Assessing and reducing listening effort of listening to speech in adverse conditions

Amy Hall[1], Jan Rennies-Hochmuth[1], Axel Winneke[1]

[1] *Hearing Speech and Audio Technology, Fraunhofer IDMT, 26129, Oldenburg, Germany Email: amy.jane.hall@idmt.fraunhofer.de*

Abstract

Background Normally hearing listeners successfully compensate for speech distortions in everyday environments, but can become fatigued as a result. AdaptDRC is a near-end-listening-enhancement algorithm that alters speech signals for playback, dependent on environmental noise, for improved intelligibility [17].

Aim In this electroencephalography (EEG) study I measured neurophysiological correlates of listening effort (LE) by comparing unprocessed speech to AdaptDRC-processed speech.

Method I recorded EEG while normally hearing participants (N=27) listened to unprocessed or AdaptDRC-processed sentences in noise, then rated the subjective listening effort. I also measured speech intelligibility, hearing and cognitive abilities.

Results For intelligible speech, subjective LE decreases with increasing SNR and is lower for AdaptDRC speech than unprocessed speech. Spectral alpha power (8-12Hz) analyses suggest that peak cognitive effort occurs at 0dB SNR. Spectral alpha also increases with task duration, indicating an association with fatigue.

Conclusions This experiment provides insight into the neurophysiological correlates of effortful listening in adverse conditions, and the benefits of near-end-listening-enhancement technology.

Introduction

Humans are excellent at understanding speech, even in challenging listening conditions. However, when a speech signal has been compromised by environmental noise, additional cognitive processes are required to decipher the original message. This compensatory processing can be understood as cognitive effort: The deliberate allocation of resources to overcome obstacles in goal pursuit when carrying out a task [13].

The concept of listening effort has become increasingly relevant in the hearing and audiology literature. Reports show that hearing aid users experience higher levels of fatigue in the workplace and switch their devices off in busy environments to prevent tiredness, and children with hearing impairment miss school due to fatigue [6] & [4]. This occurs despite excellent hearing performance in clinical settings, demonstrating that there are factors beyond speech comprehension that should be considered in audiological settings.

Listening effort can be a difficult construct to reliably measure. Researchers can use behavioural, physical and neurophysiological measures to understand how changes in effort reflect differences in stimuli and task requirements. Spectral alpha power, as measured by electroencephalography (EEG), may modulate the suppression of task irrelevant information when listening to speech in noise [19], and spectral alpha peak may increase with working memory load [8]. In a study by Winneke et al. (2016) [20], individualised hearing support significantly decreased both the subjective ratings of listening effort and spectral alpha power between 8-10Hz.

AdaptDRC is a near-end-listening-enhancement algorithm which alters speech signals dependent on environmental noise. Full details regarding this algorithm can be found in Schepker et al. (2015) [18] but a layman's summary follows. There are two main stages to the algorithm, processing the speech in 20ms frames. 1) Frequency shaping: the Speech Intelligibility Index for the speech and noise is calculated at each of eight octave sub-bands. This results in a weighting between 0-1, depending on which a frequency spectrum is applied. 2) Dynamic range compression: SNRs are calculated at each sub-band giving a ratio between 1:1 and 1:8. A compression ratio is then applied depending on this ratio. Finally, the speech signal is normalised such that the root-mean-square output power is the same as it was at input. For everyday noise conditions, higher frequencies will typically be amplified and soft parts are boosted relative to stronger parts [14].

The AdaptDRC algorithm increases intelligibility of speech presented in competing talker[1], speech-shaped noise, and cafeteria noise [14] and reduces self-reported effort even when intelligibility is at 100%. However, self-report measures are limited in their validity due to the subjective nature of the questions.

There is evidence that individual differences in cognitive abilities may influence the experience of listening effort, and that may be reflected in the neural processing of speech [12]. For example, the ease of language understanding model (ELU) [15] describes how listeners must temporarily store information to understand speech in adverse conditions.

In this experiment, I sought to identify reliable neural traces of listening effort in spectral alpha power, by comparing AdaptDRC processed speech with unprocessed speech at different SNRs and in two noise types. I also explored how individual differences in cognition are related to effortful listening.

Methods

27 participants were recruited from the University of Oldenburg. All participants were native German speakers with normal hearing (pure tone average <10dB HL) aged between 18-30 ($M = 23.4, SD = 2.6$). Participants were paid €10 per hour for their time and the experimental session, including all tests and EEG set-up lasted for a total of 2.5 hours. This research was approved by the Carol von Ossietzky Universität Oldenburg Kommission für Forschungsfolgenabschätzung und Ethik (Drs.

81/2017).
Materials: We used Oldenburg Satztest (OLSA) sentences; a set of phonetically balanced sentences with the structure name, verb, number, adjective, noun e.g. Peter sieht acht nasse Steine *Peter sees eight wet stones.* There are 10 possible words at each position, randomised such that they are semantically unpredictable. Sentences were recorded by a male speaker.
Listening effort was rated by participants on a modified version of the Adaptive Listening Effort Scale (ACALES)[11]. This is a 13-point scale from *mühelos* (effortless) to *extrem anstrengend* (extremely effortful) with the additional option of *nur Störgeräusch* (just noise).
Speech was either unprocessed or processed using the AdaptDRC algorithm, details of which can be found in Schepker et al., 2015. Speech level was fixed at 60dB SPL, and noise levels were adjusted to produce SNRs of -10, -5, 0, 5, 10. Two noise types were used: OLnoise and cafeteria noise. OLnoise is stationary speech-shaped noise (SSN) developed from the unprocessed speech material with the same frequency spectrum. Cafeteria noise is a real recording of a busy cafeteria including the sounds of unintelligible voices, furniture moving and cutlery [9]. The study had a 5 x 2 x 2 design with the factors SNR x noise type x processing type.
All auditory processing and mixing was performed in MATLAB and signals were presented over a single loudspeaker 1.5m from the listener at 0°.
Pure tone audiometry: Pure tone hearing thresholds were established for .25, .5, 1, 2, 4, & 8 kHz to ensure all participants had normal hearing levels.
Cognitive tasks: Participants completed three cognitive tasks, reflecting the three core executive functions: updating, inhibiting, and shifting. Working memory capacity (updating) was measured using a backwards digit span task [5], in which participants were presented with lists of spoken digits and asked to repeat them in reverse order. Participants were acoustically presented with sets of digits played at 1s intervals, initially consisting of 3 digits and increasing until participants failed to correctly recall 4/6 within a list. The Flanker task [3] is a well established measure of inhibition. In this incarnation, participants were asked to indicate as quickly and accurately as possible the direction of the middle arrow for congruent (<<<<<), incongruent (>><>>), and neutral (−− < −−) trials. We measured shifting ability with the 2 7 Ruff selective attention task [16]. In this task, participants have limited time to identify all 2s and 7s in arrays of digits or capitalised letters.
Listening effort task: Participants were instructed to maintain their gaze forward and to limit head and eye movements during the trials. The listening effort task consisted of 300 trials divided into 10 blocks of 30 with breaks in between. For each trial, participants were played two sentences and in 25/30 they were then prompted to rate the effort using the ACALES scale. For the other 5 trials, randomly distributed throughout the block, participants were prompted to repeat the last sentence that they had heard. These trials were included to ensure that participants were actively attempting to understand the speech and to gather speech intelligibility data. This portion of the experiment lasted between 50-60 minutes depending on how long participants took for breaks between blocks.
Data recording and analysis: A 24-channel EEG cap was used with a motor layout (mBrainTrain; EasyCap), meaning the highest density of electrodes is over the parietal lobe. The data were recorded using Smarting Streamer at a sampling rate of 500Hz and synchronised with the event markers using lab streaming layer [10]. Data processing was performed using EEGLAB [2]. All channels were re-referenced to an average of the mastoid electrodes (TP9 and TP10) and bandpass filtered between 1-45Hz. Independent components analysis was performed to remove eye-blinks and eye-movements, then artefact rejection removed trials containing extreme activity ($\geq \pm 150\mu V^2$). The data was epoched into 5000ms windows from the onset of the first sentence and baseline corrected (-200-0ms), and a nonequispaced fast Fourier transform was performed on this data to find the power spectrum between 3-25Hz. The average spectral alpha power (8-12Hz) was then calculated for each condition per participant. This procedure was performed for all electrodes, but the data presented below refer to CPz, taken as representative of activity recorded at the parietal lobe.

Results

Behavioural results

Figures 1 and 2 show the mean listening effort ratings and percentage words correct for AdaptDRC processed and unprocessed speech in speech-shaped noise and cafeteria noise. In both noise types, there are significant SNR by processing type interactions on subjective listening effort ratings and speech intelligibility. Subjective effort decreases and intelligibility increases with increasing SNR and AdaptDRC augments these effects. ($r = -.43, p \leq .001$)

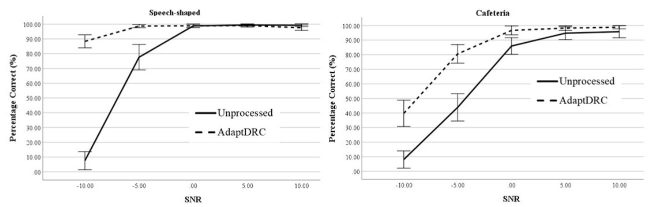

Figure 1: Mean words correct in percent. Error bars show 95% confidence intervals.

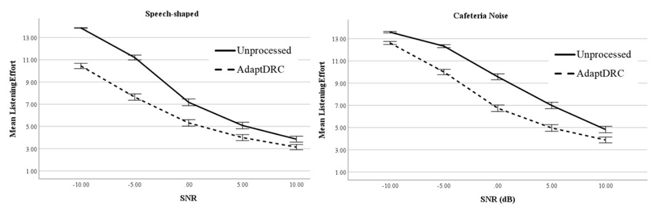

Figure 2: Mean listening effort ratings in Effort Scaling Categorical Units. Error bars show 95% confidence intervals

EEG results
Six participants were excluded from EEG data analysis ($N = 1$ due to a technical error & $N = 5$ due to extreme alpha values $\geq$ 2SD from mean). Noise types were treated separately in the following statistical analyses.
SSN: A 2x5 repeated measures ANOVA with the factors processing type and SNR showed no main effects on spectral alpha power at electrode CPz.
Cafeteria noise: A 2x5 repeated measures ANOVA with the factors processing type and SNR showed a main effect of SNR on spectral alpha power at electrode CPz ($F(9, 9) = 9, p = .047$).

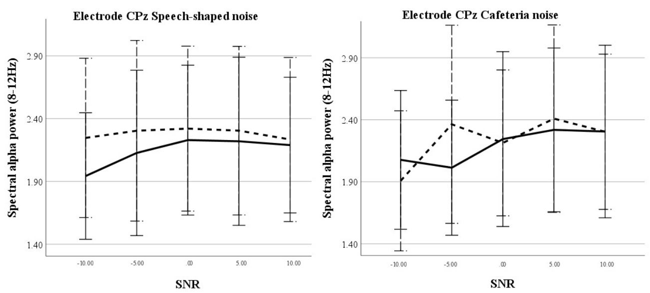

Figure 3: Mean spectral alpha power (8-12Hz). Error bars show 95% confidence intervals.

Effects of sentence: To assess whether having to remember the second sentence of each pair affected neural processing, the analyses described above were then performed on separate 2500ms epochs for the first and second sentence, respectively. For cafeteria noise, a 2x2 repeated measures ANOVA with the factors sentence and SNR showed a trend for increased spectral during the second sentence, but this did not reach statistical significance ($F(1, 43) = 3.77, p = .059$).
Effects of fatigue: To assess whether prolonged listening effort had an effect on spectral alpha power, the data were divided into two sets: the first five blocks and the second five blocks. A 2x2 repeated measures ANOVA with the factors experiment half and processing type revealed a main effect of half ($F(1, 21) = 4.47, p = .047$)

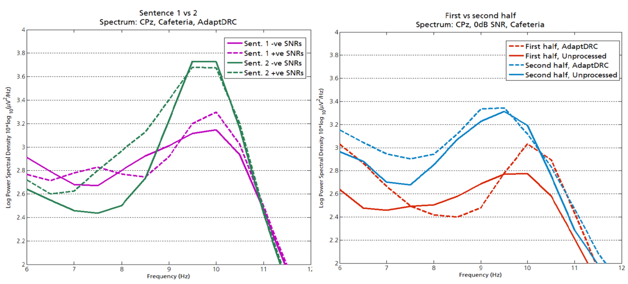

Figure 4: Left: power spectra at electrode CPz for sentences 1 and 2, processed by AdaptDRC and presented in cafeteria noise. Right: power spectra at electrode CPz for sentences in the first and second half of the experiment

Cognitive results: Working memory scores were established by taking the number of lists a participant could correctly recall. Ruff 2 7 score was calculated as the percentage of items correctly identified. Flanker effect was calculated with the equation [7]:

$$Flanker\ \ score = \frac{incongruent\ \ log(rt) - neutral\ \ log(rt)}{neutral\ \ log(rt)} \tag{1}$$

No co-correlations between working memory, inhibition, selective attention were identified, indicating that the tasks measured separate constructs. Correlational analyses revealed that, in conditions where intelligibility was below 80%, working memory capacity correlated with individual word recognition accuracy ($r = .54, p = .021$) and Flanker score correlated with individual subjective listening effort ratings ($r = .54, p = .015$).

Discussion

The speech intelligibility and self-reported listening effort data replicate previous findings for AdaptDRC processed speech [1] [14], showing that it reliably improves intelligibility and reduces subjective effort for speech presented in both stationary and non-stationary noise. This is particularly true for SSN at -10dB SNR, where AdaptDRC improved word recognition rates from 7.5% to 88.4%.
There were no significant effects of SNR or speech processing on alpha power for sentences presented in SSN, possibly due to the close to ceiling performance. For sentences presented in cafeteria noise, there was a main effect of SNR, with alpha power peaking at 0dB SNR. In the context of previous research which shows that spectral alpha power reflects effortful listening [19],[20], this implies that there is a "give-up" effect at very challenging SNRs, with peak cognitive effort occurring when good word recognition performance is possible but demanding.
In the EEG data, no benefit of AdaptDRC processing was found. However, as can be seen in Figure 3, there were very large individual differences in spectral alpha power. For this reason, further analyses are required to account for this variance, by calculating the percentage change in alpha for each condition relative to each individual's baseline alpha levels.
If an effect of sentence is found to be significant with this subsequent analysis, it would imply that actively listening to remember increases cognitive load relative to passive listening. However, further research is required to establish whether this is an effect of task or simply sentence order.
Alpha power was found to be higher in the second half of the experiment, implying that there may be a cumulative effect of listening to speech in adverse conditions that is reflected in the cognitive processing. Further research is required to establish a clear relationship between listening effort and fatigue, and if speech enhancement algorithms such as AdaptDRC could relieve this effect over longer time periods.
The results of the cognitive test battery imply that having a larger working memory capacity improves word recognition accuracy. In respect to the ELU model, this supports the idea that being able to temporarily store more speech information supports speech understanding. Individuals who are better at inhibiting visual information also experienced less effort in challenging conditions.

This implicates a domain general cognitive strategy associated with the suppression of task irrelevant information. Future studies will explore whether this is reflected in spectral alpha power and how pressures on these cognitive resources could potentially be relieved using speech enhancement.

Acknowledgments

ENRICH has received funding from the EU's H2020 research and innovation programme under the MSCA GA 675324

References

[1] Cooke, M., Mayo, C., & Valentini-Botinhao, C. (2013, August). Intelligibility-enhancing speech modifications: the hurricane challenge. In Interspeech (pp. 3552-3556).

[2] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods, 134(1), 9-21.

[3] Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. Visual Cognition, 2(2-3), 101-118.

[4] Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. Journal of Speech, Language, and Hearing Research.

[5] Hilbert, S., Nakagawa, T. T., Puci, P., Zech, A., & Bühner, M. (2014). The digit span backwards task. European Journal of Psychological Assessment.

[6] Hua, H., Anderzén-Carlsson, A., Widén, S., Möller, C., & Lyxell, B. (2015). Conceptions of working life among employees with mild-moderate aided hearing impairment: A phenomenographic study. International journal of audiology, 54(11), 873-880.

[7] Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. The Quarterly Journal of Experimental Psychology, 65(8), 1563-1585.

[8] Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cerebral cortex, 12(8), 877-882.

[9] Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. EURASIP Journal on Advances in Signal Processing, 2009, 6.

[10] Kothe, C. (2014). Lab streaming layer (LSL). https://github. com/sccn/labstreaminglayer. Accessed on October, 26, 2015.

[11] Krueger, M., Schulte, M., Brand, T., & Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. The Journal of the Acoustical Society of America, 141(6), 4680-4693.

[12] Obleser, Wöstmann, Hellbernd, Wilsch, & Maess (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. The Journal of Neuroscience, 32(36), 12376-12383.

[13] Pichora-Fuller, Kramer, Eckert, Edwards, Hornsby, Humes, ... & Naylor (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening. Ear and Hearing, 37, 5S-27S.

[14] Rennies, Pusch, Schepker, & Doclo (2018). Evaluation of a near-end listening enhancement algorithm by combined speech intelligibility and listening effort measurements. The Journal of the Acoustical Society of America, 144(4), EL315-EL321.

[15] Rönnberg, Rudner, Lunner, & Zekveld (2010). When cognition kicks in: Working memory and speech understanding in noise. Noise and Health, 12(49), 263.

[16] Ruff, R. (2011). Ruff 2&7 Selective Attention Test. Encyclopedia of Clinical Neuropsychology, 2202-2203.

[17] Schepker, Rennies, & Doclo (2013). Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression. In INTERSPEECH (pp. 3577-3581).

[18] Schepker, H., Rennies, J., & Doclo, S. (2015). Speech-in-noise enhancement using amplification and dynamic range compression controlled by the speech intelligibility index. The Journal of the Acoustical Society of America, 138(5), 2692-2706.

[19] Strauß, Wöstmann, & Obleser, (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. Frontiers in human neuroscience, 8, 350.

[20] Winneke, Meis, Wellmann, Bruns, Rahner, Rennies, Wallhoff, & Goetze (2016). Neuroergonomic assessment of listening effort in older call center employees. Proceedings 9. AAL Kongress in Frankfurt/Main vom 20.-21. April 2016.