Time is of the essence in speech perception!

Get it fast, or think about it

Shahram Moradi
Linköping Studies in Arts and Science • No. 635
Studies from the Swedish Institute for Disability Research No. 68

At the Faculty of Arts and Science at Linköping University, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in Arts and Science. This thesis comes from the Swedish Institute for Disability Research at the Department of Behavioral Sciences and Learning.

Distributed by:
Department of Behavioral Sciences and Learning
Linköping University
SE-581 83 Linköping
Sweden

Shahram Moradi

Time is of the essence in speech perception! Get it fast, or work hard

Edition 1:1
ISBN 978-91-7519-188-1
ISSN 0282-9800
ISSN 1650-1128

©Shahram Moradi
Department of Behavioral Sciences and Learning, 2014

Printed by: LiU-Tryck, Linköping 2014
Abstract

Speech recognition occurs when attending to speech stimuli in auditory, visual, or audiovisual modalities under optimum (e.g., in silence) or degraded listening conditions (i.e., in background noise or in individuals with hearing impairment). The present thesis contains details of the first study to show how background noise (steady-state white noise) delayed the identification of different types of speech stimuli (consonants, words, and final words in sentence) in auditory and audiovisual modalities by calculating the isolation points (IPs, the shortest time from the onset of an speech stimulus required for correct identification of that speech stimulus). The cognitive demands of perceiving speech stimuli under each type of listening condition (silence vs. noise) were also tested by measuring the correlations between IPs of each type of speech stimulus (in each type of listening condition) and tests of working memory capacity (i.e., the reading span test, Rönnberg, Arlinger, Lyxell, & Kinnefors, 1989) and attentional capacity (i.e., the Paced Auditory Serial Attention Test [PASAT], Gronwall, 1977). In addition, the speechreading of words in both silent and noisy (babble background noise and steady-state white noise) conditions was also investigated in above-average speechreaders. The present thesis also compared elderly hearing-aid (EHA) users and elderly normal-hearing (ENH) individuals, in terms of IPs and accuracy in the identification of different types of speech stimuli in silence. Furthermore, the extent to which the identification of auditory speech stimuli was cognitively demanding (for both EHA users and ENH individuals) was investigated by measuring the correlations between IPs and accuracies of different types of speech stimuli and working memory capacity.

The results showed that background noise resulted in later IPs (in the auditory and audiovisual modalities) and reduced accuracy (in the auditory, visual, and audiovisual modalities). In visual speech recognition (speechreading), the adverse effect of noise on accuracy was dependent on the type of background noise: babble background noise reduced accuracy but the steady-state white noise did not. In addition, the audiovisual presentation of speech stimuli, relative to auditory presentation, resulted in earlier IPs and more accurate identification of speech stimuli in both silence and noise. In the silent condition, the IPs for perceiving different types of auditory and audiovisual speech stimuli were not correlated with performance on the tests of working memory and attention. In the noise condition, the IPs of consonants and
words in the auditory-only condition were significantly correlated with tests for measuring explicit cognitive resources. Furthermore, there was a significant correlation between the results on a test of speech-in-noise performance (the HINT, Häggren, Larsby, & Arlinger, 2006) and the results on tests of working memory (the reading span test, Rönnberg et al., 1989) and attention (PASAT, Gronwall, 1977). A secondary finding was that the subjects who had been exposed to audiovisual speech stimuli subsequently had better performance in a speech-in-noise test than those who had been exposed to auditory speech stimuli only. Furthermore, EHA users showed later IPs (for consonants, words, and final words in LP sentences) and reduced accuracy (for consonants and words) than ENH individuals. In EHA users, the IPs and accuracy scores for identification of consonants and words were significantly correlated with the participants’ working memory scores. These correlations were not observed in the ENH individuals.

The novel findings of the present thesis were as follows: 1) Noise delayed the identification of speech stimuli and made it cognitively demanding in the absence of visual cues or prior semantic context. 2) Even though the EHA participants used hearing aids, identification of auditory speech stimuli in the absence of a supportive context was delayed, less accurate, and more cognitively demanding for the EHA users than for the ENH individuals. 3) Prior exposure to audiovisual speech stimuli improved subsequent speech-in-noise identification, whereas prior exposure to auditory speech stimuli did not. Together, these findings support and extend models of speech perception that incorporate auditory and cognitive resources of listeners in perceiving degraded speech stimuli (ELU model, Rönnberg et al., 2013; cognitive spare capacity hypothesis, Mishra, 2014). In addition, the findings of the present thesis are in agreement with interactive models of speech perception assuming that top-down feedback (e.g., lexical or contextual), in addition to the feed-forward incoming signal, influences the recognition of speech stimuli (e.g., TRACE model, McClelland & Elman, 1986). Finally, the present thesis suggests using audiovisual training, instead of auditory training, to improve speech-in-noise identification.
List of Papers

The current thesis is based on the following papers:


## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHA</td>
<td>elderly hearing aid [users]</td>
</tr>
<tr>
<td>ENH</td>
<td>elderly normal-hearing [individuals]</td>
</tr>
<tr>
<td>ELU</td>
<td>Ease of Language Understanding [model]</td>
</tr>
<tr>
<td>HINT</td>
<td>Hearing In Noise Test</td>
</tr>
<tr>
<td>HP</td>
<td>High-predictability [sentences]</td>
</tr>
<tr>
<td>IP</td>
<td>Isolation point</td>
</tr>
<tr>
<td>LP</td>
<td>Low-predictability [sentences]</td>
</tr>
<tr>
<td>PASAT</td>
<td>Paced Auditory Serial Addition Test</td>
</tr>
<tr>
<td>RAMBPHO</td>
<td>Rapid Automatic Multimodal Binding of PHOnology</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SPL</td>
<td>Sound presentation level</td>
</tr>
</tbody>
</table>
Table of contents

An interdisciplinary approach to disability................................................. 1

Introduction ............................................................................................. 2
Impact of background noise on speech recognition .................................... 3
Impact of hearing loss on speech recognition ............................................. 4
Perceiving different types of speech stimuli under adverse listening conditions ........ 5
  Consonant identification ............................................................................... 6
  Word identification ..................................................................................... 7
  Sentence-based word identification ............................................................ 9
The role of explicit cognitive functioning in the recognition of speech stimuli ........ 11

General aims ............................................................................................ 17
General Methods ....................................................................................... 18
  Procedure .................................................................................................... 18
  Tests ........................................................................................................... 20
  Participants ................................................................................................. 22

Summary of the Papers ............................................................................ 24
Paper 1 ......................................................................................................... 24
Paper 2 ......................................................................................................... 24
Paper 3 ......................................................................................................... 25
Paper 4 ......................................................................................................... 26
Paper 5 ......................................................................................................... 27

General Discussion .................................................................................. 29
Main findings and conclusions ................................................................. 29
Adverse effects of noise on speech recognition depend on the modality of presentation of speech stimuli and the type of speech stimuli ........................................ 30
The use of hearing aids does not fully compensate for the effects of hearing loss in terms of auditory speech recognition abilities: comparisons of EHA users with elderly normal-hearing, age-matched, participants (ENH) .............................................. 32
The cognitive demands associated with perceiving speech stimuli greatly depend on the listening conditions, type of speech stimuli, modality of presentation, and hearing loss ........................................................................................................ 33
Audiovisual presentation (relative to auditory presentation) expedites IPs and improves accuracy for the identification of speech stimuli ........................................ 37
Prior exposure to audiovisual (rather than auditory-only) speech stimuli improves speech-in-noise performance .............................................................. 38

Future Directions for Interdisciplinary Research and Applications .............. 39
An interdisciplinary approach for rehabilitation of hearing-impaired individuals ... 39
Audiovisual training as a rehabilitation method for improving speech recognition abilities in individuals with hearing difficulties (e.g., hearing aid users and children with cochlear implants) ................................................................. 39
A comparison of audiovisual gated speech recognition abilities in EHA users versus ENH individuals ......................................................................................... 40

Conclusions ............................................................................................ 41
Acknowledgements ................................................................................... 42
References ............................................................................................... 43
An interdisciplinary approach to disability

Current disability research addresses a variety of social, medical, psychological, political, and economic aspects of disability. In the past, disability research, which focuses on the etiology and rehabilitation of individuals with disability, mainly conducted one the basis of one single theoretical discipline (i.e., the medical; Bickenbach, Chatterji, Badley, & Üstün, 1999; Goodley, 2011). However, such restricted perspectives are not successful in providing a holistic view of a disability. Typically, disability represents a complex problem that requires input from a combination of different disciplines, such as the medical, psychological, and social disciplines (cf. Danermark, 2002). The ICF (International Classification of Functioning, Disability, and Health; World Health Organization, 2001) contextualizes disability and functioning as the interaction between an individual’s health conditions, environmental factors, and personal factors within a bio-psycho-social approach.

The general framework of the present thesis is based on cognitive hearing science (Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009), an interdisciplinary research area that combines different disciplines (i.e., psychology, audiology, linguistics, physiology, etc.) to understand the interaction between hearing and cognition. In particular, cognitive hearing science studies this interaction under degraded listening conditions (e.g., in noise or with hearing-impaired participants) with the ultimate goal of devising or improving rehabilitative interventions for given populations (e.g., hearing-aid users and children with cochlear implants). Within this general framework, the present thesis investigates the effect of background noise on the speed and accuracy of speech recognition, and also assesses the cognitive demands of perceiving speech stimuli presented in different modalities for young normal-hearing listeners.

With regard to disability research, the present thesis also uses an interdisciplinary analysis of the horizontal and vertical dimensions of disability (Rönnberg, 2005). In the vertical dimension, medical, psychological, and social perspectives are combined and the horizontal dimension is used to refer to the characterization a given phenomenon across different population. In this thesis, the EHA users and ENH individuals (two separate groups that vary in terms of hearing impairment) are used as examples of the horizontal dimension. Vertically, the linguistic level (linguistically varied types of speech stimuli) and psychological level
(cognitive capacity) are studied as an interactive whole. The present thesis combines the vertical and horizontal dimensions of knowledge generation and integration from the disciplines audiology, psychology, and linguistics. This enables comparison of the speech recognition abilities of EHA users and ENH individuals, and gives further insight into the generality/precision of particular cognitive demands on the recognition of different kinds of speech stimuli for EHA users and ENH individuals.

Introduction

We perceive speech signals in different modalities: either as an auditory or a visual signal, or as a combination of both auditory and visual inputs. Whereas auditory speech identification refers to the identification of meaningful sounds, achieved via the aural detection of vibrations in the surrounding medium over time, speechreading (i.e., visual speech recognition or lipreading) refers to the visually perceived articulation of sounds. Speechreading is hardly ever used by normally-hearing people as the sole modality for speech recognition. This is because speechreading is rarely perfect, due to the fact that some components of articulated speech signals (e.g., consonants or vowels) do not show sufficiently distinguishable visible features. In face-to-face conversation, people benefit from combined auditory and visual inputs (an audiovisual speech signal) for perceiving speech stimuli. The identification of audiovisual speech stimuli is more accurate than the identification of auditory-only speech stimuli, even when the auditory signal is clearly presented to listeners (Arnold and Hill, 2001; Reisberg, McLean, & Goldfield, 1987). This implies a superiority of audio-visual speech recognition relative to auditory or visual presentation alone.

In everyday life, the identification of speech signals rarely occurs under optimum listening conditions. Such identification often takes place in conditions that adversely affect one’s ability to correctly identify speech stimuli (for instance, background noise or hearing impairment can impair the recognition of speech signals). In such circumstances, listeners require more (cognitive) effort to make sense of an impoverished target signal than would be required in optimum listening conditions (Hervais–Adelman, Carlyon, Johnsrude, & Davis, 2012; Mishra, 2014).

1 Specific social aspects (e.g., job performance, social integration with others) are not explicitly studied in the present thesis. However, human social contact occurs mostly via speech (i.e., words and sentences). Thus, a social level may also be implied in the vertical dimension.
Rönnberg, Rudner, Foo, & Lunner, 2008; Rönnberg et al., 2013). The intelligibility of speech in degraded listening conditions (noise or hearing impairment) depends greatly on the modality of speech presentation. For instance, audiovisual presentation of speech stimuli has been reported to lower the SNRs in speech recognition relative to auditory-only presentation (e.g., Bernstein, Auer, & Takayanagi, 2004; Tye-Murray, Spehar, Myerson, Sommers, & Hale, 2011) and to reduce the cognitive effort required to detect speech stimuli in noisy conditions (Mishra, Lunner, Stenfelt, Rönnberg, & Rudner, 2013).

**Impact of background noise on speech recognition**

In realistic listening situations, various sounds, sometimes embedded in a scene of background sounds, will reach our ears. In such cases, listeners have to focus on a particular target sound or voice while ignoring the sound(s) of no interest. The distracting or masking sounds can be classified in different ways, but one kind of classification that has gained widespread acceptance is the one that distinguishes energetic from informational maskers (Brungart & Simpson, 2002; Pollack, 1975). Energetic masking refers to a listening condition in which the irrelevant acoustic signal overlaps in time and frequency with the target speech signal such that the listener is more or less unable to detect the masked components of the target speech signal (Brungart & Simpson, 2002). Informational masking, on the other hand, refers to a listening condition in which the competing and target speech signals are presented simultaneously to a listener such that the listener experiences difficulty in separating the target speech from similar-sounding distracter speech (Freyman, Helfer, McCall, & Clifton, 1999; Watson, Kelly, & Wroton, 1976). While the effectiveness of energetic masking greatly depends on its interaction with the acoustic characteristics of the target signal in terms of SNR, the effectiveness of informational masking greatly depends on the number of distracters (e.g., two, four, or eight talkers in the background, Simpson & Cooke, 2005), the similarity of the distracters with the target speech signal (e.g., the sound of distracter is from a male but the sound of the target speech signal is from a female, Brungart, 2001), the uncertainty of the target (e.g., the duration of target signal differs from the masker or maskers, Durlach et al., 2003), and the semantic similarity between target speech and distractor or distracters (Ng, Rudner, Lunner, & Rönnberg, 2014).

The adverse effects of energetic and informational masking also greatly depend on the modality of the speech signals; both types of masking have been
reported to have an adverse effect on the identification of auditory and audiovisual speech signals (Helfer & Freyman, 2005), but only informational masking has been shown to have an adverse effect on speechreading performance (Brungart & Simpson, 2005) or to activate explicit cognitive resources during speechreading (Lyxell & Rönnberg, 1993).

**Impact of hearing loss on speech recognition**

Hearing loss is one of the most common impairments in elderly people (Pleis & Lethbridge-Cejku, 2007). In Sweden, the prevalence of hearing loss is 42% among elderly people aged 60–70 years and 74% among those aged 70–80 years (Johansson & Arlinger, 2003). The main consequence of hearing loss is an obvious difficulty in clearly perceiving auditory speech signals. Research has also shown that hearing loss is associated with cognitive decline (Lin et al., 2013; Rönnberg et al., 2011; Wong, Yu, Chan, & Tong, 2014), structural changes in the brain (Boyen, Langers, de Kleine, & van Dijk, 2013; Eckert, Cute, Vaden, Kuchinsky, & Dubno, 2012; Lin et al., 2014; Peelle, Troiani, Grossman, & Wingfield, 2011), mental health problems such as depression (e.g., Li et al., 2014) and social isolation (e.g., Weinstein & Ventry, 1982).

Currently, the most commonly used method to improve speech recognition ability for hearing-impaired individuals is the amplification and processing of sounds using hearing aids. A hearing aid is a small electronic device that consists of a microphone, an amplifier, and a loudspeaker that transfers the amplified sounds into the ear. Digital hearing aids offer more features and flexibility to meet the varied needs of hearing-impaired individuals by using non-linear signal processing algorithms (e.g., wide dynamic range compression), directional microphones, and noise reduction algorithms. Studies have shown that aided speech recognition is better than unaided speech recognition (e.g., Dimitrijevic, John, & Picton, 2004; Walden, Grant, & Cord, 2001). Importantly, however, hearing-aid users are still reported to display inferior performance in auditory speech recognition compared to their counterparts with normal hearing (cf. Dimitrijevic et al., 2004). The adverse effects of age-related hearing loss on auditory speech recognition also greatly depend on the type of speech stimuli (Lash, Rogers, Zoller, & Wingfield, 2013; Whitmer & Akeroyd, 2011). When speech stimuli are presented without a supportive semantic context, hearing-impaired individuals show inferior performance compared to their counterparts with normal hearing on the identification of auditory speech stimuli.
(Lash et al., 2013). In speechreading ability, Lyxell and Rönnberg (1989) reported no differences between hearing-impaired and normal-hearing listeners in lip-reading ability. The audiovisual presentation of speech stimuli, even in an unaided condition, resulted in better accuracy than an auditory-aided condition in hearing-impaired individuals (Walden et al., 2001). However, Bernstein and Grant (2009) reported that hearing-impaired individuals needed higher SNRs in audiovisual identification of keywords in sentences than their counterparts with normal hearing.

In sum, individuals with hearing loss show inferior performance relative to normal-hearing individuals in aided auditory identification of speech stimuli in the absence of a supportive semantic context. The addition of visual cues to the auditory speech signals results in better identification than auditory-alone condition for normal as well as hearing-impaired participants.

**Perceiving different types of speech stimuli under adverse listening conditions**

Speech consists of various acoustic, lexical, and sentential components. According to interactive models of speech perception (McClelland & Elman, 1986; McClelland, Mirman, & Holt, 2006), top-down lexical or semantic knowledge modulates phonological processing of the speech signal, early on in the processing of the signal. That is, speech perception occurs in a dynamic and interactive manner, such that the incoming signal interacts with feedback from phonological or lexical representations from the semantic long-term memory (Rönnberg et al., 2013). While the identification of phonemes out of word context (such as in a /vCv/ or /cVc/ format) is solely based on the physical (i.e., acoustical or visual) speech cues, the identification of single words is based on an interaction between physical cues and top-down lexical knowledge. Listeners can benefit from top-down lexical and semantic knowledge, in addition to the physical characteristics of speech cues, when comprehending the meaning of a given sentence or guessing the identity of target words embedded in a congruent sentence.

Contrary to interactive models of speech perception, modular (or autonomous) models of speech perception argue that top-down phonological or lexical feedback is irrelevant or redundant in the process of speech recognition (Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2000). That is, according to modular models of speech perception, top-down knowledge does not modulate the early acoustic processing of speech stimuli but top-down lexical knowledge and acoustic
information are united at a late judgment point when a phonological (or lexical) decision has been formed.

When listening occurs in degraded conditions, the results from recent studies are in favor of interactive models of speech perception, such as top-down semantic knowledge (Sohoglu, Peelle, Carlyon, & Davis, 2014), or cognitive capacity of listeners (Wild et al., 2012; Zekveld, Rudner, Johnsrude, & Rönnberg, 2013) influencing the intelligibility of degraded speech signal (for reviews, see Rönnberg et al., 2013; Zion Golumbic, Poeppel, & Schroeder, 2012), even at the level of the auditory brain stem response (Sörqvist, Stenfelt & Rönnberg, 2012).

Consonant identification

Acoustic speech signals contain cues, such as voicing, manner, and place of articulation, which, in specific combinations, constitute consonants. These critical features of consonants are spread across the duration of the articulation of the consonant (Smits, 2000), and listeners identify consonants when these particular cues are available (Sawusch, 1977). Under noisy conditions, the acoustical properties of consonants are combined with the acoustical characteristics of noise, which make it difficult to correctly identify consonants in the auditory (Miller & Nicely, 1955; Phatak & Allen, 2007) and audiovisual (Grant & Walden, 1996) modalities.

Whereas auditory cues provide information about the manner of articulation, visual cues provide information about the place of articulation. Therefore, the combination of both auditory and visual cues in audiovisual presentation results in more accurate identification of consonants relative to auditory-only or visual-only presentation (e.g., Tye-Murray, Sommers, & Spehar, 2007a).

Although the amplification of sounds by hearing aids has been shown to improve consonant identification in comparison to unaided conditions (Walden et al., 2001; Stecker et al., 2006; Yund, Roup, Simon, & Bowman, 2006), the degree of hearing aid benefit varies greatly across hearing aid users (Davies-Venn, Souza, Brennan, & Stecker, 2009; Souza & Gallun, 2010) and across consonants (Stecker et al., 2006; Yund et al., 2006). For instance, Simpson, Hersbach, and McDermott (2006) compared consonant identification in quiet using two types of amplification: frequency compression amplification and linear compression amplification. They observed some consonant identification errors in both amplification types. Of all consonants, /t/ had the lowest correct identification level across both amplification
types. Moreover, in both amplification types, the number of errors for /b n dʒ ʤ ʤ/ was markedly increased compared to other consonants.

Ahlstrom, Horwitz, and Dubno (2014) suggested that variations in consonant identification among hearing aid users may partly result from individual differences in cognitive capacity or central auditory processing, or a combination of the two. Ahlstrom et al. provided sufficient amplification to improve high-frequency speech audibility for hearing-impaired individuals. Their results showed that the benefits provided by the addition of higher-frequency speech information were relatively limited. The authors proposed that the lack of significant hearing aid benefit might be due to factors beyond simple audibility, such as central auditory deficits or poorer cognitive ability. Harkrider, Plyler, and Hedrick (2009) reported that hearing loss, independent of aging, adversely affected underlying neural spectral cues and resulted in deficient consonant identification among hearing-impaired individuals. In a similar vein, they stated that compensation by improving audibility does not reduce the deleterious effects of hearing loss on consonant identification.

In sum, although hearing aids generally improve consonant identification for hearing-impaired individuals, the benefits provided by hearing aids vary extensively across hearing aid users, types of signal processing in the hearing instrument, and across consonants. This variability most likely arises from the hearing aids themselves (e.g., differences in types of signal processing), and from individual differences in either central auditory or cognitive systems in the brain, or a combination.

**Word identification**

Word identification occurs when an input acoustic signal matches with a phonological-lexical representation in long-term memory (Lively, Pisoni, & Goldinger, 1994; McClelland & Elman, 1986; Norris et al., 2000). Two types of temporal process have been proposed for word identification (for a review, see Weber & Scharenborg, 2012). The first process is termed “the incremental effect,” in which lexical word activation continuously increases through growing phonetic input support. The second mechanism is termed “the competing effect,” wherein words in the mental lexicon compete with each other over time until one of them finally matches the speech signal. The probability of activating a word depends on the lexical frequency, the degree of similarity between the lexical candidates in the mental lexicon and the bottom-up information provided by the phonetic signal, the phonological neighborhood density, and the semantic/contextual information (cf.
Luce & Pisoni, 1998). Words with low lexical frequency or higher phonological neighborhood density require longer acoustic duration for correct identification than words with higher lexical frequency or lower phonological neighborhood density (cf. Dufor & Frauenfelder, 2010).

Identification of words does not only occur in the auditory modality. Some words display sufficient visual cues for correct speechreading (e.g., pet, net). Studies have shown that variables such as word frequency, neighborhood density, and the similarity between lexical candidates (similar to auditory word recognition) also influenced speechreading of words (Auer, 2002; Auer & Reed, 2008). Fort et al. (2013) also demonstrated that viewing the initial articulatory gestures of a word facilitates lexical access (for low-frequency words but not for high-frequency words). The findings by Fort et al. (2013) showed the importance of visual cues in the identification of words when lexical access is difficult (i.e., under noisy conditions).

Audiovisual presentation of words increases the accuracy of identification of words in comparison with auditory-only presentation of words (Kaiser, Kirk, Lachs, & Pisoni, 2003). One explanation for the rapid and more accurate identification of words presented in the audiovisual modality relative to an auditory-alone or visual-alone condition can be explained by a concept introduced by Tye-Murray, Sommers, and Spehar (2007b): that fewer phonological neighborhoods are left in the overlap of aurally and visually presented words in comparison with words presented in an auditory-only or visual-only manner.

Although hearing aids improve word identification accuracy in comparison with unaided conditions (Dimitrijevic et al. 2004; Shanks, Wilson, Larson, & Williams, 2002), the word identification accuracy of hearing aid users is lower than that of their counterparts with normal hearing (Dimitrijevic et al., 2004). For instance, Dimitrijevic et al. reported that the accuracy ratio under aided conditions in silence was higher than that under unaided conditions (76% vs. 56%); however, the percentage of words accurately identified for hearing aid users was lower than that for individuals with normal hearing (76% vs. 97%, respectively). This reduced word identification ability among hearing-impaired individuals was explained primarily by the effects of reduced audibility due to hearing loss (Jerger & Chimel, 1997). The effect of hearing loss can also cause a long time consequences (Classon, Löökvist, Rudner, & Rönnberg, 2013; Rönnberg et al., 2011) in that the phonological representations deteriorate, hence reducing the efficiency of the matching process.
between the phonological aspects of the signal and its corresponding phonological representation in long-term memory. Interactive models of word identification (e.g., the “TRACE” model, McClelland & Elman, 1986) claim that lexical knowledge feeds back at a prelexical level to inhibit the activation of one or more irrelevant words, hence facilitating the identification of words (top-down lexical guidance).

In sum, hearing aid users exhibit inferior word identification performance than their counterparts with normal hearing. This finding may be related to poorly received auditory signals or may represent the consequences of hearing loss on the central auditory system and lexical knowledge, which in turn may reduce word identification among hearing-aid users.

**Sentence-based word identification**

In everyday conversation, people usually use sentences to convey their intentions, information, and knowledge to others. In such cases, listeners can benefit from semantic context and syntactic structure within the sentences to understand the speaker’s message. Marslen-Wilson (1989) proposed three stages for the identification of words in sentences: the activation of a set of lexical candidates based on the semantic context, the selection of a target lexical candidate, and the integration of a selected word into the overall meaning of the sentence. The identification of target words in congruent sentences is faster than the identification of words presented alone (Grosjean, 1980; Salasoo & Pisoni, 1985). This is because the semantic context activates lexical candidates in the mental lexicon that are compatible with the meaning of a sentence, and inhibits the activation of those that are not. The number of activated lexical candidates depends on the degree of expectancy provided by the semantic context (for a review see Aydelott, Leech, & Crinion, 2010). A sentence with high-predictability (HP) may activate only one word (i.e., “a pigeon is a kind of bird”), while a sentence with low-predictability (LP) is likely to activate another set of words apart from “bird” (i.e., “she pointed at the xxxxx”) in the mental lexicon.

The importance of prior semantic context is evident when speech stimuli are presented in noisy conditions. Prior semantic context alleviates the deleterious effects of background noise on the identification of target words embedded in sentences in comparison with target words presented alone (Boothroyd & Nittrouer, 1988; Grant & Seitz, 2000). As the semantic predictability of sentences increases, the adverse effects

The effect of the addition of visual speech cues on sentence comprehension is more evident when sentences are presented under noisy conditions (Kim, Davis, & Groot, 2009). Fraser, Gagné, Alepins, and Dubois (2010) compared sentence comprehension in auditory and audiovisual modalities. Their results showed that at the same SNR, audiovisual sentence comprehension is more accurate and less effortful than the auditory one. In addition, Gordon and Allen (2009) reported that under informational masking (12-talker babble), audiovisual presentation improves the accuracy for the identification of final words in both LP and HP sentences relative to auditory presentation.

Note that the speech tasks testing identification of the final words in sentences under noise conditions are totally different from tests of sentence comprehension in noise (cf. the Hagerman test, Hagerman & Kinnefors, 1995; the HINT, Hällgren et al., 2006). In the latter, listeners must correctly repeat the whole of an aurally presented sentence. In the sentence comprehension in noise tests, SNR is the criterion; this is typically determined in an adaptive manner (on the basis of algorithms that determine the SNR required for a 50% and 80% level of performance), according to the number of correctly recalled key words in each sentence. In tasks assessing the identification of final words in sentences, the criterion is the IP or accuracy at a constant SNR.

Gordon-Salant and Fitzgibbons (1997) reported that a HP semantic context could alleviate adverse word identification effects in hearing-impaired individuals, such that there were no significant differences between the hearing-impaired and normal-hearing listeners for final word identification in HP sentences. However, the hearing-impaired group showed poorer final word identification performance for LP sentences. Lash et al. (2013) studied gated auditory word recognition in elderly people with impaired hearing, elderly people with normal hearing, and young adults with normal hearing. They revealed that older adults with poor hearing acuity required a greater amount of word-onset information (i.e., had later IPs) for word recognition in a neutral context compared with older adults with good hearing acuity. This difference gradually decreased as the semantic contextual probability of a word increased.

In sum, a supportive semantic context reduces the adverse effect of hearing loss on word identification in hearing-impaired people. The facilitating effect of semantic
context greatly depends on the degree of predictability of prior context; as the predictability level of a sentence increases, the difference between hearing-impaired and normal-hearing individuals’ performance on target word identification decreases.

**The role of explicit cognitive functioning in the recognition of speech stimuli**

Language understanding in real life presumably requires collaboration between the auditory and cognitive systems. In many instances, working memory is thought to act as an interface between input signals and phonological or lexical representations (Rönnberg et al., 2008, 2013). According to the Ease of Language Understanding (ELU) model (Rönnberg et al., 2008, 2013), the matching process between the input signal and corresponding phonological representation in semantic long-term memory is assumed to demand a minimal working memory load for normal-hearing individuals under optimal listening conditions. In this case, acoustic signals and phonological representations in long-term memory are connected rapidly and automatically. However, in degraded listening conditions, such as in noise conditions or for people with hearing loss, this rapid, automatic, and implicit processing fails to match the signal with the target in the lexicon. Explicit cognitive resources (i.e., working memory and attention) are supposedly required to infer meaning from the impoverished signal. This process of filling in missing pieces of information is assumed to be conscious, effortful, and cognitively demanding in the absence of a top-down semantic context (Rönnberg et al., 2013). According to the cognitive spare capacity notion (Mishra, 2014), perceiving speech stimuli under adverse listening conditions and in the absence of visual cues, places higher demands on explicit cognitive resources. However, when auditory speech stimuli are presented under silent conditions or presented audiovisually under noisy conditions, more residual cognitive resources would be available. Mishra et al. (2013) showed that the association of visual cues with auditory speech stimuli reduces the cognitive demands on speech perception under degraded listening. Independent studies support the view that audiovisual presentation reduces cognitive demands in the processing of speech stimuli (Fr tusova, Winneke, & Phillips, 2013; Pichora-Fuller, 1996). Gosselin and Gagné (2011) showed that aging is a key variable in the impact of audiovisual presentation on listening effort. They showed that older adults expended more listening effort than younger adults in audiovisual speech identification in noise.
A key variable in audiovisual speech identification is audiovisual integration, which refers to the process of combining information that has been extracted from the auditory and visual channels (Grant, 2002). Audiovisual integration has been reported to be predominantly implicit or automatic (cf. McGurk & MacDonald, 1976; Rudner & Rönnberg, 2008; Rönnberg et al., 2008; Soto-Faraco, Navarra, & Alsius, 2004). However, Navarra, Alsius, Soto-Faraco, and Spence (2010), in a review, challenged the implicit nature of audiovisual integration and suggested that attentional resources are recruited during the integration of audio and visual speech stimuli, indicating that audiovisual integration is explicit. According to the ELU model, the binding of aurally and visually presented speech stimuli occurs at the level of an episodic buffer of working memory (Baddeley, 2000), which in the Rönnberg et al. model (2008, 2013) model is assumed to be automatic and implicit.

For hearing-impaired people who cannot clearly perceive audible acoustic signals, the process of matching the degraded acoustic signals with the corresponding phonological representations in the mental lexicon presumably requires deliberate and explicit cognitive resources (e.g., “inference-making”, which in this context refers to perceptual completion of ambiguous sounds and which is supposedly an attentionally demanding process; Rönnberg et al., 2008, 2013). Campbell and Sharma (2013), in an electroencephalography (EEG) study, supported this notion by showing that hearing loss is associated with the reallocation of cortical resources, from temporal areas to frontal cognitive areas, which presumably aid the processing and reconstruction of auditory stimuli, even in passive listening conditions.

As noted, hearing aids do not fully restore speech recognition ability. The auditory signal perceived via hearing aids is probably not as clear or undistorted as that received by normal-hearing individuals, most likely causing a mismatch (or match failure) between the input signal and phonological representations in long-term memory. There is compelling evidence that working memory supports speech perception among EHA users (Foo, Rudner, Rönnberg, & Lunner, 2007; Humes, 2002; Humes, Kidd, & Lentz, 2013; Lunner, 2003; Rudner, Foo, Sundewall-Thorén, Lunner, Rönnberg, 2008, Rudner, Rönnberg, & Lunner, 2011; Rönnberg et al., 2013). Humes (2007) reviewed the importance of audibility and cognitive factors for aided speech recognition, and concluded that once audibility was restored by hearing aids, age and cognitive factors (e.g., intelligence quotient and digit span) accounted for approximately 50% of the variance in speech recognition in older adults. In addition,
several studies have noted the relationship between the cognitive capacity of hearing aid users and their level of benefit from specific advanced signal processing algorithms in hearing aids (Arehart, Souza, Baca, & Kates, 2013; Gatehouse, Naylor, & Elberling, 2006; Ng, Rudner, Lunner, Pedersen, & Rönnberg, 2013).

One important point is that in most of the previous studies (Foo et al., 2007; Rudner et al., 2008, 2011), hearing-impaired individuals wore hearing aids with new compression settings during the experiment that were different from the settings in their own hearing aids, thus introducing a mismatch between the non-habitual signal processing modes and the phonological representations in semantic long-term memory (e.g., Foo et al., 2007). As noted by Lunner, Rudner, Rönnberg (2009), coping with different hearing aid compression settings may force hearing-impaired individuals to call on their explicit cognitive resources to perceive speech stimuli.

There are contradicting findings about how hearing-impaired individuals integrate auditory and visual speech input into coherent audiovisual speech stimuli. On the one hand, hearing-impaired individuals have been reported to integrate auditory and visual speech cues appropriately so that their performance on audiovisual stimuli was better than auditory-alone or visual-alone speech identification (Bosman & Smoorenburg, 1997; Walden et al., 2001), similar to normally-hearing individuals. On the other hand, Musacchia, Arum, Nicol, Garstecki, and Kraus (2009) showed that audiovisual integration ability was reduced in elderly people with hearing loss. In addition, Grant, Walden, and Seitz (1998) reported that the ability to integrate the auditory and visual cues varied across hearing-impaired individuals for identification of consonants. However, to date, no study has investigated the extent to which audiovisual integration in hearing-impaired or hearing-aid users is cognitively effortful (explicit) or effortless (implicit or automatic).

In summary, degraded listening conditions, resulting from noise or hearing loss, adversely affect the correct identification of speech stimuli. The deleterious effect of noise or hearing loss on auditory speech recognition mean that the recognition of speech stimuli is less accurate and needs more cognitive effort than optimum listening conditions. The audiovisual presentation of speech stimuli increases accuracy and may reduce cognitive effort in noisy conditions (e.g., Fraser et al., 2010). The adverse impact of noise on speechreading greatly depends on the type of masker. Although hearing aids improve speech recognition in hearing-impaired individuals, hearing aid
users still show inferior performance in perceiving speech stimuli than their counterparts with normal hearing.

Prior to the present thesis, no previous study had explored the extent to which degraded listening conditions (noise or hearing loss) delay the identification of different types of speech stimuli. The gating paradigm (Grosjean, 1980) has been used in the present thesis in order to calculate the IP (the shortest time from the onset of a speech stimulus required for correct identification) in both silent and noisy conditions in the auditory and audiovisual modalities for young, normally-hearing participants (Papers 1 & 2), and in silent condition for EHA users and ENH individuals in the auditory modality condition. In fact, measures like response time or IPs are presumably more sensitive and effective than accuracy to elucidate the extent to which noise or hearing loss influence the identification of different types of speech stimuli in different modalities (e.g., auditory or audiovisual). Gatehouse and Gordon (1990) studied the benefit provided by hearing aids in terms of response time and accuracy in the identification of words and sentences. While there was no benefit of the aided condition over the unaided condition using accuracy, a faster response time was observed in the aided condition. Gatehouse and Gordon concluded that measures like response time are more effective and sensitive in showing benefit than conventional measures like accuracy. In addition, the IPs of different types of speech stimuli vary greatly across participants even under optimum listening condition in which the accuracy of those speech stimuli is at the ceiling. This variability in IPs is also useful in correlational analyses (such as the relationship between IPs in silence or noise and scores on cognitive tests).

Regarding speechreading, prior research has studied how background noise influences the speechreading of participants. Note that speechreading is a difficult speech task as the accuracy of non-cued speechreading is reported to be around 7% correctly rendered words in normal-hearing subjects in Swedish studies (Samuelsson & Rönnberg, 1993). Chance level (i.e., no presentation at all, just random guesses) is around 5%. Because of this floor effect in speechreading, there is still concern that acoustic background noise (energetic and informational masking conditions) may not impact upon visual speech recognition, or conversely, may enhance visual speech recognition (Gleiss & Kayser, 2014; Lugo, Doti, & Faubert, 2008). In addition, we also tested the assumption by Brungart and Simpson (2005) that maskers need to be simultaneous to adversely influence speechreading performance. In our study, we
used continuous streams of maskers, as the onset and offset of the maskers were not simultaneous to the onset and offset of the visual speech stimuli. Furthermore, Brungart and Simpson (2005) used the coordinate response measure (CRM) as a measure of speechreading. The CRM is call-sign-based task of speech perception with a closed set of response alternatives of colors and numbers (e.g., “ready (call sign) go to (color) (number) now”). All possible combinations are four colors (red, blue, green, and white) and eight numbers (1–8). However, the speechreading stimuli in our study were the most common Swedish words in daily use.

The present thesis aimed to investigate the extent to which background noise (Papers 1, 2, & 5) or hearing loss (Paper 3) affected the identification of different speech stimuli (i.e., consonants, words, and final words in LP and HP sentences in terms of IPs and accuracy (Paper 5 only used accuracy among above-average speechreaders), presented in different modalities (auditory, visual, and audiovisual). In addition, the IPs and accuracy of audiovisual speech stimuli were compared to the IPs and accuracy of auditory speech stimuli to investigate the extent to which the addition of visual cues to auditory speech stimuli influenced the identification of different types of speech stimuli in both silent and noisy conditions (at the same SNR in auditory and audiovisual modalities).

A cognitive hearing science point of view was adopted to investigate the contribution of explicit cognitive resources (i.e., working memory and attention) to the IPs (by young normal-hearing university students) of gated speech task stimuli presented in silent versus noisy conditions, in the auditory (Paper 1) and audiovisual (Paper 2) modalities. This cognitive hearing science point of view was also employed in Paper 3, in which we measured the extent to which the IPs and accuracy rates for identified gated speech stimuli, in the silent condition, were correlated with the working memory capacities of EHA users and ENH individuals. All together, the present thesis sought to investigate how the matching process between signals (under different modalities) and their phonological representations is disturbed by noise or hearing loss and the extent to which this disturbance results in a greater demand being placed on explicit cognitive resources in order to make sense of what is heard or seen.

A secondary finding in Paper 2 is that prior exposure to audiovisual gated speech stimuli improved the signal-to-noise ratio (SNR) in a speech-in-noise test (the HINT [Hearing In Noise Test]). Paper 4, using a randomized-control design, therefore aimed to replicate and to investigate the extent to which prior exposure to
audiovisual- or auditory-gated speech stimuli influenced speech-in-noise identification.
General aims

Summarizing the argument, the present thesis had two general aims. The primary aim was to explore the extent to which degraded listening conditions (caused by noise or hearing loss) influenced the speed of identification of different types of speech stimuli in terms of IPs. The secondary aim was to examine the extent to which the processing of speech stimuli demands explicit cognitive resources under different listening conditions (noise vs. silence), with different stimulus modalities (audiovisual vs. auditory), given different levels of contextual support, and in particular groups (EHA users vs. ENH individuals). Thus, the present thesis sought to explore and specify the extent to which the process of perceiving speech is fast and automatic, and to identify when it demands more explicit cognitive processing.

To accomplish these aims, a gating paradigm (Grosjean, 1980), in shorter gate sizes than prior studies, was employed to measure the IPs and accuracy for the identification of different types of speech stimuli in auditory (Papers 1 & 3) and audiovisual modalities (Paper 2), in young normal-hearing listeners (Papers 1 & 2), and elderly hearing aid (EHA) users and elderly normal-hearing (ENH) individuals (Paper 3). Paper 5 investigated how two types of background noise (steady-state white noise vs. babble background noise) influenced the accuracy of speechreading of words when compared to speechreading of words under a quiet condition. As noted earlier, the steady-state white noise was used in addition to the babble background noise to test the hypothesis that informational masking primarily distracts via phonological competition and not via a more general effect on attention. Paper 4 used both auditory and audiovisual gating tasks to examine the extent to which prior exposure to audiovisual or auditory speech tasks subsequently influenced speech-in-noise identification.
General Methods

Procedure

Paper 1

Please see Paper 1 for information about the preparation of the gated stimuli used in Papers 1–4, and for a detailed description of the procedure. A brief description is provided here.

Each subject heard all of the gating materials (consonants, words, and sentences) in one session. All participants started with the identification of consonants, followed by words, and finally sentences. The type of condition (silence or noise) was counterbalanced across participants, such that half of the participants started with consonant identification in silence and then proceeded to consonant identification in noise, and the other half performed the tasks in the opposite order. In order to prevent any order effects, there were eight different presentation orders (i.e., different lists) of the speech stimuli. Each list consisted of consonants (silence and noise), words (silence and noise) and final words in sentences (silence and noise). The order of items within each list varied across different lists. Furthermore, the lists used for silence and noise were counterbalanced across participants.

Participants received written instructions about the task requirements, which varied depending on the type of material being presented (consonants, words, and sentences). Before any gated task was administered, the participants had a training session to familiarize them with the stimuli, specifically with the presentation of the stimuli in the gating paradigm, and to conduct some trial runs. In the training session, the subjects practiced three consonants (/k v ŋ/) and two words (/tum [inch] and /bil [car]/). Feedback was provided during the practice, but not during the experiment. After the practice session, the gating paradigm tasks started. Participants gave their responses aloud, and the experimenter recorded the responses. The presentation of gates continued until the target was correctly identified on six consecutive presentations. If the target was not correctly identified, presentation continued until the entire target was presented, even if six or more consecutive responses were identical. The next trial was then started. When a speech stimulus was not identified correctly, even after the presentation of the whole speech stimulus, its total duration plus one gate size was estimated as an IP (Elliott, Hammer, & Evan, 1987; Hardison, 2005; Metsala, 1997; Walley, Michela, & Wood, 1995).
Each subject was tested individually in a quiet room. The gating tasks (consonants, words, and sentences) were completed in one session (the first session) with short rest periods to prevent fatigue. In the second session, participants completed the reading span test, the PASAT, and the HINT. The order of tests was counterbalanced across the participants.

**Paper 2**

A detailed description of the synchronization of auditory and visual stimuli is available in Lidestam (2014). Please see Paper 2 for a detailed description of the presentation of the audiovisual speech stimuli to participants.

The gating procedure, the scoring of the gated speech stimuli, and the manner of conducting the HINT and cognitive tests were identical to those described in Paper 1.

**Paper 3**

Participants were tested individually in a soundproof booth. The gated speech signals were delivered via a laptop computer routed to the input of two loudspeakers. The speakers were located within the soundproof booth, 70 cm in front of the participants, at 20° azimuth to the left and right of the participants. The computer was outside the booth, in front of the experimenter, enabling the experimenter to control the stimulus presentation. The experimenter recorded the participants’ verbal responses. The presentation level of the gated stimuli was calibrated to 65 dB sound presentation level (SPL) at the level of the participant’s head.

The first session started with a pure-tone hearing threshold (125–8000 Hz) measurement. Then, the participants underwent the gating tasks. The gating procedure and scoring of IPs was similar to that described in Papers 1 and 2. In the second session, the reading span and word comprehension tests were administered.

**Paper 4**

Participants were tested individually in a quiet room. They sat in front of a CRT monitor. The screen of the monitor was turned off during auditory-only presentation. The auditory stimuli were delivered binaurally via earphones (Sennheiser HDA200). The experimenter recorded the participant’s verbal responses. Before and after the training phase (i.e., auditory-only and audiovisual gating tasks, respectively, and the no-training control task [watching a movie clip]), the speech-in-noise identification ability of participants was tested with the HINT. The presentation level for the gated stimuli was calibrated to 67 dB SPL with earphones. In addition, after the training phase, the participants rated their own effort on a questionnaire.
**Paper 5**

*Screening phase:* The screening phase took part in the lecture halls at Linköping University in order to select the students with the best speechreading ability. The stimuli were video-recordings of the best-identified 30 words that have been used by Lidestam and Beskow (2006). The speech stimuli were presented on large screens via video projectors. The participants’ task was to speechread the words (presented on the large screens) and then write down their responses on the response sheet.

*Experimental phase:* The best-performing participants from the screening phase were invited to the main experiment. The participants were seated in front of the monitor at a distance of approximately 60 cm. A response sheet with numbered lines (one line for each stimulus) was given to participants. The participants underwent a training procedure to familiarize themselves with the speechreading of words. The experimenter asked the participants to respond to all presented words and encouraged them to guess if needed. After each listening condition (i.e., silence, steady-state white noise, and four-talker babble), the participants completed a questionnaire, which focused on their subjective experiences (in terms of effort, performance, and distraction) of each of the listening conditions.

**Tests**

Table 1 shows the speech and cognitive tests used in Papers 1–5. Full descriptions of the tests are available in the papers.
Table 1: The speech and cognitive tests used for the present thesis

<table>
<thead>
<tr>
<th>Test</th>
<th>Modality of the test</th>
<th>Specifics</th>
<th>Dependent variable</th>
<th>Papers utilizing the test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonant identification</td>
<td>Auditory &amp; audiovisually</td>
<td>18 Swedish consonants</td>
<td>IP &amp; accuracy</td>
<td>1, 2, 3, and 4</td>
</tr>
<tr>
<td>Word identification</td>
<td>Auditory, visually, &amp; audiovisually</td>
<td>Swedish monosyllabic nouns</td>
<td>IP &amp; accuracy</td>
<td>1, 2, 3, 4, and 5 (only accuracy used in Paper 5)</td>
</tr>
<tr>
<td>Final word identification</td>
<td>Auditory &amp; audiovisually</td>
<td>Swedish monosyllabic nouns</td>
<td>IP &amp; accuracy</td>
<td>1, 2, and 3</td>
</tr>
<tr>
<td>Hearing In Noise Test (HINT)</td>
<td>Auditory</td>
<td>Sentence comprehension in noise</td>
<td>SNR</td>
<td>1, 2, and 4</td>
</tr>
<tr>
<td>Reading span test</td>
<td>Visually</td>
<td>Working memory capacity</td>
<td>Number of correctly recalled items</td>
<td>1, 2, and 3</td>
</tr>
<tr>
<td>Paced Auditory Serial Addition Test (PASAT)</td>
<td>Auditory</td>
<td>Attentional capacity</td>
<td>Number of correct sums of the last two digits in a sequence of presentation which varied in interval presentations (2 and 3 seconds intervals)</td>
<td>1 and 2</td>
</tr>
<tr>
<td>Ordförståelse B (Word comprehension test B)</td>
<td>Visually</td>
<td>Vocabulary comprehension</td>
<td>Number of correct answers</td>
<td>3</td>
</tr>
</tbody>
</table>
Participants

Paper 1
Twenty-one university students (9 females, 12 males) at Linköping University participated in this study. They were native speakers of Swedish and their ages ranged from 20 to 33 years ($M = 24.6$ years). All of them reported having normal hearing and vision (or corrected-to-normal vision), with no neurological or psychological pathologies.

Paper 2
Twenty-four university students (13 females, 11 males) at Linköping University, who were independent from the students in Paper 1, participated in this study. All of the students were native speakers of Swedish and their ages ranged from 19 to 32 years ($M = 23.3$ years). All reported having normal hearing and vision (or corrected-to-normal vision), with no neurological or psychological pathologies.

Paper 3
The participants in this study were divided into two groups: EHA users and ENH individuals.

EHA users: Twenty-four native speakers of Swedish (11 men and 13 men) with a symmetrical bilateral mild to moderate age-related hearing loss, who had been habitual hearing aid users for at least two years, took part in this study. Their ages ranged from 67 to 79 years ($M = 72.4$ years). The participants were chosen from the patients’ list of an audiology clinic at Linköping University Hospital. The participants wore their own hearing aids with no change in the setting of their hearing aids during testing. Please see Paper 3 for the brands and models of hearing aids used by the EHA users. The inclusion criteria for the EHA group were: 1) bilateral age-related hearing loss with an average threshold of $> 35$ dB for pure tone frequencies of 500 Hz, 1000 Hz, 1500 Hz, and 2000 Hz; 2) age over 65 years; and 3) Swedish as their native language.

ENH individuals: Twenty-four native speakers of Swedish with normal hearing (13 men and 11 women) participated in the study. Their ages ranged from 66 to 77 years ($M = 71.5$ years). ENH individuals were selected from the general population in the region corresponding to the hearing clinic catchment area. They were recruited into the study via flyers and invitation letters sent to their addresses. In addition, four of the ENH individuals were spouses of hearing aid users. The inclusion criteria for the ENH group were: 1) an average threshold of $< 20$ dB for frequencies of 500 Hz,
1000 Hz, 1500 Hz, and 2000 Hz; 2) age over 65 years; and 3) Swedish as the native language.

The participants in both the EHA and the ENH groups reported themselves to be in good health. They were free of tinnitus, seizures, middle-ear pathology, Parkinson’s disease, dementia, or other neurological or psychological disorders that might impact on their performance in the cognitive and speech tasks.

**Paper 4**

Sixty university students (33 females, 27 males) from Linköping University took part in this study. All were native speakers of Swedish with a mean age of 23.2 years. The participants were randomly allocated to 3 groups (auditory-only, audiovisual, and no-training control group) with 20 participants in each (9 males and 11 females in each). The mean ages of participants in the auditory-only, audiovisual, and no-training (control) groups were 23.5, 23.2, and 22.7 years, respectively. All of them reported having normal hearing and vision (or corrected-to-normal vision), with no neurological or psychological pathologies.

**Paper 5**

A total of 147 native speakers of Swedish who were students at Linköping University (53 males, 90 females, and 4 who did not divulge age and gender) participated in a screening test, which enabled the researchers to select a group of participants who had above-average speechreading ability and reported having normal hearing and vision (or corrected-to-normal vision), with no neurological or psychological pathologies. Out of these 147 students, the 43 students with the best scores in the screening phase were invited to the main experiment. Twenty-three of these students agreed to participate in the main experiment.
Summary of the Papers

Paper 1

Purpose
This study predicted that noise delayed IPs and reduced accuracy for the identification of different types of auditory speech stimuli (consonants, words, and final words in sentences). According to the ELU model (Rönnberg et al., 2008), significant correlations between the IPs of different types of auditory speech stimuli in noise and tests of participants’ working memory and attentional capacities were predicted.

Method
Twenty-one university students were presented with auditory-gated stimuli (consonants, words, and final words in HP and LP sentences) in both silent and noisy conditions. The HINT, the reading span test, and the PASAT were administered to measure the participants’ speech-in-noise ability, working memory capacity, and attention capacity, respectively.

Results and discussion
The results showed that noise delayed and reduced the identification of consonants, words, and final words in HP and LP sentences. HINT performance was correlated with the working memory and attentional capacities of participants. In the noise condition, there were correlations between the IPs of consonants and words, and the cognitive capacity of the participants. In silence, there were no correlations between the auditory and cognitive tasks. Significant correlations were observed between the HINT, the reading span test, and PASAT 2. In conclusion, although noise delayed and reduced accuracy in the identification of different types of speech stimuli, the cognitive demands depended on the type of speech stimuli. The identification of auditory speech stimuli under noisy conditions is cognitively demanding only when a supportive semantic context is lacking.

Paper 2

Purpose
This study sought to replicate and extend the study in Paper 1, using the same stimuli but presented in an audiovisual modality and in another sample of participants. We predicted that noise would delay and reduce the accuracy for the identification of different types of audiovisual speech stimuli (consonants, words, and final words in LP and HP sentences). In addition, we predicted that the audiovisual presentation of
speech stimuli would result in faster and more accurate identification of speech stimuli than auditory presentation of speech stimuli. Finally, it was predicted that audiovisual presentation of speech stimuli would reduce the cognitive demands of perceiving speech stimuli, compared to the same SNR of auditory-only presentation (Paper 1).

**Method**

The audiovisually-gated stimuli (consonants, words, and final words in LP and HP sentences) were presented in both silence and noise to 24 young university students. The HINT, the reading span test, and the PASAT were also administered to measure the participants’ speech-in-noise ability, working memory capacity, and attentional capacity, respectively.

**Results and discussion**

The results indicated that noise delays the audiovisual identification of consonants and words, but not final word identification in HP and LP sentences. There were no relationships between the IPs of audiovisual stimuli, in either silence or noise, and the working memory and attention test results. In comparison with the auditory-gated study (Paper 1), it can be concluded that the provision of visual cues (at an equivalent SNR to auditory-gated speech stimuli) increased intelligibility for the identification of speech stimuli in both silence and noise. In addition, audiovisual presentation at the same SNR as auditory presentation reduced the cognitive demands associated with perceiving speech stimuli. Similar to Paper 1, significant correlations were observed between the HINT, and the reading span test and PASAT 2. In conclusion, audiovisual presentation accelerates and improves the identification of speech stimuli, in both silent and noisy conditions, compared to auditory presentation of speech stimuli. In addition, the audiovisual identification of speech stimuli is less cognitively demanding than auditory identification of speech stimuli under noisy conditions.

**Paper 3**

**Purpose**

This study aimed to compare the IPs and accuracy rates for the identification of different types of auditory speech stimuli (consonants, words, and final words in sentences) presented in silence between EHA users and ENH individuals, and to examine the relationships between working memory capacity and the IPs and accuracy rates for different types of auditory speech stimuli, in EHA users and ENH
individuals. We predicted that, due to suboptimal amplification of sounds by hearing aids or the adverse consequences of hearing loss on higher-order processing of speech stimuli, the EHA users would have inferior performance to ENH individuals in the speed and accuracy of speech stimuli in the absence of a supportive semantic context. As a consequence, the EHA users with better working memory capacities would identify speech stimuli earlier and more accurately when a supportive semantic context is lacking.

**Method**

Auditory-gated speech stimuli were presented to 24 EHA users (with mild to moderate hearing loss) and 24 ENH individuals. The reading span test was used to measure participants’ working memory capacity.

**Results and discussion**

The EHA users showed delayed IPs and lower accuracy for the identification of consonants and words compared to the ENH individuals. The EHA users also had delayed IPs for final word-identification in LP sentences; however, no significant between-group difference in accuracy was observed. There were no significant differences between the two groups in terms of IPs or accuracy for final-word identification in HP sentences. In EHA users, greater working memory capacity was associated with shortened IPs and better accuracy in the identification of consonants and words. Opposite to our predictions, the EHA users had inferior performance relative to ENH individuals in auditory identification of speech stimuli when a supportive semantic context is lacking. This is most likely due to poor sensory coding of auditory speech stimuli and or adverse effects of age-related hearing loss on higher-order processing of speech stimuli in the brain. As a consequence, the processing of auditory speech stimuli in conditions that lack prior semantic context may be cognitively demanding for the EHA users.

**Paper 4**

**Purpose**

This study aimed to compare the effects of audiovisual versus auditory pre-training on subsequent auditory speech-in-noise identification performance. This study predicted that exposure to multisensory speech stimuli would result in better performance on subsequent unisensory processing of speech stimuli.
Method
Sixty young university students were randomly divided into three groups (n = 20 per group): auditory-only training, audiovisual training, and a control group. During auditory-only and audiovisual training, the gated consonants and words were presented at 0 dB SNR to be identified either audiovisually or aurally-only; the control group watched a movie clip without performing a speech identification task. The HINT was completed before and after the training (or control session).

Results and discussion
Prior audiovisual training improved speech-in-noise performance in comparison with auditory-only training and no training (control group), suggesting a superiority over auditory-only training for speech-in-noise identification performance.

Paper 5
Purpose
This study aimed to compare the effects of two types of noise (four-talker babble vs. steady-state white noise) tested against a baseline (silence) on speechreading accuracy for words. The four-talker babble was included to provoke competition for phonological processing, while an effect of steady-state white noise would reflect a more general effect of attentional processing. We predicted no influence of steady-state white noise on the speechreading accuracy, while four-talker babble would reduce the speechreading accuracy in comparison with steady-state white noise or quiet conditions, by its presumed elicitation of competition for phonological processing.

Method
Twenty-three participants with above-average speechreading ability took part in this study. Their task was to speechread highly frequent Swedish words, and to subjectively rate their own performance and effort in each type of listening condition (silence, four-talker babble, and steady-state white noise).

Results and discussion
Only the four-talker babble adversely affected the speechreading of words. This finding suggested the existence of phonological competition or interference between the contents of the four-talker babble and the visual articulation of words, which impeded the speechreading of words in this condition compared to speechreading in quiet and steady-state white noise conditions. The findings were in line with our
prediction that the type of background noise is a key factor that can affect the speechreading of words.
General Discussion

Main findings and conclusions

The first and foremost main finding was that background noise adversely affected, in terms of IPs and accuracy, the identification of all types of speech stimuli (consonants, words, and final words in sentences) when presented in the auditory modality. In addition, noise adversely affected the IPs for all types of audiovisual speech stimuli. In terms of accuracy, noise only adversely affected the identification of consonants and words, but not final words in sentences. Moreover, the results of Paper 2 showed that audiovisual presentation improved the recognition of speech stimuli, in terms of both IPs and accuracy, relative to auditory-only presentation, for all types of speech stimuli. A secondary finding of Paper 2 was that the listeners who had been exposed to the audiovisual gated tasks had better speech-in-noise identification than those exposed to the auditory-only gated tasks. Paper 4 was conducted to investigate whether this audiovisual advantage on speech-in-noise identification was a genuine effect. The findings of Paper 4 corroborated the secondary effect revealed in Paper 2, and confirmed that prior exposure to audiovisual stimuli can result in better auditory speech-in-noise identification than early exposure to auditory stimuli only.

Paper 5 showed that in the visual-only modality (or speechreading), only the four-talker babble background noise (informational masking) adversely affected the accuracy for the identification of words, while steady-state white noise (energetic masking) did not, suggesting an effect of phonological distraction. Further, the results showed that the distracter signal (maskers) did not need to be simultaneous in terms of onset and offset, and thus do not support the assumption made by Brungart and Simpson (2005).

Note that in the present thesis, only steady-state white noise was used as a background noise in the gating paradigm tasks (Papers 1–4). The nature of the gating paradigm task did not allow us to use informational masking (e.g., four-talker babble). Because we used a small gate size (i.e., 16.67 ms), it would have been possible, using fluctuating talker babble, that there could be noise in one gate, but not in the following gate. Hence, we chose a steady-state white noise to be sure that noise was equally present at all gate points.

The findings of Paper 3 showed that the EHA users needed more time (as indicated by later IPs) for the identification of consonants, words, and final words in
LP sentences (but not for final words in HP sentences) when compared with ENH individuals. Furthermore, EHA users had lower accuracy for the identification of consonants and words, but not for final words in LP and HP sentences.

In addition, our findings in Paper 1 showed that IPs for consonants and words presented aurally in the noise condition were significantly correlated with the test results for working memory and attentional capacities. However, when visual cues were added to the auditory cues (i.e., stimuli were presented in an audiovisual modality), no significant correlation was observed between the IPs of consonants and words (in the noise condition) and test results for working memory and attention. Furthermore, in both Papers 1 and 2, participants’ working memory and attentional capacities were significantly correlated with HINT performance (a measure of speech-in-noise identification). The findings of Paper 3 showed that the IPs and accuracy rates for consonants and words (presented in an auditory modality under conditions of silence) were significantly correlated with the working memory capacity of EHA users, but not for ENH individuals.

Adverse effects of noise on speech recognition depend on the modality of presentation of speech stimuli and the type of speech stimuli.

**Consonants.** Noise significantly reduced accuracy and delayed the identification of consonants in both auditory and audiovisual modalities. Auditory noise (e.g., steady-state white noise) in combination with only the acoustic properties of consonants, causes a perceptual change that makes their identification difficult (as illustrated by extended IPs) or it leads to the morphing of a given consonant into one or more other consonants, which in turn reduces the chance of a consonant being correctly identified. In the audiovisual modality, the correspondence between visual and auditory articulations of consonants is not always one-to-one. Visual articulation of some consonants such as /b p/, /d t/, or /r l/ are the same. Noise in combination with the auditory component of consonants presented in an audiovisual modality necessitated longer consonant viewing and listening time (greater segments of consonants were needed) for correct identification, compared with the silent condition. In addition, the combination of noise with the auditory component of audiovisually presented consonants is likely to have led to morphing into one or more other consonants. This is supported by the accuracy rate of
audiovisually presented consonants identified in the noise condition, which was decreased compared with that in the silence condition.

The addition of visual cues to auditory cues in the identification of consonants is proposed to contribute “complementarity (greater benefit resulting from combination of audio and video)” or “redundancy (less benefit resulting from combination of auditory and visual speech information)” (see Grant & Walden, 1996; Grant et al., 1998). These complementarity or redundancy effects are dependent on the information provided by each modality for the identification of a given consonant. These effects in the previous studies were based on the analysis of confusion matrices of consonants (using accuracy). IP results may be more sensitive than accuracy (i.e., greater variation, cf. Gatehouse & Gordon, 1990) in demonstrating the complementarity or redundancy effects in the identification of consonants. When comparing IPs for audiovisual identification of consonants relative to auditory identification in silence (Paper 2), the addition of visual cues to auditory cues for all consonants, except /r m/ were complementary, as the audiovisual presentation significantly shortened the IPs (except for /r m/) relative to auditory presentation. Thus, there are both complementarity (for some consonants other than /r m/) and redundancy effects (for /r m/) in silent conditions for the identification of Swedish consonants. The IP results in the noise condition revealed that the addition of visual cues to auditory ones significantly shortened the IPs for all consonants, indicating only complementarity effects. All together, the complementary or redundancy effects in the identification of Swedish consonants depend on the type of listening condition: we did not see any evidence of redundancy when consonants were presented with noise.

**Words.** In the auditory and audiovisual modalities, noise (steady-state white noise) increased the IPs and reduced accuracy for the identification of words. During visual-only speechreading, only four-talker babble noise (and not steady-state white noise) reduced the accuracy for the identification of words. White noise in combination with the acoustic cues of vowels and consonants, as the main constituents of words, in the audiovisual modality, presumably activates other phonological-lexical candidates and makes it difficult to move from one lexical candidate into another one (Singer, Bronstein, & Miles, 1981). Listeners therefore need to hear or watch target words for longer parts of the stimulus words in the noise condition (as reflected by later IPs) than in the silence condition, in order to
accumulate enough information to settle for one particular candidate. However, in some cases, later IPs were not sufficient for the correct identification of words, and listeners of words presented in auditory or audiovisual modalities sometimes perceived other, non-target words. This resulted in a reduction in the correct identification of words in the noise condition, particularly for those presented in the auditory modality.

In the speechreading of words, because there is no combination of noise with the visual articulation of words, no adverse effect of energetic masking was observed. Informational masking, however, impaired the speechreading of words not at the sensory level, but at a higher-order level, by drawing the listeners’ attention to the meaningful contents of masking, which was triggered by the four-talker babble (thereby distracting the listeners).

**Final words in Sentences.** Noise negatively affected the IPs and accuracy for final words in sentences presented in an auditory modality but not those presented in an audiovisual modality. That is, the combination of visual cues and preceding context appears to cancel the deleterious effects of noise on the identification of final words in sentences.

The **use of hearing aids does not fully compensate for the effects of hearing loss in terms of auditory speech recognition abilities: comparisons of EHA users with elderly normal-hearing, age-matched, participants (ENH).**

The findings of Paper 3 demonstrated that the use of hearing aids did not fully compensate for speech recognition difficulties experienced by hearing-impaired individuals at the acoustical (consonants) and lexical (words) levels, when compared with the ENH individuals. However, when target speech stimuli were embedded within a congruent semantic context (in particular in HP sentences), the difference between the two groups disappeared. This inferior performance of EHA users relative to ENH might be because of poorer sensory coding (that is not compensated for by hearing aid use) or may be caused by the adverse effects of age-related hearing loss on the brain areas responsible for processing auditory speech stimuli (Boyen et al., 2013; Eckert et al., 2012; Husain et al., 2011; Peelle et al., 2011). Structural changes in the brains of hearing-impaired individuals may explain why amplification of sounds by hearing aids was not fully successful to compensate for the hearing difficulty of hearing aid users to the level of their matched normal hearing controls.
This issue is important for the rehabilitation of hearing-impaired individuals. The dominant view is that the speech recognition difficulties experienced by hearing-impaired individuals are solely due to poor sensory coding, and that the amplification of sounds by hearing aids can compensate for these difficulties. Perhaps other rehabilitation approaches need to be devised (in addition to hearing aids) in order to fully compensate for the speech difficulties of hearing-impaired individuals.

The present thesis suggests that a post-fitting evaluation of phonemic and lexical abilities of hearing-aid users should be compared with the norms for normal-hearing individuals in the same age range. IP norms at both phonemic and lexical levels may be helpful for a precise and optimized post-fitting evaluation of hearing aid users (cf. Gatehouse & Gordon, 1990). Deviations from these norms may indicate that a given hearing aid user needs further rehabilitation (e.g., perceptual training to improve syllable identification, Stecker et al., 2006) in order for his or her speech recognition abilities to be closer to the level of age-matched counterparts with normal hearing.

The cognitive demands associated with perceiving speech stimuli greatly depend on the listening conditions, type of speech stimuli, modality of presentation, and hearing loss.

The auditory gating findings (Paper 1) showed that the IPs for different types of speech stimuli presented in silence were not correlated with measures of explicit cognitive function. However, in the noise condition, the IPs for consonants and words were correlated with measures of explicit cognitive function. Our results suggest that the identification of consonants and words is cognitively demanding, which is in agreement with previous studies that have shown effortful recognition of phonemes (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Dehane-Lambertz et al., 2005; Dufor, Serniclaes, Sprenger–Charolles, & Démonet, 2007) and words (Hervais–Adelman et al., 2012) in degraded listening conditions. In addition, the findings of the present thesis regarding the auditory identification of speech stimuli under degraded listening conditions are not in agreement with autonomous models of speech perception (Norris et al., 2000), as the involvement of explicit cognitive resources is a key variable for the early identification of consonants and words under degraded listening conditions.
All together, these results are in agreement with a recent version of the ELU model (Rönnberg et al., 2013). The model predicts that under optimal listening conditions, the matching process between the incoming signals and their corresponding phonological representations is rapid and implicit. However, in degraded listening situations and in the absence of prior semantic context (i.e., consonants and words), the matching process between the speech signal and its corresponding phonological representation is slow and cognitively demanding.

In addition, our findings in Papers 1 & 2 are in agreement with the general assumption of cognitive spare capacity (Mishra, 2014). The processing of speech stimuli under noisy conditions and in the absence of visual speech cues needs the contribution of explicit cognitive processes in order to extract meaning from degraded speech stimuli (Paper 1). According to the cognitive spare capacity hypothesis (Mishra, 2014), the addition of visual cues is the only way to free up cognitive spare capacity. Our findings in Papers 1 and 3 revealed that prior semantic context acts as another way to free up spare capacity. Hence, our findings extend the cognitive spare capacity hypothesis by showing that not only visual cues but also prior semantic context free up spare cognitive resources to aid listeners in perceiving stimuli under degraded listening conditions. In fact, semantic context has been assumed to act as a “top-down repair of word’s meaning” (Sivonen, Maess, & Friederici, 2006) or alleviates “bottom-up decline” in older adults with hearing loss (Wingfield, Tun, McCoy, Stewart, & Cox, 2006). Thus, prior semantic context facilitates the identification of final words in degraded listening conditions by constricting the number of words that could appear at the end sentences (cf. Obleser, Wise, Dresner, & Scott, 2007). As a consequence, the identification of final words in sentences (at the same SNR relative to words-alone) becomes less cognitively demanding than the identification of words presented in isolation.

In Papers 1 and 2, significant correlations between the cognitive tests (reading span test and PASAT 2) and the HINT (a measure of sentence comprehension in noise test) have been observed in young normally-hearing listeners. As noted earlier, the performance in the HINT differs from the identification of final words in sentences, as the HINT requires participants to correctly repeat the entire of sentences in an adaptive SNR. These findings were recently supported by Kilman, Zekveld, Hällgren, and Rönnberg (2014) who reported a significant relationship between the HINT (steady-state noise) and reading span test in normal-hearing listeners. These
findings are also consistent with neuroimaging studies demonstrating that a top-down cognitive effort is required for comprehension of degraded sentences (Wild, Davis, & Johnsrude, 2012; Wild et al., 2012).

Note that in Papers 1 and 2, there was no significant correlation between PASAT 3 and the HINT, denoting that the extent to which a cognitive test is tapping into a given cognitive capacity of a participant is a key variable to illustrate the cognitive demands of speech recognition in degraded listening conditions (cf. Akeroyd, 2008).

Obtaining the IPs of audiovisual speech stimuli and then calculating the correlations between the IPs and tests of working memory and attentional capacities would allow us to investigate the extent to which the integration of auditory and visual components of audiovisual speech stimuli is explicit or implicit. Our findings showed no correlations between IPs and tests of working memory and attentional capacities in both silence and noise, suggesting that audiovisual integration is implicit in both silence and noise. This findings support the RAMBPHO in the ELU model (Rudner & Rönnberg, 2008; Rönnberg et al., 2008, 2013), as multimodal binding of incoming signals is not assumed to depend on explicit cognitive resources.

One further explanation for implicit audiovisual speech identification in noise can be derived from the multimedia dual-processing theory (Mayer & Moreno, 1998). According to this model, there are two separate processing systems for the identification of audiovisual stimuli in working memory: one for visual information and one for auditory information. Each of these systems has a limited processing capacity. In audiovisual processing of information, the auditory information is processed in the phonological subsystem, while the visual information is processed in the visual subsystem. In such a case, the combined resources of both subsystems are used to process audiovisual information. Hence, there is presumably more capacity available for the processing of audiovisual information than for auditory or visual information alone. Whereas there is a high demand on the limited capacity of the phonological subsystem in working memory for the processing of degraded auditory stimuli, and this results in explicit (effortful) identification of auditory speech stimuli in noise (Paper 1), there appears to be a relatively lower demand on processing of audiovisual speech stimuli in noise due to the increased capacity by the visual subsystem. This results in implicit (effortless) identification of audiovisual speech stimuli (Paper 2).
One explanation for the implicit audiovisual integration in the noise condition in Paper 2 might be the fact that the SNR (0 dB) in the noise condition was not low enough to invoke explicit cognitive resource (i.e., working memory or attention) for processing of the audiovisual speech stimuli. The auditory component of audiovisual speech signals is a key variable in audiovisual speech identification in hearing-impaired (Corthals, Vinck, De Vel, & Van Cauwenberg, 1997) and normal-hearing (Baart, Vroomen, Shaw, & Bortfeld, 2014) listeners. Future studies are needed to further investigate the cognitive demands of audiovisual word identification at lower SNRs (e.g., −5, or −10 SNR dB) than the SNR used in the Paper 2 (0 dB SNR), or under conditions of informational masking. In such cases, individuals with greater explicit cognitive resources may demonstrate better audiovisual identification of consonants and words.

In addition, the findings of Paper 3 showed that in the absence of prior semantic context, the identification of consonants and words was cognitively demanding for EHA users, but not for ENH individuals. These findings can be explained by two mechanisms: first, poor sensory coding of consonants and words by EHA users places higher processing demands on perception (ELU model, Rönnberg et al., 2013); second, changes in the structural and functional auditory system of EHA users make the processing of consonants and words cognitively demanding, even when hearing aids are used (cf. Campbell & Sharma, 2013).

The present thesis suggests that in addition to the current use of audiograms and speech audiometry for fitting hearing aids to hearing-impaired individuals, audiologists should measure the cognitive demand of perceived speech stimuli (at the phonemic and lexical levels) after an acclimatization time period (typically 1–3 months after hearing aids have been fitted). For instance, pupillometry (measuring pupil dilation) is a valid measure to estimate cognitive effort required to perceive speech stimuli under degraded listening conditions (Zekveld, Kramer, & Festen, 2010, 2011). Using pupillometry during listening to consonants or words may help an audiologist to estimate the extent to which the current settings of hearing aids for a given hearing-impaired individual are cognitively demanding. If pupillometry shows a great cognitive effort during listening to consonants and words with the current hearing aid setting, perhaps it is a sign to try another setting in order to fit the hearing aids according to the patient’s cognitive capacity.
Paper 5 did not examine the cognitive demands of speechreading for words presented in both silence and noise conditions. However, Lyxell and Rönnberg (1993) reported that better performance on a test of working memory capacity (reading span test) was correlated with better performance in speechreading under the informational masking condition. It can be assumed that listeners with greater explicit cognitive resources (i.e., working memory and attention) were better able to suppress phonological competition caused by irrelevant speech (informational masking) than those with lower explicit cognitive resources.

**Audiovisual presentation (relative to auditory presentation) expedites IPs and improves accuracy for the identification of speech stimuli.**

**Consonants.** Our results showed that the audiovisual presentation of consonants resulted in rapid and more accurate identification of consonants in both silence and noise. Summerfield (1987) suggested that the addition of visual cues to the auditory presentation of consonants facilitates the identification of consonants by: 1) revealing an auditory source of speech (talker localization) and 2) providing complementary cues about the place of articulation. In a noise condition, listeners most likely focus more on visible articulation as observed on a talker’s face (such as tongue, teeth, and lip movements) to acquire complementary information (cf. Buchan, Parè, & Munhall, 2008), which is probably matched with the corresponding auditory cues of a given consonant.

**Words.** Our findings showed that the audiovisual presentation of words resulted in rapid and more accurate identification in both silence and noise. Tye-Murray et al. (2007b) proposed that the audiovisual presentation of words resulted in fewer phonological neighborhoods compared to an auditory-alone modality. This facilitates the process of matching between an incoming speech signal and its corresponding phonological representation in the mental lexicon. Paper 2 showed that the effect size of the audiovisual presentation relative to auditory presentation was larger in noise than in silence (Cohen’s $d = 7.85$ vs. $3.87$), corroborating prior studies that showed the importance of visual cues in word identification under difficult conditions (cf. Fort et al. 2013).

**Final words in sentences.** Our results showed that final-word identification in an audiovisual modality is rapid and more accurate (than identification in an auditory-only modality) for target words in both LP and HP sentences, in silence and noise conditions. In fact, for the audiovisual identification of final words in sentences, there
is an added benefit from the preceding context and the provision of visual cues, which improves listeners’ abilities to correctly identify the target word at the end of a given sentence.

**Prior exposure to audiovisual (rather than auditory-only) speech stimuli improves speech-in-noise performance.**

One might intuitively expect that auditory-only training in perceiving speech-in-noise should improve speech-in-noise identification more than audiovisual training, if the criterion is auditory-only speech identification ability. The findings of Paper 2 and those of the subsequent randomized-controlled study in Paper 4 challenged this notion by showing that prior exposure to audiovisual rather than auditory-only speech stimuli resulted in better performance in a speech-in-noise task. The interesting aspect of these findings is that the advantage of audiovisual presentation was not restricted to the idiosyncrasies of the talkers (familiarity of the talkers), as the talkers in the audiovisual or auditory gating tasks and subsequent speech-in-noise test were not the same. This finding is important and sets the studies apart from others that have shown that familiarity with talkers is a key component for subsequent speech-in-noise identification (Johnsrude et al., 2013; Rosenblum, Miller, & Sanchez, 2007; Wu, Cao, Wu, & Li, 2013). The fact that learning appeared to generalize from one talker to another in our studies has practical and clinical consequences in hearing-impaired individuals (e.g., hearing-aid users, children with cochlear implants) in order to improve their speech-in-noise performance.

One explanation as to why the audiovisual presentation of stimuli in Papers 2 and 4 resulted in better speech-in-noise identification might be that audiovisual stimuli presented in a gating format probably focuses the attention of listeners on the visual articulation of speech by the talker. Future studies are needed to compare the effects of gated versus normal presentation of audiovisual speech stimuli on subsequent speech-in-noise performance. This comparison will help us to elucidate whether the gated presentation of audiovisual speech results in better speech-in-noise performance, or whether exposure to audiovisual speech stimuli (regardless of gated or normal presentation) is sufficient for improving speech-in-noise identification.
Future Directions for Interdisciplinary Research and Applications

An interdisciplinary approach for rehabilitation of hearing-impaired individuals

The present thesis recommends an interdisciplinary approach (e.g., a bio-psycho-social approach) in the rehabilitation of hearing-impaired individuals. At present, the dominant view in the rehabilitation of hearing-impaired individuals is a medical approach; for example, audiograms are obtained and speech audiometry is performed with the aim of providing hearing-impaired individuals with hearing aids. However, psychological and social approaches can also be taken into account.

At the psychological level, obtaining information regarding the cognitive capacity of hearing-impaired individuals may ensure that hearing-aid settings are matched with the cognitive capacity of a given individual (cf. Ng et al., 2013). At present, little is known about the extent to which personality traits or the mental health status of hearing-impaired individuals may impact on the benefit from hearing aids. The present thesis recommends that future studies examine the effects of these variables on the benefit yielded by hearing aids. At the social level, an understanding of how hearing loss (or wearing hearing aids) may impact on job performance and on the likelihood of social integration (with family, friends, or colleagues) or social rejection would be valuable for professionals who work with hearing-impaired individuals.

Contextualizing the biological, psychological, and social aspects of hearing loss in the rehabilitation of hearing-impaired individuals would improve health functioning (ICF, World Health Organization, 2001) compared to using only the medical approach.

Audiovisual training as a rehabilitation method for improving speech recognition abilities in individuals with hearing difficulties (e.g., hearing aid users and children with cochlear implants).

Hearing aids do not fully restore the speech recognition abilities of hearing aid users to the same level as in their counterparts with normal hearing (cf. Dimitrijevic et al., 2004). Auditory training programs have been devised to improve speech-in-noise performance among special groups, such EHA users (Lavie, Attias, & Karmi, 2013). Independent studies have shown that such auditory training programs are associated with improved gap-detection thresholds (Kishon-Rabin, Avivi-Reich, & Roth, 2013).
and improved neural representation of auditory cues important for speech-in-noise identification (Song, Skoe, Banai, & Kraus, 2012). Generalization of learning to an unfamiliar talker, as observed in Papers 2 and 4 may inspire audiologists, speech therapists, or professionals who work with individuals with hearing difficulties to use audiovisual training instead of auditory-only training for improving speech capabilities.

A comparison of audiovisual gated speech recognition abilities in EHA users versus ENH individuals

The addition of visual cues to auditory signals amplified by hearing aids resulted in better identification of speech stimuli compared to unaided audiovisual or aided auditory-only modalities (Walden et al., 2001). A preliminary finding by Moradi, Lidestam, and Rönnberg (2014) showed that despite EHA users and ENH individuals having ceiling or near to ceiling performance in terms of accuracy for audiovisual-gated speech stimuli (consonants, words, and final words in sentences), the EHA users needed longer IPs for identification than ENH individuals in the absence of a prior semantic context. The extent to which the addition of visual cues to amplified sounds facilitates identification and reduces the cognitive demands of speech stimuli in EHA users relative to ENH individuals needs further investigation.
Conclusions

1. In young normally-hearing listeners, the deleterious effects of noise on the speed and accuracy of identification of speech stimuli greatly depend on the type of speech stimuli presented, the modality of presentation of the speech stimuli, and the type of background noise used. In the auditory modality, (steady-state white) noise delays and reduces the accuracy in the identification of all types of speech stimuli (i.e., consonants, words, and final words in HP and LP sentences). In the audiovisual modality and at the same SNR relative to the auditory modality, noise (steady-state white) only delays and reduces the accuracy in identification of consonants and words (but not final words in HP and LP sentences). In the visual modality, only four-talker babble (and not steady-state white noise) reduces the speechreading of words.

2. In young normally-hearing listeners, the cognitive demands of perceiving speech stimuli greatly depend on the listening conditions, type of speech stimuli, and modality of presentation. At the same SNR, in the auditory and audiovisual conditions, only the auditory identification of consonants and words were correlated with the cognitive capacity of listeners. The addition of visual cues or a supportive semantic context reduces cognitive demands.

3. In young normally-hearing listeners, audiovisual presentation of speech stimuli results in faster and more accurate identification of speech stimuli in both silence and noise relative to auditory presentation of those speech stimuli. In addition, prior exposure to audiovisual speech stimuli subsequently results in better speech-in-noise identification than does prior exposure to auditory speech stimuli.

4. When comparing EHA users and ENH individuals, the EHA users had inferior performance in the identification of consonants, words, and final words in LP sentences (in terms of speed). In terms of accuracy, the EHA users had inferior performance only in the identification of consonants and words.

5. In EHA users and ENH individuals, the cognitive demands of perceiving speech stimuli depend on hearing status (EHA vs. ENH group) and the type of speech stimuli. Only the identification of consonants and words correlate with the cognitive capacity of the EHA users.

6. The findings reported in the present thesis lead us to recommend applying an interdisciplinary approach to the rehabilitation of hearing-impaired individuals.
Acknowledgements

First, I would like to thank my supervisors, Jerker Rönnberg (main supervisor) and Björn Lidestam (co-supervisor), for their superb supervisions, guidance, continuing support, encouragements, and great patience throughout my PhD training. Your office doors were always open to me for my questions and when I needed advice and support.

In addition, special thanks to my friends and colleagues at HEAD Graduate School for having you around, knowing that I can count on you for everything: Wycliffe Yumba, Elaine Ng, Cecilia Henricsson, Sushmit Mishra, Josefine Andin, Cecilia Nakeva Von Mentzer, Elisabet Classon, Emelie Nordqvist, Håkan Hua, Jakob Dahl, Emil Holmer, Lisa Kilman, Rachel Ellis, and Carine Signoret.

I have a special thank to Maria Hugo-Lindén for her great administrative help during my PhD education. I wish to acknowledge the support from Mary Rudner, Björn Lyxell, Henrik Danielsson, Anna Levén, and Örjan Dahlström, Berth Danemark, and Kerstin Möller.

Furthermore, I would like to thank Amin Saremi for his technical help and close friendship, Mathias Hällgren for his technical and professional support, and Helena Torlofson for her assistance in parts of my research.
References


Studies from the Swedish Institute for Disability Research

1. Varieties of reading disability
   Stefan Gustafson

2. Cognitive functions in drivers with brain injury – anticipation and adaptation
   Anna Lundqvist

3. Cognitive deafness
   Ulf Andersson

4. Att lära sig leva med förvärvad hörselnedsättning sett ur par-perspektiv
   Carin Fredriksson
   ISBN 91-7373-105-6, 2001

5. Signs, Symptoms, and Disability Related to the Musculo-Skeletal System
   Gunnar Lundberg

6. Participation – Ideology and Everyday Life
   Anette Kjellberg

7. Föräldrar med funktionshinder – om barn, föräldraskap och familjeliv
   Marie Gustavsson Holmström
   ISBN 91-7203-500-5, 2002

8. Active wheelchair use in daily life
   Kersti Samuelsson

9. Två kön eller inget alls. Politiska intentioner och vardaglivets realiteter i den arbetslivsinriktade rehabiliteringen
   Marie Jansson
10. **Audiological and cognitive long-term sequelae from closed head injury**  
Per-Olof Bergemalm  
ISBN 91-7668-384-2, 2004

11. **Att vara i särklass – om delaktighet och utanförskap i gymnasiesärskolan**  
Martin Molin  
ISBN 91-85295-46-9, 2004

12. **Rättvis idrottsundervisning för elever med rörelsehinder – dilemma kring omfördelning och erkännande**  
Kajsa Jerlinder  
Licentiate Degree, 2005

Per-Ingge Carlsson  
ISBN 91-7668-426-1, 2005

14. **Hearing and cognition in speech comprehension. Methods and applications**  
Mathias Hällgren  
ISBN 91-85297-93-3, 2005

15. **Living with deteriorating and hereditary disease: experiences over ten years of persons with muscular dystrophy and their next of kin**  
Katrin Boström  
ISBN 91-7668-427-x, 2005

16. **Disease and disability in early rheumatoid arthritis**  
Ingrid Thyberg  
ISBN 91-85299-16-2, 2005

17. "**Varför får jag icke följa med dit fram?**” Medborgarskapet och den offentliga debatten om dövstumpa och blinda 1860-1914  
Staffan Bengtsson  
ISBN 91-85457-06-X, 2005

18. **Modalities of Mind. Modality-specific and nonmodality-specific aspects of working memory for sign and speech**  
Mary Rudner  
ISBN 91-85457-10-8, 2005
19. Facing the Illusion Piece by Piece. Face recognition for persons with learning disability
Henrik Danielsson
ISBN 91-85497-09-6, 2006

20. Vuxna med förvärvad traumatisk hjärnskada – omställningsprocesser och konsekvenser i vardagslivet. En studie av femton personers upplevelser och erfarenheter av att leva med förvärvad traumatisk hjärnskada
Thomas Strandberg

21. Nycklar till kommunikation. Kommunikation mellan vuxna personer med grav förvärvad hjärnskada och personernas närstående, anhöriga och personal
Pia Käcker

22. ”Aspergern, det är jag”. En intervjustudie om att leva med Asperger syndrom
Gunvor Larsson Abbad

23. Sounds of silence – Phonological awareness and written language in children with and without speech
Janna Ferreira

24. Postponed Plans: Prospective Memory and Intellectual Disability
Anna Levén

25. Consequences of brain tumours from the perspective of the patients and of their next of kin
Tanja Edvardsson

26. Impact on participation and service for persons with deafblindness
Kerstin Möller

27. Approaches to Audiological Rehabilitation with Hearing Aids: studies on prefitting strategies and assessment of outcomes
Marie Öberg
28. **Social Interaction and Participation in Activities of Everyday Life Among Persons with Schizophrenia**  
   Maria Yilmaz  
   Licentiate Degree, 2009

29. **Focus on Chronic Disease through Different Lenses of Expertise Towards Implementation of Patient-Focused Decision Support Preventing Disability:**  
   The example of Early Rheumatoid Arthritis  
   Örjan Dahlström  

30. **Children with Cochlear Implants: Cognition and Reading Ability**  
   Malin Wass  

31. **Restricted participation: Unaccompanied children in interpreter-mediated asylum hearings in Sweden**  
   Olga Keselman  

32. **Deaf people and labour market in Sweden. Education – Employment – Economy**  
   Emelie Rydberg  

33. **Social rättvisa i inkluderande idrottsundervisning för elever med rörelsehinder – en utopi?**  
   Kajsa Jerlinder  
   ISBN: 978-91-7668-726-0, 2010

34. **Erfarenheter av rehabiliteringsprocessen mot ett arbetsliv – brukarens och de professionellas perspektiv**  
   Helene Hillborg  

35. **Knowing me, knowing you – Mentalization abilities of children who use augmentative and alternative communication**  
   Annette Sundqvist  
36. **Lärare, socialsekreterare och barn som far illa – om sociala representationer och interprofessionell samverkan**
   Per Germundsson
   **ISBN:** 978-91-7668-787-1, 2011

37. **Fats in Mind**
   Effects of Omega-3 Fatty Acids on Cognition and Behaviour in Childhood
   Ulrika Birberg Thornberg
   **ISBN:** 978-91-7393-164-9, 2011

38. **”Jobbet är kommunikation”**
   Om användning av arbetshjälpmedel för personer med hörselnedsättning
   Sif Bjarnason
   Licentiate Degree. **ISBN:** 978-91-7668-835-9, 2011

39. **Applying the ICF-CY to identify everyday life situations of children and youth with disabilities**
   Margareta Adolfsson

40. **Tinnitus – an acceptance-based approach**
    Vendela Zetterqvist
   **ISBN:** 978-91-7393-040-6, 2011

41. **Applicability of the ICF-CY to describe functioning and environment of children with disabilities**
    Nina Klang
    **ISBN:** 978-91-7668-864-9, 2012

42. **Bringing more to participation**
    Participation in school activities of persons with Disability within the framework of the International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY)
    Gregor Maxwell
    **ISBN:** 978-91-628-8484-0, 2012

43. **From Eye to Us.**
    Prerequisites for and levels of participation in mainstream school of persons with Autism Spectrum Conditions
    Marita Falkmer
    **ISBN:** 978-91-637-2091-8, 2013
44. **Otosclerosis, clinical long-term perspectives**  
Ylva Dahlin-Redfors  

45. **Tinnitus in Context - A Contemporary Contextual Behavioral Approach**  
Hugo Hesser  

46. **Hearing and middle ear status in children and young adults with cleft palate**  
Traci Flynn  

47. **Utrymme för deltagande, beslutsprocesser i möten mellan patienter med ospecifika ländryggsbesvär och sjukgymnaster i primär vård**  
Iréne Josephson  

48. **"Man vill ju klara sig själv” Studievardagen för studenter med Asperger syndrom i högre studier**  
Ann Simmeborn Fleischer  

49. **Cognitive erosion and its implications in Alzheimer’s disease**  
Selina Mårdh  

50. **Hörselscreening av en population med utvecklingsstörning**  
Utvärdering av psykoakustisk testmetod och av OAE-registrering som komplementär metod  
Eva Andersson  

51. **Skolformens komplexitet – elevers erfarenheter av skolvardag och tillhörighet i gymnasiesärskolan**  
Therése Mineur  

52. **Evaluating the process of change: Studies on patient journey, hearing disability acceptance and stages-of-change**  
Vinaya Kumar Channapatna Manchaiah  
53. **Cognition in hearing aid users: Memory for everyday speech**  
   Hoi Ning (Elaine) Ng  

54. **Representing sounds and spellings Phonological decline and compensatory working memory in acquired hearing impairment**  
   Elisabet Classon  

55. **Assessment of participation in people with a mild intellectual disability**  
   Patrik Arvidsson  

56. **Barnperspektiv i barnavårdsutredningar – med barns hälsa och barns upplevelser i fokus**  
   Elin Hultman  

57. **Internet Interventions for Hearing Loss Examining rehabilitation Self-report measures and Internet use in hearing-aid users**  
   Elisabet Sundewall Thorén  

58. **Exploring Cognitive Spare Capacity: Executive Processing of Degraded Speech**  
   Sushmit Mishra  

59. **Supported employment i en svensk kontext – förutsättningar när personer med funktionsnedsättning når, får och behåller ett arbete**  
   Johanna Gustafsson  

60. **Effects of Specific Cochlear Pathologies on the Auditory Functions: Modelling, Simulations and Clinical Implications**  
   Amin Saremi  
61. **Children with profound intellectual and multiple disabilities and their participation in family activities**  
   Anna Karin Axelsson  

62. **Lexical and Semantic Development in Children With Cochlear Implants**  
   Ulrika Löfkvist  

63. **Rethinking sound. Computer-assisted reading intervention with a phonics approach for deaf and hard of hearing children using cochlear implants or hearing aids**  
   Cecilia Nakeva von Mentzer  

64. **Assessing cognitive spare capacity as a measure of listening effort using the Auditory Inference Span Test**  
   Niklas Rönnberg  

65. **Employees with Aided Hearing Impairment: An Interdisciplinary Perspective**  
   Håkan Hua  

66. **Prosthetic and Orthotic Services in Developing Countries**  
   Lina Magnusson  

67. **Dealing with digits - Arithmetic, memory and phonology in deaf signers**  
   Josefine Andin  