Age-related Changes in Listening Effort for Children and Teenagers with Normal Hearing and Cochlear Implants

Benson Cheng-Lin Hsua, b,Filiep Vanpouckeb, Margreet Langereisc, Ann Dierckxd, & Astrid van Wieringen a

aDepartment of Neurosciences, Experimental ORL, Catholic University of Leuven, Leuven, Belgium

bCochlear Technology Centre, Mechelen, Belgium

cDepartment of Otorhinolaryngology, Head and Neck Surgery, Hearing and Implants, Radboud University Medical Center, Nijmegen, the Netherlands

bDepartment of Otorhinolaryngology, Head and Neck Surgery, University Hospitals Leuven, Leuven, Belgium

1 STRUCTURED ABSTRACT

2 Objectives: A clinically viable measure of listening effort constitutes a critical component of a
3 comprehensive strategy safeguarding the educational success of hard-of-hearing children
4 enrolled in mainstream schools. To this end, a novel behavioural paradigm of listening effort
5 targeting school-age children has been proposed and tested in Hsu et al. (2017). The current
6 paper consists of two follow-up experiments investigating the effects of age in a similar
7 paradigm in a group of normal-hearing (NH) participants before attempting to evaluate listening
8 effort in a group of school-age cochlear implant (CI) users using the same method. The main
9 objectives of these experiments include the construction of usable normative references as well
10 as comparing the results from the CI users to their NH peers.

11 Design: In Experiment 1, the listening effort paradigm was evaluated in a group of 90 NH
12 participants with roughly even age distribution between 6 and 26 years old. Test materials were
13 presented in quiet as well as two levels of speech-shaped static noise. The secondary task
14 consisted of two categorization questions “animal” and “dangerous” representing two levels of
15 semantic processing depth. Custom iPad-based software was developed to enable streamlined
16 test administration both at home and in a clinic. Two outcome measures were obtained for each
17 experimental condition: a percentage-correct word recognition score (WRS) and the response
18 time (RT) measured during the semantic categorization task. Additionally, each participant’s
19 long-term memory retrieval speed and working memory capacity were assessed through

20 standardized tests. It was hypothesized that adding noise would negatively affect both WRS and
21 RT, whereas advancing in age, in our young cohort, would improve both measures. A subsequent
22 Experiment 2 administered a shortened version of the listening effort assessment to 14 school-
age CI users between 5 and 14 years old at a university 1 clinic. The pattern of results from the CI
2 group was expected to be similar to the NH group except with larger between-subject variability.
3 Results: For the NH group, while WRS was significantly affected by age and noise levels, RT
4 was significantly affected by age, noise levels, and depth of processing. RT was significantly
5 correlated with long-term memory retrieval speed but not with working memory capacity. There
6 was also a significant interaction effect between age and noise levels for both WRS and RT.
7 From the NH RT data set, 95% confidence and prediction intervals were calculated to serve as
8 reference normative values for clinical populations needing hearing devices. For the CI group,
9 the effect of age on the two outcome measures was more visible when target words were
10 presented in “quiet”. The WRS outcomes from CI participants were strongly correlated to their
11 RT results when they engaged in the “animal” categorization task in “quiet”.
12 Conclusion: Listening effort appears to decrease with age from early school-age years to late
13 teenage years. The effects of background noise and processing depth are comparable with those
14 reported in Hsu et al. (2017). Future studies interested in expanding the paradigm’s clinical
15 viability should focus on the reduction of testing time while maintaining or increasing the
16 sensitivity and external validity of its outcome measures.

17 INTRODUCTION

18 Listeners suffering from hearing loss often describe the experience of having a conversation in a
19 noisy situation with words such as “challenging”, “draining”, and “effortful”. That is, listening in
20 a noisy environment seemingly drains their cognitive energy (Pichora-Fuller et al. 2016).
21 Sustained periods of strenuous hearing have been shown to lead to feelings of fatigue and
22 exhaustion at the end of the day (Hétu et al. 1988; Hornsby 2013). In terms of school-age
23 children, excessive exposure to noise in an educational setting has been identified as a source of
fatigue and headache 1 (Wållinder et al. 2007). It has also been shown that prolonged periods of
2 effortful learning (Downs & Crum 1978) or mental fatigue (Bess & Hornsby 2014) are both risk
3 factors for poor educational outcomes, since this increased mental demands could become an
4 interference whenever a school-age child attempts to memorize or comprehend new information

5 (McGarrigle et al. 2019).

6 2.1 Cochlear implants and listening effort

7 A recent review article by Ohlenforst et al. (2017) provides a good overview of the ways with
8 which hearing loss and hearing amplification affects listening effort. A special form of hearing
9 amplification is the use of cochlear implants (CI). CI technologies partially restore auditory
10 abilities to deafened or near-deafened individuals (Wilson & Dorman 2008), but a decrease in
11 spectral resolution as a side-effect of CI-assisted hearing has been shown to lead to increased
12 listening effort (Pals et al. 2013). While having two CI units (as opposed to a unilateral setup)
13 generally imparts advantages to a child's speech recognition in noise (Steffens et al. 2008), the
14 extent to which binaural inputs from their bilateral implants can be perceived in a synchronous
15 fashion has equally an impact on listening effort in children (Steel et al. 2015).

16 As CIs become widely accepted as a standard treatment over the past two decades (Kral &
17 O'Donoghue 2010), school-age CI users without additional physical or mental complications are
18 increasingly being enrolled in mainstream rather than specialized schools (Venail et al. 2010; De
19 Raeve & Lichtert 2012). Despite this current trend, there have been surprisingly few research
20 studies dedicated to exploring listening effort in children with CI. Mainstream schools are known
21 to be noisy learning environments, and the daily difficulties hard-of-hearing children face are
22 well documented in the literature (Preisler et al. 2005; van Wieringen & Wouters 2015). A more
23 recent study by Busch et al. (2017) suggests that CI users tend to spend more time in noisy

4

situations during their primary- and secondary-1 school years compared to the various stages of
2 adulthood. It seems that school-age CI users, in addition to the disadvantages associated with CI
3 sound processing distortions (e.g., Bierer 2007; Wouters et al. 2015), must now contend with the
4 daily reality of listening and learning in acoustically challenging environments. It is also known
5 that, much more than the normal-hearing (NH) general population or hearing-aid users, post6
implant speech perception performance outcomes also tend to vary considerably among CI
7 recipients (Pisoni et al. 2017). Since the amount of "effortful listening" is not readily measured
8 by the common sets of diagnostic tools employed in audiology clinics today, Hsu et al. (2017),
9 proposed a novel behavioural paradigm combining traditional speech intelligibility test with a
10 top-down cognitive component to help clinicians identify which children with CI's require
11 greater listening effort and therefore are in need of more rehabilitative attention. The current

12 study intends to build upon this behavioural listening effort paradigm by 1) investigating how
13 age-related developmental changes affect its outcome measures, and 2) comparing
14 paradigm/assessment results between age- and linguistically-matched NH and CI participants.

15 2.2 Measuring listening effort behaviourally

16 Although a universally accepted definition of listening effort is yet to be agreed upon by the
17 scientific community, a working definition adopted by the present study characterizes listening
18 effort as the expenditure of limited cognitive resources (such as working memory capacity) to
19 successfully understand speech (Hicks & Tharpe 2002; Gosselin & Gagné 2010; Gustafson et al.
20 2014; Pichora-Fuller et al. 2016). Past studies of listening effort such as Hicks and Tharpe
21 (2002), Gosselin and Gagné (2010), and Gustafson et al. (2014) have all specified a similar
22 definition in the formulation of their methodologies. Successive review articles by Gosselin and
23 Gagné (2010), McGarrigle et al. (2014), and Gagné et al. (2017) provide excellent comprehensive
overviews of the 1 various approaches for measuring listening effort. The

2 techniques can be broadly categorized into subjective, behavioural, and physiological methods.

3 Among them, behavioural experiments evoke subjects' responses in a typically predictable
4 manner and have been utilized in numerous psychological studies.

5 The application of behavioural paradigms to studies of psychological aspects such as effort and
6 attention often cites Kahneman's (1973) seminal theory that there is a limited supply of cognitive
7 capacity for an individual to perform any kind of mental work. Because the cognitive resources
8 available for the task of "listening" are limited, a behavioural experiment can be designed to
9 measure "effort" by introducing an additional task that competes for the scarce resources. This is
10 often referred to as a "dual-task paradigm," where the primary task involves listening and
11 repeating speech materials and the secondary task involves an additional activity that further
12 taxes the system (Gosselin & Gagné 2010). When the primary listening task becomes more
13 difficult, one might be able to maintain an equal level of speech recognition as when the task of
14 listening is easier by devoting a greater share of his/her top-down cognitive resources. This
15 comes with the cost of not being able to do other tasks equally well as before. A dual-task
16 paradigm thus aims to capture this decrement through the addition of a secondary task. The
17 maintenance of the performance of the primary (listening) task at the expense of the
performance

18 at a secondary behavioural task is then taken as an indication of an increase in the level of
19 listening effort (see, for examples, Hicks & Tharpe 2002; Sarampalis et al. 2009; Houben et al.

20 2013; Pals et al. 2013; Picou & Ricketts 2014).

21 2.3 Depth of semantic processing

22 Employing the “depths of processing” framework (Craik & Lockhart 1972) in a dual-task

23 paradigm is one way of tapping into the domains of human cognition relevant to the
measurement of listening effort. This stipulates that a shallow level of processing a piece of an
2 auditory, linguistic signal would be the rapid analysis of acoustic characteristics such as pitch or
3 rhythm, whereas deeper levels of processing vary from rapid mapping syllabic information to
4 pre-existing phonological patterns (Rönnberg et al. 2013) to the extraction and elaboration of
5 semantic meaning and associations (Eysenck & Eysenck 1979). Recent studies such as Picou and
6 Ricketts (2014), Picou et al. (2017), and Hsu et al. (2017) have applied the depth of processing
7 framework in their design of behavioural paradigms investigating the effects of background
8 noise levels as well as changing secondary tasks on listening effort outcome. These paradigms
9 challenged the participants to recognize spoken speech in unfavourable conditions (such as in
10 noise) and thereby required the activation of explicit cognitive resources such as working
11 memory (Rönnberg et al. 2013). Additionally, these recent studies of listening effort also tasked
12 the participants to put auditorily perceived words into categories, which required deliberate and
13 explicit access to long-term memory stores containing knowledge about the world. This
14 approach not only mimics how the cognitive system needs to function in a realistic learning
15 environment but also potentially increase the sensitivity of the assessment to finer fluctuations in
16 the level of listening effort.

17 Picou and Ricketts (2014) presented behavioural dual-task results showing that changes in
18 reaction time when listening in the absence or the presence of background noise were
19 significantly greater when the secondary task involved judging the grammatical property of a
20 word than when the secondary task was a visual task unrelated to the word. The lengthening of
21 the reaction time is interpreted as a consequence of an increase in listening effort. This effect was
22 demonstrated in both normal-hearing (NH) and hard-of-hearing adults. A subsequent study by
23 Hsu et al. (2017) modified the dual-task paradigm presented in Picou and Ricketts (2014) and
made the deeper-processing conditions suitable for school-age children. The main effects of
2 background noise and processing depth on semantic categorization response times were
3 consistent between Hsu et al. (2017) and Picou and Ricketts (2014). However, two outstanding
4 questions remained to be explored: 1) how would general cognitive maturation due to an increase
5 in chronological age affect the outcomes of such dual-task assessment of listening effort, and 2)

6 how would children with CI's perform on this test?

7 2.4 Age effect on listening effort

8 Effortless access to spoken language during the formative elementary school years is generally
9 considered to be an important factor in the language and intellectual development of children. It
10 is well established in the literature that children 10 years and younger generally display lower
11 performance during auditory word-recognition-type tasks when compared to older
12 children/teenagers and adults, especially when the input acoustic signals are degraded due to
13 noise or other factors (e.g., Eisenberg et al. 2000; Corbin, et al. 2016). Because children
14 experience significant linguistic and cognitive maturation during their elementary school years
15 (Cartwright 2002; Piaget & Inhelder 2008), the same behavioural paradigm may yield different
16 patterns of results between children and adults. However, most of the published studies on how
17 listening effort changes with age focused on younger versus older adults (e.g., Degeest et al.
18 2015; Wu et al. 2016; Ward et al. 2017) with only limited attention given so far to the
19 comparison of a listening effort measure between children and adults directly (Picou et al. 2017).
20 Regarding performance during a dual-task paradigm, another recent study by Grieco-Calub et al.
21 (2017) suggested that baseline secondary task performance, rather than chronological age,
22 appeared to be a stronger predictor in school-age children for changes in secondary task
23 performance in response to a more challenging listening condition. Perreau et al. (2017), on the
other hand, did find a significant correlation between age and the level of listening effort
2 exhibited during a dual-task experiment among their adult participants. Indeed, little is known on
3 how listening effort changes with age in NH children. It has been speculated that, due to a
4 diminishing supply of cognitive resources, older adults tend to process information at a shallow
5 level of processing rather than a more elaborate level and thus exhibit poorer long-term retention
6 of the processed information compared to younger adults (Craik & Simon 1980). How would
7 this difference in the capacity of limited cognitive resources play out in children? Would children
8 exhibit more listening effort as measured by a behavioural paradigm due to their cognitive
9 system being a state a growth and maturing? By examining the pattern of listening effort change
10 from young childhood to young adulthood, it would help us add another aspect of nuance when
11 measuring this phenomenon. Furthermore, studying how listening effort changes with age in a
12 larger group of data than what was presented in Hsu et al. (2017) would also provide reference
13 data to allow meaningful comparison when testing children with cochlear implants in a clinical

14 setting.

15 2.5 Purposes and hypotheses

16 While the ultimate purpose of devising such a behavioural paradigm is the correct identification
17 of children who experience greater listening effort in a mainstream school environment and
18 thereby allowing timely intervention, there are at least two intermediate steps that must be
19 achieved first. To begin, the effects of various factors, especially those related to chronological
20 age, on the paradigm's outcome variables should be thoroughly investigated in a sizable sample.
21 Secondly, it should be demonstrated that the listening effort paradigm could be successfully
22 implemented in a clinical environment, where target populations such as school-age CI users
23 could be most likely expected to be tested. To these ends, the present article consists of two
separate but related experiments. In Experiment 1, data were collected from a group of NH
2 participants using a variation of the listening effort paradigm first described in Hsu et al. (2017),
3 which included a traditional speech intelligibility test coupled to secondary tasks requiring
4 further processing of the words perceived. Experiment 1's participants covered the entire age
5 span from early elementary-school students to working young adults, a data set that would
6 complement the findings from Hsu et al. (2017) by allowing a systematic analysis of the effects
7 of cognitive development and maturation due to age on aspects of the listening effort paradigm
8 such as noise tolerance and semantic categorization. It was hypothesized that listening effort
9 increases with greater background noise and a deeper level of cognitive/semantic processing,
10 similar to what was reported in Hsu et al. (2017). Advancing in age, on the other hand, would
11 generally improve word recognition performance and semantic categorization speed, at least
12 until young adulthood. Experiment 2, on the other hand, applied the same but somewhat
13 abbreviated version of our novel listening effort assessment to a group of school-age CI users as
14 part of their annual audiology appointment in a "pilot-like" follow-up study. We predicted that
15 some, but not all, of the CI participants would exhibit worse speech intelligibility and listening
16 effort outcomes when compared to their age-matched NH peers from Experiment 1 due to the
17 nature of CI sound processing accentuating the effects of signal masking from background noise
18 and "deeper" processing from a more cognitively demanding secondary task. Moreover, it was
19 hypothesized that there would be considerable variation in assessment outcomes within the CI
20 group, especially when one divided them into two sub-groups based on age: the "younger"
21 school-age CI users versus the "older" school-age CI users.

10

10

3 EXPERIMENT 1: EFFECTS OF AGE, NOISE, PROCESSING DEPTH, AND

2 COGNITIVE CAPACITIES IN NH PARTICIPANTS

3 3.1 Methods

4 3.1.1 Participants

5 During the first stage of the study, we recruited a total of 90 NH participants aged 6.15 to 26.82
6 years old (mean age = 13.80 years old; median age = 12.21 years old). More specifically, 51 out
7 of the 90 participants were primary-school-aged (6 to 12 years old), 22 participants were
8 secondary-school-aged (13 to 18 years old), and the remainder 17 participants were young adults
9 (19 years and older). Each participant was screened with standard pure-tone audiometry before
10 any data was collected to ensure bilateral normal hearing (< 20 dB HL) at all prominent speech
11 frequencies (500, 1000, 2000, and 4000 Hz). All pediatric participants were current students
12 attending a local, mainstream primary or secondary school, and no participants reported any
13 history of language or developmental disorders. For participants 12 years or younger, an
14 informed consent letter was signed by their parents, explaining the purposes and procedures of
15 the study. Direct verbal assent was also obtained from each child after explaining that they were
16 free to terminate participation at any time without any consequences. The older participants (13
17 years or older) signed the informed consent letter themselves. Every participant was offered 10
18 Euros at the end of the experiment as a gesture of appreciation for their participation. The
19 medical ethical committee of UZ Leuven/KU Leuven approved the participant selection criteria
20 and test procedures (Approval No. B322201523779).

21 3.1.2 Cognitive-language tests

22 In essence, our experimental listening effort assessment induces listening effort by creating a
23 situation where degraded speech signals not only need to be understood, but the meaning of the
“correct answer” also needs at least one more step of further processing to make connections to
2 pre-existing knowledge of the world. This calls for the active consumption of finite cognitive
3 resources such as working memory and long-term memory store retrieval mechanism. Therefore,
4 quantifiable cognitive entities such as working memory capacity and long-term memory retrieval
5 speed could potentially serve as predictors of how well one would perform on our listening effort
6 assessment. A recent large-scale validation study by Strand et al. (2018) also indicated that the

7 ability to use cognitive resources more efficiently could counterbalance the negative effects of
8 background noise when processing speech. To explore these potentially important aspects of
9 cognition, two validated and established cognitive language tests from the Fourth Edition of the
10 Dutch-language Clinical Evaluation of Language Fundamentals (referred to as “CELF-4-NL” or
11 simply as “CELF” from here onward; see Kort et al. 2010) were included in the experiment. The
12 first is a “Digit Recall” working memory measure, where a participant was asked to repeat an
13 increasingly longer sequence of spoken digits in both “Forward” (i.e., in the same order) and
14 “Backward” fashions. On the other hand, the long-term memory retrieval speed was evaluated
15 through the “Word Associations” test, where each participant was asked to verbally say, as many
16 examples as they could in 60 seconds, items belonging a pre-defined category (e.g., “food” or
17 “clothing items”).

18 3.1.3 Signal-to-noise ratios

19 To vary the amount of cognitive resources required on the word recognition side of the current
20 listening effort test, three signal-to-noise ratios (SNR) conditions were set up when presenting
21 words auditorily: “quiet”, “+3 dB SNR,” and “–3 dB SNR.” These SNR levels were chosen in
22 part to be within the limits of real-life mainstream classroom noise level of between +5 to –7 dB
23 (Crandell & Smaldino 1994; Howard et al. 2010). While an SNR of –3 dB was chosen to provide a
reasonably challenging condition for 1 the NH participants, the suitability of +3 dB as an
2 intermediate SNR condition was determined through pilot testing with a small group of children
3 and teenagers (n = 6) with CI’s. Using an adaptive SNR method, when word recognition
4 performance target level was fixed at 70-80%, +3 dB turned out to be the average SNR level for
5 this group of pilot participants with CI. Therefore, “+3 dB SNR” was considered a suitable
6 presentation condition when directly comparing the outcomes between NH and CI subjects,
7 where a fixed SNR would be required.

8 3.1.4 Depth of processing levels

9 The two semantic categorization tasks used in the current study, namely “animal” and
10 “dangerous”, were adopted from Hsu et al. (2017). They represented two levels of depth of
11 processing placing demands on the limited cognitive resources through increased retrieval to
12 long-term memory stores. While the “animal” task asked each participant to indicate if a

13 perceived word represented an animal or not, the “dangerous” task asked if a perceived word
14 represented something potentially harmful (to themselves and/or others) or not. The position of
15 the choices of “Yes” or “No” responses, or the two “buttons” on the touch screen, remained
16 constant throughout the experiment for both of these two tasks. The “animal-or-not”
17 categorization question involved a familiar and mildly interesting subject appropriate for
18 children and adults alike and represented a moderate increase to the access of long-term memory
19 facilities. On the other hand, asking about the “dangerousness” of each target word was
20 expected, compared to the simpler “animal” question, to elicit greater stimulus elaboration and
21 enrichment through imagination and association with past experience (Craik & Lockhart 1972)
22 and therefore represented the deepest level of semantic processing in the current paradigm.

1 3.1.5 Materials and equipment

2 The present study used the same pool of 146 Dutch nouns as developed and first utilized in Hsu
3 et al. (2017). Four categories of words comprised this pool: harmless animals, harmless non-
4 animals, dangerous animals, and dangerous non-animals. These words had been verified to be
5 suitable for use by children as young as 6 years old. For the selection, validation, recording, and
6 post-processing of these speech materials, please refer to Hsu et al. (2017) for further detail.
7 From these 146 words, six lists of 30 words each were chosen for the “animal” categorization
8 and another six lists of 30 words each for the “dangerous” condition. The generation of these
9 word lists was done by a computer algorithm according to pre-defined rules of distribution and
10 randomization. Speech-shaped stationary noise was also created based on the long-term average
11 spectrum of the full pool of 146 words to ensure spectral compatibility between target words and
12 added energetic masker.

13 The custom software used in the study was developed in the Objective-C language as an iOS app
14 designed for an Apple iPad device. Upon launching the app, its main setup screen allows a tester
15 to enter the participant ID, specify parameters for the current test run, and select different
16 subtests to run. For the main listening effort paradigm, the software also contains a “training
17 mode”, which constitutes a mandatory part of the test protocol. When the assessment is in the
18 “training mode”, each of the 30-item word list could be played with 10 words in each segment
19 (the first 10, middle 10, or the last 10) to familiarize a participant to the paradigm without adding
20 too much time at the beginning of each test session. The Dutch words and the speech-weighted

21 noise were first digitally mixed by the software and then routed to the participant through
22 professional-grade supra-aural headphones (beyerdynamic DT 770 PRO 250 ohms) binaurally.
23 According to the manufacturer, this model of headphones provides approximately 18 dBA of
14

14

ambient noise isolation when properly fitted over the ears on both sides (beyerdynamic 2014).

2 The output of these headphones was measured and calibrated through a clamped-down artificial
3 ear (Type 4153, Brüell & Kjær) connected to a sound level meter before the first participant was
4 tested. The computer software computes the output levels according to the desired signal-to-5
noise ratio (SNR) of both speech materials and noise to maintain a fixed 65 dB SPL presentation
6 level throughout the experiment.

7 3.1.6 Eye-to-hand speed test

8 In addition to the main dual-task paradigm, an additional measurement was also made to see how
9 fast the NH participants could move their responding hand to the touch screen in the absence of
10 any cognitive component. Obtaining this “baseline” measure would allow us to control for any
11 skeletomuscular maturation aspect when analyzing the age factor. To this end, each NH
12 participant was asked to touch as quickly as possible a picture of a wasp which appears randomly
13 either on the right or left side of the touch screen. The interval between each “wasp appearance”
14 on the screen varied between 1 and 3 seconds to reduce guessing through repetitive, rhythmic
15 movement. Each run of the eye-to-hand speed test consisted of 12 trials, out of which the first
16 two trials were always omitted from final analysis as “preparatory trials” (i.e., when a participant
17 was still adjusting his/her body posture to optimize their speed during the first couple of trials).
18 From the remaining 10 trials, an averaged Eye-to-Hand Reaction Time (abbreviated as RT_{EH}
19 from this point onward) could be calculated for each speed test run. This time-efficient eye-hand
20 coordination test was administered both before and after the main listening effort conditions as
21 well as both with and without the presence of speech-shaped static noise at a constant level
22 corresponding to approximately 0 dB SNR. This variety of conditions was to control for any test
fatigue effect as well as to measure the acoustic environment in which the main dual-task paradigm
2 was conducted.

3 3.1.7 Test procedures

4 All testing was carried out in a quiet room inside a participant’s own home. A constant

5 presentation level of 65 dB SPL (which constitutes a supra-threshold level for NH participants)
6 combined with the aforementioned acoustic shielding from cushioned supra-aural headphones
7 minimized the audibility of occasional low-level background noise typical of a home
8 environment. The exact speech and noise dB SPL levels were calculated by the computer after a
9 desired SNR (i.e., “quiet”, “+3”, or “-3”) was selected by the tester. This technical detail was
10 implemented to ensure that the total sound output would never be higher than what some of our
11 participants with CI devices could tolerate in case a more negative SNR was chosen. Each
12 session with one participant took about one hour of time. The participant and the tester were
13 seated at a 90° angle to each other. The upper tip of the participant’s responding hand at rest
14 position was kept approximately 4 cm from the bottom edge of the touch screen throughout the
15 session, which always began with pure-tone audiometry followed by the aforementioned two
16 CELF-4-NL cognitive-language tests and the first two runs of the eye-hand coordination speed
17 test (the first run in “quiet” immediately followed by a second run in “noise”). The tester
18 provided clear verbal instructions for each section of the experiment, and practice runs were
19 carried out the dual-task paradigm was grasped unambiguously by the participant. The order of
20 the listening effort conditions was systematically counterbalanced to minimize any order/training
21 effect. The response time of the NH participants (RTNH) between the end of a word presentation
22 and the moment when a participant indicated a categorization choice on the touch screen was
23 automatically captured by the test software. The correctness of verbal repetitions was judged by
the tester directly and entered on a prepared Excel spreadsheet. 1 The percentage of correct verbal
2 repetitions out of each list of 30 words was later calculated to obtain the word recognition score
3 for each condition for each NH participant (WRSNH). While the word recognition phase had no
4 particular time limit, the categorization phase had a response window of 5 seconds for each
5 presented word. A no-touch event during the categorization phase was marked as “no response”,
6 whereas any phonemic replacements or non-responses during the word recognition phase were
7 both recorded as “wrong”. At the very end of an experimental session, two more runs (i.e., the
8 third and fourth runs) of eye-hand coordination speed test would be conducted: the third run in
9 “noise” immediately followed by the final run in “quiet”.

10 3.2 Results

11 3.2.1 WRSNH

12 Figure 1 shows the entire WRSNH data set divided between the six experimental conditions as

13 well as participants' age groups. Two apparent observations can be drawn. First of all, there was
14 an obvious ceiling effect at the "quiet" conditions (and possibly at "+3 dB SNR" for adults, too).
15 Secondly, there did not appear to be any large difference in WRSNH between "animal" and
16 "dangerous" across all SNR conditions and age groups. To further investigate these observations
17 with statistics, a generalized linear mixed-effect (GLM) model dedicated to a binomial
18 distribution (given that all WRSNH was calculated based on dichotomic "true-or-false" responses)
19 was subsequently constructed. Three fixed-effect factors, namely "Age" as a continuous variable,
20 "SNR" ("quiet", "+3 SNR", and "-3 SNR"), and "Task" ("animal" and "dangerous"), were fitted
21 along with one random-effect factor "ID" to account for the fact that each participant might have
22 a somewhat different "default" performance level (i.e., each one having a different Y-intercept).
23 The R package used to analyze all GLM models discussed in this paper is called "lme4" (Bates

17

17

et al. 2015). An effect size estimate for each GLM 1 analysis presented in this paper is represented
2 by the numerical "shift" of Akaike information criteria (AIC; Akaike 1974) between a "full"
3 GLM model and a "reduced" model without the factor/component under investigation. An AIC
4 measures how well a model fits a given set of data, taking into consideration the inherent
5 complexity of the model in question. A greater AIC shift thus represents a greater effect of
6 removing (or adding) a particular factor/component to a model. Maximum likelihood analysis
7 involving pairwise comparisons of different iterations of the GLMWRS suggested that, while
8 participants' "Age" ($\chi^2(1) = 48.32$, $p < 0.001$, AIC shift = 46.3) and "SNR" levels ($\chi^2(2) =$
9 911.32, $p < 0.001$, AIC shift = 907.3) significantly impacted WRS performance, changing the
10 categorization "Task" from "animal" to "dangerous" did not significantly affect WRS outcomes.
11 Further analyses indicated that, as far as the WRSNH data were concerned, each of the three
12 "SNR" levels was significantly different from the others when subjected to post hoc pairwise
13 comparisons with Bonferroni corrections: "quiet" versus "+3 dB" ($p < 0.001$), "quiet" versus "-3
14 dB" ($p < 0.001$), as well as "+3 dB" versus "-3 dB" ($p < 0.001$). Moreover, there was also a
15 significant interaction between "Age" and "SNR" ($\chi^2(2) = 10.00$, $p < 0.01$, AIC shift = 6.0). This
16 finding suggested that the rate of improvement in WRSNH due to age-related maturation from
17 young childhood to young adulthood was greater when background noise level increased.

18 3.2.2 RTetH

19 Figure 2 shows a boxplot representing our NH participants' RTeH grouped by the time of testing
20 ("start" versus "end") as well as by the listening condition ("quiet" versus "noise") in a 2 x 2
21 experimental design. A mixed-effect regression analysis was carried out on the entire RTeH data
22 set using "Age", "Time of Testing", and "Listening Condition" as fixed variables and "Subject"
23 as a random variable. Again, this was done to account for the fact that each participant might
typically exhibit a different "starting" eye-to-hand coordination speed (i.e., each one having a
2 different Y-intercept). The result of this analysis indicated that only the fixed variable "Age" had
3 a main effect on RTeH ($p < 0.001$). While the RTeH values were related to a participant's age as
4 expected, there did not seem to be much impact on the NH participants' eye-to-hand speed
5 regardless of whether the tests were performed before or after the main listening effort paradigm
6 or whether there was auditory noise in the background or not. In this case, all four of the RTeH
7 outcomes from the same participant could be considered equivalent and were affected by age and
8 other subject-specific factors. We could then average all RTeH values from each NH participant
9 into a sort of "eye-to-hand coordination correction factor" and subtract it from the same
10 participant's raw RTNH outcomes to create the "adjusted-RTNH" data set. Doing so might allow
11 us to at least partially eliminate the part of RTNH that was attributable to physical maturation that
12 enabled the older participants to move their hand faster in response to a visual stimulus.
13 Hypothetically, the adjusted-RTNH could be more purely auditory and cognitive in nature than
14 the unadjusted raw RTNH. However, preliminary mixed-effect regression analysis on the
15 adjusted-RTNH set using "Age", "SNR", and "Task" as fixed variables and "Subject" as a
16 random variable resulted in no noticeable differences from those results produced when an
17 identical analysis with identical factors was performed on the raw RTNH data set. Given the facts
18 that 1) unadjusted and adjusted RTNH yielded the same mixed-effect analysis results, 2) we were
19 not able to conduct the eye-to-hand speed test with any of the CI participants during Experiment
20 2 (to be described in the second half of this article) due to the time constraint of real-life clinical
21 appointments, and 3) comparing the RT outcomes between the CI and NH groups was one of the
22 primary objectives of Experiment 2, it was decided to include only raw RTNH in all analyses for
23 the remainder of this article.

1 3.2.3 RTNH

2 Figure 3 presents a summary boxplot of the raw RTNH data set divided between the six

3 experimental conditions as well as five different age groups. The pattern shown on this figure
4 suggests that the median RTNH not only tends to decrease with age, it also tends to increase with
5 background noise levels. Moreover, median RTNH values from “animal” tend to be lower than
6 those from “dangerous” across all age groups and SNR conditions. The nature and magnitude of
7 the effects of “Age”, “SNR”, and “Task” on RTNH, and how useful these factors would be in
8 predicting future data points (including those from participants with CI) could be evaluated
9 through the construction of a robust statistical model.

10 Further examination of Figure 3 suggests an apparent non-linear relationship between RT
11 outcome variable and “Age”. That is, RTNH seems to decrease with age up to a point where it
12 begins to “plateau out” for the age range we have data from young school-age children to young
13 adults. Mathematically, RTNH, as a measure of listening effort, exhibits an inverse exponential
14 relationship with age with marked plateauing observed for participants older than 18 years.

15 Finally, since pre-modeling Shapiro-Wilk tests revealed the non-normality of our RTNH data set,
16 a simple log₁₀ transformation was applied to RTNH. Subsequent histograms and Shapiro-Wilk
17 tests confirmed that the transformed data set did not violate normality. Taking into account the
18 aforementioned considerations and adding the categorical, fixed variables of “SNR” and “Task”,
19 we constructed the following generalized linear mixed-effect model for RT (GLMRT) showing
20 how future listening effort performance log₁₀RT could be predicted from these relevant variables:

21

$\log_{10} RT$

RT_{ref}

$/ \sim \alpha_{ASYM} + \beta_1 \cdot e$

$-Age$

$/ TC_{22} + \beta_2 \cdot SNR + \beta_3 \cdot Task + (1 | ID) + \varepsilon$

23

20

20

In this formula, model parameters α_{ASYM} , β_1 , β_2 , and β_3 , are estimated as part of the GLMRT
2 analysis. Please note that the theoretical “RT_{ref}” term that forms part of the outcome variable has
3 a constant value of “1” and exists by necessity to cancel out the original unit of RTNH “msec” so
4 that the resultant “log-transformed RT” could be a dimensionless quantity. The two fixed-effect

5 terms have the same levels as those specified in the WRSNH analysis presented above: “quiet,”
6 “+3 SNR,” and “-3 SNR” for “SNR” and “animal” and “dangerous” for “Task”. The “ID” term
7 in brackets (1 | ID) denotes a random effect for the GLMRT: it acknowledges that each participant
8 could have a characteristic response time performance related to factors such as baseline hand
9 movement speed that was not controlled in the present experiment. In other words, the model
10 allows for a random y-intercept for each NH participant regardless of their age or the two
11 manipulated variables “SNR” and “Task”. The term ϵ denotes the remainder error terms not
12 captured by the preceding mixed-effect model. Of special interest is the parameter “TC” for the
13 “Age” term, which is a place holder for a time constant that mediates the rate of log₁₀RT change
14 according to “Age”. The time constant was estimated by applying a least-squares (i.e.,
15 linearization) method on the original data set with the inverse exponential equation of “Age”
16 supplied above. The result of this exercise gave us a “TC” value of around 8.74 as the point of
17 maximum model fit.

18 Maximum likelihood analysis involving pairwise comparisons of different iterations of the
19 GLMRT indicated a significant main effect of the non-linear, continuous factor “Age”: $\chi^2(1) =$
20 59.84, $p < 0.001$, AIC shift = 57.8. The two fixed-effect predictors were also significant: $\chi^2(2) =$
21 116.33, $p < 0.001$, AIC shift = 112.3 for “SNR” and $\chi^2(1) = 193.01$, $p < 0.001$, AIC shift = 191.0
22 for “Task”. Further analyses indicated that, as far as the RTNH data were concerned, each of the
23 three “SNR” levels was significantly different from the others when subjected to post hoc

21

21

pairwise comparisons with Bonferroni corrections: 1 “quiet” versus “+3 dB” ($p < 0.001$), “quiet”
2 versus “-3 dB” ($p < 0.001$), as well as “+3 dB” versus “-3 dB” ($p < 0.001$). The mean RTNHs
3 from the two levels of “Task”, “animal” and “dangerous”, were also found to be significantly
4 different from each other in post hoc pairwise comparison ($p < 0.001$). Additionally, like we saw
5 with the WRSNH analysis, there was significant interaction between “Age” and “SNR” ($\chi^2(2) =$
6 6.82, $p < 0.05$, AIC shift = 2.8). Taken together, RTNH was significantly impacted by changes in
7 both background noise levels and depth of processing, a finding that echoes those reported in
8 Hsu et al. (2017), where a similar listening effort paradigm was administered to a smaller group
9 of NH school-age children. On the other hand, not only did categorization speed tend to increase
10 with age, the magnitude of change also tended to become progressively smaller with advancing

11 age due to the inverse exponential relationship between RTNH and “Age”. Moreover, the
12 magnitude of noise-related increase in RTNH decreased as one advanced in age (i.e., the
13 interaction between “Age” and “SNR”).

14 3.2.4 RTNH versus CELF

15 Outcome scores from the standardized CELF tests as conducted in our NH group offered an
16 interesting opportunity to study the links, if any, between listening effort and individual
17 cognitive characteristics. They also allowed one to consider the claim of categorization RT as a
18 valid measure of listening effort by comparing its outcomes directly with established external
19 measures. To this end, correlation analyses were carried out to search for any potential
20 relationship between averaged RTNH and four types of CELF raw scores: “Word Associations”
21 (WA), “Digit Recall – Forward only” (DRFw), “Digit Recall – Backward only” (DRBw), and
22 “Digit Recall – Forward and Backward” (DRTotal). Preliminary analyses revealed the non23
normality of the distribution of all four types of CELF scores. All of the CELF raw scores were

22

22

also found to have a significant correlation with participants’ age. Due to these factors, a partial
2 correlation method was selected to effectively control for the influence of age while conducting
3 correlation analysis between two interested variables. Additionally, Kendall's tau (τ) was chosen
4 as the correlation statistic for its robustness against non-Gaussian distribution of data points. The
5 result of the partial correlation procedure revealed that, out of the four CELF raw scores, WA (τ
6 = -0.26, $p < 0.001$), DRBw ($\tau = -0.20$, $p < 0.01$), as well as DRTotal ($\tau = -0.15$, $p < 0.05$) all
7 exhibited a significant, inverse correlation with averaged RTNH.

8 Since the CELF test package we used mainly targets children and teenagers as evidenced by its
9 provision of age-normed scores only up to and including 18 years of age, one may well question
10 whether the inclusion of adult CELF data could unfairly skew the correlation results in one
11 direction or another. To investigate this possibility, we removed the data from participants 19
12 years and older and re-ran the partial correlation analyses. Results suggested that both WA ($\tau = -$
13 0.25, $p < 0.01$) and DRBw ($\tau = -0.18$, $p < 0.05$) still exhibited a significant inverse correlation
14 with averaged RTNH when data from only NH children and teenagers were included. Further
15 analyses separating the RTs from the two categorization tasks confirmed that WA had significant
16 inverse correlations with both “animal”-RTNH ($\tau = -0.27$, $p < 0.001$) as well as “dangerous”-

17 RTNH ($\tau = -0.22$, $p < 0.01$) in children and teenagers. In other words, the following two
18 conditions seemed to correlate with a reduction in RT when NH children and teenagers were
19 asked to semantically categorize auditorily perceived words in quiet or noise during our listening
20 effort assessment: a) when one was better at retrieving information from long-term memory
21 stores as measured by the CELF “Word Associations” test (Kinsella-Ritter 2016); b) when one
22 was better at performing challenging mental tasks due to a high working memory capacity as
23 measured by the CELF “Digit Recall – Backward only” subtest. Moreover, this inverse

23

23

relationship between RT and the a) 1 forementioned condition “a” seemed to exist for different
2 levels of semantic processing complexity as represented by the two categorization tasks “animal”
3 and “dangerous” administered during our behavioural listening effort paradigm.

4 3.3 Interim summary of Experiment 1

5 The analyses of the WRSNH and RTNH data sets above helped us better understand the effects of
6 background noise and semantic processing levels when it came to the current listening effort
7 assessment paradigm. In short, the addition or increase of noise adversely affected word
8 recognition performance as well as lengthened categorization response time, while an increase in
9 the depth of processing demands lengthened categorization response time but had no significant
10 impact on word recognition performance. We are interested to see if these effects would still
11 hold true in a CI population. This led us to Experiment 2.

12 4 EXPERIMENT 2: LISTENING EFFORT IN SCHOOL-AGE CI USERS

13 4.1 Methods

14 4.1.1 Participants

15 The second phase of the current study was to apply the same listening effort assessment in a
16 group of children with CI's to assess the clinical viability of the paradigm in a fast-paced
17 regional cochlear implantation centre. To this end, 14 school-age CI users (age range = 5.70 to
18 14.42 years; mean age = 9.49 years) were successfully recruited and tested through collaboration
19 with the University Medical Center of Radboud University Nijmegen (Radboud UMC).
20 Inclusion criteria included full-time users of at least one cochlear implant for at least one year.
21 Table 1 gives an overview of the individual characteristics of each of these 14 participants
22 labelled with ID numbers “1” to “14”. Listening effort data were collected from all consenting

23 school-age CI users who satisfied the inclusion criteria as part of their regular annual follow-up
24

24

visit at the Radboud UMC audiology clinic 1 during an approximately four-month period. In this
2 way, our CI participants could be considered a form of “convenience sample”, which was
3 deemed appropriate given the pilot-study nature of Experiment 2. In addition to collecting data
4 with our experimental listening effort assessment, the Radboud UMC also supplied the latest
5 Woordbegripsquotiënt (“WBQ” in short, “Word Understanding Quotient” in English) they had
6 on file for each child. The WBQ is a standardized score signifying receptive language ability
7 derived directly from outcome scores on the Peabody Picture Vocabulary Test (PPVT). Similar
8 to the intelligence quotient (IQ), the WBQ is normally distributed score with a mean of 100 and a
9 standard deviation of 15. The only child with CI for whom the WBQ was missing was CI
10 participant #5. All procedures pertaining to data collection, data transmission, and patient
11 counselling were conducted in adherence to Radboud UMC ethical and privacy standards. Two
12 children with CI’s (n = 2) were able to focus and complete the “quiet” conditions (i.e., half of the
13 planned listening effort assessment) for about 10 minutes but not the “noisy” conditions due to
14 reported fatigue and decreased concentration. Nevertheless, data from the portion of the
15 assessment that these two children did complete were deemed reliable by the administering
16 clinicians and were therefore included in the final analysis.

17 4.1.2 Test procedures

18 The dual-task listening effort assessment used in the second part of this article was adapted from
19 the same paradigm developed and validated with NH participants. Aspects such as general
20 paradigm design, test order, training and instructions, speech materials, and technical equipment
21 were not changed to keep methodology as similar as possible to allow for meaningful outcome
22 comparisons between the NH and CI groups. Please refer to the detailed description of methods
23 and materials in NH sections above. Nevertheless, the assessment itself still needed to be

25

25

shortened to make it possible 1 to be included as part of the audiological test battery within the
2 time frame of a typical CI follow-up appointment. This was mainly achieved through the
3 exclusion of the noisiest “-3 dB SNR” condition, which would have been a much more

4 challenging listening condition for the CI participants compared to their NH counterparts and
5 potentially produce “floor effects” or “give-up performance” outcomes for both word recognition
6 and categorization tasks. Likewise, the CELF tests were not administered in the CI group, who
7 already received a variety of other language/cognitive capacity tests during their follow-up
8 appointments. Owing largely to logistical considerations (i.e., the availability of clinicians and
9 CI-specific equipment), for the CI group the listening effort assessment was carried out in a quiet
10 therapy room at the university hospital rather than at the participant’s own home. Due to the
11 supra-threshold nature of the listening task conducted, it was felt that a quiet, undisturbed room
12 situated in a university hospital and a quiet, undisturbed room situated in a participant’s home
13 both provided a similar acoustic environment as far as the current study was concerned. For
14 school-age CI users, words and noise were delivered through accessory cables attached directly
15 to the sound processors rather than through headphones. The output from a CI device when
16 either types of accessory cables (a regular type for unilateral CI and a “forked” type for bilateral
17 CIs) was utilized to route the input directly from the iPad to the sound processor was measured
18 and calibrated via simulation software before any data collection began.

19 4.2 Results

20 4.2.1 Comparing CI to NH

21 For word recognition performance, individual WRSCI points were plotted directly over individual
22 WRSNH data for each of the four conditions as shown in Figure 4 where data existed for both
23 groups, namely “animal” at “quiet”, “animal” at “+3 dB SNR”, “dangerous” at “quiet”, and
26

26

“dangerous” at “+3 dB SNR”. The 1 age dimension was also restricted to between 5 and 15 years,
2 which was the range in which we had CI data for. WRS, being a percentage-correct score,
3 exhibited extreme clustering at or near 100% (i.e., a ceiling effect) especially for the “quiet”
4 conditions. This coupled with the “absence” of any variance observed at several age points (i.e.,
5 when everyone scored 100%, then the SD became practically “0”; cf. Figure 1) prevented any
6 meaningful application of regression analysis on WRSNH to generate “predictive ranges”.
7 For word recognition performance, a score of 78% or above traditionally indicates anywhere
8 from “slight difficulty” to “excellent” performance, whereas a score of 52% or lower is
9 considered “very poor” performance (Schoepflin 2012). In our CI data set, 71.7% of all

10 individual recognition scores were above 78%, and none of our children with CI obtained a WRS
11 of or lower than 52%. From Figure 4, the WRSCI from participants 10 years or older were
12 generally comparable with the corresponding WRSNH values, whereas the WRSCI from younger
13 participants tended to be worse. Except for a couple of notable cases (e.g., participant “5”), the
14 general distribution of WRSCI was similar both between “quiet” and “+3 dB SNR” as well as
15 between “animal” and “dangerous”. Furthermore, a cursory examination of the interaction
16 between WRSCI and the available WBQ did not reveal any apparent relationship between a child
17 with CI’s word recognition performance and receptive language ability as far as our listening
18 effort paradigm was concerned.

19 In terms of the RTNH data set from Figure 3 in the preceding section, one sees that the significant
20 and systematic RTNH change with age especially between 5 and 15 years calls for the application
21 of robust inferential methods to create estimated confidence and prediction ranges of future RT
22 values that would allow meaningful comparison with the RTCI data. To this end, the same model
23 formulated above for the maximum likelihood analysis of RTNH was utilized in conjunction with
27

27

a data resampling technique called 1 “bootstrapping with replacements” accomplished through the
2 R package “merTools” version 0.4.1. Bootstrapping has been considered a gold standard in
3 deriving prediction intervals from mixed-effect regressions short of a full Bayesian approach
4 (Knowles & Frederick 2018). From the raw RTNH data set, we were then able to estimate the
5 “predicted means” (i.e., the central tendency of future observations), the “95% confidence
6 intervals” (i.e., how reliable are the predicted means if the same experiment was repeated many
7 times on the same population), and, most crucially for the purpose of the present study, the “95%
8 prediction intervals” (a range of possible values where 95% of all observations are predicted to
9 occur). Future observations falling outside of a given prediction interval are interpreted as
10 “outliers” or, in clinical terms, an outcome measure that is outside the norms. Figure 5 presents
11 these predicted values for the age range of 5 to 15 years with an overlay of individual RTCI and
12 WBQ data for comparison purposes. Similar to what was observed in WRSCI, there did not seem
13 to be any apparent relationship between whether an RTCI fell within the normal range (i.e., the
14 95% prediction interval) and the same child with CI’s receptive language ability as indicated by
15 their last WBQ score.

16 The homogeneity of WRSCI distribution previously discussed did not seem to carry over to the
17 response time data from the CI group, as shown in Figure 5. Comparing the patterns of RTCI
18 points from each of the four “SNR” + “Task” conditions, the biggest between-condition change
19 in RTCI appeared to be between “quiet-animal” and “+3 SNR-animal”. This observation
20 suggested that the lengthening of RT by the introduction of competing noise was more
21 pronounced for the “animal” task compared to “dangerous”, whereas SNR-related changes in RT
22 from the NH group showed no significant difference between the two categorization tasks. This
28

28

is reflected in the absence of a significant interaction between “SNR” and “Task” both when
2 analyzing the current RTNH data set as well as the results presented in Hsu et al. (2017).

3 4.2.2 Comparing within CI

4 To further investigate any performance difference among children with CI, Figure 6 was drawn
5 up plotting a CI participant’s own WRS against RT from the same condition. Each condition-plot
6 has dotted guidelines visually showing what is considered “good” versus “poor” performance”:
7 while the vertical WRS guideline was derived from literature (Schoepflin 2012), the horizontal
8 age-dependent RT guidelines were derived from the upper bounds of bootstrapped 95%
9 prediction interval first presented in Figure 5. Individual data points from CI participants were
10 sub-divided into two groups: “younger” children (those younger than the mean age of 9.49 years
11 at the time of testing) versus “older” children (those older than 9.49 years of age). A linear
12 correlation trend line was also included for visual reference purposes. Kendall's analysis of
13 correlation between WRSCI and RTCI in each of the experimental conditions revealed that only
14 the “animal-quiet” condition exhibited a significant inverse correlation ($\tau = -0.56$, $p < 0.05$). In
15 other words, the children with CI who were faster (i.e., shorter RT) at “animal” categorization
16 tended to score higher with word recognition when the words were presented without noise. The
17 other three “Task” x “SNR” conditions did not produce significant Kendall's tau’s, suggesting no
18 clear relationship between WRSCI and RTCI when words were presented in noise or when the
19 secondary task was the “dangerous” question.

20 The analysis of the NH data set above gave us the following three interim conclusions with
21 regard to the effects of background noise and semantic processing levels: (1) the addition or
22 increase of noise adversely affected word recognition performance, (2) the addition or increase

23 of noise lengthened categorization response time, and (3) increasing depth of processing

29

29

demands lengthened categorization response time. The extent to which these effects were also
2 true in the CI group was more obviously seen in the data from older children with CI (the red
3 triangles) when one examines the changes of WRSI and RTCI values relative to the conditions as
4 shown in Figure 6. Specifically, effect (1) was present in older children with CI for both
5 “animal” and “dangerous”. Effect (2) was only apparent in older children with CI when they
6 engaged the “animal” task, while effect (3) was seen in this group of participants when words
7 were presented in “quiet”. Changes in WRS and RT in the younger children with CI, on the other
8 hand, did not provide strong support for these three effects. This observation could be partially
9 explained by the wide variability of outcome measures in the younger CI group. Moreover, the
10 three youngest (i.e., less than 7 years old) participants from the CI group, namely “1”, “5”, and
11 “13” deserved further consideration. While participant “13” could complete all experimental
12 conditions without any issue, participants “1” and “5” could only complete the “quiet” conditions
13 before their attention was lost. Whereas participant “1” achieved average RTs that were visibly
14 longer than those of the other participants in both “animal” and “dangerous”, participant “5”
15 scored WRSs that were visibly lower in at least the “dangerous-quiet” condition. The inclusion
16 of these very young CI participants was in the interest of seeking the “lower age bound” of our
17 novel listening effort assessment. Not only did the test results from these three participants add
18 to the variability of data from the young CI group, the high test-abortion rate in this group (i.e., 2/3)
19 also called for further work to determine if the current listening effort paradigm remained
20 suitable for kindergarten-age children in a clinical setting.

30

30

1 5 DISCUSSION

2 5.1 Regarding the NH group

3 The present study strived to make important steps toward the clinical viability of a novel
4 behavioural assessment of listening effort based on a dual-task setup. We first evaluated the
5 effects of changing background noise level during presentation, varying the depth of processing

6 level during semantic categorization, and the age of participants on the outcome measures while
7 statistically controlling for individual differences in a large group of NH school-age children,
8 teenagers, and young adults. The inclusion of these age groups gave us a unique opportunity to
9 explore and trace the pattern of change of listening effort as measured by behavioural RT as
10 one's cognitive system develops and matures from early elementary-school years all the way to
11 young adulthood. The deployment of generalized mixed-effect models helped us control for
12 individual differences while investigating the effects of these interesting variables.

13 As described in the preceding section, the primary effects of background noise and processing
14 level on how well NH participants could repeat back words and the speed with which they could
15 categorize the perceived words were generally consistent with the findings reported in Hsu et al.
16 (2017). As seen in the analysis results from Experiment 1, chronological age had an impact on an
17 NH participant's categorization RTs whether his/her averaged eye-to-hand coordination speed
18 was accounted during statistical analyses of paradigm outcome measures or not. While both
19 WRSNH and RTNH were found to improve with age, mixed-effect regression analyses revealed
20 that there was a significant interaction between NH participants' age and the amount of masking
21 noise during word presentation for both outcome measures. These results suggested that, as a
22 person goes through life from early elementary-school years to young adulthood, their ability to
23 recognize words in background noise gradually improves and the accompanying listening effort

31

31

gradually decreases, and 1 these changes become the most pronounced when noise is at the highest
2 level. This finding resonates with previous work such as the one by Leibold and Buss (2013),
3 who found that adults had an advantage over children when asked to identify consonants in the
4 presence of background noise, especially when the noise was of an informational masker type
5 rather than a simpler energetic masker type. It must be mentioned that this interplay between
6 SNR and age for the word recognition part could be at least partially explained by the extreme
7 ceiling of WRSNH at 100% at the "quiet" conditions. That is, there was not much room for
8 WRSNH improvement at "quiet", hence the observed interaction effect. Nevertheless, there is
9 accumulating evidence that the auditory processing pathway matures in development and allows
10 for an increasingly better speech recognition performance in noise as a child grows up (e.g.,
11 Eisenberg et al. 2000). At higher ages, receptive language skills benefit from more efficient

12 bottom-up processes, which in turn preserve more finite cognitive resources such as working
13 memory for more advanced top-down tasks such as making “animal” or “dangerous”
14 distinctions. Given our working definition of listening effort as the amount of cognitive resources
15 required to understand perceived speech, it is of little surprise that listening effort level tends to
16 decrease at higher ages as manifested by faster RTNH.

17 Even after accounting for the effect of age, the significant inverse correlation between RTNH
18 and the ability to recall words based on a given category provides support for the notion that the
19 secondary (categorization) task of our listening effort paradigm requires some form of retrieval
20 of semantic knowledge from long-term memory stores. Additionally, working memory capacity
21 was also shown to exhibit an inverse relationship with categorization RT’s in at least NH school22
age children and teenagers after correcting for the effect of age. This is perhaps not a surprise
23 given that, from the onset of our investigation, we have come to define listening effort as a
32

32

phenomenon linked to the “limited cognitive resources” theories (Kahneman 1973; Pichora-
2 Fuller et al. 2016), and that an “intervention” of active “working memory power” is often
3 necessary when the task of understanding speech becomes challenging (Rönnberg et al. 2013).
4 All in all, the significant relationships between RTNH and several CELF test scores may be one of
5 the first steps in establishing the general validity of our assessment paradigm by comparing its
6 outcomes with independent, established measurements of language and cognitive abilities.

7 Since the current article deals primarily with the effects of age on RT as a behavioural measure
8 of listening effort, a fair point could be raised regarding whether our proposed method in fact
9 measures (changes in) listening effort after all. The NH data set as presented in Figure 3 shows
10 that children under 15 years of age tended to exhibit longer categorization RT than their older
11 counterparts in all conditions. The availability of the eye-to-hand coordination speed
12 measurements suggests that this difference could not be explained solely by the fact that older
13 teenagers and adults could move their arms faster. However, given the way our paradigm was set
14 up, it remains not entirely clear if this age difference in RTNH could be explained by a difference
15 in cognitive resource allocation capacity when the auditory system was being challenged (i.e.,
16 listening effort) or simply an age-related development of the general language/cognitive system
17 or even a mixture of these two factors. Future studies wishing to pursue this fine distinction

18 would be well advised to include a baseline measure of the categorization tasks without auditory
19 inputs to tease out the non-auditory language/cognitive level of each and every participant.

20 5.2 Regarding the CI group

21 This brings us to Experiment 2 where we applied the listening effort paradigm to a group of
22 school-age CI users. The resultant behavioural data from Experiment 2 were then compared to
23 the same metrics from those NH participants from Experiment 1 whose age fell in the age range
33

33

of our group of CI participants. The vast majority of these CI participants (12 out of 14) appeared
2 to possess good receptive language skills as evidenced by their records of above-average PPVT3
WBQ scores (see Table 1). Therefore, these two comparing groups, CI and NH, were not only
4 similar in age but also in receptive language ability. The loss of the “temporal fine structure”
5 when speech signals are processed through CI algorithms (Wouters et al. 2015) may
6 disadvantage CI recipients in the sense that they require more from the finite cognitive resources
7 during an everyday scenario of speech-and-language understanding. Therefore, one might
8 suspect that, when their cognitive system is further burdened with a top-down task of putting
9 words into categories, all school-age children with CI would generally need more time to do the
10 categorization compared to their NH counterparts. This, however, was not the case when the
11 current RTCI data set was plotted against predictive normative values derived from our RTNH
12 data. From Figure 5, while some children with CI indeed obtained longer RTCI values that were
13 outside the 95% prediction interval, others obtained a RTCI that was well within the normative
14 range, especially during the condition “animal-quiet”. One might rightly ask that, at this point, if
15 those that had slower RTCI tended to have poorer WRSCI, too? From the correlation analysis as
16 represented by Figure 6, a significant inverse relationship between a school-age CI user’s word
17 recognition performance and listening effort measure only existed at the “easiest” condition,
18 namely “animal-quiet”. As soon as the dual-task assessment’s difficulty was enhanced, either by
19 adding competing noise or by raising the depth of processing, no more significant correlation
20 could be observed between WRSCI and RTCI. The patterns with which WRSCI and RTCI changed
21 when either noise or processing level was varied also appeared to be different in different CI
22 participants. Note that any (non-)existence of significant relationship between WRSCI and RTCI
23 could be partially mediated by age-related maturation of the auditory processing pathway similar

34

34

to what has been discussed above with regard to the NH data set. Indeed, when we roughly divided our CI participants into younger children with CI versus older children with CI as shown in Figure 6, one might start to see how the age factor interacted with noise or processing levels in terms of word recognition or listening effort in the CI group. However, due to the small size of the CI sample, any such observations from the sub-division of the group could only be further analyzed and substantiated through future studies employing a significantly larger group of CI recipients.

5.3 Limitations

The home is a typically low-anxiety environment where any test-taker could be at their optimal emotional state when engaged in a behavioural assessment involving some degree of mental exertion. Due in part to the CI participants' nature being a "convenience sample", the heterogeneity of the testing environments between Experiments 1 and 2 (i.e., the NH data set was collected in a home setting while the CI data were collected in a university clinic) could therefore be problematic if the analyzed RTNH results were to be used as norms for future data points not collected at home. It is possible that future participants who are tested in, for example, a mainstream school environment would be at an increased risk of being (falsely) identified as having higher levels of listening effort simply because the normative data were collected in a home situation. On the flip side of the coin, it remains uncertain if testing a normal-functioning (per teachers' reports) child in a quiet and familiar room in their own school would indeed be much more stressful than conducting the same test at their own home. The presence of "semi-authority" figures such as teachers, audiologists, or speech-language pathologists could also "encourage" a child/patient to pay more attention and therefore do "better" on a behavioural listening effort assessment than being tested in a relaxed environment of their own home. It is

35

35

important for future scientific or clinical work to note the potential impacts of testing environments when attempting to apply the normative values reported in this paper. As far as Experiment 2 was concerned, we were not able to conduct the CELF cognitive capacity

4 tests on any of the CI participants due to the time constraints of real-life audiology appointments.
5 This was despite the fact that such test results could well serve as a proxy for someone's working
6 memory capacity and long-term memory retrieval speed. Future studies could gain insights by
7 investigating the effects of memory access speed, attention, and other central executive functions
8 on the outcome measures of a behavioural listening effort paradigm involving deep processing
9 tasks. Moreover, it should be noted that the absolute speech and noise levels varied to a small
10 extent between different SNR conditions in order to achieve a constant total output of 65 dB
11 SPL. For school-age CI users, any small change in the absolute speech level could potentially
12 mean a significant advantage or disadvantage when it comes to the amount of effort expended in
13 order to comprehend speech, whether in noise or in quiet. This technical feature, while some may
14 consider controversial from a methodological point of view, was implemented as a reflection of
15 our overriding concern for the safety of the participants with CI devices. Future work involving
16 active CI users should explore innovation ways of presenting materials using more realistic
17 speech and noise dB SPL levels while simultaneously ensuring the integrity and safety of their
18 participants' hearing systems.

19 5.4 Conclusion

20 The experiments described in this paper were designed and conducted with an overarching goal
21 of finding the right tool of measuring listening effort and thereby increasing the probability of
22 positive educational outcome for school-age children who rely daily on their CI devices for
23 learning and communication purposes. The proposed solution was a behavioural test of listening

36

36

effort that challenged the 1 auditory system through both bottom-up and top-down tasks and
2 thereby mimicked one aspect of learning (i.e., listen and make inferences) found in a typical
3 school environment. As a result of our investigation detailed above, we felt that the current
4 method had the potential to achieve such a goal due to the fact that it could identify those
5 children with slower categorization RTCI who did not necessarily had poorer WRSCI and vice
6 versa. The usefulness of the current listening effort assessment is further enhanced by the
7 availability of calculated normative ranges based on "background" data acquired from a
8 moderately large group of NH participants. Future research effort could bring the test to full
9 clinical viability by additional trials at a variety of clinical settings with different audiologists,

10 speech-language pathologists, and even clinical psychologists. A central issue of this endeavour
11 would be to determine the mixture of conditions in the final clinical setup to achieve an optimal
12 balance between efficiency (i.e., minimal testing time) and sensitivity (i.e., the ability to
13 distinguish those with heightened listening effort from those within the normal range). A
14 comprehensive roadmap that includes referral system and additional testing as well as
15 intervention options such as cognitive training and technological upgrades should also be in
16 place to ensure a maximum chance of improvement. Follow-up appointments should also include
17 the same listening effort assessment to evaluate the efficacy of the selected intervention
18 strategies.

19 6 ACKNOWLEDGEMENTS

20 We are especially grateful to Cochlear Technology Centre colleagues Stefan Lievens for
21 developing the core test software and Tobias Busch for assisting in the statistical modeling of the
22 NH data set. KU Leuven audiology students Fleur Vandeputte and Sarah Valvekens contributed
23 tremendously by collecting both NH and pilot CI data as part of their thesis work. KU Leuven
37

37

students Elien Vanluydt and Maud Lobel also helped greatly by collecting additional NH data.

2 Special thanks go to Leonie Goffree, Ingeborg Sanders, and Nanda Leijsten for administering the
3 listening effort test to their pediatric patients at Radboud University Medical Center in the
4 Netherlands.

5 Each of the five authors contributed significantly to this work: Benson Cheng-Lin Hsu designed
6 and implemented the experiment, analyzed the data, and wrote the paper; Filiep Vanpoucke and
7 Astrid van Wieringen provided scientific supervision as well as critical reviews during the
8 drafting and revision of the manuscript; Margreet Langereis and Ann Dierckx organized and
9 provided access to CI recipients, clinicians, and relevant clinical records.

10 This work was partially supported by the European Commission through the Marie Curie
11 Actions (FP7-PEOPLE-2013-ITN) “iCARE” Initial Training Network.

12 Part of the work was presented at the Ci2018 15th International Conference on Cochlear
13 Implants and Other Implantable Auditory Technology in Antwerp, Belgium, on June 28th, 2018.

14 The authors declare no other conflict of interest.

15 Address for correspondence: Benson Cheng-Lin Hsu, Research Group ExpORL, Department of

16 Neurosciences, KU Leuven, Herestraat 49 bus 721, 3000 Leuven, Belgium. Telephone: +32
17 15795532. E-mail: bhsu@cochlear.com.

18 7 REFERENCES

19 Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on*
20 *Automatic Control*, 19(6), 716-723.

21 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
22 Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

38

38

Bess, F. H., & Hornsby, B. W. (2014). Commentary: Listening 1 can be exhausting — Fatigue in
2 children and adults with hearing loss. *Ear and Hearing*, 35(6), 592-599.

3 beyerdynamic GmbH & Co. KG. (2014, July). DT 770 PRO: Dynamic Headphone [Technical
4 specifications]. Retrieved from north-america.beyerdynamic.com/amfile/file/download/
5 file_id/3878/product_id/2912

6 Bierer, J. A. (2007). Threshold and channel interaction in cochlear implant users: evaluation of
7 the tripolar electrode configuration. *The Journal of the Acoustical Society of America*,
8 121(3), 1642-1653.

9 Busch, T., Vanpoucke, F., & van Wieringen, A. (2017). Auditory environment across the life
10 span of cochlear implant users: insights from data logging. *Journal of Speech, Language,*
11 *and Hearing Research*, 60(5), 1362-1377.

12 Cartwright, K. B. (2002). Cognitive development and reading: The relation of reading-specific
13 multiple classification skill to reading comprehension in elementary school
14 children. *Journal of Educational Psychology*, 94(1), 56-63.

15 Choi, S., Lotto, A., Lewis, D., Hoover, B., & Stelmachowicz, P. (2008). Attentional modulation
16 of word recognition by children in a dual-task paradigm. *Journal of Speech, Language,*
17 *and Hearing Research*, 51(4), 1042-1054.

18 Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word
19 recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and*
20 *Hearing*, 37(1), 55-63.

21 Craik, F., & Simon, E. (1980). Age differences in memory: The roles of attention and depth of
22 processing. *New Directions in Memory and Aging*, 95-112.

39

39

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A 1 framework for memory research.

2 Journal of Verbal Learning and Verbal Behavior, 11(6), 671-684.

3 Crandell, C. C., & Smaldino, J. J. (1994). An Update of Classroom Acoustics for Children with

4 Hearing Impairment. Volta Review, 96(4), 291-306.

5 Degeest, S., Keppler, H., & Corthals, P. (2015). The Effect of Age on Listening Effort. Journal

6 of Speech, Language, and Hearing Research, 58(5), 1592-1600.

7 Downs, D. W., & Crum, M. A. (1978). Processing demands during auditory learning under

8 degraded listening conditions. Journal of Speech, Language, and Hearing

9 Research, 21(4), 702-714.

10 Eisenberg, L., Shannon, R., Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech

11 recognition with reduced spectral cues as a function of age. The Journal of the Acoustical

12 Society of America, 107(5), 2704-2710.

13 Eysenck, M. W., & Eysenck, M. C. (1979). Processing depth, elaboration of encoding, memory

14 stores, and expended processing capacity. Journal of Experimental Psychology: Human

15 Learning and Memory, 5(5), 472-484.

16 Gagné, J. P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a

17 dual-task paradigm: A review. Trends in Hearing, 21, 1-25.

18 Gosselin, A. P., & Gagné, J.-P. (2010). Use of a Dual-Task Paradigm to Measure Listening

19 Effort. Canadian Journal of Speech-Language Pathology and Audiology, 34, 43-51.

20 Grieco-Calub, T. M., Ward, K. M., & Brehm, L. (2017). Multitasking during degraded speech

21 recognition in school-age children. Trends in Hearing, 21, 1-14.

40

40

Gustafson, S., McCreery, 1 R., Hoover, B., Kopun, J. G., & Stelmachowicz, P. (2014). Listening

2 effort and perceived clarity for normal hearing children with the use of digital noise

3 reduction. Ear and Hearing, 35(2), 183-194.

4 Héту, R., Riverin, L., Lalande, N., Getty, L., & St-Cyr, C. (1988). Qualitative analysis of the

5 handicap associated with occupational hearing loss. British Journal of Audiology, 22(4),

6 251-264.

7 Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with
8 and without hearing loss. *Journal of Speech, Language, and Hearing Research*, 45(3),
9 573-584.

10 Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue
11 associated with sustained speech processing demands. *Ear and Hearing*, 34(5), 523-534.

12 Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech
13 as a measure for listening effort. *International Journal of Audiology*, 52(11), 753-761.

14 Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that
15 are typical of the school classroom. *International Journal of Audiology*, 49(12), 928-932.

16 Hsu, B. C. L., Vanpoucke, F., & van Wieringen, A. (2017). Listening Effort Through Depth of
17 Processing in School-Age Children. *Ear and Hearing*, 38(5), 568-576.

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

19 Kinsella-Ritter, A. (2016). All about Your CELF: CELF-P2, CELF-4 and CELF-5
20 Metalinguistics FAQs [PowerPoint slides]. Retrieved from Pearson Clinical's Australian
21 website: www.pearsonclinical.com.au

22 Knowles, J., & Frederick, C. (2018). Prediction Intervals from merMod Objects. Retrieved from
23 cran.r-project.org/web/packages/merTools/vignettes/Using_predictInterval.html

Kort, W., Schittekatte, M., & Compaan, E. (2010). CELF-4-NL | Test voor diagnose en evaluatie
2 van taalproblemen (3rd ed.). Amsterdam: Pearson.

3 Kral, A., & O'Donoghue, G. M. (2010). Profound deafness in childhood. *New England Journal*
4 of Medicine, 363(15), 1438-1450.

5 Leibold, L. J., & Buss, E. (2013). Children's identification of consonants in a speech-shaped
6 noise or a two-talker masker. *Journal of Speech, Language, and Hearing Research*, 56(4),
7 1144-1155.

8 McGarrigle, R., Gustafson, S. J., Hornsby, B. W., & Bess, F. H. (2019). Behavioral measures of
9 listening effort in school-age children: Examining the effects of signal-to-noise ratio,
10 hearing loss, and amplification. *Ear and Hearing*, 40(2), 381-392.

11 McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S.
12 (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of
13 Audiology Cognition in Hearing Special Interest Group "white paper". *International*
14 *Journal of Audiology*, 53(7), 433-440.

15 Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., ... & Kramer, S.
16 E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort:
17 A systematic review. *Ear and Hearing*, 38(3), 267-281.

18 Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant
19 simulations. *Journal of Speech, Language, and Hearing Research*, 56(4), 1075-1084.

20 Perreau, A. E., Wu, Y. H., Tatge, B., Irwin, D., & Corts, D. (2017). Listening effort measured in
21 adults with normal hearing and cochlear implants. *Journal of the American Academy of*
22 *Audiology*, 28(8), 685-697.

Piaget, J., & Inhelder, B. (2008). *The Psychology 1 of the Child* (H. Weaver, Trans.). New York:
2 Basic Books. (Original work published 1966)

3 Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L.
4 E., ... Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework
5 for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37(Suppl. 1), 5S-27S.

6 Picou, E., Charles, L., & Ricketts, T. (2017). Child–Adult Differences in Using Dual-Task
7 Paradigms to Measure Listening Effort. *American Journal of Audiology*, 26(2), 143-154.

8 Picou, E. M., & Ricketts, T. A. (2014). The effect of changing the secondary task in dual-task
9 paradigms for measuring listening effort. *Ear and Hearing*, 35(6), 611-622.

10 Pisoni, D. B., Kronenberger, W. G., Harris, M. S., & Moberly, A. C. (2017). Three challenges
11 for future research on cochlear implants. *World Journal of Otorhinolaryngology-Head and*
12 *Neck Surgery*, 3(4), 240-254.

13 Preisler, G., Tvingstedt, A.-L., & Ahlström, M. (2005). Interviews with deaf children about their
14 experiences using cochlear implants. *American Annals of the Deaf*, 150(3), 260-267.

15 De Raeve, L., & Lichtert, G. (2012). Changing Trends within the Population of Children who are
16 Deaf or Hard of Hearing in Flanders (Belgium): Effects of 12 Years of Universal Newborn
17 Hearing Screening, Early Intervention, and Early Cochlear Implantation. *Volta*
18 *Review*, 112(2), 131-148.

19 Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... & Rudner, M.
20 (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and
21 clinical advances. *Frontiers in Systems Neuroscience*, 7, Article 31.

Sarampalis, A., Kalluri, S., Edwards, B., & 1 Hafter, E. (2009). Objective measures of listening

2 effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and*
3 *Hearing Research*, 52(5), 1230-1240.

4 Schoepflin, J. R. (2012). Back to Basics: Speech Audiometry. Retrieved from
5 www.audiologyonline.com/articles/back-to-basics-speech-audiometry-6828

6 Steel, M. M., Papsin, B. C., & Gordon, K. A. (2015). Binaural fusion and listening effort in
7 children who use bilateral cochlear implants: A psychoacoustic and pupillometric
8 study. *PLoS ONE* 10(2): e0117611.doi:10.1371/journal.pone.0117611.

9 Steffens, T., Lesinski-Schiedat, A., Strutz, J., Aschendorff, A., Klenzner, T., Rühl, S., ... &
10 Lenarz, T. (2008). The benefits of sequential bilateral cochlear implantation for hearing11
impaired children. *Acta Oto-laryngologica*, 128(2), 164-176.

12 Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring
13 listening effort: Convergent validity, sensitivity, and links with cognitive and personality
14 measures. *Journal of Speech, Language, and Hearing Research*, 61(6), 1463-1486.

15 van Wieringen, A., & Wouters, J. (2015). What can we expect of normally-developing children
16 implanted at a young age with respect to their auditory, linguistic and cognitive skills?.
17 *Hearing Research*, 322, 171-179.

18 Venail, F., Vieu, A., Artieres, F., Mondain, M., & Uziel, A. (2010). Educational and employment
19 achievements in prelingually deaf children who receive cochlear implants. *Archives of*
20 *Otolaryngology–Head & Neck Surgery*, 136(4), 366-372.

21 Wålinder, R., Gunnarsson, K., Runeson, R., & Smedje, G. (2007). Physiological and
22 psychological stress reactions in relation to classroom noise. *Scandinavian Journal of*
23 *Work, Environment & Health*, 33(4), 260-266.

Ward, K. M., Shen, J., Souza, P. E., & Grieco-Calub, T. M. (2017). Age-Related Differences in
2 Listening Effort during Degraded Speech Recognition. *Ear and Hearing*, 38(1), 74–84.

3 Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: a remarkable past and a brilliant
4 future. *Hearing Research*, 242(1), 3-21.

5 Wouters, J., McDermott, H. J., & Francart, T. (2015). Sound coding in cochlear implants: From
6 electric pulses to hearing. *IEEE Signal Processing Magazine*, 32(2), 67-80.

7 Wu, Y.-H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric Functions of
8 Dual-Task Paradigms for Measuring Listening Effort. *Ear and Hearing*, 37(6), 660-670.