University of Groningen

# Listening Effort

Pals, Carina

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*
Publisher's PDF, also known as Version of record

*Publication date:*
2016

Link to publication in University of Groningen/UMCG research database

*Citation for published version (APA):*
Pals, C. (2016). *Listening Effort: The hidden costs and benefits of cochlear implants*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

**Copyright**
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

**Take-down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

Download date: 12-09-2024

# CHAPTER 3

## Effects of simulated electric acoustic hearing on listening effort and perception of speech in quiet and in noise

Carina Pals, Anastasios Sarampalis, Mart van Dijk, Deniz Başkent

## ABSTRACT

*Purpose.* Although many cochlear implant (CI) users achieve good speech understanding in quiet listening conditions, CI-mediated hearing is degraded compared to normal hearing. Interpreting a degraded speech signal requires increased cognitive processing, i.e. listening effort, to compensate for the signal degradations and fill in missing information. Previous research shows that CI users with residual acoustic hearing may benefit from electric-acoustic stimulation (EAS) in increased intelligibility and improved tolerance to noise. We hypothesize that the availability of low frequency acoustic speech cues may also reduce listening effort. This study systematically investigated this hypothesis in normal-hearing listeners using acoustic simulations of CI hearing and EAS.

*Methods.* We examined the potential listening effort benefits of simulated EAS for speech understanding at three different, fixed intelligibility levels. Experiment 1 was conducted in quiet at near ceiling intelligibility. Experiment 2 and 3 were conducted in steady state, speech shaped noise at 50% and 79% sentence intelligibility, respectively. Listening effort was measured both subjectively, using a rating scale, and objectively, using a dual-task paradigm. In the dual-task, listening effort for the primary sentence intelligibility task is reflected in performance on the secondary visual response-time (RT) task.

*Results.* In quiet, with intelligibility fixed near ceiling for all conditions, simulated EAS significantly reduced the RTs on the secondary task compared to one of the two simulated CI conditions. In noise, the simulated EAS conditions produced 50% intelligibility at on average 2.7 dB lower SNR than the simulated CI conditions, and also resulted in significantly lower RTs on the secondary task. Simulated EAS produced 79% intelligibility at on average 5.4 dB lower SNR than simulated CI, with no change in RTs.

*Conclusion.* The quiet condition with near ceiling intelligibility showed the improvement in RTs expected based on the hypothesis. For speech in noise, simulated EAS allowed the desired intelligibility levels to be reached at less favorable SNRs, as can be expected from literature. Interestingly, this came without the cost of increased listening effort; at 50% intelligibility even a reduction in listening effort on top of the benefit in SNR was observed. These results suggest that in addition to the benefits in speech intelligibility and the increased tolerance to noise, EAS can also provide a benefit in reducing listening effort compared to CI listening alone.

# INTRODUCTION

Even in the most favorable listening conditions, cochlear implant (CI) mediated hearing is degraded compared to normal hearing (NH), due to factors related to the device, the electrode nerve interface, and the health of the auditory system (Başkent, Gaudrain, Tamati, & Wagner, 2016). Interpreting a degraded speech signal requires increased top-down cognitive processing (Classon, Rudner, & Rönnberg, 2013; Gatehouse, 1990; Pichora-Fuller, Schneider, & Daneman, 1995; Wingfield, 1996). The ease-of-language-understanding model (ELU) proposes a mechanism for this recruitment of cognitive resources to interpret degraded speech; when a signal is degraded, the missing or incomplete segments of the input stream cannot be automatically matched to existing phonological and lexical representations in long term memory, triggering a loop of explicit cognitive processing to fill in the missing information or to infer meaning (Rönnberg, 2003; Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008). This explicit processing that occurs when the incoming speech signal is degraded increases the cognitive load of speech understanding, which is referred to as listening effort. It stands to reason, then, that interpreting the degraded speech heard through a CI may thus be effortful for the listener, and processing strategies or device configurations that improve CI signal quality may reduce listening effort for CI users.

In support of this idea, NH listeners experience increased listening effort when presented with CI simulated speech compared to clear speech (Wagner, Toffanin, & Başkent, 2016; Wild et al., 2012) and listening effort has been shown to decrease for simulated CI speech of increased spectral resolution (Pals, Sarampalis, & Başkent, 2013 (see also Chapter 2); Winn, Edwards, & Litovsky, 2015). The device configuration known as electric-acoustic stimulation (EAS), i.e. the combination of a CI with acoustic hearing in either the implanted or the contralateral ear (amplified if necessary) may similarly provide such an improvement in signal quality that can lead to a reduction in listening effort.

Research on the effects of EAS consistently shows benefits in speech intelligibility in quiet and increased tolerance to masking noise. Although the frequency range of residual hearing in CI users is often very limited and the acoustic input alone, without the CI, does not provide much intelligibility (Dorman & Gifford, 2010), the low-frequency sound does carry additional (likely complementary) acoustic speech cues that are not transmitted well through CIs, such as voice pitch, consonant voicing, or lexical boundaries (Brown & Bacon, 2009). Even as little as 300 Hz low pass filtered (LPF) speech already provides a significant improvement in signal-to-noise ratio (SNR) for CI users with a usable level of residual hearing (Büchner et al., 2009). In addition to this benefit in speech intelligibility or noise tolerance, EAS also improves subjective hearing device benefit (Gstoettner et al., 2008), and speech perceived with EAS is generally reported to sound more natural and pleasant than a CI alone (Kiefer et al., 2005; Turner, Gantz, Lowder, & Gfeller, 2005; von Ilberg et al., 2000).

We hypothesize that the additional speech cues provided by the low frequency acoustic sound of the EAS signal, can reduce the need for explicit cognitive processing to aid the interpretation of the degraded CI speech signal, thus reducing cognitive load and freeing up cognitive resources for concurrent tasks. In the current study, this hypothesis is tested using a dual-task paradigm that combines a speech intelligibility task with a secondary visual response-time (RT) task. If EAS reduces listening effort and, therefore, results in more cognitive resources being available for the secondary task, this should be reflected as improvements in the response times on the RT task. Previous research using a similar dual-task paradigm has shown that changes in signal quality, such as increased spectral resolution (Pals et al., 2013), or noise reduction (Sarampalis, Kalluri, Edwards, & Hafter, 2009), can result in improved listening effort even if no change in intelligibility is observed.

The current study investigates whether EAS provides a benefit in listening effort in addition to the already documented benefits of EAS in intelligibility or noise tolerance, and therefore focused on conditions that lead to equal levels of intelligibility. In a series of three experiments, we examine the effects of EAS on listening effort for intelligibility fixed at three different levels: Experiment 1 for speech in quiet at near-perfect intelligibility, Experiment 2 for noise-masked speech at 50% intelligibility, and Experiment 3 for noise-masked speech at 79% sentence intelligibility. In the latter two experiments, the two different intelligibility levels were chosen to investigate effects on listening effort at different parts of the psychometric function. At the 50% sentence intelligibility level, the slope of the psychometric function is at its steepest, and small changes in signal quality will result in larger changes in intelligibility than at the 79% point in the psychometric function. Similarly, small changes in signal quality at 50% versus 79% intelligibility may also affect listening effort differently.

In order to systematically explore how low frequency acoustic input in the EAS signal affects listening effort, we manipulate the presence and amount of low pass filtered (LPF) speech in addition to noise-vocoded CI simulations (referred to as simulated EAS) on listening effort in NH listeners both in quiet listening conditions and in noise. The use of simulated CI stimuli and NH listeners allows for studying effects of a subset of device-related factors in a systematic manner, while controlling for much of the demographic and etiological factors that contribute to the large variability observed in speech intelligibility in CI users (Blamey et al., 2013; Lazard et al., 2012).

## EXPERIMENT 1: SPEECH IN QUIET AT NEAR CEILING INTELLIGIBILITY

*Motivation*

In Experiment 1, we examine how the addition of LPF acoustic information affects listening effort when there is no background noise and intelligibility is near ceiling, by simulating the most common configuration in CI users with residual hearing, namely, the use of a CI in combination with residual hearing in the contralateral ear, commonly referred to as "bimodal" listening (Dooley et al., 1993; Mok, Grayden, Dowell, & Lawrence, 2006; Seeber, Baumann, & Fastl, 2004). The different processing conditions of the stimuli were chosen based on Pals et al. (2013), such that we expect near-ceiling intelligibility; 6- and 8-channel noise-vocoded CI simulations are combined with either 300 or 600 Hz LPF speech in the contralateral ear (Qin & Oxenham, 2006). With intelligibility near ceiling, we expect little to no further improvement in intelligibility, however, we hypothesize that EAS may still serve to reduce listening effort.

*Methods*

*Participants.* Twenty NH, native Dutch speaking, young adults (age range: 18–21 years, mean 19 years; 5 female, 15 male) participated in this experiment. Participants were recruited via posters at university facilities and were screened for normal hearing thresholds of 20 dB HL or better at audiometric frequencies between 250 and 6000 Hz, measured in both ears. Dyslexia or other language or learning disabilities were exclusion criteria in this and subsequent experiments.

We provided written information about the experiment to all participants, explained the procedure in person during the lab visit, and gave the opportunity to ask questions before signing the informed consent form. Participants received a financial reimbursement of €8 per hour, plus traveling expenses, for their time and effort. The local ethics committee approved this and the subsequent experiments.

*Speech stimuli.* The sentences used for the primary intelligibility task were taken from the Vrije Universiteit (VU) corpus (Versfeld, Daalder, Festen, & Houtgast, 2000), which consists of conversational, meaningful, and unambiguous Dutch sentences, rich in semantic context, and each with eight to nine syllables (average duration is 1.8s). The corpus is organized into 78 unique lists of 13 sentences, half recorded with a female speaker, and half with a male speaker. The lists are balanced such that the phoneme distribution of each list approximates the mean phoneme distribution of the full corpus, and each sentence is of approximately equal intelligibility in noise (Versfeld et al., 2000). In this experiment we used the 39 lists spoken by the female speaker, the last 6 of these lists were used for training and a random selection of the remaining lists was used in each experiment, such that each sentence was presented no more than once to each participant.

In this experiment, four device configurations were simulated and compared; a single CI in one ear alone, a CI on both sides, a CI combined with limited residual hearing in the contralateral ear (300Hz LPF), and a CI combined with significant residual hearing in the contralateral ear (600Hz LPF). The binaural CI condition was included to distinguish between effects due to binaural versus monaural hearing and effects due to the presence of the low frequency acoustic signal in the ear contralateral to the CI signal. See Table 1 for an overview of all the experimental conditions. The CI simulations were generated using a noise-band vocoder implemented in MATLAB. Simulations of 6 or 8 spectral channels were used, as it was for these conditions that we observed changes in listening effort independent of changes in intelligibility in a previous study (Pals et al., 2013). The original signal was filtered into 6 or 8 spectral bands (analysis bands) between 80 and 6000 Hz using 6$^{th}$-order Butterworth band-pass filters with cutoff frequencies that simulate equal cochlear distance. The envelopes of analysis bands, extracted with half-wave rectification and 3$^{rd}$-order low-pass Butterworth filter with -3 dB cut-off frequency of 160 Hz, modulated carrier bands (synthesis bands), generated with white noise filtered by the same analysis band-pass filters. The modulated noise-bands were post-filtered using the original synthesis band-pass filters and added together to form the final CI simulation signal.

The low-frequency residual hearing was simulated by low-pass filtering at 300 and 600 Hz, values similar to earlier EAS simulation studies (Başkent, 2012; Qin & Oxenham, 2006; Zhang, Spahr, & Dorman, 2010), using 3rd-order Butterworth low-pass filters.

*Table 1. Summary of the experimental conditions for Experiment 1*

| *Label* | *Factor: Listening mode* | *Factor: Spectral resolution* | *Left ear (dBA)* | *Right ear (dBA)* |
|---|---|---|---|---|
| MonCI6 | Monaural CI | 6 channels | | 6 Ch. CI (65) |
| MonCI8 | | 8 channels | | 8 Ch. CI (65) |
| BinCI6 | Binaural CI | 6 channels | 6 Ch. CI (60) | 6 Ch. CI (60) |
| BinCI8 | | 8 channels | 8 Ch. CI (60) | 8 Ch. CI (60) |
| EAS6/300 | EAS LPF300 | 6 channels | 300 Hz LPF (60) | 6 Ch. CI (60) |
| EAS8/300 | | 8 channels | 300 Hz LPF (60) | 8 Ch. CI (60) |
| EAS6/600 | EAS LPF600 | 6 channels | 600 Hz LPF (60) | 6 Ch. CI (60) |
| EAS8/600 | | 8 channels | 600 Hz LPF (60) | 8 Ch. CI (60) |

Note: Conditions are divided into factors 'listening mode' and 'spectral resolution', showing the stimuli presented to the left and right ear, including the presentation levels (in dBA).

The right ear was always presented with the simulated CI signal. In the binaural CI condition, the same simulated CI signal was presented to the left ear as well. In the bimodal conditions the LPF sound was presented to the left ear. In the monaural CI conditions, the stimulus was presented at 65 dBA. In conditions where stimuli were presented to both ears (binaural CI simulation or bimodal EAS simulation), each stimulus was presented at 60 dBA to compensate for binaural loudness summation and to prevent any potential confounds from loudness (Epstein & Florentine, 2012). The presentation level of the stimuli was calibrated using the speech-shaped noise provided with the VU corpus, which matches the long-term speech spectrum of the sentences spoken by the female speaker (Versfeld et al., 2000).

*Visual stimuli.* The secondary task in the dual-task paradigm was a visual rhyme-judgment task. The stimuli used for this task were the same monosyllabic meaningful Dutch words used by Pals et al. (2013). For each of the five Dutch vowels (a, e, i, u, o) Pals et al. (2013) created lists of monosyllabic rhyme words with several word endings (e.g. [stok, vlok, wrok] or [golf, kolf, wolf]). They excluded words that could be pronounced in more than one way, as well as the 25% least frequently occurring words, according to the CELEX lexical database of Dutch

(Baayen, Piepenbrock, & Gulikers, 1995). Due to the nature of the Dutch language it was not possible to control for orthographic similarity. For each trial two words were simultaneously displayed one above another, centered on a computer monitor in large, black capital letters on a white background, each letter approximately 7 mm wide and 9 mm high, with 12 mm vertical whitespace between the words.

*Equipment.* Participants were seated in a soundproof booth, approximately 50 cm from a wall-mounted computer screen. The presentation of the speech stimuli for the primary task and the visual stimuli for the secondary task was coordinated by a MATLAB program, using the Psychophysics Toolbox Version 3, and run on an Apple Mac Pro computer. The verbal responses on the primary listening task were recorded using a PalmTrack 24-bit digital audio recorder of Alesis, L.P. (Rhode Island, USA). The digital audio stimuli were routed via the AudioFire 4 external soundcard of Echo Digital Audio Corporation (California, USA) to the Lavry digital-to-analog converter and on to the open-back HD600 headphones of Sennheiser Electronic GmbH & Co. KG (Germany).

*Procedure.* Before each new task, the experimenter explained the procedure in detail to ensure that the participant understood the task. The participants were first given three minutes to practice the rhyme-judgment task alone, during which the experimenter monitored their performance to see whether they understood the task and provided additional instructions if this proved necessary. Following that, in a 20-minute intelligibility training session (based on Benard and Başkent, 2013), participants familiarized themselves with the different processing conditions of the speech stimuli. During training, processed sentences in six of the eight processing conditions (the two monaural CI and the four EAS conditions) were presented in random order. One list of 13 sentences was used per condition, and the participant's task was to repeat the sentences as best they could. After each response, feedback was given by presenting the sentence visually and auditorily, once unprocessed and one processed. Sentence lists used during training were not used again in the rest of the experiment.

During data collection, each listening condition was presented once as a single task (intelligibility) using one set of sentences, and once as a dual-task (intelligibility and visual rhyme) using two sets of sentences (to allow for a sufficient number of visual trials to be presented during an auditory stimulus). The presentation order of the conditions was randomized using the MATLAB random permutation function seeded to the system clock.

The primary intelligibility task was to listen to processed sentences presented in quiet and repeat each sentence as accurately as possible. The sentence onsets were eight seconds apart. As the average duration of sentences was about 1.8 seconds, this timing left about 6.2 seconds available for the verbal response. The verbal responses were recorded for later scoring by a native Dutch speaker. Speech intelligibility was scored based on the percentage of full sentences repeated entirely correct.

The secondary rhyme-judgment task was to indicate as quickly as possible whether the word pair presented on the monitor rhymed or not. The accuracy of responses and the RTs were recorded by the experimental software. The RT was defined as the interval from visual stimulus onset to the key-press by the participant. The participant was instructed to look at a fixation cross in the middle of the screen. At the onset of each trial a randomly chosen pair of words would appear on the screen, one above the other. The chance of a rhyming word-pair being selected was set to 50%. The words would stay on the screen until either the participant had pressed the response key or the time-out duration of 2.7 seconds was reached, the latter of which would be logged as a 'miss'. After completion of a trial, the fixation cross would reappear for a random duration between 0.5 and 2.0 seconds before the next word pair would appear. The timing of the presentation of the visual rhyme words was not coupled to the timing of the auditory stimulus, therefore a secondary task trial could start at any time during or between auditory stimuli for the primary task.

After completing each test with one of the processing conditions, either single- or dual-task, the participants were instructed to fill out a multi-dimensional subjective workload rating scale, the NASA Task Load indeX (NASA TLX; Hart & Staveland 1988). The NASA TLX provides a subjective measure of effort associated with the task, and was also used in our previous study (Pals et al., 2013).

The procedure for Experiment 1, including audiometric tests and training, lasted approximately 2 hours.

*Results*

Looking first at the average speech intelligibility scores for Experiment 1 (Figure 1, top-left panel), intelligibility, in percentage of sentences correctly repeated, was comparable across all conditions, at just below ceiling as intended. The RTs on the secondary rhyme judgment task for Experiment 1 are shown in the middle-left panel of Figure 1. Incorrect trials for the visual rhyme-judgment task were excluded from analysis of the RTs; they accounted for about 4% of the responses. Due to the nature of the secondary rhyme judgment task, the dataset consisted of unequal trial numbers for each cell. Therefore, the data were analyzed using linear mixed-effects (LME) models in R (lme4-package version 1.1-7). Factors were added to the model incrementally and only included if they significantly improved the fit of the model. Random intercepts for both participant and sentence ID were included in the model, to account for differences in baseline performance between participants and between sentences. Including presentation order as a factor in the model in order to account for learning effects over the course of the experiment, significantly improved the fit of the model ($\chi^2(1) = 83.55$, $p < 0.001$). The factors of interest were 'listening mode' (MonCI, BinCI, EAS300, and EAS600) and 'spectral resolution' (6-channel and 8-channel). However, including spectral resolution in the model did not show a significant main effect of spectral resolution, no significant interactions, and did not improve the fit of the model ($\chi^2(1) = 2.6358$, $p = 0.6205$). Spectral resolution was therefore not included in the model.

To see if individual differences in intelligibility scores per condition can explain some of the observed differences in RT, a model was constructed including the intelligibility scores as a factor. However, including speech intelligibility in the model did not improve the fit ($\chi^2(1) = 3.5461$, $p = 0.05969$) and was therefore not included.

The preferred model therefore includes the factor 'listening mode' (with four levels: MonCI, BinCI, EAS300, and EAS600) and the numeric factor 'presentation order' and random intercepts for 'participant' and 'sentence'. In case of a non-numeric factor such as 'listening mode', the summary of a linear model estimates the value of the reference level, and lists the estimated differences between each of the other levels and the reference level. In our experiment design it makes sense to compare the BinCI, EAS300, and EAS600 to the reference level MonCI. However, as can be seen in Figure 1, the RTs were longest for BinCI, and to see whether the RTs for the EAS conditions were significantly shorter than for BinCI, the model summary is also shown with BinCI as the reference level.

*Table 2. Summary of linear models for dual-task RT results for Experiment 1*

| *Dual-task RT results* | *Estimate (ms)* | *Std. Error* | *df* | *t value* | *Pr(>\|t\|)* |
|---|---|---|---|---|---|
| MonCI (Intercept) | 1.086 | 0.032 | 24 | 33.58 | < 0.001 *** |
| OrderNR | -0.012 | 0.001 | 1.365e+04 | 1.85 | < 0.001 *** |
| BinCI | 0.016 | 0.008 | 1.369e+04 | -1.96 | 0.065 |
| EAS300 | -0.017 | 0.008 | 1.368e+04 | -1.76 | 0.050 |
| EAS600 | -0.015 | 0.008 | 1.362e+04 | -9.32 | 0.078 |
| BinCI (Intercept) | 1.102 | 0.032 | 24 | 34.0 | < 0.001 *** |
| OrderNR | -0.012 | 0.001 | 1.365e+04 | -9.3 | < 0.001 *** |
| MonCI | -0.016 | 0.008 | 1.369e+04 | -1.8 | 0.064 |
| EAS300 | -0.032 | 0.008 | 1.362e+04 | -3.8 | < 0.001 *** |
| EAS600 | -0.030 | 0.008 | 1.364e+04 | -3.6 | < 0.001 *** |

Note: Both models included the factor 'listening mode' (levels: MonCI, BinCI, EAS300, and EAS600) and the numeric factor 'presentation order'. The top half of the table shows the results for the model using the listening mode MonCI as the reference level and the bottom half of the table shows the results for the model using listening mode BinCI as reference level.

The model with the MonCI listening mode as reference level is summarized in the top half of Table 2, the same model with BinCI as the reference is summarized in the bottom half of Table 2. When comparing to MonCI as the reference, adding either simulated electric or acoustic signal in the other ear did not significantly change the RTs. The RTs for MonCI are on average halfway between the RTs for BinCI (which are estimated to be 16 ms longer than the RTs for MonCI) and the RTs for both EAS conditions (RTs for EAS300 and EAS600 are estimated to be 17ms and 15 ms faster than MonCI, respectively). In order to examine the differences between BinCI and the EAS conditions, the model was also examined using BinCI as the reference level. The intercept of the model corresponds with the listening mode 'BinCI' and was estimated at 1.102s ($\beta$ = 1.102, *SE* = 0.032, $t$ = 34.0, $p$ < 0.001). The difference between this estimate and the actual mean RT for the BinCI listening modes as shown in Figure 1 stems from the inclusion of the random intercept for sentence ID in the model. The effect of presentation order is significant and estimated at -12 ms ($\beta$ = -0.012, *SE* = 0.001, $t$ = -9.3, $p$ < 0.001), implying that participants' RTs become 12 ms faster with each consecutive task. The estimates for the other listening modes are all relative to the intercept, the estimated RT for BinCI. Both EAS listening modes resulted in significantly faster response times than BinCI; EAS300 resulted in 32 ms faster response times ($\beta$ = -0.032, *SE* = 0.008, $t$ = -3.8, $p$ <

0.001), EAS600 in 30 ms faster response times ($\beta$ = -0.030, *SE* = 0.008, *t* = -3.6, *p* < 0.001). Response times for MonCI appear to be slightly faster than for BinCI, however, this difference is not significant ($\beta$ = -0.016, *SE* = 0.008, *t* = -1.8, *p* = 0.064).

The bottom-left panel of Figure 1 shows the average NASA TLX scores for Experiment 1, for single-task and dual-task presentation. Since the NASA TLX scores for the dual-task conditions can be interpreted as an effort rating for the combined listening and secondary rhyme judgment task rather than the listening task alone, the analysis of the NASA TLX results focuses on the single task NASA TLX scores. To be able to compare the NASA TLX results with the RT results the analysis was also performed using LME models. A random intercept for participant was included in the model, however, since the NASA TLX scores consisted of one value per whole test block, no random intercept per sentence was included. Including the single task speech intelligibility significantly improved the model ($\chi^2(1) = 20.923$, $p < 0.001$). Including presentation order ($\chi^2(1) = 0.3839$, $p = 0.5355$) or spectral resolution ($\chi^2(1) = 6.1077$, $p = 0.1912$) in the model did not significantly improve the fit.

*Table 3. Summary of the linear model for the NASA TLX results for Experiment 1*

| *ST NASA TLX results* | *Estimate* | *Std. Error* | *df* | *t value* | *Pr(>\|t\|)* |
|---|---|---|---|---|---|
| MonCI (Intercept) | 85.6926 | 11.0292 | 154.8800 | 7.770 | < 0.001 *** |
| SpeechScore | -0.6301 | 0.1185 | 136.3200 | -5.316 | < 0.001 *** |
| BinCI | -3.7406 | 1.9879 | 137.0200 | -1.882 | 0.062 |
| EAS300 | -3.1712 | 2.0273 | 139.0400 | -1.564 | 0.120 |
| EAS600 | -3.1712 | 2.1602 | 150.8900 | -1.448 | 0.150 |

Note: The model included the factor 'listening mode' (levels: MonCI, BinCI, EAS300, and EAS600), the numeric factor 'presentation order', and used the listening mode MonCI as the reference level.

The best model for the NASA TLX data includes the factors 'speech score' and 'listening mode' and random intercepts for 'participant', this model is summarized in Table 3. The intercept corresponds to the estimated NASA TLX score for MonCI, extrapolated for a speech score of 0% sentence correct, this is estimated at a score of 85.7 out of 100 ($\beta$ = 85.6926, *SE* = 11.0292, *t* = 7.770, *p* < 0.001). The effect of speech score is significant and estimated at -0.63 ($\beta$ = -0.6301, *SE* = 0.1185, *t* = 05.316, *p* < 0.001), meaning that an estimated NASA TLX score for MonCI at 100% intelligibility would be 85.7 − 63.0 = 22.7. None of the listening modes differed significantly from the reference level MonCI.

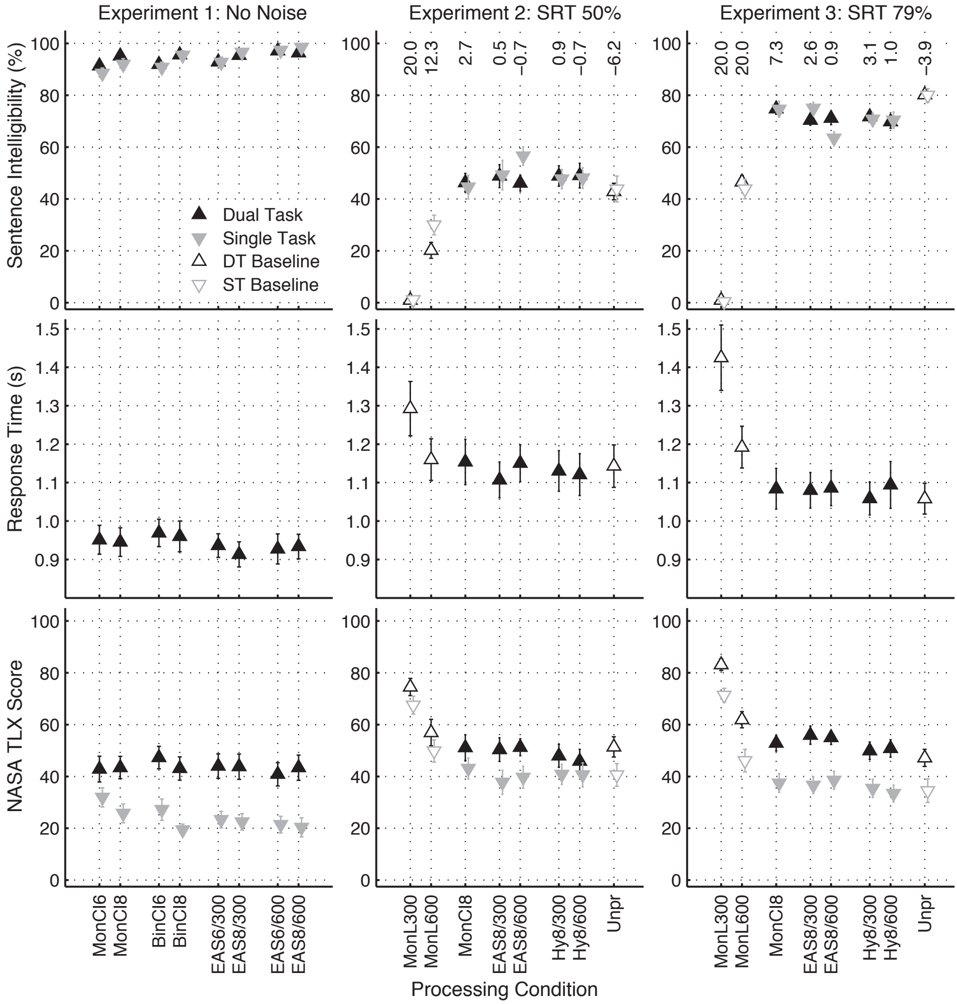


*Figure 1:* The results for Experiments 1, 2, and 3 are shown in the left, middle, and right column, respectively, with experimental conditions on the x-axes (experimental conditions are summarized in Table 1 for Experiment 1 and in Table 4 for Experiment 2 and 3). Up triangles show dual-task results, and down triangles show single-task results, error bars represent one standard error. Closed symbols show conditions of interest that are included in the analysis, open symbols show conditions that were tested for reference but not included in the analysis. The top row shows the single and double task speech intelligibility scores in percentage of sentences correctly repeated, with for Experiment 2 and 3 the SNRs at which each of the conditions were presented at the very top of the figure in dB SNR, the middle row shows the dual-task response times on the secondary task, and the bottom row shows the NASA TLX ratings (higher scores indicate more effort).

To summarize, speech intelligibility was near ceiling for all conditions, although exact speech scores varied slightly across participants and conditions. The dual-task results of Experiment 1 showed a significant benefit of EAS (i.e., faster RTs), with both 300 and 600 Hz LPF speech, compared to binaural CI, however, monaural CI was not significantly different from either binaural CI or EAS. The subjective measure of listening effort, the NASA TLX, showed no effect of listening mode. Any difference in NASA TLX ratings between conditions or participants could be entirely contributed to effects of small individual differences in intelligibility.

## EXPERIMENT 2: SPEECH IN NOISE AT 50% INTELLIGIBILITY

*Motivation*

In Experiments 2 and 3, the effect of simulated EAS, compared to CI alone, on listening effort was examined in interfering noise at equal intelligibility levels. In Experiment 2, 50% sentence intelligibility was used. Equal intelligibility across conditions was achieved by presenting the different processing conditions at different SNRs. We hypothesized that even at equal intelligibility, EAS may provide an additional benefit in reduced listening effort.

Since the results of Experiment 1 revealed no effect of spectral resolution between the 6- and 8-channel CI and EAS conditions, the 6-channel conditions were dropped in favor of including additional EAS configurations. In Experiment 1 we observed significant differences in the dual-task measure of listening effort between binaural CI and the EAS conditions. Listening effort for monaural CI did not differ significantly from either binaural CI or EAS. We believe, however, that since most CI users wear monaural CI, the comparison between monaural CI and EAS is a more meaningful comparison. Therefore, for Experiments 2 and 3, we chose to compare listening effort for speech in noise in the following simulated device configurations: a) contralateral EAS in which the LPF sound is presented to the ear contralateral to the simulated CI (the same as in Experiment 1), b) Hybrid in which the LPF sound is presented to both ears in the ear with the CI simulation replacing the overlapping lower frequency channels of the CI (new compared to Experiment 1).

*Methods*

The procedure for Experiment 2 was similar to Experiment 1, therefore only the differences are described below.

*Participants.* Twenty new participants were recruited for participation in Experiment 2. All were NH, native Dutch speaking, young adults (age range: 18–33 years, mean: 20 years; 11 female). The results of one participant were excluded from the analysis of the NASA TLX, because the questionnaire was not filled out completely.

*Stimuli.* The same auditory and visual stimuli as in Experiment 1 were used. In these experiments the 6-channel CI simulation conditions were dropped in favor of additional listening modes. Besides the monaural 8-channel CI conditions and 8-channel EAS conditions used in Experiment 1, monaurally presented acoustic simulations of LPF speech (MonL300, MonL600), as well as 'ipsilateral EAS' also referred to as 'Hybrid CI' simulations (CI simulation combined with LPF speech presented to the same ear; Hy8/300, Hy8/600) were added (see Table 4). The 8-channel simulations were preferred over the 6-channel simulations to ensure that the desired SRTs would be attainable at reasonable SNRs. A baseline, unprocessed-speech condition was also added for comparison.

The noise used in both the speech in noise test and the actual experiment was a speech-shaped steady-state noise that was provided with the VU speech corpus (Versfeld et al., 2000).

*Table 4. Summary of listening conditions for Experiments 2 and 3*

| *Label* | *Left ear* | *Right ear* | *Exp 2 SRT 50% SNR (SD)* | *Exp 3 SRT 79% SNR* |
|---|---|---|---|---|
| MonL300 | 300 Hz LPF | - | 20.0* | 20.0* |
| MonL600 | 600 Hz LPF | - | 12.3 (3.71) | 20.0* |
| MonCI8 | - | 8 Ch. CI | 2.7 (1.76) | 7.3 |
| EAS8/300 | 300 Hz LPF | 8 Ch. CI | 0.5 (1.40) | 2.7 |
| EAS8/600 | 600 Hz LPF | 8 Ch. CI | -0.7 (1.07) | 0.9 |
| Hy8/300 | 300 Hz LPF | 300 Hz + 6/8 Ch. | 0.9 (1.47) | 3.2 |
| Hy8/600 | 600 Hz LPF | 600 Hz + 5/8 Ch. | -0.7 (0.99) | 1 |
| Unpr | 80-6000 Hz | 80-6000 Hz | -6.2 (0.73) | -3.9 |

Note: The first two columns show the stimuli that were presented to the left and to the right ear, respectively, in each of the conditions. The last two columns show the average SNRs at which the desired SRTs were obtained. Values in brackets indicate standard deviations. The entries marked by asterisks show the conditions where the target intelligibility level could not be reached, and therefore the SNR was set to a nominal value of 20 dB.

*Presentation levels*. The noise was presented continuously throughout each task, and at the same level (50 dBA) for all participants and all conditions. The presentation levels of sentences for each condition were determined by an adaptive speech-in-noise test prior to the experiment. Presentation levels were determined for each participant individually, prior to the experiment, by means of a speech-in-noise test using a 1-down-1-up adaptive procedure. The speech-in-noise procedure used to determine the participants' individual SRTs was similar to the speech audiometry used in clinics in the Netherlands (Plomp, 1986). Each test used one list of 13 sentences. The first sentence was used to quickly converge on the approximate threshold of intelligibility. Starting at 8 dB below the noise and increasing the level in steps of 4 dB, the sentence was repeatedly played until the entire sentence was correctly reproduced. From this level the adaptive procedure started, where the SNR was increased or decreased by 2 dB after an incorrect or correct response, respectively. A list of 13 sentences was thus sufficient for at least 6 reversals (often about 8), which results in a reliable estimate of the 50% SRT (Levitt, 1971). The average SRTs (in dB SNR) for all 20 participants are listed in Table 4, second column from right.

Attaining the desired 50% intelligibility levels was not possible for 300 Hz LPF speech. Therefore, we chose to present sentences for this condition at 20 dB SNR.

*Procedure*. At the start of the experiment the appropriate presentation levels for each individual participant were determined using the adaptive speech-in-noise test. This additional test increased testing time by about 15 minutes, and provided some additional familiarization with the sentence material and stimulus processing. Therefore, training was done without feedback, to reduce testing time, and lasted 10 min. For the rest, the procedure was identical to Experiment 1. The entire session lasted around 2 hours.

*Results*

The speech intelligibility results for Experiment 2 are shown in the top-middle panel of Figure 1. The LPF300 and LPF600 conditions were included as a reference, and to show that LPF speech by itself produces limited intelligibility. The unprocessed speech condition was included as a normal-hearing reference point. In Experiment 2, the desired intelligibility level of 50% sentence recognition was achieved by determining the appropriate SNRs for each condition using an adaptive procedure at the start of the experiment. These SNRs are included in the figure. On average, the intelligibility scores were indeed close to 50% for the conditions of interest in this experiment.

The center panel of Figure 1 shows the RTs on the secondary rhyme judgment task for Experiment 2. Incorrect trials for the visual rhyme-judgment task were excluded from analysis of the RTs; they accounted for about 5% of the trials. As the goal of this study was to examine the effect of providing LPF speech to complement CI simulated speech, the conditions of interest are CI, EAS300, EAS600, Hy300 and Hy600; the analysis therefore focuses on these five conditions. The analysis was performed using LME models.

The results were modeled in a design that most closely resembles the contrasting dimensions in this design. Included in the model are: the effect of EAS on average compared to CI alone, the contrast between contralateral EAS and Hybrid configuration (listening mode), and the contrast between 300 and 600 Hz LPF acoustic sound. Including task order in the model significantly improved the fit ($\chi^2(1) = 27.258$, $p < 0.001$). Speech scores were included in the model to account for differences in speech scores between participants and conditions and to see how much of the observed differences in RT can be attributed to differences in intelligibility. Including speech scores did significantly improve the model ($\chi^2(1) = 38.418$, $p < 0.001$). Each condition was presented at an individually determined SNR different for each participant, however, including presentation SNR in the model was not warranted ($\chi^2(1) = 0.604$, $p = 0.437$).

*Table 5. Summary of the Linear Model for the Dual-task RT Results for Experiment 2*

| *Dual-task RT results* | *Estimate (ms)* | *Std. Error* | *df* | *t value* | *Pr(>\|t\|)* |
|---|---|---|---|---|---|
| CI (Intercept) | 1.362 | 0.052 | 34 | 25.973 | < 0.001 *** |
| SpeechScore | -0.002 | 0.000 | 7968 | -6.207 | < 0.001 *** |
| OrderNR | -0.014 | 0.003 | 7976 | -5.360 | < 0.001 *** |
| EAS | -0.030 | 0.013 | 7956 | -2.243 | 0.025 * |
| EAS:Mode | -0.002 | 0.012 | 7958 | 0.131 | 0.896 |
| EAS:LPF | -0.017 | 0.012 | 7954 | 1.412 | 0.158 |
| EAS:Mode:LPF | -0.017 | 0.024 | 7958 | 0.719 | 0.472 |

Note: The model included the factors 'speech score' and 'presentation order', EAS (the contrast between CI alone and EAS regardless of configuration or LPF cut-off), and within the EAS conditions: the factor 'listening mode' (levels: EAS and Hybrid) and the factor LPF cut-off frequency (levels: 300 Hz and 600 Hz).

Table 5 summarizes the model. The intercept of the model corresponds to the RT for CI simulated speech alone extrapolated for 0 % sentence intelligibility, and is estimated at 1.362 seconds ($\beta = 1.362$, $SE = 0.052$, $t = 25.973$, $p < 0.001$). The effect of speech score is significant and estimated at – 2 ms ($\beta = -0.002$, $SE = 0.000$, $t = -6.207$, $p < 0.001$), suggesting a decrease in RT of 2 ms for each 1-percentage point increase in intelligibility. The estimated RT for CI alone at 50% intelligibility is therefore 1.362 − 0.100 = 1.262 seconds. The model shows a significant effect of presentation order, estimated at – 14 ms ($\beta = -0.014$, $SE = 0.003$, $t = -5.360$, $p < 0.001$), implying 14ms faster RTs for each consecutive task. The effect of EAS in general compared to CI alone was significant and estimated at – 30 ms ($\beta = -0.030$, $SE = 0.013$, $t = -2.243$, $p = 0.025$) suggesting on average 30 ms faster RTs for EAS conditions than for simulated CI alone. Between the four different EAS conditions no significant differences were found.

The average NASA TLX ratings for Experiment 2, for both dual and single tasks, are shown in the bottom-middle panel of Figure 1. The NASA TLX results were analyzed in the same manner as the RT results. Adding presentation order to the model was not warranted ($\chi^2(1) = 0.1712$, $p = 0.6791$). Including presentation speech scores did significantly improve the fit of the model ($\chi^2(1) = 46.427$, $p < 0.001$).

*Table 6. Summary of the linear model for the NASA TLX results for Experiment 2*

| *ST NASA TLX results* | *Estimate* | *Std. Error* | *df* | *t value* | *Pr(>\|t\|)* |
|---|---|---|---|---|---|
| CI (Intercept) | 59.907 | 4.506 | 41.51 | 13.294 | < 0.001 *** |
| SpeechScore | -0.378 | 0.049 | 72.66 | -7.675 | < 0.001 *** |
| EAS | -0.805 | 2.089 | 71.06 | -0.385 | 0.701 |
| EAS:Mode | -0.390 | 1.870 | 71.07 | 0.209 | 0.835 |
| EAS:LPF | 2.484 | 1.856 | 71.04 | 1.338 | 0.185 |
| EAS:Mode:LPF | -2.906 | 3.690 | 71.02 | -0.787 | 0.434 |

Note: The model included the factors 'speech score', EAS (the contrast between CI alone and EAS regardless of configuration or LPF cut-off), and within the EAS conditions: the factor 'listening mode' (levels: EAS and Hybrid) and the factor LPF cut-off frequency (levels: 300 Hz and 600 Hz).

The model is summarized in Table 6. The intercept corresponds to the estimated NASA TLX score for CI simulations alone at 0% intelligibility, and is estimated at a score of 60 out of 100 ($\beta = 59.907$, $SE = 4.506$, $t = 13.294$, $p < 0.001$). There is a significant effect of speech score estimated at -0.378 ($\beta$ = -0.378, $SE$ = 0.049, $t$ = -7.675, $p$ < 0.001), implying a 0.378 decrease in NASA TLX score for each 1-percentage point increase in speech intelligibility. This means that the estimated NASA TLX score for CI alone at 50% sentence intelligibility is 60 − 19 = 41. For the NASA TLX results, none of the effects of EAS were significant.

In short, speech intelligibility was successfully fixed at 50% sentence recognition for the conditions of interest, at different SNRs for each condition (see Table 4). The dual-task results for Experiment 2 showed a significant benefit of EAS compared to monaural CI (i.e., faster RTs), and no difference between the different EAS configurations. The NASA TLX results showed no significant difference in ratings between CI and EAS conditions, suggesting that CI simulated speech and each of the four EAS conditions in noise were rated as equally effortful.

## EXPERIMENT 3: SPEECH IN NOISE AT 79% INTELLIGIBILITY

*Motivation*

Similar to Experiment 2, listening effort was evaluated for speech in noise. However, in Experiment 3, speech intelligibility level was fixed at 79 %, in order to compare effects in listening effort at fixed intelligibility level at a different, shallower point in the psychometric function. The same simulated device configurations as in Experiment 2 were tested in this experiment. The conditions as well as the SNRs to achieve the 79% sentence intelligibility level are listed in Table 4.

*Methods*

The procedure for Experiment 3 was similar to Experiment 2, therefore, only the differences are described below.

*Participants.* Twenty new participants were recruited for participation in Experiment 3. All were NH, native Dutch speaking, young adults (age range: 19–26 years, mean: 21 years; 8 female).

Furthermore, ten additional new participants were recruited for a short test to determine the SRTs for 79% sentence intelligibility. All were NH, native Dutch speaking, young adults (age range: 19–24 years, mean: 22 years; 6 female).

*Presentation levels*. Presentation levels were determined with a 3-down-1-up adaptive procedure (Levitt, 1971), similar to Experiment 2, except that the SNR was decreased by 2 dB after 3 consecutive correct responses instead of after each correct response. This procedure requires a substantial amount of time and a large number of sentences to obtain 6 to 8 reversals. Therefore, it was not feasible to determine SRTs for each participant individually prior to the experiment. Thus, for this experiment SRTs were determined beforehand with 10 new participants, similar in age and hearing levels to the participants of the experiment. The average SRTs, listed in the rightmost column of Table 4, were used in the experiment.

Attaining the desired 79% sentence recognition with 300 Hz and 600 Hz LPF speech was not feasible. Therefore, we chose to present sentences during these conditions at 20 dB SNR.

*Procedure*. As the presentation levels were determined with a different participant group, there was no concern of additional testing time (as was the case in Experiment 2). The participants of Experiment 3 therefore received the same 20-minute training (with feedback) as participants in Experiment 1 and were tested in an identical procedure to Experiment 1. The entire session lasted around 2 hours.

*Results*

The speech intelligibility scores for Experiment 3 are shown in the top-right panel of Figure 1. As in Experiment 2, the conditions LPF 300, LPF 600 and Unprocessed were included as reference points and therefore excluded from the analysis. In Experiment 3 the desired intelligibility level of 79% sentence recognition was achieved by presenting the conditions at SNRs determined with a group of 10 participants similar in age and hearing level to the participants in this experiment. These SNRs are included in the figure. On average, the intelligibility scores were around 75% and speech intelligibility in the dual task did not vary significantly across the conditions of interest.

The middle-right panel shows the RTs on the secondary rhyme judgment task for Experiment 1. Incorrect trials for the visual rhyme-judgment task were excluded from analysis of the RTs; they accounted for about 4% of the responses for Experiment 3. Including presentation order in the model significantly improved the fit ($\chi^2(1) = 50.084$, $p < 0.001$), as did including speech score ($\chi^2(1) = 29.189$, $p < 0.001$).

*Table 7. Summary of the linear model for the dual-task RT results for Experiment 3*

| *Dual-task RT results* | *Estimate (ms)* | *Std. Error* | *df* | *T value* | *Pr(>\|t\|)* |
|---|---|---|---|---|---|
| CI (Intercept) | 1.493 | 0.069 | 97 | 21.753 | < 0.001 *** |
| SpeechScore | -0.004 | 0.001 | 8131 | -5.404 | < 0.001 *** |
| OrderNR | -0.016 | 0.002 | 8256 | -6.430 | < 0.001 *** |
| EAS | -0.011 | 0.013 | 8207 | -0.838 | 0.402 |
| EAS:Mode | -0.011 | 0.012 | 8216 | -1.010 | 0.312 |
| EAS:LPF | 0.017 | 0.012 | 8224 | 1.521 | 0.128 |
| EAS:Mode:LPF | -0.005 | 0.023 | 8247 | -0.220 | 0.826 |

Note: The model included the factors 'speech score' and 'presentation order', EAS (the contrast between CI alone and EAS regardless of configuration or LPF cut-off), and within the EAS conditions: the factor 'listening mode' (levels: EAS and Hybrid) and the factor LPF cut-off frequency (levels: 300 Hz and 600 Hz).

The model is summarized in Table 7. The intercept corresponds to RTs to CI simulated speech alone in noise at 0% intelligibility and is estimated at 1.493 sec ($\beta = 1.493$, $SE = 0.069$, $t = 21.753$, $p < 0.001$). The effect of speech score is significant and estimated at -4 ms ($\beta = -0.004$, $SE = 0.001$, $t = -5.404$, $p < 0.001$), implying a 4 ms reduction in RT for each 1-percentage point increase in speech score. This means that the RT for CI simulated speech in noise at 79% intelligibility is estimated at 1.493 − (0.004 * 79 = 0.316) = 1.177 seconds. Presentation order has a significant effect on RT and is estimated at -16 ms ($\beta = -0.016$, $SE = 0.002$, $t = -6.430$, $p < 0.001$), suggesting a 16 ms decrease in RT for each consecutive task. None of the modeled contrasts between simulated CI/EAS, listening mode and LPF conditions revealed any significant differences.

*Table 8. Summary of the linear model for the NASA TLX results for Experiment 3*

| *ST NASA TLX results* | *Estimate* | *Std. Error* | *df* | *t value* | *Pr(>\|t\|)* |
|---|---|---|---|---|---|
| CI (Intercept) | 56.107 | 7.693 | 93.49 | 7.294 | < 0.001 *** |
| SpeechScore | -0.250 | 0.092 | 81.20 | -2.707 | 0.008 ** |
| EAS | -2.649 | 2.385 | 75.25 | -1.111 | 0.270 |
| EAS:Mode | -2.838 | 2.101 | 75.06 | -1.351 | 0.181 |
| EAS:LPF | -1.532 | 2.168 | 75.45 | -0.707 | 0.482 |
| EAS:Mode:LPF | -1.094 | 4.319 | 75.40 | -0.253 | 0.800 |

Note: The model included the factors 'speech score', EAS, and within the EAS conditions: the factor 'listening mode' (levels: EAS and Hybrid) and the factor LPF cut-off frequency (levels: 300 Hz and 600 Hz).

The average NASA TLX ratings for Experiment 3 are shown in the bottom-right panel of Figure 1. The NASA TLX data were modeled in a similar manner as for Experiment 2. Adding presentation order to the model was not warranted ($\chi^2(1) = 1.3535$, $p = 0.2447$). Including speech score in the model did significantly improve the fit ($\chi^2(1) = 7.4108$, $p = 0.006$). The model is summarized in Table 8. The NASA TLX score for CI alone at 0% intelligibility is estimated at 56 out of 100 ($\beta = 56.107$, $SE = 7.693$, $t = 7.294$, $p < 0.001$). The effect of speech score was significant and estimated at -0.25, implying a decrease in NASA TLX score of 0.25 per 1 percentage point increase in speech intelligibility. The NASA TLX score is therefore estimated to be 56 − (79 * 0.25 = 19.75) = 36 out of 100. Between the different listening conditions, simulated CI and the four EAS conditions, effort was not rated any differently.

To summarize, speech intelligibility was successfully fixed at 79% for the conditions of interest, at different SNRs for each condition (see Table 4). The dual-task results for Experiment 3 showed no difference in listening effort for any of the conditions of interest. The NASA TLX showed no benefits in listening effort between any of the simulated CI and EAS conditions.

## DISCUSSION

The goal of this study was to examine the potential benefits of providing low-frequency acoustic speech in addition to the electronic signal of a CI (i.e. EAS) in terms of listening effort. We hypothesized that EAS compared to CI hearing alone would, in addition to improving speech understanding in noise, also reduce listening effort. To allow for a systematic approach investigating several different device configurations (monaural or binaural CI listening, hybrid or bimodal EAS configurations with varying amounts of simulated residual hearing, and different levels of spectral resolution), unhindered by CI users' individual differences in, for example, residual hearing, we used acoustic CI simulations with young normal-hearing participants. The effect of EAS on listening effort was investigated at fixed intelligibility levels in order to separate effects of EAS from effects of speech intelligibility. We conducted three dual-task experiments, each with speech intelligibility fixed at a different point on the psychometric function (Experiment 1 at near ceiling intelligibility, Experiment 2 at 50% intelligibility, and Experiment 3 at 79 % intelligibility, with the first experiment conducted in quiet and the latter two in background noise). Listening effort was measured objectively in a dual-task paradigm with a secondary, speeded rhyme-judgment task in a dual-task paradigm, and subjectively using the NASA TLX workload rating scale. The expectation was that these measures would show an improvement in listening effort for the simulated EAS configurations compared to simulated CI alone, even for fixed intelligibility.

Because speech intelligibility was fixed at 3 different levels for the three different experiments, we cannot comment on the effects of EAS on intelligibility observed in these experiments compared to the literature. However, research shows that EAS improves speech understanding in noise, both for NH listeners presented with simulated EAS (Brown & Bacon, 2009; Dorman, Spahr, Loizou, Dana, & Schmidt, 2005; Kong & Carlyon, 2007) and for CI users with residual hearing (Kiefer et al., 2005; Kong, Stickney, & Zeng, 2005). This EAS benefit for speech perception in noise suggests that EAS will allow the desired intelligibility levels to be achieved at lower SNRs. Our results are in line with this expectation. In Experiment 3, 79% intelligibility was achieved by presenting the simulated CI condition at 7.3 dB SNR, and the EAS conditions on average at 1.9 (range 0.9 to 3.1) dB SNR, a difference in SNR of on 5.4 dB. In Experiment 2, 50% intelligibility was achieved by presenting simulated CI alone at 2.7 dB SNR, and the EAS conditions on average at 0 (range -0.7 to 0.9) dB SNR a difference in SNR of on average 2.7 dB. These values are very similar to between-group values reported for actual CI users: Dorman and Gifford (2010) showed that speech reception thresholds (implying 50% intelligibility) were on average 2.62 dB better for EAS listeners than for unilateral CI users.

The aim of this study was to investigate effects of EAS on listening effort independently of speech intelligibility, and thus at fixed intelligibility levels. In two out of the three experiments, the dual-task results showed such a benefit of EAS on listening effort. In both Experiment 1, for speech in quiet at near ceiling intelligibility, and Experiment 2, for speech in noise at 50% intelligibility, the dual-task measure of listening effort, the RTs on the secondary task, were significantly shorter for the EAS conditions than for CI. This is in line with what we expected based on research that shows that EAS improves subjective hearing device benefit (Gstoettner et al., 2004). Nevertheless, the subjective measure of listening effort in the current study, the NASA TLX, showed no difference in subjective effort rating between EAS and CI alone conditions. The difference in findings between the subjective and objective measure of listening effort is in line with our previous research (Pals et al., 2013), as well as research by others (Feuerstein, 1992; Gosselin & Gagné, 2011; Zekveld, Kramer, & Festen, 2010).

In our previous study, the NASA TLX did show significant effects for those conditions that also resulted in significant differences in intelligibility. However, when intelligibility reached ceiling, the NASA TLX no longer showed significant changes while the dual-task did still capture further changes in listening effort (Pals et al., 2013). In the current study we specifically investigated conditions at equal intelligibility. The NASA TLX results did show a significant effect of intelligibility: even though the differences in intelligibility were small, participants did rate conditions that were slightly less intelligible as more effortful. The objective measure of listening effort, dual-task RTs, appears better suited for showing differences in listening effort at equal levels of intelligibility than the subjective self-report scale the NASA TLX. This suggests that using an objective measure can uncover benefits that speech intelligibility and subjective self-report do not reveal.

For speech in quiet, at near-ceiling intelligibility (Experiment 1), the dual-task RT results showed a significant benefit of simulated EAS compared to binaural CI but not compared to monaural CI. While the RTs for the simulated monaural RTs were on average longer than the RTs for EAS conditions, they were shorter than the average RTs for simulated binaural CI, about halfway between the two and not significantly different from either. Intuitively, one would expect monaural CI speech to be more effortful to understand than binaural CI, rather than less effortful as shorter RTs suggest, although, this difference was not significant and could thus have been coincidental. What could have affected the results for the monaural CI condition is a difference in presentation level; to account for binaural loudness summation (Epstein & Florentine, 2012), the monaural CI conditions were presented at a slightly higher sound level (65 dBA) than the binaural CI and EAS conditions (at 60 dBA in each ear). Whether this resulted in exactly equal perceived loudness for the monaural compared to the other conditions is not necessarily due to the difference in frequency content between the CI simulated and LPF signals. Differences in level and perceived loudness could possibly have affected the dual-task outcomes.

For speech in noise, at 50% intelligibility (Experiment 2), the dual-task RT results show a significant effect of EAS in general compared to CI alone, however at 79% intelligibility

(Experiment 3), no significant difference in listening effort between the conditions of interest was found. In noise, EAS allows listeners to reach a target level of speech understanding at less favorable SNR, as has been documented by previous research (Büchner et al., 2009; Dorman & Gifford, 2010; Qin & Oxenham, 2006). In our Experiments 2 and 3, the simulated EAS listening conditions were presented at lower SNRs than the CI alone listening conditions to achieve equal speech understanding across conditions. Prior research has shown that a lower SNR can result in higher listening effort (Zekveld et al., 2010). Therefore, if EAS does not affect listening effort at all, one would expect increased listening effort at these lower SNRs. Our results, however, show the opposite; in Experiment 2, EAS improved listening effort compared to CI alone, despite being presented at lower SNRs. In Experiment 3, while the results did not show an improvement in listening effort for the EAS conditions compared to CI, neither did they show an increase in listening effort due to the 5.4 dB lower SNR for the EAS conditions.

In summary, from the results of this study we conclude that simulated EAS does provide a benefit in listening effort compared to simulated CI alone, at least in conditions in which the effect of EAS on listening effort is not overshadowed by the counter-directional effect of background noise on listening effort. Whether the same holds true for cochlear implant users should be addressed in future research.