

University of Groningen

## Effect of Spectral Channels on Speech Recognition, Comprehension, and Listening Effort in Cochlear-Implant Users

Pals, Carina; Sarampalis, Anastasios; Beynon, Andy; Stainsby, Thomas; Başkent, Deniz

*Published in:*  
Trends in hearing

*DOI:*  
[10.1177/2331216520904617](https://doi.org/10.1177/2331216520904617)

**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:*  
2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Pals, C., Sarampalis, A., Beynon, A., Stainsby, T., & Başkent, D. (2020). Effect of Spectral Channels on Speech Recognition, Comprehension, and Listening Effort in Cochlear-Implant Users. *Trends in hearing*, 24, 1-15. <https://doi.org/10.1177/2331216520904617>

### 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.

# Effect of Spectral Channels on Speech Recognition, Comprehension, and Listening Effort in Cochlear-Implant Users

Trends in Hearing  
Volume 24: 1–15  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2331216520904617  
journals.sagepub.com/home/tia  


Carina Pals<sup>1,2</sup> , Anastasios Sarampalis<sup>3</sup>, Andy Beynon<sup>4</sup> ,  
Thomas Stainsby<sup>5</sup>, and Deniz Başkent<sup>1,2</sup>

## Abstract

In favorable listening conditions, cochlear-implant (CI) users can reach high speech recognition scores with as little as seven active electrodes. Here, we hypothesized that even when speech recognition is high, additional spectral channels may still benefit other aspects of speech perception, such as comprehension and listening effort. Twenty-five adult, postlingually deafened CI users, selected from two Dutch implant centers for high clinical word identification scores, participated in two experiments. Experimental conditions were created by varying the number of active electrodes of the CIs between 7 and 15. In Experiment 1, response times (RTs) on the secondary task in a dual-task paradigm were used as an indirect measure of listening effort, and in Experiment 2, sentence verification task (SVT) accuracy and RTs were used to measure speech comprehension and listening effort, respectively. Speech recognition was near ceiling for all conditions tested, as intended by the design. However, the dual-task paradigm failed to show the hypothesized decrease in RTs with increasing spectral channels. The SVT did show a systematic improvement in both speech comprehension and response speed across all conditions. In conclusion, the SVT revealed additional benefits in both speech comprehension and listening effort for conditions in which high speech recognition was already achieved. Hence, adding spectral channels may provide benefits for CI listeners that may not be reflected by traditional speech tests. The SVT is a relatively simple task that is easy to implement and may therefore be a good candidate for identifying such additional benefits in research or clinical settings.

## Keywords

cochlear implants, speech perception, cognition

Received 25 January 2019; Revised 29 November 2019; accepted 17 December 2019

## Introduction

Everyday verbal communication requires the listener to perceive, comprehend, and reason about the message conveyed by the speaker before responding. Successful speech comprehension involves perceptual and cognitive processing, as well as the appropriate allocation of attentional and processing resources (effort), especially when the acoustic speech signal is compromised (Wingfield & Tun, 2007). In ideal listening conditions, speech is perceived clearly and comprehension is nearly effortless (Mattys, Davis, Bradlow, & Scott, 2012; Wild et al., 2012). In nonideal listening conditions, however, degradations of the speech signal limit the effectiveness of bottom-up perceptual processes, increasing reliance on top-down cognitive processes for compensation (e.g., Başkent, Clarke, et al., 2016; Broadbent, 1958; Downs & Crum, 1978; Rönnberg, 2003). Degraded speech

perception can be facilitated by, for example, top-down repair mechanisms to restore interrupted speech (e.g., Bhargava, Gaudrain, & Başkent, 2014; Miller & Licklider, 1950; Samuel, 1981), the use of linguistic

<sup>1</sup>Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, the Netherlands

<sup>2</sup>Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, the Netherlands

<sup>3</sup>Department of Experimental Psychology, University of Groningen, the Netherlands

<sup>4</sup>Department of Otorhinolaryngology, Head and Neck Surgery, Hearing and Implants, Radboud University Medical Centre, Nijmegen, the Netherlands

<sup>5</sup>Cochlear CTC, Mechelen, Belgium

### Corresponding Author:

Carina Pals, University of Utah, Asia Campus, Incheon, South Korea.  
Email: contact@carinapals.com



knowledge (e.g., Benard, Mensink, & Başkent, 2014; Hannemann, Obleser, & Eulitz, 2007), or the use of situational or linguistic context (e.g., Dahan & Tanenhaus, 2004; Sheldon, Pichora-Fuller, & Schneider, 2008; Wingfield, Aberdeen, & Stine, 1991). While the recruitment of higher order cognitive processes can aid, and thus enhance, the comprehension of degraded speech, it may come at the cost of increased cognitive load (e.g., Hornsby, 2013; Pals, Sarampalis, & Başkent, 2013; Wingfield & Tun, 2007; Winn, Edwards, & Litovsky, 2015; Zekveld, Kramer, & Festen, 2010). This may in turn reduce the cognitive resources available for concurrent tasks (Sarampalis, Kalluri, Edwards, & Hafter, 2009), lead to fatigue (Hornsby, 2013), affect the ability to remember the speech (McCoy et al., 2005; Rabbitt, 1966), and lead to slower speech comprehension (Mattys & Wiget, 2011; Wagner, Pals, de Blecourt, Sarampalis, & Başkent, 2016).

For cochlear-implant (CI) users, signal degradation is an everyday occurrence. The quality of the CI-transmitted speech signal is affected by many factors, including, but not limited to, electrode placement, auditory nerve survival, as well as device-related factors such as front-end processing or electrode design (e.g., Başkent, Gaudrain, Tamati, & Wagner, 2016; Blamey et al., 1992). One of the most notable consequences is a severe reduction in spectral resolution as channel interactions limit the effective number of spectral channels (Stickney et al., 2006). The effect of spectral resolution on speech recognition, that is, the ability to repeat back what was heard, has been studied extensively over the decades since the introduction of multichannel CIs (e.g., Eddington, 1980; Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Başkent, & Wang, 2001; Fu, Shannon, & Wang, 1998; Schwartz, Chatterjee, & Gordon-Salant, 2008; Winn, Chatterjee, & Idsardi, 2012). Research has shown, for example, that thresholds for phoneme recognition in noise continue to improve with increasing numbers of active electrodes up to, and possibly beyond, 16 electrodes (Fu et al., 1998), while sentence recognition reaches a plateau around 10 active electrodes in speech-shaped noise (Friesen et al., 2001) and continues to improve beyond 12 electrodes when presented with a competing talker at both low and high signal-to-noise ratios (Croghan, Duran, & Smith, 2017). While earlier research has not been able to show a similar benefit for recognition in quiet beyond 4 to 7 active electrodes (Fishman et al., 1997; Friesen et al., 2001), research with more recently implanted CI users has shown a clear benefit of 16 compared with 8 active electrodes for speech in quiet (Berg et al., 2019). Past research with normal-hearing (NH) listeners has shown that even with speech recognition at or near ceiling, further increasing the number of spectral channels could still further improve indirect measures of

listening effort, such as pupil diameter (Winn et al., 2015), and response times (RTs) on a secondary task in a dual-task paradigm (Pals et al., 2013).

This study aims to investigate the effect of number of spectral channels for CI users on aspects of the listening experience beyond speech recognition (repetition accuracy), specifically, on listening effort and speech comprehension. While traditional sentence recognition tasks assess the listener's ability to simply repeat aloud what was heard, a measure of comprehension assesses the ability to determine the meaning of the sentence (Ralston, Pisoni, Lively, Greene, & Mullennix, 1991; Wingfield et al., 2007). One such measure of comprehension is the sentence verification task (SVT) in which listeners have to determine whether a sentence is true or false, thus forcing them to process the meaning of the sentence. In this study, the same group of CI users participated in two experiments investigating the effect of number of active electrodes: dual-task experiment measuring sentence recognition and secondary-task RTs and a sentence verification experiment measuring comprehension and sentence verification RTs. We hypothesize that increasing the number of active electrodes can benefit speech comprehension and processing speed, our indirect measure of listening effort, even when speech recognition is at a plateau.

In Experiment 1, a dual-task paradigm first designed and used in our earlier study in NH listeners (Pals et al., 2013) is employed to measure speech recognition and secondary-task RTs, interpreted as listening effort, simultaneously. The current dual-task paradigm was successfully used by Pals et al. (2013) in support of the present hypothesis using acoustic simulations in a homogenous group of young adult NH listeners. The question remains, however, whether the method is suitable for use with CI users, especially given that performing the two tasks simultaneously can be challenging for some participants, and a range of different factors can affect performance in CI users (Başkent, Gaudrain, et al., 2016), including effects of age as CI users tend to be older (Bhargava et al., 2014; Bhargava, Gaudrain, & Başkent, 2016).

In Experiment 2, the SVT (Adank & Janse, 2009; Baddeley, Emslie, & Nimmo-Smith, 1992; Baer, Moore, & Gatehouse, 1993; May, Alcock, Robinson, & Mwita, 2001; Pisoni, Manous, & Dedina, 1987; Saxton et al., 2001) is used to measure comprehension and processing speed. While this task has not been previously used with CI users, a version of this task has successfully been applied in previous research to reveal effects of hearing-aid processing on listening effort in elderly (age 60+) hearing-impaired participants (Baer et al., 1993). In the SVT, participants listen to sentences that are either unmistakably true or false/nonsense. The task requires the listener to respond via key press indicating whether the sentence they heard was true or

false/nonsense, producing both accuracy scores and RTs. As an increase in cognitive load leads to slower comprehension (Gibbon, Moore, & Winski, 1997; Mattys & Wiget, 2011; Wagner, Pals, et al., 2016), the sentence verification accuracy and RTs can be interpreted to reflect comprehension and cognitive processing load, that is, listening effort, respectively.

Overall, we hypothesize that reduced spectral resolution in CI users will have a detrimental effect not only on speech understanding but also on listening effort. Crucially, similar to the findings with NH listeners (Pals et al., 2013), we expect that listening effort can be improved further with increasing spectral resolution even when recognition accuracy appears unchanged.

### Experiment 1: Dual-Task Approach: Speech Recognition and Listening Effort

In Experiment 1, to be able to compare our CI user data with our previous noise-band vocoder NH listener data, we used the same dual-task paradigm as our previous study (Pals et al., 2013), with sentence identification as the primary task, and visual rhyme judgment as the secondary task. A few minor modifications were made to the design to accommodate for expected differences in speech recognition and response speed between the young NH participants of the previous study and the adult and elderly CI user participants of this study. Specifically, easier sentence materials were used and the RT-out was longer; these changes, and the rationale behind them, are described in more detail later.

### Methods

**Participants.** Initially, a total of 34 CI users were recruited for participation, 17 through the Audiology Department at the University Medical Center Groningen and 17 through the Audiology Department at the Radboud University Medical Center in Nijmegen. Of the participants recruited in Groningen, three served as pilot participants, two could not come back for the second session due to health reasons, two could not complete the experiment due to a technical problem, and one was unable to follow the test instructions. The data from the remaining nine participants were included in the final analyses. From the participants recruited in Nijmegen, 1 did not return for the second session and the data from the remaining 16 were included in the final analysis. This resulted in a total of 25 participants (14 females, mean age 58 years, range 34–76) who completed the two experiments fully without any problems.

The participants were all native Dutch speaking, post-lingually deafened adults, implanted with the Cochlear Nucleus device and using the CP810 processor. Two participants had been hearing impaired since birth (marked

by asterisk in Table 1); however, all learned their native language in audio-verbal mode. As the goal of this study was to investigate listening effort and comprehension at high levels of speech recognition, only CI users with high clinical speech test scores were chosen. Inclusion criteria were clinical consonant-nucleus-consonant word recognition scores of 80% or higher, a minimum of 1-year experience with CI use, and no known cognitive disabilities. All participants had normal, or corrected-to-normal, vision. All-but-one of the participants had complete intracochlear electrode array insertion, and all were fitted with at least 15 active electrodes in their daily speech processor maps. All-but-two of the participants used the perimodiolar CI24RE Contour Advanced electrode array, and all-but-three used the ACE coding strategy. Demographic and hearing-related information for these participants is summarized in Table 1. This and the subsequent experiments were approved by the local ethical committee (University Medical Center Groningen, Medisch Ethische Toetsing commissie, dossier number METc2010.328).

**Speech stimuli for the primary task.** In our previous study with NH participants, we used sentences from the VU corpus (Vrije Universiteit; Versfeld, Daalder, Festen, & Houtgast, 2000). These materials are carefully prepared to consist of complete, grammatically correct, and semantically neutral sentences reflective of everyday communication and spoken at normal conversational speed. However, as the sentences for this corpus are selected mostly from digitized newspaper articles, they can be relatively difficult for CI users to interpret, especially at conversational speed. Our own earlier research had indeed indicated that even CI users selected for high clinical speech test scores could still show relatively poor sentence understanding for the VU corpus speech materials (Bhargava et al., 2014, 2016). In this study, it was essential that sentence recognition by CI users was high. The speech stimuli for the primary speech recognition task were therefore taken from a different speech corpus, namely, the Leuven intelligibility sentences test (LIST) corpus (Van Wieringen & Wouters, 2008). This corpus is specifically optimized to provide speech reception thresholds for Dutch and Flemish hearing-impaired listeners and CI users in quiet and in noise: The sentences are clearly enunciated and spoken at a slower speed. The corpus consists of 35 lists of 10 everyday conversational Dutch sentences, each spoken by the same female speaker. The lists are balanced for equal difficulty. The total number of syllables in each list of 10 sentences is 90. The lists are structured such that the first sentence is short (between 4 and 6 syllables), and each consecutive sentence is one or two syllables longer than the previous one, ending with a long sentence (between 12 and 15 syllables).

**Table 1.** Summary of the CI Participants' Demographic and Hearing-Related Information.

Participant ID	Gender	Age at experiment (years)	Age of HL (years)	CI use (years)	Etiology	Electrode array	Coding strategy
304	M	38	3	2.3	Usher	CI24RE CA	MP3000
307	M	64	46	1	Progressive	CI24RE CA	ACE
310 <sup>a</sup>	F	54	49	5	Wegener	CI24RE CA	ACE
311	M	59	31	2	Meningitis	CI24RE CA	ACE
313 <sup>b</sup>	M	60	0	7	Mother rubella	CI24RE CA	ACE
314	M	51	7	12	Osteoporosis	CI24R k	ACE
315	F	69	33	7	Progressive	CI24RE CA	ACE
316	F	41	6	2	Hereditary	CI24RE CA	MP3000
317	M	76	10	2	Otitis media	CI24RE CA	MP3000
321	F	51	10	8	Progressive	CI24RE CA	ACE
322	F	59	54	2	Schwannoma	CI24RE CA	ACE
323	F	67	38	7	Stapedectomy	CI24RE CA	ACE
324	F	66	38	3	Progressive	CI24RE CA	ACE
325	F	52	26	2	Progressive	CI24RE CA	ACE
326	M	62	38	3	Progressive	CI24RE CA	ACE
327	F	70	14	4	Progressive	CI24RE CA	ACE
328	M	34	65	4	Progressive	CI24RE CA	ACE
329	M	65	16	7	Progressive	CI24RE CA	ACE
330	F	58	48	6	Progressive	CI24RE CA	ACE
331	M	67	43	3	Progressive	CI24RE CA	ACE
332	F	59	58	7	Progressive	CI24RE CA	ACE
333	M	65	40	4	Progressive	CI24RE CA	ACE
334	F	58	34	4	Otosclerosis	CI24RE CA	ACE
335 <sup>b</sup>	F	49	0	17	Hereditary	CI24M	ACE
336	F	62	30	3	Ototoxicity	CI24RE CA	ACE

Note. M = male; F = female; HL = hearing loss; CI = cochlear implant.

<sup>a</sup>CI user who did not have a fully inserted electrode array.

<sup>b</sup>CI users who were hearing impaired since birth.

*Visual stimuli for the secondary task.* The visual stimuli for the secondary rhyme-judgment task were monosyllabic Dutch words. The lists of words used in this experiment were compiled by Pals et al. (2013) and consist of rhyme words for several word endings for each of the five basic Dutch vowels (a, e, i, u, and o). Each word list was examined by a native Dutch speaker, and words with multiple possible pronunciations, as well as the 25 least common words according to the CELEX lexical database of Dutch (Baayen, Piepenbrock, & van Rijn, 1993) were excluded (Pals et al., 2013). In the experiment, the words were presented one above the other in black capital letters on a white background on a computer monitor approximately 50 cm in front of the participant. The letters were approximately 9 mm high and 7 mm wide, with 12 mm whitespace between the two words.

*Stimulus presentation and equipment.* The experiment was programmed in MATLAB using Psychtoolbox Version 3 and ran on a Macbook Pro 2010 laptop. The program coordinated the presentation of the speech and visual stimuli and logged the responses and RTs on the secondary task. The verbal responses on the primary speech

task were recorded using a digital audio recorder to be scored later by a native Dutch speaker. The experiment was conducted in a sound-isolated booth. All speech stimuli were presented directly from the experimental computer via personal audio cable to the CI processor, to avoid small differences in residual hearing potentially affecting the outcome. All stimuli were presented at a comfortably loud level, individually determined for each participant at the start of the experiment, using a visual analog scale.

*Experimental conditions.* Experimental maps were created by altering the number of active electrodes of the CI by disabling electrodes and redistributing the frequencies assigned to them to the remaining electrodes. Previous research has shown that on average, CI users' speech recognition performance in quiet reaches a plateau from about seven active electrodes (Fishman et al., 1997; Friesen et al., 2001). A core question of this study was whether changes in listening effort occur when speech recognition no longer improves and therefore the experimental conditions were chosen to cover the range between 7 electrodes and the CI participants'

full arrays (15–22 active electrodes). Specifically, four experimental maps were generated with 7, 9, 11, and 15 active electrodes because these numbers allowed for the active electrodes to be either evenly spaced or distributed in a regularly recurring pattern across a full 22-electrode array (Figure 1). The experimental maps were generated based on the participant’s own preferred map using Cochlear Corp’s Custom Sound Software (version 4.0), and the frequencies were redistributed over the active electrodes as suggested by the software. All other parameters (T and C values, stimulation rate, pulse width, coding strategy) were left unchanged. The participant’s preferred SmartSound features, such as noise reduction, AutoSens, Adaptive Dynamic Range Optimization (ADRO®), and so on, were also left as is.

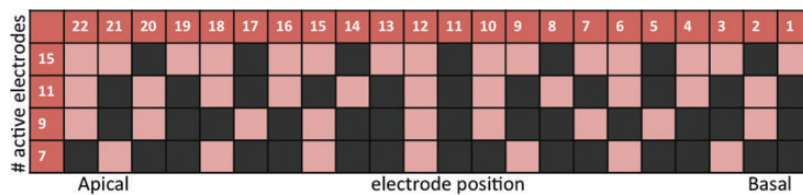
**Procedure.** The experiment consisted of two testing sessions in which the participants performed both experiments, with a 1-month training period in-between. During the 1-month training period between the two sessions, the participants received the experimental processor with the four experimental maps to take home. They were instructed to practice listening with one of the maps for 1 hr on 1 day, rotating to the next map the next day, thus cycling between the four maps every 4 days. This served to familiarize the listener with the experimental maps before the actual testing session, thus minimizing acute effects of new, unfamiliar stimulation patterns and training effects over the course of the experiment. Research has shown that, in the case of spectral mismatch, familiarization occurs relatively fast over the first few days or weeks when the experimental processor is used all day long (Fu, Shannon, & Galvin, 2002). As the reduced number of spectral channels of our experimental programs may negatively impact the CI participants’ listening abilities, for example, at the workplace, we decided instead to limit familiarization to 1 hr a day, but for the relatively long period of 1 month. To verify whether the participants had been practicing with the experimental processor, they were asked a few questions at the start of the second session. The participants were asked about their experiences with the experimental processor, whether they had experienced any difficulties, and whether they had noticed distinct differences between the programs. All participants indicated

that the reduced-channel maps were less pleasant to listen to than their own device, most notably the four-channel map was perceived as harsh and difficult to understand. Some participants expressed that they had experienced difficulty in understanding television with the experimental maps.

The first session lasted 1 hr or less, during which the participants were tested using their preferred map on their own processor to serve as a baseline measurement, while simultaneously the experimental processor was programmed. The second session lasted approximately 2 hr, during which the participants were tested with each of the four experimental maps, in counterbalanced order (in a  $4 \times 4$  balanced Latin-square design).

At the start of the first session, after explaining the procedure and allowing for questions, the presentation level for the speech stimuli was determined, following a method similar to the clinical procedure. A sample sentence was played repeatedly, starting at a very low presentation level and increasing in steps of 2.5 dB. Each time the sentence was presented, the participants were asked to indicate the perceived loudness on a visual scale ranging from *imperceptibly soft* to *uncomfortably loud*. When a comfortably loud level was reached, the stimulus was presented another 3 or 4 times, alternately increasing and decreasing in level by 2.5 dB to confirm that the selected level was loud and clear, yet still comfortable. After this, while the participants performed the experimental tasks with their own processor using their preferred map, the experimental processor was programmed based on this preferred map.

At the start of each session, the procedures of the two tasks were explained and participants performed a 3-min training session for the rhyme-judgment task before starting the actual experiment. Each condition was tested in a series of four task blocks. First, the speech recognition task was presented twice alone (single task), one training block and one experimental block, then the speech recognition task and secondary rhyme-judgment task were presented twice simultaneously (dual task), first a training block and then an experimental block. For each of the experimental conditions, the participants completed the full series of four task blocks before moving on to the next condition.



**Figure 1.** The distribution of active electrodes along the full array is shown for each of the experimental conditions. A light pink square denotes an active electrode, and a dark gray square denotes a deactivated electrode.

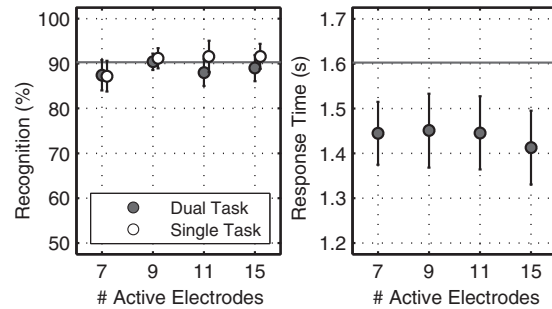
The primary speech recognition task required the participants to listen to the sentences and repeat them out loud, giving their best guess when they were not sure what they heard. When the speech recognition task was presented alone, one list of 10 sentences was used. When presented simultaneously with the secondary task, one list of 10 sentences was used for training and two lists of 10 sentences each were used for the experiment. The sentences varied considerably in duration, unlike the sentences used by Pals et al. (2013), and therefore needed a different strategy for silent interval duration than the study by Pals et al. (2013). The sentences in this study were followed by a silent interval of the duration of the sentence recording plus an additional 2.5 s. This provided the participants sufficient time to repeat the sentence before the next sentence was presented.

In the secondary visual rhyme-judgment task, a pair of words was presented on the screen. The task was to answer as fast as possible whether the word pair rhymed or not, by pressing either “v” for yes or “n” for no on a keyboard. These keys were chosen for their convenient position at the front edge of the keyboard. The word pair was randomly chosen by the MATLAB program, with a 50% chance of a rhyming pair. The stimuli were presented until a key was pressed, or until the time-out of 5 s was reached. The time-out was longer than in our previous study to accommodate the more advanced age of some of the participants of this study. If after these 5 s no key was pressed, this was logged as *unanswered*. After each stimulus, a fixation cross was presented on the screen for a random duration between 0.5 and 2.0 s before moving on to the next word pair.

In the dual task, the participants were instructed to perform the listening task and the rhyme-judgment task simultaneously. Following the design of the previous study, participants were instructed to prioritize the primary listening task over the secondary rhyming task and to respond to the secondary task as fast as possible. Because of the independent timing of the two tasks, secondary rhyme-judgment task trials could occur both during and between the presentations of sentences.

## Results

The left panel of Figure 2 shows the speech recognition accuracy scores for the primary listening task, in percentage of correctly repeated sentences, both for single task (open symbols) and for dual task (filled symbols). The baseline included in the graph reflects the average speech recognition accuracy score when the CI users were tested with their own preferred map using the full electrode array. Because the baseline scores were recorded in the first session of the experiment, and not as part of the actual data collection (i.e., within the counter-balanced test conditions), these were not included as a



**Figure 2.** The left panel shows the speech recognition in percentage sentences correctly repeated, for both single task (open symbols) and dual task (filled symbols), as a function of number of spectral channels. The right panel shows the response times in seconds on the dual-task secondary task. Error bars in both panels denote standard errors. The lines show the average baseline performances for the participants when tested with their own device in the first session of the study.

condition in the analysis. They are shown here as a reference level, as well as to confirm that our CI participants did indeed perform well with their own device and speech recognition performance was indeed near the participants’ performance with their own device for most of the experimental conditions. To verify that speech recognition was indeed at a plateau for all experimental conditions, the speech recognition accuracy scores were analyzed using a two-way repeated-measures analysis of variance (ANOVA) using R and the ez package (version 4.2-2) including the main factors spectral resolution (four levels: 7, 9, 11, 15 active electrodes) and task type (two levels: single or dual task), and presentation order as a covariate. The ANOVA revealed no significant effects of spectral resolution or task type on speech recognition accuracy and no significant interaction.

The right panel of Figure 2 shows the RTs on the secondary rhyme-judgment task in the dual task. For the RTs on the secondary rhyme-judgment task, the number of observations per participant per condition varied depending on the response speed and accuracy. The analysis method of choice for data with different number of observations per cell is linear mixed effect (LME) models. The RTs were analyzed using R and the lme4 package (version 1.1-7, lmerTest-package version 2.0-11). To approximate a normal distribution, the data were log-transformed by taking the natural logarithm of the RTs. The log-transformed RTs (lnRTs) approximated a normal distribution for RTs between 0.35 and 3 s but deviated from normal outside that range. Extremely short and extremely long RTs could have been introduced for a number of reasons, such as accidental button-press or a lapse of attention, that do not necessarily reflect actual processing speed for the task, and therefore RTs below 0.35 and over 3 s were excluded from analysis (5.9% of all trials). Accuracy on the rhyme-judgment task varied

slightly, between 94% and 96%, and only trials with correct responses were included in the analysis of RTs. However, to account for differences in accuracy between participants and conditions, the accuracy scores were included as a factor in the model. Age is known to affect cognitive processing speed (Salthouse, 1996) and has been shown in the past to affect dual-task response latency (Verhaeghen, Steitz, Sliwinski, & Cerella, 2003). Comparing RT data of individual participants, however, did not reveal any correlation with age, and including age as a factor in the model did not improve the fit. The participants' baseline RTs recorded in the first session, on the other hand, did contribute significantly to the fit of the model and were therefore included,  $\chi^2(1) = 36.202$ ,  $p < .001$ .

The final model included the factors spectral resolution, presentation order, accuracy, and baseline RT. A random intercept was included for participant ID, and random slopes and intercepts were included for all within-subject factors. The intercept of the model did not differ significantly from 0 ( $\beta = -0.1194$ , standard error [SE] = 0.0769,  $t = 1.554$ ,  $p = 1.256$ ). The effect of presentation order on lnRT was significant (estimate = 0.0182,  $\beta = 0.0182$ , SE = 0.0083,  $t = -2.208$ ,  $p = .038$ ), that is, RTs for later conditions decreased logarithmically starting with a 16 ms decrease for the second condition ( $e^{(-0.1194 - 0.0182)} - e^{-0.1194} = -.0160$ ). The effect of baseline RT was also significant (estimate = 0.2487,  $\beta = 0.2487$ , SE = 0.0297,  $t = 8.375$ ,  $p < .001$ ), participants with longer baseline RTs also have longer RTs in the experiment overall. The model showed no significant effect of number of active electrodes ( $\beta = -0.0037$ , SE = 0.0022,  $t = -1.670$ ,  $p = .109$ ), or accuracy ( $\beta = -0.0062$ , SE = 0.0064,  $t = -0.971$ ,  $p = .336$ ) on RT.

## Experiment 2: SVT Approach: Speech Comprehension and Listening Effort

In Experiment 1, we used the dual-task paradigm, as it had been previously tested and validated with NH participants listening to noise-band vocoded speech (Pals et al., 2013). The SVT we used in Experiment 2 had not been used with CI-simulated speech before. Therefore, an additional group of NH participants was recruited for Experiment 2 only, to evaluate this specific task as a measure of listening effort in NH listeners and to examine how it reflects the effects of number of spectral channels for NH listeners presented with noise-band vocoded speech.

### Methods

**Participants.** Experiment 2 was performed by two groups of participants: a group of 24 young adult NH listeners and the same 25 CI users who participated in Experiment 1.

Initially, 25 NH listeners were recruited for this experiment, all students of the Psychology Department of the University of Groningen, and they received partial course credit for their participation. One of the participants was excluded because of missing data due to a technical error during the experiment. The remaining 24 participants were all native Dutch speakers and young adults (four males; mean age 21 years, range 19–27). All NH participants had hearing thresholds of 20 dB HL or better at all audiometric frequencies between 250 and 6000 Hz. Exclusion criteria were self-reported dyslexia and other language disabilities.

**Speech stimuli.** The Dutch sentence material used for the SVT was created by Adank and Janse (2009) using the same systematic approach used by Baddeley et al. (1992) to create the English-language material for the speed and capacity of language processing test (Adank & Janse, 2009; Baddeley et al., 1992; Saxton et al., 2001). The corpus created by Adank et al. consists of in total 180 sentences, all spoken at a normal conversational speaking rate by the same male native Dutch speaker. The sentences are all syntactically correct; however, 90 are unarguably true and make sense (e.g., *Tijgers hebben een staart*, Tigers have a tail), and the other 90 are obviously false or nonsense (e.g., *Een aap is een soort vis*, A monkey is a type of fish). All sentences start with the subject noun followed by a predicate. The false sentences were constructed by combining a subject noun with a nonmatching predicate from a different sentence. Due to the nature of the Dutch language, the resolving word for true/false judgment is not always in sentence final position. However, even for sentences that did end in the resolving word, the number of syllables of the resolving words varied. For those sentences not ending in the resolving word, the number of syllables to the end of the sentence was within a similar range as the rest of the sentences. All stimuli were at least three words long (min. 4 syllables), and the longest sentence was eight words long (max. 14 syllables). The RTs were calculated as the time between the onset of the resolving word and the button-press responses.

**Stimulus presentation and equipment.** The experiment was programmed, presented, and logged in the same manner as Experiment 1. For the NH participants, the speech stimuli were presented via an AudioFire 4 external soundcard of Echo Digital Audio Corporation (Santa Barbara, CA, USA) and a DA10 digital-to-analog converter of Lavry Engineering, Inc. (Poulsbo, WA, USA) to the open-back HD600 headphones of Sennheiser electronic GmbH & Co. KG (Wedemark, Germany) at 65 dBA. For the CI users, stimuli were presented in the same way and at the same level as for Experiment 1.



**Experimental conditions.** For the NH listeners, the listening conditions were created by varying the number of bands of noise-band vocoded speech. The auditory stimuli were presented in six conditions; 4-, 6-, 8-, 12-, and 16-band noise-vocoded speech, and an unprocessed baseline condition. This was a subset of the same conditions used in our previous dual-task study (Pals et al., 2013). The noise-band vocoded speech was generated using the method as described by Shannon, Zeng, Kamath, Wygonski, and Ekelid (1995), and in a manner similar to our previous study (Pals et al., 2013). All speech stimuli, including the unprocessed condition, were first band-pass filtered to 80 to 6000 Hz. For each of the vocoder conditions, this frequency range was divided into the desired number of bands such that the bands, from lower to upper  $-3$  dB cut-off frequency, spanned approximately equal distances in the average cochlea according to the Greenwood function (Greenwood, 1990). The speech recording was band-pass filtered into the desired number of analysis bands using sixth-order Butterworth band-pass filters. The noise carriers were generated by filtering white noise into bands using the same band-pass filters. From each of the analysis bands, the envelope was extracted using half-wave rectification and low-pass filtering at 160 Hz using a third-order Butterworth filter. The carrier noise bands were modulated using the envelopes of the corresponding analysis bands and postfiltering using the original band-pass filters, and finally the resulting bands were combined to form the noise-band vocoded speech signal. For the CI users, the experimental conditions of varying spectral resolution were the same as in Experiment 1, described earlier.

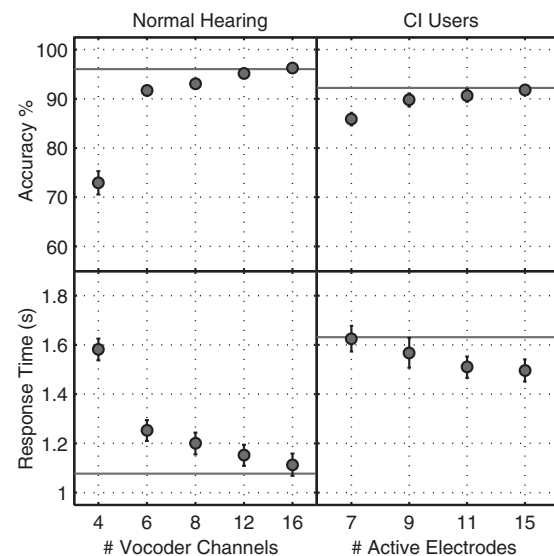
**Procedure.** All NH and CI participants were tested with a similar procedure. They were instructed to listen to one sentence at a time and to indicate whether the sentence was true or false/nonsense by pressing either “v” for true or “n” for false/nonsense. The participants were instructed to respond as accurately and fast as possible. Whether a true or false sentence was played was determined randomly by MATLAB, with a 50% chance for either. The experimental program logged the responses and recorded the RTs from the end of the stimulus to the button-press, following previous procedure described by Adank et al. (2009), therefore negative RTs were possible. If no key was pressed 5 s after start of the sentence, the program logged this as a *miss* and moved on to the next sentence. A silent interval of random duration between 1.5 and 3.0 s was used between the end of the trial and the presentation of the next sentence stimulus.

The NH participants performed Experiment 2 in one session, which lasted approximately 1 hr. The CI users performed Experiment 2 in two sessions, with a 1-month training period in-between, similar to Experiment 1. Session one lasted about 1 hr, and session two about

2 hr, as described previously. They performed Experiments 1 and 2 one after the other in session 1 with their own processor and after the training period in session 2 with the experimental maps on the experimental processor in an interleaved fashion; for each of the 4 experimental maps, the tasks for both Experiments 1 and 2 were performed before moving on to the next map. To minimize any effects of condition order, one half of the participants performed the dual task first, followed by the SVT, and the other half did the opposite. At the start of each session, the task was explained verbally, followed by one training block consisting of 15 sentences for the first session and 10 sentences for the second session. The experimental blocks were presented in counterbalanced order and consisted of 30 sentences each, of which the first 5 sentences were considered training and were not included in the performance score of the task, resulting in 25 sentences per condition.

## Results

**NH listeners.** Figure 3, top-left panel shows the accuracy in percentage correct for the SVT for the NH listeners. The baseline included in the graph reflects the average accuracy using unprocessed speech stimuli. A one-way repeated-measures ANOVA with spectral resolution (4-, 6-, 8-, 12-, and 16-band noise-vocoded speech) as a numerical within-subject factor and covariate task



**Figure 3.** Results of the sentence verification task shown for NH participants (left-side panels) and CI participants (right-side panels). The top panels show accuracy scores in percentage correct and the lower panels show RTs. Error bars show standard error. The baselines included in each figure show the average score for unprocessed speech for NH participants and the average score for the CI users when tested with their own device. CI = cochlear implant.

order revealed a significant effect of spectral resolution,  $F(1, 23) = 36.696, p < .001$ .

To examine the relationship between spectral resolution and accuracy, the results were modeled using a linear model including the within-subject factors spectral resolution (4-, 6-, 8-, 12-, and 16-band noise-vocoded speech) and task order, and a random intercept for participant ID as well as a random slope for spectral resolution per participant ID. Including baseline score did not contribute to the fit of the model,  $\chi^2(1) = 0.4865, p = .4865$ , and was therefore, for the sake of simplicity, not included in the final model.

The final model's intercept, corresponding to the average accuracy (in percentage correct) for four-channel conditions, was estimated at approximately 82% ( $\beta = 82.12, SE = 2.013, t = 40.801, p < .001$ ) and the effect of number of channels at 1.5% ( $\beta = 1.473, SE = 0.184, t = 7.990, p < .001$ ), suggesting a 1.5% increase in accuracy for every additional channel in the vocoded speech. No significant effect of task order was found ( $\beta = 0.007, SE = 0.451, t = 0.014, p = .988$ ).

Because the relationship between the spectral resolution of the noise-band vocoded speech and SVT accuracy scores appears to be linear from six spectral channels up, but with a sharp decrease in accuracy from six to four channels, the results were remodeled excluding the four-channel condition, in order to see whether the effect would still be significant. The new model's intercept was estimated at approximately 92% ( $\beta = 92.50, SE = 1.001, t = 92.436, p < .001$ ) and the effect of number of channels at 0.5% ( $\beta = 0.459, SE = 0.093, t = 4.912, p < .001$ ), suggesting a 0.5% increase in accuracy for every additional channel in the vocoded speech. No significant effect of task order was found ( $\beta = 0.153, SE = 0.215, t = -0.709, p = .48$ ).

The lower left panel of Figure 3 shows the RTs on the SVT for the NH listeners. The RTs approximated a normal distribution between  $-0.1$  and  $2.15$  s, deviating from normal outside that range. Therefore, RTs under  $-0.1$  and above  $2.15$  s were excluded from the analysis. This amounted to 2.7% of the responses. Because only correct responses were included and the longer and very short RTs were excluded, the number of observations varied per participant per condition. The RT data were therefore analyzed using LME models. The best fitting model for the RTs included the factors spectral resolution, presentation order, and baseline RT, as well as random intercepts for participant ID and sentence ID, and random slopes for spectral resolution for both participant ID and sentence ID.

The model's intercept was estimated at 1,076 ms ( $\beta = 1.076, SE = 0.0578, t = 18.616, p < .001$ ) and corresponds to the estimated average difference in RTs compared with baseline for the four-band noise-vocoded speech when presented as the first task of the experiment.

The model showed a significant effect of spectral resolution, estimated at  $-26$  ms ( $\beta = -0.0256, SE = 0.0032, t = -8.066, p < .001$ ) suggesting a 25 ms decrease in RT for each additional spectral channel. The model also revealed a significant effect of baseline RT, estimated at 573 ms ( $\beta = 0.5730, SE = 0.0890, t = 6.436, p < .001$ ), suggesting that participants with longer baseline RTs responded more slowly during the experiment as well (1 s longer baseline RT predicts on average 573 ms longer RTs in the experiment). The effect of presentation order was not significant ( $\beta = -0.0025, SE = 0.0041, t = -0.611, p = .541$ ).

Because the relationship between the spectral resolution of the noise-band vocoded speech and RT on the SVT appears to be linear from six spectral channels up, but with a sharp increase in RTs from six to four channels, the results were remodeled excluding the four-channel condition, in order to see whether the effect would still be significant. The new model's intercept was estimated at 932 ms ( $\beta = 0.9320, SE = 0.0567, t = 16.441, p < .001$ ) and corresponds to the estimated average difference in RTs compared with baseline for the six-band noise-vocoded speech when presented as the first task of the experiment. The model showed a significant effect of number of channels, estimated at  $-12$  ms ( $\beta = -0.0122, SE = 0.0023, t = -5.239, p < .001$ ), and a significant effect of baseline RT, estimated at 604 ms ( $\beta = 0.6042, SE = 0.0917, t = 6.587, p < .001$ ). The effect of presentation order was again not significant ( $\beta = -0.0064, SE = 0.0041, t = -1.545, p = .123$ ).

**CI users.** The top-right panel of Figure 3 shows the accuracy in the SVT with CI users in percentage correct. The baseline reflects the average accuracy recorded in the first session with the full electrode array. A one-way repeated-measures ANOVA with numerical within-subject factor spectral resolution and covariate task order showed a significant effect of spectral resolution on accuracy,  $F(1, 24) = 15.510, p < .001$ . To examine the effect of spectral resolution on accuracy, the results were modeled using a linear model.

The final model included within-subject factors spectral resolution (7, 9, 11, 15 active electrodes) and task order, a random intercept for participant ID as well as random slope for spectral resolution per participant ID. The model estimated the intercept at approximately 85% ( $\beta = 85.402, SE = 0.0197, t = 43.370, p < .001$ ), corresponding with the estimated accuracy for seven active electrodes when presented as the first task of the session. The model showed a significant effect of spectral resolution on accuracy of 0.66% ( $\beta = 0.664, SE = 0.175, t = 3.783, p < .001$ ), suggesting a 0.66% increase in accuracy for each additional active electrode. The effect of task order was not significant ( $\beta = 0.517, SE = 0.446, t = 1.158, p = .251$ ).

The lower right panel of Figure 3 shows the RTs in the SVT with CI users, with the average RT recorded in the first session, with the full electrode array, included as a baseline. Only RTs for correct trials were included in the analysis. The RTs approximated a normal distribution between  $-0.2$  and  $3.2$  s. RTs outside the range  $-0.2$  and  $3.2$  s deviated from the normal distribution and were therefore excluded from the analysis. This amounted to 0.5% of the responses. The best fitting LME model for the RTs included the factors spectral resolution, presentation order, and baseline RT, as well as random intercepts for participant ID and sentence ID, and random slopes for spectral resolution for both participant ID and sentence ID.

The model's intercept was estimated at 1,336 ms ( $\beta = 1.3356$ ,  $SE = 0.1308$ ,  $t = 10.213$ ,  $p < .001$ ) and corresponds with the estimated difference in RT compared with baseline for the seven active electrodes condition when presented as the first task of the experiment. The effect of number of channels was estimated at  $-17$  ms ( $\beta = -0.0170$ ,  $SE = 0.0059$ ,  $t = -2.906$ ,  $p = .007$ ) suggesting a 17 ms decrease in RTs for each additional active electrode. The effect of presentation order was estimated at  $-58$  ms ( $\beta = -0.0584$ ,  $SE = 0.0100$ ,  $t = -5.824$ ,  $p < .001$ ), suggesting a 58 ms decrease in RTs for each consecutive block in the experiment. The effect of baseline RT was estimated at 409 ms ( $\beta = 0.4093$ ,  $SE = 0.1010$ ,  $t = 4.051$ ,  $p < .001$ ).

## Discussion

The goal of this study was to investigate how number of spectral channels affects speech recognition accuracy, speech comprehension, and listening effort for CI users. We hypothesized that for CI users increasing numbers of active electrodes may improve listening effort and speech comprehension, even when speech recognition is already high. This hypothesis was evaluated in two separate experiments: in Experiment 1 using a dual-task paradigm combining a conventional speech identification task and a secondary visual RT task as an indirect measure of listening effort, and in Experiment 2 using an SVT to reflect comprehension and processing speed. The results in brief: Experiment 1 showed no effect of number of active electrodes on secondary task RTs, that is, listening effort, Experiment 2, on the other hand, showed a clear effect on both sentence verification accuracy and RTs for NH as well as CI participants. Each of these findings will be discussed in more detail later.

In Experiment 1, speech recognition was at a plateau, as intended by our design. The effect of spectral resolution on speech recognition has already been studied extensively (e.g., Chatterjee, Peredo, Nelson, & Başkent, 2010; Fishman et al., 1997; Friesen et al., 2001; Fu et al., 1998; Henry, Turner, & Behrens, 2005; Schwartz

et al., 2008; Won, Drennan, & Rubinstein, 2007) and speech recognition measures are regularly used in both clinical and research settings. The main interest of this study was, therefore, investigating potential additional benefits of increased number of spectral channels that are not directly evident from conventional speech recognition measures, such as benefits in listening effort. Contrary to our hypothesis, the secondary task RTs did not decrease further for upward of seven active electrodes, that is, no further improved listening effort. Although our previous study with NH participants did successfully use the same dual-task paradigm to show effects of listening effort when speech recognition is at or near ceiling (Pals et al., 2013), recent dual-task studies with CI users, also report no significant dual-task effects of listening effort either within-subject with and without a directional microphone (Sladen et al., 2018), with or without noise reduction (Purdy et al., 2017), or between groups for unilateral, bilateral, or hybrid CI users (Perreau, Tatge, Irwin, & Corts, 2018). Whether this is due to a lack of effect in CI users, or due to a lack of sensitivity of secondary task measures of listening effort is difficult to distinguish. A recent systematic review of a range of listening effort measures in NH and hearing-impaired participants shows mixed results across studies (Ohlenforst et al., 2017) and suggests that a dual-task measure may not always find effects, where other measures, or even other dual-task paradigms, do.

In this study with CI users, we can identify two important differences with our previous study with NH listeners that could potentially have affected the dual-task results: firstly the larger within-group variability between CI participants due to differences, for example, in age, educational background, hearing ability, and etiology of hearing loss and secondly the speech materials used.

Although the group-average dual-task RTs for the older CI participants were indeed longer (ages: 34 to 76 years; RTs approximately 1.4 s) compared with the young NH adults of our previous study (ages: 19–25 years; RTs approximately 0.9 s; Pals et al., 2013), the variability between the CI user participants was quite large. Individual average RTs ranged from 0.9 s (similar to our young NH listeners) to up to 2.3 s, and these between-participant differences in RT did not appear to correlate with age. Advancing age is generally associated with an overall decline in cognitive abilities that can be attributed to the combined effects of certain neurophysiological and cognitive changes with age, such as a decrease in processing speed (Kail & Salthouse, 1994; Salthouse, 1996), and the moderating effects of, for example, education (for review, see Drag & Bieliauskas, 2010). As each of these contributing factors vary between individuals, the variability in cognitive performance between individuals increases with advancing age. In our specific task, duration of hearing impairment could have

introduced additional across participant variability in rhyme-judgment task performance specifically, as even postlingually deafened adults show lower performance than NH participants on tasks that rely on phonological representations (Lyxell et al., 1996). The lack of correlation between age and RTs might thus be attributed to the wide range of educational backgrounds of our CI participants (Drag & Bieliauskas, 2010) as well as the inherent interindividual variability between CI users due to factors related to the device–nerve interface and etiology of the hearing loss (Başkent, Gaudrain, et al., 2016).

In addition to the between-participant variability, the speech materials used in the current CI user study were different from our previous NH study. For this study, the speech materials used were optimized for hearing-impaired and CI listeners (Van Wieringen & Wouters, 2008): The sentences were everyday conversational Dutch sentences spoken with clear articulation and, most importantly, at a slow speaking rate. Listeners can use the context provided by such everyday sentences to compensate for speech signal degradations (Pichoraffuller, 2008; Saija, Akyürek, Andringa, & Baskent, 2014; Wingfield et al., 1991) and reduce listening effort for the remainder of the sentence (Winn, 2016). However, spectral degradation has been shown to lead to slower speech processing, potentially limiting the ability to use context (Wagner, Pals, et al., 2016), or at least delay the “release from listening effort” (Winn, 2016). This increased processing time may be accommodated by slowing down the speech: Older adults show a remarkable ability to use top-down processes to compensate for degradations in a speech signal, especially for slowed down speech (Saija et al., 2014). The use of the speech materials with a slower speaking rate may have allowed our CI participants the extra time required to utilize their linguistic knowledge and the sentence context and thus have diminished the detrimental effects of the reduced number of spectral channels.

In short, the results of Experiment 1 did not show improvements in secondary task RTs, that is, listening effort, for CI users with increased spectral resolution from seven active electrodes up. However, we have insufficient data to conclude whether this result reflects a general lack of improvement in listening effort or is due to limiting factors of the design.

In Experiment 2, the SVT was used as a measure of comprehension (accuracy) and speed of comprehension (RTs; Adank et al., 2009; Baer et al., 1993). In addition to the CI participants, an extra group of young NH participants was recruited for a validation experiment. A measure of comprehension requires the listener to understand and reason about the meaning of the speech (Ralston et al., 1991; Wingfield et al., 2007), closely reflecting the requirements of everyday verbal communication. In the SVT, the RTs reflect the

processing time required to comprehend the speech and judge whether the sentence was true or false. Pisoni et al. (1987) have successfully shown differences in sentence verification speed for synthetic speech compared with natural speech, equated in speech recognition performance, and attribute the difference in processing speed to differences in cognitive processing requirements. Wagner, Toffanin, and Başkent (2016) show a more direct link between processing speed and effort: They combined an eye-tracking measure of lexical processing speed with pupil dilation measures as an indirect measure of listening effort. Their results showed that a delay in lexical disambiguation for degraded speech was paired with an increase in pupil dilation, suggesting that the delay is due to increased processing load. We argue that these studies and others suggest that increased listening effort results in longer processing time required to understand the speech (Gatehouse & Gordon, 1990; Pals, Sarampalis, van Rijn, & Başkent, 2015) and that the sentence verification RTs can thus be interpreted to reflect listening effort.

The results of Experiment 2 showed improved SVT accuracy scores, that is, improved comprehension, with increasing numbers of spectral channels for both NH listeners and CI users. The traditional speech recognition task that was used in our dual-task paradigm, in contrast, only showed improved speech recognition up to six spectral channels for NH participants (Pals et al., 2013) and was at a plateau for all experimental conditions, seven active electrodes and up, for CI users (Experiment 1). Comprehension is suggested to rely heavily on cognitive capacity (Just & Carpenter, 1992; Ralston et al., 1991). In the SVT, the understanding and reasoning about the heard speech that is needed to judge whether the sentence is true or false, requires further cognitive processing than does the simple repeating what was heard in a speech recognition task. Accuracy on the SVT may therefore be more constrained by cognitive capacity and thus more sensitive to changes in the processing requirements of the degraded speech than traditional speech recognition scores are. However, another possible explanation lies with the difference in speech materials used for both tasks. We will explore this later on in the discussion.

In addition to the improvement in accuracy scores, the SVT showed a clear linear trend of improved RTs with increased number of spectral channels for both NH and CI participants. For the NH listeners, both sentence verification accuracy and RTs improved systematically with increasing numbers of spectral channels, all the way up to 16 channels (see Figure 3). For the CI users, however, the accuracy scores continued to improve up to 15 active electrodes, while the RTs systematically improved up to 11 active electrodes, after which the benefit of additional active electrodes was noticeably smaller (see Figure 3).

The main takeaway from Experiment 2 is that, while the dual task in Experiment 1 failed to show any significant improvement in speech recognition accuracy or secondary task RTs, the SVT in Experiment 2 revealed that further increased numbers of spectral channels could still further improve sentence verification accuracy, that is, comprehension, and RTs, that is, processing speed, both in NH and CI listeners.

While the difference in effects revealed by the dual task compared with the SVT may be due to differences in the tasks, they could also be due to differences in the speech materials used. The speech stimuli used in Experiment 1 were taken from the LIST corpus that is optimized for hearing-impaired and CI listeners (Van Wieringen et al., 2008), chosen to allow the CI participants to achieve near ceiling performance on the primary listening task. In Experiment 2, the sentences were spoken by a native Dutch speaking, young-adult male speaker, speaking at normal conversational speed, and therefore likely more challenging to understand for CI users than the speech materials used in Experiment 1. The difficulty of speech materials has been shown to affect the maximum benefit of increasing spectral channels, that is, at which number of channels speech recognition plateaus (Shannon, Fu, & Galvin, 2004). Speaker style is one specific factor that has been shown to influence speech understanding (Mattys et al., 2012) and might interact with additional challenges such as reduced number of spectral channels. Wingfield, McCoy, Peelle, Tun, and Cox (2006) suggest that effects on speech comprehension become apparent only after a certain threshold of processing difficulty has been crossed and therefore both the nature of the speech material and task can affect the outcome of such tests. Perhaps in Experiment 2, the more challenging speech materials resulted in a stronger effect of spectral resolution on task performance.

However, the difference in results between the dual task and the SVT may also in part, be due to the nature of the tasks themselves. In a previous study (Pals et al., 2015), we found a similar difference in effects shown by the dual-task paradigm and a simple verbal RT measure of listening effort, in an experiment with young adult NH participants listening to speech in various noise conditions. In this previous study, both tasks were performed by the same participants, and using the same speech materials, which were sentences used in a sentence identification in both measures of RTs. The differences in outcomes between those two tasks can therefore not be attributed to differences between the participants, or to differences in speech materials, suggesting that they must stem from differences between the two measures themselves, that is, the difference between a dual-task requiring divided attention, and a single-task RT measure of listening effort while listening to, and repeating, sentences from

the same corpus. In this study, the difference in outcomes between the dual task and the SVT may also, in part, be due to differences in the nature of these two tasks: in this case, the difference between a dual-task paradigm and a single-task SVT. However, in order to tease apart the effects of the task and the speech materials, we would need to perform further experiments comparing these two tasks using the same sentence materials.

Regardless of the reason for the differences between the dual task and the SVT outcomes, the core finding of this study is this: The SVT has shown improved speech comprehension and listening effort in CI users for 7 up to 15 active electrodes, conditions in which the traditional speech recognition measures may show no change when testing in quiet. The same manipulation of spectral resolution in Experiment 1 showed no effect on speech recognition accuracy and listening effort as measured using the dual-task paradigm. Other research also shows a plateau in speech recognition in quiet listening conditions for spectral resolution beyond seven active electrodes in CI users (e.g., Fishman et al., 1997; Friesen et al., 2001), although more recent studies have been able to show improved speech recognition in quiet for 16 compared with 8 active electrodes (Berg et al., 2019). In other words, the SVT has shown a benefit of spectral resolution that may go undetected by the clinical speech recognition tests and can therefore be a valuable measure to complement the traditional speech recognition measures and reveal some of the cognitive processing underlying speech understanding.

In conclusion, spectral resolution does affect speech comprehension and listening effort in CI users. Even in highly idealized listening conditions (speech presented without background noise, through personal audio cable, and in a sound proof room), the SVT showed both improved speech comprehension and listening effort with increased numbers of active electrodes. This finding highlights the benefit of increased spectral resolution for CI users even when this benefit is no longer evident from speech recognition measures as well as the added value of a measure such as the SVT to complement traditional measures of speech recognition to uncover such potential benefits. Our specific dual-task paradigm may not be the method of choice for measuring listening effort in CI users. The SVT shows clear effects of changes in spectral resolution on both speech comprehension and listening effort, the task is easier to explain to participants, easier to perform, and easier to implement than the dual task, making it an attractive method for use in both research and for clinical purposes.

### Acknowledgments

The authors gratefully acknowledge Filiep Vanpoucke for commenting on an earlier version of this manuscript, and Bert Maat, Frits Leemhuis, Emile de Kleine, Sander Ubbink,

Esmee van der Veen, and Maraike Coenen for their help seeing this project through.

### Authors' Note

Preliminary results of this study have been presented as a podium presentation at the Association for Research in Otolaryngology 37th Annual Midwinter Meeting (San Diego, CA, 2014) and are described in one chapter in the PhD thesis "Listening effort: The hidden costs and benefits of cochlear implants" by Carina Pals (2016).

### Data Accessibility Statement

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was financially supported by Cochlear Ltd, Dorhout Mees Stichting, Stichting Steun Gehoorgestoorde Kind, the Heinsius Houbolt Foundation, a Rosalind Franklin Fellowship from the University of Groningen, the Netherlands Organization for Scientific Research (VIDI Grant 016.096.397), and is part of the research program of the University Medical Center Groningen: Healthy Aging and Communication.

### ORCID iDs

Carina Pals  <https://orcid.org/0000-0002-4417-2870>  
Andy Beynon  <https://orcid.org/0000-0002-3191-6113>

### References

- Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America*, *126*(5), 2649–2659. doi:10.1121/1.3216914
- Baayen, R., Piepenbrock, R., & van Rijn, H. (1993). *The {CELEX} lexical data base on {CD-ROM}*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Baddeley, A. D., Emslie, H., & Nimmo-Smith, I. (1992). *The speed and capacity of language-processing test*. Bury St Edmunds, England: Thames Valley Test Company.
- 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*(1), 49–72. Retrieved from <https://www.rehab.research.va.gov/jrrd/>
- Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J. D., . . . Gaudrain, E. (2016). Cognitive compensation of speech perception in hearing loss: How and to what degree can it be achieved? *Trends in Hearing*, *20*, 1–16. doi:10.1177/2331216516670279
- Başkent, D., Gaudrain, E., Tamati, T., & Wagner, A. E. (2016). Perception and psychoacoustics of speech in cochlear implant users. In A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk (Eds.), *Scientific foundations of audiology: Perspectives from physics, biology, modeling, and medicine* (p. 285). San Diego, CA: Plural Publishing.
- Benard, M. R., Mensink, S. J., & Başkent, D. (2014). Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. *The Journal of the Acoustical Society of America*, *135*(2), EL88–EL94. doi:10.1121/1.4862879
- Berg, K. A., Noble, J. H., Dawant, B. M., Dwyer, R. T., Labadie, R. F., Gifford, R. H., & Dwyer, R. T. (2019). Speech recognition as a function of the number of channels in perimodiolar electrode recipients, 145, 1556. doi:10.1121/1.5092350
- Bhargava, P., Gaudrain, E., & Başkent, D. (2014). Top-down restoration of speech in cochlear-implant users. *Hearing Research*, *309*, 113–123. doi:10.1016/j.heares.2013.12.003
- Bhargava, P., Gaudrain, E., & Başkent, D. (2016). The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners. *Journal of the Association for Research in Otolaryngology*, *17*, 475–491. doi:10.1007/s10162-016-0565-9
- Blamey, P. J., Pyman, B. C., Gordon, M., Clark, G. M., Brown, A. M., Dowell, R. C., & Hollow, R. D. (1992). Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients. *The Annals of Otology, Rhinology, and Laryngology*, *101*(4), 342–348. doi:10.1177/000348949210100410
- Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY: Pergamon Press. doi:10.1037/10037-000
- 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
- Croghan, N. B. H., Duran, S. I., & Smith, Z. M. (2017). Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility. *The Journal of the Acoustical Society of America Express Letters*, *142*(6), EL537–EL543. doi:10.1121/1.5016044
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498–513. doi:10.1037/0278-7393.30.2.498
- Downs, D. W., & Crum, M. A. (1978). Processing demands during auditory learning under degraded listening conditions. *Journal of Speech and Hearing Research*, *21*(4), 702–714. doi:10.1044/jshr.2104.702
- Drag, L. L., & Bieliauskas, L. A. (2010). Contemporary review 2009: Cognitive aging. *Journal of Geriatric Psychiatry and Neurology*, *23*(2), 75–93. doi:10.1177/0891988709358590
- Eddington, D. K. (1980). Speech discrimination in deaf subjects with cochlear implants. *The Journal of the Acoustical Society of America*, *68*(3), 885. doi:10.1121/1.384827

- 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*(5s), 1201–1215. doi:10.1044/jslhr.4005.1201
- 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*(2), 1150. doi:10.1121/1.1381538
- Fu, Q.-J., Shannon, R. V., & Galvin, J. J. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *The Journal of the Acoustical Society of America, 112*(4), 1664. doi:10.1121/1.1502901
- Fu, Q.-J., Shannon, R. V., & Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *The Journal of the Acoustical Society of America, 104*(6), 3586–3596. doi:10.1121/1.423941
- Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology, 24*(1), 63–68. doi:10.3109/03055369009077843
- Gibbon, D., Moore, R., & Winski, R. (Eds.) (1997). *Handbook of standards and resources for spoken language systems* (p. 886). Berlin, Germany: De Gruyter Mouton.
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87*(6), 2592–2605. doi:10.1121/1.399052
- Hannemann, R., Obleser, J., & Eulitz, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research, 1153*, 134–143. doi:10.1016/j.brainres.2007.03.069
- Henry, B. A., Turner, C. W., & Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *The Journal of the Acoustical Society of America, 118*(2), 1111. doi:10.1121/1.1944567
- Hornsby, B. W. Y. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing, 34*, 523–534. doi:10.1097/AUD.0b013e31828003d8
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149. doi:10.1037//0033-295x.99.1.122
- Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86*(2-3), 199–225. doi:10.1016/0001-6918(94)90003-5
- Lyxell, B., Andersson, J., Arlinger, S., Bredberg, G., Harder, H., & Rönnberg, J. (1996). Verbal information-processing capabilities and cochlear implants: Implications for preoperative predictors of speech understanding. *Journal of Deaf Studies and Deaf Education, 1*(3), 190–201. doi:10.1093/oxfordjournals.deafed.a014294
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953–978. doi:10.1080/01690965.2012.705006
- Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language, 65*(2), 145–160. doi:10.1016/j.jml.2011.04.004
- May, J., Alcock, K. J., Robinson, L., & Mwita, C. (2001). A computerized test of speed of language comprehension unconfounded by literacy. *Applied Cognitive Psychology, 15*(4), 433–443. doi:10.1002/acp.715
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A, 58*(1), 22–33. doi:10.1080/02724980443000151
- Miller, G., & Licklider, J. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America, 22*(2), 167–173. doi:10.1121/1.1906584
- Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., ... Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and Hearing, 38*, 267–281. doi:10.1097/aud.0000000000000396
- Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research, 56*(4), 1075–1084. doi:10.1044/1092-4388(2012/12-0074)
- Pals, C., Sarampalis, A., van Rijn, H., & Başkent, D. (2015). Validation of a simple response-time measure of listening effort. *The Journal of the Acoustical Society of America, 138*(3), EL187–EL192. doi:10.1121/1.4929614
- Pals, C. (2016). *Listening effort: The hidden costs and benefits of cochlear implants* (Unpublished doctoral dissertation) University of Groningen, Groningen, The Netherlands.
- Perreau, A. E., Tatge, B., Irwin, D., & Corts, D. (2018). Listening effort measured in adults with normal hearing and cochlear implants. *697*(2017), 685–697. doi:10.3766/jaaa.16014
- Pichora-fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47*(sup2), S72–S82. doi:10.1080/14992020802307404
- Pisoni, D. B., Manous, L. L. M., & Dedina, M. J. M. (1987). Comprehension of natural and synthetic speech: Effects of predictability on the verification of sentences controlled for intelligibility. *Computer Speech & Language, 2*, 303–320. doi:10.1016/0885-2308(87)90014-3
- Purdy, S. C., Welch, D., Giles, E., Louise, C., Morgan, A., Tenhagen, R., & Kuruvilla-mathew, A. (2017). Impact of cognition and noise reduction on speech perception in adults with unilateral cochlear implants. *Cochlear Implants International, 18*(3), 162–170. doi:10.1080/14670100.2017.1299393
- Rabbitt, P. M. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science, 6*(8), 383–384. doi:10.3758/bf03330948
- Ralston, J. V., Pisoni, D. B., Lively, S. E., Greene, B. G., & Mullennix, J. W. (1991). Comprehension of synthetic speech produced by rule: Word monitoring and sentence-by-sentence listening times. *Human Factors, 33*(4), 471–491. doi:10.1177/001872089103300408

- Rönnerberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*, 42(Suppl 1), S68–S76. doi:10.3109/14992020309074626
- Saija, J. D., Akyürek, E. G., Andringa, T. C., & Baskent, D. (2014). Perceptual restoration of degraded speech is preserved with advancing age. *Journal of the Association for Research in Otolaryngology*, 15(1), 139–148. doi:10.1007/s10162-013-0422-z
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403–428. doi:10.1037/0033-295X.103.3.403
- Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474–494. doi:10.1037//0096-3445.110.4.474
- 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(5), 1230–1240. doi:10.1044/1092-4388(2009/08-0111)
- Saxton, J. A., Ratcliff, G., Dodge, H., Pandav, R., Baddeley, A., & Ganguli, M. (2001). Speed and capacity of language processing test: Normative data from an older American community-dwelling sample. *Applied Neuropsychology*, 8(4), 193–203. doi:10.1207/s15324826an0804\_1
- Schvartz, K. C., Chatterjee, M., & Gordon-Salant, S. (2008). Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. *The Journal of the Acoustical Society of America*, 124(6), 3972–3988. doi:10.1121/1.2997434
- Shannon, R. V., Fu, Q., & Galvin, J. J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngologica*, 552, 50–54. doi:10.1080/03655230410017562
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304. doi:10.1126/science.270.5234.303
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, 123(1), 489–499. doi:10.1121/1.2783762
- Sladen, D. P., Nie, Y., Berg, K., Sladen, D. P., Nie, Y., & Berg, K. (2018). Investigating speech recognition and listening effort with different device configurations in adult cochlear implant users. *Cochlear Implants International*, 19, 119–130. doi:10.1080/14670100.2018.1424513
- Stickney, G. S., Loizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., & Opie, J. M. (2006). Effects of electrode design and configuration on channel interactions. *Hearing Research*, 211(1-2), 33–45. doi:10.1016/j.heares.2005.08.008
- Van Wieringen, A., & Wouters, J. (2008). LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *International Journal of Audiology*, 47(6), 348–355. doi:10.1080/14992020801895144
- Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A meta-analysis. *Psychology and Aging*, 18, 443. doi:10.1037/0882-7974.18.3.443
- 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(3), 1671–1684. doi:10.1121/1.428451
- Wagner, A. E., Pals, C., de Blecourt, C. M., Sarampalis, A., & Başkent, D. (2016). Does signal degradation affect top-down processing of speech? *Advances in Experimental Medicine and Biology*, 894, 297–306. doi:10.1007/978-3-319-25474-6\_31
- Wagner, A. E., Toffanin, P., & Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers in Psychology*, 7(March), 1–14. doi:10.3389/fpsyg.2016.00398
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(40), 14010–14021. doi:10.1523/JNEUROSCI.1528-12.2012
- Wingfield, A., Aberdeen, J. S., & Stine, E. A. (1991). Word onset gating and linguistic context in spoken word recognition by young and elderly adults. *Journal of Gerontology*, 46(3), P127–P129. doi:10.1093/geronj/46.3.p127
- Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, L. C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the American Academy of Audiology*, 17(7), 487–497. doi:10.3766/jaaa.17.7.4
- Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology*, 18(7), 548–558. doi:10.3766/jaaa.18.7.3
- Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, 20, 1–17. doi:10.1177/2331216516669723
- Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America*, 131(2), 1465–1479. doi:10.1121/1.3672705
- Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, 36(4), e153–e165. doi:10.1097/AUD.0000000000000145
- Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 8(3), 384–392. doi:10.1007/s10162-007-0085-8
- 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(4), 480–490. doi:10.1097/AUD.0b013e3181d4f251