant starting at +10 dB SNR. The initial step size was set to 2 dB, decreasing the SNR until the first reversal and the step size changed to 1 dB thereafter. The SRT was calculated from those SNRs at which a reversal appeared. The remaining blocks contained the counterbalanced, combined conditions of N_{pos} and $N_{\Delta\text{SRT}}$.

3. RESULTS

For the analysis of reaction times and error rates, the training sequence and the SRT determination sequence were removed from the data. Further, the first trial in every block as well as every trial where either the response time in the primary task or the reaction time in the secondary task exceeded ±3 standard deviations from the individual's mean response/ reaction time was excluded from the analysis (total exclusion of 6.7 % of all trials). Additionally, in case of analysis of response and reaction times, every trial with an error was eliminated (total exclusion of 39.5 % of all trials). Error rates, response and reaction times were averaged for each participant and each condition.
Statistical analysis was done using *IBM SPSS Statistics*. Due to the fact that a mixed design had been chosen for this study, the problem of missing values in noise types for some participants needed to be solved. In this case, a simple imputation had been chosen to substitute all missing values with the mean of all participants who have a non-missing value.
An analysis of variances, ANOVA, was conducted for the main effects age group, noise type N_{type}, noise position N_{pos} and SNR in relation to the SRT $N_{\Delta\text{SRT}}$ for the error rate (PE_{T1}) and the response time (RT_{T1}) of the primary task as well as the reaction time (RT_{T2}) of the secondary task.

3.1. Error rate

The ANOVA revealed a non-significant main effect of age group, $F(1, 70) = .018$, $p > .05, \eta_p^2 = .000$, with almost the same mean error rate for both groups (adults: 33.7 % vs. children: 33.9 %).
The main effect noise type was significant, $F(3, 186) = 300.445, p < .001, \eta_p^2 = .809$ (Greenhouse-Geisser corrected [G-G corr.]). A pairwise t-test showed a significant difference between the no noise condition (1.6 %) and all other noise conditions (SSN: 43.7 % vs. IMTB: 45.0 % vs. RMTB: 45.0 %). No significant difference between SSN, IMTB and RMTB was found.
The ANOVA yielded a significant main effect of noise position, $F(2, 123) = 703.611$, $p < .001, \eta_p^2 = .908$ (G-G corr.), indicating a larger error rate for the S0N0 condition than the S0N90 condition than the no noise condition (means: 1.6 % vs. 32.7 % vs. 56.5 %). A pairwise t-test revealed significance between all noise position combinations.
There was a significant main effect of SNR in relation to the SRT, $F(1, 103) = 695.671$, $p < .001, \eta_p^2 = .907$ (G-G corr.). A pairwise t-test indicated significant differences in mean error rates (no noise vs. p3: 39.1 % vs. p0: 50.0 %). Figure 4 shows the mean error rate in percentage in adults and children.

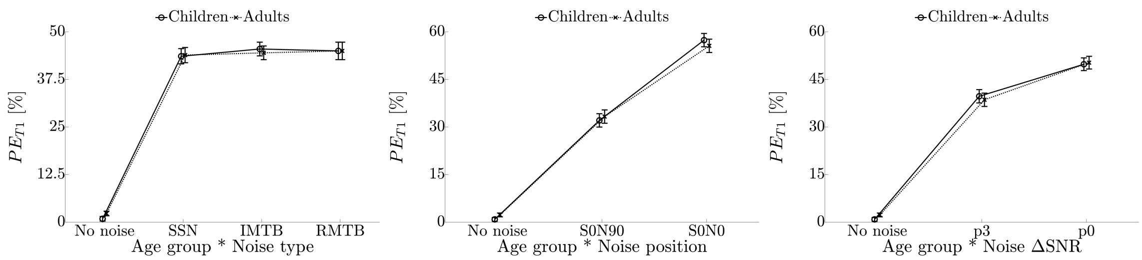

Figure 4: Error rate (in %) as a function of N_{type}, N_{pos}, $N_{\Delta SRT}$ and age group (adults vs. children). Error bars indicate standard errors.

The error rate results met the expectations since the no noise condition always showed significant differences to all other conditions. The comparison of the different noise positions indicating the benefit of spatial hearing when the distractors were spatially separated from the target speaker. Also the p0 condition of $N_{\Delta \mathrm{SRT}}$ revealed an error of 50.0 % which showed that the adaptive procedure found a sufficient SRT.

3.2. Response time in primary task

For response times in the primary task, the ANOVA yielded a significant main effect of age group, $F(1, 70) = 49.876, p < .001, \eta_p^2 = .416$, indicating longer mean response times in children than in adults (2880 ms vs. 2226 ms).
The ANOVA showed a significant effect of noise types, $F(3, 213) = 37.040, p < .001$, $\eta_p^2 = .343$ (G-G corr.). A pairwise t-test revealed significant difference between the no noise condition (mean: 2017 ms) and all other noise conditions. Between these noise conditions, no significant differences were found in the mean response time (SSN: 2813 ms vs. IMTB: 2694 ms vs. RMTB: 2689 ms).
A significant main effect of noise position was yielded, $F(2, 112) = 87.063, p < .001$, $\eta_p^2 = .551$ (G-G corr.). Significant differences in main response time were found between all combination of conditions using a pairwise t-test (no noise vs. S0N90: 2250 ms vs. S0N0: 3012 ms).

The ANOVA revealed a significant main effect of SNR in relation to the SRT, $F(1,95) = 84.441, p < .001, \eta_p^2 = .543$ (G-G corr.). A pairwise t-test showed significant differences in mean response time between all conditions (no noise vs. p3: 2584 ms vs. p0: 2878 ms). The mean response time in the primary task in adults and children is shown in Figure 5.

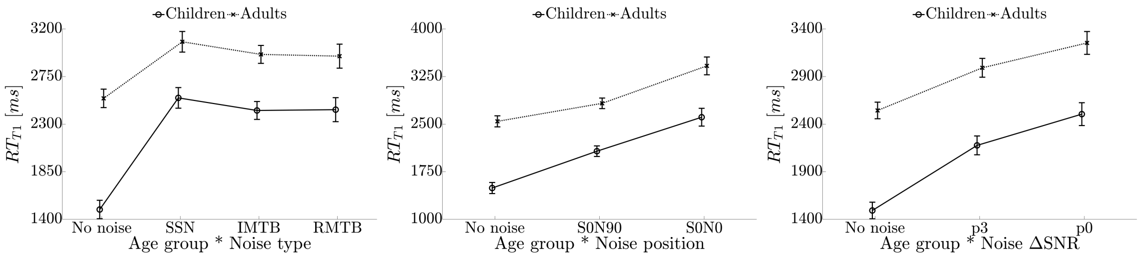

Figure 5: Response time (in ms) as a function of N_{type}, N_{pos}, $N_{\Delta SRT}$ and age group (adults vs. children). Error bars indicate standard errors.

Similar results to the error rates weere found for the response time in the primary task. The significant difference of 654 ms between the two age groups indicates the difference in speech processing time. It seems natural that children's speech processing ability is less developed in comparison to adults. A more detailed analysis point could give more insight.

3.3. Reaction time in secondary task

The ANOVA yielded a main effect of age group in reaction time of the secondary task, $F(1,70) = 25.761, p < .001, \eta_p^2 = .269$, indicating a shorter reaction time in adults than in children (mean: 748 ms vs. 1030 ms). The mean reaction time in the secondary task in adults and children is shown in Figure 6.

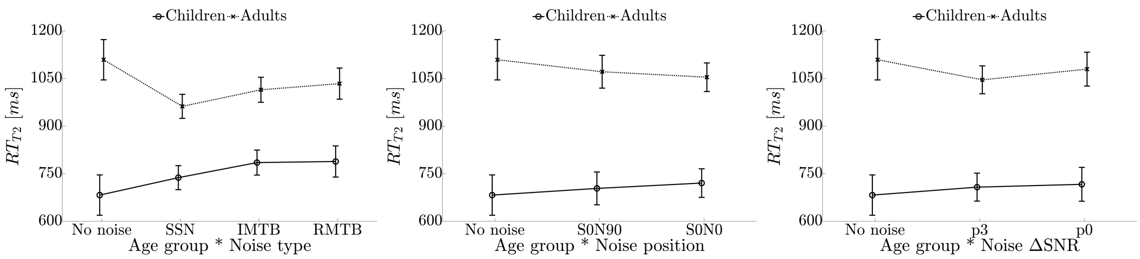

Figure 6: Reaction time (in ms) as a function of N_{type}, N_{pos}, $N_{\Delta SRT}$ and age group (adults vs. children). Error bars indicate standard errors.

No significant main effect was found in noise type, $F(2,159) = 1.270, p > .05, \eta_p^2 = .018$ (G-G corr.). A pairwise t-test only revealed a significant difference between SSN and RMTB (mean: 850 ms vs. 911 ms). However, no differences were found between the other combinations of conditions.
There was no significant main effect of noise position, $F(1,92) = .178, p > .05, \eta_p^2 = .003$ (G-G corr.). Furthermore, no significant main effect of SNR in relation to the SRT was found, $F(1,94) = .889, p > .05, \eta_p^2 = .012$ (G-G corr.).

A similar result to the response time in the primary task was expected, but no effect in the reaction time was found in the secondary task. However, the plots in Figure 6 indicate a reverse behavior of reaction time in children in comparison to adults which can not be explained without further analysis. Notable is the higher reaction times in children during the no noise condition in comparison to the noise condition, which is contrary to the perceptual load theory[9].

4. CONCLUSION

In this work, a dual-task paradigm, including a closed word recognition task as the primary task and a reaction time task as the secondary task, was adapted to test listening effort of children in comparison to adults. Acoustically, realistic classroom scenes were simulated in a virtual acoustic environment and presented binaurally with aurally-accurate methods referring to the anthropometric head measures of children and adults. The listening experiment was conducted in the mobile hearing booth "MobiLab" nearby the cooperating school. Response time and error rates from the primary task as well as reaction times were evaluated considering the capability of the dual-task to test different levels of complexity in acoustic scenes. Results from the primary task indicate sufficient performance of the speech recognition task. Spatial benefit of distractor position could be found and the results further revealed differences in children and adults with longer response times in children. Nevertheless, results from the secondary task are irritating since the mean reaction times in children regarding the different noise conditions showed a reverse pattern to the reaction times in adults. It could either indicate noise coping strategies adopted by the children or an insufficient choice of the secondary task for children. A detailed analysis on factor interaction effects could give more insight.

5. ACKNOWLEDGMENTS

This work received funding support from the HEAD-Genuit-Stiftung under the project ID P-16/17-W with the title "Lärmexposition in Kindertagesstätten, Kindergärten und Grundschulen: Messung mit kindgerechten Verfahren, Analysen und Bewertungen". The authors wish to thank Prof. Maria Klatte and her team for developing the adjusted word lists and providing the stimuli material and Prof. Jonas Brännström for providing the Swedish children multi-talker babble used in this study. Furthermore, the authors would like to acknowledge the involved educational institution, as well as the students Julian Burger and Roser Batlle for assisting paradigm design and data collection.

6. REFERENCES

[1] Bomhardt, R. and Fels, J. (2014). Analytical interaural time difference model for the individualization of arbitrary Head-Related Impulse Responses. In *Audio Engineering Society Convention 137*.

[2] Fels, J., Buthmann, P., and Vorländer, M. (2004). Head-related transfer functions of children. *ACTA ACUSTICA united with ACUSTICA*, 90(5):918–927.

[3] Fels, J. and Vorländer, M. (March/April 2009). Anthropometric Parameters Influencing Head-Related Transfer Functions. *ACTA ACUSTICA united with ACUSTICA*, 95(2):331–342.

[4] Feuerstein, J. F. (1992). Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort. *Ear and hearing*, 13(2):80–86.

[5] Gosselin, P. A. and Gagné, J.-P. (2010). Use of a Dual-Task Paradigm to Measure Listening. *Revue canadienne d'orthophonie et d'audiologie-Vol*, 34(1).

[6] Institute of Technical Acoustics, RWTH Aachen University (2018). Virtual Acoustics - A real-time auralization framework for scientific research. Accessed on 2018-03-16.

[7] Kahneman, D. (1973). *Attention and Effort*, volume 1063. Citeseer.

[8] Klatte, M., Spilski, J., Mayerl, J., Möhler, U., Lachmann, T., and Bergström, K. (2017). Effects of aircraft noise on reading and quality of life in primary school children in Germany: Results from the NORAH study. *Environment and Behavior*, 49(4):390–424.

[9] Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, 21(3):451.

[10] Masiero, B. and Fels, J. (2011). Perceptually robust headphone equalization for binaural reproduction. In *Audio Engineering Society Convention 130*.

[11] Rennies, J. and Kidd Jr, G. (2018). Benefit of binaural listening as revealed by speech intelligibility and listening effort. *The Journal of the Acoustical Society of America*, 144(4):2147–2159.

[12] Schmitz, A. (1995). Ein neues digitales Kunstkopfmeßsystem. *Acta acustica united with acustica*, 81(4):416–420.

[13] Shield, B., Conetta, R., Dockrell, J., Connolly, D., Cox, T., and Mydlarz, C. (2015). A survey of acoustic conditions and noise levels in secondary school classrooms in England. *The Journal of the Acoustical Society of America*, 137(1):177–188.

[14] Shield, B. and Dockrell, J. (2003). The effects of noise on children at school: A review. *Building Acoustics*, 10(2):97–116.

[15] Sjödin, F., Kjellberg, A., Knutsson, A., Landström, U., Lindberg, L., et al. (2012). Noise exposure and auditory effects on preschool personnel. *Noise and Health*, 14(57):72.

[16] Stansfeld, S. and Lercher, P. (2003). Non-auditory physiological effects of noise: Five year review and future directions. In *ICBEN 2003. Proceedings of the 8th International Congress on Noise as a Public Health Problem, Rotterdam. Foundation ICBEN*, pages 84–90.

[17] von Lochow, H., Lyberg-Åhlander, V., Sahlén, B., Kastberg, T., and Brännström, K. J. (2018). The effect of voice quality and competing speakers in a passage comprehension task: Perceived effort in relation to cognitive functioning and performance in children with normal hearing. *Logopedics Phoniatrics Vocology*, 43(1):32–41.