LISTENING EFFORT

The hidden costs and benefits of cochlear implants

Carina Pals

Publication of this thesis was financially supported by

School of Behavioural and Cognitive Neuroscience (BCN)

University of Groningen

Cochlear Ltd.

Advanced Bionics Benelux B.V.

MED-EL Deutschland GMBH

Oticon Medical – Because Sound Matters

Prof. Dr. Eelco Huizinga Stichting

EmiD audiologische apparatuur

Daleco Pharma B.V.

Cover art by Gonny Jannink, Kunstenaarscollectief Deep Yellow

Cover design by Carina Pals

Printed by & Lay Out by Gildeprint || Carina Pals

ISBN gedrukte versie: 978-94-6233-403-8

ISBN digitale versie: 978-94-6233-404-5

rijksuniversiteit
groningen

Listening Effort

The hidden costs and benefits of cochlear implants

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 26 oktober 2016 om 14.30 uur

door

Carina Pals

geboren op 30 oktober 1977
te Utrecht

Promotor
Prof. dr. D. Başkent

Copromotor
Dr. A. Sarampalis

Beoordelingscommissie
Prof. dr. S.E. Kramer
Prof. dr. M. Rudner
Prof. dr. L.C. Verbrugge

TABLE OF CONTENTS

CHAPTER 1

General introduction

Carina Pals

PREFACE

Imagine sitting in a crowded café listening to your friend telling you about their recent travels, while around you the other guests are chatting loudly, the music is playing, and the bartender is stacking beer glasses. You have to strain to pick your friend's voice out of the mixture of sounds, try hard not to get distracted by the woman with the loud voice at the table next to you, and rely on the context of the story to make out everything your friend says. While it may be possible to understand everything, it does require a considerable amount of *effort*. Situations such as these are not uncommon in daily life and are already quite mentally demanding for normal-hearing (NH) listeners. For hearing-impaired (HI) listeners or deaf people with a cochlear implant (CI) such noisy listening conditions can be even more challenging and effortful.

Some CI users anecdotally report avoiding settings such as described above, because the effort it takes to try and keep up with the conversation can leave them exhausted. Although the regained hearing ability after implantation significantly improves quality of life (e.g. Klop, Briaire, Stiggelbout, & Frijns, 2007; Vermeire et al., 2005), the listening effort, especially in more challenging listening conditions, and the resulting fatigue, can still influence the lives of CI users. This is for example reflected in the results of a survey by the Dutch society for the hearing impaired (Nederlandse vereniging voor slechthorenden, NVVS). The survey was sent out to all 567 known CI users among the members of the NVVS, about 50% of which responded. The results showed that a CI improved quality of life for 87% of the respondents, while fatigue was improved for only 49%, with 17% reporting increased fatigue after implantation (van Hardeveld, 2010). This hearing-related fatigue may be due to the effort required to interpret the incoming sound, i.e. *listening effort* (Hornsby, 2013). Hearing related strain and fatigue can have serious consequences, such as leading to increased sick-leave from work among HI individuals compared to NH employees (Kramer, Kapteyn, & Houtgast, 2006). Alleviating listening effort for CI users may therefore further improve quality of life. Unlike speech intelligibility, however, listening effort is not directly observable and at the outset of this project little research had addressed this topic in CI users. The research described in thesis therefore, delves into listening effort in CI users.

THEORETICAL BACKGROUND

In normal hearing, the hair cells in the inner ear (the cochlea) transform the incoming sound waves into a neural signal. By far the most common form of hearing loss results from damage to the hair cells or nerves, either congenitally or, for example, because of exposure to (sudden, loud, or prolonged) noise or aging (e.g. Angeli et al., 2005; Uus & Bamford, 2006). When the damage is severe and only few hair cells remain intact, acoustic amplification using conventional hearing aids no longer produces a usable neural signal, resulting in profound hearing impairment or deafness. If the auditory nerve is sufficiently healthy, then partial hearing may be restored by means of direct electric stimulation of the nerve via cochlear implantation.

A CI consists of a behind-the-ear processor, a transmitter worn on the head, attached with a magnet to the receiver that is embedded in the skull, and an electrode array that is inserted in the cochlea (see Figure 1). The processor mimics the hearing of the healthy ear using the tonotopic arrangement of the auditory nerve endings in the inner ear. The incoming acoustic signal is filtered into frequency bands and the envelopes extracted from each of these bands are used to modulate a series of electrical pulses. This electrical signal is then transmitted via the electrode array to the auditory nerve, thus bypassing the damaged hair cells and producing the sensation of hearing, in a way that approximates, but not quite replicates, normal hearing.

The current multiple electrode devices provide an auditory signal rich enough to allow speech communication without the visual aid of lip reading for many CI users (Loizou, 1998). Improved devices, speech processing strategies, surgical procedures, and selection for implantation candidacy have resulted in more and more CI users achieving very good speech intelligibility results (Blamey et al., 2013; Lazard et al., 2012). CI hearing, however, is not equivalent to NH. Limitations of the device, the peripheral auditory system, such as dead regions in the cochlea (i.e. regions of non-functional inner hair cells or nerves; Moore, 2004), and the transfer of the electrical signal from the electrode to the auditory nerve, result in a perceptually degraded signal compared to NH (Başkent, Gaudrain, Tamati, & Wagner, 2016). The most notable form of degradation of the auditory signal for CI users is the loss of frequency information, i.e. reduced spectral resolution of the signal. The loss of spectral resolution cannot be attributed to the limited number of electrodes alone. A number of factors further limit the effective use of the spectral information available in the electrical signal for CI users, such as auditory nerve survival and the way the electric current from one electrode spreads and stimulates a wide range of auditory nerve fibers, at times leading to cross-talk between distinct electrodes (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Başkent, & Wang, 2001; Fu, Shannon, & Wang, 1998; Stickney et al., 2006).

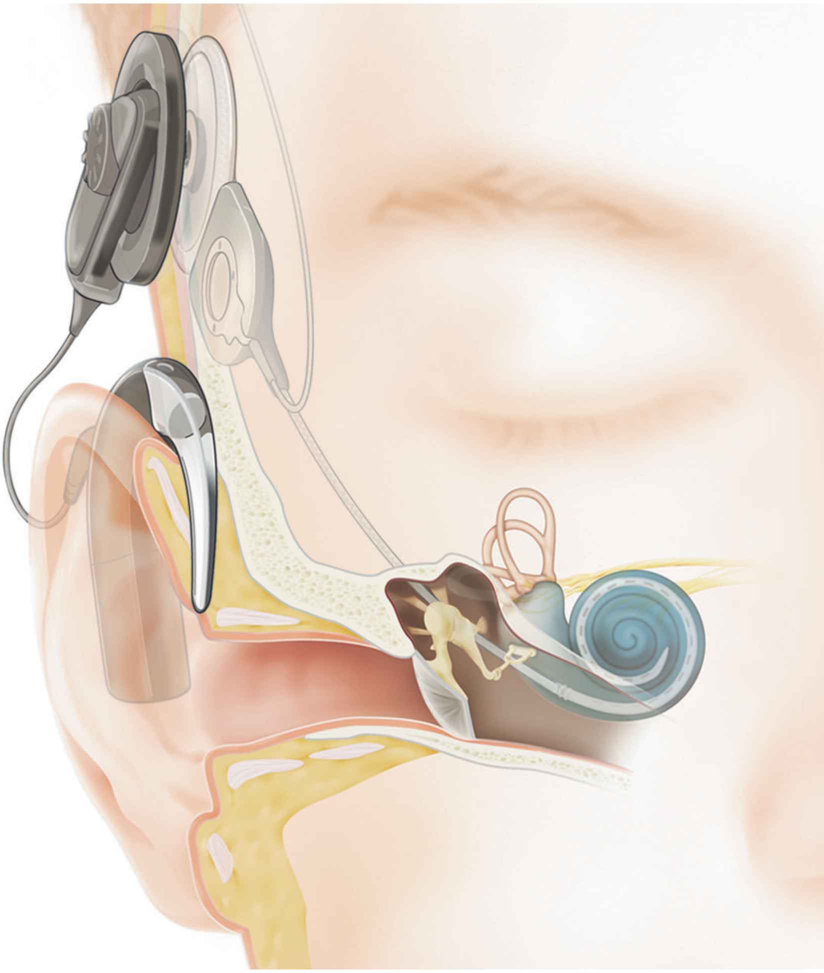


Figure 1: Illustration of a right ear with a cochlear implant. Hooked behind the ear (or pinna) is the speech processor, which connects to the transmitter that sits on the skull (dark gray). The transmitter is held in place by a magnet that connects it to the receiver embedded in the skull (translucent), which in turn connects to the electrode array inserted in the cochlea. Image Copyright Cochlear Limited ©

Reduced spectral resolution contributes to CI users' difficulty understanding speech in noise (Fu et al., 1998; Henry, Turner, & Behrens, 2005; Won, Drennan, & Rubinstein, 2007). Specifically, when listening to speech masked by modulated noise, CI users show reduced ability to benefit from the 'glimpses' of the speech signal that are available when the masker is less intense (Chatterjee, Peredo, Nelson, & Başkent, 2010; Fu & Nogaki, 2005; Nelson & Jin, 2004). Spectral resolution can be easily manipulated using a vocoder algorithm (Dudley, 1939), a method often used to simulate speech heard through a CI. Similar to CI processing, the acoustic signal is filtered into spectral bands, the envelopes are extracted, and then used to modulate noise-band carriers (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Studies examining the effect of spectral resolution using such CI simulations in NH listeners suggest that the reduced spectral resolution may lead to increased processing load (Schvartz, Chatterjee, & Gordon-Salant, 2008; Winn, Edwards, & Litovsky, 2015). This cognitive processing load of speech understanding is referred to as listening effort, which is defined in this thesis as the proportion of shared and limited cognitive resources that is used for the task of speech understanding.

The rest of this section will provide a more detailed background on cognitive resource capacity, cognitive processing load, cognitive processing of degraded speech, and how individual differences affect speech understanding and listening effort.

Limited capacity cognitive resources

The assumption that cognitive resources are limited and shared across tasks is commonly accepted, although how exactly is still a matter of debate. There is no consensus, for example, about whether resources are shared across modalities, or modality specific. On the one hand there is research that provides evidence for modality-free limitations, showing interference between visual and auditory attention (Dyson, Alain, & He, 2005) or memory (Morey & Cowan, 2004). While other research shows attentional interference only within the same modality, suggesting modality-specific resources (Duncan, Martens, & Ward, 1997; Dyson et al., 2005; Morey & Cowan, 2004). Yet other research shows that task interference depends both on modality and working memory load (Nijboer, Taatgen, Brands, Borst, & van Rijn, 2013). Another point of debate is whether the cognitive resource capacity limit is fixed or modulated by arousal, stress, or fatigue (Hockey, 1997; Kahneman, 1973). Kahneman (1973) suggested that increased arousal may temporarily increase cognitive resource capacity. Hockey (1997) described how the effects of increased workload and stress on performance can differ across individuals depending on coping strategies. This suggests that, while resources are assumed to be limited, how increased workload for one task affects performance is perhaps not quite straightforward.

Several models exist describing the limited cognitive resources either in terms of attentional resources (Broadbent, 1958; Kahneman, 1973) or a working memory system limited in both storage and processing capacity (Baddeley & Hitch, 1974). Baddeley and Hitch (1974) proposed a model consisting of a 'central executive' that coordinates the execution of complex tasks and the distribution of resources, and two short-term memory stores that allow for temporary storage and manipulation (such as active rehearsal to maintain the information) of auditory and visual information respectively. In a more recent version of the model, Baddeley (2000) introduced an extra component, "the episodic buffer". The episodic buffer operates outside the executive system and interacts with long-term memory to form chunks or 'episodes', thus facilitating more efficient use of storage and processing. Listening effort, then, depends on the processing requirements of the incoming speech signal, knowledge in long-term memory that can facilitate more efficient processing, and the cognitive resource capacity of the listener. Thus when the signal is degraded, it requires increased cognitive processing, which can be compensated by the listener's linguistic knowledge or knowledge of the topic of conversation (e,g, Sohoglu, Peelle, Carlyon, & Davis, 2012; Wingfield, 1996).

Baddeley's model aims to explain working memory and cognitive processing capacity in general. While it does include an auditory short-term memory component, it is not specifically tailored to explain the cognitive processing involved in language understanding. The Ease of Language Understanding (ELU) model (Rönnberg, Rudner, Foo, & Lunner, 2008; Rönnberg, 2003) proposes a mechanism to explain how language comprehension can lead to increased cognitive processing demand. In the ELU model, the (multimodal) sensory input is bound into (syllabic) phonological representations in the episodic buffer to be subsequently matched with phonological representations in long-term memory. If the incoming signal is clear and the appropriate representations are available in the listener's lexicon, i.e. the listener is proficient in the language spoken and familiar with the accent, the matching occurs immediately and implicitly, giving direct access to the associated lexical representations and their meaning. If the incoming signal is compromised (due to masking noise, or hearing loss for example), the phonological elements may fail to match existing representations (Mattys, Davis, Bradlow, & Scott, 2012). This mismatch will trigger a loop of explicit processing to either restore missing information and retry matching to representations in long-term memory, or, if no match can be found, to infer meaning (Rönnberg et al., 2013). The implicit and explicit processing components of the ELU model resemble the episodic buffer and the central executive system (including short term memory storage) from Baddeley's working memory model, respectively (Baddeley, 2000; Rönnberg et al., 2013, 2008).

The ELU model thus predicts that speech understanding in ideal listening conditions is fast, effortless, automatic, and independent of working memory capacity, while interpreting a degraded speech signal requires slow, effortful, and explicit cognitive processing and does depend on individual working memory capacity. In the next section, each part of this prediction will be examined and compared to the literature.

Cognitive processing in ideal vs. adverse listening conditions

For ideal listening conditions (i.e. speech clearly articulated by a healthy native speaker, unhindered by background noise or reverberation, and percieved by a normal-hearing, native listener), the ELU predicts fast, effortless, automatic speech understanding *independent of individual cognitive capacity.* This raises the question: can language be comprehended without relying on limited cognitive processing capacity? Caplan and Waters (1999) presented a systematic review of research on the role of working memory in language comprehension. They discuss a number of studies in healthy subjects under memory load, patients impaired in working memory capacity, and patients impaired in executive control. Each of these studies shows evidence that comprehension of simple, frequently used syntactic structures is not affected by memory load or reduced working memory capacity. Alzheimer's patients, for example, a population typically impaired in working memory and executive control, show normal speech comprehension when the task allows for implicit processing, but impaired comprehension when the task forces explicit processing (Kempler, Almor, Tyler, Andersen, & MacDonald, 1998).

The studies described above show support for fast, automatic, and effortless speech processing in ideal listening conditions. Is there any evidence in support of such effortless, automatic speech processing? This mechanism of effortless speech processing is referred to as implicit language processing in the ELU model. According to the model, implicit language processing relies on the rapid, automatic matching of sensory input with representations in long-term memory. Shtyrov, Kujala, and Pulvermüller (2010) suggest that strong memory traces for known words allow for automatic lexical activation. In an fMRI study they show that early lexical processing of known words does not to suffer from attentional load while processing of pseudo-words does, suggesting automatic lexical activation for known words, but not for pseudo-words (Shtyrov et al., 2010). This evidence suggests that under favorable conditions, language comprehension can indeed function automatically and independent of explicit attention and cognitive resources, and that this process depends on the automatic activation of long-term memory traces.

In adverse listening conditions on the other hand, the ELU predicts slow, effortful, explicit cognitive processing that does depend on individual working memory capacity. This raises the question: when does language comprehension require explicit cognitive processing? Research shows that for older listeners with age-related hearing loss and age-related decline in language processing, good speech comprehension depends on the recruitment of additional cognitive resources to compensate for these age-related deficits (Getzmann & Falkenstein, 2011), suggesting that older listeners may depend more on explicit processing for successful speech comprehension. This is supported by research that shows that older listeners rely increasingly on conscious rather than automatic processing (Alain, McDonald, Ostroff, & Schneider, 2004). Comprehension of spectrally degraded speech may similarly require explicit processing.

A recent neuroscience study shows that, while NH listeners appear to process ideal speech automatically and regardless of attention, the processing of spectrally degraded, yet highly intelligible, CI simulated speech, does require explicit attention (Wild et al., 2012). Interpreting spectrally degraded speech compared to clear speech results in increased activation in certain brain regions (including for example Broca's area) associated with grammar and speech motor control, suggesting that higher-order cognitive processes are recruited (Wild et al., 2012) or articulatory (motoric) representations of speech are accessed (Hervais-Adelman, Carlyon, Johnsrude, & Davis, 2012) to aid comprehension. This supports the prediction of the ELU model that loss of signal quality, such as the reduced spectral resolution for CI hearing or age-related hearing loss, increases the need for explicit cognitive processing for speech comprehension. The following section will go into further detail on the cognitive processing involved in speech comprehension.

Cognitive processing for speech comprehension

Even the comprehension of clear speech can require a certain amount of cognitive processing, for example, to disambiguate between words with similar onsets (e.g. Dahan & Tanenhaus, 2004; Salverda, Dahan, & McQueen, 2003), or to resolve complex syntactic structure (e.g. Piquado, Isaacowitz, & Wingfield, 2010). While lexical activation appears to be rapid and automatic (e.g. Aydelott & Bates, 2004; Shtyrov et al., 2010), the process of resolving lexical competition is slow and effortful (e.g. Aydelott & Bates, 2004; Wagner, Pals, de Blecourt, Sarampalis, & Baskent, 2015). Lexical decision can be facilitated by using prosodic cues, i.e. the pattern of pitch changes that, among other things, indicates the boundaries of words and sentences (e.g. Salverda et al., 2003; Wingfield, Lindfield, & Goodglass, 2000), or by using linguistic context (e.g. Dahan & Tanenhaus, 2004). Such strategies for facilitating lexical decision either reduce the number of lexical entries that are activated, or introduce a bias in favor of a subset of the activated lexical entries, thus reducing processing time and effort. Degradation of the speech signal, however, delays the semantic integration of context information, thus diminishing the benefit of context (Wagner et al., 2016). When lexical decision is no longer facilitated by context, lexical processing becomes slower and more effortful (Goy, Pelletier, Coletta, & Pichora-Fuller, 2013; Kuchinsky et al., 2012; Wagner et al., 2016).

Similarly, when speech is partially masked by noise, interpreting the incomplete parts of the bottom-up perceptual signal requires increased explicit processing. The perception of interrupted speech can be facilitated by expectations derived from linguistic context (Boothroyd & Nittrouer, 1988; Samuel, 1981a, 1981b). When the audible parts of the interrupted speech are spectrally degraded, however, the benefit of this top-down restoration mechanism is diminished (Başkent, 2012; Bhargava, Gaudrain, & Başkent, 2014; Chatterjee et al., 2010), suggesting that signal degradation impairs access to the available linguistic context. Evidence for reduced benefit of linguistic context has been shown for a range of different signal degradations including for uninterrupted, spectrally degraded CI simulated speech (Wagner et al., 2016), for energetically masked speech (Mattys, Brooks, & Cooke, 2009), as well as for time-compressed speech and for low-pass filtered speech (Aydelott & Bates, 2004;

Goy et al., 2013). Although the availability of sentence context has been shown to benefit perception of noise-vocoded speech (Sheldon, Pichora-Fuller, & Schneider, 2008), when this context information is contained in the degraded signal it may not be fully accessible for the listener's benefit. Strauß and colleagues (2013) suggest that such a reduced benefit of context may be explained from a limited cognitive resources perspective. Processing the incoming degraded speech signal uses cognitive resources that would otherwise be available to form hypotheses based on the linguistic context.

To summarize, even in ideal listening conditions, ambiguity and syntactic complexity inherent in language can introduce the need for increased cognitive processing. When, in addition to this, the signal is degraded, the need for cognitive processing is increased. Speech understanding can be facilitated by context, however, if this context information is embedded in the degraded signal itself its benefit seems to be reduced. The reduced access to context in a degraded speech signal can be explained from a limited cognitive resources perspective, which will be explained in more detail in the next section.

Individual cognitive capacity, speech comprehension, and listening effort

Speech comprehension and listening effort depend on the interaction between a number of factors. On the one hand, speech understanding and effort depend on factors related to the *speech signal*, such as the phonetic and the contextual cues available in the speech signal. When the incoming speech signal is degraded, increased cognitive processing is required for interpretation. However, contextual cues available in the sentence, discourse, or setting can help to form hypotheses about the meaning of the speech and thus facilitate more efficient processing. On the other hand, speech understanding and effort also depend on factors related to the *listener*, such as individual cognitive capacity and linguistic abilities (e.g. vocabulary, knowledge of grammar, common expressions). Larger cognitive capacity will allow the listener to allocate more resources to interpret the degraded speech, leading to better speech comprehension. Earlier in this introduction we have defined listening effort as the proportion of limited cognitive resources engaged in the task of speech understanding. This definition implies that, in otherwise equal listening situations, a listener's perceived listening effort depends on their individual cognitive capacity. And finally, better linguistic ability will allow the listener to make better use of context information to interpret a degraded signal, thus improving intelligibility.

The previous sections have described how signal quality affects the cognitive processing required for speech understanding. The next few paragraphs will address how individual cognitive and linguistic ability affect speech understanding and listening effort, starting with cognitive ability.

Research shows that better working memory capacity is indeed related to better speech-in-noise perception (Arehart, Souza, Baca, & Kates, 2013; Koelewijn, Zekveld, Festen, Rönnberg, & Kramer, 2012; Lunner, 2003; Rudner, Rönnberg, & Lunner, 2011), as well as the ability to benefit from contextual cues to facilitate better speech understanding (Zekveld, Rudner, Johnsrude, & Rönnberg, 2013). Memory constraints also limit the ability to benefit from downstream context, i.e. context that follows *after* the part of the speech that needs to be resolved (Wingfield, 1996). As mentioned before, listening effort is assumed to be relative to cognitive capacity. This is supported, for example, by research that shows that better working memory is related to less perceived effort for speech-in-noise (Rudner, Lunner, Behrens, Thorén, & Rönnberg, 2012), and low working memory capacity results in increased effort when interpreting speech that is inconsistent with the preceding context (Otten & Van Berkum, 2009). Working memory, or cognitive capacity is thus related to speech understanding and listening effort, and even the listeners' ability to use linguistic context. Linguistic ability (such as vocabulary, knowledge of grammar, etc.) can therefore be expected to predict the listener's ability to use context and thus speech comprehension and listening effort.

Research shows that linguistic ability is indeed associated with the ability to interpret interrupted speech (Benard, Mensink, & Başkent, 2014). How linguistic ability and the use of context relate to listening effort, however, is less clear. Research using pupillometry, a method that uses dilation of the pupil as a measure of cognitive effort, shows that listeners with larger vocabulary and better language processing skills are better able to utilize linguistic context to aid comprehension, although at the cost of *increased listening effort* as reflected by pupil dilation (Koelewijn et al., 2012; Zekveld, Kramer, & Festen, 2011). This suggests that accessing context information requires increased cognitive processing, rather than facilitating more efficient processing. Research on lexical access (the process of linking sound to meaning), on the other hand, suggests that the use of context does facilitate faster and *less effortful* lexical disambiguation, although this benefit is diminished if the speech carrying the context information is degraded (Goy et al., 2013; Wagner et al., 2016). The larger pupil response associated with better linguistic skills found by Koelewijn et al. (2012) also showed a positive correlation with a measure that reflects both working memory capacity and the ability to suppress irrelevant linguistic information. Hence, perhaps the larger pupil response reflects the suppression of irrelevant information while interpreting the masked speech, and not necessarily increased processing load related to the use of context information.

To summarize, better cognitive capacity is associated with better speech intelligibility, better ability to use context, and reduced listening effort. Similarly, better linguistic ability improves speech perception in noise and the ability to use context. The use of context information may require increased effort to process the context information, while on the other hand relieving effort for the interpretation of subsequent speech.

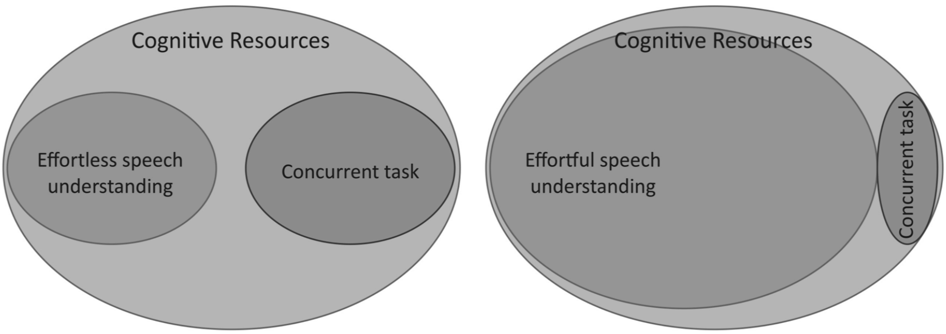


Figure 2: Cognitive resources and the interaction between the task demand of speech understanding and resources available for a concurrent task. So long as the cognitive resources required for the task of speech understanding do not exceed the available resources, full intelligibility can be achieved (left panel), however, the more resources are needed for speech understanding, the fewer resources will be available for concurrent tasks (right panel).

Consequences of effortful listening

In the previous section, the effects of individual cognitive capacity and signal quality on speech understanding and listening effort have been discussed. However, this is not the complete story: effortful listening in turn can also affect cognitive processes. The increased cognitive processing load for speech understanding due to a degraded signal reduces the cognitive resources available for simultaneous tasks or downstream processing of the speech message (e.g. Tun, McCoy, & Wingfield, 2009). As long as the processing demand of speech comprehension does not exceed the available resources, full intelligibility can be reached (see Figure 1). Words heard in noise, for example, while they can be repeated accurately at the moment they are heard, are later recalled less accurately than words heard without interfering noise (Rabbitt, 1966). Rabbitt suggests that this may be due to the effort required to interpret the speech in noise, which reduces the cognitive resources available for committing the words to memory. This effect of listening effort on memory may, perhaps in part, explain the apparent forgetfulness associated with old age. As age-related hearing loss increases, listening effort and the resulting reduction in cognitive resources available for concurrent tasks leads to difficulty remembering even the speech that was understood correctly (McCoy et al., 2005; Pichora-Fuller, Schneider, & Daneman, 1995; Rabbitt, 1991).

In addition to effects on memory, the slower explicit processing of degraded speech can reduce the ability to switch selective attention from one speaker to another (Shinn-Cunningham & Best, 2008). The reduced ability to benefit from context information for degraded speech may be due to the longer processing time for effortful speech comprehension (Wagner et al., 2016), or due to reduced cognitive resources available to form hypotheses based on context (Strauß et al., 2013). High listening effort can thus become a vicious circle; increased listening effort limits the listener's ability to use linguistic context to help interpret the next segment of the discourse as the conversation continues, which then increases the need to recruit yet additional cognitive processes to aid understanding. This recruitment of additional cognitive processes requires conscious, explicit attention and increases listening effort. This increased listening effort, in turn, can reduce the cognitive resources available for concurrent tasks (Sarampalis, Kalluri, Edwards, & Hafter, 2009), lead to slower speech comprehension (Mattys & Wiget, 2011; Wagner et al., 2016), and fatigue (Hornsby, 2013).

In summary, when listening to a degraded signal, the listener may still be able to fully understand the speech. However, maintaining speech intelligibility may require increased cognitive processing, i.e. increased listening effort. The increased processing load reduces the cognitive resources available for concurrent tasks and can increase the processing time required to decode the meaning of the speech signal. Thus, increased listening effort can, for example, adversely affect memory for the speech that was heard, lead to difficulties switching attention between speakers, or reduce the effective use of linguistic context presence in the speech. When effortful listening is sustained for a longer period of time it can ultimately lead to fatigue. Thus, even if the effects of a degraded speech signal are not directly apparent from reduced intelligibility, it may lead to increased listening effort which can have a number of undesirable consequences for the listener.

Summary

The neural signal resulting form speech delivered by a cochlear implant is degraded compared to normal hearing, most notably in terms of spectral resolution. Compared to a clear signal, interpreting a degraded signal requires increased cognitive processing, resulting in increased cognitive load. In the context of speech understanding, this increased cognitive load is referred to as *listening effort*. According to the Ease of Language Understanding (ELU) model, when the incoming speech signal is degraded, elements of the speech input may fail to match phonological representations, resolving these mismatches requires explicit cognitive processing thus increasing listening effort. Cognitive processes and strategies that can be called upon to aid the comprehension of degraded speech include articulatory representations for speech production, using prosody and pitch cues, knowledge of grammar and vocabulary, and situational or linguistic context. The effective use of these strategies for understanding degraded speech depends on the listener's cognitive capacity, as well as linguistic ability. Through increased cognitive processing the listener can, to some extent, maintain speech understanding. However, the increased listening effort limits the cognitive resources available for simultaneous tasks or further processing of the speech message and can ultimately lead to fatigue.

All in all, the literature suggests that speech understanding may be effortful for CI users, which can have undesirable consequences for the listener both immediately and over a longer period of time. The aim of this thesis is, therefore, to investigate speech comprehension and listening effort in CI users; how listening effort can be measured, if it changes independently of speech intelligibility, and factors that affect intelligibility and listening effort.

THIS THESIS

This thesis aims to systematically investigate listening effort with cochlear implant (CI) hearing. A series of experiments, with NH participants using CI simulations and with CI users, aim to address the following questions. Is CI-mediated speech more effortful to understand than normal hearing (NH)? Do changes in listening effort occur when no changes in speech intelligibility are observed? Do changes in the spectral resolution of CI hearing affect speech understanding and listening effort? Does wearing a hearing aid to complement the CI signal, as in electric acoustic simulation (EAS), reduce listening effort? In order to address these research questions, first of all, a reliable measure of listening effort is needed.

Measuring listening effort

In clinical settings the quality of fit of a CI is often assessed by pure-tone and speech audiometry, measuring hearing thresholds and speech intelligibility respectively. However, effortful speech understanding does not inherently mean loss of intelligibility. When listening effort is high for a longer period of time it can lead to fatigue. For some hearing-impaired listeners or CI users, this mental fatigue can be a serious problem, leading to increased sick-leave from work compared to NH employees (Kramer et al., 2006). Some CI participants in the study described in the final chapter anecdotally reported that hearing-related fatigue was the main reason for them to decide to work part-time or quit working altogether. These CI users did not perform particularly poorly, on the contrary, they were selected for their exceptionally high speech recognition scores in clinical tests. This suggests that reduced listening effort can mean a significant improvement in quality of life for CI users such as these. The quality of CI-mediated communication is thus not only reflected by the proportion of speech that can be understood, but also by the amount of effort invested to reach this level of understanding. Measures of listening effort can therefore complement the traditional measures of speech intelligibility (e.g. Gosselin & Gagné, 2010; Houben, van Doorn-Bierman, & Dreschler, 2012). An easy to administer and reliable method for measuring listening effort could be a valuable tool for use in hearing research as well as in clinical settings.

A wide variety of methods for measuring listening effort have been used in research, ranging from subjective rating scales to behavioral and physiological measures, each with its own advantages and disadvantages. Subjective rating scales are easy to administer, however, comparisons between individuals are difficult, since people may differ in what they consider ‘normal effort’ to be, or in their interpretation of effort altogether (McGarrigle et al., 2014). Therefore, objective measures are preferred. Physiological measures have proven to be a promising objective measures of listening effort, however, these typically require expensive equipment and the procedures can be cumbersome. These drawbacks are easily overcome in research settings, however, they make physiological measures less suitable for use in the clinic. Most behavioral tests do not rely on expensive equipment and may therefore be suitable candidates for an objective measure of listening effort that is widely applicable in any setting.

In order to explore measures that can potentially be used for routine fitting in clinical settings as well as for research purposes, behavioral measures for listening effort are used in this thesis. The main method for measuring listening effort was the dual-task paradigm, which will be explained in more detail below. In each of the individual chapters, this method was complemented with another, simpler measure of listening effort. In the first two chapters a subjective rating scale was used. In Chapter 3 and 4, the dual-task measure was complemented by two different simple response time measures; a verbal response time measure in Chapter 3, and a sentence verification task in Chapter 4. These measures will also be introduced briefly below.

Dual-task paradigm

A long established method for quantifying cognitive effort is the dual-task paradigm (e.g. Broadbent, 1958; Kahneman, 1973; Tyler, Hertel, McCallum, & Ellis, 1979). The dual-task paradigm is based on the limited cognitive capacity assumption. In a dual-task paradigm two tasks, one primary and one secondary, are performed simultaneously and compete for the limited cognitive resources. Participants are instructed to prioritize the primary task, while still performing the secondary task as best they can. As the primary task becomes more effortful, fewer resources are available for the secondary task and reduced performance on the secondary task, therefore, reflects increased effort on the primary task (Wu, Stangl, Zhang, Perkins, & Eilers, 2016).

The dual-task paradigm has been used in hearing research to quantify listening effort in a number of studies (e.g. Gosselin & Gagné, 2010; Sarampalis et al., 2009). Sarampalis and colleagues (2009), for example, investigated the effect of hearing-aid-like noise reduction on speech understanding in background noise in NH listeners. The effects on speech intelligibility and listening effort were investigated in two dual-task experiments. For both experiments, the primary task was to listen to sentences or words and repeat back what was heard. In one experiment, the secondary task was to hold words in memory, and in the other experiment, a visual response-time (RT) task. Both these experiments showed that at low signal-to-noise ratios (SNRs), noise reduction did not improve intelligibility but did improve performance on the secondary task. In the noise-reduction conditions, words were recalled better and responses to the visual RT task were faster.

In this thesis a similar dual-task paradigm is used to measure listening effort. The primary task was to listen to conversational sentences and repeat back what was heard. The secondary task was a visual response-time task.

The dual-task paradigm shows promise as a measure of listening effort in research settings. For use in a clinical setting, on the other hand, it may not be the method of choice. The procedure of performing two tasks simultaneously may be difficult to explain to certain populations, such as children or the elderly. In addition to this, the balance between the primary and secondary task difficulty has to be carefully chosen to have the right amount of interaction, but this may greatly depend on individual patients' cognitive, and auditory, abilities. In a research setting, when testing a group of NH young adults of similar age and educational level, e.g. first year Psychology students, this does not pose much of a problem. In a clinical setting, however, one may need to test patients of a wide range of ages, and from a wide range of social-, and educational backgrounds. Two tasks that are well balanced for one group of patients (i.e. showing interference in secondary task performance when the primary task becomes more effortful) may be too easy or too difficult for another group (thus resulting in floor or ceiling performance and showing no changes in secondary task performance).

Verbal response-time task

A measure that may be more suited for use in clinical settings is a verbal response time (VRT) to sentences (Gatehouse & Gordon, 1990). A number of studies have shown that when listening to degraded speech compared to clear speech lexical access and lexical decision is slower and delayed (Goy et al., 2013; Kuchinsky et al., 2012; Wagner et al., 2016). The ELU model predicts that degraded speech input will result in mismatches with phonological representations in long-term memory, and thus require slow, effortful, explicit cognitive processing. Such a mismatch may result in more lexical candidates being activated, and thus increased lexical competition, which has been proposed to be time-consuming and effortful to resolve (e.g. Aydelott & Bates, 2004; Wagner, Pals, de Blecourt, Sarampalis, & Baskent, 2015).

Based on this, the VRTs are expected to be longer for degraded speech perception, and within-subject changes in VRT are assumed to reflect changes in listening effort.

The VRT task is simple: the participant is instructed to listen to sentences and repeat them out loud, hence it can be implemented as part of a clinical speech intelligibility task. The VRT is defined as the time between the offset of the sentence stimulus and the onset of the verbal response. These measurements are easy to acquire in a clinical setting and the task is easy to explain to the patient.

Sentence verification task

Another potential candidate as a clinical measure is the sentence verification task (Adank & Janse, 2009; Baer, Moore, & Gatehouse, 1993). The task is to listen to sentences that are either unmistakably true or false/nonsense, and press a button as soon as possible to indicate whether the sentence was true or false. This test is again both easy to implement and the task is easy to explain. Similar to the VRT, we assume the response time to this task to reflect listening effort as effortful cognitive processing is time consuming. However, the difference with the VRT tasks is that the sentence verification task requires the participant to *comprehend* and reason about the meaning of the sentence, whereas the VRT allows the listener to repeat the sentence as soon as each word was heard correctly, though not necessarily comprehended.

Chapter outline

The aim of this thesis is to examine how CI processing affects listening effort. First this will be examined in normal-hearing participants listening to CI simulated speech and finally, in Chapter 5, in CI users.

Chapter 2

How does spectral resolution of CI simulated speech affect speech intelligibility and listening effort?
In Chapter 2, listening effort is measured using the dual-task paradigm. CI hearing is simulated using a noise-band vocoder, and the spectral resolution is manipulated by varying the number of spectral bands of the simulations. The effect of spectral resolution on intelligibility is already well established (e.g. Fishman et al., 1997; Friesen et al., 2001), and the

conditions are chosen such that a number of conditions provided enough spectral resolution to reach full intelligibility. Does listening effort change when intelligibility is near or at ceiling? The study examines how changes in spectral resolution affect the outcomes of a speech task, the dual-task measure of effort, and a subjective measure of effort.

Chapter 3

How does providing low frequency sound to complement CI simulated speech affect speech intelligibility and listening effort?
In Chapter 3 the same dual-task paradigm and subjective scale as in the previous chapter are used to measure listening effort. The CI simulated conditions are chosen for near ceiling intelligibility, and are complemented with either 300 Hz or 600 Hz low pass filtered speech (based on Qin & Oxenham, 2006), to simulate acoustic input from residual hearing.

Chapter 4

This chapter introduces a new simple and straightforward behavioral method for measuring listening effort, verbal response times; the time it takes to start repeating a sentence after hearing it. The dual-task paradigm from before and the verbal response times are compared for their sensitivity to the presence of masking noise, noise type, and noise level.

Chapter 5

Chapter 5, finally, returns to the question of Chapter 1; how do changes in spectral resolution affect intelligibility and listening effort?
Chapter 5 examines how spectral resolution affects listening effort in CI users, manipulating spectral resolution by changing the number of active electrodes of the CI. In addition to intelligibility and listening effort this study addresses an extra question; how does spectral resolution affect speech *comprehension*. Comprehension requires further cognitive processing than plain speech perception, and may therefore reflect both speech perception and cognitive processing requirement in one measure. The same dual-task paradigm as in the previous chapters is again used in this study, as well as a sentence verification task that serves as a measure of comprehension and processing speed.

CHAPTER 2

Listening Effort with Cochlear Implant Simulations

Carina Pals, Anastasios Sarampalis, Deniz Başkent

This chapter is published in: Journal of Speech, Language and Hearing Research

2013 | Volume 56 | Issue 4 | Pages 1075-84

ABSTRACT

Purpose. Fitting a cochlear implant (CI) for optimal speech perception does not necessarily minimize listening effort. This study aims to show that listening effort may change between CI processing conditions for which speech intelligibility remains constant.

Methods. Nineteen normal-hearing participants listened to CI simulations with varying numbers of spectral channels. A dual-task paradigm combining an intelligibility task with either a linguistic or a non-linguistic visual response-time (RT) task measured intelligibility and listening effort. The simultaneously-performed tasks compete for limited cognitive resources; changes in effort associated with the intelligibility task are reflected in changes in RT on the visual task. A separate self-report scale provided a subjective measure of listening effort.

Results. All measures showed significant improvements with increasing spectral resolution up to 6 channels. However, only the RT measure of listening effort continued improving up to 8 channels. The effects were stronger for RTs recorded during listening than for RTs recorded between listening.

Conclusion. The results suggest that listening effort decreases with increased spectral resolution. Moreover, these improvements are best reflected in objective measures of listening effort such as RTs on a secondary task, rather than intelligibility scores or subjective effort measures.

Keywords: Cochlear Implants, Computer Simulation, Listening Effort, Dual Task

INTRODUCTION

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

Research on effort in general is based on the historical work of Broadbent (1958), Baddeley & Hitch (1974), and Kahneman (1973), each of whom proposed a shared, limited cognitive resource (later commonly referred to as working memory) that can be allocated to various tasks, as necessary. A more recent version of Baddeley's theory proposes a phonological loop for storing and manipulating incoming auditory information, a visuospatial sketchpad for visual information, an episodic buffer which stores and retrieves information from long term memory, and a central executive which coordinates the execution of complex tasks (Baddeley, 2012). An effortful task requires a large proportion of the resources relevant to the task or a considerable involvement of the central executive, or both. Listening effort can then be defined as the proportion of limited cognitive resources engaged in interpreting the incoming auditory signal. It has been suggested that the presence of noise or distortions in a speech signal increases cognitive demand and thus listening effort (Schneider & Pichora-Fuller, 2000; Stenfelt & Rönnberg, 2009). Spectral degradation of a speech signal, such as in CI processing or CI simulations, has been shown to tax top-down cognitive processes involved in speech perception (Başkent, 2012; Chatterjee, Peredo, Nelson, & Başkent, 2010). Thus, we argue that especially when the fitting of a CI is less than optimal, interpreting the impoverished signal requires substantial cognitive resources, making listening for CI users effortful.

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

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

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

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

METHOD

Participants

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

Speech stimuli

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

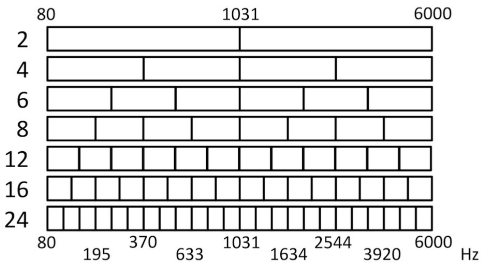


Figure 1. Graphical representation of the frequency bands used in the CI simulations. The vertical axis shows the number of bands, and the horizontal axis shows some of the -3 dB cut-off frequencies (based on the Greenwood formula).

The vocoder processing was implemented as follows: first the original acoustical signal was separated into a number of spectral bands (the analysis bands) as determined by the experimental condition, using 6th order Butterworth band-pass filters. From each analysis band the slow-changing envelope was extracted by means of half-wave rectification and filtering with a 3rd order low-pass Butterworth filter with -3 dB cut-off frequency of 160 Hz. A set of noise-band carriers (the synthesis bands) were constructed by similarly dividing white noise into spectral bands using 6th order Butterworth band-pass filters. For this experiment, the center frequencies and bandwidths of the analysis bands were the same as those of the synthesis bands in order to simulate matching frequency-to-place mapping of the CI electrode array (Başkent & Shannon, 2007; Greenwood, 1990). The CI simulations were then constructed by modulating each synthesis band with the envelope extracted from the corresponding analysis band and then adding these modulated noise bands together to form the final stimulus.

Visual stimuli

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

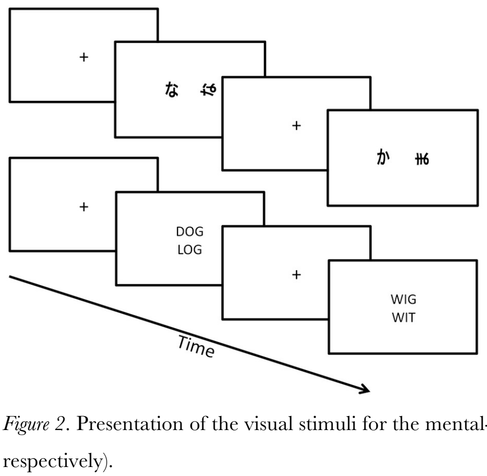


Figure 2. Presentation of the visual stimuli for the mental-rotation and rhyme-judgment task (upper and lower half, respectively).

Equipment

The participants were seated in a soundproof booth, in front of a wall-mounted computer screen at approximately 50 cm distance. A computer program, implemented using the Psychophysics Toolbox Version 3 for MATLAB and run on an Apple Macintosh computer, coordinated the presentation of both the auditory stimuli for the primary task and the visual stimuli for the secondary task. The verbal responses on the primary listening task were recorded for later scoring on a PalmTrack 24-bit digital audio recorder of Alesis, L.P. (Rhode Island, USA). The key-press responses and the RTs on the secondary task were automatically logged by the experimental program. The digital audio stimuli were sent via the AudioFire 4 external soundcard of Echo Digital Audio Corporation (California, USA) to the open-back HD600 headphones of Sennheiser electronic GmbH & Co. KG (Germany), to be presented to the participant diotically. The participants were instructed to adjust the volume to a comfortably loud, clearly audible level, within the range of 65 – 75 dB SPL. The calibration was done using the processed stimuli, measuring root-mean-square sound pressure with integration time-constant of 1 second.

Procedure

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

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

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

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

Subjective rating scale. For subjective assessment of listening effort, the NASA Task Load Index (TLX) was used (Hart & Staveland, 1988). The NASA TLX is a multi-dimensional scale which measures a range of aspects contributing to perceived work load; mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). Each dimension is rated on a visual analog scale and the final score is the weighted mean of the scores from the different dimensions. The weights 0 to 5 are assigned to each of the dimensons after the experiment by means of pair-wise comparison. For all possible pairs of dimensions the participants are asked to indicate which of two contributed most to the overall workload of the tasks. This procedure of weighting the ratings is designed to reduce inter-subject variability due to differences in individual interpretation of workload and its factors.

RESULTS

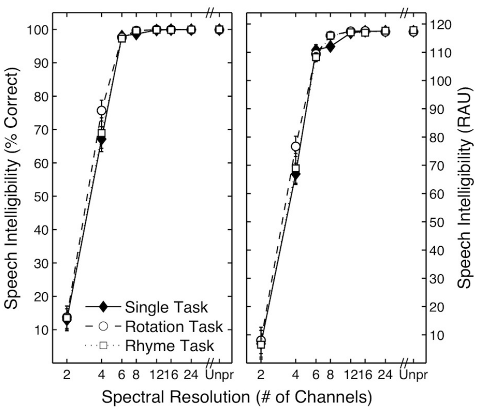


Figure 3. Average speech intelligibility scores in percent correct (left panel) and in rationalized arcsine units (RAU, right panel), as a function of spectral resolution. The different lines with open symbols show results for the two dual-task setups, the solid line with filled symbols for single task.

The average speech intelligibility scores are depicted in Figure 3. In each panel, the scores are plotted separately for listening task only, listening task combined with rotation task, and listening task combined with rhyme-judgment task. In the left panel of 3 the intelligibility scores are plotted in units of percentage correct, as a function of spectral resolution. In the right panel of 3 the scores are plotted in Rationalized Arcsine Units (RAU; Studebaker, 1985), as a function of spectral resolution. The conversion to RAU was performed to allow a closer examination of the effects near ceiling; RAU scores are easier to interpret since, unlike with proportional scores, the variance is independent of the mean, thus differences in percentage scores on different parts of the scale (for example the difference between 50% and 60% is not comparable to the difference between 90% and 100%) are not comparable, while differences in RAU are. Since the maximum possible value in RAU depends on the number of items in a task, the RAU scores were calculated based on an average number of words per list (80 words) and a proportion of words repeated correctly for each task. This ensures that a score of 100% correct always corresponds to the same RAU value, in this case, 117.83. The left panel of 3 shows that, in terms of percentage correct, speech intelligibility appears to reach a plateau at 6 channels. The right panel shows that there might still be some improvement in intelligibility between 6 and 8 channels. To examine these effects and differences between percentage correct and RAU scores further analyses were carried out on both sets of scores.

Two repeated-measures ANOVAs were performed with task and spectral resolution as factors (with 3 and 8 levels, respectively), one on the percentage correct scores and one on the RAU scores. Both ANOVAs showed a significant main effect of spectral resolution (percentage correct: $F(7, 126)=396.84$, $p<0.001$; RAU: $F(7, 126)=412.89$, $p<0.001$) and a significant interaction between task and spectral resolution (percent correct: $F(14, 252)= 2.11$, $p=0.012$; RAU: $F(14, 252)= 1.85$, $p=0.033$). A post-hoc analysis using Tukey's Honestly Significant Difference (HSD) indicated that the sole cause of the significant interaction was a significant difference in intelligibility between the mental-rotation task and the single-task for the 4-channel condition, while there was no significant difference between the single task and the rhyme-judgment task. To confirm that there was no difference in performance between the two dual-tasks, two-way ANOVAs were performed for these two tasks over the eight levels of spectral resolution, again for both the percentage correct scores and the RAU scores. These ANOVAs showed no significant interaction between task and processing. From this we conclude that there is no significant difference between the two dual tasks in terms of speech intelligibility. Therefore, the data for the two dual tasks were then grouped together to examine the differences between listening conditions. As expected, speech intelligibility with 2 channels was very low, about 13% of the words identified correctly. Increasing the number of channels to 4 provided a dramatic improvement in intelligibility; participants scored on average 70% correct. For 6 channels speech intelligibility was near perfect, on average 98% correct. The significance of these differences was determined using Tukey's HSD, which showed that there are significant improvements in intelligibility from 2 to 6 channels of CI simulations, and that further increases in spectral resolution resulted in no significant improvement in intelligibility. This was true for both the percentage correct scores and the RAU scores.

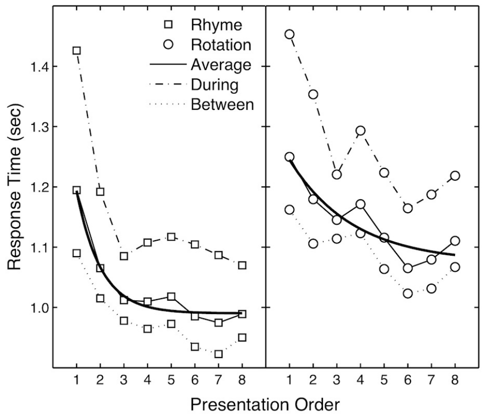


Figure 4. Average response times (RTs) on the rhyme-judgment task (left panel) and the rotation task (right panel) as a function of presentation order. The training effect is shown by decreasing RTs, over the course of the experiment. The solid lines show the exponential fits to the average RT data for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

Upon examining the RTs from individual participants, we discovered a reduction over time, regardless of the listening conditions, suggesting training effects. Figure 4 shows the mean dual task RT data as a function of presentation order, with each visual task in a separate panel. Despite each participant being presented with the listening conditions in a different, randomized order, the fit-lines in these figures do show that there were strong learning effects during the course of the experiment. In order to reduce the between-subject variance introduced by these training effects, they were modeled and compensated for using the following procedure. First an exponential function was fitted to the overall mean RT data for each of the two secondary tasks. The proportion of variance accounted for (R^2) by these fits was 0.975 for the rhyme-judgment task and 0.841 for the mental-rotation task. The horizontal asymptote of the exponential fit-line is interpreted as the value the RT converges on when all training effects have stabilized. For each individual participant, the learning effects were then compensated for by subtracting the deviation from the asymptote predicted by the fit-line for each condition based on the order of presentation. These manipulations of the data had no visible effect on the shape of the RT data as a function of spectral resolution. They did, however, considerably reduce the variance; for the raw RT data, standard deviations (SDs) of the mean RTs per level of spectral resolution ranged from 0.35 to 0.63 s; compensating for learning effect reduced the SDs to values between 0.15 and 0.27 s. Further analyses of the RTs were performed on the data corrected for learning effects.

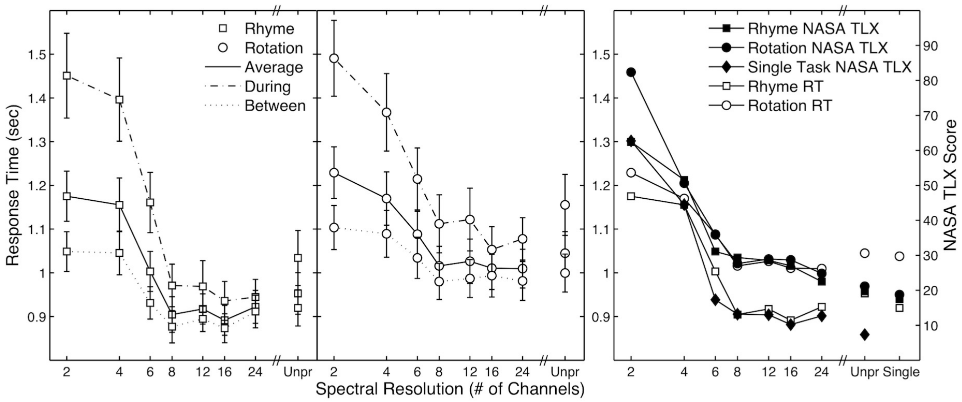


Figure 5. The first two panels show the average response times (RTs) on the two secondary tasks (solid lines) as a function of spectral resolution, as well as the RTs split into those recorded during sentence presentation (dashed line) and those between sentences (dotted line). The third panel shows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

The first two panels of 5 show the RT data (adjusted for learning effects) for the rhyme-judgment task and the mental-rotation task. Since wrong answers may be the result of a strategy or accidental button-press, these RTs may be unrealistically short and thus distort the data. Therefore, RTs for trials where a wrong answer was given were left out of the analysis. In both panels the mean RTs are shown for trials presented during listening (dashed line) and for trials presented between the auditory stimuli (dotted line). The overall mean RTs, recorded during and between auditory stimuli grouped together, are represented by the solid line. These two plots show that the reduction in RTs from 2 to 8 channels, is greater for trials presented during a sentence, than for trials presented in-between. A three-way repeated measures ANOVA, with the factors task, visual stimulus timing, and spectral resolution, indicated that the interaction between stimulus timing and processing is indeed significant ($F(7, 126)= 18.37$, $p<<0.01$). The results of the ANOVA also showed a significant main effect of processing on RT ($F(7, 126)= 28.78$, $p<<0.01$). Comparison between mean RTs for consecutive processing conditions using Tukey's HSD showed a significant decrease in RT for listening conditions up to 8 channels. Interesting to note is that there are significant RT improvements even for conditions where speech intelligibility had reached a plateau.

The third panel in 5 shows the mean NASA TLX scores for the listening task performed alone and with both secondary tasks. Both dual tasks were consistently judged more effortful than the single task conditions. It can also be seen from the figure that the NASA TLX scores, like the RTs, decrease with increasing number of channels CI simulations, at least for low numbers of channels. However, the decrease in effort between 6 and 8 channels CI simulations for NASA TLX scores was not significant. A two-way ANOVA of the NASA TLX scores showed significant main effects of spectral resolution, as well as task, which we attribute to the difference between single and dual task. An ANOVA when performed only on the NASA TLX data for the two dual tasks did not show a significant effect of task. Since there is no significant difference between the two dual tasks, the mean RTs were averaged over both dual tasks to be further examined. The mean RTs for consecutive listening conditions were compared using Tukey's HSD, which showed a significant decrease in NASA TLX scores for listening conditions up to 6 channels.

DISCUSSION

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

The results of the primary speech intelligibility task, in line with the findings of previous research (e.g. Başkent, 2006; Friesen et al., 2001), showed an increase in intelligibility with increased spectral resolution. The present study closely reproduced the speech intelligibility results reported by Friesen et al. (2001) for normal-hearing listeners presented with CI-simulated English sentences in Quiet. In both studies, a marked increase in intelligibility was observed between 2- and 6-channel CI simulation. Intelligibility appeared to reach about 98% for 6 spectral channels, and further increases in spectral resolution produced no significant improvement.

The main interest of the present study was listening effort. The results from the subjective workload questionnaire, the NASA TLX, showed consistently higher scores for dual task compared to single task. This is not surprising since the NASA TLX is designed to measure overall task-load, and a dual task can be considered to be cognitively more demanding than a listening task alone (Wickens, 2008). For all three tasks (the single listening task, the rhyme-judgment dual task, and the mental-rotation dual task) the NASA TLX showed a significant decrease in workload from 2 to 6 spectral channels. For an increased number of spectral channels beyond 6, no significant decrease in subjective workload was found. The results of the two visual decision-making dual tasks showed that, while both speech intelligibility and NASA TLX scores improved only from 2 up to 6 spectral channels, RTs on both secondary tasks improved significantly from 2 up to 8 channels. In other words, the RT measures captured an improvement, or benefit, of increasing spectral resolution from 6 to 8 spectral channels that the intelligibility task and the NASA TLX did not capture.

One can argue whether the benefit captured by a decrease in RTs is indeed due to reduced listening effort. Recall that what we call ‘listening effort’ is the proportion of a shared, limited cognitive resource that is allocated to the listening task. The larger the effort, the larger is this proportion assigned to the listening task, and thus the less of this resource is available to perform another task simultaneously. The RTs recorded between presentations of auditory stimuli showed a significantly shallower effect of spectral resolution than the ones recorded during presentation of auditory stimuli (fig 5). This observation supports the idea that the effects of simulated CI processing present in the RT data are indeed caused by changes in demands on shared resources due to these differences in processing. In short, the observed pattern suggests that the reduced RTs are caused by a decrease in listening effort associated with the increase in spectral resolution. Literature shows that effects of effort are rather elusive and effects in response time, while significant, can be as small as about 50 milliseconds (e.g. Baer, Moore, & Gatehouse, 1993; Sarampalis, Kalluri, Edwards, & Hafter, 2009). Although the results show only a significant effect in RT for conditions with constant intelligibility between 6 and 8 channels, this effect is observed for both secondary tasks, therefore we are convinced that it is a persistent and repeatable effect. While also intelligibility and subjective workload measures are likely to reflect changes in listening effort to some degree, as these two measures showed a pattern similar to that of the RT measures, they appear to be less sensitive to changes in listening effort; they showed no significant improvement between 6 and 8 spectral channels while the RT measures did.

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

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

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

In short, auditory processing, working memory and speed of processing seem to interactively affect both speech understanding and the resources available for additional tasks (Lunner, 2003). In this light, changes in the effort needed to interpret the auditory signal can be reflected in both measures of working memory performance and speed of processing, such as the RTs on a secondary task used in this study. In their study, Sarampalis et al. (2009) showed a benefit of noise reduction strategies reflected both in better working memory performance and faster RTs, even for conditions where noise reduction provides no benefit in speech intelligibility. The current study shows similar results; a significant decrease in RTs was found for increasing the number of channels for CI simulations from 6 to 8, while this produced no significant increase in intelligibility.

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

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

CHAPTER 3

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

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

ABSTRACT

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

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

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

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

INTRODUCTION

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

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

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

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

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

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

EXPERIMENT 1: SPEECH IN QUIET AT NEAR CEILING INTELLIGIBILITY

Motivation

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

Methods

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

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

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

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

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

Table 1. Summary of the experimental conditions for Experiment 1

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

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

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

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

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

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

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

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

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

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

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

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

Results

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

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

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

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

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

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

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

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

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

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

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

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

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

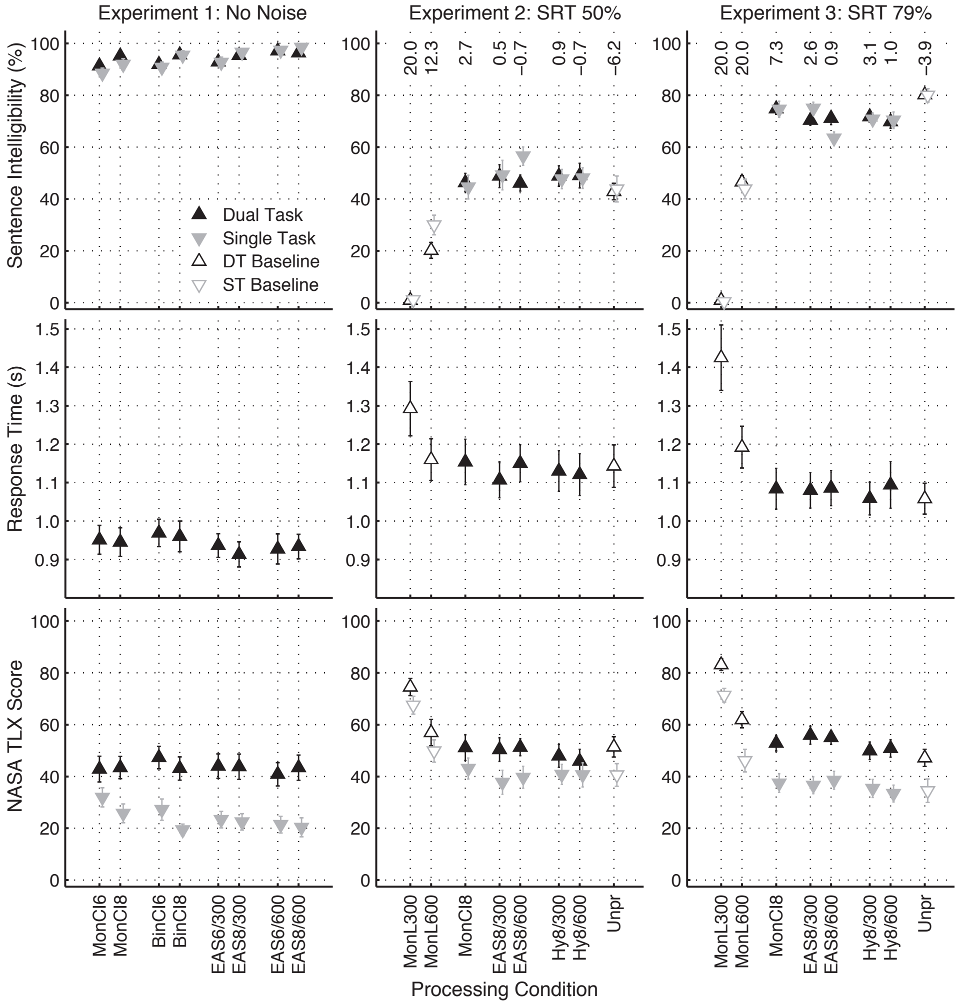


Figure 1: The results for Experiments 1, 2, and 3 are shown in the left, middle, and right column, respectively, with experimental conditions on the x-axes (experimental conditions are summarized in Table 1 for Experiment 1 and in Table 4 for Experiment 2 and 3). Up triangles show dual-task results,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

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

EXPERIMENT 2: SPEECH IN NOISE AT 50% INTELLIGIBILITY

Motivation

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

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

Methods

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

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

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

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

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

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

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

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

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

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

Results

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

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

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

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

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

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

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

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

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

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

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

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

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

EXPERIMENT 3: SPEECH IN NOISE AT 79% INTELLIGIBILITY

Motivation

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

Methods

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

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

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

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

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

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

Results

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

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

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

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

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

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

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

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

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

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

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

DISCUSSION

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

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

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

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

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

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

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

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

CHAPTER 4

Validation of a simple response-time measure of listening effort

Carina Pals, Anastasios Sarampalis, Hedderik van Rijn, Deniz Başkent

This chapter is published in: Journal of the Acoustical Society of America

2015 | Volume 138 | Issue 3 | Pages EL187-192

ABSTRACT

This study compares two response-time measures of listening effort that can be combined with a clinical speech test for a more comprehensive evaluation of total listening experience: verbal response times to auditory stimuli (RT_{aud}) and response times to a visual task (RTs_{vis}) in a dual-task paradigm. The listening task was presented in five masker conditions; no noise, and two types of noise at two fixed intelligibility levels. Both the RTs_{aud} and RTs_{vis} showed effects of noise. However, only RTs_{aud} showed an effect of intelligibility. Due to its simplicity in implementation, RTs_{aud} may be a useful effort measure for clinical applications.

PACS numbers: 43.71.Gv, 43.72.Dv, 43.71.Sy

INTRODUCTION

Speech understanding heavily depends on the cognitive processing required to interpret the (degraded) speech signal in everyday listening environments (Stenfelt & Rönnberg, 2009), perhaps even more so for hearing-impaired individuals. Measures of listening effort (LE) can therefore complement traditional speech intelligibility measures by providing additional information about the listening experience (McGarrigle et al., 2014). Different methods have been suggested for quantifying LE, ranging from subjective self-report (Rudner, Lunner, Behrens, Thorén, & Rönnberg, 2012), to behavioral measures, such as memory tasks (Rabbitt, 1966), speech response-times (RTs; Baer, Moore, & Gatehouse, 1993; Gatehouse & Gordon, 1990; Hecker, Stevens, & Williams, 1966) or dual-task paradigms (Pals, Sarampalis, & Başkent, 2013 (see also Chapter 2); Sarampalis, Kalluri, Edwards, & Hafter, 2009), and physiological measures, such as pupillometry (Koelewijn, Zekveld, Festen, & Kramer, 2012). An easy-to-administer method for measuring LE could be a valuable tool in research and clinical settings.

The current study compares two behavioral measures of LE that can be combined with the traditional clinical speech intelligibility test; the dual-task paradigm and verbal RTs to a speech task. Dual-task paradigms are an established method for quantifying LE (Pals et al., 2013; Sarampalis et al., 2009), and are based on the assumption that cognitive resources are limited and shared across tasks (Baddeley & Hitch, 1974; Kahneman, 1973). The resources needed for the primary task reduce the resources available for the secondary task (Nijboer, Taatgen, Brands, Borst, & van Rijn, 2013). Therefore, when the primary task is given precedence, secondary task performance is assumed to indirectly reflect the processing demands of the primary task. The verbal response times to auditory stimuli (RTs_{aud}), proposed as early as in the 1960s as a tool for discriminating between seemingly comparable speech communication systems (Hecker et al., 1966), and later used to quantify hearing device benefit (Baer et al., 1993; Gatehouse & Gordon, 1990), reflect cognitive processing time and index the cognitive effort required to interpret and respond to an incoming auditory signal (Baer et al., 1993; Gatehouse & Gordon, 1990).

In this study, a speech intelligibility task similar to clinical tests used in the Netherlands was performed either by itself to provide the RT_{aud}, or simultaneously with a secondary visual rhyme-judgment task (Pals et al., 2013) to provide visual response-times (RT_{vis}). To manipulate listening effort and intelligibility separately, and based on previous observations that LE can vary depending on the noise type (Koelewijn et al., 2012), participants listened to sentences in quiet, and in two different types of noise, each at two different intelligibility levels.

METHODS

Participants

Nineteen native Dutch speakers (age=18 to 25 years, mean=19), all students of University of Groningen, participated in exchange for partial course credit. Exclusion criteria were self-reported dyslexia or other language or learning disabilities, and pure tone thresholds above 20 dB HL at any of the audiometric frequencies (250 Hz - 6 kHz). The study was approved by the local ethical committee.

Stimuli

The speech stimuli used for the listening task were taken from the female speaker set of the Vrije Universiteit (VU) corpus (Versfeld, Daalder, Festen, & Houtgast, 2000). The corpus consists of 39 balanced lists of 13 conversational Dutch sentences, each 8 to 9 syllables long. A random subset of 24 lists was used per participant, two lists for each experiment or training block. A steady-state, speech-shaped noise (SSN; provided with the VU corpus) and an 8-talker babble in English were used as background noises. The sentences were presented in both noise types, each at two signal-to-noise ratios (SNRs), resulting in two levels of intelligibility; approximately 79% or near ceiling (NC).

Individual SNRs to achieve 79% intelligibility were determined for each participant at the start of the experiment, using sentences from the same corpus that were not included in the main experiment. This was done separately for SSN and babble, following a 3-down-1-up adaptive procedure (Levitt, 1971), which typically results in 79% accuracy. Each sentence-in-noise was presented at an overall level of 70 dB A. The first sentence was played repeatedly until the sentence was correctly understood, starting at -8 dB SNR and increasing the SNR in steps of 4 dB. After this, the adaptive procedure ran for 8 reversals at a step size of 2 dB. The resulting mean SNRs from last 8 reversals that were used in the experiment were: SNR = -1.20 dB (SD = 1.00) for SSN and SNR = 2.30 dB (SD = 1.10) for babble. A pilot experiment showed that increasing the 79% SNR by 5 dB resulted in NC speech understanding and this was therefore used as the SNR for the NC intelligibility conditions.

For the secondary, visual rhyme-judgment task, pairs of Dutch monosyllabic words (the same as used in Pals et al., 2013) were displayed in large, black capital letters on a white background, one above another, horizontally centered on a computer monitor placed ~60 cm from the participant. Each letter was approximately 7 mm wide and 9 mm high, with 12 mm vertical whitespace between the words.

Experimental procedure

Before the start of the main experiment, two cognitive tests were administered: the symbol search test from the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2012), to measure cognitive processing speed, and the standard computerized version of the reading span test (RST; van den Noort, Bosch, Haverkort, & Hugdahl, 2008), to measure working memory capacity.

The experimental procedure consisted of 2 training blocks and 11 experimental blocks. Training consisted of one single-task rhyme-judgment task and one dual-task combining the listening task and the rhyme-judgment task. The experimental blocks consisted of six single-task blocks: five times a listening task, and one visual rhyme-judgment task; and five dual-task blocks combining the listening task and the rhyme-judgment task. The listening tasks, in both single and dual task, were presented in 5 listening conditions: in no noise and in two noise types (babble and SSN) both at two intelligibility levels (79%, NC). Presentation order of the experimental blocks was counterbalanced using a Latin-square design.

In the listening task, participants listened to sentences and repeated them out loud. The sentence recordings were on average 1.8 s in duration and were presented 8 s apart, giving the participants 6.2 s between sentences to respond. The responses were recorded for later scoring of RTs_{aud} and accuracy. The RTs_{aud} were calculated from the offset of the stimulus, as logged by the experimental program, to the onset of the verbal response, as marked by a native Dutch speaker upon visual inspection of the recorded waveform in Audacity. A second native Dutch speaker re-scored a random sample of the recordings to test for inter-rater reliability (Pearson's $r > 0.99$).

In the secondary, visual, rhyme-judgment task, participants pressed one of two buttons as fast as possible to indicate whether two words rhymed or not. Chance of a rhyming pair was 50%. The words were presented on a monitor for a maximum of 2.7 s, or until the participant responded. In case no key was pressed, a 'miss' was logged. A fixation cross appeared for a randomly varied interval between 0.5 and 2.0 s between stimuli.

For the dual task, the listening task and the visual rhyme-judgment task were presented simultaneously, but with independent timing to prevent expectation-driven preparation (Pals et al., 2013). Note that this meant that the secondary-task stimuli could be presented during or between auditory stimuli.

RESULTS

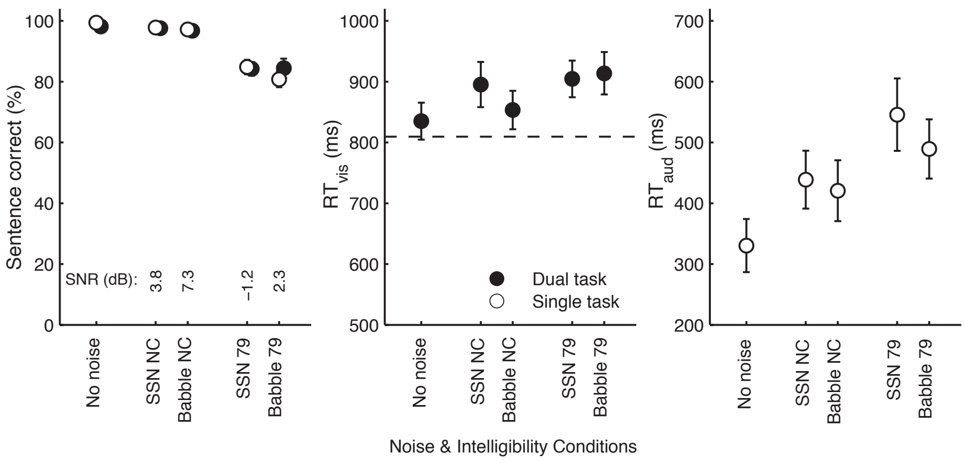


Figure 1: Left panel: Mean intelligibility in % sentences correctly repeated on the listening task in dual task (closed circles) and single task (open circles). Middle panel: Mean dual-task RTs_{vis} in ms, with single-task RT_{vis} performance indicated by the dashed reference line. Right panel: Mean single-task RTs_{aud} in ms. In all panels, the error bars show ±1 standard error.

The left panel of Figure 1 shows the speech intelligibility results in percentage of sentences correctly repeated, and confirms that the desired intelligibility levels were achieved.

The middle panel of Figure 1 shows the dual-task RTs_{vis} per condition, with average single-task RT_{vis} included as a baseline. Data from incorrect secondary-task trials were excluded from the analysis. Due to the nature of the rhyme-judgment task, with the number of trials depending on both response speed and response accuracy, the number of secondary task trials varied per participant per condition. As ANOVAs are less suitable for analyses based on different number of trials per cell, linear mixed-effects (LME) models were used (lme4-package version 1.1-7; lmerTest-package version 2.0-11) to analyze the RT_{vis} data. As the RT_{vis} were not normally distributed, we log-transformed the response times and excluded reaction times below .35 and over 2 s (1.80% of all trials), yielding a reasonably normal $lnRTs_{vis}$ distribution (assessed using QQNorm). Cognitive test score results were: WAIS (mean = 43.7, SD = 7.0) and RST scores (mean = 65.1, SD = 11.6).

The model of the dual-task $lnRT_{vis}$ results took into account all experimental manipulations: the overall effect of the presence or absence of noise, and for speech in noise, the effects of intelligibility and of noise type. Furthermore, visual stimulus timing (either during or in between the auditory presentation of sentences) and participants' WAIS and RST scores were included as factors. Random intercepts and slopes were included for all within-subject factors, and for stimulus timing (Barr, Levy, Scheepers, & Tily, 2013). A random intercept for sentence ID was not included, as no sentence can be assigned to RTs_{vis} responses recorded in-between auditory stimuli. Two different contrast-coding strategies were used to reflect the experiment design. The difference between noise and quiet was treatment-coded, setting quiet to 0 and noise to 1. The contrasts between SSN and babble and between 79% and NC intelligibility were effect-coded, setting one of the two to -0.5 and the other to 0.5. The p-values reported are obtained using the Satterthwaite approximation as reported by the lmerTest package.

The model of the $lnRT_{vis}$ is summarized in the top half of Table 1. The intercept corresponds to the average $lnRT_{vis}$ for speech in quiet, and is estimated at 0.323, although due to large variance it was not significant ($\beta = 0.323$, $SE = 0.221$, $t = 1.465$, $p = 0.162$). The model shows an effect of Noise, estimated at exp(.323+.041)-exp(.323)=.7 s ($\beta = 0.041$, $SE = 0.013$, $t = 3.174$, $p = 0.005$) when compared to the intercept. For speech in noise, the effects of noise type and intelligibility were not significant, nor was the interaction between noise type and intelligibility. RT_{vis} were significantly longer for secondary task trials presented simultaneously with an auditory stimulus than for trials in-between auditory stimuli, the effect in $lnRT_{vis}$ was estimated at 0.055 ($\beta = 0.055$, $SE = 0.009$, $t = 6.161$, $p < 0.001$). From the two cognitive measures collected before the experiment, only the WAIS score showed significant predictive value: the effect of WAIS score on $\ln RT_{vis}$ is estimated at -0.007 (β = -0.007, *SE* = 0.004, *t* = -2.138, *p* = 0.048), suggesting on average lower RT_{vis} for participants with a higher score on the WAIS symbol search, i.e. better cognitive processing speed predicts lower RT_{vis}.

Table 1. Summary of the LME model for Dual-task RTs_{vis} and Single-task RTs_{aud}

Dual-task $\ln RT_{vis}$ model	*Estimate (ms)*	*Std. Error*	*df*	*t value*	*Pr(>|t|)*	
(Intercept)	323.82	221.09	16.24	1.465	0.162	
Noise	41.97	13.23	18.16	3.174	0.005	**
N:Intelligibility	25.61	16.93	17.96	1.513	0.148	
N:NoiseType	-1.18	14.54	17.86	-0.081	0.936	
N:Intel:NoiseType	16.76	20.19	19.19	0.830	0.417	
Timing	55.64	9.03	26.00	6.161	< 0.001	***
WAIS	-7.67	3.59	16.14	-2.138	0.048	*
RST	-3.61	2.17	16.06	-1.667	0.115	
Single-task RT_{aud} model	*Estimate (ms)*	*Std. Error*	*df*	*t value*	*Pr(>|t|)*	
(Intercept)	556.82	191.12	16.27	2.913	0.010	*
Noise	131.09	20.85	17.89	6.284	< 0.001	***
N:Intelligibility	72.21	17.46	17.77	4.137	< 0.001	***
N:NoiseType	-24.72	12.66	17.77	-1.952	0.067	
N:Intel:NoiseType	-6.34	22.73	17.78	-0.279	0.783	
WAIS	-4.57	3.09	15.87	-1.480	0.158	
RST	-0.48	1.87	15.90	-0.259	0.799	

Note: The intercept estimates the RT_{vis} for No noise. *Noise* lists the average effect for speech in noise compared to No noise. Effects of *Intelligibility*, *NoiseType* and their interaction are only present in *Noise* and estimated relative to *Noise* (signified by 'N:').

The right panel of Figure 1 shows the average RT_{aud} per listening condition. Only RTs_{aud} for sentences that were repeated correctly were included in the analysis, therefore, similar to the dual-task RT_{vis} data, the RT_{aud} data contained unequal numbers of trials per cell depending on speech recognition accuracy. RTs_{aud} were analyzed using the same methodology as the dual-task RTs_{vis}. The RT_{aud} were approximately normally distributed for durations up to 1 s duration, with a skewed tail above 1 s. Therefore, RTs_{aud} of over 1 s were excluded from the analysis (1.85% of all trials). All factors relevant to the RT_{aud} were included as fixed effects, and a maximal random effects structure was used, accounting for individual intercepts and slopes for all within subject factors, as well as random intercepts for sentenceID.

The results of the model are summarized in the bottom half of Table 1. In quiet listening conditions, the verbal response was estimated to start 557 ms after stimulus offset (β = 556.82, *SE* = 191.12, *t* = 2.913, *p* = 0.010). In noise, averaged across the noise conditions, RT_{aud} were significantly longer by 131 ms (β = 131.09, *SE* = 20.85, *t* = 6.284, *p* < 0.001) implying an average RT_{aud} in noise of 688 ms. The average RTs_{aud} for speech in noise at 79% intelligibility was 72 ms longer than at NC intelligibility (β = 72.21, *SE* = 17.47, *t* = 4.137, *p* < 0.001) suggesting that the average RT_{aud} in noise at NC intelligibility was 652 ms, and the average RT_{aud} in noise at 79% intelligibility was 724 ms. The effect of noise type was not significant, suggesting that RTs_{aud} averaged over both intelligibility levels was no different for speech in SSN compared to babble. Finally, the interaction between noise type and intelligibility was not significant either. The cognitive measures taken before the experiment, the WAIS and the RST, were both included in the model as factors, however neither showed a significant effect.

DISCUSSION

The goal of this study was to compare RT_{aud} and RT_{vis} for suitability as measures of LE, especially as a complementary test next to a speech intelligibility test. Speech intelligibility, RTs_{aud} (for a simple speech intelligibility task), and RTs_{vis} (on a secondary rhyme-judgment task in a dual-task paradigm) were measured in five listening conditions: in no noise, and in SSN and babble, each at 79% and NC sentence intelligibility. Both RTs_{vis} and RTs_{aud} showed a clear effect of the presence of noise, similar to what literature suggests. However, RTs_{aud} showed a significant effect of intelligibility, while the RTs_{vis} did not.

The dual-task paradigm is a powerful tool for understanding the challenges listeners face in every day settings when combining speech communication with other tasks, or for showing the consequences of increased LE on cognition (Pals et al., 2013; Sarampalis et al., 2009). Hockey (1997) proposed that individual differences in coping strategies in demanding situations result in differences in the total amount of resources allocated to the tasks at hand. Dual-task measures have been suggested to reflect the proportion of the allocated resources needed for the primary task, while physiological measures, such as pupillometry, can reflect the magnitude of resource allocation (Karatekin, Couperus, & Marcus, 2004). It could well be that an increase in dual-task demands results in allocation of more resources to the combination of tasks, therefore not showing a difference in the proportional use of the allocated resources. However, if the goal is to find a measure suitable for clinical purposes, physiological measures would present drawbacks as they require expensive equipment and the procedures can be cumbersome.

The single-task RTs_{aud} showed a significant difference between the two intelligibility levels while the dual-task RTs_{vis} did not. On top of this, the RTs_{aud}, as measured in this experiment, have several advantages over the dual task for potential use in clinical settings and with a wide range of patients, for example, children and elderly. The RT_{aud} can be collected from recordings made during a simple speech-understanding test, already used in clinics, without the need for additional tests or expensive equipment. While the patient listens to sentences and repeats them out loud, the RT_{aud} can be collected by recording the responses for offline analysis, using software for automated speech onset detection (Jansen & Watter, 2008), or online using a simple, inexpensive voice-activated trigger. With its ease of implementation, RT_{aud} seems to be a good candidate for a measure of LE, complementing speech tests, in research and clinical settings.

CHAPTER 5

Effect of spectral resolution on speech intelligibility, comprehension, and listening effort in cochlear-implant users

Carina Pals, Anastasios Sarampalis, Andy Beynon,

Thomas Stainsby, Deniz Başkent

ABSTRACT

Objectives. Previous research has shown speech intelligibility in cochlear implant (CI) users to improve with increasing spectral resolution up to 7 active electrodes, and plateau thereafter. Here we hypothesized that further increased spectral resolution may still further improve listening effort, even if intelligibility remains unchanged.

Design. Spectral resolution was manipulated by varying the number of active electrodes of the CI between 7 and 15. After a one-month familiarization period, the CI users performed two experiments. In Experiment 1, a dual-task paradigm was used to measure speech intelligibility and listening effort, reflected in the accuracy scores on the primary listening task and response times on the secondary visual task, respectively. In Experiment 2, a sentence verification task was used to measure speech comprehension and listening effort, reflected in accuracy scores and response times, respectively.

Results. In line with literature, speech intelligibility did not improve beyond 7 active electrodes. In contrast, speech comprehension, as reflected by the sentence verification task, improved up to 11 active electrodes. The dual-task measure of listening effort showed no improvement beyond 7 active electrodes, while the sentence verification task measure of listening effort revealed a systematic improvement for increased spectral resolution up to 11 active electrodes.

Conclusion. The sentence verification task results revealed a benefit of increased spectral resolution for both comprehension and listening effort, for conditions that typically show no improvement in speech intelligibility. This highlights both the potential benefits of improving spectral resolution for CI users, and the added value of clinical assessment tools that can reveal such benefits when traditional speech intelligibility measures show no improvement. The sentence verification task may be a good candidate for such a clinical tool.

Keywords: Cochlear implant, speech perception, listening effort, spectral resolution

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 resources and processing (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 et al., 2012; Wild et al., 2012). In non-ideal 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 knowledge (e.g. Benard, Mensink, & Başkent, 2014; Hannemann, Obleser, & Eulitz, 2007), or 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 et al., 2007; Winn, Edwards, & Litovsky, 2015; Zekveld, Kramer, & Festen, 2010). This may in turn reduce the cognitive resources available for concurrent tasks (Sarampalis et al., 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 et al., 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 et al., 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 intelligibility has been studied extensively over the decades since the introduction of multichannel CIs (e.g. Eddington, 1980; Fishman, Shannon, & Slattery, 1997;

Friesen et al., 2001; Fu, Shannon, & Wang, 1998; Schvartz, Chatterjee, & Gordon-Salant, 2008; Winn, Chatterjee, & Idsardi, 2012). Research has shown, for example, that the recognition of individual phonemes in quiet improves up to 7 electrodes (Fishman et al., 1997) and thresholds for phoneme recognition in noise continue to improve up to, and possibly beyond, 16 electrodes (Fu et al., 1998). Sentence intelligibility, on the other hand, reaches ceiling performance at about 10 active electrodes for speech in noise (Friesen et al., 2001). While in quiet, sentence intelligibility plateaus around 4 active electrodes for the average CI users (Fishman et al., 1997), and around 7 active electrodes for high-performing CI users (Friesen et al., 2001). This increased tolerance for reduced spectral resolution when listening to full sentences can be explaned by the availability of sentence context and suggests the involvement of effortful top-down processing to enhance intelligibility (Sheldon et al., 2008). This implies that while intelligibility has reached ceiling, listening effort may still be high. Effects of spectral resolution on listening effort for CI users, however, have not been previously documented.

The current study aims to investigate the effect of spectral resolution in CI users on not only speech understanding, but also listening effort. Specifically, we hypothesized that when spectral resolution increases, this could provide a benefit in listening effort, even after intelligibility performance has reached a plateau. Indirectly supporting this idea is the fact that in normal hearing (NH) listeners presented with acoustic, noise-band vocoder simulations of CI speech in quiet, speech intelligibility improved up to 6 spectral channels and plateaued thereafter, while listening effort continued to improve up to 8 spectral channels (Pals et al., 2013, see also Chapter 2). Similar results have been shown using pupil dilation as a measure of effort: in NH listeners, spectral resolution affects pupil dilation, i.e. listening effort, even when intelligibility is at 100% (Winn et al., 2015). In the present study, we systematically investigate whether, similar to NH listeners, CI users benefit from increased spectral resolution in reduced listening effort even when changes in intelligibility are not observed or expected.

In the present study, a dual-task paradigm first designed and used in our earlier study in NH listeners (Pals et al., 2013), was employed to measure intelligibility (primary listening task) and listening effort (secondary task) simultaneously. The dual-task paradigm is an established method for measuring cognitive load, and has been used to quantify listening effort in a number of studies (e.g. Anderson Gosselin & Gagné, 2010; Fraser, Gagné, Alepins, & Dubois, 2010; Gosselin & Gagné, 2010; Pals et al., 2013; Rakerd, Seitz, & Whearty, 1996; Sarampalis et al., 2009). The assumption that cognitive resources are limited and shared across tasks (Broadbent, 1958; Kahneman, 1973), implies that when two tasks are performed simultaneously, the execution of the primary task uses resources that would otherwise have been available for the secondary task. Performance on the secondary task thus reflects the cognitive processing load of the primary task (Broadbent, 1958; Rabbitt, 1966). 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 whether the method is suitable for use with CI users, since a range of different factors can affect performance in CI users (Başkent, Gaudrain, et al., 2016), and effects of age may further affect the results as CI users tend to be older (Bhargava et al., 2014; Bhargava, Gaudrain, & Başkent, 2016).

Therefore, the dual-task paradigm (Experiment 1) was complemented with a simpler, single-task measure of comprehension and processing speed (Experiment 2); the sentence verification task (Adank & Janse, 2009; Baer, Moore, & Gatehouse, 1993). While this task was not 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 sentence verification task, 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 response times (RTs). As an increase in cognitive load leads to slower comprehension (Gibbon, Moore, & Winski, 1997; Mattys et al., 2011; Wagner et al., 2016), the sentence verification accuracy and RTs can be interpreted to reflect comprehension and cognitive processing load, i.e. 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 in NH listeners (Pals et al., 2013), we expect that listening effort could be improved further with increasing spectral resolution even when intelligibility appears unchanged.

EXPERIMENT 1: DUAL-TASK APPROACH: SPEECH INTELLIGIBILITY AND LISTENING EFFORT

To be able to compare our results with NH listeners and CI users, the same dual-task paradigm we had designed for our previous study (Pals et al., 2013) was used. A few minor modifications were made to the design to accommodate for expected differences in speech understanding 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 response time-out was longer; these changes are described in more detail below.

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 9 participants were included in the final analyses. From the participants recruited in Nijmegen one 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 female, mean age 58 years, range 34 - 76) who completed the two experiments fully without any problems.

The participants were all native Dutch speakers, and postlingually 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 the Table 1), however, all learned their native language in audio-verbal mode. As the goal of this study was to investigate listening effort for speech understanding at ceiling, only the best performing CI users were chosen. Inclusion criteria were a minimum of 1 year experience with CI use, clinical consonant-nucleus-consonant word recognition scores of 80% or higher, and no known cognitive disabilities. All participants had either normal vision, or vision corrected to normal by spectacles. All-but-one of the participants had complete intra-cochlear electrode array insertion and all were fitted with at least 15 active electrodes in their daily speech processor maps. This and the subsequent experiments were both approved by the local ethical committee. Demographic and hearing-related information for these participants is summarized in Table 1.

Table 1. Summary of the CI participants' demographic and hearing-related information.

Subject ID	*Gender*	*Age during experiment*	*Age of first HL*	*Duration of CI use*	*Etiology*
304	M	38	3	2.3	Usher
307	M	64	46	1	Progressive
310**	F	54	49	5	Wegener
311	M	59	31	2	Meningitis
313*	M	60	0	7	Mother rubella
314	M	51	7	12	Osteoporosis
315	F	69	33	7	Progressive
316	F	41	6	2	Hereditary
317	M	76	10	2	Otitis media
321	F	51	10	8	Progressive
322	F	59	54	2	Schwannoma
323	F	67	38	7	Stapedectomy
324	F	66	38	3	Progressive
325	F	52	26	2	Progressive
326	M	62	38	3	Progressive
327	F	70	14	4	Progressive
328	M	34	65	4	Progressive
329	M	65	16	7	Progressive
330	F	58	48	6	Progressive
331	M	67	43	3	Progressive
332	F	59	58	7	Progressive
333	M	65	40	4	Progressive
334	F	58	34	4	Otosclerosis
335*	F	49	0	17	Hereditary
336	F	62	30	3	Ototoxicity

Note: * CI users who were hearing impaired since birth. ** CI user who did not have a fully inserted electrode array.

Speech stimuli. In our previous study with normal-hearing (NH) participants we used sentences from the VU corpus (Vrije Universiteit; Versfeld et al., 2000). However, even CI users selected for high phoneme scores may still show poor sentence intelligibility for the VU corpus speech materials (Bhargava et al., 2014, 2016). In the current study, the speech stimuli for the primary intelligibility task were therefore taken from the LIST corpus (van Wieringen & Wouters, 2008). This corpus is specifically optimized to provide accurate speech reception thresholds for Dutch and Flemish hearing-impaired listeners and CI users in quiet and in noise. 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).

Visual stimuli. 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 5 basic Dutch vowels (a, e, i, u, 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 50cm in front of the participant. The letters were approximately 9mm high and 7mm 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 response times 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 affecting the outcome. As a result, the stimulus presentation level was not controlled for in dB SPL but set to a comfortably loud level at the start of the experiment and kept the same throughout data collection.

Experimental conditions. Spectral resolution was manipulated by altering the number of active electrodes of the CI by disabling electrodes and redistributing the frequencies assigned to them to the remaining electrodes (Friesen et al., 2001). Previous research has shown that, on average, CI users' speech intelligibility performance in quiet is near ceiling from about 7 active electrodes (Fishman et al., 1997; Friesen et al., 2001). Because a core question of this study is whether changes in listening effort occur when intelligibility no longer improves, the experimental conditions were therefore chosen to cover the range between 7 electrodes and the full 22-electrode array. Specifically, four experimental maps were generated with 7, 9, 11, and 15 active electrodes, chosen because these numbers allowed for the active electrodes to be either evenly spaced or distributed in a regularly recurring pattern across the 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, which resulted in a redistribution of frequencies similar to that used by Friesen et al. (2001). 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, ADRO, etc., were also left as is.

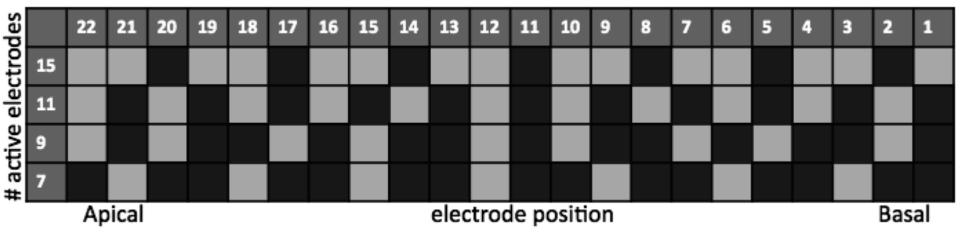


Figure 1. The distribution of active electrodes along the full array is shown for each of the experimental conditions. A light gray square denotes an active electrode, a dark gray square a deactivated electrode.

Procedure. The experiment consisted of two testing sessions in which the participants performed both Experiment 1 and 2 (Experiment 2 will be described later), with a one-month training period in-between. During the one-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 the maps for one hour a day, rotating between the four different maps on the processor. This served to familiarize the listener with the experimental maps before the actual testing session, thus minimalizing acute effects of new, unfamiliar stimulation patterns and training effects over the course of the experiment. Research shows 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 spectral resolution of our experimental programs may negatively impact the CI participants' listening abilities for example at the workplace, we decided instead to limit familiarization to one hour a day, for one 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.

The first session lasted at most 1 hour, during which the participants were tested using their preferred map on their own processor, while simultaneously the experimental processor was programmed. The second session lasted approximately 2 hours, during which the participants were tested with each of the 4 experimental maps, in counterbalanced order (in a 4x4 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. A sample sentence was played repeatedly, starting at a very low presentation level and increasing in steps of 2.5 dB. Following clinical procedure, 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 three or four times, alternately increasing and decreasing in loudness 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-minute 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 intelligibility task was presented twice alone (single task), one training block and one experimental block, then the intelligibility 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 4 task blocks before moving on to the next condition.

The primary intelligibility task required the participants to listen to the sentence stimuli and repeat them out loud, giving their best guess when they were not sure what they heard. When the intelligibility 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 seconds. 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 seconds was reached. The time-out was longer than in our previous study to accommodate the more advanced age of some of the participants of the present study. If after these 5 seconds 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 seconds 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 intelligibility scores for the primary listening task, both in single task (open symbols) and in dual task (filled symbols). The baseline included in the graph reflects the average intelligibility 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 counterbalanced test conditions), these were not included as a condition in the analysis. They are shown here purely as a reference level. The speech intelligibility scores from experimental conditions were analyzed using a two-way repeated-measures ANOVA using R and the ez package (version 4.2-2) including the main factors spectral resolution (4 levels: 7, 9, 11, 15 active electrodes) and task type (2 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 intelligibility and no significant interaction.

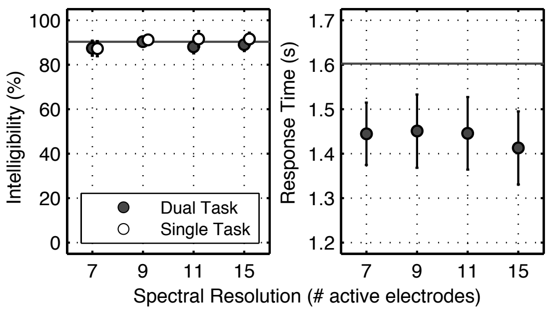


Figure 2: The left panel shows the speech intelligibility in percentage sentences correctly repeated, for both single task (open symbols) and dual task (filled symbols), as a function of spectral resolution. The right panel shows the response times in seconds on the dual-task secondary task. Error bars 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.

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, therefore RTs below 0.35 and over 3 s were excluded (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. As with the speech intelligibility scores, the baseline RTs recorded in the first session were not included as a condition in the analysis, however, when included as a factor in the model they contributed significantly to the fit of the model and were therefore included ($\chi^2(1) = 36.202$, $p < 0.001$).

The final model included the factors spectral resolution, presentation order, accuracy, and baseline RT. A random intercept was included for participantID, and random slopes and intercepts were included for all within-subject factors. The intercept of the model corresponds to average difference in RT compared to baseline on the secondary task while listening to speech using 7 active electrodes as the first task of the experiment, and did not differ significantly from 0 (β = -0.1194, SE = 0.0769, t = 1.554, p = 1.256). The model revealed a significant effect of presentation order on lnRT, estimated at 0.0182 (β = 0.0182, SE = 0.0083, t = -2.208, p = 0.038), suggesting a decrease in RTs of $e^{(-0.1194-0.0182)}-e^{-0.1194}$ =-.0160 s for each consecutive condition in the experiment, and a significant effect of baseline RT, estimated at 0.2487 (β = 0.2487, SE = 0.0297, t = 8.375, p < 0.001), suggesting that participants with higher baseline RTs also have longer RTs in the experiment overall ($e^{(-0.1194+0.2487)}-e^{-0.1194}$ =0.2506 s longer RTs in the experiment per 1 second longer baseline RTs). The model showed no significant effect of spectral resolution (β = -0.0037, SE = 0.0022, t = -1.670, p = 0.109), or accuracy (β = -0.0062, SE = 0.0064, t = -0.971, p = 0.336) on RT.

EXPERIMENT 2: SVT APPROACH: SPEECH COMPREHENSION AND LISTENING EFFORT

In Experiment 1, we had used the dual-task paradigm, as it had been previously tested and validated with NH participants listening to CI simulated speech (Pals et al., 2013). The sentence verification task 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 reduced spectral resolution in NH listeners presented with CI-simulated 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 that 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 (4 male) were all native Dutch speakers and young adults (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 sentence material used for the sentence verification task was created by Adank and Janse (2009), and the same recordings were used for both the NH and the CI participants. The corpus 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, are at least 3 words long (min. 4 syllables), and the longest sentence is 8 words long (max. 14 syllables).

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 (California, USA) and a DA10 digital-to-analog converter of Lavry Engineering, Inc. (Washington, USA) to the open-back HD600 headphones of Sennheiser electronic GmbH & Co. KG (Wedemark, Germany) at 65 dB A.

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, spectral resolution was manipulated by varying the number of bands of noise-vocoded CI simulation. The auditory stimuli were presented in 6 conditions; 4-, 6-, 8-, 12-, and 16-band noise-vocoded CI-simulated 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). All speech stimuli, including the unprocessed condition, were band-pass filtered to 80 − 6000 Hz. The CI simulations were generated using the method as described by Shannon at al. (1995). For each of the CI simulation conditions, the 80 − 6000 Hz 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 cochlea according to the Greenwood function (Greenwood, 1990). The speech recording was band-pass filtered into the desired number of analysis bands, using 6th 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 3rd order Butterworth filter. The carrier noise bands were modulated using the envelopes of the corresponding analysis bands and post-filtering using the original band-pass filters, and finally the resulting bands were combined to form the noise band vocoded CI simulation speech signal.

For the CI users, the experimental conditions of varying spectral resolution were the same as in Experiment 1, described above.

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 research using the same paradigm (Adank et al., 2009), this implies that RTs were possible. If no key was pressed 5 seconds 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 seconds 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 hour. The CI users performed Experiment 2 in two sessions, with a one-month training period in-between, similar to Experiment 1. Session one lasted about 1 hour, and session two about 2 hours. They performed Experiment 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 Experiment 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 sentence verification task, 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. The accuracy data were converted to the sensitivity measure d', because this provides a bias-free measure of accuracy, i.e. it is not affected by individual preferences for either 'yes' or 'no' answers that may distort a % correct accuracy score. Figure 2, top-left panel shows the accuracy in d' scores for the sentence verification task 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 vocoder CI simulation) as a numerical within-subject factor and covariate task order, revealed a significant effect of spectral resolution ($F(1, 23)= 36.696$, $p<0.001$).

In order 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, 16 channel CI simulations) 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 ($\chi^2(1)=0.0151$, $p=0.9021$) 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 d') for 4 channel conditions, was estimated at approximately 2.86 (β = 2.8564, *SE* = 0.2844, *t* = 10.044, *p* < 0.001) and the effect of number of channels at 0.20 (β = 0.2021, *SE* = 0.0259, *t* = 7.807, *p* < 0.001), suggesting a 0.20 d' increase in accuracy for every additional channel in the CI simulation. No significant effect of task order was found (β = 0.0778, *SE* = 0.0647, *t* = -1.203, *p* = 0.232).

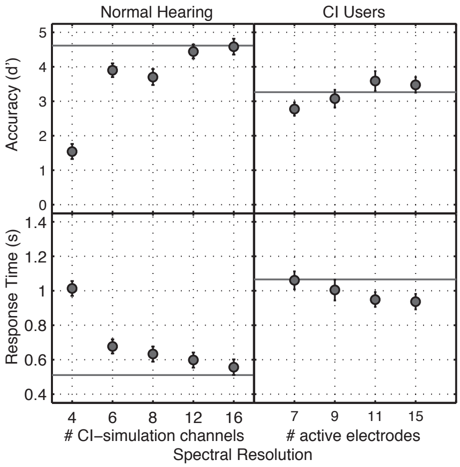


Figure 3: Results of the sentence verification task shown for NH participants (left-side panels) and CI participants (right-side panels). The top panels show accuracy scores in *d'* and the lower panels show RTs. Error bars show standard error. The baselines included in each figure show the average score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

The lower-left panel of Figure 3 shows the RTs on the sentence verification task for the NH listeners. The RTs approximated a normal distribution between -0.1 and 2.15 seconds, 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 512 ms (β = 0.5123, *SE* = 0.0574, t = 8.926, p < 0.001) and corresponds to the estimated average difference in RTs compared to baseline for the 4-channel CI simulation when presented as the first task of the experiment. The model showed a significant effect of spectral resolution, estimated at -25 ms (β = -0.0252, *SE* = 0.0031, t = -8.073, p < 0.001) suggesting a 25 ms decrease in RT for each additional spectral channel. The model also revealed a significant effect of baseline RT, estimated at 569 ms (β = 0.5693, *SE* = 0.0892, t = 6.381, p < 0.001), suggesting that participants with longer baseline RTs responded more slowly during the experiment as well (1 s longer baseline RT predicts on average 569 ms longer RTs in the experiment). The effect of presentation order was not significant (β = -0.0025, *SE* = 0.0041, t = -0.604, p = 0.546).

Because the relationship between the spectral resolution of the CI simulation and RT on the sentence verification task appears to be linear from 6 spectral channels up, but with a sharp increase in RTs from 6 to 4 channels, the results were re-modeled excluding the 4 channel condition, in order to see whether the effect would still be significant. The new model's intercept was estimated at 367 ms (β = 0.3672, *SE* = 0.0570, t = 6.444, p < 0.001) and corresponds to the estimated average difference in RTs compared to baseline for the 6-channel CI simulation when presented as the first task of the experiment. The model showed a significant effect of number of channels, estimated at -12 ms (β = -0.0118, *SE* = 0.0023, t = -5.166, p < 0.001), and a significant effect of baseline RT, estimated at 603 ms (β = 0.6032, *SE* = 0.0922, t = 6.537, p < 0.001). The effect of presentation order was again not significant (β = -0.0062, *SE* = 0.0041, t = -1.509, p = 0.132).

CI users. The top-right panel of Figure 3 shows the accuracy in the sentence verification task with CI users in d' scores. 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<0.001). In order to examine the effect of spectral resolution on accuracy, the results were modeled using a linear model. Including baseline RT as a factor did not improve the model fit (χ^2(1)=3.7594, p=0.053) and was therefore not included in the model.

The final model, with 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, estimated the intercept at an accuracy score of d' = 2.297 (β = 2.2967, SE = 0.3986, t = 5.762, p < 0.001), corresponding with the estimated accuracy for 7 active electrodes when presented as the first task of the session. The model showed a significant effect of spectral resolution on accuracy of 0.129 (β = 0.1291, SE = 0.0405, t = 3.187, p = 0.002), suggesting an increase in d' of 0.129 for each additional active electrode. The effect of task order was not significant (β = 0.1291, SE = 0.0405, t = 3.187, p = 0.072).

The lower-right panel of Figure 3 shows the RTs in the sentence verification task 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 seconds. 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 774 ms (β = 0.7741, SE = 0.1298, t = 5.964, p < 0.001), and corresponds with the estimated difference in RT compared to baseline for the 7 active electrodes condition when presented as the first task of the experiment. The effect of number of channels was estimated at -17 ms (β = -0.0169, SE = 0.0058, t = -2.884, p = 0.007) suggesting a 17 ms decrease in RTs for each additional active electrode. The effect of presentation order was estimated at -58 ms (β = -0.0582, SE = 0.0100, t = -5.800, p < 0.001), suggesting a 58 ms decrease in RTs for each consecutive block in the experiment. The effect of baseline RT was estimated at 407 ms (β = 0.4074, SE = 0.1003, t = 4.062, p < 0.001).

DISCUSSION

The current study investigated how spectral resolution affects speech intelligibility, speech comprehension, and listening effort in CI users. In our previous study in NH listeners, we observed that, even when intelligibility had already reached ceiling, further improved spectral resolution could still further improve listening effort. Based on this observation, we hypothesized that for CI users, listening effort may similarly improve with increased spectral resolution, even when speech intelligibility is near ceiling. Experiment 1 examined the effect of spectral resolution on speech intelligibility and listening effort in CI users, using a dual-task paradigm that was validated in our previous study with NH participants listening to noise-vocoded CI simulated speech (Pals et al., 2013). However, this dual-task paradigm had not been used with CI users before. The differences between the NH participants of our previous study and the CI participants of this study, both in hearing ability and age, might affect the dual-task outcome, possibly resulting in floor or ceiling effects. We therefore included a second experiment, using a simple, single-task, response-time measure of listening effort. Experiment 2 examined the effects of reduced spectral resolution on speech comprehension and processing speed in both NH and CI listeners, using a sentence verification task. The results in a nutshell: Experiment 1 showed no effect of spectral resolution on either intelligibility or listening effort; Experiment 2, on the other hand, showed a clear effect of spectral resolution on both speech comprehension and processing speed in NH as well as CI participants. Each of these findings will be discussed in more detail below.

The results from Experiment 1 showed that for CI users, further increased spectral resolution from 7 active electrodes upwards, did not lead to further improved sentence intelligibility in quiet listening conditions. These findings were as intended by our design, which was based on the literature. In this study, spectral resolution was manipulated by limiting the number of active electrodes. The experimental conditions (7, 9, 11 and 15 active electrodes) were chosen based on earlier research that had shown a plateau in speech intelligibility for CI users from 7 active electrodes and up (Fishman et al., 1997; Friesen et al., 2001). The effect of spectral resolution on speech intelligibility has been extensively studied (Chatterjee et al., 2010; Fishman et al., 1997; Friesen et al., 2001; Fu et al., 1998; Henry, Turner, & Behrens, 2005; Schvartz et al., 2008; Won, Drennan, & Rubinstein, 2007) and measures of intelligibility are regularly used in both clinical and research settings. The main interest of this study was therefore primarily effects of increased spectral resolution on listening effort when intelligibility is near ceiling, as this is when measures of listening effort can reveal potential benefits that are not directly evident from intelligibility measures.

The results from Experiment 1 showed that for CI users, further increased spectral resolution from 7 active electrodes upwards, did not lead to decreased secondary task RTs. These results are not in line with our expectations based on studies in NH listeners that show improved listening effort for increased spectral resolution. In prior research, we have successfully used this same dual-task paradigm to show improvements in listening effort for increased spectral resolution in NH listeners presented with CI simulated speech, even for conditions that resulted in equal intelligibility (Pals et al., 2013). Other research similarly shows that reduced spectral resolution leads to more effortful speech understanding for NH listeners, as reflected by pupil dilation (Winn et al., 2015). Eye tracking data further suggests that spectral degradation leads to slower speech processing, thus reducing the benefit of sentence context, which can further increase listening effort (Wagner et al., 2016). The speech materials used for the CI users, however, were optimized for use with hearing impaired and CI listeners (van Wieringen et al., 2008): they are spoken with clear articulation and at a slow speaking rate. This may have diminished the detrimental effects of reduced spectral resolution on the effective use of sentence context, as the slower speaking rate may have accommodated for the slower speech processing. In short, the results of this study suggest, that for CI users, specifically in the experimental setting of this study, i.e. when listening to slow-spoken and carefully articulated speech presented without interfering noise and over personal audio cable in a sound isolated booth, improved spectral resolution from 7 active electrodes up does not improve effort as measured by our dual-task paradigm.

Perhaps the lack of effect of spectral resolution on secondary task performance may also be partially explained by a difference in motivation between the NH participants of the previous study and the CI users in the current study. The NH participants were university students participating in a number of studies for course credit and the experimenters observed in these participants a lack of intrinsic motivation to perform well in the experiment. The CI participants on the other hand, were, in our experience, generally very grateful for the improvements scientific research has provided for CI technology, from which they directly benefit, and therefore quite motivated to contribute to scientific research and perform their very best in the experiments. This difference in motivation could have affected the dual-task outcome. The dual-task paradigm reflects effort as a proportion of the total capacity of available resources. However, even if cognitive resources are limited, it has been suggested that the total capacity may be temporarily increased depending on individual strategy, motivation, and determination to enhance performance by exerting extra effort (Hockey, 1997; Kahneman, 1973; Wingfield et al., 2007). Therefore, if the highly motivated CI users expended extra effort to temporarily increase the cognitive resources available for the dual-task, then, even if the processing load of the primary task increases, it may not necessarily be reflected in performance on the secondary task (Pals et al., 2015, see also Chapter 4).

Another notable difference between the two studies is the ages of the NH versus the CI participant groups. The NH listeners in our previous study (Pals et al., 2013) ranged in age from 19 to 25 years, and the CI listeners in this study ranged in age from 34 to 76 years. Age is known to affect cognitive ability, and could in our CI participant group have resulted in reduced cognitive capacity compared to their younger NH counterparts. However, reduced cognitive capacity alone cannot explain the lack of interaction between the primary listening task and the secondary response time task. Perhaps the problem lies with the nature of the secondary task; a response-time task that requires a fast motor response. Research suggests that cognitive motor control is also affected by age, and more importantly, that divided attention over a cognitive and sensory-motor task, such as in our dual-task paradigm, greatly affects motor control in older adults (Li & Lindenberger, 2002; Seidler et al., 2010). This may explain the fact that the RTs on the secondary task in this study, are around half a second longer than the RTs for the young NH participants in our previous study. This increased cost of divided attention and cognitive motor control may have resulted in a floor effect for the rhyme-judgment task, and could therefore be a reason why the RTs showed no additional effect of improved spectral resolution of the stimuli.

In Experiment 2, a simple, single-task, response-time measure was used as a measure of listening effort; the sentence verification task (Adank et al., 2009; Baer et al., 1993). In addition to the CI users participating in both Experiment 1 and 2, an extra group of young NH participants was recruited for a validation experiment. The sentence verification task is considered a measure of comprehension (accuracy), and speed of comprehension (RTs). A measure of comprehension requires the listener to understand the meaning of the speech and reason about it (Ralston et al., 1991; Wingfield et al., 2007), and may therefore more closely reflect the requirements of everyday verbal communication. In the sentence verification task, the RTs reflect the processing time required to comprehend the speech and judge whether the sentence was true or false. Research suggests that increased listening effort results in longer processing time required to understand the speech (Gatehouse & Gordon, 1990; Gibbon et al., 1997; Pals et al., 2015; Wagner et al., 2016). We therefore interpret longer RTs as increased listening effort.

The results of Experiment 2 showed improved sentence verification task accuracy scores, i.e. comprehension, with increased spectral resolution for NH listeners at least up to 12 spectral channels, and for CI users up to 11 active electrodes. In our previous study, the dual-task intelligibility results in NH listeners, improved only up to 6 spectral channels (Pals et al., 2013). Other research similarly shows ceiling performance on sentence intelligibility in quiet for NH listeners for spectral resolution of around 5 to 6 spectral channels, sentence intelligibility in noise, however, continues to improve with increased spectral resolution up to 12 to 20+ spectral channels depending on the SNR (Dorman et al., 1998; Dorman, Loizou, & Rainey, 1997; Friesen et al., 2001). Similarly, sentence intelligibility for CI users improves with increased spectral resolution, in quiet up to 4 to 7 active electrodes, and in noise up to 10 active electrodes (Fishman et al., 1997; Friesen et al., 2001). The sentence verification task comprehension scores in quiet appear more similar to the speech intelligibility results in noise reported in these studies. This might suggest that the sentence material used for this sentence verification task is more challenging than the sentences typically used for intelligibility tasks. Alternatively, comprehension, the understanding and reasoning about the heard speech as needed to judge the truth value of a sentence, may be more affected by changes in cognitive processing load than a measure of intelligibility is. As speech comprehension requires further cognitive processing than the first step of speech perception that is measured in an intelligibility task, comprehension is suggested to rely more on cognitive capacity (Just & Carpenter, 1992; Ralston et al., 1991). Comprehension may thus be more affected by changes in listening effort than intelligibility.

In addition to the improvement in comprehension, the NH results showed a clear linear trend of improved RTs up to at least 16 spectral channels. Similar to our previous dual-task study in NH listeners, the sentence verification task RTs show more clearly and convincingly an improvement with increased spectral resolution from 12 to 16 channels, while the accuracy scores, i.e. comprehension, appeared to reach ceiling around 12 channels (see Figure 3). In CI users, on the other hand, both accuracy and RTs appeared to reach ceiling at around the same level of spectral resolution; around 11 active electrodes. The results of Experiment 2 show that, even when the dual-task results for speech in quiet no longer show improved sentence intelligibility or listening effort, the sentence verification task suggests that further improved spectral resolution can still further improve speech comprehension and listening effort, both in NH and CI listeners.

In addition to the potential problems with our dual-task paradigm and the older CI participants as discussed above, the different speech materials used for the dual task and for the sentence verification task may have contributed to the differences in outcomes between the two measures. 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 material used in Experiment 1. Wingfield et al. (2006) suggest that effects on speech comprehension become apparent only after a certain threshold of processing difficulty has been crossed, and therefore the nature of the speech material and task affect the outcome of such tests. Perhaps in Experiment 2, the more challenging speech materials result in a stronger effect of spectral resolution on task performance.

However, the difference in results between the dual task and the sentence verification task 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. The differences in outcomes between the two tasks can therefore not be attributed to differences in age or motivation of the participants, or to differences in speech materials, suggesting that they must stem from differences between the two measures themselves. In the current study, the difference in outcomes between the dual task and the sentence verification task may also, in part, be due to differences in the nature of the tasks.

Regardless of the reason for the differences between the dual task and the sentence verification task outcomes, the core finding of this study is this: the sentence verification task has shown improved speech comprehension and listening effort in CI users for improved spectral resolution between 7 – 11 active electrodes, conditions in which speech intelligibility measures typically show no change. The same spectral manipulation in Experiment 1 showed no effect on speech intelligibility and listening effort as measured using the dual-task paradigm, and other research also shows a plateau in speech intelligibility in quiet listening conditions for spectral resolution beyond 7 active electrodes in CI users (e.g. Fishman et al., 1997; Friesen et al., 2001). In other words, the sentence verification task has shown a benefit of spectral resolution, that is not likely to be detected by the clinical speech intelligibility tests and may therefore be a valuable measure to complement the traditional intelligibility 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, clear speech presented without background noise, through personal audio cable in a sound proof room, the sentence verification task showed both improved speech comprehension and listening effort for increasing spectral resolution up to 11 active electrodes. This finding shows both the benefit of increased spectral resolution for CI users even when this benefit is no longer evident from speech intelligibility measures, and thus also the added value of a measure such as the sentence verification task to complement traditional measures of intelligibility 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 sentence verification task 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.

CHAPTER 6

General discussion

Carina Pals

The title of this thesis is "Listening effort, the hidden costs and benefits of cochlear implant hearing." So what are these *hidden* costs and benefits of hearing with a cochlear implant (CI)? The obvious *visible* benefit of a CI is the restored hearing ability, which allows the CI user to participate more comfortably in our predominantly hearing society. Some of the better performing CI users are even able to communicate over the telephone, without the visual aid of lip reading. These visible benefits are measurable in terms of speech understanding, and can be easily observed by friends and coworkers. The *hidden* costs and benefits, on the other hand, are internal to the listener and less easily measured or observed. Even if speech understanding performance is similar, there may still be differences between normal hearing (NH) listeners and CI users, between individual CI users, or within a single CI user for different device settings, configurations, or different listening situations. Differences, for example, in *listening effort*.

Listening effort refers to the cognitive processing load associated with listening. In the context of this thesis the focus is specifically on the effort related to speech understanding. For NH listeners, speech understanding in ideal listening conditions seems to be effortless and automatic (Mattys, Davis, Bradlow, & Scott, 2012; Wild et al., 2012). Speech that is degraded, whether due to factors internal or external to the listener, however, does not match the listener's phonological representations in long-term memory, and requires increased cognitive processing for the interpretation of the message (e.g. Lunner & Sundewall-Thorén, 2007; Rönnberg, Rudner, Foo, & Lunner, 2008; Wild et al., 2012; Wingfield & Tun, 2007). CI mediated speech, due to both technical and neural limitations of electric stimulation, results in perceptual representations that are spectro-temporally degraded, i.e. sparser in information regarding the frequency content and timing of the signal, and distinctly different from NH (e.g. Blamey et al., 1992; Stickney et al., 2006). Especially for postlingually deafened CI users, who formed their speech representations in long-term memory based on normal acoustic speech input before losing their hearing, this degraded input from electric stimulation can lead to a mismatch between the incoming speech and representations in long-term memory. Speech understanding in otherwise favorable listening conditions may therefore already be more effortful for CI users than for NH listeners, which is supported by research using CI simulated speech in NH listeners (Wagner, Pals, de Blecourt, Sarampalis, & Başkent, 2016). When the speech signal is additionally degraded, for example due to interfering background noise, resolving the mismatch may be even more effortful.

Increased listening effort serves the purpose of maintaining speech understanding, but at what cost? The additional recruitment of limited cognitive resources for effortful listening reduces the spare capacity available and limits, therefore, the performance on simultaneous tasks and cognitive processes. Even between listening conditions that result in similar intelligibility, differences in listening effort have been shown to affect performance on a secondary task (e.g. Broadbent, 1958; Sarampalis, Kalluri, Edwards, & Hafter, 2009), short-term memory for the correctly heard speech (e.g. McCoy et al., 2005; Pichora-Fuller, Schneider, & Daneman, 1995; Rabbitt, 1966, 1991), which in turn affects discourse comprehension (Pichora-Fuller, 2003), and long-term episodic memory of the speech (Sörqvist & Rönnberg, 2012). In other words, while the listener appears to understand what is said at that moment, the actual comprehension and later memory for the message may be compromised. One can imagine, then, effortful listening could potentially affect academic or professional performance of the listener (Ljung, Sörqvist, Kjellberg, & Green, 2009; Van Engen & Peelle, 2014). In addition to these cognitive consequences, sustained periods of effortful listening, for example in a noisy work environment, can lead to mental fatigue (Hornsby, 2013; McGarrigle et al., 2014), and has been shown to correlate with stress-related sick-leave from work (Kramer, Kapteyn, & Houtgast, 2006). In short, effortful listening can lead to a broad array of negative consequences for the listener and their active participation in society (Hua et al., 2014).

To come back to the population of interest in this thesis: how much is actually known about listening effort in CI users? Although a large body of scientific work had traditionally explored effects of CI processing and device configurations on speech intelligibility (e.g. Fu, Shannon, & Wang, 1998; Spriet et al., 2007; Wilson et al., 1991), at the outset of this project, little had been published on *listening effort* and CIs. Measures of speech intelligibility, such as the percentage of correctly repeated words or sentences, or the signal-to-noise ratio (SNR) that results in a certain level of intelligibility, reflect the end result of all the perceptual and cognitive processes involved in speech understanding. Yet, they do not reveal the nature or magnitude of cognitive processing, or listening effort, that was invested to reach this level of speech understanding. Measures of listening effort can complement intelligibility measures to reveal the otherwise hidden cost of increased cognitive processing load for speech understanding in challenging conditions (e.g. Gatehouse & Gordon, 1990; Larsby, Hällgren, Lyxell, & Arlinger, 2005; McGarrigle et al., 2014), and can therefore, potentially, shed new light on how the process of speech perception is affected by different aspects of hearing with a CI, specifically when intelligibility measures fail to show an effect.

To summarize, when the incoming auditory signal does not match the phonological representations in long-term memory, as it may well be the case for CI listeners, speech understanding requires increased cognitive processing, i.e. listening effort. This increased processing load does not necessarily lead to a loss of intelligibility, and may therefore go undetected. However, even if the end result intelligibility is not affected, increased listening effort can have detrimental consequences for the listener, such as reduced memory for the heard speech, which can, for example, affect discourse comprehension, and for example, lead to poor academic performance. Listening effort can, therefore, have a significant impact on the lives of CI users. Yet relatively little is known about how CI processing affects listening effort. This gives rise to the following questions. At a scientific level; how do intelligibility and listening effort interrelate for different CI configurations, individuals, or listening situations? At a more practical level; can a measure of listening effort uncover hidden benefits of CI processing that are not revealed by measures of intelligibility? Finally at a clinical level: does any such measure of listening effort seem promising for clinical applications?

THIS THESIS

The aim of this thesis was to systematically explore how differences in CI processing or device configurations can affect intelligibility, which was extensively studied before, as well as listening effort, which had not been studied before. More specifically, this thesis explores whether and how different CI settings might affect listening effort when intelligibility shows no change. The studies progressed from simple and more controlled, e.g. testing NH listeners using CI simulations quiet, to more complex and closer to real-life, e.g., including speech perception in background noise, and eventually examining effects of actual CI settings in actual CI users. The first two studies were conducted using normal-hearing participants presented with CI simulated speech, to control for much of the between-CI-user variability. Thus, these studies could focus strictly on the effects of spectral resolution on listening effort for speech in otherwise optimal listening conditions (Chapter 2) or the effect of added low-frequency acoustic speech to simulated CI speech, thus simulating electric-acoustic stimulation (EAS), on speech perception in quiet and in noise (Chapter 3). The next study explored the effect of background noise and noise-type on the perception of clear, unprocessed speech for NH listeners (Chapter 4). In the final chapter, the insights gained in the first three studies were combined and applied to investigate how spectral resolution affects listening effort in CI users, manipulating spectral resolution by varying the number of active electrodes of the CI (Chapter 5).

The main method chosen to objectively quantify listening effort was a dual-task paradigm combining a primary intelligibility task with a secondary visual response-time task. In a dual-task paradigm two tasks compete for cognitive resources and performance on the secondary task therefore reveals the cognitive processing load for the primary listening task (e.g. Baddeley & Hitch, 1974; Kahneman, 1973; Tyler, Hertel, McCallum, & Ellis, 1979), much in the same way effortful listening can interfere with concurrent tasks in real-life situations. The dual-task paradigm's long history of use as a measure for cognitive effort in hearing research (e.g. Broadbent, 1958), relative ease of implementation, the fact that it requires no expensive equipment, and its ability to measure intelligibility and listening effort simultaneously, made it a good starting point in the search for a suitable scientific, as well as clinical, tool. In the earlier studies described in this thesis (Chapter 2 and 3) the dual-task was complemented with a subjective self-report measure of listening effort to compare the perceived effort with the objectively measured cognitive processing load. In the later studies (Chapter 3 and 4), simple response-time measures, such as verbal response-times, reflecting cognitive processing time (Gatehouse & Gordon, 1990) and the sentence verification task, reflecting comprehension and processing time (Baer, Moore, & Gatehouse, 1993), compared with the dual-task paradigm as potential measures of listening effort. Such simple response time measures, due to the simplicity of the tasks, might be more widely applicable in clinical settings.

As the main interest of this thesis was to investigate 'hidden' effects, i.e. effects that are not revealed by traditional intelligibility measures, the studies were designed such that the effects of experimental parameters on listening effort could be compared at equal levels of intelligibility. To recapitulate, the parameters investigated in this thesis were spectral resolution in NH listeners (Chapter 2) and in CI users (Chapter 5), simulated EAS compared to CI alone in NH listeners (Chapter 3), and background noise and noise type in NH listeners (Chapter 4).

THE FINDINGS

In general, the *intelligibility results* confirmed our expectations based on the literature. Spectral resolution, in both NH listeners and CI users, has been shown to improve sentence intelligibility in quiet until reaching ceiling performance at a relatively low number of spectral channels (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Başkent, & Wang, 2001). In this thesis, for NH listeners increased spectral resolution improved sentence intelligibility in quiet up to 6 spectral channels (Chapter 2), and in CI users, 7 active electrodes produced near-ceiling intelligibility, showing no improvement for further increased spectral resolution (Chapter 5). As for the effect of simulated EAS compared to CI, the literature shows improved speech perception in noisy listening conditions (e.g. Brown & Bacon, 2009; Dorman, Spahr, Loizou, Dana, & Schmidt, 2005; Kong & Carlyon, 2007). In this thesis, adding low-frequency acoustic speech to CI simulated speech to simulate EAS improved speech perception in background noise, thus allowing for equal intelligibility at lower SNRs than for CI alone (Chapter 3). In Chapter 4, the effects of background noise and noise type were investigated in NH listeners for speech masked by steady-state, speech-shaped noise (SSN) and 8-talker babble. For NH listeners, understanding speech masked by 8-talker babble required a higher SNR than speech masked by steady-state, speech-shaped noise (SSN) to achieve the same level of target speech intelligibility, in line with previous research that shows better speech intelligibility in SSN compared to 8-talker babble (Chapter 4; Lecumberri & Cooke, 2006). In summary, in each of the studies in this thesis the intelligibility results were in line with expectations based on the literature, and intelligibility was successfully fixed at the desired levels for the conditions of interest.

In the first two chapters a *subjective self-report measure of listening effort* was administered alongside with the objective dual-task measure. In clinical settings the fit of hearing devices is usually mainly evaluated using intelligibility measures and for additional information, clinicians rely on subjective reports made by the patient. Therefore, if a subjective self-report measure could reliably reveal ‘hidden’ effects, effects not reflected by the traditional intelligibility measures, this could be useful in clinical settings. In this thesis, however, the results of the subjective measures were observed to closely follow the intelligibility results and revealed no additional effects (Chapter 2 and 3), while, as will be further described below, the objective measures of listening effort did reveal additional effects. Prior research often shows that subjective effort measures appear not to correlate with objective measures (e.g. Anderson Gosselin & Gagné, 2010; Feuerstein, 1992; Fraser, Gagné, Alepins, & Dubois, 2010). In fact, subjective estimates of effort often appear to reflect perceived accuracy (e.g. Feuerstein, 1992), as was also the case in this thesis. Subjective self-report, while convenient, may therefore not be the most suitable measure for revealing effects that are not already reflected in intelligibility measures.

THE DUAL-TASK MEASURE OF LISTENING EFFORT, RESULTS AND IMPLICATIONS

Unlike the subjective measure of listening effort, the *objective, dual-task measure of listening effort* did successfully reveal effects of simulated CI processing and device configurations on listening effort in NH listeners, even at equal intelligibility. In Chapter 2, increased spectral resolution for speech in quiet resulted in significantly improved secondary task performance for NH listeners up to 8 spectral channels while intelligibility reached ceiling at 6 channels. In Chapter 3, simulated EAS compared to CI alone, improved listening effort in quiet listening conditions at near-ceiling intelligibility and in SSN at 50% intelligibility. In a dual-task paradigm, improved performance on the secondary task implies that the primary task required fewer cognitive resources, thus leaving more resources available to allocate to the execution of the secondary task (e.g. Baddeley & Hitch, 1974; Kahneman, 1973; Tyler, Hertel, McCallum, & Ellis, 1979). These dual-task results thus show that, compared to high spectral resolution, poor spectral resolution requires increased cognitive processing, even if speech intelligibility is at ceiling. Furthermore, the results show that both increasing spectral resolution, or providing additional low frequency acoustic speech, as in EAS, can potentially reduce the cognitive processing load of speech understanding, even when these improvements do not appear to further benefit intelligibility.

These findings may be explained using the ease-of-language-understanding (ELU) model (Rönnberg et al., 2013, 2008). The ELU predicts little or no interference with concurrent tasks for the perception of well-formed speech. However, when the auditory signal does not match representations in long-term memory, explicit cognitive processing is required to resolve this mismatch (Rönnberg et al., 2013, 2008). The more mismatches occur, the more cognitive processing is required, thus depleting resources for concurrent tasks or cognitive processes. Empirical evidence does indeed suggest that only those portions of the bottom-up speech signal that cannot be matched with top-down predictions are passed on for further, higher-order cognitive processing (Sohoglu, Peelle, Carlyon, & Davis, 2012; Van Engen & Peelle, 2014). While intelligibility for 6-channel CI simulated speech is at ceiling, the interaction with the secondary task thus suggests that the reduced spectral resolution of the speech signal does still result in mismatches with the speech representations in the listener's long-term memory, and therefore requires increased cognitive processing for interpretation. Increasing spectral resolution of the CI simulated speech up to 8 spectral channels significantly reduced interfere between the speech task and the secondary task (Chapter 2), suggesting the 8-channel CI simulated speech is more similar to the listener's representations. Providing low frequency acoustic speech in addition to the CI simulated speech signal similarly improved secondary task performance (Chapter 3).

Improving spectral resolution beyond 8 spectral channels, however, did not lead to further increase in secondary task performance (Chapter 2), and while EAS improved secondary task performance compared to simulated CI alone, no distinction could be made between the different EAS simulations combining 300 Hz or 600 Hz low-pass filtered speech with CI simulated speech (Chapter 3). However, this absence of dual-task interaction with the secondary task does not necessarily imply 'effortless' listening. The secondary-task performance for listening conditions with 8 spectral channels and up resulted in similar performance as when the task was performed without a simultaneous speech task suggesting ceiling performance. Interpreting the speech may still require increased cognitive processing compared to NH. As long as the combined processing load of the primary listening task and the secondary task do not exceed the limit of available processing resources, this increased processing load of the listening task will not affect the secondary task.

While the dual-task paradigm successfully revealed effects of spectral resolution on listening effort in NH participants, in CI users it was not as successful. In CI users, changing the spectral resolution by varying the number of active electrodes between 7 and 15 did not affect secondary task performance (Chapter 5). While this could be due to a number of limitations of the dual-task paradigm, which were discussed in chapter 5, as well as the same ceiling effect discussed above, the ELU might provide a different explanation. For the CI users, due to their experience listening with a CI for at least one year, and some participants even many more years, their phonological representations may have adapted to the input they hear every day.

The change from their normal, full electrode array to a limited subset of those electrodes may have been less drastic than the change for NH participants from normal acoustic hearing to the 6-channel, noise-band vocoded CI simulations. The 7-active-electrode speech input may therefore not have required increased cognitive processing to resolve the mismatch with representations in long-term memory to the degree that this processing requirement interfered with secondary task performance.

However, as will be further detailed below, the results of the sentence verification task, did suggest the decreased spectral resolution for CI users does affect speech comprehension and the required processing time. This need not contradict the explanation above, since as long as the limit of the total available resources has not been reached, increases in listening effort do not affect the secondary task.

SIMPLE RESPONSE TIME MEASURES OF LISTENING EFFORT, RESULTS AND IMPLICATIONS

In Chapters 4 and 5, two simple response time measures of listening effort were compared with the dual-task paradigm. Effortful listening has been suggested to increase not only the need for cognitive processing resources, but also lead to increased processing time (e.g. Francis & Nusbaum, 2009; Rönnberg et al., 2013; A. E. Wagner, Toffanin, & Başkent, 2016). Previous research has successfully used response times to speech as a measure that complements intelligibility measure, and reflects 'ease of listening' (Baer et al., 1993; Gatehouse & Gordon, 1990).

In Chapter 4, 'verbal response times' were used, i.e. the time required to start repeating a sentence after it was heard (Gatehouse & Gordon, 1990). This measure can be collected during a traditional speech intelligibility task by recording the verbal responses and scoring manually or using speech detection software. The effects of noise and noise type at different intelligibility levels on listening effort were measured using both the dual-task paradigm and the verbal response times. The dual-task results showed reduced secondary task performance when noise was present, but did not differentiate between the noise types or between 79% and near-ceiling intelligibility. Similar to the dual-task paradigm, the verbal response times showed an effect of the presence of noise; responses were slower for speech in noise than for speech in quiet. Unlike the dual-task paradigm, the verbal response times also showed a significant difference between the 79% and near-ceiling intelligibility. Neither measure showed an effect of noise type at fixed intelligibility levels (Chapter 4). These results suggest that verbal response times may be at least as suitable for measuring listening effort as the dual-task paradigm employed in this thesis, and its ease of implementation could make it a practical method for clinical use.

In Chapter 5, a sentence verification task was used. In this task the participant is presented with sentences that are either unmistakably true or nonsense, and they respond by pressing a key indicating whether the sentence was true or false (Adank & Janse, 2009; Baer et al., 1993). The sentence verification task accuracy and response times reflect both comprehension and processing speed, respectively. The task is simple, easy to explain and perform, and the response collection is easily automated, making it an appealing candidate for a clinical tool. In CI users, increased spectral resolution from 7 active electrodes up did not affect the dual-task measures of intelligibility or listening effort. The sentence verification task, on the other hand, revealed a clear improvement in both speech comprehension and speed-of-processing for increasing the number of active electrodes from 7 to 11 (Chapter 5). The same trend of improved comprehension and speed-of-processing was found in NH listeners for improved spectral resolution up to 16 spectral channels (Chapter 5), while in a different group of (similar) NH participants the dual-task measure showed improvement only up to 6 channels for intelligibility and 8 channels for listening effort (Chapter 2). In part, these differences in results could be attributed to the different speech materials used for the sentence verification task and the dual-task paradigm. Nevertheless, these results for both NH and CI listeners establish the sentence verification task as a likely candidate for a clinical measure of listening effort.

One challenge arises from the sentence verification task results: not only the response times improved with increased spectral resolution when the dual-task measures of listening effort and intelligibility had not, but so did the accuracy scores. The accuracy scores can be interpreted to indirectly reflect intelligibility; when the sentence is not heard correctly and the answer must be guessed, there is a 50% chance of getting a wrong answer. Thus, does this mean that for the speech materials used in the sentence verification task, intelligibility was affected by the decreased spectral resolution even when intelligibility in the dual-task was at ceiling? Possibly. However, the accuracy scores on the sentence verification task do not purely reflect intelligibility. In order to correctly answer whether the sentence was true or false/nonsense, the listener has to *comprehend* the sentence. Comprehension goes beyond just perception, in that the meaning has to have been understood and the participant has to reason about the meaning of the sentence (Ralston, Pisoni, Lively, Greene, & Mullennix, 1991; Wingfield & Tun, 2007). This requires further cognitive processing of the speech than only individually perceived phonemes or words, and therefore likely relies more on cognitive capacity (Just & Carpenter, 1992). The reduction of cognitive resources due to increased listening effort may therefore also affect comprehension.

Regardless of whether the accuracy scores on the sentence verification task reflect intelligibility of the sentence materials, listening effort, or both, the results show the value of additional measures. Based on the intelligibility or dual-task listening effort results, one would have had to conclude increased spectral resolution beyond 7 active electrodes in CI users, or 8 spectral channels for NH listeners, no longer improves speech perception. The sentence verification task revealed both improved comprehension and processing speed for further increased spectral resolution in both participant groups.

SUMMARY AND CONCLUSIONS

The results for the dual-task paradigm and the simple response-time measures of listening effort illustrate the added value of measures that tap into effort and cognitive processes involved in speech understanding to complement the traditional intelligibility measures. The studies described in this thesis were designed such that the effects of CI processing and device configurations on listening effort could be compared at equal intelligibility. Subjective self-report closely followed intelligibility and revealed no additional effects. The dual-task did show that secondary task performance could be affected differently for conditions that did produce similar intelligibility, thus illustrating the value of a measure of listening effort to complement intelligibility measures. The sentence verification task further revealed effects where the dual-task did not. This may be due to a limitation of the dual-task; when the listening task and secondary task combined do not require all available cognitive resources to be performed simultaneously, the tasks will not interact, and changes in listening effort will not affect secondary task performance. All in all, the sentence verification task appears to be a useful behavioral measure to investigate speech comprehension and listening effort in both NH listeners and CI users, and may prove to be a suitable tool for clinical use.

As for CI processing and device configurations, our studies in NH listeners suggest that reduced spectral resolution, results in increased processing load for speech perception. Both increasing spectral resolution (Chapter 2) and providing low pass filtered acoustic speech to simulate EAS (Chapter 3) can improve listening effort. No distinction could be made between the different EAS conditions, however, for future research it might be interesting to revisit this question using the sentence verification task to see if further hidden effects can be uncovered. While these strategies appear to work well in simulations and NH listeners, they may not be easily realizable in CI users. The source of reduced spectral resolution in CI users is not only related to the device, but also due to factors such as dead regions in the cochlea and current spread, which limits the CI users' effective use of spectral information contained in the signal (Fu et al., 1998; Stickney et al., 2006). Similarly, EAS may not be an option for many CI users, as they may not have any residual hearing. The results described in Chapter 5, however, are hopeful; the sentence verification task did show improved comprehension for increased spectral resolution from 7 up to 11 active electrodes in CI users.

REFERENCES

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), 2649–59. doi:10.1121/1.3216914

Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. a. (2004). Aging: a switch from automatic to controlled processing of sounds? *Psychology and Aging*, *19*(1), 125–33. doi:10.1037/0882-7974.19.1.125

Angeli, S. I., Yan, D., Telischi, F., Balkany, T. J., Ouyang, X. M., Du, L. L., … Liu, X. Z. (2005). Etiologic diagnosis of sensorineural hearing loss in adults. *Otolaryngology--Head and Neck Surgery : Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, *132*(6), 890–5. doi:10.1016/j.otohns.2005.03.001

Arehart, K. H., Souza, P., Baca, R., & Kates, J. M. (2013). Working memory, age, and hearing loss: susceptibility to hearing aid distortion. *Ear and Hearing*, *34*(3), 251–60. doi:10.1097/AUD.0b013e318271aa5e

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, *19*(1), 29–56. doi:10.1080/01690960344000099

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

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. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

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

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

Baddeley, A. D., & Salamé, P. (1986). The unattended speech effect: perception or memory? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *12*(4), 525–9.

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Başkent, D. (2006). Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *The Journal of the Acoustical Society of America*, *120*(5 Pt 1), 2908–25. doi:10.1121/1.2354017

Başkent, D. (2012). Effect of Speech Degradation on Top-Down Repair: Phonemic Restoration with Simulations of Cochlear Implants and Combined Electric-Acoustic Stimulation. *Journal of the Association for Research in Otolaryngology : JARO*, *13*(5), 683–92. doi:10.1007/s10162-012-0334-3

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*.

Başkent, D., Gaudrain, E., Tamati, T., & Wagner, A. E. (2016). Perception and psychoacoustics of speech in cochlear implant users. In *Scientific Foundations of Audiology*.

Başkent, D., & Shannon, R. V. (2007). Combined effects of frequency compression-expansion and shift on speech recognition. *Ear and Hearing*, *28*(3), 277. doi:10.1097/AUD.0b013e318050d398

Benard, M. R., & Başkent, D. (2013). Perceptual Learning of Interrupted Speech. *PloS One*, *8*(3). doi:10.1371/journal.pone.0058149

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

Bhargava, P., Gaudrain, E., & Başkent, D. (2014). Top-down restoration of speech in cochlear-implant users. *Hearing Research*, *309*, 113–23. 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 : JARO*. doi:10.1007/s10162-016-0565-9

Blamey, P. J., Artières, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., … Lazard, D. S. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. *Audiology & Neuro-Otology*, *18*(1), 36–47. doi:10.1159/000343189

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), 342–8.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of* ….

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

Brown, C. a, & Bacon, S. P. (2009). Low-frequency speech cues and simulated electric-acoustic hearing. *The Journal of the Acoustical Society of America*, *125*(3), 1658–65. doi:10.1121/1.3068441

Büchner, A., Schüssler, M., Battmer, R. D., Stöver, T., Lesinski-Schiedat, A., & Lenarz, T. (2009). Impact of low-frequency hearing. *Audiology & Neuro-Otology*, *14 Suppl 1*(suppl 1), 8–13. doi:10.1159/000206490

Caissie, A. F., Vigneau, F., & Bors, D. a. (2009). What does the Mental Rotation Test Measure? An Analysis of Item Difficulty and Item Characteristics. *The Open Psychology Journal*, *2*(1), 94–102. doi:10.2174/1874350100902010094

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *The Behavioral and Brain Sciences*, *22*(1), 77–94; discussion 95–126.

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), EL37–41. doi:10.1121/1.3284544

Classon, E., Rudner, M., & Rönnberg, J. (2013). Working memory compensates for hearing related phonological processing deficit. *Journal of Communication Disorders*, *46*(1), 17–29. doi:10.1016/j.jcomdis.2012.10.001

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

Dooley, G. J., Blamey, P. J., Seligman, P. M., Alcántara, J. I., Clark, G. M., Shallop, J. K., … Menapace, C. M. (1993). Combined electrical and acoustical stimulation using a bimodal prosthesis. *Archives of Otolaryngology--Head & Neck Surgery*, *119*(1), 55–60.

Dorman, M. F., & Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *International Journal of Audiology*, *49*(12), 912–9. doi:10.3109/14992027.2010.509113

Dorman, M. F., Loizou, P. C., Fitzke, J., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. *The Journal of the Acoustical Society of America*, *104*(6), 3583–5.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *The Journal of the Acoustical Society of America*, *102*(4), 2403–11.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., & Schmidt, J. S. (2005). Acoustic simulations of combined electric and acoustic hearing (EAS). *Ear and Hearing*, *26*(4), 371–80.

Downs, D. W., & Crum, M. a. (1978). Processing demands during auditory learning under degraded listening conditions. *Journal of Speech and Hearing Research*, *21*(4), 702–14.

Dudley, H. (1939). The automatic synthesis of speech. *Proceedings of the National Academy of Sciences of the United States of America*, *25*(7), 377. doi:10.1073/pnas.25.7.377

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*(6635), 808–10. doi:10.1038/42947

Dyson, B. J., Alain, C., & He, Y. (2005). Effects of visual attentional load on low-level auditory scene analysis. *Cognitive, Affective & Behavioral Neuroscience*, *5*(3), 319–38.

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

Epstein, M., & Florentine, M. (2012). Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. *The Journal of the Acoustical Society of America*, *131*(5), 3981–8. doi:10.1121/1.3701984

Feuerstein, J. F. (1992). Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort. *Ear and Hearing*, *13*(2), 80–6.

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 Researchlanguage, and Hearing Research : JSLHR*, *40*(5), 1201–15.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception & Psychophysics*, *71*(6), 1360–74. doi:10.3758/APP.71.6.1360

Fraser, S., Gagné, J.-P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research : JSLHR*, *53*(1), 18–33. doi:10.1044/1092-4388(2009/08-0140)

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. (2002). Temporal processing and speech recognition in cochlear implant users. *Neuroreport*, *13*(13), 1635–9.

Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology : JARO*, *6*(1), 19–27. doi:10.1007/s10162-004-5024-3

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), 3586–96.

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

Gatehouse, S. (1990). The role of non-auditory factors in measured and self-reported disability. *Acta Oto-Laryngologica. Supplementum*, *476*(3), 249–56.

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology*, *24*(1), 63–8.

Getzmann, S., & Falkenstein, M. (2011). Understanding of spoken language under challenging listening conditions in younger and older listeners: a combined behavioral and electrophysiological study. *Brain Research*, *1415*, 8–22. doi:10.1016/j.brainres.2011.08.001

Gibbon, D., Moore, R., & Winski, R. (1997). *Handbook of standards and resources for spoken language systems*. (D. Gibbon, R. Moore, & R. Winski, Eds.) (p. 886). Berlin, DE: DE GRUYTER MOUTON.

Gosselin, P. A., & Gagné, J.-P. (2010). Use of a dual-Task paradigm to measure listening effort. Utilisation d'un paradigme de double tâche pour mesurer l'attention auditive. *Revue Canadienne d'Orthophonie et d'Audiologie*, *34*(1), 43–51.

Gosselin, P. A., & Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research : JSLHR*, *54*(3), 944–58. doi:10.1044/1092-4388(2010/10-0069)

Goy, H., Pelletier, M., Coletta, M., & Pichora-Fuller, M. K. (2013). The effects of semantic context and the type and amount of acoustic distortion on lexical decision by younger and older adults. *Journal of Speech, Language, and Hearing Research : JSLHR*, *56*(6), 1715–32. doi:10.1044/1092-4388(2013/12-0053)

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), 2592–605.

Gstoettner, W. K., Kiefer, J., Baumgartner, W.-D., Pok, S., Peters, S., & Adunka, O. F. (2004). Hearing preservation in cochlear implantation for electric acoustic stimulation. *Acta Oto-Laryngologica*, *124*(4), 348–52. doi:10.1080/00016480410016432

Gstoettner, W. K., Van de Heyning, P. H., O'Connor, A. F., Morera, C., Sainz, M., Vermeire, K., … Adunka, O. F. (2008). Electric acoustic stimulation of the auditory system: results of a multi-centre investigation. *Acta Oto-Laryngologica*, (792024350), 1–8. doi:10.1080/00016480701805471

Hannemann, R., Obleser, J., & Eulitz, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research*, *1153*, 134–43. doi:10.1016/j.brainres.2007.03.069

Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload*, *1*, 139–183.

Hecker, M. H., Stevens, K. N., & Williams, C. E. (1966). Measurements of reaction time in intelligibility tests. *The Journal of the Acoustical Society of America*, *39*(6), 1188–9.

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, *28*(3), 376–85.

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

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, *27*(7-8), 1145–1166. doi:10.1080/01690965.2012.662280

Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. W. (2007). Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiology & Neuro-Otology*, *12*(4), 254–64. doi:10.1159/000101473

Hockey, G. R. (1997). Compensatory control in the regulation of human performance under stress and high workload; a cognitive-energetical framework. *Biological Psychology*, *45*(1-3), 73–93.

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*. doi:10.1097/AUD.0b013e31828003d8

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2012). Measuring listening effort with digits in noise.

Hoyek, N., Collet, C., Fargier, P., & Guillot, A. (2012). The Use of the Vandenberg and Kuse Mental Rotation Test in Children. *Journal of Individual Differences*, *33*(1), 62–67. doi:10.1027/1614-0001/a000063

Hua, H., Emilsson, M., Kähäri, K., Widén, S., Möller, C., & Lyxell, B. (2014). The impact of different background noises: effects on cognitive performance and perceived disturbance in employees with aided hearing impairment and normal hearing. *Journal of the American Academy of Audiology*, *25*(9), 859–68. doi:10.3766/jaaa.25.9.8

Jansen, P. a., & Watter, S. (2008). SayWhen: An automated method for high-accuracy speech onset detection. *Behavior Research Methods*, *40*(3), 744–751. doi:10.3758/BRM.40.3.744

Just, M. a, & Carpenter, P. a. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122–49.

Kahneman, D. (1973). Attention and effort. In *Measurement*. Englewood Cliffs, N.J., Prentice-Hall.

Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*(2), 175–185. doi:10.1111/j.1469-8986.2003.00147.x

Kempler, D., Almor, A., Tyler, L. K., Andersen, E. S., & MacDonald, M. C. (1998). Sentence comprehension deficits in Alzheimer's disease: a comparison of off-line vs. on-line sentence processing. *Brain and Language*, *64*(3), 297–316. doi:10.1006/brln.1998.1980

Kiefer, J., Pok, M., Adunka, O. F., Stürzebecher, E., Baumgartner, W.-D., Schmidt, M., … Gstoettner, W. K. (2005). Combined electric and acoustic stimulation of the auditory system: results of a clinical study. *Audiology & Neuro-Otology*, *10*(3), 134–44. doi:10.1159/000084023

Klop, W. M. C., Briaire, J. J., Stiggelbout, A. M., & Frijns, J. H. M. (2007). Cochlear implant outcomes and quality of life in adults with prelingual deafness. *The Laryngoscope*, *117*(11), 1982–7. doi:10.1097/MLG.0b013e31812f56a6

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291–300. doi:10.1097/AUD.0b013e3182310019

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology*, *2012*, 865731. doi:10.1155/2012/865731

Kong, Y.-Y., & Carlyon, R. P. (2007). Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *The Journal of the Acoustical Society of America*, *121*(6), 3717–27. doi:10.1121/1.2717408

Kong, Y.-Y., Stickney, G. S., & Zeng, F.-G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *The Journal of the Acoustical Society of America*, *117*(3 Pt 1), 1351–61. doi:10.1121/1.1857526

Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *International Journal of Audiology*, *45*(9), 503–12. doi:10.1080/14992020600754583

Kramer, S. E., Zekveld, A. A., & Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: the value of using the Text Reception Threshold test as a visual equivalent of the SRT test. *Scandinavian Journal of Psychology*, *50*(5), 507–15. doi:10.1111/j.1467-9450.2009.00747.x

Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. a. (2012). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*. doi:10.1111/j.1469-8986.2012.01477.x

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology*, *44*(3), 131–143. doi:10.1080/14992020500057244

Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P. H., Truy, E., Sterkers, O., … Blamey, P. J. (2012). Pre-, Per- and Postoperative Factors Affecting Performance of Postlinguistically Deaf Adults Using Cochlear Implants: A New Conceptual Model over Time. *PloS One*, *7*(11), e48739. doi:10.1371/journal.pone.0048739

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445. doi:10.1121/1.2180210

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2), Suppl 2:467+.

Li, K. Z. ., & Lindenberger, U. (2002). Relations between aging sensory/sensorimotor and cognitive functions. *Neuroscience & Biobehavioral Reviews*, *26*(7), 777–783. doi:10.1016/S0149-7634(02)00073-8

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging*, *9*(3), 339–55.

Ljung, R., Sörqvist, P., Kjellberg, A., & Green, A. (2009). Poor Listening Conditions Impair Memory for Intelligible Lectures : Implications for Acoustic Classroom Standards, *16*(3), 257–265.

Loizou, P. C. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine*, *15*(5), 101–130. doi:10.1109/79.708543

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, *42 Suppl 1*(s1), S49–58. doi:10.3109/14992020309074624

Lunner, T., Rudner, M., & Rönnberg, J. (2009). Cognition and hearing aids. *Scandinavian Journal of Psychology*, *50*(5), 395–403. doi:10.1111/j.1467-9450.2009.00742.x

Lunner, T., Rudner, M., Rosenbom, T., Ågren, J., & Ng, E. H. N. (2016). Using Speech Recall in Hearing Aid Fitting and Outcome Evaluation Under Ecological Test Conditions. *Ear and Hearing*, *37 Suppl 1*(Wilson 2003), 145S–54S. doi:10.1097/AUD.0000000000000294

Lunner, T., & Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology*, *18*(7), 604–17.

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

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–43. doi:10.1016/j.cogpsych.2009.04.001

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

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. A, Human Experimental Psychology*, *58*(1), 22–33. doi:10.1080/02724980443000151

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper." *International Journal of Audiology*, (November 2013), 1–13. doi:10.3109/14992027.2014.890296

Miller, G., & Licklider, J. (1950). The Intelligibility of Interrupted Speech. *The Journal of the Acoustical Society of …*, *22*(2).

Mok, M., Grayden, D., Dowell, R. C., & Lawrence, D. (2006). Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. *Journal of Speech, Language, and Hearing Research : JSLHR*, *49*(2), 338–51. doi:10.1044/1092-4388(2006/027)

Moore, B. C. J. (2004). Dead Regions in the Cochlea: Conceptual Foundations, Diagnosis, and Clinical Applications. *Ear and Hearing*, *25*(2), 98–116. doi:10.1097/01.AUD.0000120359.49711.D7

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, *11*(2), 296–301.

Nelson, P. B., & Jin, S.-H. (2004). Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, *115*(5 Pt 1), 2286–94. doi:10.1121/1.1703538

Nijboer, M., Taatgen, N. A., Brands, A., Borst, J. P., & van Rijn, H. (2013). Decision making in concurrent multitasking: do people adapt to task interference? *PloS One*, *8*(11), e79583. doi:10.1371/journal.pone.0079583

Otten, M., & Van Berkum, J. J. a. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101. doi:10.1016/j.brainres.2009.07.042

Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research : JSLHR*, *4388*. 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

Pfingst, B. E., Zwolan, T. a, & Holloway, L. a. (1997). Effects of stimulus configuration on psychophysical operating levels and on speech recognition with cochlear implants. *Hearing Research*, *112*(1-2), 247–260. doi:10.1016/S0378-5955(97)00122-6

Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *International Journal of Audiology*, *42 Suppl 2*, 2S26–32.

Pichora-Fuller, M. K., Schneider, B. a, & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–9. doi:10.1111/j.1469-8986.2009.00947.x

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research*, *29*(2), 146–54.

Qin, M. K., & Oxenham, A. J. (2006). Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. *The Journal of the Acoustical Society of America*, *119*(4), 2417–26. doi:10.1121/1.2178719

Rabbitt, P. M. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science*, *6*(8), 383–384.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*(3), 241–8. doi:10.1080/14640746808400158

Rabbitt, P. M. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica*, *111*(s476), 167–176. doi:10.3109/00016489109127274

Rakerd, B., Seitz, P., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing*, *17*(2), 97.

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), 471–91. doi:10.1177/001872089103300408

Rönnberg, 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*, S68–76.

Rönnberg, J., Lunner, T., Zekveld, A. A., Sörqvist, P., Danielsson, H., Lyxell, B., … Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(July), 31. doi:10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47 Suppl 2*(February), S99–105. doi:10.1080/14992020802301167

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Applied Psychology*, *53*(1), 61–86. doi:10.1111/j.1464-0597.2004.00161.x

Rudner, M., Foo, C., Rönnberg, J., & Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scandinavian Journal of Psychology*, *50*(5), 405–18. doi:10.1111/j.1467-9450.2009.00745.x

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, *23*(8), 577–89. doi:10.3766/jaaa.23.7.7

Rudner, M., Rönnberg, J., & Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *Journal of the American Academy of Audiology*, *22*(3), 156–67. doi:10.3766/jaaa.22.3.4

Ruthruff, E., Miller, J., & Lachmann, T. (1995). Does mental rotation require central mechanisms? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 552–570. doi:10.1037/0096-1523.21.3.552

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. doi:10.1016/S0010-0277(03)00139-2

Samuel, A. (1981a). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474–494.

Samuel, A. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. … *of Experimental Psychology: Human Perception and …*, *7*(5), 1124–1131.

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 : JSLHR*, *52*(5), 1230–40. doi:10.1044/1092-4388(2009/08-0111)

Schneider, B. a, & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. *The Handbook of Aging and Cognition*, *2*, 155–219.

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), 3972–88. doi:10.1121/1.2997434

Seeber, B. U., Baumann, U., & Fastl, H. (2004). Localization ability with bimodal hearing aids and bilateral cochlear implants. *The Journal of the Acoustical Society of America*, *116*(3), 1698–709. doi:10.1121/1.1776192

Seidler, R. D., Bernard, J. A., Burutolu, T. B., Fling, B. W., Gordon, M. T., Gwin, J. T., … Lipps, D. B. (2010). Motor control and aging: links to age-related brain structural, functional, and biochemical effects. *Neuroscience and Biobehavioral Reviews*, *34*(5), 721–33. doi:10.1016/j.neubiorev.2009.10.005

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), 489–99. doi:10.1121/1.2783762

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, *12*(4), 283–99. doi:10.1177/1084713808325306

Shtyrov, Y., Kujala, T., & Pulvermüller, F. (2010). Interactions between language and attention systems: early automatic lexical processing? *Journal of Cognitive Neuroscience*, *22*(7), 1465–78. doi:10.1162/jocn.2009.21292

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

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(25), 8443–53. doi:10.1523/JNEUROSCI.5069-11.2012

Sörqvist, P., & Rönnberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: what is the role for working memory capacity? *Journal of Speech, Language, and Hearing Research : JSLHR*, *55*(1), 210–8. doi:10.1044/1092-4388(2011/10-0353)

Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., van Dijk, B., … Wouters, J. (2007). Speech understanding in background noise with the two-microphone adaptive beamformer BEAM in the Nucleus Freedom Cochlear Implant System. *Ear and Hearing*, *28*(1), 62–72. doi:10.1097/01.aud.0000252470.54246.54

Stenfelt, S., & Rönnberg, J. (2009). The Signal-Cognition interface: interactions between degraded auditory signals and cognitive processes. *Scandinavian Journal of Psychology*, *50*(5), 385–93. doi:10.1111/j.1467-9450.2009.00748.x

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

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–95. doi:10.1162/jocn_a_00389

Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455–62.

Tun, P. A., McCoy, S. L., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, *24*(3), 761–6. doi:10.1037/a0014802

Turner, C., Gantz, B. J., Lowder, M. W., & Gfeller, K. E. (2005). Benefits seen in acoustic hearing+ electric stimulation in same ear. *The Hearing Journal*, *58*(11), 53.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory*, *5*(6), 607–617. doi:10.1037/0278-7393.5.6.607

Uus, K., & Bamford, J. (2006). Effectiveness of population-based newborn hearing screening in England: ages of interventions and profile of cases. *Pediatrics*, *117*(5), e887–93. doi:10.1542/peds.2005-1064

Van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A Standard Computerized Version of the Reading Span Test in Different Languages. *European Journal of Psychological Assessment*, *24*(1), 35–42. doi:10.1027/1015-5759.24.1.35

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*(August), 577. doi:10.3389/fnhum.2014.00577

Van Hardeveld, R. Het belang van Cochleaire Implantatie voor gehoorbeperkten - resultaten van een enquete gehouden in 2010 (2010).

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), 348–55. doi:10.1080/14992020801895144

Vermeire, K., Brokx, J. P. L., Wuyts, F. L., Cochet, E., Hofkens, A., & Van de Heyning, P. H. (2005). Quality-of-life benefit from cochlear implantation in the elderly. *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *26*(2), 188–95.

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), 1671–84. doi:10.1121/1.428451

Von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Stürzebecher, E., & Klinke, R. (2000). Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss. *ORL; Journal for Oto-Rhino-Laryngology and Its Related Specialties*, *61*(6), 334–40.

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

Wechsler, D. (2012). Wechsler Adult Intelligence Scale, 4th ed. (WAIS-IV-NL) Dutch version.

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

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. doi:10.1518/001872008X288394

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), 14010–21. doi:10.1523/JNEUROSCI.1528-12.2012

Wilding, J., & White, W. (1985). Impairment of rhyme judgments by silent and overt articulatory suppression. *The Quarterly Journal of Experimental Psychology Section A*, *37*(1), 95–107. doi:10.1080/14640748508400953

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature*, *352*(6332), 236–8. doi:10.1038/352236a0

Wingfield, A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, *7*(3), 175–82.

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), P127–9.

Wingfield, A., Lindfield, K. C., & Goodglass, H. (2000). Effects of Age and Hearing Sensitivity on the Use of Prosodic Information in Spoken Word Recognition. *Journal of Speech, Language, and Hearing Research*, *43*(4), 915.

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), 487–97.

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., 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), 1465–79. 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), e153–65. 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 : JARO*, *8*(3), 384–92. doi:10.1007/s10162-007-0085-8

Wu, Y., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric Functions of Dual-Task Paradigms for Measuring Listening Effort, 1–11.

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), 480–90. doi:10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear and Hearing*, 1–13. doi:10.1097/AUD.0b013e31820512bb

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, *134*(3), 2225–2234. doi:10.1121/1.4817926

Zhang, T., Spahr, A. J., & Dorman, M. F. (2010). Frequency overlap between electric and acoustic stimulation and speech-perception benefit in patients with combined electric and acoustic stimulation. *Ear and Hearing*, *31*(2), 195–201. doi:10.1097/AUD.0b013e3181c4758d

NEDERLANDSE SAMENVATTING

Luisterinspanning, de verborgen voor- en nadelen van horen met een cochleair implantaat

Carina Pals

THEORETISCHE ACHTERGROND

Wanneer je luistert geeft je betekenis aan de geluiden die je hoort. In dit proefschrift kijk ik specifiek naar het luisteren naar, en verstaan van, spraak. Bij spraakverstaan kan de luisteraar van een groot aantal cognitieve mechanismes gebruik maken en uit een grote rijkdom aan kennis en informatie putten. De informatie die de luisteraar ter beschikking staat kan deel uitmaken van het signaal, zoals fluctueringen in frequentie en intensiteit, intonatie, zinsopbouw, en de context van het gesprek, maar kan ook deel uitmaken van de situationele context. Verder kan de luisteraar gebruik maken van kennis; kennis van de taal die wordt gesproken, zoals de grammatica of welke woorden vaker samen voorkomen, of kennis over het gespreksonderwerp. In situaties waarin de spraak moeilijk te verstaan is kunnen nog extra mechanismes gerekruteerd worden om bijvoorbeeld op basis van de verstaanbare delen van de spraak hypotheses te vormen over de ontbrekende delen en zo te achterhalen wat de spreker gezegd kan hebben. Zo kan een luisteraar, zelfs in een rumoerige omgeving, nog steeds verstaan wat haar gesprekspartner zegt. Het aanwenden van al deze kennis om moeilijk hoorbare spraak te verstaan, kost echter wel inspanning. Dit is wat ik in dit proefschrift "luisterinspanning" noem; de mentale inspanning die een luisteraar aanwendt om spraak te verstaan in uitdagende luistersituaties.

De 'uitdagende luistersituaties' die ik zojuist noemde kunnen aan de ene kant voortkomen uit de omgeving, doordat er bijvoorbeeld veel rumoer op de achtergrond is, of de ruimte een erg holle akoestiek heeft waardoor de echo het moeilijk maakt de spraak te verstaan. Aan de andere kant kunnen ook factoren die te maken hebben met de luisteraar zelf bijdragen aan hoe moeilijk spraakverstaan is, bijvoorbeeld doordat de luisteraar slechthorend is. Sommige zeer slechthorende of dove mensen die voldoen aan een aantal criteria zijn mogelijk kandidaat voor een cochleair implantaat (CI). Zo'n implantaat deelt het binnenkomende geluid op in frequentiebanden en vertaalt deze naar series elektrische pulsen die via elektrodes geïmplanteerd in het binnenoor de gehoorzenuw stimuleren (zie Figuur 1). Zo wordt een neuraal signaal gegenereerd dat normaal gehoor nabootst. Echter, door eigenschappen van CIs en bijvoorbeeld de gezondheid van de gehoorzenuw, is zelfs met CI het spraaksignaal minder rijk aan informatie dan normaal gehoor. Hoewel veel CI gebruikers spraak in ideale situaties vaak prima verstaan, is het goed mogelijk dat CI gebruikers toch meer inspanning moeten leveren dan normaalhorenden. Als een CI gebruiker meer inspanning moet leveren om spraak te verstaan, maar deze uiteindelijk wel goed verstaat, is deze extra inspanning niet direct zichtbaar voor de buitenwereld. Vandaar dat de titel van dit proefschrift spreekt van 'verborgen voor- en nadelen'.

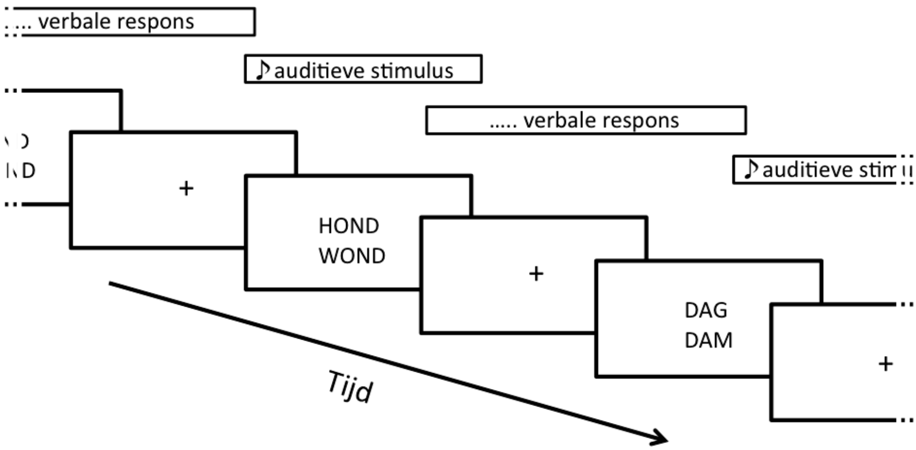
Figuur 1. Schematische weergave van een rechter oor met cochleair implantaat. De processor achter de oorschelp stuurt het vertaalde signaal via de zendspoel op de schedel (donkergrijs) naar de ontvangstspoel onder de huid (transparant grijs), welke het signaal doorstuurt via de elektrodes in het binnenoor naar de gehoorzenuw. Afbeelding Copyright Cochlear Limited ©

Maar wat zijn de gevolgen van hoge luisterinspanning? Waarom beschouw ik dit als een nadeel als het duidelijke voordeel is dat de spraak met de extra inspanning wordt verstaan? De mentale inspanning die we op een bepaald moment kunnen leveren is niet onuitputtelijk. Dit heeft tot gevolg dat als luisteren veel inspanning kost, er minder 'mentale reserve' over is om tegelijkertijd aan andere taken of cognitieve processen te wijden. Dit kan problemen opleveren bij het combineren van spraakverstaan met gelijktijdige taken, denk bijvoorbeeld aan autorijden; letten op het overige verkeer terwijl je luistert naar de bijrijder die routeaanwijzingen geeft. Maar ook minder direct waarneembaar; het luisteren naar een boodschap en deze opslaan in het geheugen. Als spraakverstaan veel inspanning kost kan het zijn dat de luisteraar de boodschap wel verstaat, maar zich later niet meer goed kan herinneren. Naast deze directe gevolgen kan langdurige luisterinspanning ook leiden tot vermoeidheid en ziekteverzuim van werk. Reden dus, om te kijken of luisterinspanning voor CI gebruikers te verbeteren valt.

In dit proefschrift wordt een serie experimenten beschreven die aan de ene kant als doel hebben een breed toepasbare en betrouwbare maat voor luisterinspanning te vinden, en aan de andere kant systematisch in kaart brengen hoe bepaalde aspecten van geluid gehoord via een CI spraakverstaan en luisterinspanning beïnvloeden.

METHODOLOGIE

De bijeffecten van hoge luisterinspanning kunnen verklaard worden vanuit modellen die uitgaan van de aanname dat cognitieve capaciteit gelimiteerd is, zoals bijvoorbeeld het werkgeheugenmodel van Baddeley. De centrale aanname in dit soort modellen is dat een eindige capaciteit aan cognitieve middelen moet worden verdeeld over alle gelijktijdige cognitieve taken en processen. In het onderzoek beschreven in dit proefschrift maak ik gebruik van deze 'gelimiteerde cognitieve capaciteit' aanname om luisterinspanning te meten. Dit doe ik door middel van een 'dual-task paradigm', dat wil zeggen dat een proefpersoon twee taken gelijktijdig moet uitvoeren. Als de primaire taak veel inspanning vergt, blijft er daardoor minder cognitieve restcapaciteit over voor het uitvoeren van de secundaire taak. De score op de secundaire taak geeft dus indirect weer hoe inspannend de primaire taak is; is deze inspannend dan zal de uitvoering van de secundaire taak daaronder lijden. In mijn onderzoek ben ik specifiek geïnteresseerd in luisterinspanning, en de primaire taak betreft daarom dus een luistertaak. Door de stimuli voor deze luistertaak te manipuleren kan ik onderzoeken hoe bepaalde factoren gerelateerd aan horen met een CI spraakverstaan de luisterinspanning beïnvloeden. Als secundaire taak heb ik een visuele rijm taak gebruikt; steeds verschijnen twee woorden op een beeldscherm en de proefpersoon moet zo snel mogelijk aangeven of deze woorden rijmen of niet. Het idee is dat de luistertaak en de rijmtaak beide gebruik maken van cognitieve middelen specifiek voor het verwerken van taal en dus concurreren voor deze middelen. De reactietijd op de rijmtaak dient als maat voor luisterinspanning op de primaire taak. Zodra de luisterinspanning hoger wordt zal de proefpersoon minder cognitieve capaciteit ter beschikking hebben om de rijmtaak uit te voeren en zal dus trager antwoorden. De timing van de stimuli voor beide taken is niet aan elkaar gekoppeld, zo kan de proefpersoon niet anticiperen op de komst van een paar rijmwoorden en zo een strategie ontwikkelen om deze snel te beantwoorden zonder dat de luistertaak daaronder lijdt (Zie Figuur 2).



Figuur 2: Dual-task procedure. De stimuli voor de luistertaak en de visuele rijmtaak worden tegelijkertijd gepresenteerd, maar niet aan elkaar gekoppeld, zo kan dus een paar rijmwoorden zowel tijdens een zin of tussen twee zinnen van de luistertaak door gepresenteerd worden.

Het voordeel van het dual-task paradigma is dat het duidelijk inzichtelijk maakt hoe luisterinspanning de cognitieve capaciteit beperkt voor overige taken. Echter, voor een klinisch toepasbare maat is de dual-task mogelijk minder geschikt. Voor bijvoorbeeld kinderen of ouderen zou het uitvoeren van twee taken gelijktijdig ingewikkeld kunnen zijn. Vandaar dat ik in de latere hoofdstukken alternatieve, eenvoudigere taken introduceer, en deze vergelijk met de dual-task. Het onderzoek beschreven in hoofdstuk 4 focust op de methodologie en evalueert specifiek zo'n eenvoudigere maat. In een experiment met normaal-horende proefpersonen die luisteren naar spraak met of zonder achtergrondruis worden de dual-task en de ‘verbale reactie tijd’ (de tijd die het kost om na het horen van een zin te beginnen deze hardop te herhalen) als maat voor luisterinspanning met elkaar vergeleken. De resultaten van dit experiment laten zien dat zowel de dual-task als de verbale reactie tijd een effect laten zien van de aan- of afwezigheid van achtergrondruis. De verbale reactie tijd laat daarnaast ook een effect zien van de verstaanbaarheid van de spraak; de beter verstaanbare spraak kan sneller worden herhaald. Dit suggereert dat deze verbale reactietijd mogelijk inderdaad geschikt is als maat voor luisterinspanning.

Zoals de titel van dit proefschrift suggereert was ik met name geïnteresseerd in ‘verborgen’ effecten van luisteren met een CI op luisterinspanning. Het doel van de onderzoeken gepresenteerd in dit proefschrift was dan ook om verschillende factoren die bijdragen aan de geluidskwaliteit van een CI te testen op hun effect op luisterinspanning – *specifiek voor condities die geen verschil teweegbrengen in spraakverstaan*.

SPECTRALE RESOLUTIE EN LUISTERINSPANNING

Een van de belangrijkste aspecten van de kwaliteit van CI geluid is de ‘spectrale resolutie’, dat is de rijkheid aan frequentie-informatie in het signaal. Door de eigenschappen van een CI, maar ook de verminderde gezondheid van de gehoorzenuw, is voor CI gebruikers deze frequentieresolutie merkbaar minder dan voor normaal horenden. Het effect van verminderde spectrale resolutie op spraakverstaan is al veel onderzocht en goed in kaart gebracht. Het is duidelijk dat onder een bepaalde grens de spraakverstaanbaarheid omlaag gaat met verder gereduceerde frequentieresolutie. Echter, hoe dit de luisterinspanning beïnvloedt was tot voorheen niet onderzocht. Het lijkt aannemelijk dat als de frequentieresolutie dusdanig laag is dat de verstaanbaarheid omlaag gaat, dat dan ook de luisterinspanning om nog zo veel mogelijk te verstaan omhoog gaat. Maar waar ik met name in geïnteresseerd was, is of de luisterinspanning nog verder verbeterd met hogere frequentieresolutie als de spraakverstaanbaarheid al optimaal is.

Aangezien bij CI gebruikers meerdere factoren invloed hebben op de frequentieresolutie van een spraaksignaal en er dus tussen CI gebruikers verschillen kunnen zijn, is het eerste onderzoek, beschreven in hoofdstuk 2, uitgevoerd met normaalhorenden en CI simulaties. Op deze manier kon ik zoveel mogelijk factoren constant houden tussen proefpersonen. CI-geluid werd gesimuleerd met verschillende aantallen frequentiebanden en, zoals gezegd, gepresenteerd aan normaalhorende proefpersonen. De resultaten laten duidelijk zien dat voor simulaties met meer frequentiebanden, dat wil zeggen met hogere frequentieresolutie, de verstaanbaarheid van de spraak omhoog gaat, totdat vanaf 6 frequentiebanden een plateau, of eigenlijk plafond, wordt bereikt. De luisterinspanning, gemeten in gemiddelde reactietijd op de rijmtaak, verbetert ook met toenemende frequentieresolutie, met als belangrijkste vinding dat luisterinspanning nog verbetert tot 8 frequentiebanden. Daarmee toont dit onderzoek aan, dat zelfs als spraakverstaan op 100% zit, een luisteraar toch nog baat kan hebben bij verder verbeterde frequentieresolutie, dit kan namelijk de luisterinspanning verbeteren. Hiermee wordt ook geillustreerd waarom het belangrijk is om naast spraakverstaan ook luisterinspanning te meten: daarmee wordt deze 'verborgen' ruimte voor verbetering zichtbaar gemaakt.

Voor normaalhorende luisteraars is duidelijk dat als het signaal een hogere frequentieresolutie heeft, zij dit ook daadwerkelijk kunnen benutten. Voor CI gebruikers is dat niet altijd vanzelfsprekend, aangezien niet alleen het signaal geproduceerd door de CI een rol speelt, maar ook de overdracht van dat signaal naar de zenuw en de gezondheid van de zenuw. Het kan dus zijn dat het CI signaal weldegelijk een hogere frequentieresolutie heeft, maar de CI gebruiker dit niet effectief kan benutten. Om te achterhalen of ook voor CI gebruikers een verschil in frequentieresolutie van het CI signaal leidt tot verschillen in luisterinspanning is het bovenstaande onderzoek gerepliceerd met CI gebruikers. In dit onderzoek, beschreven in hoofdstuk 5, wordt in de verschillende experimentele condities het geluid aangeboden via een beperkt aantal elektrodes. De dual-task resultaten van dit onderzoek tonen geen verschil in spraakverstaan en ook niet in luisterinspanning met toenemend aantal elektrodes. Echter, een alternatieve taak, de 'sentence verification task' waarin de luisteraar met een druk op een knop moet aangeven of een zin waar is of onzin, laat zien dat de verschillende aantallen actieve elektrodes weldegelijk het vermogen over de gehoorde spraak te redeneren beïnvloeden. Voor het feit dat het dual-task experiment geen verschil toont tussen de verschillende experimentele condities kunnen een aantal verklaringen opgevoerd worden. Echter het feit dat de alternatieve taak, de sentence verification task, wel een duidelijk effect laat zien van frequentieresolutie sluit de mogelijkheid dat frequentieresolutie gewoonweg geen effect heeft uit. Hieruit kunnen we concluderen dat zowel normaal-horende proefpersonen als CI gebruikers baat kunnen hebben bij hogere spectrale resolutie, zelfs als spraakverstaan al optimaal lijkt.

BIMODAAL HOREN EN LUISTERINSPANNING

Onder CI gebruikers zijn er mensen die nog bruikbaar restgehoor hebben in een of beide oren, meestal in de lage frequenties. Onderzoek heeft aangetoond dat dit acoustische gehoor gecombineerd met het elektrische gehoor van een CI, ook wel genaamd bimodaal gehoor, spraakverstaanbaarheid kan verbeteren met name in achtergrondlawaai. Over het effect van restgehoor op luisterinspanning was nog weinig bekend. Het onderzoek beschreven in hoofdstuk 3 bestudeert de toegevoegde waarde van laagfrequent restgehoor zowel in situaties met als zonder achtergrondgeluid, wederom in een simulatie experiment met normaal horende proefpersonen. In dit onderzoek werden de verschillende CI-alleen en bimodale condities vergeleken op verschillende niveaus van spraakverstaanbaarheid. De spraakverstaanbaarheid werd gemanipuleerd door de spraak in dusdanig luide achtergrondruis aan te bieden dat de beoogde verstaanbaarheid werd bereikt. De resultaten laten zien dat de aanwezigheid van (gesimuleerd) restgehoor inderdaad leidt tot verminderde luisterinspanning voor spraak in stilte op 100% verstaanbaarheid, evenals voor spraak in achtergrondruis op 50% verstaanbaarheid. Er lijkt geen meetbaar verschil in toegevoegde waarde voor restgehoor in de lage frequenties tot 600Hz vergeleken met tot 300Hz. Deze vinding suggereert dat zelfs beperkt restgehoor, wat op zichzelf (zonder CI) niet tot significante spraakverstaanbaarheid leidt, in combinatie met een CI al meerwaarde op kan leveren in de vorm van gereduceerde luisterinspanning.

CONCLUSIES

De onderzoeken gepresenteerd in dit proefschrift laten zien dat een verbetering in de kwaliteit van het spraaksignaal, door bijvoorbeeld hogere spectrale resolutie of toegevoegde laagfrequent akoestisch geluid, luisterinspanning kunnen verlagen. Deze verbeteringen in luisterinspanning kunnen zich zelfs nog voordoen als spraakverstaanbaarheid onveranderd lijkt, ofwel omdat alles wordt verstaan, of omdat het (kunstmatig) door ruis op een bepaald percentage wordt gehouden, luisterinspanning toch kan verschillen. Dit illustreert het belang van het meten van luisterinspanning naast de gebruikelijke maten van spraak verstaan. Een goede maat voor luisterinspanning kan nuttig zijn voor gebruik in onderzoek, om zo beter inzicht te krijgen in welke factoren invloed hebben op spraakverstaan en luisterinspanning en welke strategieën werken om luisterinspanning te verbeteren, en voor gebruik in de kliniek, om de CI te kunnen afregelen voor optimale luisterinspanning. Tot slot laten vergelijkingen tussen de dual-task en eenvoudigere taken zoals de verbale reactie tijden en de 'sentence verification task' zien dat een goede maat voor luisterinspanning niet ingewikkeld hoeft te zijn. De 'sentence verification task' is mogelijk een goede kandidaat voor klinische toepassing aangezien de taak makkelijk te implementeren is en eenvoudig uit te leggen aan een grote verscheidenheid aan patiënten.

DANKWOORD

Een promotie project tot een goed einde brengen en een proefschrift afleveren is niet niks. Het zou mij dan ook niet zijn gelukt zonder de steun, aanmoediging, en bij tijden welkome afleiding verzorgd door vrienden, familie en collega's. Een aantal van deze mensen zou ik hier graag expliciet willen bedanken. Allereerst mijn ouders, Lia en Nico, en mijn zusje Roos en hun partners Dirk, Aaf en Mendel bedankt voor alle steun en de fijne etentjes, spelletjes avonden, en saunabezoeken. Deze momenten hebben mij steeds weer nieuwe energie gegeven om door te gaan. De vernieuwde motivatie die mij heeft geholpen het proefschirft eindelijk af te ronden, om daarna straks een nieuw avontuur aan te gaan in Zuid Korea, is te danken aan mijn vriend YoungHwan. Bedankt!

Verder wil ik de mensen die het promotieproject mogelijk hebben gemaakt bedanken, te beginnen met mijn promotor. Bedankt Deniz, voor alles wat ik van je heb geleerd, voor je vertrouwen in mijn kunnen, en voor de altijd gezellige (en lekkere!) etentjes. Dankjewel Tassos, voor de stimulerende gesprekken en je altijd nuttige feedback. Thomas Stainsby, Filiep Vanpoucke en andere medewerkers van Cochlear LTD, bedankt voor de prettige samenwerking, ondersteuning met materiaal en cursus, de feedback op de artikelen, en de financiering van het project. Pim van Dijk, bedankt, voor de steun in het eerste jaar van mijn promotieproject en voor de gezellige etentjes met de collegas van KNO. Andy Beynon en Hedderik van Rijn, bedankt voor de prettige samenwerking, de goede ideeen en feedback.

Bert, Emile, Sander, Marnix, Gijs, en Ruben, bedankt voor alle antwoorden op mijn eindeloze vragen op het audiologische vlak. Gerda, bedankt voor de hulp met het contacteren van CI gebruikers. Frits, Paolo en Leon hartelijk bedankt voor de technische ondersteuning. Mijn dank, ook voor alle organisatorische en administratieve hulp van de mensen bij BCN, in het bijzonder Diana, en bij KNO, met name Ria, Sandra, Gerlinde en Nadine. Ik wil ook graag al mijn studenten en student assistenten van harte bedanken voor alle hulp met experimenten runnen, audio-opnames transcriberen en meer. Mart, Marica, Marije, Annemieke, Esmée, Floor, Maraike, Nico, en Matthias hartelijk bedankt voor alle hulp!

En dan wil ik graag mijn collega's bedanken, het was niet alleen prettig samenwerken, jullie maakten de afdeling tot een fijne, gezellige plek waar ik altijd wel bij iemand terecht kon voor een praatje, een kop thee, of even inhoudelijk brainstormen. Lisette, hartelijk dank voor de

heerlijk nuchtere kijk op alles en je fijne vertrouwde droge gevoel voor humor. Amarins, ik heb met veel plezier het kantoor met jou gedeeld, bedankt voor alles! Pranesh, ik geniet altijd weer van de gesprekken met jou en je slimme vragen over de Nederlandse taal, bedankt! Karin, Jeanne, Annika, Jefta, Enja, Nawal, Anita, Sonja, Etienne, Dave, Kris, en alle andere collega's en ex-collega's van de onderzoeksgroep bij KNO, bedankt! Ook mijn collega's bij Psychologie, met name Judith, Maraike en Evelien, bedankt voor de gezelligheid en steun tijdens het laatste staartje proefschrift schrijven!

En niet te vergeten, zonder mijn uitlaatklep tangodansen was het proefschrift waarschijnlijk niet tot stand gekomen, ik wil dan ook al mijn tango-vrienden en vriendinnen van harte bedanken voor de fijne dansjes en gesprekken. Méus, Bettie, Rob, Inez, Teodora, Hester, Anouchka, Anne Myrrthe, grote- en kleine Frank, Erwin, Rogier, en vele anderen, bedankt!

En als laatste, maar zeker niet het minste, wil ik een paar speciale vrienden bedanken. Nicole, bedankt voor de heerlijke etentjes en saunabezoeken samen! Hester, bedankt, het is altijd gezellig samen theedrinken of wandelen in het noorderplantsoen. Tanja, bedankt voor de goede gesprekken en het fijn samen eten aan de Marwixstraat. Emma, bedankt voor de heerlijke pizzas en de gezellige avonden Koreaanse tv-series kijken. Lisette Mol, al zien we elkaar niet vaak, ik geniet altijd van onze heerlijk open gesprekken samen, bedankt! Bart, bedankt voor de random chats en youtube linkjes tot laat in de nacht. En mijn andere vrienden, in het speciaal Urias, Asbjørn, Eelco, en Sebastiaan, bedankt voor de fijne discussies en gezellige tijden.

LIST OF ABBREVIATIONS

CI	Cochlear implant
EAS	Electric-acoustic stimulation (i.e. CI + residual (acoustic) hearing)
ELU	Ease of Language Understanding
HI	Hearing impaired
LE	Listening effort
LPF	Low-pass filtered
NASA TLX	NASA Task Load Index
NC	Near ceiling
NH	Normal hearing
RAU	Rationalized arcsine unit
RST	Reading span test
RT	Response time
SNR	Signal-to-noise ratio
SRT	Speech reception threshold
SSN	Steady state noise (in this case: steady state, speech shaped noise)
WAIS	Wechsler Adult Intelligence Scale