University of Groningen

Listening Effort

Pals, Carina

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

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

Publication date:
2016

Link to publication in University of Groningen/UMCG research database

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

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

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

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

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

Download date: 12-09-2024

CHAPTER 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).

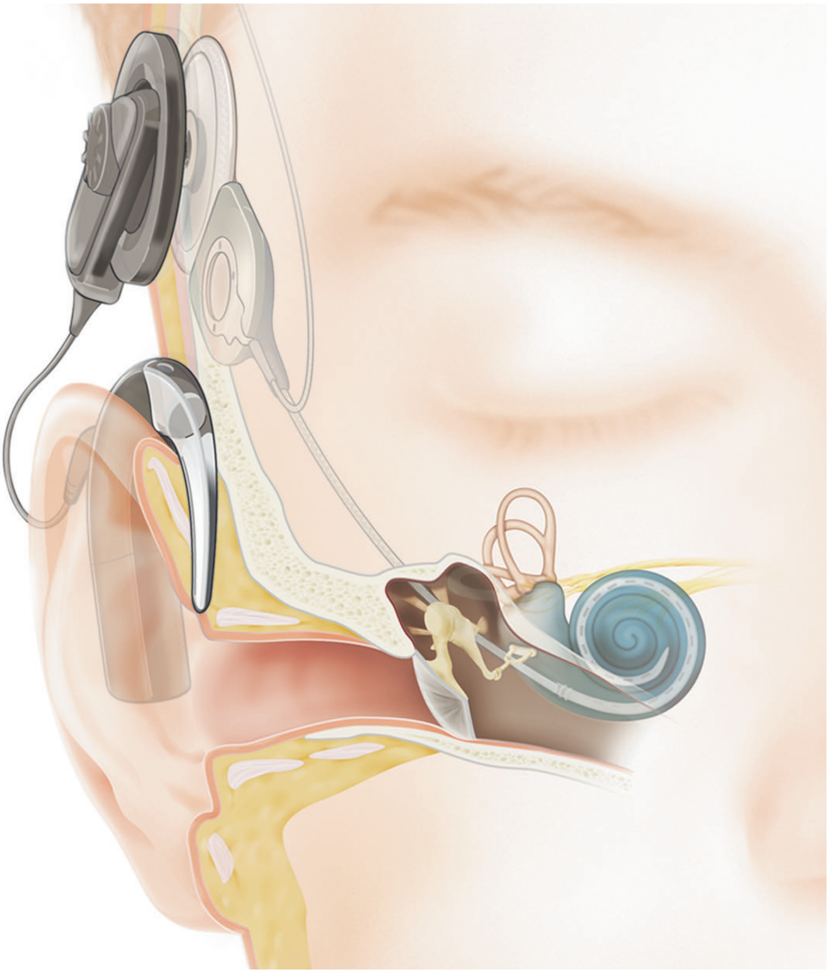

Figure 1: Illustration of a right ear with a cochlear implant. Hooked behind the ear (or pinna) is the speech processor, which connects to the transmitter that sits on the skull (dark gray). The transmitter is held in place by a magnet that connects it to the receiver embedded in the skull (translucent),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

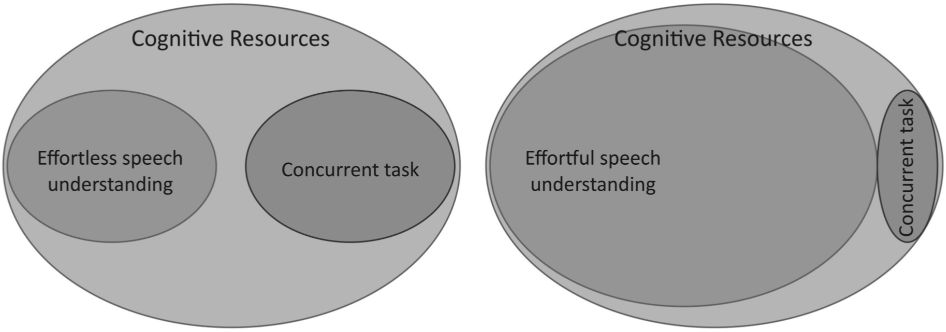

Figure 2: Cognitive resources and the interaction between the task demand of speech understanding and resources available for a concurrent task. So long as the cognitive resources required for the task of speech understanding do not exceed the available resources, full intelligibility can be achieved (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.