JSLHR

Article

Listening Effort With Cochlear Implant Simulations

Carina Pals,[a,b] Anastasios Sarampalis,[c] and Deniz Başkent[a,b]

Purpose: Fitting a cochlear implant (CI) for optimal speech perception does not necessarily optimize listening effort. This study aimed to show that listening effort may change between CI processing conditions for which speech intelligibility remains constant.
Method: Nineteen normal-hearing participants listened to CI simulations with varying numbers of spectral channels. A dual-task paradigm combining an intelligibility task with either a linguistic or nonlinguistic visual response-time (RT) task measured intelligibility and listening effort. The simultaneously performed tasks compete for limited cognitive resources; changes in effort associated with the intelligibility task are reflected in changes in RT on the visual task. A separate self-report scale provided a subjective measure of listening effort.
Results: All measures showed significant improvements with increasing spectral resolution up to 6 channels. However, only the RT measure of listening effort continued improving up to 8 channels. The effects were stronger for RTs recorded during listening than for RTs recorded between listening.
Conclusion: The results suggest that listening effort decreases with increased spectral resolution. Moreover, these improvements are best reflected in objective measures of listening effort, such as RTs on a secondary task, rather than intelligibility scores or subjective effort measures.

Key Words: cochlear implants, listening effort, dual task, reaction time, computer simulation, hearing, speech perception

Cochlear implants (CIs) are implantable auditory prostheses that partially restore hearing to people with profound hearing impairment. To accomplish this, a sound processor translates the incoming acoustic signal to electrical pulse trains, which are transmitted to the auditory nerve by an electrode array inserted in the cochlea. From the early days of CI research, the primary focus has been on improving the ability to understand speech (e.g., Fishman, Shannon, & Slattery, 1997; Fu, 2002; Manrique et al., 1999; Pfingst, Zwolan, & Holloway, 1997; Skinner et al., 1994; Wilson et al., 1991). In this context, CI benefit has typically been measured using speech intelligibility tests. Research on hearing impairment, however, has shown that cognitive measures (e.g., the response times on a verbal sentence verification test [Baer, Moore, & Gatehouse, 1993], the dual-task paradigm [e.g., Anderson Gosselin & Gagné, 2011; Sarampalis, Kalluri, Edwards, & Hafter, 2009], and pupillometry [Zekveld, Kramer, & Festen, 2010]) can provide an additional layer of information to complement intelligibility measures. The additional performance information these measures provide has been linked to ease of listening (e.g., Baer et al., 1993), or listening effort (e.g., Anderson Gosselin & Gagné, 2011), which was the focus of the present study.

Research on effort in general is based on the historical work of Broadbent (1958), Baddeley and Hitch (1974), and Kahneman (1973), each of whom proposed a shared, limited cognitive resource, later commonly referred to as *working memory,* that can be allocated to various tasks, as necessary. A more recent version of Baddeley's theory proposes a phonological loop for storing and manipulating incoming auditory information, a visuospatial sketchpad for visual information, an episodic buffer that stores and retrieves information from long-term memory, and a central executive that coordinates the execution of complex tasks (Baddeley, 2012). An effortful task requires a large proportion of the resources relevant to the task or a considerable involvement of the central executive, or both. Listening effort can then be defined as the proportion of limited cognitive resources engaged in interpreting the incoming auditory signal. It has been suggested that the presence of noise or distortions in a speech signal increases cognitive demand and thus listening effort (Schneider & Pichora-Fuller, 2000; Stenfelt & Rönnberg, 2009). Spectral degradation of a speech signal, such as in CI processing or CI simulations, has been shown to tax top-down cognitive processes involved in speech

---

[a] University Medical Center Groningen, the Netherlands
[b] Research School of Behavioural and Cognitive Neurosciences, University of Groningen, the Netherlands
[c] University of Groningen, the Netherlands
Correspondence to Carina Pals: c.pals@umcg.nl

Editor: Sid Bacon
Associate Editor: Eric Healy
Received March 3, 2012
Revision received August 21, 2012
Accepted November 1, 2012
DOI: 10.1044/1092-4388(2012/12-0074)

perception (Başkent, 2010, 2012). Thus, we argue that, especially when the fitting of a CI is less than optimal, interpreting the impoverished signal requires substantial cognitive resources, making listening for CI users effortful.

If minimizing listening effort is to be taken into consideration when fitting CIs, it is essential to have a measure that reliably reflects listening effort. Traditionally, the fit of CIs and the benefit of new processing strategies have been determined using speech intelligibility measures. Baer et al. (1993) have shown, however, that a benefit of processing strategy measured in response times on a verbal task, which they linked to ease of listening, was more pronounced than the benefit expressed in improved intelligibility. This suggests that other measures may be more suitable for reflecting benefits in listening effort. Supporting this idea, Rabbitt (1968, 1991) has shown that a degraded bottom-up auditory signal, while not affecting the ability to repeat each word of a list at the moment it is heard, can have a significant effect on later recall of the words. This performance on the memory task is a measure of working memory load, which can be interpreted to reflect listening effort. Sarampalis et al. (2009) have shown, in normal-hearing participants, that hearing-aid-like noise reduction strategies can result in an improvement in performance on a secondary task, even when no improvement in speech intelligibility is seen. This finding implies that a hearing device feature, such as noise reduction, though it may be deemed not beneficial when assessed only with an intelligibility test, may in fact be beneficial due to a reduction in listening effort. Other studies in hearing aid research also suggest that signal-processing benefits may sometimes be better reflected by tests of listening effort (Lunner, Rudner, & Rönnberg, 2009; Rudner, Foo, Rönnberg, & Lunner, 2009; Sarampalis et al., 2009).

The hypothesis of the current study was that listening effort may change independently from speech intelligibility for different processing settings of the CI, and therefore an advantage in effort may not be accurately reflected by speech intelligibility measures. This hypothesis was tested using speech stimuli, which were generated using a noise-band vocoder to simulate CI processing. The use of simulations allowed intelligibility to be systematically manipulated by changing the spectral resolution (i.e., number of processing channels). Normal-hearing participants listened to the CI-simulated sentences and repeated what they heard, thus providing speech intelligibility data for each level of processing. The variations in listening effort resulting from the different processing conditions were assessed using a dual-task paradigm chosen based on Sarampalis et al. (2009). In a dual-task paradigm, a primary and a secondary task are performed simultaneously. If the tasks are similar, they compete for resources, and an increase in effort associated with the primary task will thus be reflected in decreased performance on the secondary task (Broadbent, 1958; Rabbitt, 1966). For more complex cognitive tasks, dual-task interference is less straightforward. However, simple psychometric tasks, such as an image-judgment task, appear to show most interference when performed simultaneously with such complex tasks (Hegarty, Shah, & Miyake, 2000).

In the current study, the primary task was a speech intelligibility task using the CI-simulated stimuli. The measure chosen to reflect listening effort was the response times (RTs) on a visual secondary task. This choice was based on the argument that one of the central cognitive resources relevant to speech understanding is speed of processing (Kramer, Zekveld, & Houtgast, 2009), and thus a secondary task using this resource will reflect effort associated with the primary speech intelligibility task. The secondary task of choice would need to be affected by effort associated with the speech task, while not affecting performance on the speech task itself. For this reason, we used two different secondary tasks in this experiment, which we expected to show different degrees of interference with the speech task: a rhyme-judgment task (e.g., Baddeley & Salamé, 1986; Wilding & White, 1985) and a simplified, two-dimensional version of the mental-rotation task (Caissie, Vigneau, & Bors, 2009; Hegarty et al., 2000; Hoyek, Collet, Fargier, & Guillot, 2012). Rhyme judgment and mental rotation tap verbal and visuospatial speed of processing, respectively (Heydebrand, Hale, Potts, Gotter, & Skinner, 2007). Research has shown the rhyme-judgment task to be a predictor of speech understanding (Heydebrand et al., 2007; Lunner, 2003), which suggests that this task relies at least partly on the same cognitive resources as speech perception, and thus we expected it to show strong interference with the primary task. The mental-rotation task showed no correlation with speech comprehension (Heydebrand et al., 2007), and for this reason, we expected it to interfere less with the primary task.

In addition to the dual-task paradigm, which was used as an objective measure of effort, listening effort was measured on a subjective multidimensional self-report scale. Although self-report measures of subjective effort are easy to administer, it is not certain whether they reflect the proportional demand on cognitive resources (Wickens, 1992). Objective measures of effort, such as the dual-task paradigm, are specifically designed to reflect cognitive demand and may therefore be more sensitive to small changes in listening effort. However, such measures are less practical to use in, for example, a clinical setting. Although both subjective and objective measures are used to quantify listening effort, studies combining both often report no statistical relation between the two (Anderson Gosselin & Gagné, 2011; Feuerstein, 1992; Zekveld et al., 2010), suggesting that objective and subjective measures of listening effort may tap different aspects of listening effort and may be complementary.

Method

Participants

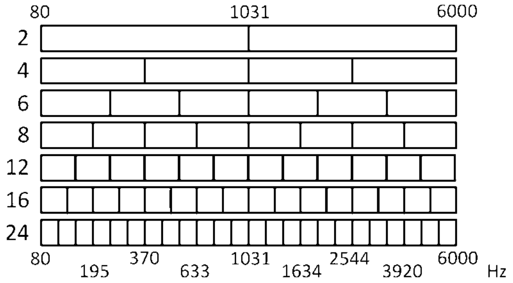
Twenty-three normal-hearing young adults were recruited for participation in this study, four of whom were excluded during data analysis because of missing values in their data sets (either due to problems with the digital voice recorder or inconsistent filling out of the subjective workload questionnaire). The remaining 19 ranged in age from 19 to 25 years (average age about 22 years). Three were male, 16 were female. All participants were native Dutch speakers, and they reported having no dyslexia or other language impairment. Prior knowledge of Japanese or similar scripts was an exclusion criterion based on the stimuli used in one of the secondary tasks. Normal hearing was confirmed by pure-tone thresholds (below 20 dB HL at audiometric frequencies between 250 and 6000 Hz). All participants were given sufficient explanation about the experiment and voluntarily signed an informed consent form prior to data collection and were reimbursed for their time and effort.

Speech Stimuli

The stimuli used for the intelligibility task were full sentences, eight to nine syllables in length, of on average 1.8-s duration. Using sentences rather than single words would allow for a full secondary task trial, from stimulus presentation until response, to be completely contained within the presentation of one auditory stimulus. The sentences of the VU corpus (Vrije Universiteit, Amsterdam, the Netherlands; Versfeld, Daalder, Festen, & Houtgast, 2000), of the female speaker set, were used. These are digitally recorded (sampled at 44.1 kHz) and organized in 39 balanced lists, each list consisting of 13 Dutch sentences. The sentences were processed using a noise-band vocoder (Dudley, 1939), implemented in MATLAB, to simulate CI processing (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). The experimental variable in the listening task was the spectral resolution of the simulated speech, manipulated by using different numbers of spectral channels in the vocoder. Normal-hearing listeners can usually understand CI-simulated speech quite well with four to eight channels (Başkent, 2006; Friesen, Shannon, Başkent, & Wang, 2001). On the basis of these studies, the conditions for the listening task were chosen to cover the range from nearly unintelligible to perfectly intelligible: 2-, 4-, 6-, 8-, 12-, 16-, and 24-channel CI simulations and a control condition using unprocessed speech stimuli. The filter bandwidths and cutoff frequencies varied depending on the number of channels. The bands were chosen such that they corresponded to evenly spaced regions in the cochlea; this was achieved by calculating the –3 dB cutoff frequencies using Greenwood's frequency-to-place mapping formula (Greenwood, 1990). For some examples of –3 dB cutoff frequencies, see Figure 1.

The vocoder processing was implemented as follows: First, the original acoustical signal was separated into a number of spectral bands (the analysis bands) as determined by the experimental condition, using sixth-order Butterworth bandpass filters. From each analysis band, the slow-changing envelope was extracted by means of half-wave rectification and filtering with a third-order low-pass Butterworth filter with –3 dB cutoff frequency of 160 Hz. A set of noise-band carriers (the synthesis bands) was constructed by similarly dividing white noise into spectral bands using sixth-order Butterworth bandpass filters. For this experiment, the center frequencies and bandwidths of the analysis bands were the same as those of the synthesis bands in order to simulate matching frequency-to-place mapping of the CI electrode array (Başkent & Shannon, 2007; Greenwood, 1990). The CI simulations were then constructed by modulating each synthesis band with the envelope extracted from the corresponding analysis band and then adding these modulated noise bands together to form the final stimulus.

Figure 1. Representation of the frequency bands used in the cochlear implant simulations. The vertical axis shows the number of bands, and the horizontal axis shows some of the –3 dB cutoff frequencies (based on the Greenwood [1990] formula).



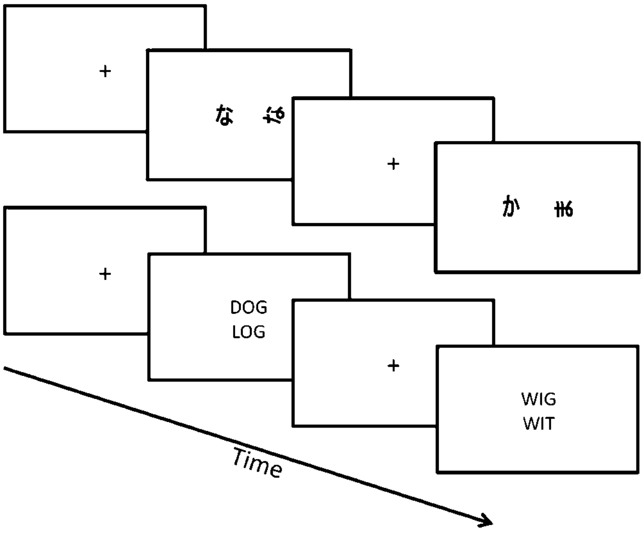
Visual Stimuli

The stimuli for the secondary tasks were rhyme words for one task and Japanese characters for the other. The reason for using Japanese phonetic symbols was to ensure that the stimuli for the mental-rotation task were linguistically meaningless to the participants. The words used in the rhyme-judgment task were monosyllabic Dutch words, vetted for their pronunciation by a native speaker of Dutch. The 75% most frequently occurring words, as determined by the CELEX lexical databases of Dutch (Baayen, Piepenbrock, & Gulikers, 1995), were used in the experiment. The words in the display were clearly visible, one above another, centered on a computer monitor in big, black capital letters on a white background, each letter approximately 7 mm wide and 9 mm high, with a 12-mm vertical white space between the words. The Japanese characters used for the mental-rotation task were taken from the hiragana, one of the two syllabaries in use in Japanese. For those pairs of characters that can easily be mistaken as the same when rotated by 90° (for example: し and つ), only one of the two characters was used. The characters in the display were clearly visible, positioned side by side, and centered on a computer monitor in black on a white background, each character approximately 3 cm wide and 3 cm high, with a 4-cm horizontal white space between the characters, as illustrated in Figure 2.

Equipment

The participants were seated in a soundproof booth, in front of a wall-mounted computer screen at approximately 50-cm distance. A computer program, implemented using the Psychophysics Toolbox Version 3 for MATLAB and run on an Apple Macintosh computer, coordinated the presentation of both the auditory stimuli for the primary task and the visual stimuli for the secondary task. The verbal responses on

Figure 2. Presentation of the visual stimuli for the mental-rotation and rhyme-judgment task (upper and lower panels, respectively).



the primary listening task were recorded for later scoring on a PalmTrack 24-bit digital audio recorder (Alesis, LP). The key-press responses and the RTs on the secondary task were automatically logged by the experimental program. The digital audio stimuli were sent via the AudioFire 4 external soundcard (Echo Digital Audio Corporation) to open-back HD600 headphones (Sennheiser electronic GmbH & Co. KG), to be presented to the participant diotically. The participants were instructed to adjust the volume to a comfortably loud, clearly audible level, within the range of 65–75 dB SPL. The calibration was done using the processed stimuli, measuring root-mean-square sound pressure with integration time constant of 1 s.

Procedure

Listening effort was measured objectively with a dual-task paradigm, consisting of a listening task (primary) and two different visual decision-making tasks (secondary); and subjectively with a multidimensional subjective rating scale.

Listening task. The primary task was designed to measure the participant's intelligibility score for sentences of varying spectral resolution. This task was presented three times for each of the eight levels of spectral resolution: once as a single task and once combined with each secondary task. The presentation order of these 24 conditions was randomized for each participant. Blocks of presentations for the single-task listening conditions consisted of one list of 13 sentences. For the dual-task conditions, no more than one RT measurement could be recorded during the presentation of each sentence. Therefore, each block of presentations included a total of 26 sentences. This way it was possible to gather a sufficient amount of RT data recorded during the presentation of an auditory stimulus for statistical analysis. The interval between the onsets of the sentences was timed 8 s apart, and the average duration of the sentences was approximately 2 s. The intelligibility task was to listen to the processed sentences and repeat out loud what was heard. The participants were encouraged to guess if they were not sure what they heard. Their responses were recorded for offline scoring by a native Dutch speaker. The percentage of correctly identified words served as a measure of intelligibility.

Visual tasks. The visual decision-making tasks were designed to measure RTs, from stimulus onset until a key was pressed by the participant. In the rhyme-judgment task, a randomly chosen pair of words was displayed, one above the other, on the computer monitor. The participant's task was to indicate whether the two words rhymed or not by pressing one of two buttons on the keyboard. In the mental-rotation task, a randomly chosen pair of Japanese characters was displayed side by side on the monitor, one of which was rotated by 90°. The location of the rotated character (left or right) and the direction of the rotation (clockwise or counterclockwise) were randomly determined by MATLAB, with equal probabilities for each possible combination. On this task, participants indicated whether the two characters were the same (except for the rotation) or different by pressing one of the same two buttons used in the rhyme-judgment task.

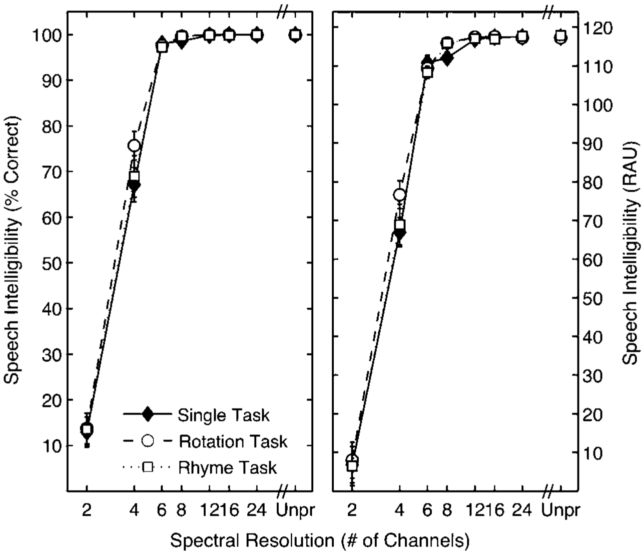
The rest of the procedure was the same for both visual tasks. The stimulus combination was chosen at random, with a 50% chance for a pair that required a "yes" answer. The stimuli were presented until a key was pressed in response or for a maximum of 2.7 s, after which the next trial would begin. Consecutive stimuli were separated by an interstimulus interval during which a fixation cross would appear in the center of the screen for the participants to focus on. The duration of this interval was pseudorandomly varied between 0.5 and 2.0 s, based on a uniform distribution. If no key was pressed, the trial was logged as a "miss." This variation ensured that the participants were unable to predict when a stimulus would appear and that in dual-task conditions the timing of the auditory and visual stimuli varied.

Subjective rating scale. For subjective assessment of listening effort, the NASA Task Load Index (TLX) was used (Hart & Staveland, 1988). The NASA TLX is a multidimensional scale that measures a range of aspects contributing to perceived workload: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). The total score is the weighted mean of the scores from the different dimensions. The weights are determined after the experiment by pairwise comparison. For all possible pairs of dimensions, the participants are asked to indicate which of two contributed most to the overall workload of the tasks. This procedure of weighting the ratings is designed to reduce intersubject variability resulting from differences in individual interpretation of workload and its factors.

Results

The average speech intelligibility scores are depicted in Figure 3. In each panel, the scores are plotted separately for

Figure 3. Average speech intelligibility scores in percentage correct (left panel) and in rationalized arcsine units (RAU; right panel), as a function of spectral resolution. The different lines with open symbols show results for the two dual-task setups, the solid line with filled symbols shows results for the single task.

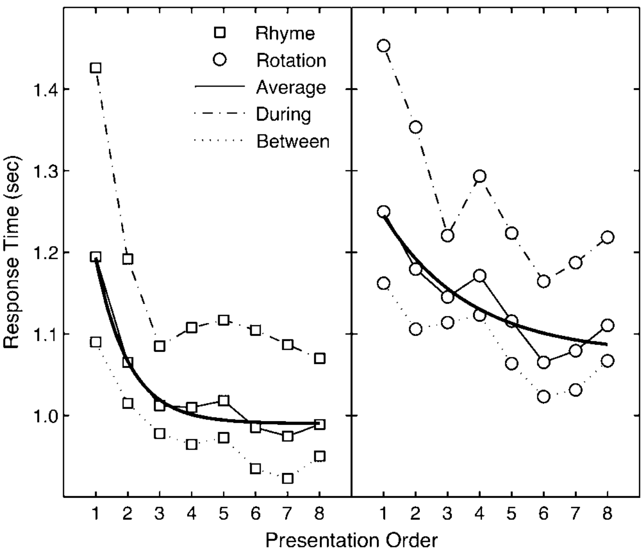


listening task only, listening task combined with rotation task, and listening task combined with rhyme-judgment task. In the left panel of Figure 3, the intelligibility scores are plotted in units of percentage correct, as a function of spectral resolution. In the right panel of Figure 3, the scores are plotted in rationalized arcsine units (RAU; Studebaker, 1985), as a function of spectral resolution. The conversion to RAU was performed to allow a closer examination of the effects near ceiling; RAU scores are easier to interpret because, unlike with proportional scores, the variance is independent of the mean, and thus differences in percentage scores on different parts of the scale (e.g., the difference between 50% and 60% is not comparable to the difference between 90% and 100%) are not comparable, whereas differences in RAU are. Because the maximum possible value in RAU depends on the number of items in a task, the RAU scores were calculated based on an average number of words per list (80 words) and a proportion of words repeated correctly for each task. This ensures that a score of 100% correct always corresponds to the same RAU value, in this case, 117.83. The left panel of Figure 3 shows that, in terms of percentage correct, speech intelligibility appears to reach a plateau at six channels. The right panel shows that there might still be some improvement in intelligibility between six and eight channels. To examine these effects and differences between percentage correct and RAU scores, we carried out further analyses on both sets of scores.

Two repeated measures analyses of variance (ANOVAs) were performed with task and spectral resolution as factors (with three and eight levels, respectively), one on the percentage correct scores and one on the RAU scores. Both ANOVAs showed a significant main effect of spectral resolution—percentage correct, $F(7, 126) = 396.84$, $p < .001$; RAU, $F(7, 126) = 412.89$, $p < .001$—and a significant interaction between task and spectral resolution—percentage correct, $F(14, 252) = 2.11$, $p = .012$; RAU, $F(14, 252) = 1.85$, $p = .033$. A post hoc analysis using Tukey's honestly significant difference test (HSD) indicated that the sole cause of the significant interaction was a significant difference in intelligibility between the mental-rotation task and the single task for the four-channel condition, while there was no significant difference between the single task and the rhyme-judgment task. To confirm that there was no difference in performance between the two dual tasks, we performed two-way ANOVAs for these two tasks over the eight levels of spectral resolution, again for both the percentage correct scores and the RAU scores. These ANOVAs showed no significant interaction between task and processing. From this, we concluded that there was no significant difference between the two dual tasks in terms of speech intelligibility. Therefore, we then grouped the data for the two dual tasks together to examine the differences between listening conditions. As expected, speech intelligibility with two channels was very low, with about 13% of the words identified correctly. Increasing the number of channels to four provided a dramatic improvement in intelligibility; participants scored on average 70% correct. For six channels, speech intelligibility was near perfect, on average 98% correct. The significance of these differences was determined using Tukey's HSD, which showed that there were significant improvements in intelligibility from two to six channels of CI simulations and that further increases in spectral resolution resulted in no significant improvement in intelligibility. This was true for both the percentage correct scores and the RAU scores.

Upon examining the RTs from individual participants, we discovered a reduction over time, regardless of the listening conditions, suggesting training effects. Figure 4 shows the mean dual-task RT data as a function of presentation order, with each visual task in a separate panel. Despite each participant being presented with the listening conditions in a different, randomized order, the fit lines in these figures do show that there were strong learning effects during the course of the experiment. In order to reduce the between-subject variance introduced by these training effects, they were modeled and compensated for using the following procedure. First, an exponential function was fitted to the overall mean RT data for each of the two secondary tasks. The proportion of variance accounted for (R^2) by these fits was 0.975 for the rhyme-judgment task and 0.841 for the mental-rotation task. The horizontal asymptote of the exponential fit line is interpreted as the value on which the RT converges when all training effects have stabilized. For each individual participant, the learning effects were then compensated for by subtracting the deviation from the asymptote predicted by the fit line for each condition on the basis of the order of presentation. These manipulations of the data had no visible effect on the shape of the RT data as a function of spectral resolution. They did, however, considerably reduce

Figure 4. Average response times (RTs) on the rhyme-judgment task (left panel) and the mental-rotation task (right panel) as a function of presentation order. The training effect is shown by decreasing RTs, over the course of the experiment. The solid lines show the exponential fits to the average RT data for each of the two tasks.

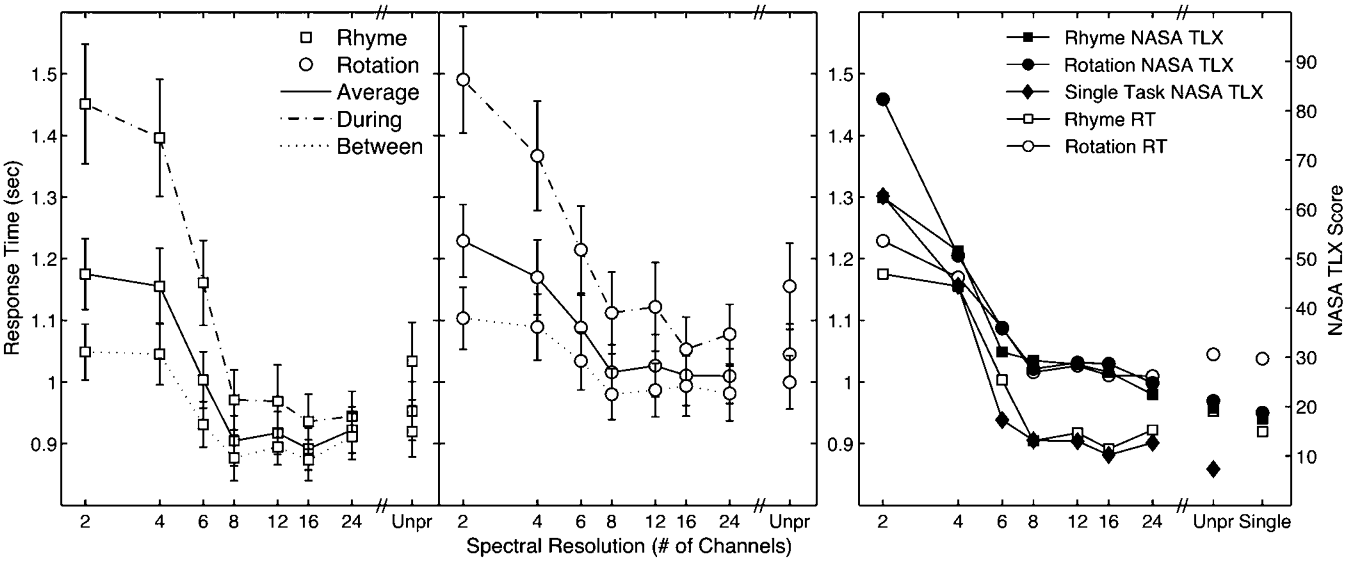


the variance; for the raw RT data, standard deviations of the mean RTs per level of spectral resolution ranged from 0.35 to 0.63 s; compensating for learning effect reduced the standard deviations to values between 0.15 and 0.27 s. Further analyses of the RTs were performed on the data corrected for learning effects.

The first two panels of Figure 5 show the RT data (adjusted for learning effects) for the rhyme-judgment task and the mental-rotation task. Because wrong answers may be the result of a strategy or accidental button press, these RTs may be unrealistically short and thus distort the data. Therefore, RTs for trials on which a wrong answer was given were left out of the analysis. In both panels, the mean RTs are shown for trials presented during listening (dashed line) and for trials presented between the auditory stimuli (dotted line). The overall mean RTs, recorded during and between auditory stimuli grouped together, are represented by the solid line. These two plots show that the reduction in RTs from two to eight channels was greater for trials presented during a sentence than for trials presented in between. A three-way repeated measures ANOVA, with the factors task, visual stimulus timing, and spectral resolution, indicated that the interaction between stimulus timing and processing was indeed significant, $F(7, 126) = 18.37$, $p < .01$. The results of the ANOVA also showed a significant main effect of processing on RT, $F(7, 126) = 28.78$, $p < .01$. Comparison between mean RTs for consecutive processing conditions using Tukey's HSD showed a significant decrease in RT for listening conditions up to eight channels. Interesting to note is that there were significant RT improvements even for conditions in which speech intelligibility had reached a plateau.

The third panel in Figure 5 shows the mean NASA TLX scores for the listening task performed alone and with both secondary tasks. Both dual tasks were consistently judged more effortful than the single-task conditions. It can

Figure 5. The first two panels show the average RTs on the two secondary tasks (solid lines) as a function of spectral resolution, as well as the RTs split into those recorded during sentence presentation (dashed line) and those between sentences (dotted line). The third panel shows the NASA Task Load Index (NASA TLX) scores (filled symbols) for all three tasks plotted together with the average RTs (open symbols), as a function of spectral resolution.



also be seen from the figure that the NASA TLX scores, like the RTs, decrease as the number of channels CI simulations increases, at least for low numbers of channels. However, the decrease in effort between six- and eight-channel CI simulations for NASA TLX scores was not significant. A two-way ANOVA of the NASA TLX scores showed significant main effects of spectral resolution as well as task, which we attribute to the difference between the single and dual task. An ANOVA performed only on the NASA TLX data for the two dual tasks did not show a significant effect of task. Because there was no significant difference between the two dual tasks, we averaged the mean RTs over both dual tasks to further examine them. The mean RTs for consecutive listening conditions were compared using Tukey's HSD, which showed a significant decrease in NASA TLX scores for listening conditions up to six channels.

Discussion

The hypothesis of the present study was that, with different settings of CI processing, listening effort may change differently from speech intelligibility. Furthermore, the conventional speech intelligibility tests may not be sufficient to capture these effects accurately. To explore this hypothesis, listening effort was assessed using an objective and a subjective measure. These measures were then compared with speech intelligibility scores to examine whether they were sensitive to differences in listening effort between conditions in which no improvement in intelligibility was seen. The participants were presented with speech stimuli processed to simulate CI output with different levels of spectral resolution and asked to repeat back what they heard. This was the primary task of the dual-task paradigm used, and this task resulted in speech intelligibility scores for each level of spectral resolution. Two different secondary tasks—a rhyme-judgment task and a mental-rotation task—served to provide an objective measure of the listening effort associated with each level of processing, in the form of RTs on a visual decision-making task performed simultaneously with the intelligibility task. In addition to this, a multidimensional workload questionnaire was administered after each task to serve as a subjective measure of listening effort.

The results of the primary speech intelligibility task, in line with the findings of previous research (e.g., Başkent, 2006; Friesen et al., 2001), showed an increase in intelligibility with increased spectral resolution. The present study closely reproduced the speech intelligibility results reported by Friesen et al. (2001) for normal-hearing listeners presented with CI-simulated English sentences in quiet. In both studies, a marked increase in intelligibility was observed between two- and six-channel CI simulation. Intelligibility appeared to reach about 98% for six spectral channels, and further increases in spectral resolution produced no significant improvement.

The main interest of the present study was listening effort. The results from the subjective workload questionnaire, the NASA TLX, showed consistently higher scores for dual task compared with single task. This is not surprising because the NASA TLX is designed to measure overall task load, and a dual task can be considered to be cognitively more demanding than a listening task alone (Wickens, 2008). For all three tasks (the single listening task, the rhyme-judgment dual task, and the mental-rotation dual task) the NASA TLX showed a significant decrease in workload from two to six spectral channels. For an increased number of spectral channels beyond six, no significant decrease in subjective workload was found. The results of the two visual decision-making dual tasks showed that, although both speech intelligibility and NASA TLX scores improved only from two up to six spectral channels, RTs on both secondary tasks improved significantly from two up to eight channels. In other words, the RT measures captured an improvement, or benefit, of increasing spectral resolution from six to eight spectral channels that the intelligibility task and the NASA TLX did not capture.

One can argue whether the benefit captured by a decrease in RTs is indeed due to reduced listening effort. Recall that what we call "listening effort" is the proportion of a shared, limited cognitive resource that is allocated to the listening task. The larger the effort, the larger is this proportion assigned to the listening task, and thus the less of this resource is available to perform another task simultaneously. The RTs recorded between presentations of auditory stimuli showed a significantly shallower effect of spectral resolution than the ones recorded during presentation of auditory stimuli (Figure 5). This observation supports the idea that the effects of simulated CI processing present in the RT data are indeed caused by changes in demands on shared resources due to these differences in processing. In short, the observed pattern suggests that the reduced RTs are caused by a decrease in listening effort associated with the increase in spectral resolution. The literature shows that effects of effort are rather elusive, and effects in RT, though significant, can be as small as about 50 ms (e.g., Baer et al., 1993; Sarampalis et al., 2009). Although the results show a significant effect in RT only for conditions with constant intelligibility between six and eight channels, this effect was observed for both secondary tasks; therefore, we are convinced that it is a persistent and repeatable effect. Although both intelligibility and subjective workload measures are likely to reflect changes in listening effort to some degree, as these two measures showed a pattern similar to that of the RT measures, they appear to be less sensitive to changes in listening effort; they showed no significant improvement between six and eight spectral channels, whereas the RT measures did.

The NASA TLX scores do not show the same sensitivity to changes in listening effort as the RT measures. This difference in sensitivity between the NASA TLX and the RT measures can be explained in two ways. As mentioned in the beginning of this article, several studies combining objective and subjective measures have reported no statistical relation between the two (Anderson Gosselin & Gagné, 2011; Feuerstein, 1992; Zekveld et al., 2010). Anderson Gosselin and Gagné (2011) suggested that these different types of measures reflect different aspects of listening effort. They referred to the distinction between "effort" and "ease" made by Feuerstein (1992) and suggested that whereas performance on the secondary tasks reflect effort, a subjective self-report measure reflects ease. Another possible explanation attributes this difference to the "performance" dimension in the NASA TLX. Rubio, Diaz, Martin, and Puente (2004) compared different subjective workload measures and concluded that, of the three measures they compared, the NASA TLX showed the highest correlation with performance. This could explain why the NASA TLX results in the present study follow the intelligibility results more closely and, like the intelligibility measures, are less sensitive to changes in listening effort.

In the present study, the rationale for using two different visual RT tasks, one linguistic in nature and one purely visual, was based on the hypothesis that these two types of secondary tasks tap different aspects of working memory, the phonological loop and the visuospatial sketchpad (Baddeley, 2012; Heydebrand et al., 2007), and thus might be affected differently by the primary intelligibility task. We originally expected that this could result in different patterns of the RT outcomes for the two tasks as a function of spectral resolution or differences in interaction with the primary speech perception task. Against our expectation, the patterns of average RTs for the two secondary tasks looked very similar, and there was indeed no significant interaction between the type of secondary task and spectral resolution. Furthermore, neither task affected the performance on the primary task. One possible explanation for these similarities between the two tasks could be that, because of the nature of the Dutch language, most rhyming word pairs were orthographically similar, whereas most nonrhyming pairs were dissimilar. Therefore, although we assumed the task to be purely linguistic, it was possible for the participants to adopt a visual strategy. Alternatively, mental rotation is such a complex operation that it is not limited to the visual modality but rather requires central processing as well (Ruthruff, Miller, & Lachmann, 1995). Thus, even though the task used in this study was a simplified version of the classical mental-rotation task, it could well be affected by a concurrent task in a different modality—such as a listening task. Regardless of the nature of the secondary task, both versions showed effects of listening effort where speech intelligibility scores and subjective effort scores did not.

Overall, we take the findings of the present study to mean that decreased spectral resolution, as manipulated by reducing the number of vocoder channels in CI simulations, results in increased listening effort, which is reflected in longer RTs on a secondary task. Supporting our observations, Lindenberger and Baltes (1994) hypothesized that, in a manner similar to the interference between tasks in a dual-task paradigm, interpreting degraded sensory input may require an increased allocation of cognitive resources, leaving fewer resources available for other cognitive tasks at hand. Schneider and Pichora-Fuller (2000) referred to this as the "information degradation hypothesis." Further support for such coupling between degraded speech and the increased cognitive resources needed for its processing was presented by Pichora-Fuller, Schneider, and Daneman (1995), who have shown that effects of age on cognitive performance can, at least partially, be explained by a decrease in sensory function; older listeners were found to have more trouble recalling lists of spoken items, whereas for both young and old listeners, decreasing the signal-to-noise ratio of auditory stimuli reduced their ability to store the items in memory. This finding suggests that a reduction in signal quality increases cognitive demand similarly in both young and old listeners. Two more recent studies showed increased cognitive demand as a result of decreased spectral resolution with both CI simulations and CI users (Başkent, 2012; Chatterjee et al., 2010).

In short, auditory processing, working memory, and speed of processing seem to interactively affect both speech understanding and the resources available for additional tasks (Lunner, 2003). In this light, changes in the effort needed to interpret the auditory signal can be reflected in both measures of working memory performance and speed of processing, such as the RTs on a secondary task used in this study. In their study, Sarampalis et al. (2009) showed a benefit of noise reduction strategies reflected both in better working memory performance and faster RTs, even for conditions in which noise reduction provides no benefit in speech intelligibility. The current study shows similar results: A significant decrease in RTs was found as the number of channels for CI simulations increased from six to eight, although this produced no significant increase in intelligibility.

To summarize, the present study used a dual-task paradigm in which normal-hearing participants were asked to simultaneously perform a speech intelligibility task using CI-simulated speech stimuli with different numbers of spectral channels and a visual RT task. The results showed that RTs decreased with an increasing number of channels, even for some conditions that showed no more improvement in speech intelligibility. This finding suggests that it is possible to further improve the listening experience for CI users, even when no improvement is observed in speech intelligibility. Currently, there is no clinical test that can show such benefits of different programs.

This line of research will help identify processing features and strategies for improving listening effort for CI users and help develop a method for measuring listening effort in a clinical setting to assist in improving CI fitting to optimize listening effort. Considering that a large proportion of Dutch CI users report increased listening effort with a CI compared with preimplantation (van Hardeveld, 2010), such optimization would be beneficial to a large population. The dual-task paradigm used in this study is not yet suitable for measuring listening effort in one individual, because of large individual variance and training effects, and is thus not suitable for use in clinical settings. However, it has proven to be sensitive enough to show effects of listening effort across a group of participants and hence presents a useful method that can be used in research settings, such as in developing new signal-processing algorithms.

Acknowledgments

This research was supported by Cochlear Ltd, Dorhout Mees, Stichting Steun Gehoorgestoorde Kind, the Heinsius Houbolt

Foundation, a Rosalind Franklin Fellowship from the University of Groningen, the Netherlands Organization for Scientific Research (NWO, VIDI Grant 016.096.397), and the Cochlear Implant Center Northern Netherlands and is part of the research program Communication Through Hearing and Speech.

We gratefully acknowledge Thomas Lunner for his insightful suggestions concerning this research; Thomas Stainsby and Ruben Benard for their comments on earlier versions of this article; and Frits Leemhuis, Annemieke ter Harmsel, Nico Leenstra, and Marije Sleurink for their help in seeing this project through.

References

Anderson Gosselin, P., & Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research, 54,* 944–958. doi:10.1044/1092-4388(2010/10-0069)

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation: Advances in Research and Theory, 8,* 47–89. doi:10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Salamé, P. (1986). The unattended speech effect: perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 525–529.

Baer, T., Moore, B. C. J., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *Journal of Rehabilitation Research and Development, 30,* 49–72.

Başkent, D. (2006). Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *The Journal of the Acoustical Society of America, 120*(5, Pt. 1), 2908–2925. doi:10.1121/1.2354017

Başkent, D. (2010). Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores. *The Journal of the Acoustical Society of America, 128,* EL169–EL174. doi:10.1121/1.3475794

Başkent, D. (2012). Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation. *Journal of the Association for Research in Otolaryngology, 13,* 683–692. doi:10.1007/s10162-012-0334-3

Başkent, D., & Shannon, R. V. (2007). Combined effects of frequency compression-expansion and shift on speech recognition. *Ear and Hearing, 28,* 277–289. doi:10.1097/AUD.0b013e318050d398

Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY: Pergamon Press. doi:10.1037/10037-000

Caissie, A. F., Vigneau, F., & Bors, D. A. (2009). What does the Mental Rotation Test measure? An analysis of item difficulty and item characteristics. *Open Psychology Journal, 2*(1), 94–102. doi:10.2174/1874350100902010094

Chatterjee, M., Peredo, F., Nelson, D., & Başkent, D. (2010). Recognition of interrupted sentences under conditions of spectral degradation. *The Journal of the Acoustical Society of America, 127*(2), EL37–EL41. doi:10.1121/1.3284544

Dudley, H. (1939). The automatic synthesis of speech. *Proceedings of the National Academy of Sciences of the United States of America, 25,* 377–383. doi:10.1073/pnas.25.7.377

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

Fishman, K. E., Shannon, R. V., & Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *Journal of Speech, Language, and Hearing Research, 40,* 1201–1215.

Friesen, L. M., Shannon, R. V., Başkent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110,* 1150–1163. doi:10.1121/1.1381538

Fu, Q.-J. (2002). Temporal processing and speech recognition in cochlear implant users. *NeuroReport, 13,* 1635–1639.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87,* 2592–2605.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload, 1,* 139–183. Available at http://humansystems.arc.nasa.gov/groups/TLX/downloads/NASA-TLXChapter.pdf.

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition, 28,* 376–385.

Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. (2007). Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiology & Neuro-Otology, 12,* 254–264. doi:10.1159/000101473

Hoyek, N., Collet, C., Fargier, P., & Guillot, A. (2012). The use of the Vandenberg and Kuse Mental Rotation Test in children. *Journal of Individual Differences, 33,* 62–67. doi:10.1027/1614-0001/a000063

Kahneman, D. (1973). *Attention and effort: Measurement*. Englewood Cliffs, NJ: Prentice-Hall.

Kramer, S. E., Zekveld, A. A., & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: The value of using the Text Reception Threshold Test as a visual equivalent of the SRT test. *Scandinavian Journal of Psychology, 50,* 507–515. doi:10.1111/j.1467-9450.2009.00747.x

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging, 9,* 339–355.

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology, 42*(Suppl. 1), S49–S58. doi:10.3109/14992020309074624

Lunner, T., Rudner, M., & Rönnberg, J. (2009). Cognition and hearing aids. *Scandinavian Journal of Psychology, 50,* 395–403. doi:10.1111/j.1467-9450.2009.00742.x

Manrique, M., Cervera-Paz, F. J., Huarte, A., Perez, N., Molina, M., & García-Tapia, R. (1999). Cerebral auditory plasticity and cochlear implants. *International Journal of Pediatric Otorhinolaryngology, 49*(Suppl. 1), S193–S197. doi:10.1016/S0165-5876(99)00159-7

Pfingst, B. E., Zwolan, T. A., & Holloway, L. A. (1997). Effects of stimulus configuration on psychophysical operating levels and on speech recognition with cochlear implants. *Hearing Research, 112,* 247–260. doi:10.1016/S0378-5955(97)00122-6

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97,* 593–608.

Rabbitt, P. M. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science, 6,* 383–384.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology, 20,* 241–248. doi:10.1080/14640746808400158

Rabbitt, P. M. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica, 476*(Suppl.), 167–176. doi:10.3109/00016489109127274

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Applied Psychology, 53,* 61–86. doi:10.1111/j.1464-0597.2004.00161.x

Rudner, M., Foo, C., Rönnberg, J., & Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scandinavian Journal of Psychology, 50,* 405–418. doi:10.1111/j.1467-9450.2009.00745.x

Ruthruff, E., Miller, J., & Lachmann, T. (1995). Does mental rotation require central mechanisms? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 552–570. doi:10.1037/0096-1523.21.3.552

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research, 52,* 1230–1240. doi:10.1044/1092-4388(2009/08-0111)

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 155–219). Mahwah, NJ: Erlbaum.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995, October 13). Speech recognition with primarily temporal cues. *Science, 270,* 303–304. doi:10.1126/science.270.5234.303

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., . . . Beiter, A. L. (1994). Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System. *American Journal of Otology, 15*(Suppl. 2), 15–27.

Stenfelt, S., & Rönnberg, J. (2009). The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scandinavian Journal of Psychology, 50,* 385–393. doi:10.1111/j.1467-9450.2009.00748.x

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28,* 455–462.

van Hardeveld, R. (2010). *Het belang van Cochleaire Implantatie voor gehoorbeperkten: Resultaten van een enquete gehouden in 2010* [The importance of cochlear implantation for the hearing impaired: Results of a survey held in 2010]. Houten, the Netherlands: NVVS—Commissie Cochleaire Implantatie. Retrieved from www.nvvs.nl/CmsData/Images/CI/resultatenenquete_CI_2010.pdf

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America, 107,* 1671–1684. doi:10.1121/1.428451

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York, NY: HarperCollins.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50,* 449–455. doi:10.1518/001872008X288394

Wilding, J., & White, W. (1985). Impairment of rhyme judgments by silent and overt articulatory suppression. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 37A,* 95–107. doi:10.1080/14640748508400953

Wilson, B. S., Finley, C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991, July 18). Better speech recognition with cochlear implants. *Nature, 352,* 236–238. doi:10.1038/352236a0

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing, 31,* 480–490. doi:10.1097/AUD.0b013e3181d4f251