Multiple perspectives on the association between cognition and speech-in-noise perception performance

By Adam Dryden

MRC Institute of Hearing Research & the School of Psychology,

PhD thesis

June 2018

Abstract

This thesis investigated the role of cognition and hearing sensitivity in Speech-in-Noise (SiN) perception across different listener groups and SiN listening conditions.

A typical approach to investigating the contribution of cognition is correlating cognitive ability to SiN intelligibility in populations controlled for or varied in age and/or hearing sensitivity. However, using this approach to advance our understanding of the contribution of cognition, and its potential interaction with age and hearing loss, for SiN perception has been limited by a combination of: A lack of systematicity in selection of SiN perception tests and a lack of theoretical rigor in selection of cognitive tests, a lack of comparability across studies due to differences in both cognitive test and SiN perception test selections, and in differences in age or hearing sensitivity ranges among tested populations, and the limitations of using a correlation study approach. Therefore, the main focus of the thesis will be to generate evidence to overcome these limitations in three purpose-designed investigations, discussed in chapters two, three and four respectively.

In chapter two I report a systematic review and meta-analyses, which took a systematic and theory driven approach to comprehensively and quantitatively assess published evidence for the role of cognition in SiN perception. The results of this chapter suggest a general association of $r \approx .3$ between cognitive performance and SiN perception, although some variability in association appeared to exist depending on cognitive domain and SiN target or masker assessed.

In chapter three I present a study, which used a theory-driven and systematic approach to investigate the contribution of cognition and listener characteristics (namely age and hearing sensitivity differences across younger and older listener groups) for SiN perception in different SiN conditions, using an association study design. The study revealed that the Central Executive contributed to SiN perception performance in older, but not younger listeners, regardless of SiN condition. Phonological Loop processing was important for both listener groups, but with a different role depending on age group and masker type. Episodic Buffer ability only contributed to SiN performance for older listeners, and was modulated by hearing sensitivity and background masker.

In chapter four, building on the association study findings, I report a dual-task study that manipulated the availability of specific cognitive abilities for SiN perception for younger adult listeners. Here I provided further evidence to show Phonological Loop ability is more important than Central Executive ability and Episodic Buffer ability for SiN perception for this listener group, using a carefully controlled experimental design.

In summary, the evidence from this thesis indicates that the role of different cognitive abilities for SiN perception can differ depending on age, hearing sensitivity and listening condition. Additionally, using a systematic approach and combining multiple methodological techniques has been informative in investigating these roles to a greater extent than has previously been achieved in the literature.

Copyright permissions

Publications

Chapter two is adapted from a peer review publication

Dryden, A., Allen, H. A., Henshaw, H., Heinrich. A. The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic review and meta-analysis. Trends in Hearing, 21: 1-21

Selected conference presentations

Data from chapter three were presented at the following selected conferences:

Dryden A., Allen H. A., Henshaw H., & Heinrich A. (2017). The contribution of cognition and hearing loss to individual differences in speech intelligibility in a variety of speech-perception-in-noise tests in adult listeners. British Society of Audiology Basic Science Meeting, Nottingham, UK

Dryden A., Allen H. A., Henshaw H., & Heinrich A. The contribution of cognition in speech-in-noise perception in normal hearing listeners (2016). Speech perception in Noise 8th Speech in Noise Workshop, Groningen, Netherlands

Dryden A., Allen H. A., Henshaw H., & Heinrich A. (2016). The contribution of cognition in speech-in-noise perception in younger and older normal hearing listeners. Society for Neuroscience 46th International Conference, San Diego, CA, USA

Data from chapter four were presented at the following selected conferences:

Dryden A, Allen H. A., & Heinrich A. (2017). Exploring the role of Working Memory for speech perception in energetic and informational masking: a dual-task study. Speech perception in Noise 9th Speech in Noise Workshop, Oldenburg, Germany

Dryden A., Allen H. A., Henshaw H., & Heinrich A. (2016). A different role for Working Memory in energetic versus informational masking of speech: a dual-task study. British Society of Audiology Basic Science Meeting, Cambridge, UK

Acknowledgements

Firstly, I would like to thank my excellent and dedicated supervisory team, Drs. Antje Heinrich, Harriet Allen, and Helen Henshaw, for their highly valuable, insightful, and constructive input, guidance and support throughout this project – they have been a true pleasure to work with.

Secondly, I would like to thank former IHR post-doc, Dr. Sarah Knight, for sharing her wealth of expertise with regards to various aspects of the project including, but not limited to, speech perception and cognitive theory, programming, and statistics.

I would also like to thank IHR statistician, Oliver Zobay, for his valued input in all things statistical, but specifically with respect to factor analysis and linear modelling methods, and using the statistical software program R.

Additionally, I would like to thank all my fellow PhD students and other colleagues at the IHR, and across various other university departments, for their general help and support in and around all the other trappings of PhD life.

Furthermore, I would like to thank Liz for her general words of wisdom and for being an invaluable and often much needed lift to the work environment at the IHR.

Finally, outside my PhD bubble, I would also like to thank my girlfriend, Clare, and all my family and friends back home in Newcastle (and in various other parts of the country), for all their unconditional support and for helping me keep perspective on life outside of work, and for keeping me on the right track!

Thank you all.

Table of Contents

Table of figures

Chapter two

Chapter three

Chapter four

Chapter five

Supplementary material

Abbreviations

3B	3-talker Babble
AIC	Akaike Information Criterion
ASL	Adaptive Sentence List
BESST	British English Semantic Sentence Test
BSA	British Society of Audiology
CE	Central Executive
CI	Confidence Interval
CSB	Corsi Span Backward
CSF	Corsi Span Forward
DSB	Digit Span Backward
DSF	Digit Span Forward
DTC	Dual-Task Cost
EB	Episodic Buffer
EFA	Exploratory Factor Analysis
ELU	Ease of Language Understanding
FA	Factor Analysis
HL	Hearing Loss
HP	High Predictability
IQ	Intelligence Quotient
LNS	Letter-Number Sequencing
LP	Low Predictability
NH	Normal Hearing
PCA	Principal Components Analysis
pDTC	proportional Dual-Task Cost
PL	Phonological Loop
PTA	Pure-Tone Average
R1	Rhyme verification task 1
R4	Rhyme verification task 4
RAU	Rationalized Arcsine Units
RST	Reading Span Test
SD	Standard Deviation
SE	Standard Error (of the mean)
SiN	Speech-in-Noise
SMN	Speech-Modulated Noise
SPIN-R	Speech Perception In Noise - Revised
SRT	Speech Reception Threshold
TEA	Test of Everyday Attention
VSS	Visuo-Spatial Sketchpad
WM	Working Memory

Chapter one – General introduction

1.1 Background

Following speech conversation can be a challenging process, especially in situations where there are background noises and/or competing talkers. A decline in the ability to perceive speech in noisy environments is a common complaint in older adults and those with Hearing Loss (HL) (Lin et al., 2011; Tun et al., 2002). A decline in the ability to perceive spoken language and thereby engage in everyday communication and related activities can have a profound adverse effect on health and quality of life (Ciorba et al., 2012; Dalton et al., 2003). In an aging population (Randell, 2017) with an ever-increasing prevalence of HL (Wilson et al., 2017) this is a growing issue and a significant burden of disease.

In the clinic, as well as in translational and basic science research, hearing sensitivity/HL is assessed using Pure-Tone Audiometry (PTA). In addition to this perception of spoken language can be assessed using speech perception tests. Speech perception tests involve listening to pre-recorded spoken language presented either over headphones or in free-field, and in the presence or absence of visual speech information. Speech perception can also be assessed with and without the presence of background noise, the former is known as Speech-in-Noise (SiN) perception. SiN perception tests vary in multiple aspects, I will outline some of these aspects in the following paragraph.

Firstly, SiN perception tests can vary in terms of the target speech, which can range in complexity from phonemes, single words, and whole sentences to short stories. Secondly, SiN perception tests can vary in the type of background noise used. Background noises can vary in terms of energetic and informational properties. Energetic masking refers to a masking signal that physically obscures a target signal and where the interference to the target is due to the physical overlap with the background signal (Kidd et al., 1994). Informational masking on the other hand refers to a masking signal that contains intelligible sounds, such as words and phonemes, and where the interference to the target is due to the distracting quality of the masker (Pollack, 1975). As such, maskers can be categorised on a continuum of these properties, ranging from static noise (energetic properties only), and speech-modulated noise (energetic and some speech properties), to multi-talker babble noise (energetic and informational), and a single competing talker (less energetic and more informational). Thirdly, SiN perception tests can also vary in terms of assessing perception at specific Signal-to-Noise Ratios (SNRs) where intelligibility level is the outcome measure. Alternatively, they can involve an adaptive staircase procedure where accuracy thresholds can be assessed at specific levels (by varying intensity of the target and/or the masker) and the SNR at the targeted threshold is the outcome measure. A fixed SNR procedure can benefit over an adaptive procedure because it is closer in simulating real-world listening, i.e., everyone is exposed to the same presentation levels of target and masker – akin to people following the same conversation in a noisy room, and it creates multiple data points, as opposed to a single data point in an adaptive procedure. However, an adaptive procedure can be preferred to a fixed procedure because a measure of perception can be quickly assessed at a desired intelligibility level.

SiN perception is of specific research interest because not only is it a common compliant in older people and those with HL, but it also can be an issue for younger people and for people without HL, as assessed using PTA (Plack et al., 2014). More specifically, hearing sensitivity does not account for all individual differences observed in SiN perception (Gordon-Salant et al., 1997; Parbery-Clark et al., 2009; Pichora-Fuller et al., 1995). Along with factors such as age and HL, cognition has emerged as another important factor associated with SiN intelligibility (Akeroyd, 2008; Arlinger et al., 2009; Pichora-Fuller et al., 1995).

1.2 Cognition and SiN perception

Cognitive abilities such as Working Memory, Inhibitory Control and Episodic Memory are theorised to be important for SiN perception (Akeroyd, 2008; Goldinger, 1996; Janse, 2012; Rönnberg et al., 2008). Working Memory, a limited-capacity process for simultaneous storage and manipulation of information to perform complex tasks (Baddeley, 2000; Daneman et al., 1980; Engle et al., 2004), is proposed to play a role in SiN perception through its restoration of the degraded target signal and inhibition of an interfering signal and this is central to the Ease of Language Understanding (ELU) model of language perception and cognition (Rönnberg et al., 2013; Rönnberg et al., 2008). The ELU model assumes that there is a short-term memory buffer (RAMBPHO), which holds speech information, while the information is matched for representations in long term memory. In favourable conditions the speech matching process is fast and implicit. However, adverse conditions, such as SiN listening, can disrupt or interfere with this matching process. In these circumstances explicit Working Memory processes (both storage and manipulation) are utilised in order for speech to be perceived. In a study of younger and older adults with normal hearing Besser et al. (2012) examined the association between Working Memory and SiN perception. They assessed two SiN conditions, recall of sentences in stationary and speech modulated noise; Working Memory was assessed using the Reading Span Test (Andersson et al., 2001; Daneman et al., 1980) (See Supplementary Figure 2.1 for a comprehensive list cognitive test descriptions). They found that better Working Memory ability was associated with better SiN intelligibility in modulated noise, but not in the stationary noise. The difference in masker conditions suggests that the additional signal interference caused by the modulated versus stationary noise may have further engaged Working Memory processing. Furthermore, when controlling for age they found the correlation coefficients declined for both conditions and were no longer significant in the case of modulated noise. This result suggests that age-related decline may drive the association between Working Memory ability and SiN perception for normal hearing listeners.

Inhibitory Control, the process by which a strong interfering factor is overcome to focus attention on a specific target or task (Diamond, 2013; Hasher et al., 1979), is thought to play in role in SiN perception because it requires focusing on a desired target in the presence of a distracting masker, particularly if the masker is intelligible and/or highly similar to the target (Janse, 2012). Veneman et al. (2013) investigated the role of Inhibitory Control for younger and older adult listeners with normal hearing. They assessed Inhibitory Control using a Visual Distracter Test (May, 1999) and SiN perception was assessed in two listening conditions, sentences in speech spectrum noise and 4-talker babble. They found that Inhibitory Control ability was significantly correlated with SiN intelligibility in both listening conditions (with better Inhibitory Control relating to better SiN intelligibility). This result suggests that Inhibitory Control may be generally important for SiN perception. However, it is not clear from this study if the association was driven by the older listener group. In another study, Stenbäck et al. (2015) examined the role of Inhibitory Control for SiN perception in younger normal hearing listeners. They assessed Inhibitory Control using the Hayling Task (Burgess et al., 1996) and SiN perception was assessed using sentences in speech-modulated noise at two intelligibility levels corresponding to 50% and 80% correct thresholds. They found better Inhibitory Control ability was associated with better SiN intelligibility in both listening conditions. Together these results indicate that better Inhibitory Control ability may generally be important for SiN perception for both younger and older listeners.

Episodic Memory, the process by which information is encoded into distinct episodes for later recall (Tulving, 1972), is thought to be important for SiN perception because it requires holding and integrating continuous and complex speech target signals for later recall (Goldinger, 1996; Rönnberg et al., 2008). A role of Episodic Memory for SiN perception has been shown for normal hearing older listeners. Meister, Schreitmuller, Grugel, Beutner, et al. (2013) found that higher intelligibility in three SiN conditions (sentences in unmodulated, modulated and babble noise) was association with better Episodic Memory ability (assessed using the Verbal Learning Memory Test (Helmstaedter et al., 1990)). Additionally, in a study investigating the role of Episodic Memory in younger and older adult listeners with normal hearing, Füllgrabe et al. (2015) found an association between Episodic Memory ability (assessed using the Digit Span test) and SiN intelligibility (sentences-in-2-talker babble) in the older adult group and in both listener groups when controlling for age. Overall these results suggest the Episodic Memory may be general important for SiN perception and that another factor other than cognitive decline may account for age-related differences.

1.3 Cognition, age, hearing sensitivity and SiN perception

Although it is widely acknowledged that cognition plays a role in SiN perception (Akeroyd, 2008; Dryden et al., 2017; Rönnberg et al., 2013), its underlying nature and its complex relation to age and hearing sensitivity is not yet fully understood. Previous studies investigating the relationship between cognition and SiN perception using a correlation approach have found seemingly inconsistent results. For example, Working Memory ability has been found to be significantly associated with SiN intelligibility in some studies (Anderson et al., 2013; Gordon-Salant et al., 2016; Rönnberg et al., 2014; Surprenant, 2007), while others showed mixed results (Füllgrabe et al., 2015; Koelewijn et al., 2012; Zekveld et al., 2014) or no significant associations at all (Carroll et al., 2016; Cervera et al., 2009; DiDonato et al., 2015). Similar inconsistencies have also been found for Inhibitory Control, some studies found significant associations (Janse, 2012; Veneman et al., 2013), others have found mixed results (Ellis et al., 2014; Stenbäck et al., 2015), and some have found no associations (Heinrich & Knight, 2016; Helfer et al., 2014), Furthermore, inconstancies have also been reported with regards to the role for Episodic Memory for SiN perception. As with the other cognitive domains some studies found an association (Meister, Schreitmuller, Grugel, Beutner, et al., 2013), others showed mixed results (Helfer et al., 2014; Uslar et al., 2013), while others reported no significant findings (Cervera et al., 2009; Heinrich et al., 2015; Heinrich, Henshaw, et al., 2016).

These apparent inconsistencies in results between studies may be explained by and related to differences in listener factors such as age and hearing sensitivity. The role of age in SiN perception is of particular interest because it is associated with declines in perception (including auditory perception) and cognitive factors. In a pair of recent reviews Roberts et al. (2016) (assessing age-related declines from an auditory and a visual perspective) and Wayne et al. (2015) (assessing age-related declines from an auditory perspective) summarise four established hypotheses, which are not mutually exclusive, for the association between cognition, SiN perception and age-related decline: *common cause*, *cognitive load on perception, information degradation*, and *sensory deprivation*.

Common cause hypothesis (CHABA, 1988; Lindenberger et al., 1994) proposes that a third common factor, e.g., cerebrovascular risk factors, can cause declines in both perception (inc. SiN perception) and cognitive processes. This process could act on central processes involved for both cognitive and perceptual tasks, but also could act on specific perceptual processes or pathways including auditory, but also visual and motor systems. This hypothesis should be carefully considered and controlled for in participant recruitment to account for comorbidities, including cognitive and sensory declines. *Cognitive load on perception hypothesis* (CHABA, 1988; Lindenberger et al., 1994) proposes that poor cognition leads to poor performance in perception (e.g. SiN) tasks. However, this hypothesis may be contested in that declines can be disproportionate between sensory systems, suggesting central cognitive processes may not be the main driving force. *Information degradation hypothesis* (CHABA, 1988; Pichora-Fuller, 2003a; Schneider et al., 2000) proposes that impoverished perception, e.g., degradation of auditory signal due to peripheral declines, impacts on performance for cognitive tasks. *Sensory deprivation hypothesis* (Baltes et al., 1997; CHABA, 1988; Lindenberger et al., 1994) proposes that over time, impoverished perception can lead to cognitive decline – this hypothesis highlights the potential importance of early detection and intervention in treating perceptual decline.

Several hypotheses highlight the role of compensatory cognitive mechanisms, which can aid when a sensory signal becomes impoverished due to peripheral decline, this is specifically relevant to *information degradation* and *sensory deprivation* hypotheses where decline is driven by perception over cognition. If such a mechanism occurs the association between cognitive ability and SiN perception would be expected to be greater in the presence of HL. However, this is a simplified view in that, firstly, there might be a specificity in which cognitive abilities are involved in the compensatory process, secondly, HL as assessed by an audiogram is not sensitive to other periphery factors important for SiN perception, e.g., sensitivity to temporal fine structure, and thirdly, all the above processes may be differently involved depending on specific SiN listening conditions.

Given the complexity in the relationship between these factors it is perhaps unsurprising that previous studies investigating the role of age, hearing sensitivity and cognition for SiN perception have shown a lack of consensus. For example, with regards to Working Memory some studies have linked Working Memory ability to SiN perception for older listeners (Anderson et al., 2013; Heinrich & Knight, 2016; Parbery-Clark et al., 2011), while others have found mixed results (Heinrich et al., 2015) or no relation (Helfer et al., 2014). However, there is emerging evidence that the role of Working Memory for SiN perception does appear to be stronger for older compared to younger listeners (Füllgrabe et al., 2016b), but these differences may also relate to age-related hearing loss or other differences in other auditory factors such as temporal fine structure (Füllgrabe et al., 2015).

Furthermore, the underlying relationship between age, hearing sensitivity and cognition (Jerger, 1992) and how they relate to specific SiN listening conditions (Moore et al., 2014) is still not clear. The specificity of these roles for specific SiN listening conditions is of importance because cognitive processes may be engaged differently depending on the properties of both the speech target and background noise. For example the perception of complex target signals such as sentences, compared to phonemes, may further engage processes such as Working Memory (Heinrich & Knight, 2016) and Episodic Memory (Anderson et al., 2013), perhaps because it requires the listener to hold a speech trace in mind for a greater amount of time in order for the trace to be processed and integrated with previously heard or retrieved information (Goldinger, 1996; Rönnberg et al., 2008). Furthermore, speech perception in informational versus energetic masking conditions may evoke different cognitive processes and to different extends. This may involve processes such as Working Memory (Koelewijn et al., 2012; Rönnberg et al., 2014), attention and Inhibitory Control (Mattys et al., 2009), which are important in separating target and distractor signals (Freyman et al., 2004). Therefore, another factor driving the inconsistencies may be differences in the selection of SiN perception tests between studies, which vary in target and masker combinations.

Additionally, although an association/correlational study approach has been beneficial in improving our understanding of the processes involved for SiN perception it has limitations in being unable to determine causal relationships between factors. Therefore other behavioural approaches are required in tandem with association studies in order to provide new insights. One such approach in the SiN literature are dual-task paradigms.

1.4 Dual-task approaches in investigating the role of cognition for SiN perception

Dual-task paradigms are underpinned by the underlying assumption that cognitive resources are limited in capacity. Classical theories of dual-task interference include the *bottleneck theory* (Broadbent, 1958; Cherry, 1953) and the *shared resource theory* (Kahneman, 1973). Both theories assume that resources are limited. However, the *bottleneck theory* proposes that interference occurs due to an attentional bottleneck where only one task can pass through an attentional filter at a time, whereas the *shared capacity theory* proposes that dual-task interference occurs because additional resources are required to perform two tasks at once. Other theories, such as the Baddeley model of Working Memory (Baddeley, 2000), place more emphasis on memory in dual-task performance. The Baddeley model consists of a Central Executive component, responsible for the task-driven focus of attention, and slave components, responsible for short-term storage. The slave components consist of a Phonological Loop, responsible for storage of verbal information (including speech), a Visuo-Spatial Sketchpad, responsible for storage of visuo-spatial information, and an Episodic Buffer, responsible for storage and binding of cross-modal information into a unitary episodic representation. During dual-task performance, the Central Executive can prioritise attention to a specific task whilst ignoring another. Interference can arise at the level of storage if the tasks require the same storage resources.

Typically, dual-tasks involve two concurrently presented tasks with a primary task, usually of specific research interest, and a secondary task, which engages shared cognitive resources used for the primary task. Performance can be measured in both tasks under dual-task conditions and measured individually in single-task performance. This allows a dual-task cost to be calculated for performance in either or both the primary and secondary tasks, depending on how the task is designed. If priority is directed to either task or divided equally between tasks, then performance will decline in the lower priority task or in both tasks respectively. In SiN perception dual-task studies performance in the SiN perception test typically remains similar under dual-task conditions relative to performance and reaction times in single-task conditions (*see Gagne et al. (2017) and Ohlenforst et al. (2017)*). It is performance, or reaction time, in the secondary cognitive task that typically declines (or increases in the case of reaction times) in dual-task relative to single-task conditions. It is this determinant that is used to assess dual-task cost.

Previous dual-task studies investigating the role of cognition for SiN perception have shown greater dual-task costs in the presence of masking versus no masking (Pals et al., 2015), and greater dual-task costs at lower versus higher SNRs (Sarampalis et al., 2009). These results agree with ELU model of speech perception, which predicts a greater involvement of cognition when a speech signal is degraded, for example in the presence of background noise, a role that may increase as the amount of degradation increases, for example at lower SNRs (Rönnberg et al., 2013; Rönnberg et al., 2008).

However, the role of cognition for different SiN listening conditions is less clear. For example, with regards to background masker, result have been mixed in determining if cognitive processes differ between maskers with more versus less informational properties (e.g. babble versus speech-shaped noise). In a study in younger normal hearing listeners Pals et al. (2015) investigated dual-task performance for the perception of sentences in three masking conditions (steady state noise, speech-shaped noise, and 8-talker babble) during a secondary rhyme judgement task. They found that reaction times in the secondary task did not differ depending on masking condition. Additionally, they assessed Working Memory ability and examined the correlations between Working Memory ability and dual-task performance in the secondary task. They found that Working Memory ability was not associated with SiN perception for any of the masking conditions.

In another study, Desjardins et al. (2013), investigated dual-task performance in a range of different listener groups, younger adults with normal hearing, older adults with normal hearing, and older adults with HL. In the study the SiN perception test comprised of six different listening conditions: low and high predictability sentences presented in either 2-talker babble, 6-talker babble, or speech-shaped noise. The secondary task was a visual-motor tracking task in which participants were required to track a moving target presented on a computer screen. Dual-task cost was assessed in the secondary task performance. The results showed that there was no effect of target predictability for any of the listener groups. However, there was an effect of masker. In the younger, but not the older groups, a greater dual-task cost was found for the 6-talker babble condition compared to speech-modulated noise.

Additionally, dual-task costs were found to be greater for both the older groups compared to the younger group in 2-talker babble and speech-modulated noise conditions. This result suggests that the younger group recruited additional cognitive resources in the more informational masking conditions, while the older listeners recruited additional cognitive resources in both informational and energetic masking. Working Memory ability was also assessed for all listener groups and it was found to be associated with dual-task costs in the 2-talker babble and speech-modulated noise, but not the 6-talker babble. This suggests that older adults and listeners with hearing loss may be recruiting additional Working Memory resources compared to younger listeners. Additionally, processes other than Working Memory, e.g., Inhibitory Control, may be more important in the case of the 6-talker babble masker.

There are some key gaps in the current literature, which warrant further investigation. Firstly, studies often use a simple and non-verbal secondary cognitive test, e.g. visuo-motor or tactile response tasks (Desjardins, 2016; Desjardins et al., 2013; Fraser et al., 2010; Gosselin et al., 2011), which do not specifically engage the verbal aspect of Working Memory, and perhaps do not sufficiently engage the Central Executive component, both of which are likely to be important for SiN perception. Secondly, few SiN-cognition dual-task studies have examined differences depending on SiN listening condition or assessed multiple secondary tasks, with limited exceptions (Bockstael et al., 2018; Gagne et al., 2017; Picou et al., 2014). Additionally, studies often take an association approach by examining correlation coefficients between dual-task costs and performance in cognitive tests (Desjardins, 2016; Desjardins et al., 2013; Tun et al., 1991) rather than manipulated the availability of hypothesised key cognitive components. Therefore, it may be of greater benefit to select (multiple) secondary tasks, which engage specific cognitive abilities to different extents, and to select SiN perception tests, which vary in properties such as speech target and/or background signals. This would allow the engagement of specific cognitive abilities for specific SiN listening conditions and/or listener groups to be experimentally manipulated. This approach would also remove the need for correlational analysis between dual-task performance and cognitive ability.

1.5 Thesis overview

Overall it is difficult to determine the complex relationships between cognition, SiN perception, hearing sensitivity, and age. One key step is to systematise the selections of both cognitive and SiN perception tests and to be as theory-guided as possible. In this thesis I tease apart some of these possible sources of inconsistency seen in the literature. I achieve this by investigating the roles of specific cognitive abilities for SiN perception in different listening conditions for younger and older listeners with a range of hearing sensitivities (normal hearing and mild HL). By taking this approach I further our understanding of the underlying cognitive processes for SiN perception and how these processes may vary depending on listening condition and listener-specific factors such as age and hearing sensitivity.

In the next three sections I provide a detailed overview of each chapter in turn.

1.5.1 Chapter two overview

The purpose of the systematic review and meta-analysis was to take a systematic and theory-driven approach to comprehensively and quantitatively assess the role of cognition in SiN perception in the published literature. Previous literature reviews have either focused on a specific cognitive ability (Füllgrabe et al., 2016b) or used a qualitative approach (Akeroyd, 2008). Here I explored a full range of cognitive abilities and made quantitative comparisons wherever possible.

As a first step I categorised all reported cognitive tests into nine cognitive sub-domains (Alerting, Orienting, Set-Shifting, Inhibitory Control, Working Memory, Episodic Memory, Fluid Intelligence, and Crystallised Intelligence) and assumed that (the degree to which they assess the same underlying concept) they will show a similar relationship to SiN perception. The cognitive sub-domains were based on multiple established cognitive theories (Baddeley, 2000; Diamond, 2013; Miyake et al., 2000; Petersen et al., 2012; Salthouse, 2000). I also systematised SiN listening conditions along two dimensions, target and masker signals to gain a better understanding of how the relationship of different cognitive sub-domains might vary depending on listening condition. Finally, I took into account hearing sensitivity by categorising participants in each study into one of two groups, shaped by the data available: normal hearing to mild hearing loss, and normal hearing to moderate HL.

This exploratory, yet systematic approach allowed me to answer several specific questions:

1) What is the overall association between cognition and SiN perception?

2) Which specific cognitive abilities are associated with SiN perception?

3) How does hearing sensitivity relate to differences previously found in various associations between cognitive ability and SiN perception?

4) Do cognitive ability-SiN perception associations differ depending on SiN listening conditions (target/masker type)?

1.5.2 Chapter three overview

This chapter reports a theory driven and systematic evaluation of the role of cognition and hearing sensitivity in SiN perception in different listening conditions using an association study design. I selected 1) a battery of cognitive tests to assess cognitive abilities based on established cognitive theory, 2) a range of different SiN listening conditions varying in linguistic complexity of the target signal (words, and low and high predictability sentences) and energetic to informational masker properties and 3) two listeners groups: younger and older adults.

To assess cognitive abilities, it was important to be as theory-guided as possible. Working Memory, the process of simultaneous storage and manipulation of information, is regarded as being of general importance for SiN perception (Rönnberg et al., 2013). However, Working Memory is not a unitary process, but instead involves multiple cognitive abilities. A number of models exist that specify the involved cognitive abilities and their relationships to each other (Baddeley, 2000; Broadbent, 1958; Craik et al., 1975; Diamond, 2013; Engle et al., 2004; Miyake et al., 2000; Petersen et al., 2012; Treisman, 1964). Out of these models I chose two as basis for my own investigations, Baddeley (2000) and Diamond (2013).

I chose the Baddeley model of Working Memory (Baddeley, 2000) because it is well established in the cognitive literature, differentiates between verbal and non-verbal storage and Executive Functions, it is intuitive to understand and easy to implement. However, a drawback of being intuitive can be that some model components are less differentiated than might be useful for a particular investigation. The Baddeley model differentiates between slave systems (Visuo-Spatial Sketchpad, Episodic Buffer, and Phonological Loop) and a single domain, Central Executive, for Attention and Executive Functions. Given the potential importance of those Central Executive Functions for SiN perception, I looked for a second cognitive model that specified these Executive Functions in further detail. The model I selected was the Diamond’s model of Executive Functions (Diamond, 2013). The model assesses Executive Functions separately as Inhibitory Control and Working Memory.

Another goal of the study was to move away from the practice of assessing cognitive ability by a single test because it conflates variability due to test-specific and ability-specific aspects ("Impurity Principle") (Surprenant et al., 2009). Therefore, multiple cognitive tests were selected to assess each of the theorised cognitive domains.

Besides being theory-guided in the choice of cognitive tests I also wanted to gain a clearer picture as to if and how the contribution of cognition differs for different listening conditions, age and hearing sensitivity. Therefore, I selected a range of SiN listening conditions. The range represented a sub-group of the range identified in the systematic review to address known gaps in the research literature. More specifically, three different target signals (varying in linguistic complexity from single words, to low and high predictability sentences) and two masking conditions (varying in informational properties) were selected.

Finally, the inclusion of different age groups (younger and older adults) and a range hearing sensitivities (younger and older adults: normal hearing; older adults: mild HL) allowed me to explore both age and hearing sensitivity-based factors in relation to cognitive ability across SiN listening conditions.

In employing this systematic and theory-driven approach, I was able to explore the following research questions:

1) Which cognitive abilities are important for SiN perception?
2) Does the role of specific cognitive abilities differ depending on SiN listening condition?
3) Do these roles differ depending on age group and hearing sensitivity?

1.5.3 Chapter four overview

Based on the results of the association study in chapter three, I devised a dual-task experiment to manipulate the availability of specific cognitive abilities for SiN perception. If cognitive abilities were as important as suggested by the results of the association experiment, then their relative unavailability should adversely affect SiN perception. Here I focus on a younger adult listener group with normal hearing to rule out any confounding effects of age and hearing sensitivity.

The Baddeley model (Baddeley, 2000) was selected as a theoretical underpinning for the selection of secondary tasks. The model was selected due to accounting for storage, differentiating between verbal and non-verbal modalities, whilst also considering executive processes. Furthermore, selecting tests on the basis of the same cognitive model will preserve continuity between this and the association study. I selected four different secondary memory tasks that engaged the four sub-domains of the Baddeley model (Central Executive, Visuo-Spatial Sketchpad, Episodic Buffer, and Phonological Loop) to different extents.

This design allowed me to investigate the following research questions:

1) Are the cognitive processes involved in SiN perception modality (verbal versus non-verbal) specific?

2) Do storage and manipulation processes differ in their role in SiN perception?

3) Do these roles differ depending on the informational properties of the masker?

1.6 Summary

Through a variety of complimentary methods, and a theory-guided and systematic approach I elucidate the role of various cognitive processes in SiN perception across different listening conditions, and across differences in both age and hearing sensitivity of the listener.

Chapter two – A systematic review and meta-analysis: the association between cognitive performance and speech-in-noise perception for adult listeners

Abstract

In this chapter, I investigated the speech-in-noise and cognition literature by conducting a systematic review and meta-analysis.

Published studies assessing the association between cognitive performance and speech-in-noise perception examine different aspects of each, test different listeners, and often report quite variable associations. By examining the published evidence base using a systematic approach, I aimed to identify robust patterns across studies and highlight any remaining gaps in knowledge. I limited my assessment to adult non-hearing aid users with audiometric profiles ranging from normal hearing to moderate hearing loss.

A total of 253 articles were assessed, with 25 meeting the criteria for inclusion. Included articles assessed cognitive measures of attention, memory, Executive Function, intelligence (IQ) and Processing Speed. Speech-in-noise perception tests varied by target (phonemes/syllables, words, sentences) and masker type (unmodulated noise, modulated noise, multi (n>2) talker babble, and n≤2 talker babble).

The overall association between cognitive performance and speech-in-noise perception was r=.31. For component cognitive domains, the association with (pooled) speech-in-noise perception were; Processing Speed (r=.39), Inhibitory Control (r=.34), Working Memory (r=.28), Episodic Memory (r=.26) and Crystallised IQ (r=.18). Similar associations were shown for the different speech target and masker types.

The results of this chapter suggest a general association of r≈.3 between cognitive performance and speech perception, although some variability in association appeared to exist depending on cognitive domain and speech-in-noise target or masker assessed. Where assessed, degree of unaided hearing loss did not play a moderating role. I also identified a number of cognitive performance and speech-in-noise perception combinations that have not been tested, and whose future investigation would enable further finer-grained analyses of these relationships.

Note: this chapter has been published, reference: Dryden et al. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic review and meta-analysis. Trends in Hearing, 21:1-22

2.1 Introduction

Following a conversation in a noisy environment is difficult, and the effort required increases with hearing impairment (Zekveld et al., 2011). Hearing Loss (HL) has been extensively investigated as a primary underlying factor for difficulties in speech perception under adverse listening conditions (Agus et al., 2009; Humes et al., 1990; Jerger et al., 1991; Smoorenburg, 1992).

While HL does explain some of the difficulties, it has also become clear that it cannot be the only driving factor given the following observations: first, listeners with similar auditory sensitivity can differ greatly in their Speech-in-Noise (SiN) performance (Anderson et al., 2011; Vermiglio et al., 2012); second, SiN difficulties can be found in the absence of HL (Gordon-Salant et al., 1993; Gosselin et al., 2011; Plack et al., 2014); and third, SiN listening difficulties can persist even when HL has been alleviated by hearing aids (Humes, 2002; Studebaker et al., 1999).

Another factor that has repeatedly been suggested to play a role in SiN perception is cognition (Roberts et al., 2016). While investigations of the association between cognitive performance and SiN perception have a long tradition (Pichora-Fuller et al., 1995; Rabbitt, 1968; Tun et al., 1999; van Rooij et al., 1990, 1992), interest and publications in the field have surged in the past 20 years, leading to the coining of "*cognitive hearing science*" as a term for the field (Arlinger et al., 2009; Rönnberg et al., 2010; Tun et al., 2012).

Despite increasing interest in the association between cognitive performance and SiN perception, the emerging picture is far from clear. Not only do measures of SiN perception and cognitive tasks vary greatly across published studies but also research participant samples vary widely and can include any combination of young and old listeners with or without hearing loss, tested under aided or unaided conditions.

One way of dealing with the great variability in the field is to use a descriptive approach when summarising results across studies. This strategy was adopted by Akeroyd (2008) in a review that explored the relationship between individual differences in cognition and SiN perception in normal and hearing-impaired adult listeners (including aided listeners) across twenty studies. He found inconsistencies between study results not only for cases where listening conditions and cognitive domains assessed varied across studies, but also for cases where the assessed cognitive domain, such as Working Memory, was constant and only the listening condition varied. Specifically, when surveying all published associations between Working Memory performance and any SiN perception tests, Akeroyd (2008) found that just over half of the associations (53/87) were statistically significant. He concluded that most of these significant associations were shown for studies using SiN perception tests with a sentence (compared to single words) as target speech signal, and modulated noise (compared to static noise) background masker.

In a more recent review and meta-analysis Füllgrabe et al. (2016b) focused on a single cognitive ability - Working Memory (as measured by the Reading Span test), and investigated its association with SiN listening in normal hearing adult listeners. Using a meta-analysis, they examined the association between the performance on the Reading Span test and SiN perception using tests with a sentence target presented in co-located background noise. Comparing 24 correlations from 16 studies they found an overall (non-significant) association of .12. As a result of their meta-analysis, the authors suggested that Working Memory contributes relatively little to individual differences in SiN perception in normally hearing younger adult (≤40 years of age) listeners.

The different findings of these two prior reviews may simply be due to differences in the populations studied. The association between Working Memory and SiN perception may not be as ubiquitous as sometimes assumed, but instead may vary substantially by age and/or hearing status of the listener. Alternatively, it is possible that the differences arose because Füllgrabe et al. (2016b) restricted their search to a single cognitive domain (Working Memory), assessed using one measure (Reading Span test).

In this chapter I explored both possibilities. First, I considered a range of hearing abilities (normal hearing to moderate hearing loss) in pre-clinical unaided listeners. Second, I extended the investigation to cognitive abilities other than Working Memory and included a range of measures for each cognitive ability. I systematised all cognitive measures used in the reviewed studies into cognitive domains and sub-domains based on well-established cognitive theories. I also systematised SiN perception tests based on the target speech signal and background masker type. These categorisations enabled me to investigate the specific associations between cognitive domain and SiN perception test and how this might contribute to the variability of previously found results.

In contrast to the previous reviews, I hoped that my systematic approach would enable me to identify similarities between published studies that use tests assessing the same cognitive domain and similar SiN perception tests and uncover differences between studies that assess different cognitive domains and/or SiN perception tests. I also aimed to highlight any gaps in the published literature by identifying under-studied combinations of SiN perception tests and cognitive domains that warrant further investigation

Here my specific research questions were:

1) What is the overall association between cognition and SiN perception?
2) Which specific cognitive abilities are associated with SiN perception?
3) How does hearing sensitivity relate to differences previously found in various associations between cognitive ability and SiN perception?
4) Do cognitive ability-SiN perception associations differ depending on SiN listening conditions (target/masker type)?

2.2 Methods

2.2.1 Categorising speech-in-noise tests

SiN perception tests can vary on foreground signal, background signal, type of response (open and closed set), signal-to-noise ratios/intelligibility levels, adaptive and non-adaptive paradigms, and signal presentation (headphones or free-field) to name but a few aspects. Each of these variations could impact on the manner and/or extent to which cognitive resources are required to perceive the speech message. As I cannot consider all aspects in this review, I will focus on the examination of the role that foreground and background signals might play for the association between cognition and SiN perception. By systematising SiN perception tests based on the foreground (target) and background (masker, i.e. the noise) signal, I investigated whether all SiN perception tests within the same category of fore- and/or background sound show a similar relationship with a particular cognitive measure.

I categorised the foreground target according to its lexical complexity from simplest to most complex into 1) phonemes and syllables, 2) words, and 3) sentences. I classified the target signal as the speech signal that the listener is instructed to respond to. This includes instances where for example a phoneme or word target is embedded in a more complex signal such as a sentence or a carrier phrase. When a participant is instructed to repeat a full sentence, but unbeknownst to them the response is scored only on the final word, this will be classified as a sentence target signal. This is because the task, not the scoring, defines the characteristics of the signal. There were no reported instances of participants' being aware of the scoring procedure for any SiN perception test in the included studies.

I chose lexical complexity as the basis for categorisation because it has been shown to be important for the manner and/or extent to which cognitive processes are engaged (Heinrich et al., 2015; Heinrich, Henshaw, et al., 2016; Heinrich & Knight, 2016; Xu et al., 2005). For example, when measuring correlations between cognition and SiN perception, Heinrich and Knight (2016) showed an increased association between the Reading Span test and the Letter-Number Substitution tests when comparing words and sentences, respectively, in a background of speech-modulated noise. Moreover, in a language comprehension fMRI study, Xu et al. (2005) mapped brain activation in single word and sentence comprehension. They found increased activation in regions including Broca's area, left middle temporal gyri, right posterior cerebellum, left putamen and ventral thalamus for sentence, compared to single word, comprehension, indicating a differing network of activation for these types of stimuli.

I conceptualised differences in the background signal by considering the extent to which the background engages energetic and informational masking. Placing background signals on a continuum between energetic and informational masking resulted in the following order of (decreasing) energetic and (increasing) informational masking: 1). unmodulated noise, 2) modulated noise, 3) multiple (>2) background talkers, and 4) a single or two-distractor voice(s). Background signals with one- and two-distractor voices were separated in this classification from multiple background voices for two reasons. First, Simpson and Cooke (2005) showed that the difference in intelligibility of foreground speech is particularly marked for one and two background talkers versus a higher number of talkers. Second, it has been suggested that increased intelligibility of background sounds (indicating increased informational masking) engages cognitive processes such Inhibitory Control and attention (Mattys et al., 2009) that help to disentangle the target signal from the masker (Freyman et al., 2004). Possibly, these processes are not engaged to the same extent by multiple background voices.

The matrix for the categorisation of the SiN perception tests used in the studies considered in this chapter is displayed in Figure 2. 1. Within these categories intelligibility levels, adaptive versus non-adaptive paradigms and signal presentation are not distinguished. I recognise this as a limitation of my categorisation system. However, due the vast heterogeneity in SiN perception tests in previous studies, some simplification was necessary, and I chose to investigate the role of foreground and background signal for this review while generalising over all other differences.

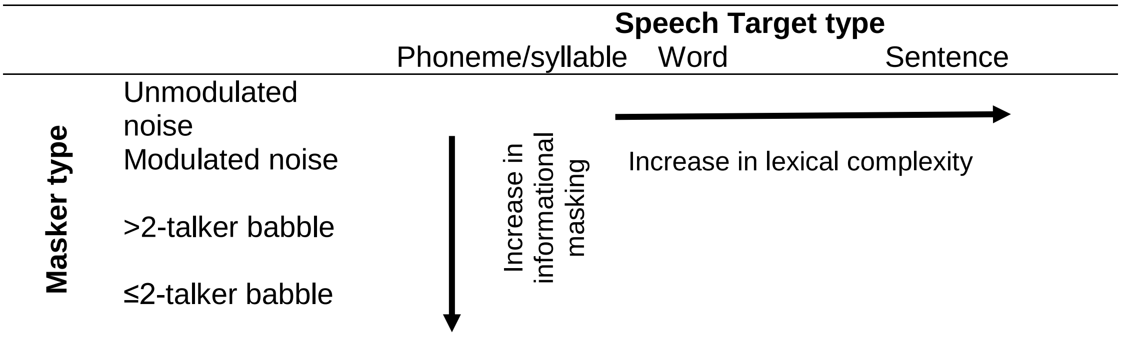


Figure 2. 1 – SiN perception test target/masker matrix

Speech-in-Noise test matrix displaying the categories for classifying speech target and masker type. >2-talker babble: speech babble consisting of more than two speakers; ≤2-talker babble: speech ‘babble’ containing two or only one distracter voice.

2.2.2 Categorising cognitive measures

Cognitive function associated with SiN perception has been assessed using a wide variety of measures. This can make the direct comparison between studies difficult. I addressed this issue by abstracting from a particular cognitive test to the tested cognitive domain and sub-domain being assessed. In total I distinguish five cognitive domains (attention, executive processes, memory, intelligence and Processing Speed) and nine cognitive sub-domains (Alerting, Orienting, Set-Shifting, Inhibitory Control, Working Memory, Episodic Memory, Fluid and Crystallised intelligence (IQ), and Processing Speed) based on contemporary cognitive theories (Baddeley, 2000; Diamond, 2013; Miyake et al., 2000; Petersen et al., 2012; Salthouse, 2000).

I define each domain and its constituting sub-domains below and briefly explain their proposed involvement in SiN perception. Although I recognise that an individual test can load on multiple cognitive domains (“Impurity Principle”) (Surprenant et al., 2009), for the purpose of this review, I categorised each test only according to the main sub-domain it is theorised to assess. I categorised cognitive performance at the level of sub-domain for two main reasons. Firstly, this level specificity allowed me to differentiate specific sub-domains of interest for SiN perception. For example, assessing Set-Shifting, Working Memory and Inhibitory Control as individual sub-domains of executive control may be of added value and interest compared to the consideration of a single executive processes domain. Secondly, by categorising cognitive performance at the level of sub-domain I hoped to reduce heterogeneity within each domain, i.e., they are complex tasks which assess multiple cognitive domains.

Supplementary Table 2. 1 in the Appendix provides a full list and description of all cognitive tests used in the reviewed studies, ordered by cognitive domain and sub-domain. Please note that a few tests, such as the text reception threshold (Zekveld et al., 2007), which is the theorised visual equivalent to the speech reception threshold test, are not included in this review because they are not readily definable within my single cognitive domain framework.

Attention

I conceptualised tests assessing attention within Posner and Petersen's (1990) framework, which considers three distinct but interconnected processes: 1) Alerting, 2) Orienting and, 3) executive control. Given the central role that executive control is assumed to play for SiN perception (Pichora-Fuller et al., 2016; Tamati et al., 2013; Zekveld et al., 2014) I considered the further sub-domains of executive processing separately from attention.

Alerting: Alerting is the ability to prepare and sustain attention to a high priority signal (Posner et al., 1990). It may be important for SiN perception because it allows listeners to focus on the speech target in an environment of other noise sources (Binder et al., 1994; Heald et al., 2014). It is possible that it plays a particularly important role for more complex target signals (such as whole sentences) because they require sustained attention for a longer period of time.

Orienting: Orienting refers to the ability to, overtly or covertly, prioritise sensory input from a particular spatial or temporal location or modality (Posner et al., 1990). It may be important for SiN perception, particularly in situations of spatial separation because it allows temporal and spatial preferential selection of a target signal (Astheimer et al., 2009; Calvert et al., 1998).

Executive processes

Executive processes control and coordinate performance of complex cognitive tasks. They are closely related to attention and are sometimes considered as one of its sub-domains (Posner & Petersen, 1990). Due to their potential importance for SiN perception I considered them as a separate domain and subdivided them further based on Miyake et al. (2000) into three sub-domains: 1) Set-Shifting, 2) Inhibitory Control, and 3) updating (synonymous with ‘Working Memory’).

Set-Shifting refers to the ability to switch between tasks, operations or mental sets (Miyake et al., 2000). Set-Shifting ability is thought to be closely related to representations of internal speech and task-specific organisation (Cragg et al., 2010). It might also be predicted that it is important when a listener has to shift from one speech target to another.

Inhibitory Control is a process by which a strong interfering factor is overcome in order to maintain focus on the desired target or task (Diamond, 2013; Hasher et al., 1979). Inhibitory Control has been suggested to play a role for SiN perception in several ways. First, poor inhibition may increase susceptibility to background noise during SiN perception, particularly in informational masking conditions (Janse, 2012); second, poor inhibition may make it harder for listeners to successfully select the target during lexical access (Sommers et al., 1999). Third, inhibition may have a general role in degraded signal restoration (Janse et al., 2014; Mattys et al., 2012).

Working Memory is a limited-capacity process by which we simultaneously store, process and manipulate information necessary to complete complex tasks (Daneman et al., 1980). Prominent Working Memory theories include the multi-component model proposed by Baddeley and Hitch (Baddeley, 2000; Baddeley et al., 1974) and the activation model by Engle et al. (2004). Both models propose a single amodal executive processing component required for a task-driven focus of attention. In addition, Baddeley (2000) also proposed amodal and modality-specific separate slave systems for information storage. The concept of Working Memory is very prominent in the SiN perception literature. It has been incorporated into a prominent framework on the involvement of cognition in speech perception, the Ease of Language Understanding (ELU) model (Rönnberg, 2003; Rönnberg et al., 2013; Rönnberg et al., 2008). The ELU model posits that Working Memory plays a role in the restoration of degraded speech signals and in the inhibition of masking signals (Rönnberg et al., 2013). However, whether Working Memory is equally important for all groups of listeners or only for those with a degraded input (e.g., listeners with hearing impairment) is a matter of considerable debate. For a task to be classed as Working Memory within this review it had to contain both a storage and a manipulation component. The type of information (verbal/non-verbal) and the modality of presentation (auditory/visual) are not differentiated here.

Memory

Memory is the faculty by which information is encoded, stored and retrieved (Atkinson & Shiffrin, 1968). There are many classifications of memory depending on the aspect of memory that is emphasised. Here I am particularly interested in **Episodic Memory**, which according to Tulving refers to the encoding of distinct episodes of information for later recall (Tulving, 1972). The distinguishing feature of Episodic Memory compared with Working Memory for the purpose of the current review is the presence (Working Memory) or absence (Episodic Memory) of a manipulation component. Episodic Memory has been hypothesised to be important for SiN perception because with longer speech signals a listener has to hold a speech trace in mind in order to integrate it with previously heard or retrieved information (Goldinger, 1996; Rönnberg et al., 2008).

Intelligence

General intelligence refers to the overall mental ability common to performance of all cognitive tasks (Spearman, 1904). Cattell (1963) differentiates between Fluid and Crystallised IQ.

Fluid Intelligence (IQ) refers to the general ability to solve problems and use abstract reasoning. It may be related to SiN perception through its link with Working Memory and executive control and may be particularly important in complex listening conditions such as dichotic listening (Engle, 2002; Meister, Schreitmuller, Grugel, Ortmann, et al., 2013). Fluid intelligence is typically assessed using non-verbal tasks.

Crystallised Intelligence (IQ) refers to language- and culture-specific knowledge and skills, which are acquired over time. It is thought to be important for SiN perception when the listening task requires increased reliance on lexical or general knowledge. Such situations may arise when the masker is informational or when target stimuli contain substantial contextual support (Schneider et al., 2016).

Processing Speed

Processing Speed is the rate at which information is processed in order to execute a task. It has been suggested to play a crucial role in explaining age-related changes in cognition (Salthouse, 2000). Processing Speed has been implicated in speech perception due to the sequential nature of the speech signal, which requires rapid and repeated recruitment of other cognitive processes such as, but not limited to, working and Episodic Memory and linguistic knowledge (Wingfield, 1996). It could be speculated that such rapid comprehensive processing is even more important when the speech is complex (e.g., long complex sentences, fast speech rate, large number of propositions) and/or the speech signal is degraded. In this case the speed with which this knowledge can be accessed determines how deeply the speech is processed and how much extra load is placed on memory processes (Gordon-Salant et al., 2001; Wingfield et al., 1999). Older adults tend to process information at a slower speed so it may well be that slowing Processing Speed is a factor for declining SiN perception in older listeners (Pichora-Fuller, 2003b).

2.2.3 Review guidelines

Although this is a review of basic research, the conduct and reporting of this systematic review and meta-analysis was informed by healthcare systematic review guidelines, including the Centre for Research and Dissemination's guidance for undertaking reviews in health care (Centre for Research and Dissemination, 2009), the Grading Quality of Evidence and Strength of Recommendations (Atkins et al., 2004) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist (Moher et al., 2009).

Systematic search strategy and study identification

This review considered all of the existing literature published to May 2017. Only published studies appearing in peer-reviewed journals were considered. The literature search was conducted using Web of Science, PubMed and Scopus. The search terms "speech" AND "cognit*" AND "noise" OR "babble" OR "talker" NOT "children" NOT "imaging" were entered across all categories and yielded 19,012 hits. The removal of duplicate studies reduced this number to 18,764 studies.

PICOS screening criteria

In the screening process each of the 18,764 studies were assessed, by reading the titles and abstracts, and included/eliminated based on the PICOS (Population, Intervention, Comparator, Outcome, Study design) criteria (Centre for Research and Dissemination, 2009). Studies which could not be assessed by the titles and abstracts were subject to a full-text search. I and a member of my supervision team (HH) independently conducted the screening and identification processes. In the full-text search I collated, removing any duplication, the studies selected in the identification.

Population

Inclusions: Studies reporting results of at least one group of adults (18+ years) with:

- Hearing in the range of normal sensitivity to moderate hearing loss measured using pure-tone audiometry (pure-tone average thresholds better than 71dB across at least three octave frequencies below 8kHz)
- No reported previous or current hearing intervention

Exclusion: Studies which are explicit in reporting listener groups which include:

- Non-native speakers
- Visual impairment not corrected to normal
- Diagnoses of neurological or psychiatric co-morbidities

Intervention

A minimum of one audio-only SiN perception measure consisting of a concurrently and co-locally presented target and masker was the intervention. A composite SiN outcome measure was only accepted if the individual measures that made up the composite assessed target/masker combinations within the same category as defined above, e.g., a composite measure that comprised two or more individual measures of sentence-in-4-talker babble.

Comparator

A minimum of one cognitive ability measure acted as comparator. A composite was only accepted if the individual measures that made up the composite measure assessed a single cognitive sub-domain (*See Categorising cognitive measures section*). Note, any cognitive test that was conducted as part of a dual-task paradigm (e.g. in competing noise) was not considered.

Outcome

A quantitative comparison between speech-in-noise intelligibility and cognitive measures (either correlation, regression, or linear model analyses) was the outcome measure.

Study design

Single time point association studies (or single time point associations taken from a larger study) were considered. SiN intelligibility measures would be presented within either an adaptive or a fixed SNR procedure across the entire intelligibility range. Other measures, e.g. reaction times, were not considered. Both the SiN perception and cognitive performance measures must have been conducted in a quiet room free from distraction, and not as part of a brain imaging paradigm. Only data collected from participants individually were considered. Data collected as part of a group testing session were not included.

2.2.4 Screening results

After initial abstract and title screening, a full-text assessment was deemed necessary for 253 studies. This process resulted in a final set of 25 articles eligible for inclusion in the review. None of the articles included in the review reported more than one study, hence the number of articles equalled the number of included studies. Figure 2. 2 shows a flow diagram of each stage of the search process. Only one study (Zekveld et al., 2011) included a group with hearing aid intervention, alongside a group with hearing thresholds ranging from normal hearing to untreated moderate HL. In this case, only the data from the untreated HL group were included in the review. In all other cases, any participant HL was assumed to be untreated. While the hearing level of listeners in all remaining studies was described as normal or age-normal, the range of pure-tone averages was considerable across studies.

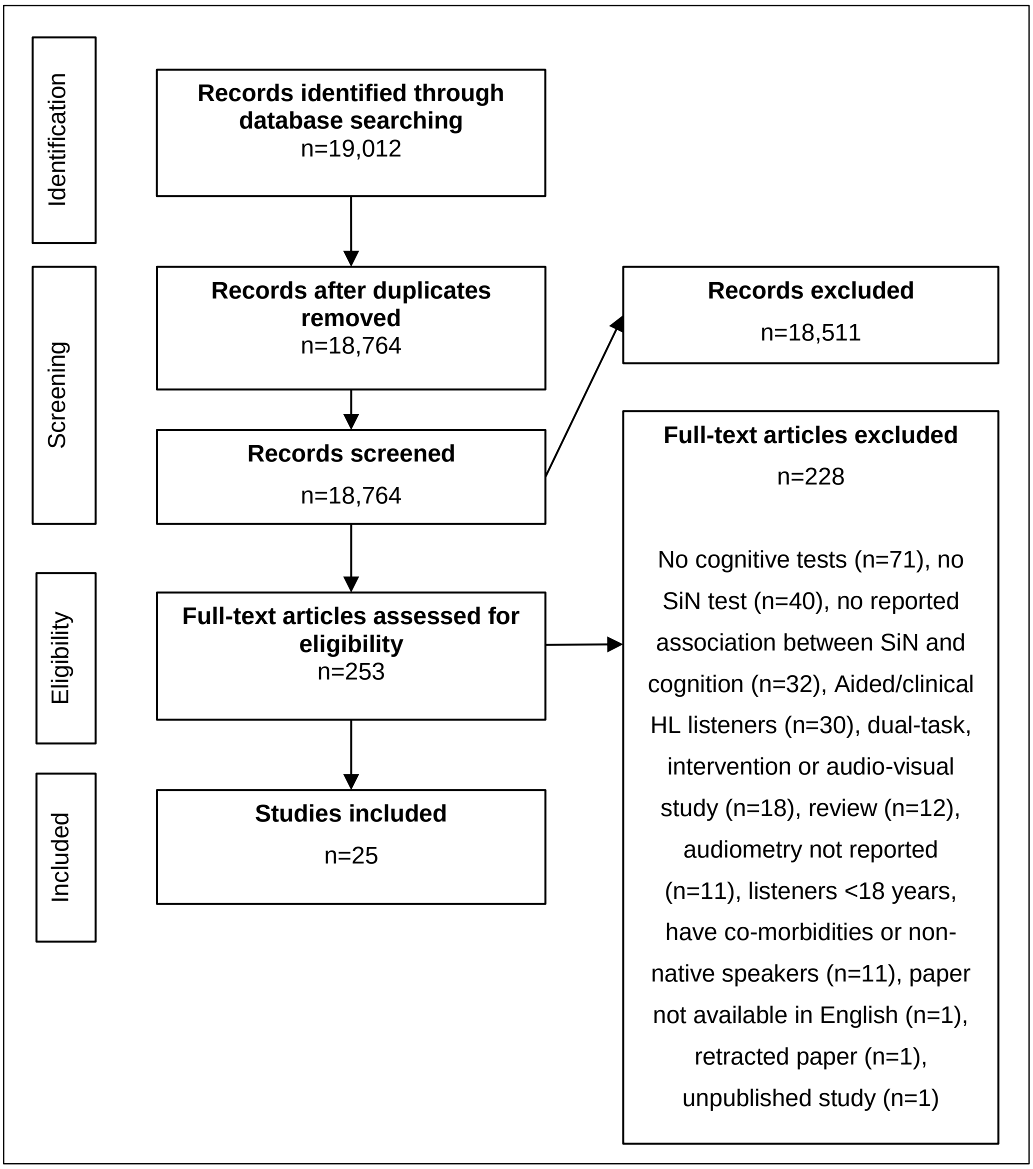


Figure 2. 2 – PRISMA flow chart

PRISMA flow chart of literature search showing the identification, screening, eligibility and inclusion phases of the search

Assessment of risk of study bias

I devised a risk of bias assessment on which each of the 25 full-text articles included in the review were assessed. This scoring system was informed by risk of bias assessments for clinical trials (Higgins et al., 2011). Although only the universal criteria were retained, we must be aware that the reporting requirements of experimental studies are not as rigorous as clinical trials and so we may not expect them to report to these standards.

Supplementary Table 2. 2 details the four questions of the risk of bias assessment (2. 2a) and the score key (2. 2b). All 25 studies were independently scored by my supervisor HAA. In addition, all of the studies were also independently scored by either myself or one of the other supervisors AH or HH. Studies whose scores diverged in more than one category were discussed between scorers until a consensus was reached on at least 3/4 questions. If a divergence remained in one question for a given study the maximum divergence allowed was one point.

2.2.5 Categorisation of studies

Each study's methods were read, and the SiN and cognitive measures were categorised according to the matrix in Figure 2. 1, cognitive measures according to Supplementary Table 2. 1.

Categorisation based on participant groups

As it has been suggested that HL may play a moderating role in the association between cognitive performance and SiN perception (Füllgrabe et al., 2016b), I considered, where possible, the association of cognitive performance and SiN perception for studies where listeners' hearing sensitivity ranged from normal hearing to mild HL and where ability ranged from normal hearing to moderate HL.

Unfortunately, it was not possible to assess associations across the categories (normal hearing, mild HL, moderate HL) independently due to the overlapping sampling methods employed by the studies included in this review. Such a differentiation needed to be balanced against the fact that the number of reviewed studies was rather small and the combination of SiN and cognitive conditions rather large. If the association between cognitive performance and SiN listening is universal, then I would expect the inclusion/exclusion of listeners with moderate HL not to make an appreciable difference to the strength of association. If on the other hand, HL moderates the relationship, then we might expect the level of association to change depending on the presence of the listeners with moderate HL.

I categorised reported audiometric thresholds according to British Society of Audiology (BSA) guidelines, BSA (2011), in normal hearing (<20dB HL average across octave frequencies .25-4kHz), mild HL (20-40dB HL, .25-4kHz), and moderate HL (41-70dB HL, .25-4kHz). I then categorised studies according to their participant group. Sixteen studies fitted into the normal hearing to mild HL category, nine into the normal hearing to moderate HL category.

Here I only considered pre-clinical, unaided listeners. Hearing intervention and aided listening may influence the association between cognitive performance and the processing of incoming (altered) acoustic signals (Ferguson et al., 2017). For a review investigating the role of cognitive sub-domains in hearing intervention/impairment see Taljaard et al. (2016).

2.2.6 Meta analyses

In order for a meta-analysis to be performed for a given cognition and SiN perception test association a minimum of four studies were required. This number was chosen to provide a balance between calculating as many meta-analyses as possible while also maintaining a minimum of statistical power. For all meta-analyses, if more than one quantitative comparison was reported in a single study (e.g. the same SiN perception test correlated with two different measures of Working Memory), the mean value was computed from the multiple correlation coefficients.

Meta-analyses and Forest plots were computed using MedCalc® version 16.8.4. A random-effect model was chosen for the calculation of pooled associations because it incorporates random variation both within and between studies. The applied model calculated weighted summaries of individual correlations based on the Hedges-Olkin method (Hedges et al., 1985). Heterogeneity between studies was assessed using the I^2 statistic (Higgins et al., 2002) with 0% showing no heterogeneity between studies and a higher percentage value indicating higher heterogeneity between studies included in the pooled association. No comparison was removed on the basis of high heterogeneity. Forest plots aid the comparison of individual studies included in the meta-analysis. Within each Forest plot marker size varies according to weight assigned to each study based on the random-effects model. Larger symbols indicate a larger contribution to the pooled (or average) associations.

2.3 Results

2.3.1 Included studies

A summary of each of the 25 articles included in the review is given in Supplementary Table 2. 3. The table includes demographic information about participants, and categorisations of SiN and cognitive measures for each study.

2.3.2 Risk of bias assessment

The results of the bias assessment are displayed below in Table 2. 1. Risk of bias was high for Q1 as the majority of these basic investigations did not include a sample size calculation to inform statistical power. For those studies that excluded participant data, adequate justification was provided in most cases (Q2). Around a third of studies did not provide sufficient information to confirm that results were reported for all included outcome measures (Q3). The majority of studies did not report any conflicts of interest (Q4). Taken together, although I can be relatively confident that the reported results are at low risk of reporting bias, I are unable to confirm whether or not the individual studies included in this review and meta-analysis include sample sizes that are sufficient to adequately detect statistically significant associations. One motivation to conduct a meta-analysis is to overcome this short-coming.

2.3.3 Speech-in-noise perception tests

The 25 studies tested a total of 1026 listeners on a total of eight different combinations of foreground (target) and background (masker) signals. Table 2. 2 shows the frequencies with which each target-masker combination was used. Relatively few studies used phonemes or words as speech target stimuli. Of those that used sentences, all types of masker were used, with unmodulated noise being the most frequent.

Study	Q1: Sample Size	Q2: Exclusion	Q3: Outcome	Q4: Conflict
Anderson et al. (2013)	?	N/A	✓	✓
Besser et al. (2012)	✗	✓	✓	#
Carroll et al. (2016)	✗	✓	✓	✓
Cervera et al. (2009)	✗	N/A	✓	?
Ellis & Rönnberg (2014)	✗	N/A	✓	✓
Gordon-Salant et al. (2015)	✗	N/A	#	✓
Gordon-Salant & Cole (2016)	✗	N/A	?	✓
Heinrich et al. (2015)	✗	?	✓	✓
Heinrich & Knight (2016)	✗	N/A	✓	✓
Helfer & Freyman (2014)	✗	N/A	#	✓
Janse (2012)	✗	N/A	✓	✓
Koelewijn et al. (2012)	✗	N/A	✓	✓
Meister et al. (2013a)	✗	N/A	?	✓
Meister et al. (2013b)	✗	N/A	✓	✓
Parbery-Clark et al. (2009)	?	N/A	#	✓
Parbery-Clark et al. (2011)	✗	N/A	✓	✓
Rönnberg et al. (2014)	✗	N/A	?	?
Slater & Kraus (2016)	✗	✓	✓	#
Stenbäck et al. (2015)	✗	N/A	?	✓
Surprenant (2007)	✗	✓	✓	✓
Tun & Wingfield (1999)	✗	✓	?	✓
Uslar et al. (2013)	✗	N/A	✓	✓
Veneman et al. (2013)	?	N/A	✓	✓
Zekveld et al. (2011)	✗	✓	✓	#
Zekveld et al. (2014)	✗	N/A	✓	✓

Table 2. 1 – Bias score summary for each study included in the meta-analysis

Bias scores for each article included in the review. Full details of the scoring questions and verbal descriptions of the response categories are in Supplementary Figure 2. 2, briefly Q1: Did the authors include a sample size justification? Q2: If any participant data were excluded from the analysis is a clear justification given? Q3: Were all the outcome measures in the methods included in the results? Q4: Were there any conflicts of interest? I.e., is the study funded or conducted by a body with vested interests in the results? Scores highlighted in red indicate a high risk of bias, green indicate low risk of bias and scores in orange indicate an unknown risk of bias. For each question the score could be, ✗ (Q1-3 Insufficient information for judgement/Q4. Clear conflict of interest), ? (Q1-3 Incomplete information/Q4 unclear), ✓ (Q1-3 Appropriate use and sufficient information/Q4 no conflict of interest) or N/A for Q2 (i.e. there were no relevant instances). Where there is a difference between the scorers this can be seen by the total being a # and is considered the equivalent risk as ?.

	Phoneme/ syllable	Word	Sentence	*total*
Unmodulated noise	2	0	13	*15*
Modulated noise	0	1	5	*6*
>2-talker babble	0	3	10	*13*
≤2-talker babble	1	*0*	5	*6*
Total	*3*	*4*	*33*	*40*

Table 2. 2 – Frequencies of SiN *perception* test target/masker combinations

Frequency of target and masker combinations across all 25 reviewed studies. Where target/masker type combinations are repeated within a study the combination is only recorded once.

2.3.4 Cognitive test measures

The 25 studies included a total of 59 cognitive measures which comprised two measures of Alerting, one of orientating, two of Set-Shifting, seven of Inhibitory Control, 26 of Working Memory, seven of Episodic Memory, two of Fluid IQ, eight of Crystallised IQ, and four measures of Processing Speed.

2.3.5 Meta-analyses

In total I carried out five sets of meta-analyses (reported in Tables 2. 3 - 2. 7).

In the first set of analyses, the overall association between all cognitive performance (collapsed across all sub-domains) and SiN categories (collapsed across all categories) was investigated. It was carried out with a sub-analysis for the groups with different amounts of hearing loss.

A second set of analyses looked at each cognitive sub-domain in turn with SiN perception tests collapsed across all categories. Sub-analyses were conducted for the two hearing loss groups where possible.

For the third and fourth set of analyses the SiN perception tests were separated along the two dimensions of target and masker type, and associations with a particular cognitive sub-domain were calculated for each dimension. For instance, when the association with SiN target types was investigated, separate group analyses with cognitive sub-domains were calculated for each SiN target type (phonemes, words and sentences) while collapsing over all types of background masker. Similarly, when the association with background masker was investigated, separate group analyses with cognitive sub-domain were calculated for each type of masker (unmodulated noise, modulated noise, >2-talker babble and ≤2-talker babble) while collapsing across all SiN target types. In a final set of analyses, the association between cognitive sub-domains and specific SiN perception measures (collapsing across target or background signals, e.g. sentences-in-modulated noise) was assessed.

Association between cognitive performance (collapsed across sub-domains) and SiN perception (collapsed across all target/masker types)

The analyses of the association between a general measure of cognitive performance (includes all cognitive subdomains and where multiple subdomains were reported in a single study an average correlation coefficient was calculated) and a general measure of SiN perception (includes all target/masker types and where multiple SiN perception tests were reported in a single study an average correlation coefficient was calculated), when considering the full range of listeners, showed an association of .31. The sub-analysis of hearing range showed associations of .31 and .32 with virtually overlapping confidence intervals. The heterogeneity statistics (I^2) showed that all three groups had low-moderate heterogeneity, ranging between 42-47%. The confidence intervals for all hearing ranges and the normal hearing to mild HL sub-groups differed from zero indicating that there was some unexplained variance between or within studies. I suggest factors such as cognitive sub-domain and SiN target and masker types may account for some of this variance. I will explore these factors in further detail in the sections following this one.

Table 2. 3 shows the full descriptive statistics of the meta-analysis for the entire group of studies and for the two sub-groups of listeners with normal hearing to mild HL and normal hearing to moderate HL. Figure 2. 3 displays the Forest plots of the individual studies contributing to, as well as the mean association of, each of the three meta-analyses. The plots show that while most associations were positive, only some reached statistical significance.

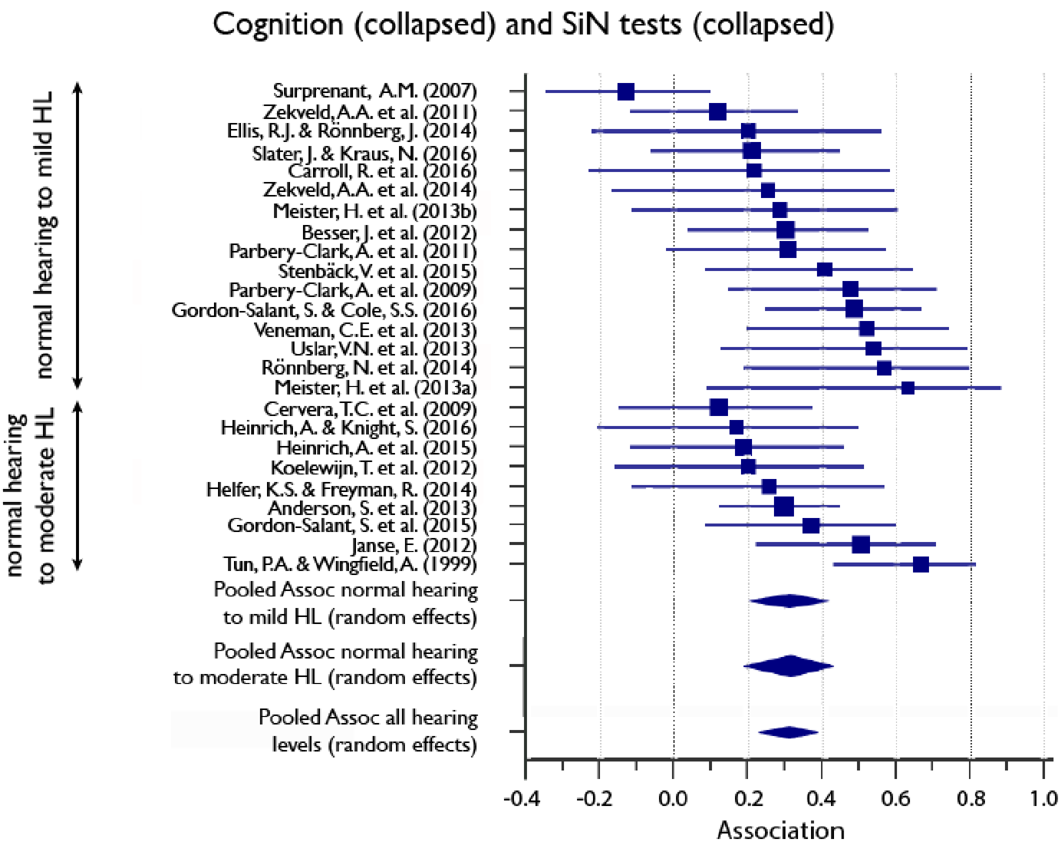


Figure 2. 3 – Forest plot of overall cognition and SiN associations

Forest plot showing the association between cognition (all sub-domains collapsed) and SiN (all conditions collapsed) for normal to mild and normal to moderate hearing loss. Marker sizes for individual studies (squares) are weighted on random-effect model weights. Whiskers represent 95% confidence interval (CI). Pooled effects, calculated using a random-effects model, are shown as diamonds with the symbols extending to 95% CI.

Cognitive sub-domain	Target	Masker	Hearing range	Pooled (n)	Pooled (r)	95% CI of r	Z statistic and p-value	I^2	95% CI of I^2	# sig. studies/ # all studies
Collapsed	Collapsed	Collapsed	All	1026	.31	.23 to .39	7.2, <.001	44%	9 to 65	12/25
			NH to mild HL	595	.31	.20 to .42	5.28, <.001	47%	6 to 71	8/16
			NH to moderate HL	431	.32	.19 to .43	4.82, <.001	42%	0 to 73	4/9

Table 2. 3 – Summary of meta-analysis for overall cognition and SiN association

Meta-analysis of the association between cognition (all sub-domains collapsed) and SiN perception (all conditions collapsed) for all listeners and subdivided for ranges ‘normal hearing (NH) to mild hearing loss (HL)’ and ‘NH to moderate HL’. CI: confidence interval, I^2: heterogeneity statistic.

Association between cognitive sub-domains and SiN perception (collapsed across all target/masker types)

Table 2. 4 shows the full descriptive statistics for the association between cognitive sub-domain and SiN perception measures, which was computed for Inhibitory Control, Working Memory, Episodic Memory, Crystallised IQ and Processing Speed. For Working Memory, the meta-analyses were also run separately for groups of listeners whose hearing ranged between normal and mild HL and normal and moderate HL. Associations ranged between .18 and .39 and were significant for all sub-domains, except Crystallised IQ. Heterogeneity was low for all cognitive submain comparisons with the exception of Crystallised IQ, and the normal hearing to mild HL group sub-analyses of Working Memory. One explanation of the heterogeneity in the normal hearing to mild hearing loss sub-group analysis for Working Memory is that factors such as SiN listening condition may be causing additional variance between studies.

Figure 2. 4 displays the Forest plots of the individual results contributing to, as well as the mean association of, each meta-analysis of the five sub-domains. The plots show that while most associations were positive, only some reached statistical significance.

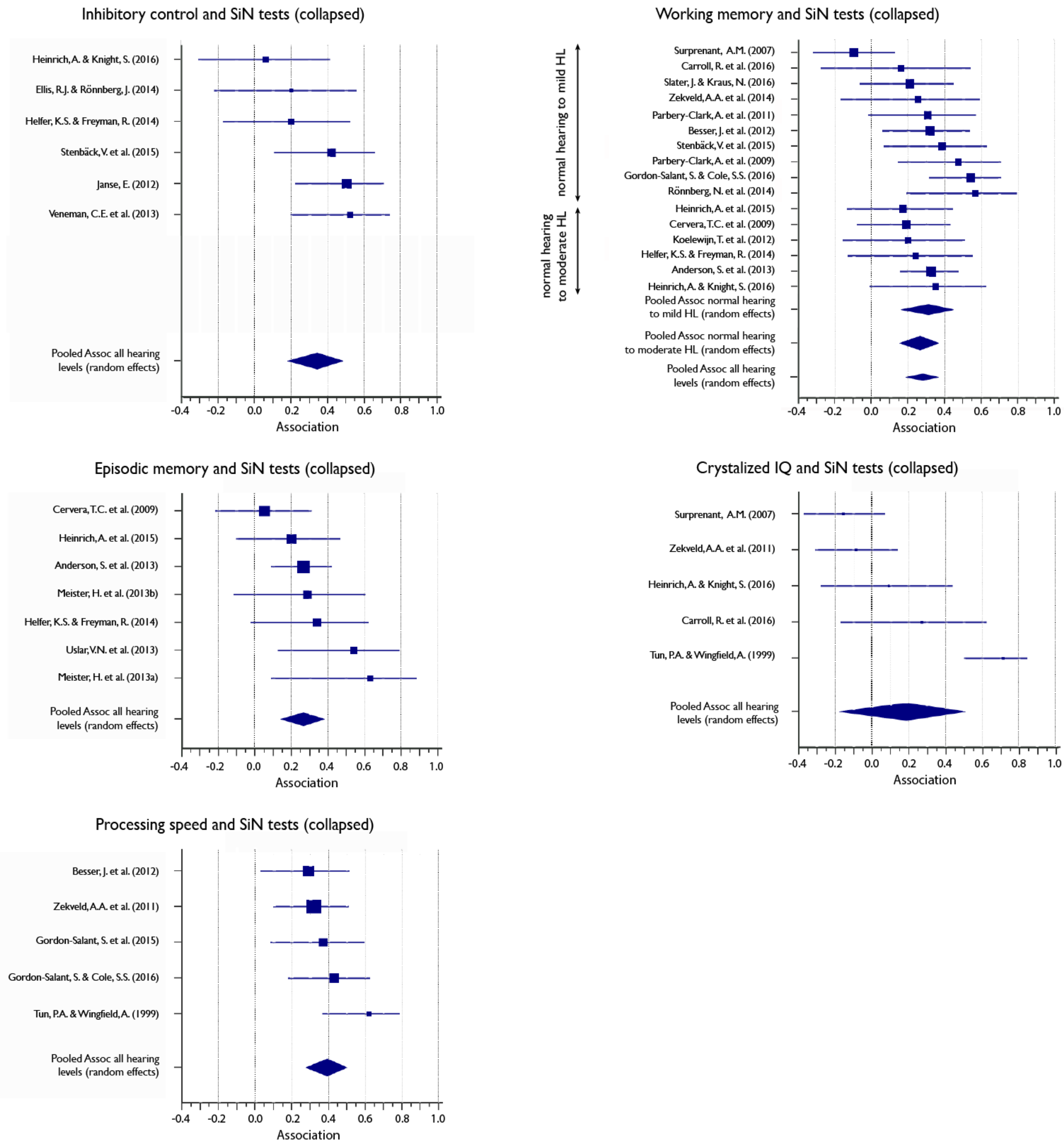


Figure 2. 4 – Forrest plots for cognitive sub-domain and SiN associations

Forest plots showing the association between cognitive sub-domain and SiN (all conditions collapsed) for all listeners unless otherwise stated. Marker sizes for individual studies (squares) are weighted on random-effect model weights. Whiskers represent 95% confidence interval (CI). Pooled effects, calculated using a random-effects model, are shown as diamonds with the symbols extending to 95% CI.

Cognitive sub-domain	Target	Masker	Hearing range	Pooled (n)	Pooled (r)	95% CI of r	Z statistic and p-value	I^2	95% CI of I^2	# sig. studies/ # all studies
Inhibitory Control	Collapsed	Collapsed	All	189	.34	.18 to .48	4.08, <.001	23%	0 to 67	3/6
Working Memory			All	720	.28	.19 to .37	5.89, <.001	34%	0 to 64	6/16
Working Memory			NH to mild HL	409	.31	.16 to .45	3.96, <.001	57%	13 to 79	5/10
Working Memory			NH to moderate HL	311	.26	.15 to .37	4.61, <.001	0%	0 to 25	1/6
Episodic Memory			All	307	.26	.14 to .38	4.12, <.001	12%	0 to 75	3/7
Crystallised IQ			All	237	.18	-.18 to .50	1.00, .32	86%	69 to 95	1/5
Processing Speed			All	263	.39	.28 to .50	6.14, <.001	11%	0 to 83	5/5

Table 2. 4 – Summary of meta-analysis for cognitive sub-domain and SiN association

Meta-analysis of the association between cognitive performance sub-domain and SiN perception tests (all target/masker conditions collapsed) for all listeners, unless otherwise stated. CI: confidence interval, I^2: heterogeneity statistic

Association between cognitive sub-domains and SiN target speech types (collapsed across maskers)

Associations ranged between .29 and .43 and were significant for all sub-domains, except Crystallised IQ (see Table 2. 5). Figure 2. 5 displays the Forest plots of the individual results contributing to, as well as the mean association of, each of the six meta-analyses. The plots show that while most associations reported by individual studies were positive, only some reached statistical significance.

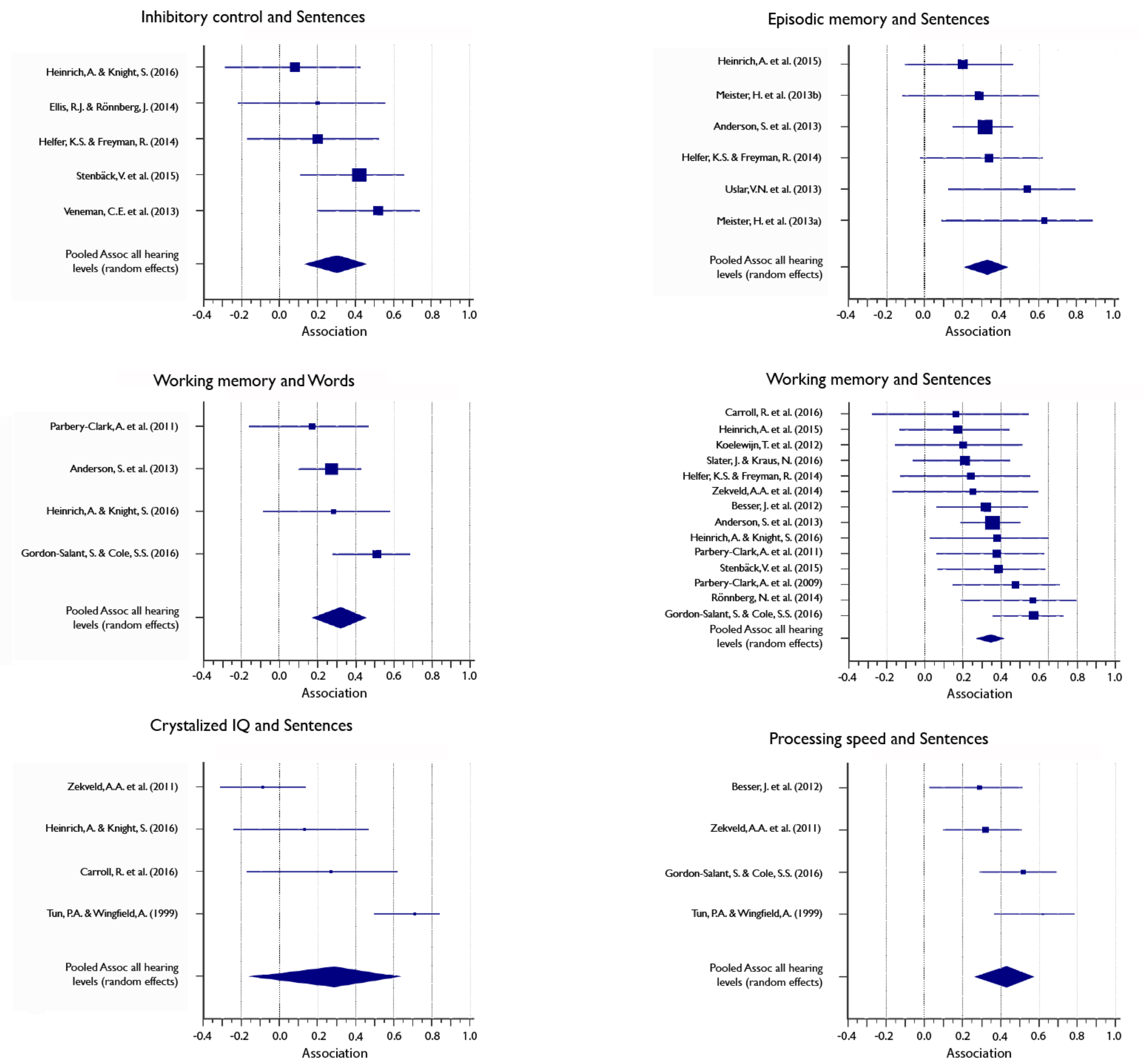


Figure 2. 5 – Forest plots for cognitive sub-domain and SiN target associations

Forest plots showing the association between SiN target types (collapsed over masker) and cognitive sub-domains for all listeners. Marker sizes for individual studies (squares) are weighted on random-effect model weights. Whiskers represent 95% confidence interval (CI). Pooled effects, calculated using a random-effects model, are shown as diamonds with the symbols extending to 95% CI.

Cognitive sub-domain	Target	Masker	Hearing range	Pooled (n)	Pooled (r)	95% CI of r	Z statistic and p-value	I^2	95% CI of I^2	# sig. studies/ # all studies
Inhibitory Control	Sentences	Collapsed	All	150	.30	.13 to .46	3.40, .001	11%	0 to 83	2/5
Working Memory	Words			240	.32	.17 to .45	4.12, <.001	24%	0 to 90	2/4
Working Memory	Sentences			590	.34	.27 to .42	8.37, <.001	0%	0 to 48	8/14
Episodic Memory	Sentences			252	.33	.21 to .44	5.23, <.001	0%	0 to 65	3/6
Crystallised IQ	Sentences			162	.29	-.16 to .64	1.27, .205	86%	67 to 94	1/4
Processing Speed	Sentences			218	.43	.27 to .57	4.83, <.001	45%	0 to 82	4/4

Table 2. 5 – Summary of meta-analysis for cognitive sub-domains and SiN target associations

Meta-analysis of the association between SiN target speech types (collapsed across maskers) and cognitive performance sub-domains for all listeners. CI: confidence interval, I^2: heterogeneity statistic.

Associations between cognitive sub-domains and masker types (collapsed across target speech types)

Associations ranged between .13 and .39 and were significant for all but one (Crystallised IQ) cognitive sub-domain (see Table 2. 6). Figure 2. 6 shows the Forest plots of the individual results contributing to, as well as the mean average association of, each of the five meta-analyses. Again, despite overall significant average association and generally positive associations, only some of the individual associations were significant.

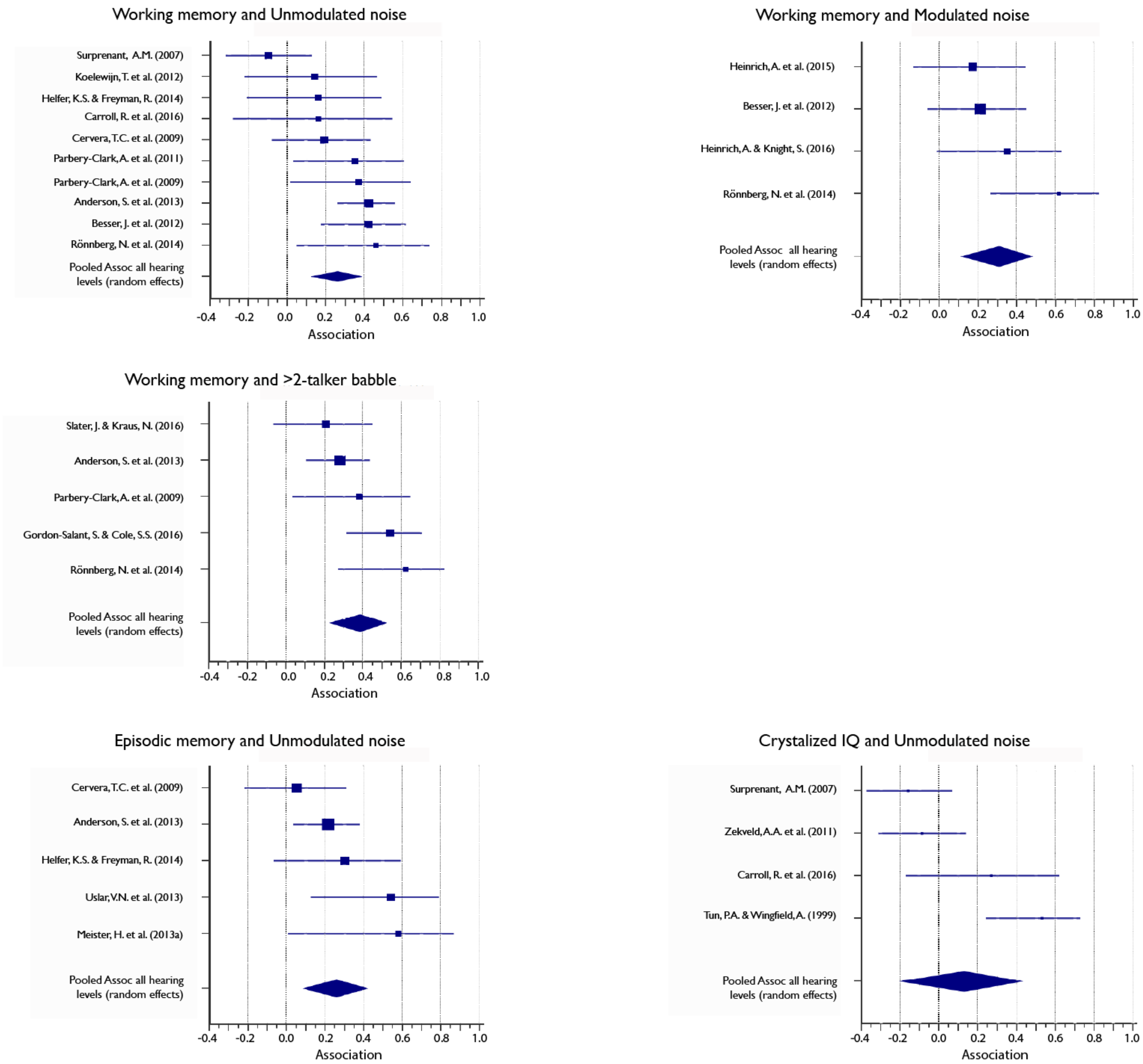


Figure 2. 6 – Forest plots of cognitive sub-domain and SiN masker associations

Forest plots showing the association between SiN masker types (collapsed over target) and cognitive sub-domains for all listeners. Marker sizes for individual studies (squares) are weighted on random-effect model weights. Whiskers represent 95% confidence interval (CI). Pooled effects, calculated using a random-effects model, are shown as diamonds with the symbols extending to 95% CI.

Cognitive sub-domain	Target	Masker	Hearing range	Pooled (n)	Pooled (r)	95% CI of r	Z statistic and p-value	I^2	95% CI of I^2	# sig. studies/ # all studies
Working Memory	Collapsed	Unmodulated noise	All	479	.26	.13 to .38	3.76, <.001	50%	0 to 76	5/10
Working Memory		Modulated noise		151	.31	.11 to .48	3.00, .003	34%	0 to 77	1/4
Working Memory		> 2-talker babble		280	.39	.23 to .52	4.54, <.001	45%	0 to 80	4/5
Episodic Memory		Unmodulated noise		237	.26	.08 to .42	2.88, .004	32%	0 to 74	3/5
Crystallised IQ		Unmodulated noise		207	.13	-.20 to .43	.75, .45	80%	48 to 93	1/4

Table 2. 6 – Summary of meta-analysis for cognitive sub-domain and SiN perception test masker associations

Meta-analysis of the association between SiN masker types (collapsed across target speech types) and cognitive performance sub-domains for all listeners. CI: confidence interval, I^2*: heterogeneity statistic.*

Associations between cognitive sub-domains and specific SiN target speech/masker type combinations

Associations ranged between .31 and .43 and all reached significance (Table 2. 7). Figure 2. 7 shows the Forest plots of the individual results contributing to, as well as the mean association of, each of the four meta-analyses. The Forest plots in Figure 2. 7 indicate that while all contributing associations were positive, there was considerable variability in size and significance of individual associations contributing to each meta-analysis.

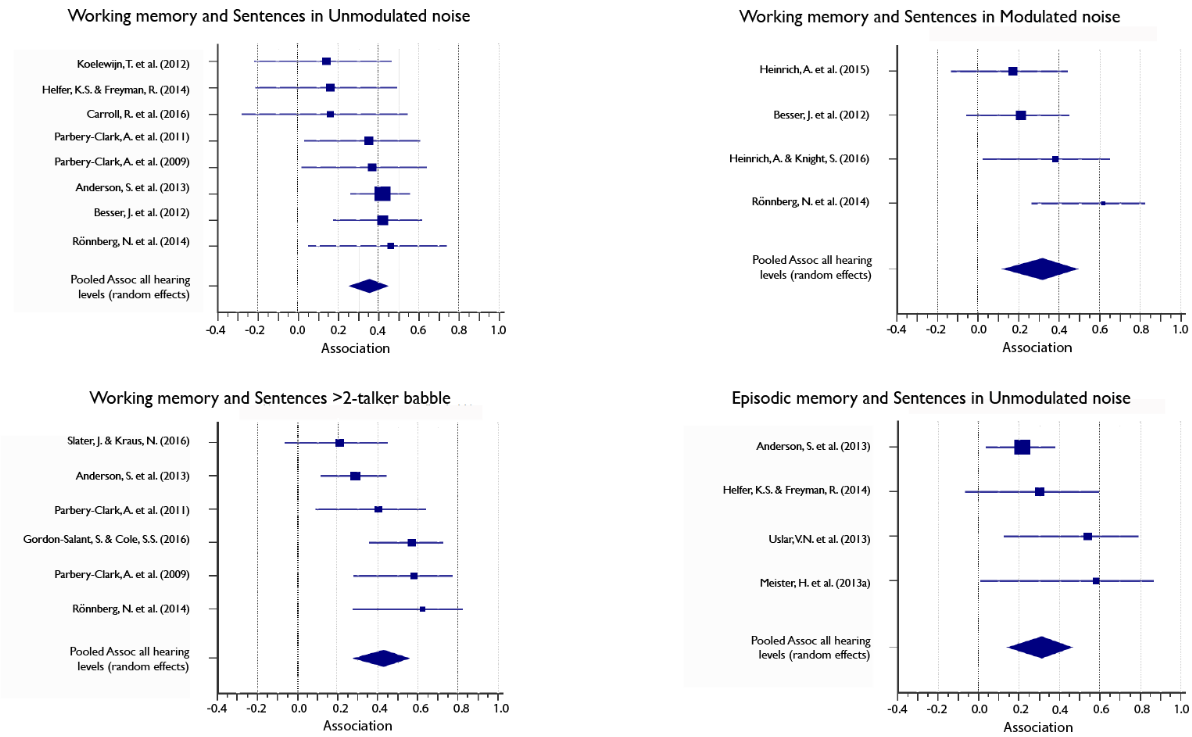


Figure 2. 7 – Forest plots for cognitive sub-domain and SiN target/maker associations

Forest plots showing the association between SiN speech and masker type combinations and cognitive sub-domains for all listeners. Marker sizes for individual studies (squares) are weighted on random-effect model weights. Whiskers represent 95% confidence interval (CI). Pooled effects, calculated using a random-effects model, are shown as diamonds with the symbols extending to 95% CI.

Cognitive sub-domain	Target	Masker	Hearing range	Pooled (n)	Pooled (r)	95% CI of r	Z statistic and p-value	I^2	95% CI of I^2	# sig. studies/ # all studies
Working Memory	Sentences	unmodulated noise	All	349	.35	.25 to .44	6.64, <.001	0%	0 to 55	5/8
Working Memory	Sentences	modulated noise		151	.32	.12 to .49	3.03, .002	36%	0 to 78	2/4
Working Memory	Sentences	>2-talker babble		317	.43	.28 to .56	5.21, <.001	50%	0 to 80	5/6
Episodic Memory	Sentences	unmodulated noise		182	.31	.14 to .47	3.44, .001	15%	0 to 89	3/4

Table 2. 7 – Summary of meta-analysis for cognitive sub-domains and SiN perception test target/masker associations

Meta-analysis on the effect size of the association between SiN target speech and masker types and cognitive performance sub-domains for all listeners. CI: confidence interval, I^2: heterogeneity statistic.

2.4. Discussion

The association between cognitive performance and Speech-in-Noise (SiN) perception has attracted increasing research interest over the past 20 years. However, at the individual study level, the outcomes have been varied and inconsistent. In this chapter I investigated three sources of variation: 1) a wide range of cognitive performance measures, 2) a wide range of SiN perception tests, and 3) variability in participants' hearing thresholds. This review addressed these issues by categorising cognitive measures into five cognitive domains and nine sub-domains according to established cognitive theories. I also categorised the speech signal according to the lexical complexity of its target signal and the extent to which the background signal engages informational masking. Finally, I calculated effects for two participant groups; listeners with normal hearing to mild HL and those with normal hearing to moderate HL. Reported data were assessed in a series of formal meta-analyses where sufficient studies were available.

Here I explored the following research questions: what is the overall association between cognition and SiN perception? Which specific cognitive abilities are associated with SiN perception? How does hearing sensitivity relate to differences previously found in various associations between cognitive ability and SiN perception? Do cognitive ability-SiN perception associations differ depending on SiN listening conditions (target/masker type)?

I will discuss the findings in relation to these questions in the sections below: in relation to general cognitive ability, SiN perception and HL (section 2.4.1), and each cognitive (sub-)domain in relation to SiN perception, and listening condition and HL, where possible (sections 2.4.2 - 2.4.4)

2.4.1 General association between cognitive performance, SiN perception and hearing loss

Collapsing across all cognitive domains and all SiN perception measures, there was an overall association of .31. Furthermore, the strength of the association did not vary depending upon hearing loss groupings. This suggests that cognitive performance is associated with SiN perception, and that this is independent of hearing loss in the ranges examined.

Although I divided cognition into nine sub-processes I was only able to conduct meta-analyses for 5 sub-processes, namely Inhibitory Control (six studies), Working Memory (16 studies), Episodic Memory (seven studies), Crystallised IQ (five studies), and Processing Speed (five studies). This result in itself is of interest because it indicates a bias particularly towards assessing Working Memory ability, while sub-domains such as Alerting (attention), Orienting (attention), Set-Shifting (executive process), and Fluid IQ (intelligence) have received little or no attention.

In the following sections I will discuss the results of each sub-domain where possible.

2.4.2 Attention

Alerting and Orienting were expected to be generally important for SiN perception (Astheimer et al., 2009; Heald et al., 2014). My review of the existing evidence shows that so far only a limited number of studies have investigated these relationships (two Alerting and one Orienting) and as a result I was unable to perform a meta-analysis for this domain.

2.4.3 Executive processes

I hypothesised that executive processing may be linked to SiN perception and that the strength of the association may vary by sub-domains. Only two of three executive processes sub-domains (Inhibitory Control and Working Memory) were reported in sufficient published studies to be (partially) assessed using meta-analyses.

Inhibitory Control has previously been suggested to be important for SiN perception, particularly under informational masking conditions (Janse, 2012; Sommers et al., 1999). It was assessed by six studies and was, with some combinations of SiN conditions, included in a meta-analysis. Overall Inhibitory Control showed a significant association with SiN perception of .34. Furthermore, the great majority of studies that assessed Inhibitory Control in connection with SiN perception used sentences as their target speech. Hence it was not surprising that when the type of target speech was considered, the pooled association between sentences and Inhibitory Control was almost identical (.30) to the overall association. There was insufficient data available to assess differences in association strength between inhibitory processes and different SiN masker types.

It has been suggested that Working Memory is of general importance for SiN perception, regardless of specific target and masker types (Rönnberg et al., 2013) and perhaps particularly so for SiN perception tests that use sentence targets and more complex background maskers (e.g. Akeroyd, 2008). As many studies had included Working Memory measures in their testing protocol, its role for various SiN perception tests could be evaluated in meta-analyses more thoroughly than the role of any other cognitive sub-domain. The general association between Working Memory and speech perception across all listeners was .28 with a slightly higher value for listeners with hearing in the range between normal to mild HL (.31 than listeners with hearing in the range between normal to moderate HL (.26). However, as the confidence intervals of both sub-groups virtually overlapped, it was not possible to conclude that the association between Working Memory and speech perception was moderated by (unaided) HL. But, the heterogeneity statistics indicate that there may be some unexplained variance in the normal hearing to mild HL and not the normal to moderate HL group. This suggests that factors other than Working Memory may be explaining variance in the normal hearing to mild HL group.

The speech target analysis showed similar and significant associations of .32 and .34 across both target stimulus categories for which enough data were available to test separately (i.e., words and sentences). When background masker types were considered separately for sub-categories that provided enough data, significant correlations ranging between .26 and .39 were found for unmodulated noise, modulated noise, and >2-talker babble. It might be interesting to note that association strength appeared to increase with an increasing amount of informational masking in the background signal.

Finally, Working Memory was one of the two cognitive sub-domains (the other was Episodic Memory) that allowed the investigation of specific sub-domain and listening condition combinations, with associations ranging between .32 and .43. While confidence intervals were again largely overlapping, it is interesting to note that mean associations appeared to be strongest when the background sound contained informational masking, and the target type was sentences.

2.4.4 Memory, intelligence and Processing Speed

Episodic Memory was expected to show an association with SiN perception particularly for more complex speech targets (Goldinger, 1996). I found that Episodic Memory showed an overall association with speech perception of .26 and that this association strength did not vary considerably where I could assess specific target speech signals or background maskers.

While there were sufficient studies assessing the association between speech perception and Crystallised IQ to conduct a meta-analysis, this was not the case for

Fluid intelligence. Crystallised IQ has been suggested to be closely linked with SiN perception in terms of comprehension and lexical access (Schneider et al., 2016). When assessing target speech and masker background types separately, some interesting patterns emerged. When Crystallised IQ was associated with SiN perception of any target speech type, masked by unmodulated noise, the pooled association was .13. However, when the target speech was sentences (collapsed across masker types), the association was numerically higher (.29). These data might suggest that the association between speech perception and Crystallised IQ might be driven by the complexity of the target speech, however there are insufficient data and studies to be confident in this conclusion.

Finally, I speculated that Processing Speed may be particularly important in situations with lexically complex speech targets due to an increase in processing required for memory retrieval (Gordon-Salant et al., 2001; Wingfield et al., 1999). Overall there was a significant association (.39) between SiN perception and Processing Speed when collapsing across all SiN categories. In terms of more fine-grained meta-analyses, SiN target type sentences showed a significant association with Processing Speed (.43).

2.4.5 Patterns of results in the literature

This chapter highlights four important patterns in the published data, which only become evident when a large number of studies are simultaneously considered.

First, it appears that the majority of associations between cognitive performance and SiN perception were of the magnitude of $r \approx .3$, although the entire range of associations across all combinations was between .13 and .43. This was seen when collapsing data across cognitive domains and SiN categories, largely regardless of hearing loss, and also when assessing specific cognitive sub-domains, in particular Inhibitory Control, Working Memory and Episodic Memory. It is striking how little the association between SiN and cognitive performance differed across cognitive sub-domains when the SiN target speech was sentences. As other types of target speech were comparatively rarely used, it is difficult to know whether a similar uniformity of associations would be seen for other types of target speech. Conversely, different combinations of cognitive sub-domains and background maskers seem to vary more. Thus, being specific about the target and background signal as well as the tested cognitive sub-domain and employing the full range of available stimuli may be a way to draw out further variability in association.

Second, it is interesting that although pooled associations were statistically significant, half of the associations from single studies that contributed to the meta-analyses (13/25) were not. This is particularly true for the cognitive sub-domains of Working Memory and Episodic Memory. In the case of Working Memory, it also appears to be a particular issue for studies with listener groups in the range of normal hearing to moderate HL. Possibly this result may highlight issues with low statistical power for individual studies (see the results of the risk of bias assessment – Table 2. 1), so that the associations only become reliably significant when data are pooled. This also has potential ramifications for the sub-processes other than Working Memory included in the meta-analyses, which included fewer studies (five to seven). The results for these sub-processes may be less robust and could potentially alter at the level of pooled effect size if more studies were included. Therefore, with specific reference to Crystallised IQ, it cannot be categorically concluded that Crystallised IQ ability does not play a role for SiN perception.

The third key result of this review is that associations between SiN perception and many of the cognitive domains have so far been under-investigated. Attention and Fluid intelligence did not feature in enough included studies to warrant meta-analyses (n<4). Even executive processes, which have been investigated in much greater detail, do not provide enough data to examine their role across the whole range of individual SiN target and background categories. For a comprehensive and detailed understanding of the relationship of cognition and SiN perception, a systematic investigation of the association between all cognitive sub-domains and SiN target/masker types, even when no significant correlations are expected, would be informative. Negative or non-significant results are just as important as significant correlations because they allow us to understand the specificity of these results.

Finally, it is worth noting that when the moderating role of hearing sensitivity was assessed I found little difference in association between studies that included listeners with relatively better or poorer average unaided hearing thresholds, given the limited categorisation I was able to apply.

2.5 Limitations

There are some limitations of this review chapter. Firstly, all cognitive tests were assigned to a specific cognitive domain to aid data categorisation for assessment and reporting. However, it is recognised that any given cognitive test may actually assess a multitude of cognitive domains, and to different extents (e.g. Surprenant et al., 2009). I note that re-assignment of complex cognitive tests to different respective cognitive domains or sub-domains may lead to minor differences in the conclusions drawn from this research.

Secondly, I did not account for differences in measurement or scoring methods across cognitive tests that assess a single sub-domain. Although I recognise its importance this is not a factor I was able to specifically assess in this review. For a review on general method test bias in psychometric tests see Podsakoff et al. (2003) and for an overview on memory span tasks see (Conway et al., 2005).

Thirdly, cognitive domains were informed by multiple cognitive theories rather than on the basis of one specific unifying framework (although this could be viewed as a more informed and considered process than using a single theory).

Fourth, I am limited in my conclusions by the available literature. For instance, I was not able to evaluate whether visual perception (perhaps indicating general differences in health or cognition) interacted with performance on cognitive tests (Scialfa, 2002) because virtually no studies measured this.

Finally, the SiN categorisation did not discriminate between adaptive and set level signal-to-noise ratio paradigms, type of response set, different intelligibility levels or modes of signal presentation, and instead assumed that methodologies would engage cognitive processes in a similar way and to a similar extent. However, this may not be the case as suggested by the results of studies which have examined associations between cognition and non-adaptive SiN perception tests at multiple SNRs (Carroll et al., 2016; Heinrich & Knight, 2016; Tun et al., 1999) or adaptive SiN perception tests at multiple levels of intelligibility (Koelewijn et al., 2012) within the same speech signal and masker type combination.

In future studies this assumption needs to be further examined, with investigations of associations between adaptive versus non-adaptive SiN perception tests and cognition being of potential interest to both basic science and clinical perspectives.

2.6 Conclusions

Summarising the results of this chapter I conclude that: 1) for cognitive performance and SiN perception r=.3 appears to be the ‘magic number’ for strength of association. 2) Inhibitory Control, Working Memory, Episodic Memory and Processing Speed are shown to be important for SiN perception, consistent with previous published evidence. These conclusions are based on a literature which is selective in the specific measures and stimuli used, such that many alternative hypotheses have not yet been sufficiently assessed.

Chapter three - An association study: the role for cognition and hearing sensitivity for SiN perception in younger and older adult listeners

Abstract

This study examined the association between cognition and speech-in-noise perception in a range of different listening conditions for younger and older adults.

I selected two listener groups, younger (n=50, 18-30 years) and older adults (n=50, 60-85 years) who varied in hearing sensitivity (assessed using pure-tone audiometry) between -10 and 15 dB HL at 0.25-4kHz (young adults) and 0 and 39 dB HL at 0.25-4kHz (old adults).To assess different listening conditions I selected speech-in-noise tests, which varied in the semantic support of the target (single words, low and high predictability sentences) and the informational properties of the background masker (speech-modulated noise, 3-talker babble). Cognitive tests were selected specifically on the basis of Baddeley's (2012) model of Working Memory (Central Executive, Episodic Buffer and Phonological Loop). However, I also applied a second model Diamond's (2013) model of Executive Functions (Working Memory and Inhibitory Control), to explore is model selection was critical or if findings could generalise across different theoretical approaches. Multiple cognitive tests were selected for each sub-domain, allowing principal components to be derived for each of the sub-domains.

Using this systematic and theory-driven approach I explored the following research questions: Which cognitive abilities are important for SiN perception? Does the role of specific cognitive abilities depend on SiN listening condition, age group or hearing sensitivity?

With respect to the Baddeley model, Central Executive ability was shown to contribute to SiN perception in the older, but not the younger listeners, regardless of listening condition. Phonological Loop processing was shown to be important for both listener groups, but revealed a different role depending on age group and masker type. Episodic Buffer ability may only contribute in older listeners and is modulated by hearing sensitivity and background masker. With regards to the Diamond model, Working Memory ability also play a role for SiN perception the older adult group and this role is moderated by hearing sensitivity and masker type.

This novel systematic investigation of the role of cognition for SiN perception suggested that younger and older listeners employ different listening strategies and these strategies can vary depending on listening condition and hearing sensitivity.

3.1. Introduction

3.1.1 Background

The aim of this chapter is to extend previous cognitive hearing association study research (Pichora-Fuller et al., 1995; Rönnberg et al., 2014; van Rooij et al., 1992; Zekveld et al., 2011), by taking a systematic approach to Speech-in-Noise (SiN) perception test selection and using well-established cognitive models as the basis for cognitive test selection. This is important in the investigation because it is a key first step in uncovering how the roles of specific cognitive abilities vary for specific SiN listening conditions.

Although cognition is recognised as being important for SiN perception (Arlinger et al., 2009; Rönnberg et al., 2010), the findings of previous studies appear to be inconsistent. Taking Working Memory ability as an example, an ability commonly linked to SiN perception (Rönnberg et al., 2008), not all studies have found a significant Working Memory ability-SiN intelligibility association (Akeroyd, 2008; Füllgrabe et al., 2016b). Some possible explanations of these inconsistencies include variation and limitations in selection of both SiN and cognitive tests. For example, with regards to SiN perception, tests can vary between studies in a number of ways including but not limited to, speech target, background masker, Signal-to-Noise Ratio (SNR), procedure (fixed versus adaptive), and scoring method. These factors make it difficult to compare results within and between studies. Therefore, it is desirable to select multiple SiN perception tests within a single experiment, systematically varying the target stimulus and masker type. Doing so will allow us to make inferences of the role of cognition for SiN in specific listening conditions.

Furthermore, with regards to cognition there are multiple tests used across the literature to assess specific cognitive abilities. However, these tests can vary in the extent to which they engage specific abilities, some assessing multiple abilities (Surprenant et al., 2009), and can also vary in terms of task in a multitude of ways, e.g., modality of stimulus (verbal/non-verbal), response measure (item recall, reaction time), task difficulty, and scoring method. These factors have implications in making inferences of the role of cognition for SiN perception within studies and also comparing results across studies which use different cognitive tests to assess a specific cognitive ability. One way to overcome the task- and ability-specific variation is to select multiple tests theorised to assess a specific ability and to extract the latent structure using a factor analysis or structural equation modelling method.

Other factors, which may contribute to the inconstancies, include age and hearing sensitivity. Age is of particular interest because it is associated with declines in sensory (including hearing sensitivity) and cognitive factors (Roberts et al., 2016). However, it is not clear how these factors interact with each other (Jerger, 1992) and how they relate to specific SiN listening conditions (Moore et al., 2014).

To address these issues, I took a systematic and theory-driven approach to investigate the role of cognition for SiN perception. I varied listening condition by target and masker, assess multiple cognitive abilities using multiple tests. This allowed me to investigate the role of cognition for SiN perception in specific listening condition, which may vary involvement and engagement of specific cognitive abilities. I also selected two age groups, younger and older adults, and assessed hearing sensitivity using Pure-Tone Audiometry (PTA) to assess how the relationship between cognition and SiN perception is mediated by age and hearing sensitivity.

In the sections below, I describe the SiN listening conditions and cognitive abilities I assessed in this study. I will also draw specific hypotheses for the role of cognitive abilities for particular SiN listening conditions, and how these roles may vary with age and hearing sensitivity.

Speech-in-Noise perception

SiN perception tests were selected based on the same speech target by masker type matrix used in chapter two. Because the selection of SiN perception tests to cover every cell of the SiN target/masker matrix would have been too resource-consuming to test, I selected sub-groups of target stimuli with specific theoretical interest (*see Figure 3. 1*), namely single words and sentences. Sentences were sub-categorises into low (LP) and high semantic predictability (HP). Further descriptions of the target and masker selections are given below. These target stimuli were selected because they vary in terms of the amount of semantic support, from least semantic support (single words), to medium semantic support (LP sentences) to highest semantic support (HP sentences). Semantic support was chosen to be varied because previous findings showed that speech targets of different complexities may recruit different cognitive resources or similar resources but to different extents (Heinrich et al., 2015; Heinrich & Knight, 2016; Xu et al., 2005).

Moreover, two types of SiN maskers were chosen, speech-modulated noise and 3-talker babble noise. In terms of masker-type matrix used previously they fell within the modulated noise and >2-talker babble categories respectively. These masking conditions were selected because they were matched for energetic masking properties but varying in informational masking properties, with the speech-modulated noise exerting less informational masking than 3-talker babble (see the *Masker* sub-section within the Methods section for further details). The more informational masker is expected to engage cognitive processes either more or at least differently compared to the less informational masker due to the presence of the intelligible distractor information (Freyman et al., 2004; Janse, 2012; Mattys et al., 2009).

The SiN perception tests were selected in such a way that they allowed me to build upon and relate results to that of the systematic review and meta-analysis in chapter two. Specifically, four of the selected SiN conditions (words in modulated noise, words in >2-talker babble, LP sentences in modulated noise, HP sentences in modulated noise) conditions were under-represented in the literature. The other two, (LP and HP sentence in a background of >2-talker babble) were less novel; however, this allowed the findings of this study to be compared to a more robustly tested area in the context of previous studies.

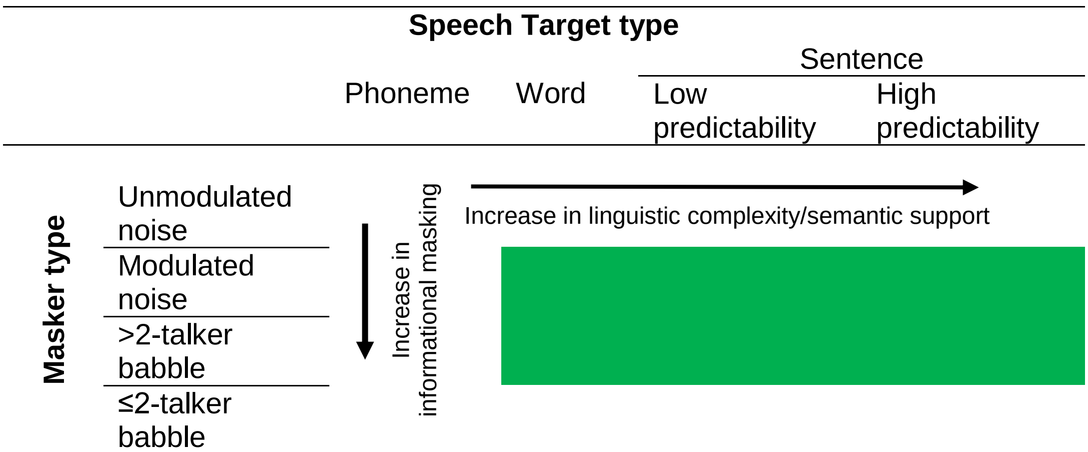


Figure 3. 1 – SiN *perception* test target/masker matrix

Adapted from Figure 2.1. Speech-in-Noise test matrix displaying the categories for classifying speech target and masker type. Highlighted cells indicate where in the matrix SiN perception tests were selected for this study.

Cognition

Several cognitive domains are theorised as being important for SiN perception, including, Attention (Heinrich et al., 2015), Executive Functions (Helfer et al., 2014; Janse, 2012; Veneman et al., 2013), Processing Speed (Gordon-Salant et al., 2015; Zekveld et al., 2011), Episodic Memory (Meister, Schreitmuller, Grugel, Beutner, et al., 2013; Uslar et al., 2013) and Working Memory (Parbery-Clark et al., 2009; Rönnberg et al., 2008; Stenbäck et al., 2015).

Working Memory, the process by which information is simultaneously stored and manipulated in complex tasks, has perhaps been most frequently cited as being important for speech perception. Working Memory processes may be particularly important in adverse conditions such as listening in noise (Akeroyd, 2008; Pichora-Fuller et al., 1995; Rönnberg et al., 2008) and maybe of increased importance for older adults (Füllgrabe et al., 2016a) who are more likely to have hearing or cognitive declines.

One model in the cognitive hearing science literature which describes the role of Working Memory for speech perception is the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013; Rönnberg et al., 2008). The model assumes that there is an Episodic Buffer, which facilitates the Rapid, Automatic and Multimodal Binding of PHOnological (RAMBPHO) information. Speech information is then matched for representations in episodic long-term memory. In favourable conditions the speech matching process is fast and implicit. However, adverse conditions, such as SiN listening, can disrupt or interfere this matching process. In these circumstances explicit Working Memory processes (storage and manipulation) and other executive processes are recruited to resolve the mismatch.

The ELU model does not focus on categorising the specific sub-domains of Working Memory (Rudner et al., 2014). However, it is of interest for the sub-domains of Working Memory that is manipulation and storage, to be examined as distinct sub-domains. This approach would allow for the roles of the sub-domains of Working Memory to be individually assessed in relation to SiN perception. Therefore, it was desirable to take a step back to the cognitive literature to use a framework which can applied to SiN perception to these listener groups.

One well established model which meets these sets of criteria is the Baddeley and Hitch’s model of Working Memory (Baddeley, 2000; Baddeley et al., 1974).

Baddeley's (2000) Working Memory model consists of four sub-domains: the Central Executive, Visuo-Spatial Sketchpad, Episodic Buffer and Phonological Loop – shown below in Figure 3. 2. The Central Executive is mainly involved in information manipulation and governs the three memory sub-domains (involved mainly in storage) to allow attentional directed task execution. The Central Executive domain includes multiple executive processes, such as Inhibitory Control, in additional to attentional processes. The Phonological Loop is responsible for storage and rehearsal of verbal information. The Episodic Buffer is an amodal storage sub-domain, which holds information in episodes and is capable of integrating information from multiple sources and modalities. The Visuo-Spatial Sketchpad relates to storage of information in the visual domain and will not be discussed any further because here SiN perception is investigated only in the verbal domain in this thesis, i.e., in the absence of any visual information or cues. It is noted that in naturalistic conditions visual speech information, e.g., lip and mouth movements, is available and could engage the Visuo-Spatial Sketchpad.

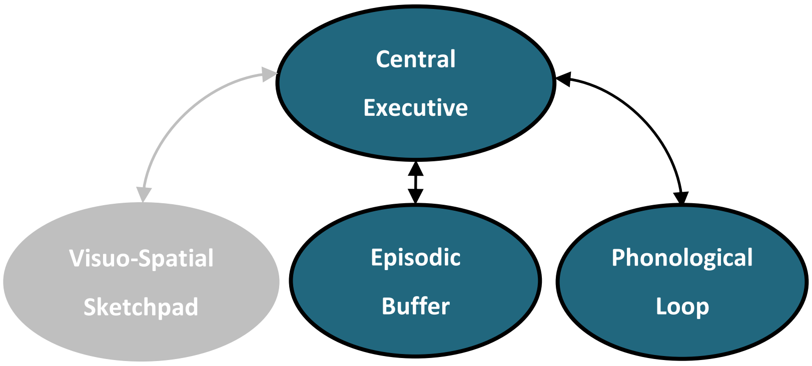


Figure 3. 2 – Flow chart of the Baddeley model of working memory

The model Baddeley consists of a Central Executive component (amodal), responsible for manipulation of information and governance of the memory sub-domains, and three memory sub-domains responsible for amodal (Episodic Buffer) or modality specific (Phonological Loop – verbal, Visuo-Spatial Sketchpad – non-verbal) storage of information. Here the visuo-spatial memory component is theorised to be irrelevant for SiN perception and appears in the figure only for completeness. This figure is adapted from Figure.I in Baddeley's (2000) model of working memory.

One limitation of the Baddeley model is that the Central Executive sub-domain does not differentiate between executive and attentional processes. Given the potentially differing role for executive and attentional processes for SiN perception it is necessary to also consider an established cognitive model, which differentiates between executive and attentional processes while acknowledging the storage aspects of the Baddeley model.

The Diamond model of Executive Functions (Diamond, 2013) relates storage aspects of Working Memory (sub-dividing verbal and non-verbal Working Memory) to Inhibitory Control, and also to cognitive flexibility and higher-level Executive Functions. Figure 3. 3 below shows a schematic of the Diamond model.

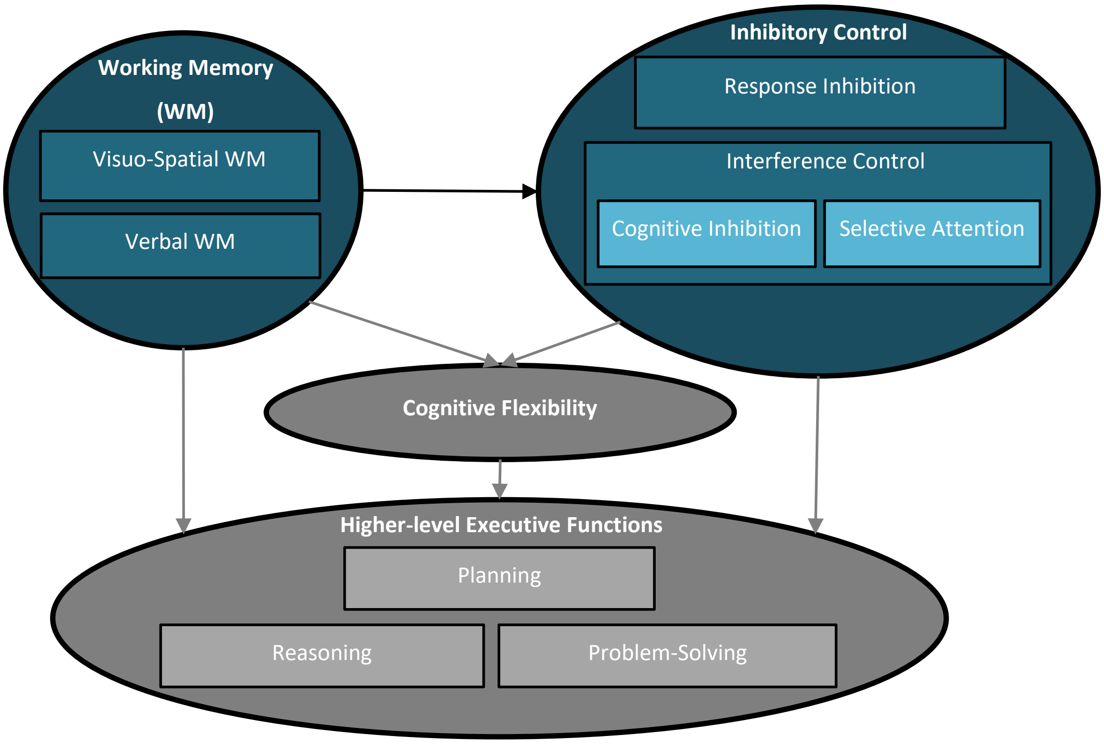


Figure 3. 3 – Flow chart of the Diamond model of executive functions

The Diamond model of executive functions, consists of four sub-domains, working memory, inhibitory control, cognitive flexibility and higher-level executive functions. Grey shaded sub-domains are deemed to be less important or irrelevant for SiN perception tasks. This figure is adapted from Figure 5 from Diamond's (2003) executive functions review.

Here I will specifically focus on the Working Memory (storage and manipulation) and Inhibitory Control (manipulation) aspects of the model because they are most relevant to the putative cognitive resources required for SiN perception tests and they most clearly relate Diamond's model to Baddeley's model. In the Diamond model, Working Memory is further divided into visuo-spatial and verbal sub-domains. Here it is specifically verbal Working Memory that is of interest for SiN perception. Inhibitory Control comprises two sub-processes, Response Inhibition and Interference Control. Response Inhibition refers to inhibition at the level of selfcontrol, whereas Interference Control involves the inhibition of thoughts/memories (cognitive inhibition) and attentional inhibition (selective attention).

Although the Baddeley and Diamond models are not directly geared towards investigating the association between cognition and SiN perception they do have a direct application in defining cognitive frameworks from which specific hypothesis for SiN perception can be made, for example, the ELU model (Rönnberg et al., 2013; Rönnberg et al., 2008). Furthermore, the overlap in cognitive processes defined between these models and the ELU model, e.g., Working Memory, executive processes, an Episodic Buffer, and phonological storage, enable general comparisons to be made. Therefore, using this combined approach allowed me to extent our understanding of the association between cognition and SiN perception taking specific focus on normal hearing/pre-clinical HL listeners.

3.1.2 Hypotheses

In assessing the role of specific cognitive abilities for SiN perception in different listening conditions in younger and older adults, I will explore the following research questions:

1) Which cognitive abilities are important for SiN perception?
2) Does the role of specific cognitive abilities differ depending on SiN listening condition?
3) Do these roles differ depending on age group and hearing sensitivity?

Baddeley model

In relation to the Baddeley model sub-domains, the Central Executive may be expected to be of general importance to SiN Perception (Binder et al., 1994; Heald et al., 2014), but perhaps to different extents depending upon listening condition, e.g., more complex target signal such as sentences, and maskers with informational properties. Furthermore, the relationship may depend on the specific executive processes (attention or inhibition) to be measured. Attentional executive processes may be particularly important for SiN listening conditions with more linguistically complex targets because those listening conditions may require higher levels of sustained attention for a greater amount of time. Inhibitory Control executive processes may be of general importance in SiN perception due to their role in signal restoration (Mattys et al., 2012), but may also be further engaged in conditions of informational masking (Janse, 2012).

The Episodic Buffer, synonymous with short-term Episodic Memory storage, is expected to be of particular importance for more linguistically complex target signals, as they tend to further engage storage capacity (Goldinger, 1996; Rönnberg et al., 2008).

The Phonological Loop, which requires both storage and rehearsal of verbal information, is expected to be of general importance for SiN perception because it requires both storage and rehearsal of verbal information. In addition, it may be particularly important for more linguistically complex stimuli and for listening in informational masking as it is conceivable that the conditions rely on verbal storage and rehearsal processes most.

General cognitive ability may show a general increased association with SiN perception in older populations, who may compensate some of their cognitive and sensory (including hearing) declines with an increase in cognitive processing (Roberts et al., 2016). Specific abilities such as attention and Episodic Memory have been associated with increased neuronal activity in older versus younger adults using imaging techniques (Cabeza et al., 1997; Grady et al., 2000; Reuter-Lorenz et al., 2008). Therefore, abilities such as attention (Central Executive) and (verbal) Episodic Memory (Episodic Buffer and Phonological Loop) might be expected to show a greater association with SiN perception in the older compared to the younger adults. Additionally, these differences may be more pronounced in the perception of more complex signals, such as sentences, and these differences may be mediated by hearing sensitivity (Peelle et al., 2011). Furthermore, in the presence of HL, SiN perception of informationally masked speech may lead to increased disruption at the level of storage and rehearsal (Episodic Buffer and Phonological Loop), and this may have a knock-on effect at the level of Executive Function (Central Executive) due to signal degradation and/or a lack of availability of cognitive resources. Alternatively, age-related cognitive decline may lead to declines in Executive Functions, which would lead to difficulties directing attention to specific targets (speech), inhibiting unwanted distracting noises (informational masking), and in restoration of a degraded signal (energetic and informational masking), causing a general decline in SiN perception, but perhaps to a greater extent in informational masking.

Diamond model

(Verbal) Working Memory is expected to be important for all listening conditions (Akeroyd, 2008) but perhaps to a greater degree in informationally masked speech, due to increased executive/attentional and/or storage demands.

As mentioned in the previous section Inhibitory Control executive processes may be of general importance in SiN perception due to their role in signal restoration and may be further engaged in conditions of informational masking.

Furthermore, Working Memory ability may play a greater role for SiN perception in older or HL populations, and may play no role for normal hearing younger adult listeners (Füllgrabe et al., 2016b). Both Working Memory and Inhibitory Control also link to the compensation hypothesis where older adults show greater cognitive engagement due to a compensatory mechanism (Reuter-Lorenz et al., 2008). Additionally, this may effect SiN perception disproportionally depending on listening condition. For example in conditions such as informational masking, which may require further engagement of Working Memory and/or inhibitory abilities, particularly for older and HL listeners who might expend a greater listening effort (Rönnberg et al., 2013).

3.2 Methods

3.2.1 Testing materials and procedure

Participant demographics and pre-testing

Participants were self-reported to meet the following criteria: 1) Aged between 18-30 years (younger adult group) or ≥60 years (older adult group), 2) Native English language speaker, 3) No diagnosed hearing loss, 4) No diagnosed neurological, psychiatric or language (including dyslexia) disorders or impairments. If all the criteria were met they were invited to take part in the study.

Younger adults

Fifty younger adult participants, with a mean age of 22.7±2.9, took part in the study (36 female and 14 male). Participants were recruited from the University of Nottingham and wider local population (Nottinghamshire) via electronic and paper advertisements. Electronic advertisements were placed on social network website (Facebook) groups and on the website callforparticpants.com. Paper advertisements were displayed in various buildings across the University Park campus of the University of Nottingham and also in a local supermarket in Beeston, Nottinghamshire.

Older adults

Fifty older adult participants, with a mean age of 69.4±5.4, took part in the study (32 female and 18 male). Participants were recruited from the local population (Nottingham and surrounding areas) via electronic and paper advertisements.

Sample sizes for both age groups (100 total, 50 younger and 50 older adults) were determined in advance based on several factors relating to the statistical approaches used in this study. For principal components analysis a minimum sample size to variable ratio of 10:1 is recommended as a rule of thumb (Nunnally, 1978). By sampling 100 participants for 5 cognitive sub-domains I exceed this ratio. For correlation analysis at a power of 0.8 to find significant two-tailed correlations at r=.4 (a value commonly seen in the literature) a minimum sample size of 44 participants is required (G*Power (v3.0.10)). My combined and separate age group samples sizes meet or exceed these criteria.

Pre-testing

Upon arrival, prior to any experimental testing, participants completed several screening tests for potential confounding factors. They included: a medical questionnaire (see Supplementary Figure 3. 1 in the Appendix), a colour vision test (Fletcher, 1998), a near vision acuity test (Bach, 2007), hearing sensitivity (pure tone audiometry), and the Mill Hill vocabulary test. The older adults also underwent three additional tests, the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005), an ear canal examination, and tympanometry. See Table 3. 1 below for a list of the screening tests and descriptions.

A medical questionnaire was included as a part of the screening process to control for any potentially cofounding factors including, neurological or psychiatric disorders, language difficulties (including dyslexia), hearing disorders (including tinnitus), and non-native English speakers.

To assess colour blindness a colour vision screening test, part one of the City University Colour Vision Test (Fletcher, 1998), where participants were asked to identify the presence and position of differently coloured spots. A Score of 9 or 10 (out of 10) was considered as normal.

A near vision acuity test, acuity C (Landolt-C) test from the Freiburg Visual Acuity and Contrast Test, version 3.9.3 (Bach, 2007) was selected to ensure participants had normal or corrected to normal version. The test was set to an eight-alterative forced choice procedure, over a total of 18 trials. Participants were sat at a viewing distance of 150cm, which allowed a maximum Visual Acuity decimal of 1.47. Visual acuity greater (better) than .3 was considered within normal range as defined in the International Classification of Diseases (ICD-10) (WHO, 1992)

Hearing sensitivity was assessed using an Interacoustics® AT235 impedance audiometer (for 43 participants) or a Grayson-Stadler® GSI 16 Audiometer (57 participants) and TDH-50P Telephonics® headphones (note: two audiometers were used due to a change in testing equipment). Nine frequencies between 0.25 and 8 kHz were assessed following the British Society of Audiology recommended procedure guidelines (BSA, 2011).

The Mill Hill vocabulary test (Raven et al., 1982) was selected to assess language ability, there was no exclusion criteria for this test.

Test	Description
Eligibility assessment	
Medical questionnaire	*(see Supplementary Figure 3. 1 in the Appendix)*
Colour vision screening test - part one of the City University Colour Vision Test (Fletcher, 1998)	*Identification of differently coloured spots, over a total of 10 trials.*
Near vision acuity (Landolt-C) test - Freiburg Visual Acuity and Contrast Test, version 3.9.3 (Bach, 2007)	*An eight-alterative forced choice procedure, over a total of 18 trials.*
Hearing sensitivity (Pure Tone Average)	*Nine frequencies at octave intervals between 0.25 and 8 kHz were assessed following the British Society of Audiology recommended procedure guidelines (BSA, 2011).*
Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) – older adults only	*Mild cognitive impairment and dementia screening test, which assesses multiple cognitive domains (visuo-spatial abilities, language, short-term memory, attention, Executive Function)*
Ear canal examination – older adults only	*Visual inspection (using an otoscope) of ear canal and ear drum for abnormalities.*
Tympanometry – older adults only	*Assessment of middle ear function.*
Auxiliary assessment	
Mill Hill vocabulary test (Raven et al., 1982)	*Identification of the correct synonym of a target word in a six-alternative multiple-choice format, over a total of 20 trials.*

Table 3. 1 – Summary of eligibility and auxiliary tests

List of eligibility and auxiliary tests with a short description of each. Note, both adult groups were assessed with each test unless otherwise indicated.

In addition to the above older participants also underwent further tests. The Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) was selected to test cognitive status, a score of ≥23(/30) was needed for inclusion (Luis et al., 2009), after years of education adjustment (for scores <30, 1 point is added if number of years in education is <12 years).

An ear canal examination (using a Keeler® Standard otoscope) and tympanometry were used to measure middle ear function using a GSI® Tympstar Middle ear analyser, following the British Society of Audiology recommended procedure guidelines (BSA, 2013).

All scores across all tests were considered within normal range and all tested participants were included in the analysis. However, some of the older adults (n=15) did fall within the mild hearing loss range (20-39 dB HL 0.25-4kHz), but were still included in the analysis.

Speech-in-Noise tests

Sentences

The sentence stimuli were taken from the British English Semantic Sentence Test (BESST) (Heinrich et al., 2014), which itself is based on the SPIN-R test (Bilger et al., 1984). The test included 96 sentence pairs, with each pair ending in the same monosyllabic word, but varying in the ease with which the final word could be predicted from the preceding part of the sentence. For example, *'Daisy wore a helmet on her <u>head</u>'* was the high predictable (HP) sentence and *'Daisy saw a feather on her <u>head</u>'* the low predictable (LP) counterpart. Sentences pairs were matched for duration, stress pattern and intonation. Sentences were recorded using a male speaker with a Standard Southern British English accent. Only one item from each sentence pair was selected for experimentation, in order to avoid repetition of target words. See Supplementary Figure 3. 2 in the Appendix for a list of the sentence stimulus set.

Words

The word stimuli consisted of 56 monosyllabic words (16 practice trials, and two blocks of 20 experimental trials), with no word repeated from the final words in the sentences. The stimuli were recorded using a male speaker (different to the sentences) with a Standard Southern British English accent. See Supplementary Figure 3. 3 in the Appendix for a list of the word stimulus set.

Maskers

The 3-talker babble (3B) was created using the phonetics software Praat® (v5.4.06) by combining the voices of three talkers reading a different text passage. Two talkers were female and third was male. All of them spoke English with different regional English accents. The 3B masker functioned as an informational masker for this study. The speech-modulated noise (SMN) was created in Matlab® (R2014b) by averaging the babble signal in chunks of 23msec. This preserved the long term average spectrum of the signal as well as the overall signal envelope, yet made the sound completely unintelligible. Thus, this masker signal functioned as an energetic masker. All sound signals were low- and high-pass filtered between 50Hz and 10,000Hz. The target and masker stimuli were combined in Matlab® with the noise stimuli starting and finishing 2 seconds before and after the speech target.

Signal-to-noise ratios

The 3B masked conditions (sentences and words) were presented at a fixed Signal-to-Noise Ratio (SNR) of -2dB for the younger adult group and a fixed SNR of 0dB for the older adult group. The SMN masked conditions (sentences and words) were presented at a fixed SNR of -7dB for the younger adults and a fixed SNR of -4dB for the older adult group. The SNRs were chosen to approximate 50% intelligibility thresholds for both age groups across all three target stimulus types (for word stimuli and averaged between LP and HP sentences). Different SNRs were required across the two age groups to approximate the same level of intelligibility. The SNR levels were based on the pilot data for the younger adults (not shown here) and based on a study using similar stimuli for the older adults (Heinrich & Knight, 2016). Intelligibility levels were matched to ensure audibility of target stimulus were as close as possible between the groups. A mismatch in intelligibility levels may lead to differences in the engagement of cognitive processes for SiN perception.

Procedure

In all the SiN perception tests participants were instructed to listen carefully to each target stimulus, whole sentences or single words, and to repeat back the stimulus as best as they could. Participants were also encouraged to guess on any words if they were unsure and were advised that they would not be expected to be able to recall every word or sentence due to the difficulty of the task. In the sentence conditions only the final word was scored, and for both sentence and word conditions the target word was scored either correct or incorrect (i.e., any response other than the exact target word was considered to be incorrect). For each participant a proportion correct score was determined for each of the six SiN conditions (HP sentences in speech-modulated noise, LP sentences in speech-modulated noise, single words in speech modulated noise, HP sentences in 3-talker babble, LP sentences in 3-talker babble, and single words in 3-talker babble).

Speech tests were presented in four blocked conditions: sentences in 3B, sentences in SMN, words in 3B, and words in SMN. Prior to each block participants completed practice trials (8 stimuli presentations per practice block) at an SNR of +5dB. This was done for the participant to familiarise themselves with the test.

Sentence blocks consisted of 40 sentence stimuli, 20 HP and 20 LP sentences. Word stimuli were blocked into sets of 20 stimuli. No final sentence words were repeated across LP/HP sentences or single word conditions within testing for individual participants and stimulus target blocks were fully counterbalanced according to a randomised Latin square design across all participants.

Cognitive tests

Cognitive tests were selected to fit with Baddeley's and Diamond's models. Multiple cognitive tests were selected to assess each cognitive ability, for which a single component representing the sub-domain was derived using Principal Components Analysis (PCA).

Here EFA (Exploratory Factor Analysis) was considered ahead of the PCA method because EFA extracts only common variance between observed variables and allows for the construction of latent factors. In only extracting common variance EFA accounts for conflation of task- and ability-specific variance – the impurity principle (Surprenant et al., 2009). Whereas PCA extracts the maximal amount of variance of observed variables and thereby does not account for task- and ability-specific variance.

However, after preliminary EFA was conducted it was deemed to be an inappropriate approach for this dataset and subsequent linear mixed model analysis. This was due to a combination of factors. Firstly, generally at least three observed variables are usually needed to be obtained for a factor to be considered reliable and stable (Costello et al., 2005). However, an exception can be permitted in the case of two observed variables per latent factor if the correlation between the observed variables within the factor is >.70 (Yong et al., 2013). There were two cases where only two observed variables were obtained for a single factor (Episodic Buffer and Phonological Loop). However, the correlations were not found to meet the r>.70 criteria and therefore could not be consider reliable. Secondly, factor scores derived using EFA techniques face a problem of indeterminacy. Where there is no unique solution for deriving factors scores and thereby an infinite number of solutions is possible, all of which would be equally valid (note: there are methods for dealing with indeterminacy – see (Devlieger et al., 2016)). This would be problematic for when applying individual factor scores as individual predictors of cognitive performance in the main analysis (assessing the role of cognitive ability for SiN perception) in this study. PCA does not encounter this problem since principle components have a unique factor solution.

It is recognised that the PCA method does not account for the impurity principle. However, I attempted to remedy through a highly theory-driven test selection for each theorised principal component/cognitive sub-domain and applied a strict inclusion criteria in retaining observed variables for each of the principal components. In addition previous studies in the literature have applied PCA methods to reduce cognitive ability and/or hearing sensitivity measures into principal components (Brannstrom et al., 2012; Heinrich et al., 2015; Heinrich, Henshaw, et al., 2016; Humes et al., 1994; Kidd et al., 2007).

Table 3. 2 below provides a summary of each of the cognitive tests and the respective Baddeley and Diamond model domains they correspond to.

Cognitive test	**Baddeley model (2000) sub-domain**	**Diamond model (2003) sub-domain**
Test of Everyday Attention, subtest 1	Central Executive	Inhibitory Control
Test of Everyday Attention, subtest 6	Central Executive	Inhibitory Control
Test of Everyday Attention, subtest 7	Central Executive	Inhibitory Control
Stroop test	Central Executive	Inhibitory Control
Reading Span Test	Central Executive	Working Memory
Letter-Number Sequencing	Central Executive	Working Memory
Corsi Span Forward	Episodic Buffer	N/A
Corsi Span Backward	Central Executive	Working Memory
Digit Span Forward	Episodic Buffer	N/A
Digit Span Backward	Central Executive	Working Memory
Word list recall	Episodic Buffer	N/A
Rhyme verification task - condition 1 (R+O-)	Phonological Loop	Inhibitory Control
Rhyme verification task - condition 4 (R-O+)	Phonological Loop	Inhibitory Control

Table 3. 2 – Summary of theorised cognitive abilities assessed by each cognitive test

List of cognitive tests and corresponding Baddeley and Diamond model sub-domains. Note, the Corsi Span Forward, Digit Span Forward, and Word list recall tasks corresponded only to a sub-domain of the Baddeley model (Episodic Buffer).

Test of Everyday Attention

The map search, subtest 1 (TEA1), the telephone search, subtest 6 (TEA6), and telephone search dual task, subtest 7 (TEA7) (Robertson et al., 1994), were used to assess the attention aspect of the Central Executive sub-domain in the Baddeley model and the Inhibitory Control sub-domain for the Diamond model.

In TEA 1, participants were asked to find symbols on a map within a set amount of time. The score was the number of symbols found (out of a possible 80) in 1 and 2 minutes. Only the number of items found in 1 minute will be reported due to ceiling effects in the 2-minute search task found during the pilot experimentation in the younger adult group.

In TEA 6, participants were instructed to look for key symbols (two matching circle, square, cross or star symbols) while searching entries in a simulated classified telephone directory for a specified target, e.g., plumbers. Participants were instructed to work as quickly and accurately as possible. Additionally, participants were instructed to only look at each entry in the phonebook column once to prevent items from being examined multiple times. Participants were scored on time taken per item found, i.e., the number of correctly identified items divided by the total time taken to complete the task.

In TEA 7, participants were again asked to search in the directory for key symbols, but with the additional task of simultaneously counting strings of tones. The tone stimulus and recorded instructions were played to participants on a loudspeaker at a comfortable listening level. The performance score was calculated by taking the time per item and dividing it by the proportion of tone strings correctly stated. Participants were scored on time taken per item found, i.e., the number of correctly identified items divided by the total time taken to complete the task.

Stroop test

The Stroop test was selected to assess inhibitory aspects of the Central Executive sub-domain of the Baddeley model and the Inhibitory Control sub-domain for the Diamond model. In a variation of the original colour-word interference test (Stroop, 1935) participants were presented with 8x6 grids printed on A4 paper. Each page was in turn placed flat on a desk in front of the participant at a viewing distance of approximately 50cm. There were three conditions, word reading, neutral colour naming and incongruent colour naming (see Figure 3. 4 below for an example of grids for each condition). In the word reading condition each tile in the grid contained the words, ‘red’, ‘blue’, ‘green’ or ‘brown’ printed in black text size 20 sans-serif font and placed in the centre of a white tile. The task was to read each word as quickly and as accurately possible. In the neutral colour naming condition, each tile within the grid contained ‘XXXX’ printed in size 20 sans-serif font and placed in the centre of a coloured tile (red, blue, green or brown). In the incongruent colour naming condition, the tiles contained a colour word incongruent to the colour of the tile, e.g., the word 'green' in a red tile. In colour naming trials (neutral and incongruent) participants were instructed to name the colour of each tile, whilst trying to ignore the text, as quickly and accurately as possible. Participants first run through a practice set of the three grids and then repeated two experimental trials. The overall time taken to read the colour words or name the background colours was measured. An average total completion time was taken for the two trials of each of the three conditions. The Stroop score was calculated by regressing the colour neutral time subtracted by the word reading time against the incongruent colour naming time (Ben-David et al., 2009). This was done because dimensional imbalance (differences in composite colour neutral and word reading times) has been proposed to vary with age-related changes in colour perception (Knight et al., 2017).

a.

green	blue	red	brown	red	brown	blue	green
brown	red	green	blue	green	blue	red	brown
blue	green	brown	green	blue	red	brown	red
green	blue	green	brown	red	blue	red	brown
blue	green	brown	red	brown	red	green	blue
red	brown	blue	green	blue	green	brown	red

b.

XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX
XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX

c.

green	blue	red	brown	red	brown	blue	Green
Brown	red	green	blue	green	blue	red	brown
blue	green	brown	green	blue	red	brown	red
green	blue	green	brown	red	blue	red	brown
blue	green	brown	red	brown	red	green	blue
red	brown	blue	green	blue	green	brown	red

Figure 3. 4 – Stroop test stimuli

Examples of the grids (a. word reading, b. neutral colour. & c. incongruent colour) used in the three conditions in the Stroop task (not shown to scale)

Reading Span Test

The Reading Span Test (RST) (Daneman et al., 1980) was selected to measure Working Memory ability as an aspect of the Central Executive of the Baddeley model and the Working Memory sub-domain of the Diamond model. In the RST participants were required to read out lists of unrelated, complex sentences displayed visually, in size 20 sans-serif font, on a computer screen approximately 50cm from the participant, while remembering the last word of each sentence for later recall. The number of sentences per trial increased by one every five trials, starting from two and ending with five sentences per trial; making 70 trials in total. Participants were instructed to read out loud each sentence and to continue to the next sentence immediately. As a further instruction, they were asked to wait for a recall prompt before repeating the final word of each sentence within each sentence block. The reading span score was taken as the proportion of total number of words correctly recalled from all 70 trials, following a partial credit scoring method was favoured over an all-or-nothing scoring method as recommended by Conway et al. (2005) to ensure and preserve comparability between span tasks. The same scoring methods was adopted for all span tasks.

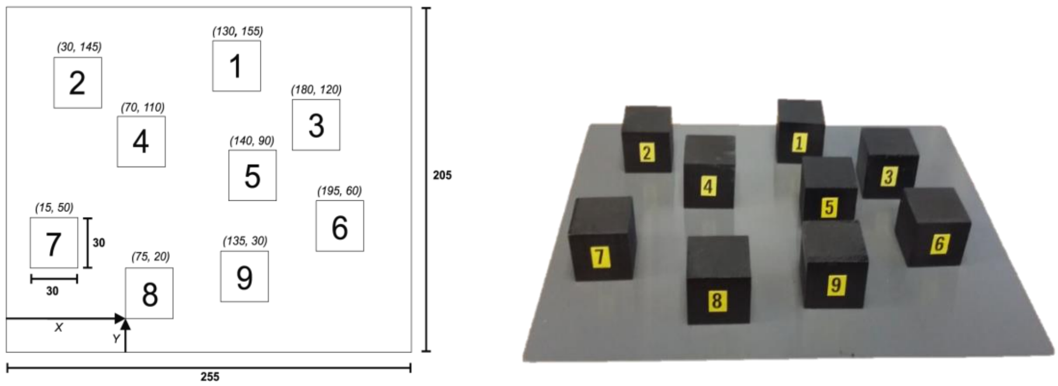
Letter-Number Sequencing

The Letter-Number Sequencing task (LNS) (Wechsler, 1997) was another measure of Working Memory selected to assess the Central Executive sub-domain of the Baddeley model in the Working Memory domain of the Diamond model. In the LNS task participants were required to listen to sequences of letter and number combinations. The task was to recall first the numbers in numerical order, then the letters in alphabetical order. Sequences began with two items and increased by one item every three trials ending with eight item lists. The letter-number recordings were spoken by a male talker with a Standard English accent. The items within each sequence were presented at an interval of 0.5 seconds. At the end of each sequence, as a prompt to begin recall, a 5kHz pure tone was played for 0.5 seconds. Sequence span was the proportion of correctly recalled items out of the total number of items on all letter-number lists. The correct recall was determined on an item level, with the correct item being required to be recalled at the correct serial position in each sequence.

Corsi block tapping task

The Corsi block tapping test (Corsi, 1972) was used to assess the Episodic Buffer of the Baddeley model (forward span) and the Central Executive and Working Memory sub-domains of the Baddeley and Diamond model respectively (backward span). It is recognised that the Corsi span tasks primarily engage the Visuo-Spatial Sketch subdomain of Working Memory. However, the Corsi span tasks may also engage the binding component of the Episodic Buffer subcomponent, particularly during the recall of longer spatial sequences, which may exceed the limited capacity of the Visuo-Spatial Sketchpad (Kessels et al., 2015). In the case of the backward span the additional manipulation component of re-ordering the to-be-recalled items in reserve order is thought to also engage the Central Executive subdomain of the Working Memory, although this is disputed (Kessels et al., 2008).

The Corsi test was conducted using a physical board (255mm x 205mm) with nine 3-dimensional blocks (3mm x 3mm x 3mm), the digits 1 to 9 were printed on one side of the blocks visible only to the experimenter. The design specifications were based upon those listed in Kessels et al. (2000) (*see Figure 3. 5 below*). The task was to repeat sequences of block taps as demonstrated by the experimenter. The experiment began with two-tap sequences and increased by one tap every second trial ending with eight- or nine-tap sequences. The Corsi test had a forward and a backward condition, Corsi Span Forward (CSF) and Corsi Span Backward (CSB). In the forward condition the participant was required to repeat the tap sequence in the correct serial order (to a maximum of a nine-tap sequence); in the backward condition the sequence was to be repeated in reverse serial order (to a maximum of an eight-tap sequence). Blocks were tapped with the index finger at a rate of one per second (with no pauses between individual blocks) and no single block was tapped twice in any single sequence. To aid in administrating each block tap sequence the experimenter listened to a verbal recording, via headphones (to ensure the participant could not overhear the recording), of each digit sequence (corresponding to each individual block-tap sequence). In all sequences the experimenter tapped the appropriate block when the digit was heard on each recording. Digits were spaced at one second intervals to ensure the blocks were tapped at an approximately constant rate. The recorded verbal digits were heard at an intensity of 60dB SPL. The sequence ordering for both the forward and backward conditions followed the sequences outlined in the appendix of Claessen et al. (2015). All sequences were always presented to each participant and scores were calculated separately for the forward and backward conditions by dividing the number of correctly recalled items by the total number of items presented.



a. *b.*

Figure 3. 5 – Corsi test stimuli

a, design specification of blocks and boards, measurements are stated in mm (image take from Kessels et al. (2000), b, photographic image of Corsi board and blocks – note the printed digits were only visible to the experimenter.

Digit span

The Forward (DSF) and Backward (DSB) digit span tasks (Wechsler, 2008) were used to assess the Episodic Buffer and the processing aspect of the Central Executive of the Baddeley models respectively and the DSB was also used to assess the Working Memory sub-domain of the Diamond model. Although it is recognised that digit span tasks engage the Phonological Loop subdomain (via storage and rehearsal) of Working Memory, they also engage the Episodic Buffer subdomain due to the binding and representation of information in serial chunks for recall. While the DSB has some emphasis on storage, it also has a manipulation component, hence its use as a marker for the Central Executive subdomain. In contrast, the DSF has no manipulation component and instead only contains a memory component, hence its use as a marker for the Episodic Buffer.

In both tasks participants were required to listen to orally presented strings of numbers; in the forward task, numbers were recalled in the order they were presented. In the backward condition the numbers were recalled in the reverse order. In the forward digit test lists began with two numbers, increasing by one number every two trials, to a maximum of eight number per list. The backward span followed the same protocol, but there were four two-number lists and the test ended with a span of seven numbers. The digits were taken from the same male speaker recordings used in the LNS task. The items within each sequence were presented at an interval of 0.5 seconds. To signal the end of each sequence and the beginning of recall a 5kHz pure tone was played (duration 0.5 seconds). The span scores for the forward and backward tests were calculated as proportion of correctly recalled items out of the total number of items on all number lists; separate scores were given for the two test. An item was scored correct only if it was recalled at the correct serial position within a given span.

Word list recall task
The word list recall task was used as a measure of the Episodic Buffer of the Baddeley model. Participants were required to listen to word lists and to immediately recall a word list after its presentation. The participants were instructed to repeat the words in any order. The word stimuli were taken from the AB word list (Boothroyd, 1968). In total, there were seven word lists, each word list contained eight words. Words were presented with a 2 second interval between them. Proportion of words recalled correctly out of the total number of words heard was calculated as an outcome score.

Rhyme verification task
A rhyme verification task (Johnston et al., 1986) was chosen as a measure of the Phonological Loop sub-domain in the Baddeley model and the Inhibitory Control domain (incongruent word pair conditions only, i.e., does rhyme/orthographically different and no rhyme/orthographically similar) in the Diamond model. The rhyme verification task is considered to assess the Phonological Loop because it requires the short-term storage of verbal information. It differs from verbal span tasks, e.g., the digit span, in that only two items are held at a time in the storage component of the Phonological Loop. Therefore, the storage capacity of the Phonological Loop is less likely to be exceeded and there is minimal requirement for information to be binded and held in the Episodic Buffer. Additionally, the incongruent conditions of the rhyme verification task requires suppression of conflicting information due to the mismatch in phonological and orthographic properties.

Participants were required to judge if visually-presented word pairs rhymed or not, indicating their response with a key press (left or right arrow keys). Participants were instructed to respond as quickly and accurately as possible. Word pairs were displayed side by side on a computer monitor. The test contained four stimulus type presentations with 20 stimulus pairs per group: does rhyme/orthographically dissimilar (R+O-) e.g., ‘make’ and ‘ache’, does rhyme/orthographically similar (R+O+), e.g., ‘fall’ and ‘tall’, does not rhyme/ orthographically dissimilar (R-O-), e.g., ‘milk’ and ‘land’, and does not rhyme/ orthographically similar (R-O+), e.g., ‘cost’ and ‘post’. The stimulus presentation order was randomised and programmed breaks were encouraged throughout the task. Rhyming ability was determined by proportion correct response for each individual condition. Only the R+O- (Rhyme 1) and R-O+ (Rhyme 4) conditions will be used in the analysis due to expected ceiling effects in the congruent conditions based on results from Johnston & McDermott's (1986) paper, which were and confirmed in the current study – *see Figure 3.6 below*.

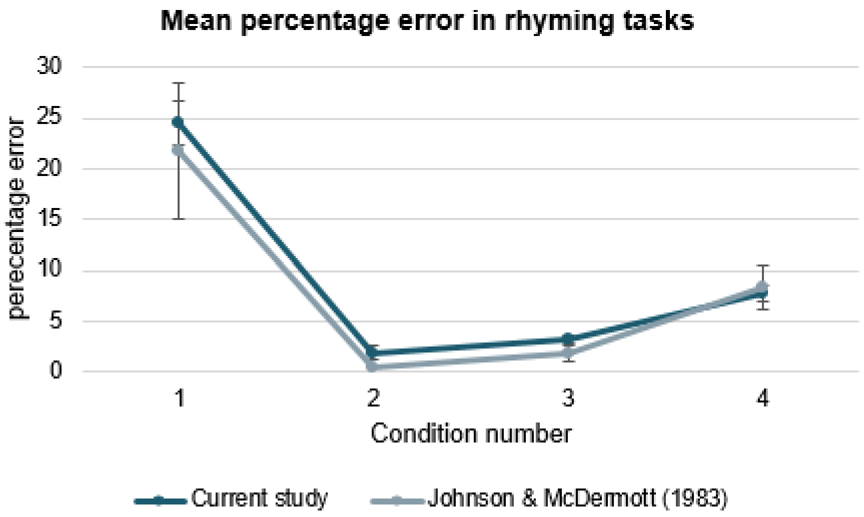


Figure 3. 6 – Mean percent error in rhyme verifiication tasks

Mean (and standard error) percentage errors for the four rhyme conditions (1 = R+O-, 2 = R+O+,3 = R-O-, 4 = R-O+), for the current study and for Johnson & McDermott (1983): control group of experiment one.

Experimental Procedure

All testing was conducted in a sound-attenuated booth. For all participants screening tests were administered first, followed by SiN perception tests and lastly cognitive tests. Separately for the cognitive and speech perception tests the experimental order was counterbalanced between participants using a fully randomised Latin square design. The total testing time per participant was approximately 2.5 hours for the younger adults and 3 hours for the older adults. Testing was always completed in a single visit for both age groups.

All auditory stimuli were presented monaurally to the left ear using over-the-ear headphones (Sennheiser® HD 280 pro 64Ω).

In the younger adult group, all auditory stimuli were set at 60dB SPL. This level was chosen because it is within the range of typical conversation levels of speech, in quiet or low noise conditions, at a distance of 1m (Olsen, 1998 ; Sengpiel).

In the older adult group, auditory stimuli were at individualised levels for each participant, 30dB SPL above their Speech Reception Threshold in quiet (mean SRT in older adults with normal hearing expected to 30dB SPL (Peters et al., 1998)). This was done in order to adjust for any auditory/periphery decline in the older compared to younger listeners and to create a level and comparable playing field in terms of SiN target audibility between the two listener groups.

The sentence stimuli in the SRT test came from the Adaptive Sentence List (ASL) sentence list (MacLeod et al., 1990). Participants were instructed to listen to each sentence and to then immediately verbally repeat back the sentence as best as they could. Each sentence contained three key words, all three key words needed to be identified correctly for a response to be scored as correct (respondents were not informed of this scoring system prior to or during the procedure). An adaptive one-up, one-down procedure was used to determine a 50% correct threshold. The initial intensity was fixed at 60dB, with an initial step size of 10dB, after one reversal the step size decreased to 5dB and after two reversals the step decreased to 2dB. The test included 10 reversals in total, the average threshold was calculated from the mean of the final 7 reversals. Thresholds were assessed in the left ear only. Participants (older group only) scored a mean SRT level of 30.6dB (SD=6.5).

The intensities of all auditory stimuli including speech and cognitive tests were verified using an artificial ear (Brüel and Kjær, type 4153). All tests that were electronically run were displayed on a Toshiba® satellite pro laptop with a 17inch, 1440 x 900 pixel, 60Hz monitor (43 participants) or Dell® 23inch, 1280 x 1024 pixel, 60Hz monitor (57 participants) using the software PsychoPy® (v2). An exception was the Reading Span Test, which was presented using the software Praat® (v5.4.06). Participants were sat approximately 60cm from the monitor during stimuli presentation.

3.2.2 Statistical methods

Statistical analyses were performed in IBM® SPSS® Statistics (v22) apart from the linear mixed modelling, which was performed using the lme4 (Bates et al., 2015) and lmtest (Zeileis et al., 2002) packages in RStudio (V1.0.153).

The speech test proportion response scores were close to floor and ceiling in some conditions. This may affect statistical power so they were converted to Rationalized Arcsine Units (RAU) (Studebaker, 1985) for further analyses. RAU scores extend upper and lower extremes of a psychometric functions making it more linear and thereby increasing statistical power.

All variables were assessed for normal distribution using the Shapiro-Wilk test of normality and by visual inspection of Q-Q plots.

For the SiN perception tests, normality was assessed after the RAU transformation and separately for the younger and older adult age groups. In the younger adults all but the words in speech-modulated noise condition was found not to have a normal distribution, as assessed by the Shapiro-Wilk statistic (W=.93, *p=.008*), where a significant p-value indicates the distribution is not normal. In older adults all but the words in 3-talker babble and HP sentences in 3-talker babble was found not to have a normal distribution, as assessed by the Shapiro-Wilk statistic (W=.94, p=.02 & W=.92, *p=.002*). However, after a visual inspection of the Q-Q plots and box plots it was deemed acceptable to bring all results forward for parametric testing without further transformation (see Supplementary Figure 3. 4 in the Appendix for Q-Q plots and box plots for the SiN perception tests, which had a significant Shapiro-Wilk statistic, i.e., did not have a normal distribution).

Statistical assessment of the cognitive tests was done as a combined group of young and old adults because cognitive test performance on all 14 tests across both age groups was reduced into principal components in the next step. Eight cognitive tests (TEA1 (W=.96, p=.049), TEA7 (W=.93, p=.012), Reading Span Test (W=.96, p=.0058), Corsi Span Backward (W=.97, p=.018), Corsi Span Forward (W=.97, p=.041), Word list recall (W=.97, p=.021), Rhyme verification task conditions 1 (W=.89, p<.0001) & 4 (W=.86, p<.0001)) were found not to have a normal distribution, as assessed by the Shapiro-Wilk statistic. Q-Q plots and box plots were also examined for the eight significant tests (see Supplementary Figure 3. 5 in the Appendix for Q-Q plots and box plots). Through the combined assessment via Shapiro-Wilks statistic, and Q-Q plot and box plot, a transformation was applied to the most extreme cases of non-normal distribution only, namely the Rhyme verification tasks (conditions 1 & 4). In both cases an inverse transform $X_t = 1/(C-X)$ (where X_t is the transformed score, C is the maximum score plus 1, and X is the original score) was used. All cognitive test measures were converted into standardised Z-scores (x – mean / standard deviation) prior to the Principal Components Analysis (PCA).

Using a single factor solution PCA was to extract a single component for each group of cognitive tests theorised to assess a specific sub-domain of the two cognitive models. The cognitive variables were weakly/moderately correlated, however because only one component was extracted no rotation was applied (see Supplementary Figure 3. 6 in the Appendix for correlation matrix of cognitive variables). Two rounds of PCA were performed to determine the makeup of each principal component. In the initial round, all cognitive tests selected to load on a respective sub-domain were entered. In the second round, only the tests which showed a communality of >.5 in the initial round were included. Hence, on final principal components all included surface tests showed a loading of at least .5. Based on this final principal component factor scores for each participant and for each cognitive sub-domain were obtained. Table 3. 4 in the results section displays the communalities and variance explained for each principal component/cognitive sub-domain, of the Baddeley and Diamond models, in rounds one and two of the PCA.

Next, I examined the predictive effects of each cognitive principal component for SiN intelligibility (as a proportion correct response (expressed in RAUs) individually for each participant for each SiN condition, i.e., not per SiN target item) by fitting mixed linear models, via maximum likelihood estimation. Separate models were run for each cognitive model (Baddeley and Diamond). As a first step models were run combined for the two age groups, inputting age group as a categorical variable. In a second step separate models for age group, younger and older adults, were run for each of the two cognitive models. This resulted in six sets of linear models in total: 3 (both age groups, younger adults only, older adults only) x 2 (Baddeley, Diamond). In all models speech and cognitive variables were coded as fixed effects. Specifically, speech was coded as two categorical variables: speech type (words, LP, HP) and masker type (SMN, 3B). Cognitive variables were coded as three continuous variables (Baddeley model: Central Executive, Episodic Buffer, phonological, loop) or as two continuous variables (Diamond model: Working Memory and Inhibitory Control). The cognitive variables factors scores were obtained from the PCA. All models also contained a continuous PTA (0.25-8kHz) variable. The outcome variable for each model was SiN intelligibility expressed in RAUs.

A backwards step approach was taken to determine the final shape of the model. Random effects were modelled first, and fixed effects second. In the combined age models, random effects were included for speech target, masker type and age group. In the separate adult group analysis (SiN intelligibility for younger and older adults assessed separately) random effects were included for speech target and masker type. In all models the fit of the full model (all possible random effects or all possible fixed effects and their interactions) was assessed first. See Supplementary Figure 3. 7 in the Appendix to see each of the full models. Then, when testing random effects, one random effect term was removed at a time, to test whether their presence significantly improved model fit. Where a simpler model, without the term in question, was found to fit equally well as a more complex model (using a likelihood ratio test) then the simpler model of carried forward. Only random effects found to significantly improve model fit were retained.

A similar backwards step approach was then taken for fixed effects. First the full model (at the level of all four-way interactions) was tested, and then one interaction term at a time was removed. Where a simpler model, without the term in question, was found to fit equally well as a more complex model (using a likelihood ratio test) then the simpler model of carried forward. This process was performed separately at each level of interaction, four- three- and two-way. All main effects were retained regardless of fit. Additionally, no lower-effect contained within a retained higher-level interaction was removed. Main effects from each model are reported using an ANOVA (type III, using the Satterthwaite method for calculating the denominator degrees of freedom), and estimates of fixed effects are also reported.

3.3 Results

3.3.1 Cognitive tests

Descriptive results

Significant age group differences (one-way ANOVA) were found for the following tests: Mill Hill vocabulary test, TEA1, TEA6, TEA7, Stroop test, Reading Span Test, Digit Span Backward, Digit Span Forward, Corsi Span Backward, Corsi Span Forward, and Word list recall. Typically, the younger adults performed better than the older adults; one exception was the Mill Hill vocabulary test where the older adults performed best. No significant age group differences were found for the Letter-Number Sequencing test, Rhyme verification task 1, and Rhyme verification task 4. Table 3. 3 below displays the mean and standard errors for the raw cognitive test data, separately for younger and older adults and combined for both age groups. The table also displays the (one-way) ANOVA statistics for the age group comparisons.

Cognitive test	Younger adults (n=50)	Older adults (n=50)	Age group differences	Both age groups (n=100)
Mill hill vocabulary test (raw score / 20) *	14.48 (.28)	16.78 (.29)	F(1, 98)=31.68, p<.001	15.63 (.23)
TEA1 (targets per minute) *	51.26 (1.56)	32.16 (.99)	F(1, 98)=107.27, p<.001	41.71 (1.33)
TEA6 (time per target - seconds) *	2.34 (.07)	3.07 (.06)	F(1, 98)=57.28, p<.001	2.70 (.06)
TEA7 (time per target - seconds) *	2.68 (.11)	3.90 (.14)	F(1, 98)=45.93, p<.001	3.29 (.11)
Stroop interference (time - seconds) *	10.3 (.4)	14.9 (.8)	F(1, 98)=24.43, p<.001	12.5 (.5)
RST (proportion correct) *	.55 (.02)	.42 (.02)	F(1, 98)=10.04, p=.002	.48 (.02)
LNS (proportion correct)	.73 (.01)	.70 (.01)	F(1, 98)=2.79, p=.098	.72 (.01)
DSB (proportion correct) *	.73 (.02)	.64 (.12)	F(1, 98)=17.13, p<.001	.69 (.01)
DSF (proportion correct) *	.77 (.02)	.68 (.01)	F(1, 98)=13.61, p<.001	.73 (.01)
CSB (proportion correct) *	.83 (.01)	.73 (.01)	F(1, 98)=35.14, p<.001	.78 (.01)
CSF (proportion correct) *	.75 (.02)	.62 (.01)	F(1, 98)=42.70, p<.001	.68 (.01)
Word list recall (proportion correct) *	.66 (.02)	.48 (.01)	F(1, 98)=63.37, p<.001	.57 (.01)
Rhyme 1 (R+O-) (proportion correct)	.71 (.03)	.83 (.03)	F(1, 98)=3.39, p=.069	.76 (.02)
Rhyme 4 (R-O+) (proportion correct)	.91 (.01)	.95 (.01)	F(1, 98)=3.36, p=.070	.92 (.01)

Table 3. 3 **–** Summary of cognitive test performance and age group differences

Cognitive test means and standard errors, displayed separately for younger adults, older adults and overall for both age groups.
TEA: Test of Everyday Attention, RST: Reading Span Test, LNS: Letter-Number Sequencing, DSB: Digit Span Backward, CSB: Corsi Span Backward, CSF: Corsi Span Forward, DSF: Digit Span Forward, R+O-: does rhyme/orthographically dissimilar, R-O+: does not rhyme/orthographically similar
**denotes significant difference (p<.05) between younger and older adults (One-way between age group ANOVA, corrected for multiple comparison (Bonferroni))*

3.1.2 Principal components analysis

Cognitive tests were reduced to principal components separately for the two cognitive models (Baddeley and Diamond) – see Table 3. 4 below. The table shows the separate PCAs run for each of the respective selected sub-domains of the Baddeley model (Central Executive, Episodic Buffer, Phonological Loop) and the Diamond model (Working Memory and Inhibitory Control).

Baddeley model

For the Central Executive component only 3 of 8 cognitive tests loaded onto a single component. Those were the three sub-tests (1, 6 and 7) of the Test of Everyday Attention; they explained 78% of the variance of the principal component. The Stroop, the Reading Span Test, the Letter-Number Sequencing, the Digit Span Backward, and the Corsi Span Backward tests each had communalities lower than .50 so were not retained. For the Episodic Buffer component two of the three cognitive tests (Corsi Span Forward & Word list recall) were retained, accounting for 78% of the variance. The Digit Span Forward test narrowly missed the communality inclusion criteria and was therefore excluded. For the Phonological Loop component both the Rhyme verification tasks (1 & 4) were retained, each with communality values of .62, and explaining 62% of the variance.

Diamond model

For the Working Memory sub-domain 3 of 4 cognitive tests were retained. The three verbal Working Memory tests (Reading Span Test, Letter-Number Sequencing, and Digit Span Backward) had high communality values with the extracted factor, whereas the visual Working Memory test (Corsi Span Backward) not meeting the communality criteria. For the Inhibitory Control sub-domain 3 of 6 of the cognitive tests were retained (TEA 1, 6 & 7), explaining 78% of the variance. Note that this sub-domain was identical in the final selection of cognitive tests as the Central Executive component of the Baddeley model.

Cognitive model	Principal component	Cognitive test	PCA round one		PCA round two	
			Communality	Variance explained	Communality	Variance explained
Baddeley	Central Executive	TEA1	.64	40%	.72	78%
		TEA6	.68		.84	
		TEA7	.58		.78	
		Stroop	.00		*Test excluded*	
		RST	.18		*Test excluded*	
		LNS	.25		*Test excluded*	
		DSB	.42		*Test excluded*	
		CSB	.43		*Test excluded*	
	Episodic Buffer	CSF	.59	61%	.78	78%
		Word list recall	.77		.78	
		DSF	.47		*Test excluded*	
	Phonological Loop	Rhyme 1	.62	62%	.62	62%
		Rhyme 4	.62		.62	
Diamond	Working Memory	RST	.55	51%	.62	63%
		LNS	.58		.61	
		DSB	.66		.67	
		CSB	.24		*Test excluded*	
	Inhibitory Control	TEA1	.74	40%	.72	78%
		TEA6	.81		.84	
		TEA7	.76		.78	
		Stroop	.02		*Test excluded*	
		Rhyme 1	.03		*Test excluded*	
		Rhyme 4	.02		*Test excluded*	

Table 3. 4 – Summary of principal components analysis of cognitive test variables

Principal Components Analysis (PCA) for Baddeley and Diamond model sub-domains/principal components. Round one and round two communalities and total variances are displayed separately. Only cognitive tests with communalities greater than .50 were included in round two. TEA: Test of Everyday Attention, RST: Reading Span Test, LNS: Letter-Number Sequencing, DSB: Digit Span Backward, CSB: Corsi Span Backward, CSF: Corsi Span Forward, DSF: Digit Span Forward, Rhyme 1 (R+O-: does rhyme/orthographically dissimilar), Rhyme 4 (R-O+: does not rhyme/orthographically similar).

3.3.2 Speech-in-Noise tests

Group mean (and standard errors) SiN intelligibility scores (expressed in RAU) for each of the six SiN conditions are displayed for the younger adults, older adults and both adult groups in Table 3. 5 below.

	SiN condition					
	Speech modulated noise			*3-talker babble*		
	Words	*LP Sentences*	*HP Sentences*	*Words*	*LP Sentences*	*HP Sentences*
Younger adults	38.5 (1.4)	30.2 (1.8)	55.7 (2.3)	50.7 (1.9)	45.2 (2.4)	60.7 (2.2)
Older adults	45.5 (2.0)	42.3 (2.6)	69.7 (2.6)	51.7 (1.9)	53.0 (2.8)	73.3 (1.7)
Both age groups	42.0 (1.3)	36.3 (1.7)	62.7 (1.9)	51.2 (1.3)	49.1 (1.9)	67.0 (1.5)

Table 3. 5 – Summary of performance in SiN perception tests

Means and standard errors (in parentheses) for each of the six SiN listening conditions. Data are shown separately and in combination for the younger and older age groups. Performance scores are expressed in Rationalised Arcsine Units (RAU)

Using a 2 (masking condition) x 3 (speech target) x 2 (age group) ANOVA SiN intelligible was assessed. Main effects were found for masking condition ($F(1, 49)=26.46$, $p<.001$) (3B>SMN), speech target ($F(2, 48)=253.41$, $p<.001$) (HP sent > words > LP sent), and age group ($F(1, 49)=25.42$, $p<.001$) (older listeners > younger listeners).

Post-hoc analysis (Paired T-tests, two tailed, multiple comparison corrected) showed that for target type intelligibility was significantly higher for HP sentences compared to single words ($t(99)=15.30$, $p<.001$) and LP sentences ($t(99)=22.26$, $p<.001$), and for single words compared to LP sentences ($t(99)=3.85$, $p<.001$).

Two-way interactions were found for speech target and masker ($F(2, 48)=12.73$, $p<.001$) (see *Figure 3. 7*) and speech target and age group ($F(2, 48)=9.76$, $p<.001$), (*see Figure 3. 8*).

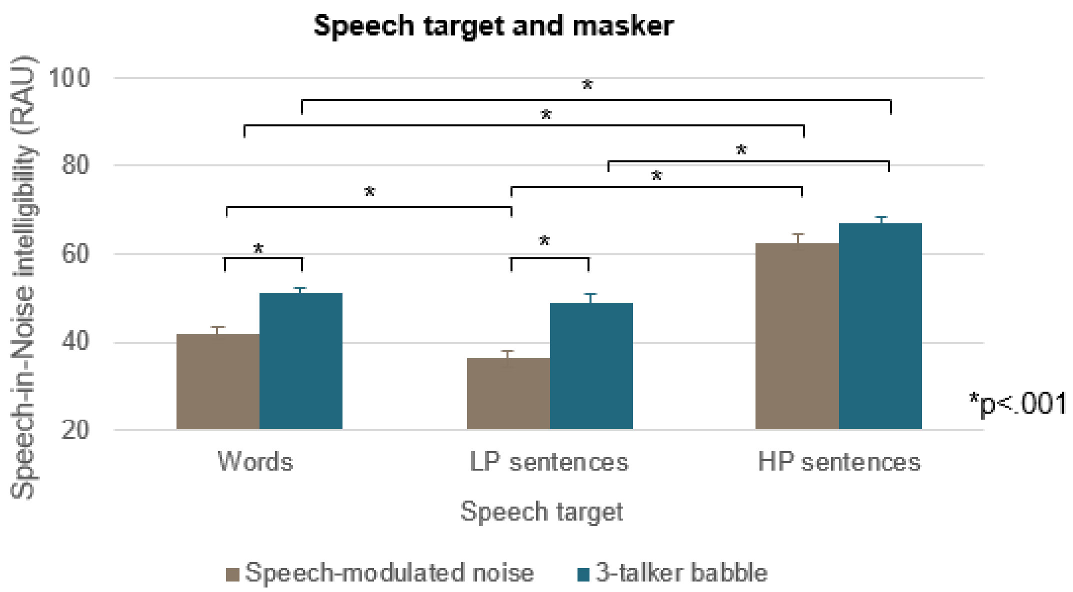


Figure 3. 7 – Bar chart of performanc in SiN perception tests (target and masker interaction)

Bar chart showing mean (and standard error) Speech-in-Noise intelligibility (expressed as RAU) speech target type (words, LP sentences, HP sentences), with separate lines for masker type (speech-modulated noise and 3-talker babble)

With respect to the speech target and masker interaction (see *Figure 3. 7*) intelligibility for words (t(99)=4.72, *p<.001*) and LP sentences (t(99)=5.02, *p<.001*) was significantly greater for 3B compared to SMN, whereas for HP sentences (t(99)=2.15, *p=.31*) there was no significant difference between 3B and SMN.

Within 3B masking condition there were significant difference between HP and LP sentences (t(99)=12.55, *p<.001*), and HP sentences and words (t(99)=9.27, *p<.001*), but no difference between words and LP sentences (t(99)=1.07, *p=.29*). Within the SMN masking condition there were significant differences between HP sentences and words (t(99)=10.55, *p<.001*), HP sentences and LP sentences (t(99)=22.55, *p<.001*), and words and LP sentences (t(99)=3.18, *p=.018*).

With respect to the speech target and age group interaction (see *Figure 3. 8*) intelligibility for HP sentences (t(49)=5.27, *p<.001*) and LP sentences (t(49)=4.66, *p<.001*) was significantly greater for the older adults compared to younger adults, whereas for words (t(49)=2.42, *p=.17*) there was no significant difference between older and younger adults. Within older adult group there were significant differences between HP and LP sentences (t(49)=18.21, *p<.001*), and HP sentences and words (t(49)=15.18, *p<.001*), but no difference between words and LP sentences (t(49)=.63, *p=.54*). Within the younger adult group there were significant differences between HP sentences and words (t(49)=8.45, *p<.001*), HP sentences and LP sentences (t(49)=13.89, *p<.001*), and words and LP sentences (t(49)=5.39, *p<0.001*).

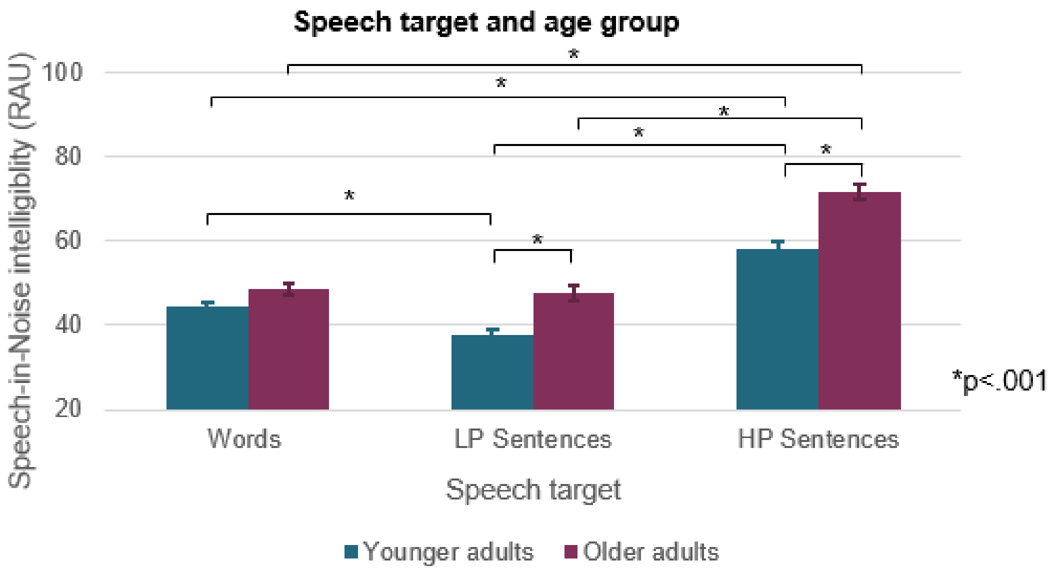


Figure 3. 8 – Bar chart of performance in SiN tests (target and age group interaction)

Bar chart showing mean (and standard error) Speech-in-Noise intelligibility (expressed in Rationalized Arcsine Units (RAU) speech target type (words, Low Predictability (LP) sentences, High Predictability (HP) sentences), with separate lines for age group (younger and older adults)

3.3.3 Speech-in-Noise intelligibly, PTA and age

Correlation analysis (Pearson's, two-tailed) revealed significant correlations between PTA(0.25-8kHz) and SiN intelligibility (averaged over all conditions) for both the younger (r=-.28, *p=.046*) and older adult (r=-.63, *p<0.001*) groups. Better PTA was associated with better SiN intelligibility for both age groups – see Figure 3. 9 below for a scatterplot of the results.

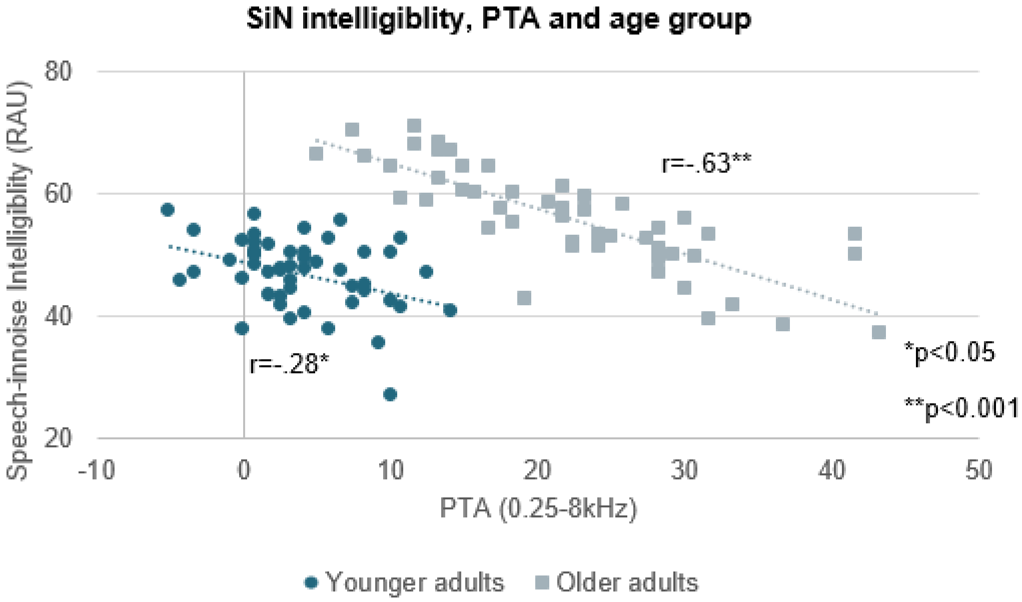


Figure 3. 9 – Scatterplot for PTA and SiN intelligibility

Scatterplot showing pure tone average against SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)). Individual points indicate observed SiN intelligibility scores, the dashed lines indicate lines of best fit for younger and older adult groups respectively. A higher PTA indicates poorer hearing sensitivity.

3.3.3 Linear mixed modelling

Baddeley model

Younger and older adults

The contributions of cognition (Baddeley sub-domains), hearing sensitivity, age group (younger and older adults), and SiN listening condition were assessed for SiN perception. *Table 3. 6* below displays the model fit (AIC (Akaike Information Criterion) value), and significant fixed effects (assessed using an ANOVA of the final model). See Supplementary Figure 3. 8for a full list of the fixed effect estimate coefficients.

The ANOVA of the mixed-model found significant main effects for speech target (F(2, 200.86)=190.79, *p<.001*), PTA (F(1, 112.93)=19.85, *p<.001*), age group (F(1, 20.69)=52.71, *p<.001*), Central Executive ability (F(1, 121.80)=4.89, *p=.029*) and Episodic Buffer ability (F(1, 121.80)=4.88, *p=.029*). The ANOVA also revealed significant interaction effects for Target and Masker (F(2, 397.94)=7.62, *p=.001*), Target and Age group (F(2, 200.86)=6.93, *p=.001*), Masker and Age group (F(1, 108.33)=14.60, *p<.001*), Central Executive and PTA (F(1, 121.80)=4.62, *p=.034*), Central Executive and Age group (F(1, 121.80)=4.67, *p=.033*), Masker, Phonological Loop and Age group (F(1, 108.05)=5.89, *p=.017*), and Masker, PTA and Age group (F(1, 108.05)=9.01, *p=.003*).

AIC value	**Random effects**	**Fixed effects**	
		Main effects	**Interactions**
4845	Participant, Target, Masker	Target, PTA, Age group, CE, EB	Target*Masker Target*Age group Masker*Age group CE*PTA CE*Age group Masker*PL*Age group Masker*PTA*Age group

Table 3. 6 – Summary of significant fixed effects from the mixed linear model assessing the Baddeley model sub-domains for younger and older adults

Linear mixed model assessing the relationship between hearing, cognition and SiN perception for both <u>younger and older adults</u> for the Baddeley model. The table displays the AIC value, random effects and significant main effects as assessed by an ANOVA of the final linear model.

AIC: Akaike Information Criterion, PTA: Pure Tone average, CE: Central Executive, EB: Episodic Buffer, PL: Phonological Loop

For the SiN target main effect, β coefficient estimates showed that SiN intelligibility was 25 (Rationalised Arcsine Units (RAU)) (β=-24.7, SE=1.98, t=-12.49, *p<.001*) lower for LP sentences compared to HP sentences. Intelligibility was also predicted to be of 16 RAU (β=-16.1, SE=2.11, t=-7.60, *p<.001*) lower for words compared to HP sentences.

The target and masker interaction β coefficient estimates revealed a masker effect of intelligibility (β=10.0, SE=3.02, t=3.30, *p=.001*), this means that intelligibility was predicted to be 10 RAU higher for HP sentences in 3-talker babble compared to speech-modulated noise. Furthermore, predicated intelligibility differences varied between LP sentences and words between the two masker types (LP sentences*3-talker babble, β=8.5, SE=2.20, t=3.89, *p<.001*) (words*3-talker babble, β=8.5, SE=2.20, t=3.89, *p=.025)*. This showed that in speech-modulated noise intelligibility was 9 RAU (24.7-16.1=8.6) greater for words compared to LP sentences, whereas in 3-talker babble intelligibility was only 5 RAU greater ((24.7-8.5)-(16.1-4.9)=4.8) for words compared to LP sentences.

For the target and age group interaction, β coefficient estimates showed an age group effect, predicting that intelligibility was 46 RAU higher (β=45.6, SE=5.05, t=9.04, *p<.001*) in the older adults compared to the younger adults. This showed that for HP sentences intelligibility was predicted to be 46 RAU higher for the older adults compared to the younger adults. Additionally, predicated intelligibility differences varied between LP sentences and words between the two listener groups (LP sentences*older adults, β=-3.4, SE=2.33, t=-1.46, p=*.147*) (words*older adults, β=-9.3, SE=2.56, t=-3.66, *p<.001)*. From this it can be deduced that for the younger listeners intelligibility was 9 RAU (24.7-16.1=8.6) greater for words compared to LP sentences, whereas for the older adults intelligibility was predicted to be 3 RAU greater ((24.7-9.3)-(16.1-3.4)=2.7) for words compared to LP sentences.

The masker and age group (3-talker babble*older adults, β=-22.9, SE=5.99, t=-3.82, *p<.001*) interaction revealed that for younger adults intelligibility was predicted to be higher (10.0 RAU) for 3-talker babble compared to speech-modulated noise. Whereas, for the older adults intelligibility was predicted to be 13 (22.9-10.0=12.9) RAU higher in speech-modulated noise compared to 3-talker babble.

In the remainder of this results section I will focus on fixed effects involving cognitive ability and complex interactions involving PTA.

With regards cognitive ability, the Central Executive and Episodic Buffer were found to significantly contribute as main effects, whereas the Phonological Loop was not. For both the Central Executive and Episodic Buffer, higher cognitive ability was predictive of lower SiN intelligibility (CE: r(100)=.25, *p=.012*; EB r(100)=-.21, *p=.041*). See Figure 3. 10 below for a scatterplot of Central Executive ability and SiN intelligibility and Figure 3. 11 for a scatterplot of Episodic Buffer ability and SiN intelligibility.

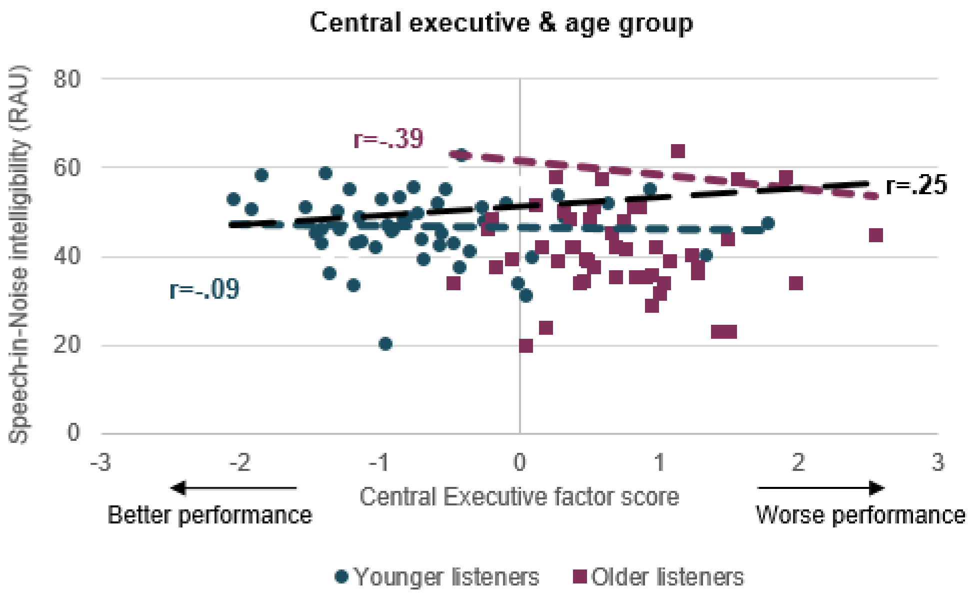


Figure 3. 10 – Scatterplot displaying the Central Executive and age group interaction

Scatterplot showing the Central Executive factor against SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)), separated by younger and older listeners groups. Individual points indicate observed SiN intelligibility scores, linear fit lines were fitted using the predicted SiN intelligibility scores from the linear mixed model, with a fit line for both age groups combined (black), and separately for younger (blue) and older (plum) adults. A lower (negative) factor score indicates better performance, higher SiN intelligibility score indicates better performance.

Next, turning to the Central Executive and age group interaction, Central Executive ability was found to be more predictive of SiN intelligibility in the older (r(50)=-.39, *p=.007*) than the younger age group (r(50)=-.09, *p=.543*) (see *Figure 3. 10*).

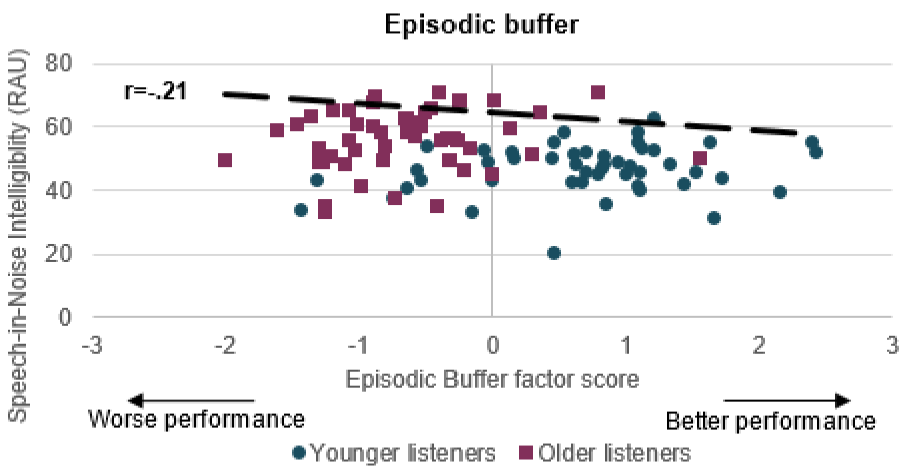


Figure 3. 11 – Scatterplot displaying the Episodic Buffer main effect

Scatterplot showing Episodic Buffer factor score against SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)), separated by younger and older listeners groups. Individual points indicate observed SiN intelligibility scores, linear fit lines for age groups together (black) were fitted using the predicted SiN intelligibility scores from the linear mixed model. A higher (positive) Episodic Buffer factor score indicates better performance.

For the Central Executive and PTA two-way interaction Figure 3. 12 displays a 3D surface plot to allow the interaction to be visualised. The plot suggests that having both good Central Executive and PTA resulted in the highest SiN intelligibly. A decrease in either Central Executive ability or PTA resulted in a sharp decrease in SiN intelligibility. However, having both poor Central Executive and hearing sensitivity resulted in better SiN intelligibility compared to a single decline in Central Executive or hearing sensitivity.

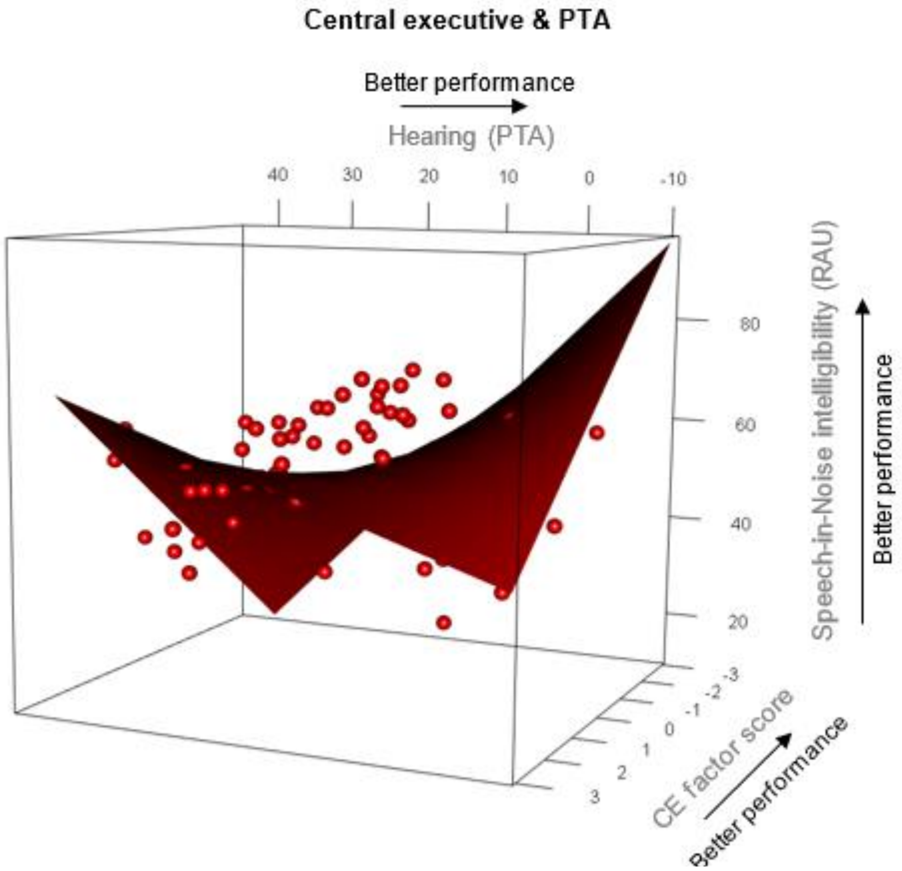


Figure 3. 12 – 3D surface plot displaying the Central Executive and PTA interaction

3D surface plot for pure tone audiometry, Central Executive factor score and SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)), combined for both younger and older listener age groups. The surface plot is based on the predicted values from the linear mixed model. Lower (negative) pure tone thresholds indicate better hearing thresholds, lower (negative) Central Executive factor score indicates better performance, and higher SiN intelligibility indicates better performance.

Figure 3. 13 displays scatterplots for the interaction between Phonological Loop ability and masker type separately for the younger (Figure 3. 13 a) and older age (Figure 3. 13 b) groups. For speech-modulated noise masking there was no significant effect of PL for younger adults (r(50)=.06, *p=.669*) or older adults (r(50)=.23, *p=.061*). For the 3-talker babble masker, there was a positive predictive effect between Phonological Loop ability and SiN intelligibility for the younger adults (r(50)=.33, *p=.019*) and a negative, but non-significant coefficient for the older adults (r(50)=-.27, *p=.061*).

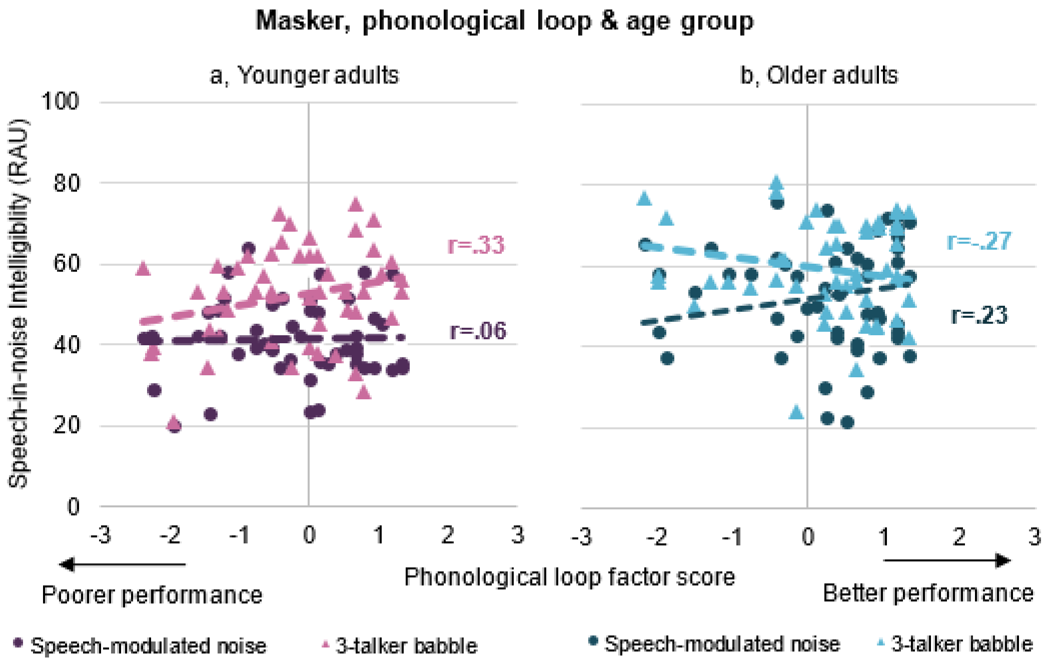


Figure 3. 13 - Scatterplots displaying the masker, Phonolological Loop and age group interaction

Scatterplots, separated for (a) younger and (b) older listener groups, showing Phonological Loop factor score against SiN intelligibility in speech-modulated noise and 3-talker babble masking conditions (averaged over target type). Individual points indicate observed SiN intelligibility scores (expressed in Rationalized Arcsine Units (RAU), linear fit lines were fitted using the predicted SiN intelligibility scores from the linear mixed model. A higher (positive) factor score indicates better performance, higher SiN intelligibility score indicates better performance.

Comparisons of the correlation coefficients (z-test statistic) showed that, comparing between younger and older adults, there was no significant difference for speech-modulated noise (z=.84, *p=.401*), but there was a significant difference for 3-talker babble (z=-3.00, *p=.003*). Furthermore, there was no significant difference between the two masking conditions for younger adults (z=1.37, *p=.171*), but there was a significant difference between the masking conditions for the older adults (z=2.48, *p=.013*).

Figure 3. 14 below displays scatterplots for PTA and masker type separately for the younger (a) and older age (b) groups. For 3-talker babble masker, similar associations were seen with SiN intelligibility for both younger (r(50)=-.46, *p=.001*) and older adults (r(50)=-.36, *p=.005*), with better hearing predicting better SiN intelligibility. In contrast, for the speech-modulated noise masker a much stronger association was seen with the older (r(50)=-.89, *p<.001*), compared to younger adults (r(50)=-.04, *p=.766*), again with better hearing sensitivity being associated with higher SiN intelligibility.

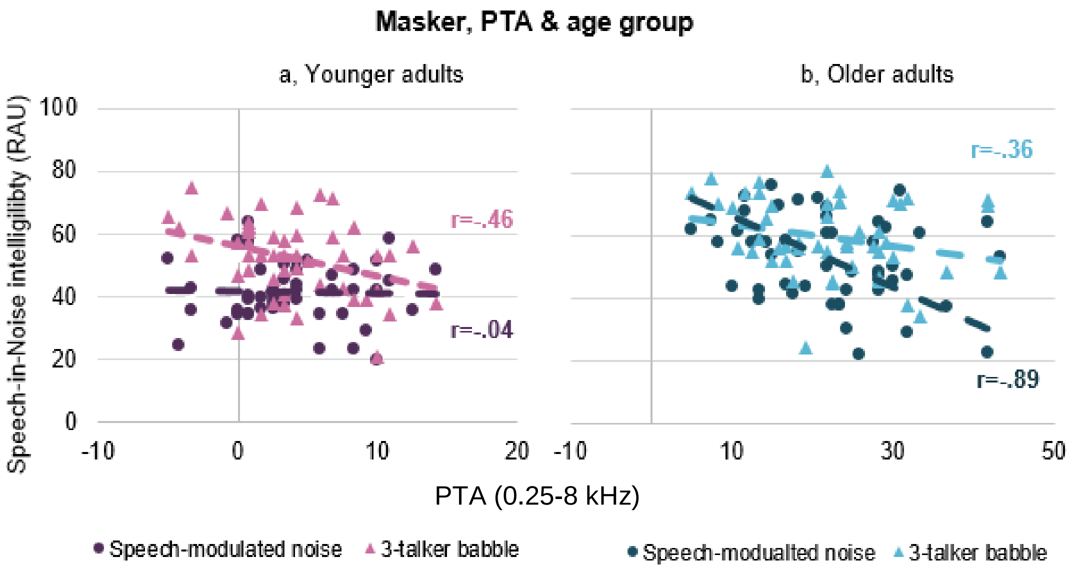


Figure 3. 14 - Scatterplots displaying the masker, PTA and age group interaction

Scatterplots, separated for (a) younger and (b) older listener groups, pure tone average against SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)) in speech-modulated noise and 3-talker babble masking conditions (averaged over target type). Individual points indicate observed SiN intelligibility scores, linear fit lines were fitted using the predicted SiN intelligibility scores from the linear mixed model.

Further analysis (z-test statistics) showed that for 3-talker babble there was no significant differences between younger and older adults (z=-.58, *p=.562*), but for speech-modulated noise there was a significant difference between younger and older adults (z=6.7, *p<.001*). Within both listener groups there was a significant difference between 3-talker babble and speech-modulated noise, with a greater difference for the older adults (z=5.07, *p<.001*) compared to the younger adults (z=-2.22, *p=.027*).

In the next two sections I will describe supplementary analysis, which was performed separately for the two age groups, younger and older adults. This supplementary analysis was performed for two main reasons. Firstly, to further explore complex interactions involving cognitive ability and age group and PTA and age. Secondly, the SiN perception test analysis in section 3.3.2 found significant differences in intelligibility between the two listener groups.

Younger adults only

Table 3. 7 below displays the results of linear mixed model for younger adults only. The table displays the model fit (AIC value), random effects, and includes all significant fixed effects (assessed using an ANOVA of the final model). The results of the ANOVA showed main effects for speech target (F(2, 96.01)=41.63, *p<.001*) and masker (F(1, 50.40)=24.98, *p<.001*), and two-way interactions for target and masker (F(2, 199.27)=6.82, *p=.001*), target and Phonological Loop (F(2, 100.75)=4.34, *p=.016*), PTA and Phonological Loop (F(1, 51.75)=5.00, *p=.030*), and a three-way interaction for Target, PTA and Phonological Loop (F(2, 100.75)=7.09, *p=.001*). See Supplementary Figure 3. 9for a full list of the fixed effect estimate coefficients.

		Fixed effects	
AIC value	**Random effects**	**Main effects**	**Interactions**
2408	Participant, Target, Masker	Target, Masker	Target*Masker Target*PL PTA*PL Target*PTA*PL

Table 3. 7 - Summary of significant fixed effects from the mixed linear model assessing the Baddeley model sub-domains for younger adults

Linear mixed model assessing the relationship between hearing, cognition and SiN perception for younger adults only for the Baddeley model. The table displays the AIC value, random effects and significant main effects as assessed by an ANOVA of the final linear model.

AIC: Akaike Information Criterion, PTA: Pure Tone Average, PL: Phonological Loop

With regards to the target main effect, the β coefficient estimates showed that intelligibility is predicted to be lower (β=-25.2, SE=2.62, t=-9.62, *p<.001*) for LP sentences compared to HP sentences. Intelligibility was also predicted to be lower for words compared to HP sentences (β=-16.6, SE=2.74, t=-6.06, *p<.001*).

For the masker main effect, the β coefficient estimates revealed that intelligibility was predicted to be 5 RAU higher for the 3-talker babble compared to speech-modulated noise (β=5.0, SE=2.69, t=1.87. p=.065).

With respect to the target and masker (LP sentences*3-talker babble β=10.0, SE=2.79, t=3.58, *p<.001*; words*3-talker babble β=7.2, SE=2.79, t=2.59, *p=.010*) interaction the β coefficient estimates revealed that intelligibility was only 5 RAU higher for HP sentences in 3-talker babble compared to speech-modulated. Whereas for LP sentences and words intelligibility was predicated to be 15 RAU (5.0+10.0=15.0) and 12 RAU (5.0+7.2=12.2) higher respectively for 3-talker babble compared to speech modulated noise.

In the remainder of this results section I will focus on fixed effects involving cognitive ability.

Target and Phonological Loop interaction post-hoc analysis (Pearson's correlation) revealed there was no significant correlation between Phonological Loop ability and SiN intelligibility in both LP (r(50)=.21, *p=.140*) and HP sentence (r(50)=.21, *p=.148*) conditions. For the word target condition on the other hand there was a moderate association between SiN intelligibility and Phonological Loop ability (r(50)=.40, *p=.004*), with better Phonological Loop ability indicative of higher SiN intelligibility. See Figure 3. 15 below for a scatterplot of the speech target and Phonological Loop interaction.

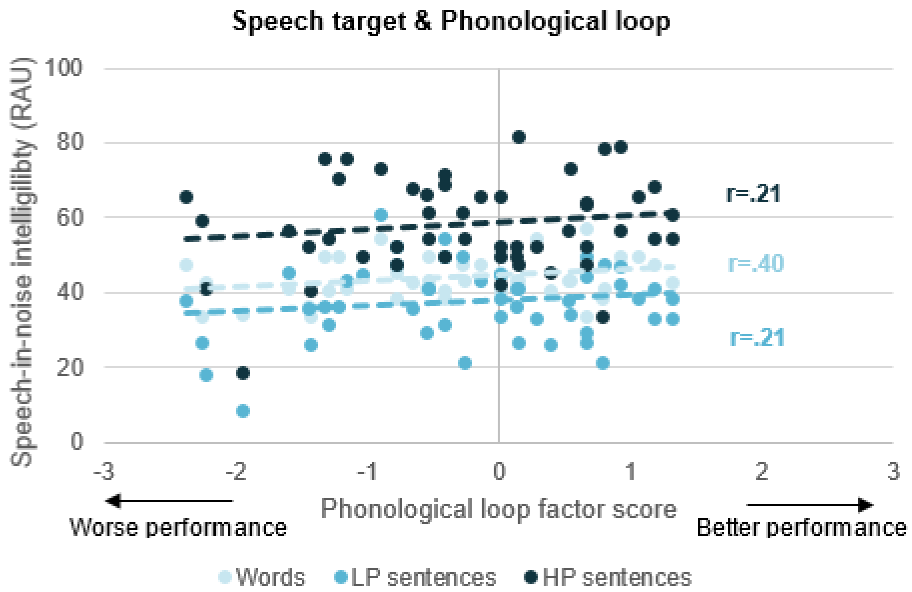


Figure 3. 15 - Scatterplot displaying the target and Phonological Loop interaction (younger adults)

Scatterplot showing Phonological Loop factor against SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)) for younger listeners only. Individual points indicate observed SiN intelligibility scores, linear fit lines were fitted using the predicted SiN intelligibility scores from the linear mixed model. A higher (positive) Phonological Loop factor score indicates better performance, higher SiN intelligibility score indicates better performance.

The two-way interaction between PTA and Phonological Loop ability (see *Figure 3. 16* below) showed that having poorer Phonological Loop ability in combination with poorer hearing sensitivity was associated with lower SiN intelligibility. Surprisingly, having better phonological ability in combination with better hearing sensitivity also was associated with poor SiN intelligibility levels, although not as low a level as the poorer Phonological Loop and poorer hearing sensitivity combination. It also appears that having either poorer Phonological Loop ability in combination with better hearing sensitivity or better Phonological Loop ability in combination with poorer hearing sensitivity was associated with the best levels of SiN intelligibility. Speculatively this may be due to increased distraction by the background information, where having both high Phonological Loop ability and good hearing causes the background, as well as the foreground, information to be clearly represented in storage. Additionally, have a higher ability in either PTA or Phonological Loop ability may lead to a poorer representation of the background noise, causing less interference in the intelligibility of the foreground information.

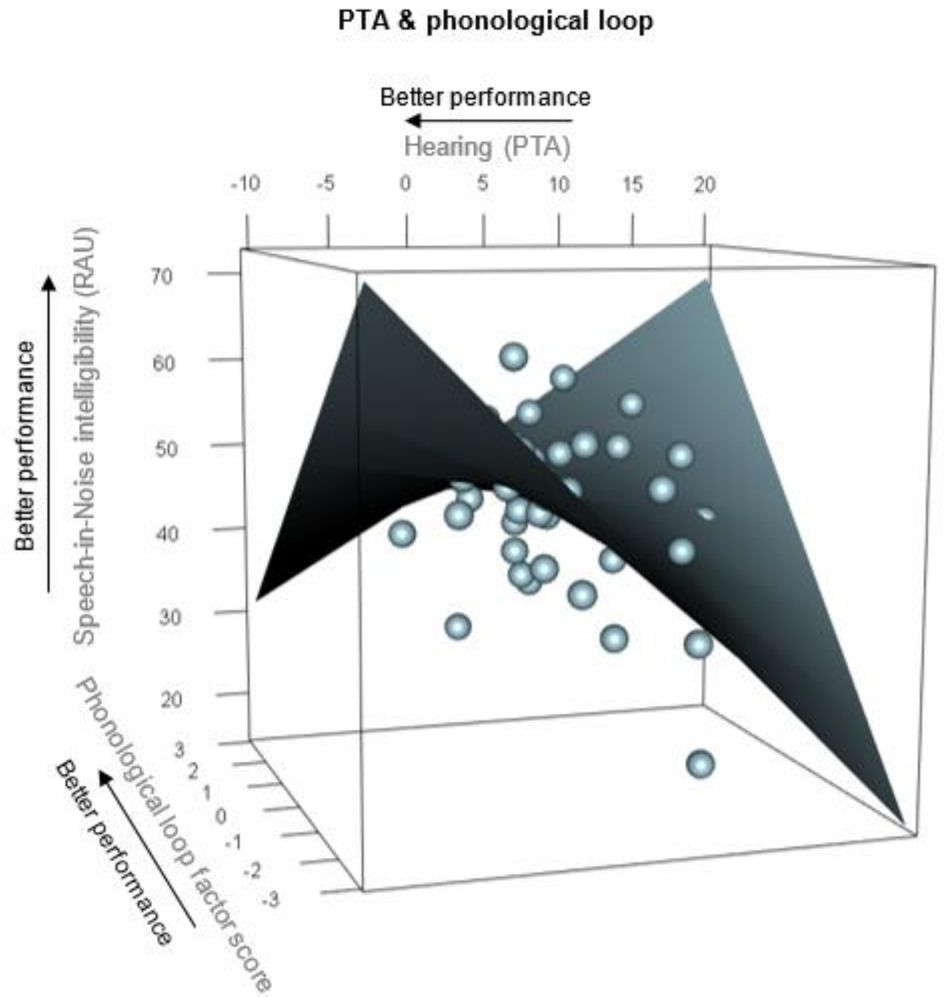


Figure 3. 16 – 3D surface plot displaying the PTA and Phonological Loop interaction (younger adults)

3D surface plot for pure tone audiometry, Phonological Loop factor score and SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)), for younger listener only. The surface plot is based on the predicted values from the linear mixed model. Lower (negative) pure tone thresholds indicate better hearing thresholds, higher (positive) Phonological Loop factor score indicates better performance, and higher SiN intelligibility indicates better performance.

The three-way interaction between PTA, Phonological Loop and speech target (see *Figure 3. 17* below), mirrored the same pattern of results for low and high predictability sentences, although performance was higher overall for high predictability sentences.

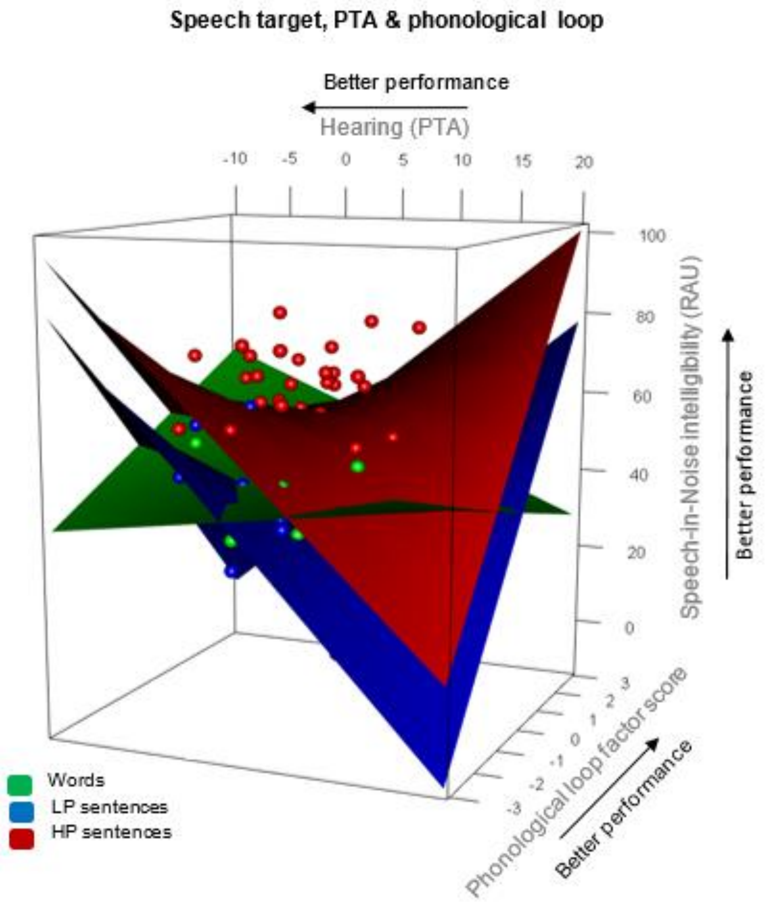


Figure 3. 17 – 3D surface plot displaying the target, PTA and Phonological Loop interaction (younger adults)

3D surface plot for pure tone audiometry, Phonological Loop factor score and SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)) separated by speech target type (words, low predictability sentences, high predictability sentences – averaged over masker type), for younger adults only. The surface plot is based on the predicted values from the linear mixed model. Lower (negative) pure tone thresholds indicate better hearing thresholds, a higher (positive) Phonological Loop factor score indicates better performance, and higher SiN intelligibility indicates better performance.

Specifically, intelligibility was best when either hearing or Phonological Loop performance was good but not when both were good. In contrast, for the single word condition SiN intelligibility was highest with a combination of better

Phonological Loop ability and hearing sensitivity. A decline in either or both phonological ability and hearing sensitivity was associated with lower levels in SiN intelligibility.

Older adults only

As for the young adults, here I present a result for the older adult group only. Table 3. 8 below displays the results of the linear mixed model for older adults only for the Baddeley model. The table displays the model fit (AIC value) random effects, and includes all significant fixed effects (assessed using an ANOVA of the final model). The results of the ANOVA showed main effects for speech target (F(2, 250)=105.28, *p<.001*) and PTA (F(1, 50)=18.16, *p<.001*). Interactions were found for masker and Phonological Loop (F(1, 250)=6.59, *p<=011*) and masker, PTA and EB (F(1, 250)=4.35, *p<.001*). See Supplementary Figure 3. 10 for a full list of fixed effect estimates.

		Fixed effects	
AIC value	**Random effects**	**Main effects**	**Interactions**
2458	Participant	Target, PTA	Masker*PL Masker*PTA*EB

Table 3. 8 - Summary of significant fixed effects from the mixed linear model assessing the Baddeley model sub-domains for older adults

Linear mixed models assessing the relationship between hearing, cognition and SiN perception for older adults only for the Baddeley model. The table displays the AIC value, random effects and significant main effects as assessed by an ANOVA of the final linear model.

AIC: Akaike Information Criterion, PTA: Pure Tone Average, PL: Phonological Loop, EB: Episodic Buffer

For the target main effect, the β coefficient estimates showed that intelligibility in the LP sentence condition was predicted to be 24 RAU lower compared to HP sentences (β=-23.9, SE 1.86, t=12.81, *p<.001*), and the intelligibility in the word condition is likely to be 23 RAU less compared to HP sentences (β=-22.9, SE 1.86, t=12.31, *p<.001*).

In the remainder of this results section I will focus on fixed effects involving cognitive ability. The Masker and Phonological Loop interaction finding here is in agreement the masker, Phonological Loop and age group interaction finding in the main analysis (see *Table 3. 6* and *Figure 3. 13*).

The masker, PTA and Episodic Buffer three-way interaction shows that the relationship between hearing sensitivity and SiN intelligibility differs depending on background masking condition (see *Figure 3. 18* below). In the 3-talker babble masking condition having both high Episodic Buffer ability and good hearing is required for high SiN intelligibility. A decline in either or both is associated with a marked decline in SiN intelligibility. In the speech-modulated noise condition SiN intelligibility is more dependent on hearing sensitivity. If hearing sensitivity is high SiN intelligibility is high, regardless of Episodic Buffer ability. As hearing sensitivity declines the contribution of the Episodic Buffer increases, but only moderately.

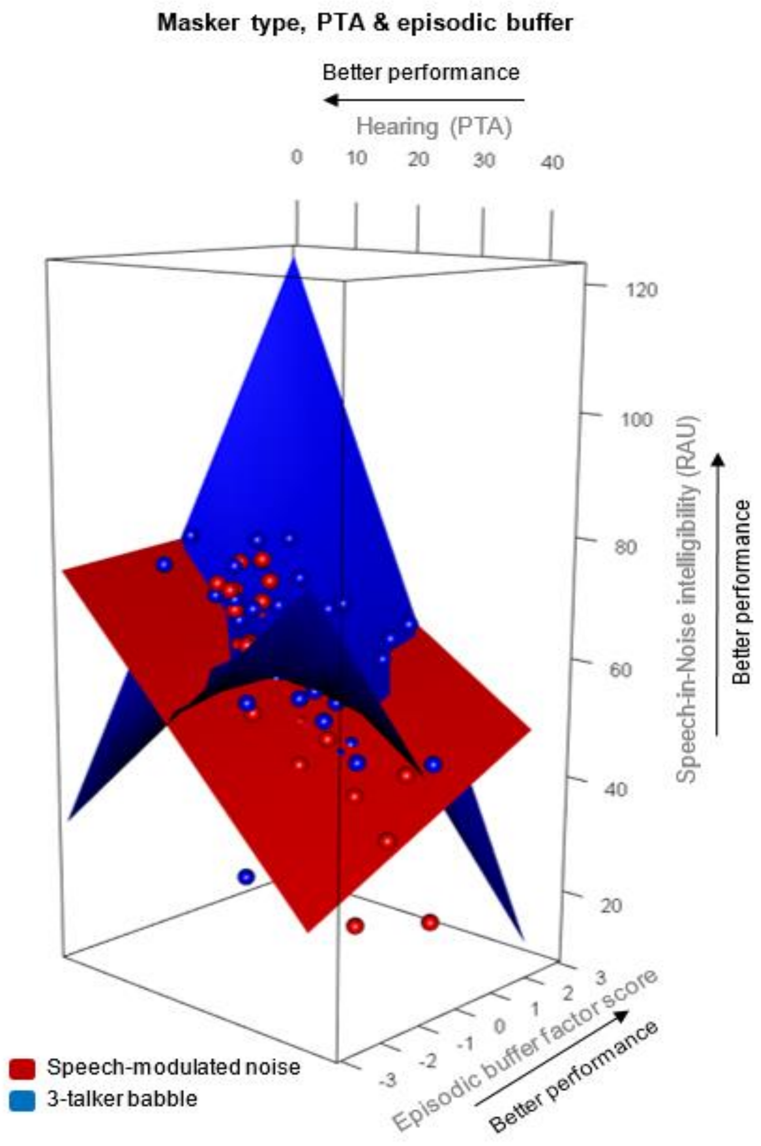


Figure 3. 18 - 3D surface plot displaying the masker, PTA and Episodic Buffer interaction (older adults)

3D surface plot for pure tone audiometry, Episodic Buffer factor score and SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)) separated by masker type (speech-modulated noise, 3-talker babble– averaged over speech target type), for older listeners only. The surface plot is based on the predicted values from the linear mixed model. A lower (negative) pure tone thresholds indicate better hearing thresholds, a higher (positive) Episodic Buffer factor score indicates better performance, and a higher SiN intelligibility indicates better performance.

Diamond model

Younger and older adults

Table 3. 9 below displays the results of linear mixed model for younger and older adults for the Diamond model. The table displays the model fit (AIC value) random effects, and includes all significant fixed effects (assessed using an ANOVA of the final model). See Supplementary Figure 3. 11for a full list of fixed effect estimates.

There were no significant main effects for cognitive abilities (Inhibitory Control and Working Memory. However, cognitive ability aside, this model replicated all the effects found when using cognitive domains based on the Baddeley model. Specifically, main effects were again found for target (F(2, 204.25)=188.37, *p<.001*), PTA (F(1, 110.72)=21.88, *p<.001*) and age group PTA (F(1, 112.92)=52.57, *p<.001*), and interactions for target and masker (F(2, 398.23)=7.56, *p=.001*), target and age group (F(2, 204.45)=6.94, *p=.001*), masker and age group (F(1, 107.11)=16.42, *p<.001*) and masker, PTA and Age group (F(1, 106.72)=9.56, *p=.003*).

The β coefficient estimates for the target and age group main effects, and target and masker, target and age group, and condition and age group interactions were found to be qualitatively similar to the predictions in the Baddeley age group combined model in the previous section. Therefore, the results will not be repeated here. Please refer to Supplementary Figure 3. 12 in the appendix for a list of the prediction coefficients.

		Fixed effects	
AIC value	**Random effects**	**Main effects**	**Interactions**
4848	Participant, Target, Masker	Target, PTA, age group	Target*Masker Target*Age group Masker*Age group Masker*PTA*Age group

Table 3. 9 - Summary of significant fixed effects from the mixed linear model assessing the Diamond model sub-domains for younger and older adults

Linear mixed model assessing the relationship between hearing, cognition and SiN perception for younger and older adults for the Diamond model. The table displays the AIC value, random effects and significant main effects as assessed by an ANOVA of the final linear model.

AIC: Akaike Information Criterion, PTA: Pure Tone Average

As in the Baddeley model analysis section here I also carried out two sets of supplementary analysis for the younger and older adult age groups. The supplementary analysis is described below.

Younger adults only

Table 3. 10 below displays the results of linear mixed model for younger adults only for the Diamond model. The table displays the model fit (AIC value) random effects, and includes all significant fixed effects (assessed using an ANOVA of the final model). No significant effects were found involving cognitive ability, Inhibitory Control and Working Memory. However, significant main effects were found for speech target ($F(2, 83.43)=78.43$, $p<.001$), masker ($F(1, 83.43)=50.39$, $p<.001$), PTA ($F(1, 83.43)=68.53$, $p=.044$) and a two-interaction term was found between target and masker ($F(2, 199.39)=6.53$, $p=.002$).

Here the β coefficient estimates did not vary qualitatively between this model and the equivalent model in the Baddeley model analysis for the target, masker and target and masker fixed effects. Therefore I will not state the results here to avoid repetition.

		Fixed effects	
AIC value	Random effects	Main effects	Interactions
2410	Participant, Target, Masker	Target, Masker, PTA	Target*Masker

Table 3. 10 - Summary of significant fixed effects from the mixed linear model assessing the Baddeley model subdomains for younger adults

Linear mixed model assessing the relationship between hearing, cognition and SiN perception for <u>*Younger adults only*</u> *for the Diamond model. The table displays the AIC value, random effects and significant main effects as assessed by an ANOVA of the final linear model.*

AIC: akaike information criterion

Older adults only

Table 3. 11 below displays the results of linear mixed model for older adults only for the Diamond model. The table displays the model fit (AIC value) random effects, and includes all significant fixed effects (assessed using an ANOVA of the final model).Significant main effects were found for speech target ($F(2, 200.00)=122.49$, *p<.001*) and PTA ($F(1, 50.19)=34.96$, *p<.001*). Additionally, there was a three-way interaction between masker, PTA and Working Memory ($F(1, 50.00)=4.06$, *p=.049*).

Here the β coefficient estimates for fixed effects did not vary between this model and the equivalent model in the Baddeley model analysis for the target main effect. Therefore, I will not state the results here to avoid repetition. See Supplementary Figure 3. 13 for a full list of the estimates of the fixed effects.

		Fixed effects	
AIC value	**Random effects**	**Main effects**	**Interactions**
2456	Participant, masker	Target, PTA	Masker*PTA*WM

Table 3. 11 - Summary of significant fixed effects from the mixed linear model assessing the Diamond model subdomains for older adults

Linear mixed model assessing the relationship between hearing, cognition and SiN perception for older adults only for the Diamond model. The table displays the AIC value, random effects and significant main effects as assessed by an ANOVA of the final linear model.

AIC: Akaike Information Criterion, PTA: Pure Tone Average, WM: Working Memory

Figure 3. 19 below displays a 3D surface plot relating to the Masker, PTA and WM interaction. As with the Episodic Buffer domain, here we see a different relationship between hearing sensitivity and SiN intelligibility depending on background masking condition. In the 3-talker babble masking condition having both high Working Memory ability and good hearing is required for high SiN intelligibility. A declines in either or both Working Memory ability or hearing sensitivity is associated with marked declines in SiN intelligibility. In the Speech-modulated masker SiN intelligibility is more dependent on hearing sensitivity. If hearing sensitivity is high SiN intelligibility is high, regardless of Working Memory ability. As hearing sensitivity declines the contribution of Working Memory increases, but only to a low level.

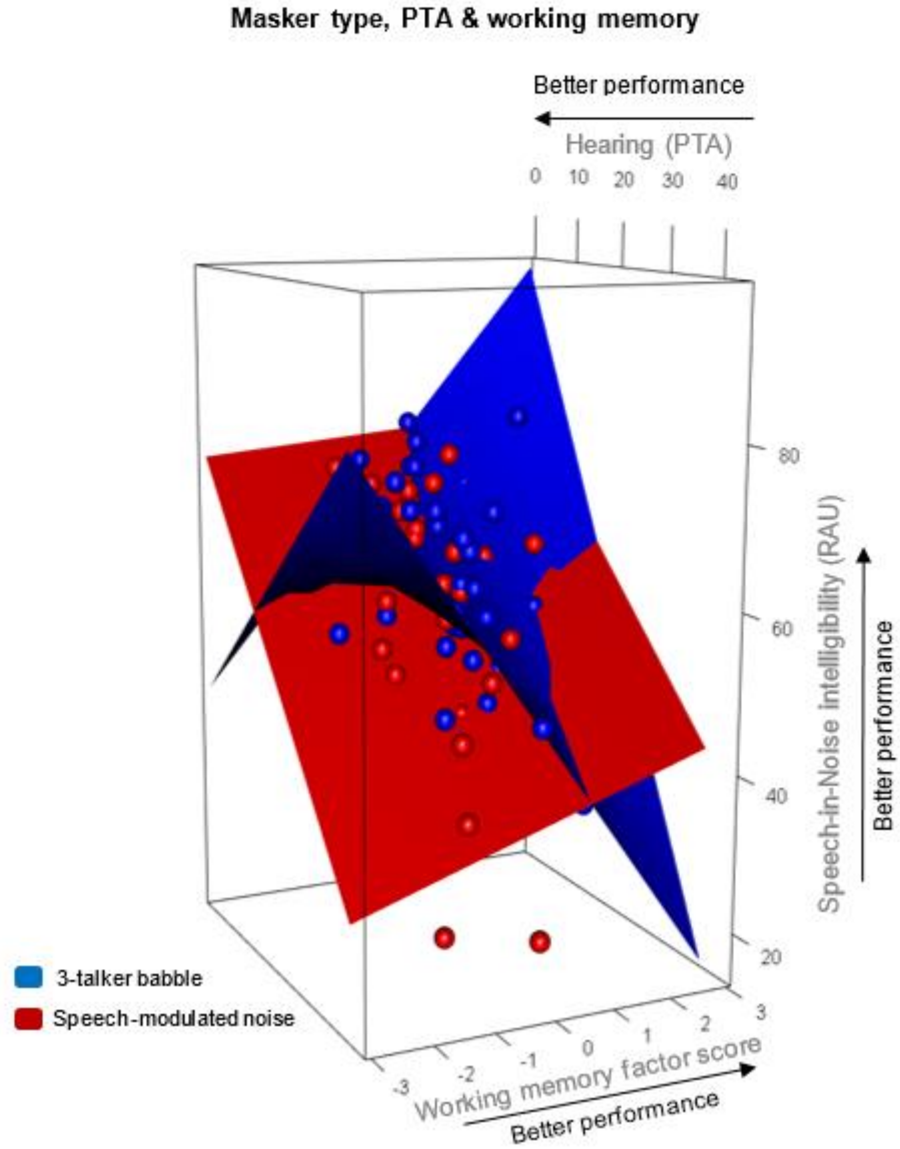


Figure 3. 19 – 3D surface plot displaying the masker, PTA and working memory interaction (older adults)

3D surface plot for pure tone audiometry, working memory factor score and SiN intelligibility (expressed in Rationalized Arcsine Units (RAU)) separated by masker type (speech-modulated noise, 3-talker babble – averaged over speech target type), for older listeners only. The surface plot is based on the predicted values from the linear mixed model. A lower (negative) pure tone thresholds indicate better hearing thresholds, a higher (positive) working memory factor score indicates better performance, and a higher SiN intelligibility indicates better performance.

3.4 Discussion

I used a theory driven and systematic approach to investigate the role of cognition and hearing sensitivity in SiN perception in different listening condition in an association study experiment.

In using this systematic and theory-driven approach I explored the following research questions:

1) Which cognitive abilities are important for SiN perception?
2) Does the role of specific cognitive abilities differ depending on SiN listening condition?
3) Do these roles differ depending on age group and hearing sensitivity?

3.4.1 Cognitive models

Baddeley model

An adapted version of the Baddeley model was used in this study, which included the Central Executive, Episodic Buffer and Phonological Loop sub-domains (excluding the Visuo-Spatial Sketchpad from the model).

Diamond model

The simplified version of the Diamond model of Executive Functions used for this study comprised of the sub-domains Inhibitory Control and Working Memory. This model was selected as a complementary model to the Baddeley model of Working Memory, but differ in assessing executive processes separately.

The Inhibitory Control sub-domain comprised of the same three cognitive tests as the Central Executive sub-domain of the Baddeley model, namely the Test of Everyday Attention subtests 1, 6, and 7 – all tests that assess sustained attention, a sub-division of Inhibitory Control within the Diamond model. The cognitive tests used to derive the Working Memory principal component were, the Reading Span Test, Letter-Number Sequencing, and the Digit Span Backward – all three tests assess verbal Working Memory, a sub-domain of Working Memory within the Diamond model.

3.4.2 Which cognitive abilities are important?

Baddeley model

In the overall model I found that performance on all three sub-domains predicted SiN performance. However, in all but one case (Episodic Buffer) the predictive effects were moderated by listening condition, age group, or PTA.

The role of the Central Executive was specific to age and PTA. Moreover, SiN intelligibility was dependant on both hearing sensitivity and Central Executive ability. Furthermore, Central Executive ability was found to predict SiN intelligibility in the older, but not the younger group. Additionally, these roles did not depend on SiN listening condition.

For the Phonological Loop SiN intelligibility was specific to background masker and age group. More specifically, for younger listeners Phonological Loop ability was associated with SiN intelligibility in 3-talker babble, but not speech-modulated noise. Whereas, for the older adults Phonological Loop ability was associated with SiN intelligibility in both 3-talker babble and speech-modulated noise.

For the Episodic Buffer the overall model did appear to show a main effect for Episodic Buffer ability for SiN intelligibility. However, this appeared to be an artefact caused by overall intelligibility being higher for the older versus the younger listeners at the age-specific SNRs used in this study. This age group difference in SiN performance appears to be driving the correlation observed between Episodic Buffer ability and SiN performance. In the absence of this difference it seems unlikely that there is a main effect of Episodic Buffer ability for SiN perception. Additionally, the supplementary age group analysis in the older listeners did show that the role of Episodic Buffer ability for SiN intelligibility may be moderated by background masker and PTA. The interpretation of the Episodic Buffer finding will need to be taken with caution, particularly with regards to making inferences to age group due to this finding not being replicated in the main analysis.

Diamond model

The Inhibitory Control sub-domain was not found to significantly contribute to SiN intelligibility. Therefore Inhibitory Control, as assessed using the Test of Everyday Attention subtests 1, 6 and 7 (all sustained attention measures), does not contribute to SiN perception in the listening conditions and for listeners included in this study. This is perhaps due to the Test of Everyday Attention showing some ceiling effects and a lack variance in the non-clinical groups tested in this study.

However, this is still a surprising results given that the Inhibitory Control sub-domain was assessed by the same surface tests (i.e., the TEA tests) as the Central Executive sub-domain of the Baddeley model. Further analysis of the Baddeley model data revealed that the Central Executive interactions with age and hearing sensitivity were lost if all the Episodic Buffer was removed as a predictor. Working Memory was also not found to contribute to SiN intelligibility in the main analysis. However, in the supplementary analysis for the older listeners Working Memory ability was found to contribute to SiN intelligibility, and that contribution was modulated by hearing loss and SiN listening condition.

3.4.3 Were the roles moderated by listening condition, age or PTA?

Baddeley model

Central Executive

The overall effect of the Central Executive sub-domain on SiN perception was moderated by age such that older but not the younger adults showed better Central Executive ability predicted better SiN intelligibility, and this was true regardless of SiN listening condition. Central Executive ability was also shown to be moderated by hearing sensitivity. The result indicates that both good hearing and good Central Executive ability are required for optimal speech SiN perception. Therefore, a decline in either hearing sensitivity or Central Executive ability could lead to a decline in SiN perception. This, in part, explains why even when poor hearing is restored by hearing intervention, some listeners still experience SiN difficulties.

The Central Executive principal component was derived from the Test of Everyday attention sub-tests 1, 6 and 7. Here all three of the included cognitive tests assess attention, specifically selective attention (Chan et al., 2006; Posner et al., 1990). Therefore, the Central Executive sub-domain assessed here can be said to be focused specifically on the attentional processes. Due to this level of specificity in my own results, and in previous work, I will relate my findings to previous studies focusing on the attentional sub-processes of the Central Executive.

My results confirm and extend previous results in the following way. In a recent study Heinrich et al. (2015) found that both hearing sensitivity and attention ability were associated with SiN performance (sentences/modulated noise) for older adults with mild HL. However, in a study investigating the role of cognition for SiN perception in normal hearing listeners, Füllgrabe et al. (2015) found that attention ability was not associated with intelligibly for sentences (in 2-talker babble). The lack of association could be explained by the selectivity in hearing sensitivity range, which only included normal hearing listeners. These result suggest that a combination of HL and Central Executive (attention) decline contributes to a decline in SiN perception, in agreement with my findings.

These results may be of particular interest in treatment for older patients with Central Executive/attentional decline, where hearing intervention could be used in a compensatory manner to improve SiN perception. Therefore, targeting hearing intervention might be the best way to improve SiN deficits, particularly for older adults with mild HL. One form of hearing intervention is fitting hearing aids, yet aided listeners still report difficulties in SiN listening. This suggests that hearing aid intervention alone is not beneficial enough to improving SiN listening (Johnson et al., 2011). More recent studies have suggested that a combination of auditory and cognitive training might be a beneficial route in improving SiN listening in aided listening (Ferguson et al., 2015; Rudner, 2016; Tremblay et al., 2016; Yu et al., 2017). However, further research is needed in understanding the link between auditory and cognitive training and SiN perception.

Phonological Loop

The overall model showed these results showed for the younger adult listener group Phonological Loop ability was only related to SiN perception in the 3-talker babble, but not the speech-modulated noise. In the older adult group Phonological Loop ability plays a role in both 3-talker babble and speech-modulated noise masking. Surprisingly, in 3-talker babble masking better Phonological Loop ability was associated with poorer SiN perception. One possible explanation for this surprising result is that older listeners, who require further Phonological Loop ability to process speech targets, are also more vulnerable to interference from intelligible background noise.

The rhyme verification tasks used to assess the Phonological Loop sub-domain also assess other sub-domains, specifically Crystallised intelligence and verbal ability. Verbal ability is known to improve with age and this finding was replicated in this study with older adults performing higher than the younger adults in the Mill Hill vocabulary test ($t(49)=6.28$, $p<0.001$). Potentially this higher verbal ability may become detrimental to SiN intelligibility for informational masking in the presence of sensory and/or cognition related-declines, which is not captured by this study.

In the younger adult supplementary analysis, a three-way interaction was found between the Phonological Loop, hearing sensitivity and speech target. The results showed, less linguistically complex target stimuli, single words, required both good hearing and Phonological Loop abilities for good SiN intelligibility, whereas for more linguistically complex targets, sentences, good ability was only required in either Phonological Loop processes or hearing sensitivity. This finding is possibly due to the contextual support, as opposed to linguistic complexity, properties of the target signals. Single words provide no contextual support therefore both hearing and processing of the signal information are of importance, whereas for sentence targets, where there is more contextual support, there might be a compensatory mechanism were good ability in either hearing sensitivity or Phonological Loop processing can compensate for poorer ability in the other.

In a previous study, Surprenant (2007), assessed performance in a rhyme verification task and of SiN perception in younger and older adults with normal hearing. The SiN perception test was a syllable identification task with a background noise of broadband noise at three SNRs (+25dB, +5dB, 0dB). The rhyme verification was a four-choice identification task where participants were required to select one word which did not rhyme with the other three, all the words in each trial were orthographically similar. The outcome measure used to assess phonological ability was reaction time, not proportion correct response due to ceiling effects. The results of the study showed that phonological ability was not associated with syllable identification (at a 70% correct threshold). Additionally, no age effects were found in the study. This suggesting that phonological abilities may not play a role for syllable-in-noise identification at favourable SNRs for both younger and older normal hearing listeners. It is difficult to make a direct comparison between this study and my own because there is no overlap in speech target or masker type in the SiN perception tests. However, it can be inferred that Phonological Loop ability may play a greater role in perception of more complex target stimuli (words and sentences), and background maskers with informational properties, and it these differences in SiN listening conditions that lead to differences in strategies for younger and older listeners.

Overall these results that Phonological Loop ability can play a different role depending on age, hearing sensitivity and SiN listening condition. In terms of hearing intervention, amplification for younger adults may be most beneficial for SiN perception in processing low-context information or listening in babble noise. For older adults amplification would potentially be generally useful. However, in persons with high Phonological Loop ability it may actually be detrimental in conditions where there are competing talkers. Therefore, measuring Phonological Loop ability might be of use in a clinical setting where fittings could be tailored in relation to a person's cognitive ability and pre-sets could be determined for different listening conditions.

Episodic Buffer

The overall model showed no effect of Episodic Buffer other than the main effect discussed in the previous section. The supplementary analysis in older adults only showed a three-way interaction with, hearing sensitivity and background masker, this effect was not observed in the younger adult group. The result revealed that in more energetic masking (speech-modulated noise) hearing sensitivity plays the greatest role, whereas in more informational masking (3-talker babble) a combination of hearing sensitivity and Episodic Buffer ability is important.

In previous studies, Füllgrabe et al. (2015) found, for older adults, an association between short-term Episodic Buffer ability (assessed using the digit span test) and perception of sentences in 2-talker babble and Anderson et al. (2013) found associations with sentences presented in both speech-shaped noise and 4-talker babble masking conditions. But neither related the role of hearing sensitivity in relation to short-term Episodic Buffer ability, masking and SiN perception. This highlighting the advances my study has brought to the literature.

The ability to store verbal information into episodes for further executive processing appears to be particularly important in the presence of intelligible distractors. Furthermore, verbal distraction can even cause a negative effect in those with good hearing, but low Episodic Buffer ability. This is perhaps due to better hearing listeners being able to hear of the intelligible background but may not have the cognitive storage abilities to appropriately deal with this information. This result further refines the previously discussed finding of an interaction between the Central Executive and hearing sensitivity where I suggested that providing hearing amplification and auditory/cognitive training might be a more successful strategy in improving SiN perception outcomes.

Diamond model

Working Memory

In the older group supplementary analysis, Working Memory moderated SiN performance in connection with hearing sensitivity for different masker types as shown by a three-way interaction between Working Memory with hearing sensitivity and masker. The results showing that for more energetic masking (speech-modulated noise) hearing sensitivity plays the greatest role, whereas in more informational masking (3-talker babble) a combination of hearing sensitivity and

Working Memory ability is important. This result suggests that the role of Working Memory is increased for speech perception when there is intelligible background information.

This pattern of results is strikingly similar to the relationship shown between SiN masker type and the Episodic Buffer (memory) in the Baddeley model. Perhaps this not unexpected in that Episodic Buffer and Working Memory ability are closely related in that both abilities have a storage aspect, but Working Memory has a manipulation component in addition to that. This comparable result between the two cognitive abilities for masker type suggests that the effect seen for Working Memory may be driven by the Episodic Buffer storage component more than manipulation or other executive processes.

In terms of previous findings, studies investigating the role for Working Memory in different SiN listening conditions for older listeners (with hearing sensitivity ranging between no HL and moderate HL) have found an association between Working Memory ability and SiN intelligibility of words and low context sentences in multiple talker babble (Anderson et al., 2013; Gordon-Salant et al., 2016) and short simple sentences in speech-shaped noise (Anderson et al., 2013). These results showing that, for older listeners with a range of hearing sensitivities, Working Memory plays a role for SiN perception for a range of different listening conditions. The current study building upon these results by revealing the differing role of hearing sensitivity in SiN perception depending on background masking.

3.5 Limitations

As with all experiments, this study has several limitations. I will outline some of these limitations in relation to cognition, SiN perception tests, listener variables, and the overall methodological approach.

Firstly, the cognitive ability factor scores used in the mixed linear models were derived using PCA, not a FA, method. Although this method circumvented the issue of factor score indeterminacy the components accounted for maximum variance in the data and not solely common variance as a FA approach would have achieved. Therefore, the principal components were not the equivalent to latent variables and did not account for the impurity principle.

Secondly, despite its advantages in approach to cognitive test selection and theory, this study was limited in assessing attentional and executive processes due to the cognitive models that were selected. One approach to more comprehensively assess attention and executive processes would have been to select both the Posner and Peterson's (2012) model of attention and Miyake's (2000) model of Executive Functions. Petersen et al. (2012) divide attention into three sub-domains, Alerting, Orienting and executive processes. Miyake et al. (2000) divide executive processes into three sub-domains, Set-Shifting, inhibition and Working Memory. The commonality between these models would allow them to be tested separately and in combination as a unified model. But to avoid limitations in assessing memory abilities a model such as Baddeley's (2000) model would also need to be selected. However, combining this combination of models would place it on a scale beyond the limited time and resources afforded within the confines a PhD project.

In terms of SiN perception test selection this study was limited in including three target types (words, LP sentences, HP sentences) and two masker types (speech-modulated noise, 3-talker babble). Referring back to the target/masker matrix used in the systematic review in chapter two, which comprises of three target categories (phonemes, words, sentences) and four masker categories (unmodulated noise, modulated noise, >2-n talker babble, ≤2-n talker babble), it would have also been of worth to examine other SiN perception test combinations within this matrix. For example, the inclusion of phoneme targets to give further breadth along the linguistic complexity continuum.

A further limitation of this study relates to listener factors such as age and hearing status. In this study age was investigated as a categorical variable in a cross-section of younger (18-30 years of age) and older adults (60+ years of age). Due to same constraints I was unable to add a middle age (30-60 years of age) group, which would have required testing a further 50 participants to match the sampling sizes of the other age groups. Alternatively, age can be investigated in a longitudinal study design. However, a longitudinal design would have been too constrained and less meaningful within the limits three-year PhD project.

With regards to hearing status it would be of interest to include younger adults with hearing loss to further explore the relation between age and HL. One way to extend the current study would be to include listeners with moderate and severe HL, and to include hearing aid groups to assess how cognitive strategies differ with HL severity and how this processes may differ between aided and unaided listening. This approach would be useful for assessing the feasibility of bringing cognitive testing to the hearing clinic and it may also have further implications for study of age-related neurodegenerative disorders such as dementia.

Finally, this study is also limited by virtue of taking an association methodological approach. Although this is undoubtedly a useful approach in the psychophysics and psychoacoustic toolkit, it is limited in not determining causality. But it does serve as a platform for more precisely guiding experimental studies.

One such approach was used in the study in the next chapter, which investigates at the role of Working Memory (storage and manipulation) in perception of informationally and energetically masked speech, using a dual-task paradigm.

3.6 Conclusions

In summary this study provides insight to the specific roles of cognitive abilities in different SiN listening conditions for younger and adult adults.

More specifically, the Central Executive (attention) is important for SiN perception regardless of listening condition, particularly for older adults who are more likely to exhibit declines in both hearing sensitivity and executive processing.

Phonological Loop processing plays a differing role in SiN perception depending on background masker and listener age. In the younger adult group Phonological Loop ability only plays a role for SiN perception in informational masking, whereas for older adults it plays a role in both energetic and informational masking. This suggests the older listeners may process SiN differently to younger adults.

The supplementary analysis in older listeners showed Working Memory and Episodic Buffer ability shared similar results in both contributing to SiN perception, with a differing role depending on hearing sensitivity and background masker. Hearing sensitivity is most important in energetic masking conditions, perception in the presence of intelligible distractors requires both good hearing and episodic/Working Memory ability. This similarity in results between Working Memory and the Episodic Buffer suggests that the masking effect is driven more by storage than the manipulation aspect of Working Memory.

These results highlight the fact that examining multiple cognitive processes using multiple surface tests is vital in determining the roles of specific cognitive processes for SiN perception. Furthermore, these complex relationships appear to differ depending on SiN listening condition, age and hearing loss.

<Blank page>

Chapter Four - A dual-task study: the role of working memory ability for speech-in-noise perception in younger adult listeners

Abstract

Based on the results of the association study in chapter three I devised a dual-task experiment that would manipulate the availability of specific cognitive abilities for SiN perception. If cognitive abilities were as important as predicted by the association experiment, then their relative unavailability should adversely affect SiN perception.

Using a novel approach within the SiN perception dual-task literature, I selected four secondary cognitive tests, (Digit Span Forward, Digit Span Backward, Corsi Span Forward, and Corsi Span Backward) which were designed to engage the Baddeley sub-domains (Central Executive, Visuo-spatial Sketchpad, Episodic Buffer, Phonological Loop) to different extents.

By presenting the cognitive tests concurrently with SiN perception tests and instructing participants to prioritise attention equally to both tasks I was able to assess dual-task performance for SiN and cognitive ability separately and as a combined measure. Furthermore, the selection of cognitive and SiN perception tests allowed me to assess the role of each Baddeley sub-domain for SiN in two masking conditions (speech-modulated noise, 3-talker babble). This approach allowed me to investigate the role of cognition for SiN perception with a greater degree of specificity and to move on from an association-testing to a causality-testing design.

I tested young listeners (18-30 years) without hearing loss (<20dB $HL_{0.25\text{-}4kHz}$). For this age group I provide evidence that verbal processing (Phonological Loop) is more important than Executive Functions (Central Executive) and episodic storage (Episodic Buffer) for SiN perception.

These results provide an explanation as to why previous studies have not found a role of Working Memory for SiN perception in young adult listeners with normal hearing.

4.1. Introduction

In the previous chapter I found that cognitive abilities contributed differently for SiN perception depending on age group, hearing sensitivity and listening condition. In the older, but not the younger adult group, Central Executive and Episodic Buffer abilities were found to be important for SiN perception. However, Phonological Loop ability was found to play a role for SiN perception for both age groups, and this role differed depending on background masker.

Here, using a dual-task, I examined the role of specific cognitive abilities for SiN perception in different listening conditions for younger adult listeners. If cognitive abilities are as important as predicted by the association experiment, then their relative unavailability should adversely affect SiN perception. Using a dual-task approach allowed me to systematically manipulate and assess the roles of different cognitive abilities for SiN perception in different background maskers. This experiment was also designed so the results will be directly comparable to the previous study and thereby strengthen the overall findings of the thesis, particularly if a consensus is reached between studies.

Dual-task paradigms are underpinned by the underlying assumption that cognitive resources are limited in capacity. Classical theories of dual-task interference include *bottleneck theory* (Broadbent, 1958; Cherry, 1953) and *shared resource theory* (Kahneman, 1973). Additionally, Baddeley's (2000) model of Working Memory suggests that dual-task interference can occur at the level of storage, where only tasks of the same modality (verbal or visuo-spatial) can interfere with each other whereas storage tasks in different modalities will not. Dual-task studies involve two concurrently presented tasks with a primary task, usually of specific research interest, and a secondary task, which engages the proposed shared cognitive resources used for the primary task.

A dual-task study in the SiN literature typically consists of a primary speech perception task and a secondary cognitive task, which can vary in sensory domain (auditory, visual, and tactile) and modality (verbal, non-verbal). Task performance (proportion correct response and/or reaction time) is usually compared between single- and dual-task conditions in secondary task performance. It is expected that under dual-task conditions accuracy will drop and reaction time will increase due to the shared nature and limited capacity of cognitive resources. As a consequence cognitive resources may be prioritised to one task - *bottleneck theory* (Broadbent, 1958; Cherry, 1953) or shared across multiple tasks – *cognitive control theory*

(Kahneman, 1973) leading to decreased performance in one or both tasks. If both tasks require storage in Working Memory (Baddeley, 2000) any interference between them may depend on the modality of the task.

Dual-task cost can be calculated using one of two methods, a difference score (single-task performance – dual-task performance) or a proportional dual-task cost (pDTC) ((dual-task performance – single-task performance) / single task performance) (Doumas et al., 2008; Somberg et al., 1982). Proportional dual task cost is preferred to difference scores when there are differences between group scores in single-task performance (Gagne et al., 2017; Somberg et al., 1982). Dual-task cost can be derived from either primary, secondary, or an aggregate of both performance scores. An aggregate score is particularly important if there are differences in dual-task versus single-task performance in both the primary and secondary scores (Plummer et al., 2015).

Here I will examine dual-task performance in the primary (SiN), secondary (cognition), and combined (SiN-cognition) performance using the proportional dual-task cost method, extending previous studies, which only report dual-task cost based on secondary task performance (Desjardins et al., 2013; Pals et al., 2015; Tun et al., 1991). Using all three performance indices I will be able to assess and compare if and how performance is affected by a specific secondary task. While dual-task studies have been previously used to measure the contribution of cognition to SiN perception, there remain some key gaps in the literature, which this study aims to fill.

Firstly, studies often use a simple non-verbal secondary cognitive test, e.g. visuo-motor or tactile response tasks (Desjardins, 2016; Desjardins et al., 2013; Fraser et al., 2010; Gosselin et al., 2011), which do not specifically engage verbal memory of Working Memory, and perhaps do not fully engage the Central Executive component, both of which are likely to be important for SiN perception. Secondly, few SiN-cognition dual-task studies have examined differences depending on SiN listening condition or assessed multiple secondary tasks, with limited exceptions (Bockstael et al., 2018; Gagne et al., 2017; Picou et al., 2014). Additionally, studies often take an association study approach by examining correlation coefficients between dual-task performance and single-task performance in cognitive tests (Desjardins, 2016; Desjardins et al., 2013; Tun et al., 1991). Therefore, it may be of greater benefit to select (multiple) secondary tasks that engage specific cognitive abilities to different extents, and to select SiN perception tests, which vary in properties such as speech target and/or background signals. This would allow the engagement of specific cognitive abilities for specific SiN listening conditions and/or listener groups to be experimentally manipulated.

In this study I addressed these issues by 1) testing two different SiN conditions, and 2) selecting multiple secondary tasks, which differently engage Working Memory sub-processes. By comparing the consequences of systematically engaging and disrupting specific cognitive abilities during SiN perception in different listening conditions, I also obviate the need for correlational analysis between dual-task performance and single-task cognitive ability.

The Baddeley model (Baddeley, 2000) was selected as a theoretical underpinning in the selection of secondary tasks because it defines and differentiates specific cognitive components, and is well established, intuitive to understand and easy to implement. Furthermore, it provides continuity to chapter three where tests were selected on the basis of the same model. In the previous chapter I demonstrated that SiN perception tests engage, the Central Executive, Episodic Buffer and Phonological Loop sub-domains of the Baddeley model in young and/or older listeners. Looking at younger listeners specifically, the Phonological Loop sub-domain was particularly engaged in 3-talker babble but not in speech-modulated noise. Hence, here I investigated SiN listening in two masking conditions that were matched on energetic properties but differed in informational properties – speech-modulated noise and 3-talker babble.

I selected four different secondary cognitive tasks that engaged the four sub-domains of the Baddeley model (Central Executive (CE), Visuo-Spatial Sketchpad (VSS), Episodic Buffer (EB), and Phonological Loop (PL)), to different extents. The cognitive tests I selected were: Digit Span Forward (DSF), Digit Span Backward (DSB), Corsi Span Forward (CSF), and Corsi Span Backward (CSB). The digit span tasks are verbal in modality and assess Episodic Buffer and Phonological Loop abilities, but the Digit Span Backwards differs by additionally assessing Central Executive abilities. Similarly, the Corsi span tasks are non-verbal in modality and assess Episodic Buffer and Visuo-Spatial Sketchpad abilities, but the Corsi Span Backward differs by additionally assessing Central Executive abilities. The cognitive tests also overlap in task type: Digit and Corsi Span Forward tasks both assess Episodic Buffer ability, and Digit and Corsi Span Backward tasks both assess Central Executive ability.

Figure 4. 1 below displays which sub-domains of the Baddeley model each of the cognitive tests assesses.

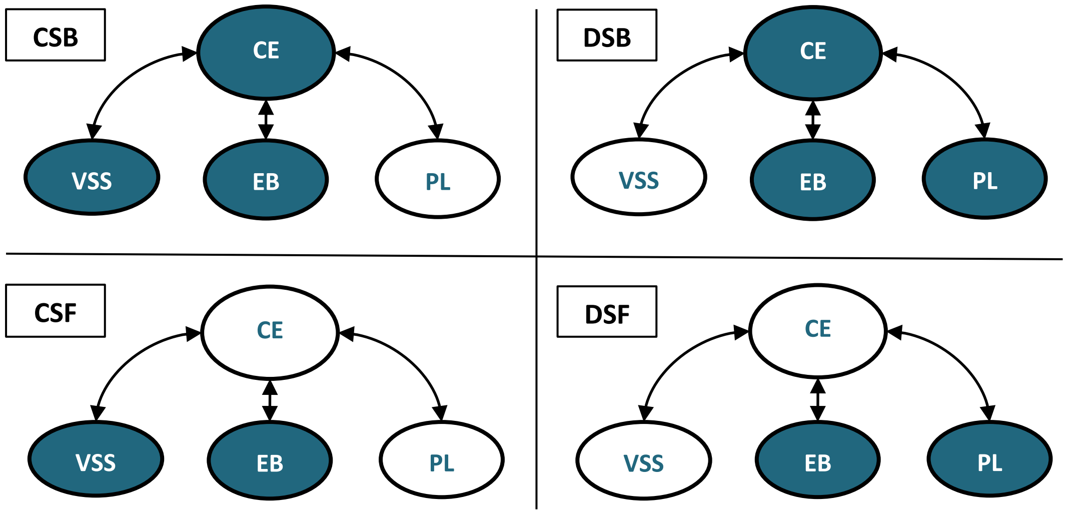


Figure 4. 1 – Baddeley model sub-domains engaged by cognitive tests

Baddeley model of working memory (Baddeley, 2000) flow charts indicating which cognitive sub-domains (CE: Central Executive, VSS: Visuo-Spatial Sketchpad, EB: Episodic Buffer, PL: Phonological Loop) are engaged by each of the four memory tasks, Corsi Span Backward (CSB), Digit Span Backward (DSB), Corsi Span Forward (CSF), and Digit Span Forward (DSF). For each memory task the sub-domains highlighted with a blue background are engaged and those with a white background are not engaged.

One important consideration for the presentation of a digit span task concurrently with a SiN perception test was auditory inference and masking between the two. In order to overcome auditory interference, a visual version of the digit span test was created. Although an auditory superiority effect has been reported in auditory versus visual versions of the digit span task, the tasks are thought to load onto the same cognitive domains in both modalities (Kemtes et al., 2008).

The design of this experiment allowed me to assess dual-task cost in two different listening conditions, for performance in each of four cognitive tests, and in performance of both the SiN perception and cognitive tasks. Additionally, I was also able to assess dual-tasks cost for specific cognitive abilities relating to the Baddeley model (instead of the four separate cognitive tasks).

Based on my previous study and the literature, for the younger adult group tested here, I have the following hypotheses:

1. Modality effect:

Verbal (digit span) tasks will show a greater proportional dual-task cost (pDTC) than non-verbal (Corsi span) tasks because they are in greater competition with the SiN perception test, which is an inherently verbal (Gagne et al., 2017).

2. Task effect:

   No effect of task (forward or backward) given the association study found no involvement of the Central Executive for SiN perception: as a consequence pDTCs should be similar between the forward and backward span, for both digit and Corsi dual-task conditions.

3. Masker effect:

   A greater pDTC for the 3-talker babble compared to speech-modulated noise masking condition in both digit span tasks, but not the Corsi span tasks because the digit span tasks cause great engagement of Phonological Loop ability.

4. Effects concerning the Baddeley domains:

   A greater role for the Phonological Loop and lesser role for the Central Executive and Episodic Buffer, would replicate my findings from the previous study.

4.2. Methods

4.2.1. Testing materials and procedure

Participants were pre-screened by email and self-reported to meet the following criteria: 1) aged between 18-30 years, 2) native English language speaker, 3) no diagnosed hearing loss, 4) no diagnosed neurological, psychiatric or language (including dyslexia) disorders or impairments. If all the criteria were met they were invited to take part in the study.

Twenty-four younger adult participants took part in the study (18 female, 6 male). They had a mean age of 23.2 (SD=3.1) years. Participants were recruited from the University of Nottingham and wider local population (Nottinghamshire) via electronic and paper advertisements.

Experimental Procedure

All testing was conducted in a sound-attenuated booth. For all participants screening tests were administered first. Then intelligibility was measured for all listeners in a word-in-noise task. These results were subsequently used to set individualised SNRs for the sentence-in-noise tasks in the main experiment. The main experiment involved single-task SiN perception of sentences, single-task cognitive tests, and finally the combined dual-task performance on SiN perception and cognitive tests. See Figure 4. 2 below for a flow chart of the experimental procedure. Note, the run order of the cognitive tasks within the single-tasks and all the dual-task conditions were randomised according to a fully randomised Latin square design to minimise any learning effects.

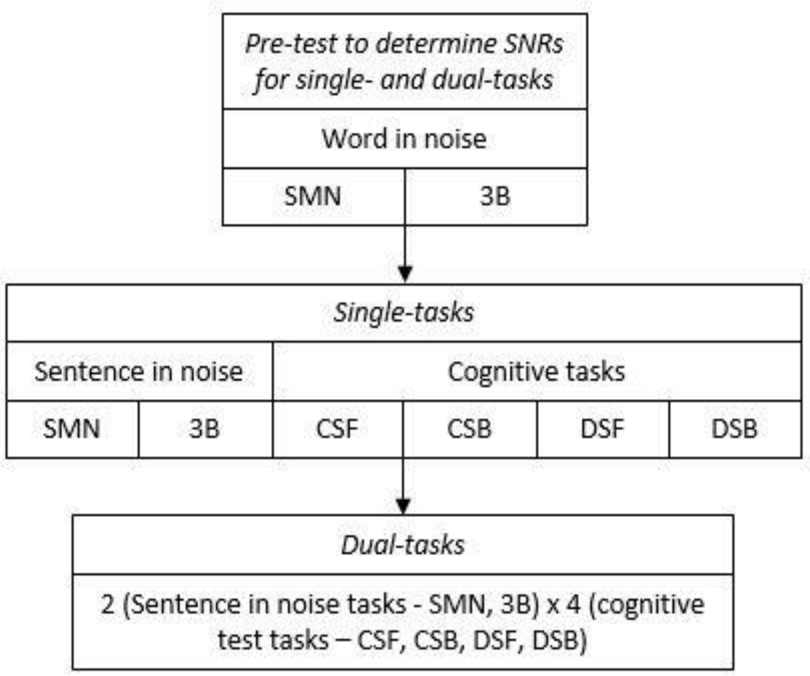


Figure 4. 2 – Experimental proceedure flow chart

Experimental procedure flow chart:

1) Pre-test to determine signal-to-noise ratios for SiN perceptiontests in the single- and dual-tasks

2) Single-tasks (2 sentence in noise tasks, 4 cognitive tasks)

3) Dual-tasks (2 sentences in noise tasks x 4 cognitive tests)

SMN: Speech-modulated noise, 3B: 3-talker babble, CSF: Corsi Span Forward, CSB: Corsi Span Backward, DSF: Digit Span Forward, DSB: Digit Span Backward

The total testing time per participant was approximately 1 hour 45 minutes. All auditory stimuli were presented monaurally to the left ear using over-the-ear headphones (Sennheiser® HD 280 pro 64Ω). The intensities of all auditory stimuli, including masking, were verified using an artificial ear (Brüel and Kjær, type 4153). All administered tests were displayed on a Dell® 23inch P2314T monitor using the software PsychoPy® (v2). Participants were sat at a viewing distance of approximately 50cm for the monitor display.

Speech-in-noise material

Words

The word stimuli consisted of 56 (16 practice trials and two experimental blocks of 20) monosyllabic words, with no word repeated from the final words in the sentences. The stimuli were recorded using a male speaker (different to the sentences) with a Standard Southern British English accent. See Supplementary Table 4.1 for a list of the word stimulus set.

Sentences

The sentence stimuli were taken from the British English Semantic Sentence Test (BESST) (Heinrich et al., 2014), which consisted of 113 high and low predictability sentences, with each sentence ending with a monosyllabic word. In this experiment I only used the low predictability sentences in order to make the task as challenging as possible. Sentences were recorded using a male speaker with a Standard Southern British English accent. See Supplementary Table 4. 2 for a list of the sentence stimulus set.

Maskers

The 3-talker babble, an informational masker, was created using the phonetics software Praat® (v5.4.06) by combining the voices of three talkers reading a different text passage. Two talkers were female and a third was male. All of them spoke English with different regional English accents. The speech-modulated noise, an energetic masker, was created in Matlab® (R2014b) from averaging the babble signal in chunks of 23 ms. All sound signals were low- and high-pass filtered between 50Hz and 10,000Hz. The target and masker stimuli were combined in Matlab® with noise stimuli starting and finishing 2 seconds before and after the speech target.

Signal-to-Noise Ratio (SNR)

For the word-in-noise pre-tasks the SNR ratios were set to the same SNRs for all participants, 0dB SNR for the 3-talker babble and -4dB SNR for the speech-modulated noise. These tasks were used as a pre-test to determine the SNRs for the sentence in noise tasks used in the main experiment.

For the sentence-in-noise tasks used in the main experiment the SNRs were individualised for each participant to approximate 70% intelligibility levels in the single-task SiN conditions based on the follow procedure:

SNR ratios were calculated based on the SiN intelligibility results from the younger adult group of the associations study (chapter three, section 3.2). In the association study, at in SNR of -2dB, mean intelligibility levels in the 3-talker babble conditions between words (51%) and low predictability sentences (45%) were not significantly different ($t(49)=2.10$, $p=.05$). In the speech-modulated noise conditions mean intelligibilities, both set at a SNR of -7dB, differed significantly between words (37%) and low predictability sentences (29%) ($t(49)=3.76$, $p<.001$). Based upon this, at equivalent SNRs, intelligibility is assumed to be approximately equal in the 3-talker babble masked conditions (for words and low predictability sentences) and mean intelligibility was assumed to differ by approximately 10% between words (higher) and low predictability sentences (lower).

A further assumption was that within any listening condition altering the SNR by 1dB would alter intelligibility levels by approximately 10%. This assumption is based on the findings of a systematic review by MacPherson et al. (2014). They found a 9% difference in SiN intelligibility per 1dB change in SNR for sentences in 3-talker babble masking, and an 8% difference for a 1dB change in SNR for modulated noise.

Using these assumptions, the word-in-noise task was used as a pre-test to determine the SNRs required to approximate 70% intelligibility in the sentence in noise task used in the main experiment. This procedure differed from the previous experiment where all participants within each age group heard each SiN condition at the same SNR. A different approach was taken here to more carefully control intelligibility levels because the extent to which cognition is engaged may vary depending upon SNR.

Table 4. 1 below displays how the intelligibility level in the word-in-noise task was used to determine the SNR level for the sentence-in-noise tasks. This was performed separately for the two masking conditions. For example, an intelligibility score of 70% in the word in 3-talker babble condition (presented at 0dB SNR) would correspond to a SNR of 0dB being selected for subsequent sentence in speech-modulated task to approximately 70%. An intelligibility score of 70% in the word in speech-modulated noise task (presented -4dB SNR) would correspond to a SNR of

-3dB (-4+1=-3) being selected in the subsequent sentence in 3-talker babble task to approximately 70%.

Masking condition			
Speech-modulated noise		3-talker babble	
Intelligibility performance for words in noise task	Required SNR (dB) for sentence in noise tasks	Intelligibility performance for words in noise task	Required SNR (dB) for sentence in noise tasks
100%	-6	100%	-3
90-99%	-5	90-99%	-2
80-89%	-4	80-89%	-1
70-79%	-3	70-79%	0
60-69%	-2	60-69%	1
50-59%	-1	50-59%	2
≤49%	0	≤49%	3

Table 4. 1 – Signal-to-noise ratio determination for main experiment

Table for selecting SNRs to approximate 70% intelligibility in low predictability *sentence target conditions based upon intelligibility level in word target conditions.*

Single-task procedure

For all SiN perception tests participants were instructed to listen carefully to each target stimulus and to repeat back out loud, the stimulus (when prompted) as best as they could. Participants were also encouraged to guess on any words if they were unsure and were advised that they would not be expected to be able to recall every word or sentence due to the difficulty of the task. Only the final word was scored for each sentence, and the target word was scored either correct or incorrect (i.e., any response other than the exact target word was considered to be incorrect). For each participant a proportion correct score was determined individually for the two background noise conditions.

The word target conditions were presented first because these tasks were used to determine the individualised SNRs for the sentence-in-noise tasks used in the main experiment. The word-in-noise tasks were tested in two separate blocks, the speech-modulated noise condition followed by the 3-talker babble condition. Each block began with a practice condition consisting of eight words, which was immediately followed by the two experimental conditions, each consisting of 20 words. Each target word was scored either correct or incorrect and a proportion response correct score was calculated separately for the two background noise conditions.

The sentence target conditions were also presented in two separate blocks, again with the speech-modulated noise masked condition proceeding the 3-talker babble. Each sentence block consisted of ten sentences. There was a practice trial for each of the two masker types, each consisting of one sentence. See Figure 4.1 above for a full chart of the experimental procedure. Sentence items within each block were fully counterbalanced between the single- and dual-task SiN conditions according to a fully randomised Latin square design.

Cognitive tests

Four cognitive tests were selected: Corsi Span Forward, Corsi Span Backward, Digit Span Forward, and Digit Span Backward. These tasks were designed to engage the sub-domains of the Baddeley model of Working Memory to different extents. The two non-verbal tasks, Corsi Span Forward and Backward both have a storage component (Visuo-Spatial Sketchpad and Episodic Buffer), but the Corsi Span Backward differs by also having a manipulation component (Central Executive). The two verbal tasks, Digit Span Forward and Backward, both have a storage component (Episodic Buffer and Phonological Loop), but differ the Digit Span Backward differs by also having a manipulation component (Central Executive). See Figure 4. 3 below for a schematic of the cognitive tests and the sub-domains they engage.

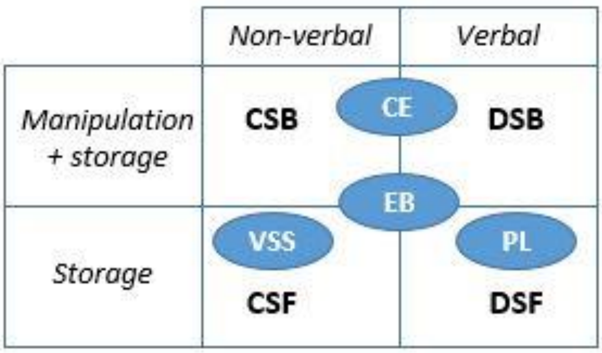


Figure 4. 3 – Baddeley model sub-domains assessed by each cogniitve test – including modality and component information

The four cognitive tests (Corsi Span Backward, Digit Span Backward, Corsi Span Forward, Digit Span Forward displayed in their relation to which of the Baddeley sub-domains (Central Executive, Visuo-Spatial Sketchpad, Episodic Buffer, Phonological Loop) they engage, and their modality (non-verbal, verbal) and task components (manipulation and storage, storage).

Corsi span (non-verbal): forward & backward

An electronic 2D version of the Corsi block tapping test was adapted from the original three-dimensional physical version. This is similar to the specifications in the electronic adaption of the Corsi span test by Brunetti et al. (2014). As the original this version consisted of nine blocks (30mm x 30mm blue frames on presented on a dark grey background) made to the same 2D configuration of the original - see Figure 3. 5a section 2.1.4 in chapter three.

To mimic the ‘tapping’ sequence, each target square flashed to the colour blue (from and returning to having no fill colour) for a duration of 0.5 seconds, with an interval of 1 second between block flashes. Sequences had a total path distance between 60 and 63mm, each sequence path containing two cross-overs. In both the forward and backward spans each sequence consisted of seven block flashes, presented at a rate of one per second with no individual block flashing more than once in each sequence. The number of items per list was set to seven in both forward and backward conditions. This ensured that both tasks shared the same taxation of the storage component (Visuo-Spatial Sketchpad and Episodic Buffer), yet both tasks differed in that the backward span had a manipulation (Central Executive) component.

A ‘recall’ prompt was presented at the bottom of the screen to signal to the participant it was time to respond. This was designed to parallel the physical version of Corsi span tests, were the participant had the benefit of a visual prompt of the experimenter’s hand moving away from the blocks when the sequence had finished being tapped.

Participants responded by clicking, using a mouse, the blocks in the correct sequence. When a participant clicked inside a block it would change to the colour blue to signify it has been clicked. In the forward condition, the participant was asked to repeat each sequence in the forward sequential order; in the backward condition participants were required to repeat each sequence in reverse sequential order. The two conditions, forward and backward, were run as separate trials, each made of 10 unique sequences. The forward condition was run prior to the backward. The span scores for the forward and backward conditions were calculated as proportion of correctly recalled items out of the total number of items (10); separate scores were given for the two conditions.

Digit span (verbal): forward and backward
The digit forward and backward spans were selected to assess verbal storage (Episodic Buffer and Phonological Loop) alone and verbal storage in conjunction with and manipulation (Central Executive) respectively (Wechsler, 2008).

In both tasks participants were required to attend to and memorise a list of visually presented strings of numbers; in the forward task, numbers were recalled in the order they were presented, and in the backward condition the numbers were recalled in reverse order.

Both conditions had a fixed length of eight digits per list. Within each sequence no one digit was repeated and no more than two consecutive digits were presented in ascending or descending order, e.g., a sequence of 1, 2, 3 or 3, 2, 1 would not be contained within any sequence. Digits were presented visually, one digit at a time (duration 0.5s) at a rate of one digit per second. Each digit was displayed in the centre of the monitor, in sans-serif font and with a height of approximately 2.8cm. The numbers of items per list was set to eight digits in both forward and backward conditions to ensure both tasked shared the same taxation of the storage component (verbal storage), but differed in only the backward span having a manipulation component.

Participants were asked to respond verbally, listing the digits in the correct sequence. The two conditions, forward and backward, were blocked as separate trials, each made of 10 unique sequences. The forward condition was run prior to the backward condition. The span scores for the forward and backward conditions were calculated as proportion of correctly recalled items out of the total number of items (10); separate scores were given for the two conditions.

Dual-task procedure
The dual-task conditions were always run after the single-task conditions. Both the same design and protocols as in the single-task. The stimulus presentation followed a concurrent design, i.e., the presentation of the SiN and cognitive task stimuli fully overlapped temporally. The run order of the dual-task conditions (cognitive task versus SiN background noise) was counter-balanced across all participants using a fully randomised Latin square design. Additionally, the sentence items within each block were fully counterbalanced between the single- and dual-task SiN conditions according to a fully randomised Latin square design.

Participants were instructed to attend to both tasks equally and to always repeat the memory span sequences (Corsi Span Forward and Backward, and Digit Span Forward and Backward) first and the target sentence second. The memory span response was required first to ensure the manipulation process in the backward span tests had taken place prior to the response to the SiN stimuli. Figure 4. 4 below shows an example of the dual-task procedure.

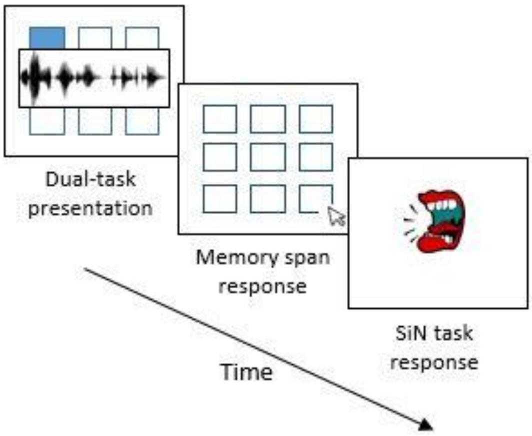


Figure 4. 4 – Dual-task proceedure

An example of dual-task procedure – Corsi span – not Corsi blocks are not shown in actual experimental configuration

4.2.2. Statistical methods

To measure the adverse effect of dual-task performance on intelligibility and cognitive performance the proportion Dual-Task Cost (pDTC) (Doumas et al., 2008) was calculated individually for each participant and each task using Equation 4. 1 stated below. Proportion dual-task cost was calculated separately for SiN and cognitive task performance. A positive proportional dual-task cost indicates a dual-task cost and a negative value indicates a dual-task advantage.

$$pDTC = -\frac{dual\ task\ accuracy - single\ task\ accuracy}{single\ task\ accuracy}$$

Equation 4. 1 – Proportion dual-task cost

The equation used to calculate proportion Dual-Task Cost (pDTC) in speech-in-noise, cognitive, SiN-cognition combined performance.

Prior to the proportion dual task cost calculation all SiN intelligibility scores were converted into Rationalized Arcsine Units (RAU) (Studebaker, 1985) for further analyses. RAU scores extend upper and lower extremes of a psychometric functions making it more linear and thereby increasing statistical power. A combined pDTC was also calculated for performance in both the primary (SiN) and secondary (cognitive) tasks because dual-task performance was found to differ with respect to single-task performance in both the SiN and cognitive tasks. The SiN cognition performance combined pDTC was calculated by taking the mean of the SiN intelligibly and cognitive performance pDTC, as stated below in Equation 4. 2. This was done individually for each participant and for each dual-task condition.

$$SiN\ Cognition\ combined\ pDTC = \frac{SiN\ pDTC + cognition\ pDTC}{2}$$

Equation 4. 2 – SiN and cognition combined proportion dual-task cost

The equation used to calculate SiN-cognition combined ability proportion Dual-Task Cost (pDTC)

All pDTC variables were assessed for normal distribution using the Shapiro-Wilk test of normality and by visual inspection of Q-Q plots and histograms. Two (of eight) SiN pDTC conditions (Digit Span Forward with speech-modulated noise masker, Corsi Span Forward with speech-modulated noise masker) and three (of eight) cognitive pDTC conditions (Digit Span Backward with 3-talker babble, Corsi Span Backward with 3-talker babble, and Corsi Span Backward with speech-modulated noise) showed significant deviations from normality based on the Shapiro-Wilk statistic. However, after a visual inspection of the Q-Q plots and box plots it was deemed acceptable to bring all results forward for parametric testing without further transformation.

Linear mixed model analysis was performed in R Studio (v1.0.153) and all other analysis was performed in IBM® SPSS® Statistics (v22). Bar and scatterplots were made using Microsoft Excel (2013).

Proportional dual-task cost was assessed in relation to the SiN intelligibility, cognitive performance, and SiN-cognition combined performance using three separate backward step linear mixed models. In a backwards step approach the full model (at the level of all four-way, and lower, interactions, and main effects) is the basis from which systematically one term is removed thereby assessing the importance of the removed term to the overall fit of the model (as expressed by the Akaike Information Criterion (AIC)). Terms which do not significantly contribute to the fit of the model are not added back in. This procedure was followed first for random effects, then for fixed effects.

Model fit was estimated using the AIC and models were compared using likelihood ratio tests. When assessing fixed effects this process was performed separately at each level. All main effects and all lower-level interactions part of a higher-level interaction were retained regardless of fit. Main effects from each model are reported using an ANOVA (type III, using the Satterthwaite method for calculating the denominator degrees of freedom), and estimates of fixed effects are also reported. Post-hoc analysis into fixed effects was assessed using paired t-tests (two-tailed and Bonferroni corrected for multiple comparisons).

In addition to standard dual-task costs for each combination for SiN and cognitive test performance (separate and combined) pDTCs were also determined for each of the sub-domains of the Baddeley model. This analysis was performed to make the results here more relatable to the results in the previous chapter (Central Executive, Episodic Buffer, and Phonological Loop). However, for cognitive and SiN-cognition combined performance, this was only performed for the Central Executive and Phonological Loop. This was calculated individually for each participant for performance in each sub-domain for SiN, cognitive, and SiN-cognition combined performance pDTC.

The equations are stated below in Equation 4. 3 a-c. For example, the SiN intelligibility pDTC for the Episodic Buffer is equal to the SiN intelligibility pDTC in the Corsi Span Forward condition since the Episodic Buffer is the only shared cognitive ability between to the two tasks.

Another example is for the Central Executive where there are two methods of isolating Central Executive ability and an average is taken of the two: SiN performance pDTC in the respective forward span tasks (Digit Span Forward and Corsi Span Forward) are subtracted from their respective backward span companions (Digit Span Backward and Corsi Span Backward). Where the backward span tasks have the addition of a Central Executive component compared to the forward span, while sharing the same Episodic Buffer and Phonological Loop or Visuo-Spatial Sketchpad components.

e.g.,

$$\textbf{CE}\ pDTC = DSB\ pDTC\ (\textbf{CE + EB + PL}) - DSF\ pDTC\ (\textbf{EB + PL})$$

and

$$\textbf{CE}\ pDTC = CSB\ pDTC\ (\textbf{CE + EB +VSS}) - CSF\ pDTC\ (\textbf{EB + VSS})$$

This was done separately for the two SiN masking conditions, 3-talker babble and speech-modulated noise. Then, three separate ANOVAs were conducted to assess sub-domain and masker main effects and interactions for SiN intelligibility, cognitive performance, and SiN-cognition combined performance pDTCs.

$$a)\ CE = \frac{(DSB - DSF) + (CSB - CSF)}{2}$$

$$b)\ EB = CSF$$

$$c)\ PL = \frac{(DSB - CSB) + (DSF - CSF)}{2}$$

Equation 4. 3 – Proprotional dual-task costs for each Baddeley sub-domain

Equations for SiN intelligibility proportional dual-task costs (pDTCs) for a) the Central Executive (CE), b) the Episodic Buffer (EB), and c) the Phonological Loop (PL) subcomponents of the Baddeley model. Note the EB equation was only used for the SiN intelligibly pDTC data. DSB (Digit Span Backward), DSF (Digit Span Forward), CSB (Corsi Span Backward), CSF (Corsi Span Forward).

4.3. Results

4.3.1 Single-task task performance

In the speech-in-noise tests mean single-task performance in speech-modulated noise was .79 (SE=.03) and .70 (SE=.03) in the 3-talker babble noise. A post-hoc T-test (two tailed) revealed this difference to be significant ($t(23)=2.29$, $p=.03$).

In the cognitive tests, the non-verbal modality (Corsi span) mean performance in the forward task was .75 (SE=.03) and in the backward task .72 (SE=.02). In the verbal modality (digit span) mean performance .69 (SE=.04) in the forward task and .55 (SE=.04) in the backward task.

A one-way ANOVA revealed a main effect of cognitive test type ($F(3, 92)=7.43$, $p<.001$). Paired t-test (two-tailed, Bonferroni corrected) analysis showed there were no significant differences between the Corsi Span Forward, Corsi Span Backward and Digit Span Forward cognitive performance. Performance in the Digit Span Backward was significantly different from all three other tests Corsi Span Forward (t(23)=6.77, *p<.001*), Corsi Span Backward (t(23)=4.85, *p<.001*) and Digit Span Forward (t(23)=5.17, *p<.001*).

Figures 4. 5 and 4. 6 below show raw single- and dual-task performance in SiN intelligibly and cognition respectively. Performance in the two SiN masking conditions (3-talker babble and speech-modulated noise) are shown separately. Dual-task performance (SiN and cognitive test) data will be presented in more detail the next three sections.

Single task performance in the SiN perception tests was also assessed for an association with PTA. Correlation analysis (Pearson's, two-tailed) showed no significant association between PTA and SiN intelligibility in either 3-talker babble (r(24)=-.21, *p=.322*) or speech-modulated noise (r(24)=-.09, *p=.658*) conditions. As there was no association between PTA and SiN perception, PTA was not included as a predictive factor in the linear mixed model analysis.

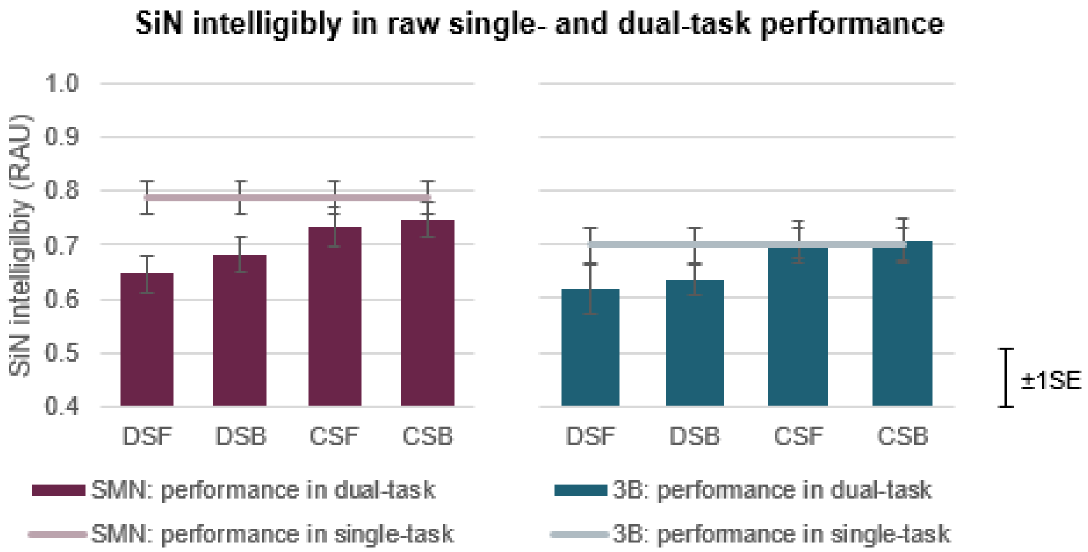


Figure 4. 5 – Bar charts displaing raw single- and dual-task performance in the SiN perception tests

Mean and Standard Error (SE) of raw single- and dual-task SiN intelligibility

SMN=Speech Modulated Noise, 3B=3-talker babble, CSF= Corsi Span Forward, CSB=Corsi Span Backward, DSF=Digit Span forward, DSB=Digit Span Backward

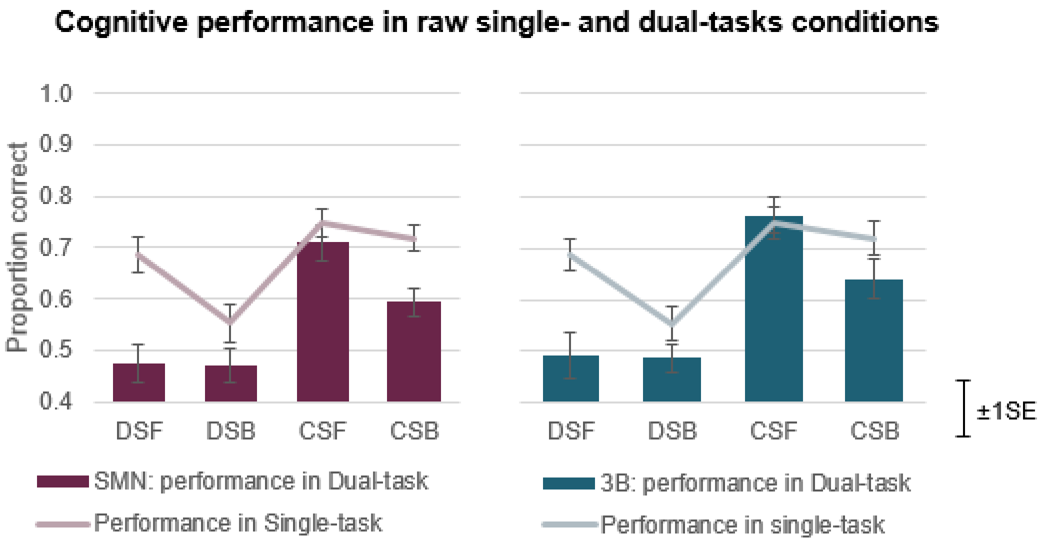


Figure 4. 6 – Bar charts displaying raw single- and dual-task performance in the cognitive tests

Mean and Standard Error (SE) of raw single- and dual-task cognitive performance

SMN=Speech Modulated Noise, 3B=3-talker babble, CSF= Corsi Span Forward, CSB=Corsi Span Backward, DSF=Digit Span forward, DSB=Digit Span Backward

4.3.2 Speech-in-Noise intelligibility dual-task cost

An ANOVA of the best fitting linear mixed model (AIC=5.64) predicting speech-in-noise intelligibility proportional dual-task cost showed a significant main effect for modality ($F(1, 144)=19.37$, $p<.001$) (verbal (digit span) > non-verbal (Corsi span)), but not task ($F(1, 144)=.80$, $p=.37$) and masker ($F(1, 24)=1.53$, $p=.23$). See Table 4. 2 below for a summary of the fixed effects and model fit, and see Figure 4. 7 for a plot of the main effect of modality.

AIC value	Fixed effects	
5.64	Condition	Interaction
	Modality	*none*

Table 4. 2 – Summary of significant fixed effects from the linear mixed model for SiN intelligibility pDTC

ANOVA of fixed effects of linear mixed model assessing the relationship between modality, task, masker and SiN intelligibility proportional dual-task cost. The model fit is indicated with the AIC (a lower number showing a better fit).

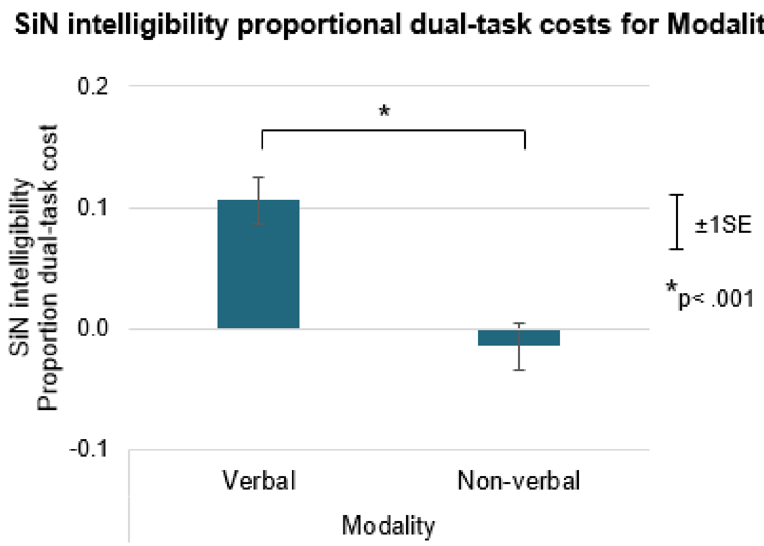


Figure 4. 7 – Bar chart displaying the modality main effect for SiN pDTC

Mean and standard error (SE) of SiN proportional dual task cost for Modality (verbal (digit span), non-verbal (Corsi span)– averaged over task (forward and backward) and masker type (speech-modulated noise and 3-talker babble)

4.3.3 Cognitive performance dual-task cost

The best fitting linear mixed model (AIC=-141.62) predicting cognitive performance proportional dual-task cost showed (ANOVA) a significant main effect for modality ($F(1, 27.83)=19.29$, $p<0.001$) (verbal/digit span>non-verbal/Corsi span), and an interaction between modality and task ($F(1, 166.29)=51.44$, $p<.001$). However, there were no significant main effect of task ($F(1, 166.29)=2.01$, $p=.158$) nor masker ($F(1, 166.29)=3.18$, $p=.077$).

With regards to the linear mixed model, β coefficients showed that pDTC was .28 less for the non-verbal forward (Corsi Span Forward) compared to the verbal forward (Digit Span Forward) task (β=-.28, SE=.04, t=-7.81, $p<.001$) and that pDTC for task was .18 less in verbal backward (Digit Span Backward) compared to verbal forward (Digit Span Forward) condition (β=-.18, SE=.03, t=-6.07, $p<.001$). Additionally, the model showed a significant interaction term for modality and task interaction, the β coefficient of the interaction was .30 (SE=.04, t=7.17, $p<.001$). This showed that in the verbal modality (digit span) pDTC was (.18) greater in the Forward versus the Backward span (DSF>DSB), whereas in the non-verbal modality (Corsi span) pDTC was .12 greater (.30 -.18=-.12) in backward versus the forward span (CSF<CSB).

See Table 4. 3 below for a summary of the fixed effects and model fit, see Figure 4. 9 for a bar chart of modality main effect (verbal and non-verbal), and for Figure 4. 8 a bar chart for modality (verbal and non-verbal) and task (forward and backward) interaction.

AIC value	Fixed effects	
-141.62	Condition	Interaction
	Modality	Modality*Task

Table 4. 3 - Summary of significant fixed effects from the linear mixed model for cognitive performance pDTC

ANOVA of fixed effects of linear mixed model assessing the relationship between modality, task, masker and cognitive test proportional dual-task cost. The model fit is indicated with the AIC (a lower or negative number showing a better fit)

Further analysis (one-sample t-tests, test value=0, two tailed) to examine if pDTC differed from zero, i.e., no dual-task cost, revealed that pDTC for the verbal forward (Digit Span Forward) ($t(23)=8.17$, $p<.001$), verbal backward (Digit Span Backward) ($t(23)=3.34$, $p=.01$), and non-verbal backward (Corsi Span Backward) ($t(23)=6.14$, $p<.001$) conditions were significantly different from zero. However, the non-verbal forward (Corsi Span Forward) condition was not significantly different from zero ($t(23)=.87$, $p=.40$).

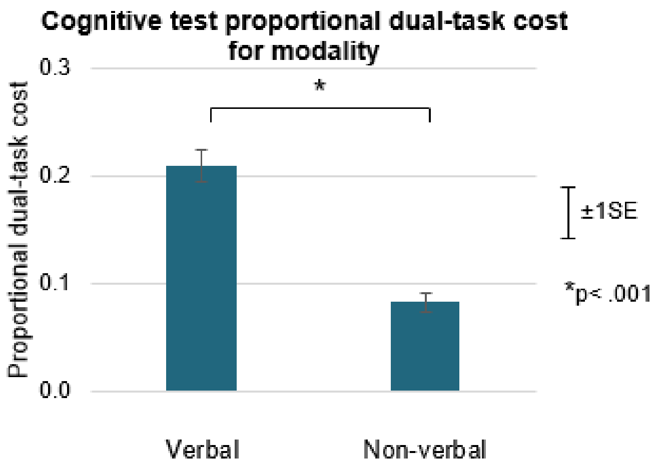


Figure 4. 9 – Bar chart displaying the modaility main effect for cognitive performance pDTC

Mean and standard error (SE) of memory task proportional dual-task cost for modality type (verbal (digit span) and non-verbal (Corsi span)), averaged over task (forward and backward) and masker type (speech-modulated noise and 3-talker babble)

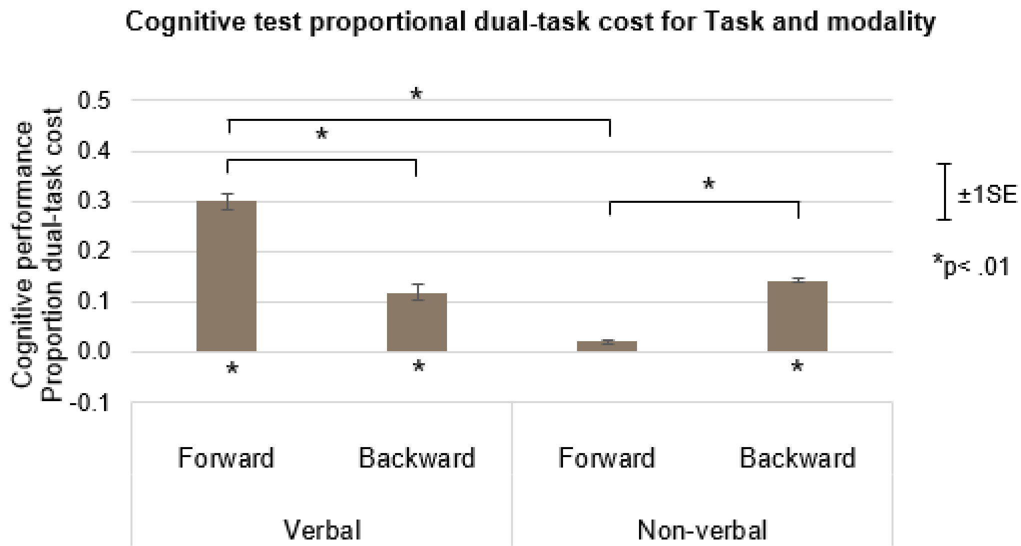


Figure 4. 8 – Bar chart displaying task and modality interaction for cognitive performance pDTC

Mean and standard error (SE) for cognitive task performance proportional dual task cost for Task (forward and backward) within Modality (verbal and non-verbal) and averaged across both (speech-modulated noise and 3-talker babble) masker types.

Verbal Forward = Digit Span Forward, Verbal Backward = Digit span Backward, Non-verbal Forward = Corsi Span Forward, Non-verbal Backward = Corsi Span Forward.

4.3.4 SiN-cognition combined performance dual-task cost

The best fitting linear mixed model (AIC=-170.73) predicting SiN-cognition performance combined pDTC showed (ANOVA) a significant main effect for modality (Verbal > non-verbal) ($F(1, 144.00)=43.53$, *p<.001*), and an interaction between modality and task ($F(1, 144.00)=22.15$, *p<.001*). There was no significant main effect for task ($F(1, 32.18)=1.56$, *p=.22*) and masker ($F(1, 24.09)=2.74$, *p=.11*).

With regards to the linear mixed model, β coefficients showed that pDTC was .20 less for the non-verbal forward (Corsi Span Forward) compared to the verbal forward (Digit Span Forward) task (β=-.20, SE=.03, t=-7.99, *p<.001*) and that pDTC for task was .11 less in verbal backward (Digit Span Backward) compared to verbal forward (Digit Span Forward) condition (β=.-11, SE=.03, t=-3.93, *p<.001*). Additionally, the model showed a significant interaction term for modality and task interaction, the β coefficient of the interaction was .17 (SE=.04, t=4.71, *p<.001*). This showed that in the verbal (digit span) modality pDTC was (.11) greater in the Forward versus the Backward span (DSF>DSB), whereas in the non-verbal modality (Corsi span) pDTC was .06 greater (.17 -.11 =.06) in backward versus the forward span (CSF<CSB).

Table 4. 4 below for a summary of the model main effects and fit, Figure 4. 10 displays a bar chart of modality main effect (verbal (digit span) and non-verbal (Corsi span)), and Figure 4. 11 displays a bar chart for modality (verbal (digit span) and non-verbal (Corsi span)) and task (forward and backward) interaction.

AIC value	Fixed effects	
-170.73	Condition	Interaction
	Modality	Modality*Task

Table 4. 4 - Summary of significant fixed effects from the linear mixed model for SiN and cognitive performance combined pDTC

ANOVA of fixed effects of linear mixed model assessing the relationship between modality, task, and masker and for SiN-cognition combined performance proportional dual-task cost. The model fit is indicated with the AIC (a lower or negative number showing a better fit).

Further analysis (one-sample t-tests, test value=0, two tailed) to examine if pDTC differed from zero, i.e., no dual-task cost, revealed that for the verbal forward (Digit Span Forward) ($t(23)=8.71$, *p<.001*) and verbal backward (Digit Span Backward) (t(23)=4.75, *p<.001*) were significantly from zero. The non-verbal forward (Corsi Span Forward) (t(23)=.38, *p=.71*) and non-verbal backward (Corsi Span Backward) (t(23)=2.23, *p=.14*) conditions were not significantly different from zero.

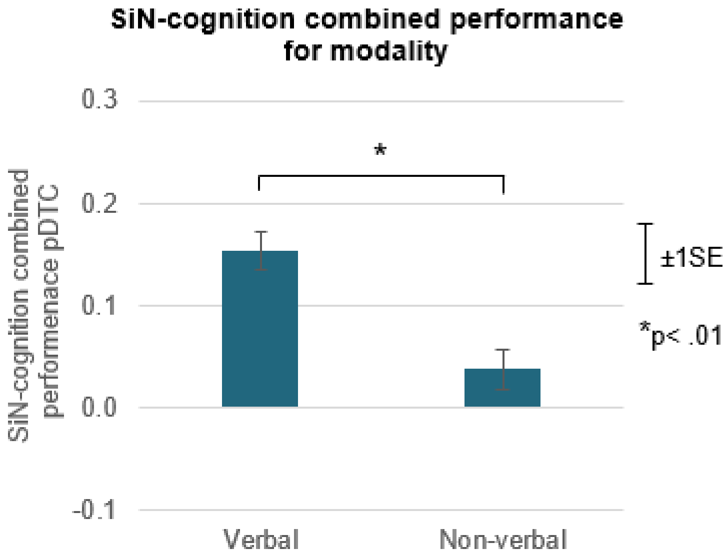


Figure 4. 10 – Bar chart displaying the modality main effect for SiN-cognition combined performance pDTC

Mean and standard error (SE) for SiN-cognition combined performance proportional dual-task cost for modality type (verbal and non-verbal), averaged over task (forward and backward) and masker type (speech-modulated noise and 3-talker babble)

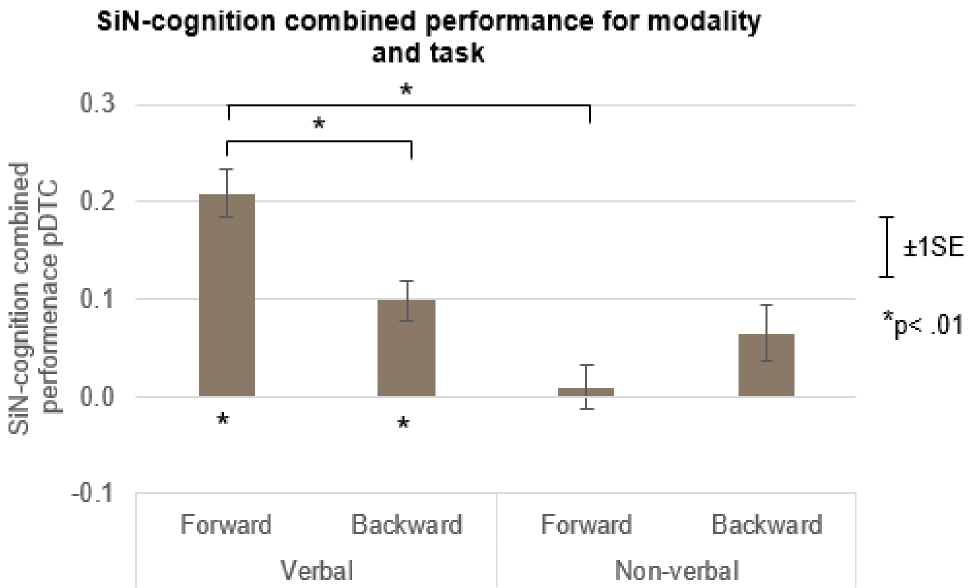


Figure 4. 11 Bar charts displaying the modality and task interaction for SiN-cognition combined performance pDTC

Mean and standard error (SE) for SiN-cognition combined performance proportional dual task cost for Task (forward and backward) within Modality (verbal and non-verbal) and averaged across both (speech-modulated noise and 3-talker babble) masker types.

Verbal Forward = Digit Span Forward, Verbal Backward = Digit span Backward, Non-verbal Forward = Corsi Span Forward, Non-verbal Backward = Corsi Span Forward.

4.3.5 Performance dual-task cost for the Baddeley model sub-domains

Figure 4. 12 below shows a bar plot of SiN intelligibility, cognitive performance, and both tasks combined performance pDTC for the Baddeley sub-domains.

SiN intelligibility proportional dual-task cost

The analysis, a 3 (Domain) x2 (Masker) ANOVA for SiN intelligibility pDTC for each of the Baddeley sub-domains (excluding the Visuo-spatial Sketchpad) showed that there was a significant main effect for sub-domain ($F(2, 22)=6.83$, *p=.005*), no significant effect for masker ($F(1, 23)=.55$, *p=.47*), and no interaction between sub-domain and masker ($F(2, 22)=.92$, *p=.41*).

The sub-domain main effect showed that the SiN intelligibility pDTC was significantly greater for the Phonological Loop than the Central Executive (*p=.004*). There were no significant differences between the Phonological Loop and Episodic Buffer (*p=.08*) and the Central Executive and Episodic Buffer (*p=1.00*). Additionally, a one-sample t-test (two-tailed) showed the pDTC to be significantly different from zero for the Phonological Loop ($t(23)=4.34$, $p<.001$), and not for the Central Executive ($t(23)=-.86$, $p=.40$) and Episodic Buffer ($t(23)=-.10$, $p=.92$).

Cognitive performance proportional dual-task cost

The analysis, a 2 (Domain) x2 (Masker) ANOVA for cognitive performance pDTC for each of the Baddeley sub-domains (excluding the Visuo-spatial Sketchpad and Episodic Buffer) showed that there was a significant main effect for sub-domain ($F(1, 23)=15.21$, $p=.001$) (Phonological Loop > Central Executive), no significant effect for masker ($F(1, 23)=.27$, $p=.65$), and no interaction between sub-domain and masker ($F(1, 23)=.22$, $p=.64$). Additionally, a one-sample t-test (2-tailed) showed the pDTC to be significantly different from zero for the Phonological Loop ($t(23)=3.69$, $p=.001$), and not for the Central Executive ($t(23)=-1.06$, $p=.30$).

SiN-cognitive combined performance proportional dual-task cost

The analysis, a 2 (Domain) x2 (Masker) ANOVA for cognitive performance pDTC for each of the Baddeley sub-domains (excluding the Visuo-spatial Sketchpad and Episodic Buffer) showed a significant main effect for sub-domain ($F(1, 23)=28.63$, $p=.001$) (Phonological Loop > Central Executive), no significant effect for masker ($F(1, 23)=1.04$, $p=.32$), and no interaction between sub-domain and masker ($F(1, 23)=.02$, $p=.88$). A one-sample t-test (two-tailed) showed the pDTC to be significantly different from zero for the Phonological Loop ($t(23)=6.07$, $p<.001$), and not for the Central Executive ($t(23)=-1.28$, $p=.21$).

Comparison between dual-task cost methods

A comparison between pDTC measures (3 (pDTC measures) x 2 (sub-domains)) ANOVA) showed that there was no significant effect of pDTC measure (F(2, 22)=.15, *p=0.86*), a significant effect of sub-domain (F(1, 23)=29.61, *p<.001*) (PL >CE), and no significant interaction between pDTC measure and sub-domain (F(2, 22)=.08, *p=0.92*).

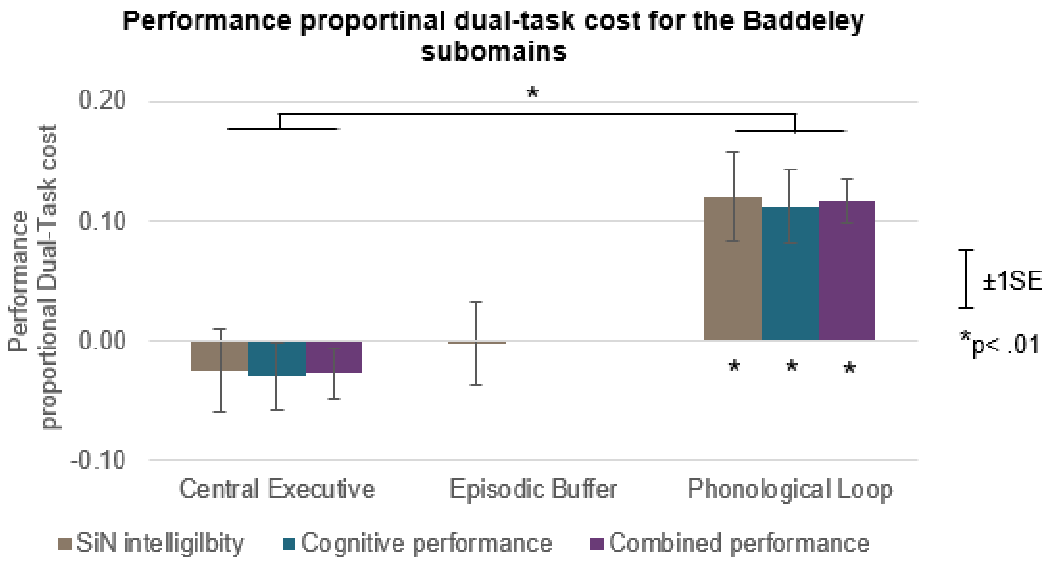


Figure 4. 12 – Bar chart displaying pDTCs for each Baddeley domain

Mean and standard error (SE) of SiN and cognitive performance proportional dual task cost for Baddeley sub-domain (average across masker type). Note a cognitive performance pDTC was not calculated for the Episodic Buffer.

4.4. Discussion

In the association study (Chapter three) I showed a number of ways in which differences in various cognitive abilities predict differences in a number of different SiN listening conditions. If these predictive differences are meaningful then it should be possible to manipulate these cognitive abilities and see an effect on SiN perception. This is what this final experiment attempted to achieve.

Using a dual-task approach I investigated the role of Working Memory for SiN listening for younger adult listeners. Here, using the Baddeley model of Working Memory (Baddeley, 2000) as a theoretical framework, I selected four different cognitive tests (Corsi Span Forward, Corsi Span Backward, Digit Span Forward, Digit Span Backward) designed to engage the sub-domains (Central Executive, Visuo-spatial Sketchpad, Episodic Buffer, and Phonological Loop) of the Baddeley model to different extents.

My hypotheses were as follows:

1) Modality effect: verbal (digit span) tasks will show a greater dual-task cost (DTC) than non-verbal (Corsi span) tasks
2) Task effect: pDTCs should be similar between the forward and backward span, for both digit and Corsi dual-task conditions
3) Masker effect: A greater pDTC for the 3-talker babble compared to speech-modulated noise masking condition in both digit span, but not the Corsi span task
4) Baddeley domain: A greater role for the Phonological Loop and lesser role for the Central Executive and Episodic Buffer

4.4.1 Speech-in-noise intelligibility dual-task cost

Modality effect

The SiN intelligibility pDTC analysis revealed a main effect for modality: verbal (digit span) > non-verbal (Corsi span). This result was in line with my prediction that there would be a greater performance dual-task cost for the digit versus the Corsi span tasks because SiN perception in an inherently verbal task. This finding is in consensus with previous studies, which have linked verbal Working Memory capacity to DTC performance in secondary cognitive tasks (Desjardins et al., 2013; Sarampalis et al., 2009).

Task effect

The SiN intelligibility pDTC revealed no significant differences in pDTCs between forward and backward span tasks. This result, in agreement with my hypothesis, suggests there is no engagement of the Central Executive since the backward span tasks have the addition of a manipulation/Central Executive component. This result also provides further explanation to recent findings that Working Memory ability is not important for younger adults for normal hearing (Füllgrabe et al., 2016b).

Masker effect

Here I did not find an effect of masker. This result suggests that the Working Memory abilities involved for SiN perception in informational (3-talker babble) versus energetic (speech-modulated noise) do not differ. However, this is in contradiction with the results of the association study, which found an association between Phonological Loop and SiN perception in 3-talker babble, but not for speech-modulated noise. I will discuss potential explanations for these difference in the limitations section.

Baddeley domains

The Baddeley domain analysis showed a significant pDTC for the Phonological Loop, but not the Episodic Buffer or Central Executive. This result supports my hypothesis that the Phonological Loop is the most important sub-domain of Working Memory and that there is limited involvement of the Central Executive for normal hearing younger listeners.

In relation to the association study literature, previous studies have found a role of verbal storage for SiN listening in older adults with normal hearing (Füllgrabe et al., 2015; Uslar et al., 2013) and a range of hearing sensitivities (Anderson et al., 2013). However, the cognitive tests used to measure verbal storage overlap in terms of engaging the Episodic Buffer and the Phonological Loop. It is in differentiating between these sub-domains that this current study is able to build on previous work.

4.4.2 Cognitive performance dual-task cost

Modality effect

The cognitive dual-task cost analysis showed a main effect of modality and an interaction between modality and task. The modality effect, as in the SiN intelligibility pDTC analysis, showed a greater cost in the verbal (digit span) versus the non-verbal (Corsi) tasks. Again, this is in agreement to previous dual-task study findings in linking verbal Working Memory with SiN perception (Desjardins et al., 2013; Sarampalis et al., 2009).

Task effect

In contrast with the SiN intelligibility pDTC results, here I saw a modality-task interaction. The interaction showed that there was a greater dual-task cost for the Digit Span Forward compared to the backward span indicating a greater role for storage (Phonological Loop, Episodic Buffer) versus the Central Executive. The reverse was true of Corsi conditions, where pDTC was higher in the backward versus the forward span. This result indicates there might be a role for the Central Executive in SiN perception, and perhaps this role is more detectible using a non-verbal method of assessing Central Executive. Finally, higher pDTCs in the Digit Span Forward versus the Corsi Span Forward condition may suggest that the Phonological Loop plays a greater role for SiN perception than the Episodic Buffer.

With regards to the digit span, pDTC were either expected to be approximately the same between the forward and backward conditions if there was no involvement of the Central Executive or higher in the backward span if there was an involvement of the Central Executive. Here, against both of those predictions I observed a higher pDTC in the forward condition. An explanation for this is that in the single-task conditions performance was at approximately 70% in the Digit Span Forward, but at approximately 50% for the Digit Span Backward. Therefore, the tasks were not matched in performance/difficulty level and in the case of the backward span task cognitive resources may have operated near capacity even in the single task condition and therefore only a limited pDTC was observable.

Furthermore, this result differed from the SiN intelligibility pDTC analysis were the non-verbal tasks were found not to disrupt SiN intelligibility and no task differences were observed. This difference may have been due to factors such as prioritisation, with participants giving priority to SiN perception regardless of the cognitive test. If non-verbal secondary tasks engaged cognitive processes important for speech perception after all, this may have been more visible in the dual-task costs of the non-prioritised tasks.

Masker effect

Again, as per the SiN pDTC analysis, I did not find an effect of masker. I expected to see a greater pDTC in the 3-talker babble versus the speech-modulated noise if the informational properties of the 3-talker babble masker further engaged Working Memory, and specifically the Phonological Loop sub-domain.

Baddeley domains

The Baddeley domain analysis showed a significant pDTC for the Phonological Loop, but not for the Central Executive. This result agrees with the findings of the SiN intelligibly pDTC analysis. Here I was unable to make a direct inference for the Episodic Buffer since the performance as measured by performance in Corsi forward condition does not differentiate between Episodic Buffer and Visuo-Spatial Sketchpad abilities.

4.4.3 SiN-cognition combined performance dual-task cost

Modality effect

The SiN-cognition combined performance findings replicated the SiN intelligibly and cognitive performance pDTC analysis in finding a main effect of modality. Again, finding a greater pDTC was found in the verbal versus the non-verbal modality.

Task effect

This result also replicated the cognitive performance pDTC results in showing an interaction between modality and task. However, the nature of this interaction differed in that pDTC performance was no longer significantly different between the Corsi span tasks and furthermore neither cost differed from zero (no pDTC). This result strengthens the argument that Central Executive and Episodic Buffer abilities may play a limited role for SiN perception for younger adult listeners. Furthermore, it again suggests the Phonological Loop to play the greatest role for SiN perception.

Masker effect

Again, as in the previous sets of pDTC analysis, I did not find an effect of masker.

Baddeley domains

The Baddeley domain analysis showed a significant pDTC for the Phonological Loop, but not for Central Executive. Again, this result agrees that the Phonological Loop plays the greater role for SiN perception and that there is only a limited role for the Central Executive.

4.5 Limitations

This current study did not find differences in the role of cognition for SiN perception in different background listening conditions (3-talker babble and speech-modulated noise). The hypothesis that the engagement of cognitive processes for SiN perception differed depending on background masker was based on previous publications suggesting an increased engagement of cognitive abilities in informational versus energetic masking (Freyman et al., 2004; Janse, 2012; Mattys et al., 2009). The specificity of the Phonological Loop ability prediction was based on my previous study where I found an association between the Phonological Loop and SiN perception in 3-talker babble, but not speech-modulated noise masker conditions.

One possible explanation for the difference in results is the cognitive test selection. For the dual-task study I selected memory span tasks while in the association study I selected rhyme verification tasks to assess the Phonological Loop. It is possible that the Phonological Loop domain assessed in the association study also captured other cognitive abilities not captured here, such as linguistic or long-term memory abilities. Here, the Phonological Loop was not assessed in the same way because I here selected tests, which both varied and overlapped, i.e., all tests shared an Episodic Buffer component, in the engagement of the four sub-domains of the

Baddeley model of Working Memory. Whereas in the association study I purposely selected the tests to engage Phonological Loop processing ability, whilst having less engagement of Episodic Buffer ability.

Another explanation for the difference in results may lie in the fact that this study did not account for the Impurity Principal (task- and test-specific variation) (Surprenant et al., 2009) in the same way that the association study did by using principal components analysis. It was not possible or indeed appropriate to apply this approach here because cognitive tests were selected to overlap in the cognitive abilities they assess and not to specifically assess the ability as a single sub-domain.

A further explanation is in the difference between the studies could be driven by the difference in approach and intelligibility levels between the studies. In the association study the SNR ratios were fixed at -2dB in the 3-talker babble masker a -7dB, and mean intelligibility levels were 30 (SE=1.8) and 45 (SE=2.4) RAU respectively. In the current study on the other hand SNRs were individualised for each participant and were selected to approximate 70% thresholds. The mean SNRs were 1.6dB (SE=.2) and -2.0dB (SE=.3) for 3-talker babble and speech modulated noise. This led to actual mean intelligibility levels of 79 (SE=3.3) and 70 (SE=3.4) RAU respectively for 3-talker babble and speech-modulated noise. The differences in intelligibility levels could have potential repercussions for the involvement and engagement of hearing sensitivity and cognitive abilities for SiN perception.

4.6 Conclusions

The results of this study revealed that verbal secondary tasks cause disruption to SiN perception and this effect is driven by the engagement of Phonological Loop ability. The results also indicated that there is limited or no contribution of Central Executive and Episodic Buffer abilities for SiN perception in the listener demographics (younger adults with normal hearing) tested in this study. However, Central Executive ability, as assessed by the Digit Span Backward test, was set at a higher level of difficulty compared to other conditions and this may have contributed to the apparent lack of or low involvement of the Central Executive domain found in this study.

Additionally, by analysing pDTC performance separately for SiN intelligibility and cognitive performance I was able to assess similarities and differences between the two approaches. pDTCs in SiN intelligibility may only be measurable if the secondary task is verbal. Here I have also demonstrated that presenting verbal information in a visual format is sufficient in showing this effect. The differences in results across the three sets of pDTC analyses show that participants may have prioritised speech perception regardless of task instruction.

Overall this study has demonstrated that using a dual-task approach can be useful in investigating the role of Working Memory/cognition for SiN perception. Furthermore, it has provided further evidence for the specific roles of Working Memory sub-processes. It would be of great interest to extend this study to older adults and listeners with hearing loss to allow the contributions of age and hearing loss to be investigated.

Chapter five - General discussion

5.1 Summary

This thesis investigated the association between cognition and SiN perception performance from multiple perspectives, including primary data and published evidence across different SiN conditions and different listener groups. The research employed systematic and theory-guided approaches to help advance the field and address key gaps in the research literature.

In section 5.1 I will highlight the key findings of each study, in section 5.2 I will discuss general limitations of the thesis, and section 5.3 I will draw general conclusion from the thesis as a whole.

5.1.1 Chapter two: Cognition, SiN perception and r≈.3

As a first step in the investigation I carried out a systematic review and meta-analysis of the cognitive hearing science literature, taking a specific focus on publications using an association study design, a popular approach in the field. This was a unique approach and built on previous review papers, which had either used a qualitative approach (Akeroyd, 2008) or focused on a single cognitive ability (Füllgrabe et al., 2016b).

In the preparation of the review it became evident that there was a vast range of cognitive and SiN perception tests, and the combinations therein. To allow for meaningful interpretation it was necessary to categorise both cognitive and SiN perception tests.

Cognitive tests were categorised into theorised cognitive domain/sub-domains based on established cognitive theory (Baddeley, 2000; Diamond, 2013; Miyake et al., 2000; Petersen et al., 2012; Salthouse, 2000).This allowed for the role of specific and theoretically conceptualised cognitive abilities to be assessed. However, although this is an important step in systematising cognitive abilities it does not account for the problem of skill-specific and task-specific variance (Surprenant et al., 2009).

The SiN perception tests were categorised based on two dimensions, foreground target and background masker. The target was categorised on a continuum of linguistic complexity, and masker on a continuum of energetic and informational properties. These features of SiN perception tests were selected because I theorised that the role and contributions of cognitive abilities may vary depending upon specific target/distractor combinations in listening conditions.

In addition to the categorisations of cognitive and SiN perception tests, hearing sensitivity of listeners in studies was also considered. Despite including only studies in the review where participants had hearing loss in a pre-clinical range, listener variation in hearing sensitivity ranged from normal hearing to moderate levels of HL. Due to the design of most studies it was not possible to discretely categorise HL groups. As a result, HL was categorised into two overlapping groups, normal hearing to mild HL, and normal hearing to moderate HL. Hearing sensitivity was of interest because it may lead to differences in the engagement of cognitive processes in SiN perception. Also, the engagement of cognition by different levels of HL may vary depending upon SiN listening condition.

The main findings of the systematic review and meta-analysis are summarised below:

1) Cognition and SiN perception have a general association of ≈.3
2) There is an imbalance in the selection of cognitive and SiN perception tests across the literature
3) Inhibitory Control, Working Memory, Episodic Memory and Processing Speed play a role for SiN perception
4) No difference was found in association between cognition (Working Memory) and SiN perception in the two groups that differed in the amount of HL: normal hearing to mild HL and normal hearing to moderate HL groups

In more detail, firstly, when assessing cognitive ability and SiN perception as single respective domains the association was approximately .3. However, this analysis also revealed a high heterogeneity suggesting that there was variance between studies. I explored cognitive abilities, SiN listening condition and hearing sensitivity as possible sources of this variance.

Secondly, due to imbalances in selections of cognitive and SiN perception tests a comprehensive assessment of the role for cognition for different SiN listening conditions was not possible. Moreover, not all theorised cognitive domains and SiN listening conditions were available to be included in the meta-analysis due to an insufficient number of studies investigating particular cognitive domains. Therefore, the review highlights an imbalance in cognition and SiN perception test selection and a need for further investigation into cognitive domains such as, attention, task switching and Fluid intelligence, and SiN listening conditions such as, single word target signal and modulated noise background signals.

Thirdly, not all cognitive abilities included in the meta-analysis were significantly associated with SiN listening. Whereas Inhibitory Control, Working Memory, Episodic Memory and Processing Speed showed a statistically significant association, Crystallised IQ did not.

Finally, the associations between SiN perception and cognition were similar between the relatively better hearing group (normal hearing to mild HL) and relatively poorer hearing group (normal hearing to moderate HL). The same was true when only Working Memory was considered. Furthermore, in the Working Memory sub-analysis there was heterogeneity in the normal hearing to mild HL group, but not in the normal hearing to moderate HL group. This suggests that while Working Memory ability may be more important for both better and poorer hearing listeners, processes other than Working Memory ability may account for more of the individual differences observed in better hearing compared to poorer hearing listeners.

One conclusion was that in order to further our understanding it would be beneficial to consider cognitive test selection within the context of standard cognitive models so that the domains could be understood in terms of underlying cognitive processes, not surface tests. Another conclusion was a lack of systematicity in the selection of target and masker signal combinations of the SiN perception tests. In order to fully and adequately explore if the role of cognition and HL differs depending on listening condition it is important to take a systematic approach in investigating different SiN target/masker combinations.

5.1.2 Chapter three: younger and older adult listeners use different listening strategies depending on SiN listening condition

In the next chapter of my thesis (chapter three) I conducted a cognition and SiN association experiment building on the conclusions from the systematic review and meta-analysis.

I built on these conclusions by systematically varying SiN listening conditions, assessing the role of cognitive abilities selected based on well-established cognitive theory, and by testing younger and older adult listeners who vary in hearing sensitivity.

SiN perception tests were systematically varied to allow the role of cognitive abilities to be assessed depending upon different target/masker listening conditions. This design feature was chosen to test the hypothesis that the role of cognitive abilities may vary depending on the characteristics of SiN listening condition (Engle, 2002; Goldinger, 1996; Heinrich et al., 2015; Janse, 2012; Meister, Schreitmuller, Grugel, Ortmann, et al., 2013; Rönnberg et al., 2013; Rönnberg et al., 2008). The specific target and masker conditions were selected from the same target/masker matrix used in the systematic review chapter. The speech targets were chosen to vary in linguistic complexity, ranging from single words, to low and high predictability sentences. The background maskers were chosen to vary in the degree of informational properties between speech-modulated noise (less) and 3-talker babble (more).

Cognitive tests were selected based on well-established cognitive models, the Baddelely model of Working Memory (Baddeley, 2000) and the Diamond model of Executive Functions (Diamond, 2013). These models were deliberately chosen because 1) they include cognitive processes theorised to be important for SiN perception, 2) they are well established in the cognitive literature, and 3) are intuitive to understand and easy to implement. Multiple cognitive tests were selected to assess each cognitive ability from which a single factor score could be derived. This procedure was chosen to tackle the conflation of test- and ability-specific variance (Surprenant et al., 2009). In extracting common variance for multiple surface tests, as opposed to using a single cognitive test score, a more reliable measure of a specific cognitive ability can be assessed.

Additionally, to assess the role of hearing threshold and age, two listener groups were selected, younger and older adult listeners. All the younger listeners had normal hearing; the older listeners ranged in hearing sensitivity from normal to mild HL, as assessed and reported using pure-tone audiometry following British Society of Audiology guidelines (BSA, 2011).

The results of the study showed several key findings concerning the role of cognitive abilities, age and hearing sensitivity for SiN listening:

1) Central Executive ability plays a role for SiN perception in the older, but not the younger listener group, and its role is moderated by HL
2) Phonological Loop ability plays a different role depending on age group and background masker
3) Supplementary analysis in the older listeners showed:
    - Episodic Buffer and Working Memory ability displayed the same pattern of results, which differed depending on HL and background masker type

The results are described below in further detail, in turn for each tested cognitive domain/ability.

Central Executive

Central Executive (attention) abilities were only associated with SiN listening in the older listener group and this was true of all tested listening conditions. Additionally, when focusing on HL for all listener groups combined, both good hearing and good Central Executive abilities were shown to be needed in combination for optimum SiN intelligibility, regardless of listening condition. These results showed that hearing sensitivity and Central Executive (attention) abilities are important for SiN perception in all listening conditions. Furthermore, this effect may be more detectable in older adults because of greater variance and declines in both hearing sensitivity and attention abilities in comparison to normal hearing younger listeners.

Phonological Loop

Phonological Loop abilities were shown to contribute differently to SiN intelligibility depending upon age group and masking condition. In the younger adult group in speech-modulated noise masking, Phonological Loop ability was not associated with SiN intelligibility. This result suggests that Phonological Loop ability does not play a role in this masking condition. Conversely, in 3-talker babble, better Phonological Loop ability predicted better SiN intelligibility, suggesting that processing background babble on a phonological level was useful. For the older group in speech-modulated noise, better Phonological Loop ability was associated with better SiN intelligibility. Given that there was no phonological information in the background noise, this implies that Phonological Loop ability played a role in processing of the target signal. For 3-talker babble, better Phonological Loop processing was associated with poorer SiN intelligibility. This result suggests that the additional phonological information available in the background led to a disruption in processing for the target speech at the level of verbal storage.

Supplementary analysis in the younger listener group revealed a difference in association between Phonological Loop ability and SiN intelligibility depending on target type, namely a greater association with single words compared to sentences. The older adult group did not show this differential effect. This result suggested that in younger adult group short-term verbal storage and/or rehearsal of information is more important for intelligibility of less linguistically complex targets. Speculatively, this may indicate that less complex signals, such as single words, are processed more on a phonological level. In contrast more complex signals, such as sentences, may be processed on a semantical level, which involves the recruitment of other cognitive resources such as long-term memory (Craik et al., 1975).

Episodic Buffer

Supplementary analysis in the older adult group showed that the role of Episodic Buffer ability varied depending on hearing sensitivity and SiN background condition. In the speech-modulated noise masking condition, SiN intelligibility was more dependent on hearing sensitivity than Episodic Buffer ability. Listeners with better hearing abilities generally had higher SiN intelligibility, with Episodic Buffer ability contributing very little. However, in listeners with poorer hearing abilities, Episodic Buffer ability made a modest contribution to SiN intelligibility. In the 3-talker babble condition, Episodic Buffer and hearing abilities showed a similar contribution. These results show that the role of Episodic Buffer ability is greater in masking with more informationally properties (3-talker babble > speech-modulated noise), whereas and hearing sensitivity is important in both background masking situations.

Working Memory

The supplementary analysis in the older adults showed that the role of Working Memory ability result closely resemble the results for Episodic Buffer ability for the different masking conditions, i.e., cognitive ability plays a greater, or at least a varying, role in informational but not energetic masking. This commonality could be driven by the fact that both Working Memory and Episodic Buffer tasks have a storage (often verbal) component. A Working Memory task differs from an Episodic Buffer task by also containing a manipulation component. Given the similarity in results between Working Memory and Episodic Buffer ability it is possible that the observed effect in Working Memory ability is driven by the storage component.

Although previous studies have compared associations between Working Memory ability and SiN perception in different masking conditions (Parbery-Clark et al., 2009; Rönnberg et al., 2014), this study is the first to find a three-way interaction between Working Memory ability, hearing sensitivity, and SiN masker.

In summary, these results show across the tested cognitive abilities that younger and older listeners employ different listening strategies and that these strategies can vary depending on listening condition and hearing sensitivity. Within both listener groups there also appear to be differences in the role of specific cognition abilities depending on listening condition, specifically with regards to background masker.

5.1.3 Chapter four: Phonological Loop, not Central Executive ability, is important for SiN perception for younger adults

In the association experiment I showed a number of ways in which differences in various cognitive abilities predict differences in a number of different SiN listening conditions. If these predictive differences are meaningful then it should be possible to manipulate these cognitive abilities and see an effect on SiN perception. This final experiment attempted to test this hypothesis in a group of normal hearing adult listeners.

As per dual-task study design, a cognitive task was presented concurrently with a SiN perception test, not separately as in the association study. A further difference to the association study was that it was not possible to select multiple tests to measure an individual cognitive ability by deriving its shared variance. Instead only one cognitive test could be presented concurrently with the listening task. So, instead of isolating cognitive ability via multiple tests I characterised all cognitive abilities contributing to a particular test and chose the cognitive tests in such a way that cognitive contributions per test varied in a systematic way.

As in the previous chapter, I based test selection on the Baddelely model of Working Memory (Baddeley, 2000). Four cognitive tasks were designed to engage the sub-components of Working Memory to different extents. 1) Corsi forward task engaged the Visuo-Spatial Sketchpad and the Episodic Buffer (non-verbal storage only), 2) Corsi backward span task engaged the Visuo-Spatial Sketchpad, Episodic Buffer and Central Executive (non-verbal storage and manipulation), 3) digit forward span engaged the Phonological Loop and the Episodic Buffer (verbal storage) and 4) digit backward span engaged the Phonological Loop, Episodic Buffer and Central Executive (verbal storage and manipulation). Using this experimental design, I was able to assess proportional Dual-Task Costs (pDTCs) separately and combined (three sets of analyses in total) for performance in the cognition and SiN perception test.

The results of this study revealed that:

1) Verbal cognitive (digit span) tasks cause disruption (pDTC) in both SiN perception and cognitive task performance

2) The effect of verbal task disruption was driven by the engagement of Phonological Loop ability for both SiN perception and verbal cognitive tasks

3) There was limited or no contribution (pDTC) of Central Executive and Episodic Buffer abilities for SiN perception in younger adults with normal hearing

This provides insight as to why previous studies have not found significant associations between performance Working Memory tasks (which have a strong Central Executive component) and SiN perception for younger normal hearing listeners.

I also demonstrated that presenting verbal information in a visual format is sufficient in showing this effect.

Furthermore, by analysing pDTC performance separately for SiN intelligibility and cognitive performance I was able to assess similarities and differences between the two performances. For the digit span dual-task conditions, pDTCs were observed in both SiN and cognitive task performance. This shows that regardless of task priority, disruption is caused at the level of verbal storage (Phonological Loop) due to both tasks having a verbal component and the limited capacity of the Phonological Loop. Whereas, for the Corsi span dual-task conditions, pDTCs were only observed in cognitive, not SiN perception task performance. Furthermore, only limited pDTC were observed for the Corsi Span Backward condition, and no pDTC was shown for the Corsi Span Forward condition. This revealed two things, firstly, the lack of pDTC in the Corsi Span Forward condition showed that the is no role for the episodic buffer for SiN perception since that is the only shared sub-domain between the role tasks and no pDTC was observed for performance in either the SiN perception or the Corsi Span Forward tasks. Secondly, pDTC observed in cognitive performance in Corsi Span Backward condition showed there may be some contribution from Central Executive component, since there was no pDTC in Corsi Span Forward condition and the Corsi Span Backward condition only differed in having the additional of a manipulation (central executive) component. Additionally, the pDTC of the Phonological Loop component is greater than that of the Central Executive component, from this it can be inferred that SiN perception primarily engages the Phonological Loop (verbal storage), and the Central Executive (manipulation) to only a limited degree. Speculatively, the engagement of the Central Executive component may increase for listening with HL or for older listeners, who are more likely to experience perceptual or cognitive declines.

Overall this study has demonstrated that using a dual-task approach can be useful in investigating the role of Working Memory/cognition for SiN perception. Furthermore, it has provided further evidence for the specific roles of Working Memory sub-processes.

5.2 General limitations

In chapter two the categorisation of cognitive tests was based on multiple cognitive theories, whereas Chapters three and four focused specific on the Baddeley model of Working Memory. The wide range and variation of the cognitive tests used in the literature reviewed in chapter two did not allow for a single cognitive theory to be applied to individually assess all theorised cognitive abilities thought to be important for SiN perception. It is desirable to select a single model, which captures all the relevant cognitive processes involved for SiN perception in order for each process to be investigated in the context of that model. Working Memory is thought to be a key process involved in SiN perception and is central to models such as the ELU (Rönnberg et al., 2013; Rönnberg et al., 2008). Therefore, specifically focusing on Working Memory was of key interest. However, Working Memory is not a unitary process and therefore it was also desirable to investigate the role of Working Memory sub-processes for SiN perception. The Baddelely model (Baddeley, 2000) was selected for this investigation because it combines and differentiates between executive and storage components. However, this approach led to limitations on the differentiation of executive processes, and hence a second model, the Diamond model of Executive Functions (Diamond, 2013), was also assessed. Despite this consideration the experiments in chapters three and four were not fully able to assess and differentiate between all attentional and executive (sub-)processes thought to be important for SiN perception. For example, Set-Shifting, divided attention, and Fluid IQ were all found to be important in the meta-analysis, yet could not be assessed in the association and dual-task studies in chapters three and four as such an exhaustive list of cognitive processes would not have been feasible.

This thesis also focused on investigating the role of cognitive abilities for SiN perception on a subset of listening tasks that varied in target and masker properties as it was not feasible to select SiN perception tests to cover the full continuum. Yet, the selection of tests were carefully considered. For example, with regards to target type, sentences and words were favoured over phonemes to preserve comparability real-world listening conditions. Additionally, SiN listening factors other than target/masker type such as spatial separation and visual speech information, which were outside the scope of this thesis to investigate, are thought to differentially engage cognition. I focused on SiN listening with co-located auditory only listening conditions because they are common in the SiN perception literature and the clinic alike. The wide range of target/masker combinations used across different studies and a lack of systematicity in the selection of target/masker combination was apparent in the systematic review and meta-analysis.

A further aspect of SiN perception I could have chosen to systematically vary would have been intelligibility level (SNR), which is thought to differently engage cognitive processes depending on listening condition. Moreover, key differences were found between studies three and four regarding the role of the Phonological Loop for SiN perception in different masker conditions. It is possible that it was not the masker conditions that were crucial but the fact that the intelligibility levels differed between the studies. In the association study the SNR ratios were fixed at -2dB in the 3-talker babble masker a -7dB, and mean intelligibility levels were 30 and 45 RAU respectively (for LP sentences). In the dual-task study SNRs were individualised for each participant to approximate 70% thresholds. The mean SNRs were 1.6dB and -2.0dB for 3-talker babble and speech modulated noise. The actual mean intelligibility levels were 79 and 70 RAU respectively for 3-talker babble and speech-modulated noise. These differences in intelligibility levels could have led to a different role of cognition given that SiN perception occurred at differing positions on the SiN intelligibility psychometric curve. In hindsight it would have been more appropriate to attempt to match intelligibility levels between the studies. However, this would not have been feasible for the LP sentences because if the SNR ratio was set to 30-45 RAU in the single-task conditions there would have been a possibility of floor effects for SiN intelligibility, particularly in the digit span dual-task conditions.

An obvious solution to this SNR mismatch would have been to select the HP sentences, which could have been more closely matched for intelligibility levels between the studies. However, LP sentences were favoured over HP sentences to minimise the benefit of contextual support and benefit of auditory glimpsing during listening in fluctuating masking conditions (Schoof et al., 2015) maximise cognitive load. If HP sentences were used instead of LP sentences in the dual-task study it is possible it could effect in the outcome in that listeners with higher Working Memory ability may able to better take advantage of context compared to listeners with lower Working Memory. Furthermore, the presence of audible interference (3-talker babble) may cause further disruption for lower Working Memory ability listeners.

Other differences between studies three and four lie in cognitive test selection and in methodological approach. In the association study Phonological Loop processing ability was assessed using a rhyme verification task, which has a smaller short-term storage component, but has more involvement with long-term storage via lexical access. The verbal storage task in the dual-task study, the forward digital span task, has a much greater storage component, but a lesser long-term storage component. Therefore, this may have accounted for the lack of masker differences between the studies. If this is true then it suggests that interference from informational masking arises from disruption of lexical access and not disruption at the level of short-term verbal storage. Nevertheless, the digit span tests were selected in the dual-task study to allow for a visuo-spatial equivalents, the Corsi spans, to be selected. Therefore, the Phonological Loop ability domain in the association study may relate more to rehearsal and the Phonological Loop ability domain in the dual-task study more relate more to storage.

5.3 General conclusions

The method of categorisation of SiN perception tests was preserved throughout the thesis making results for specific listening conditions as comparable as possible. Furthermore, although the systematic review and meta-analysis in chapter two did not explicitly use cognitive abilities categorised on the basis of the Baddeley model of Working Memory, the results are still comparable due to overlaps in the listener demographics and cognitive ability categorisations.

One advantage in the categorisation strategy used in the systematic review and meta-analysis in chapter two was the differentiation between attentional (Alerting, Orienting) and executive processes (Set-Shifting, Inhibitory Control, Working Memory), all of which all fall under the Central Executive sub-domain of the Baddeley model.

The results from chapter two highlighted a role for Working Memory in SiN perception, with similar degrees of associations regardless of listening condition and hearing sensitivity. I explored this role systematically and with more specificity in chapters three and four by investigating the roles of the sub-processes of Working Memory for SiN perception in different listening conditions. In chapter three I also examined these roles for two different age groups, younger and older adults. This adds a higher level of specificity than was possible in chapter two, which did not examine age-related differences.

Figure 5. 1 below displays a schematic summary of the result of chapters three and four. The figure shows the associations found between SiN intelligibility and cognition (sub-domains for the Baddeley model of Working Memory), SiN intelligibility and PTA. This is represented separately for listener group (younger adults, older adults) and SiN background masker (speech-modulated noise, 3-talker babble) in a 2x2 matrix. The results have been summarised in this way to demonstrate that the role of cognitive abilities and their relation to age group vary depending on SiN listening condition.

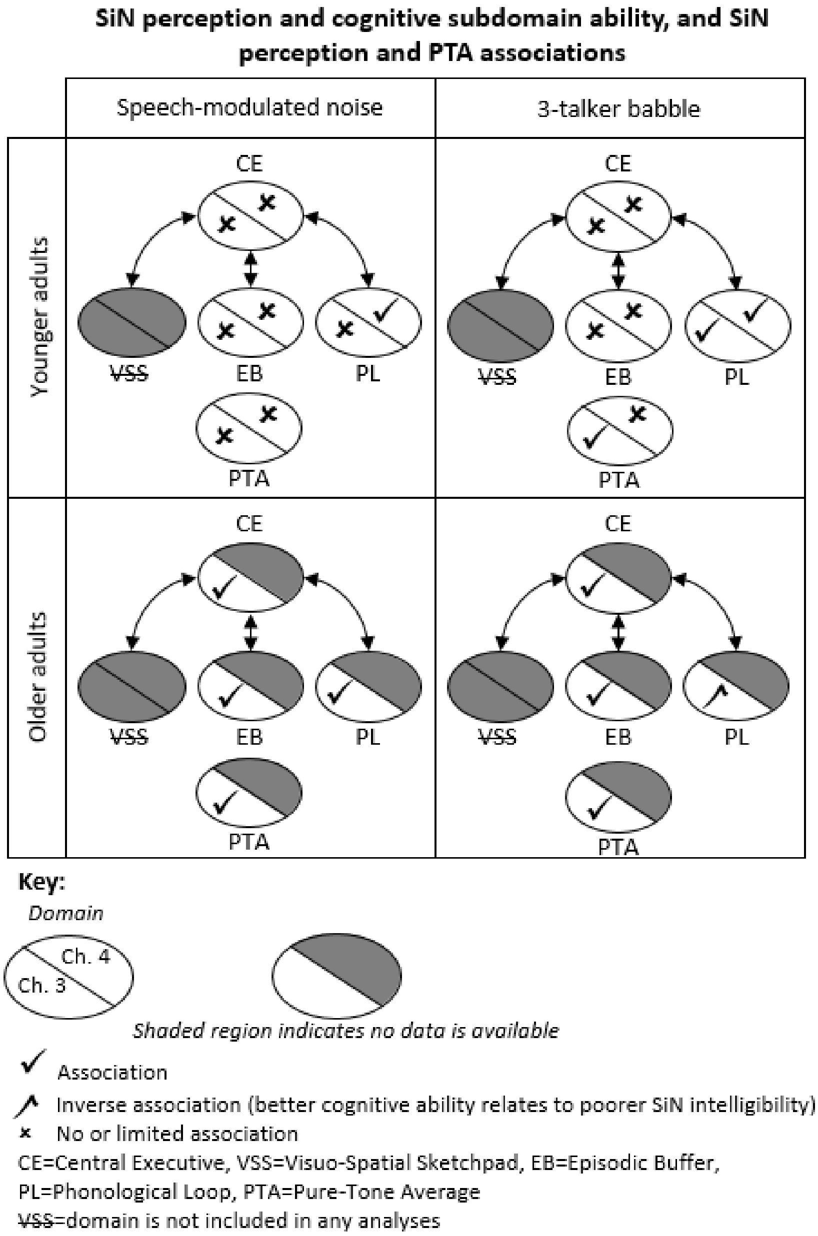


Figure 5. 1 – Summary of results of chapters three and four

Summary of cognitive sub-domain and SiN perception, and PTA and SiN perception associations, separated by listener group (young and older adults), and SiN masker condition (speech-modulated noise, 3-talker babble). Note intelligibility levels were not matched between Chapters 3 and 4 (intelligibility levels were lower in chapter 3

5.3.1 Younger listeners

For the younger listeners, both studies give evidence that Phonological Loop ability appeared to be the most important cognitive ability for SiN perception, with only a limited involvement of Central Executive and Episodic Buffer abilities. This thesis provides evidence that younger listeners can perceive speech in adverse conditions with minimal involvement of executive processes. This result suggests that SiN perception in younger listeners with normal hearing is relatively automatic and does not fully engage Working Memory. In terms of the ELU model (Rönnberg et al., 2013) it suggests that the signal representation was not too degraded for them as the model postulates that Working Memory processes (Executive Functions as opposed to short-term verbal memory/storage) only become engaged during perception of degraded signals. Hence, only in the presence of HL, SiN perception was difficult enough for Working Memory to be engaged. This may indicate that in these listening situations and this listener group a cognitive compensatory mechanism and/or a change in listening strategy occurred.

With regards to SiN listening condition, only the association study (Chapter three), but not the dual-task study (Chapter four), gave evidence to support masker (energetic versus informational) differences with regards to Phonological Loop ability. As previously discussed, this inconsistent result could be due to differences in the intelligibility levels of the SiN perception tests and/or differences in the selection of cognitive tests to assess Phonological Loop ability.

The results for the two studies suggest that for younger adults at a lower SNR level in 3-talker babble both hearing and cognition are important for SiN perception, whereas in speech-modulated noise cognition is more important. At more favourable SNRs neither is important in either masking conditions. To test this speculation, a study would need to be set up in which SNRs vary within a single study, note between two studies differing in methodologies. Such a study was conducted by Tun et al. (1999) where they investigated the contributions of PTA and Processing Speed for SiN perception of sentences in different masking conditions (inc. 2-talker babble (nearest equivalent to 3-talker babble) and 20-talker babble (nearest equivalent to speech-modulated noise) for younger and older adult listeners with normal hearing. At lower SNRs they found that in 2-talker babble both PTA and Processing Speed were predictive of SiN intelligibility, whereas in 20-talker babble only PTA was predictive of SiN intelligibility. At higher SNRs this relationship flipped, where in 2-talker babble only PTA was predictive of SiN intelligibility, and in

20-talker babble PTA and Processing Speed were predictive of SiN intelligibility. This suggests that the relative contributions of hearing and cognition can differ depending on not only masker properties but also depending on overall intelligibility as operationalised by different SNRs.

5.3.2 Older listeners

For older listeners, in contrast to the younger group, Central Executive and Episodic Buffer were associated with SiN intelligibility. One possible explanation of this observation is that in older persons, with poorer hearing sensitivity, successful perception of a degraded signal requires an increased or altered engagement of cognitive processes (such as manipulation and episodic storage of information) compared to younger persons with good hearing sensitivity. This finding is in agreement with the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013; Rönnberg et al., 2008) and a recent review (Füllgrabe et al., 2016b), which proposed that cognition/Working Memory is further engaged in the presence of hearing loss for listening in challenging conditions. Furthermore, the additional engagement of Central Executive processes may suggest the presence of a cognitive compensatory mechanism or difference in listening strategy, where those with higher Central Executive ability are best able to adapt and compensate when faced with perception of a degraded signal. Alternatively, because both better hearing sensitivity and Central Executive ability are associated with better SiN intelligibility, in circumstances of sensory degradation caused by hearing loss any level of Central Executive ability will be unable to compensate to restore a degraded signal in SiN perception.

The combined association between Central Executive ability and hearing sensitivity shows that cognition and perception are closely related and the uniform declines in both abilities may be underlined by an unknown common factor – the *common cause hypothesis* (CHABA, 1988; Lindenberger et al., 1994). However, here it is not possible to rule out either a decline in cognition impacting on perception (*cognitive load on perception hypothesis* (CHABA, 1988; Lindenberger et al., 1994)) or a decline perception driving a impacting on cognition (*information degradation hypothesis* (CHABA, 1988; Pichora-Fuller, 2003a; Schneider et al., 2000)). The parallel declines observed for both hearing sensitivity and Central Executive ability could outline the limits of cognitive compensation rather than the absence or impossibility of it. Therefore, this speaks to the possibility of *sensory* (Baltes et al., 1997; CHABA, 1988; Lindenberger et al., 1994) *and information degradation* hypotheses, where an impoverished signal leads to cognitive compensation and over a prolonged period this may lead to cognitive declines. If a decline in cognitive load was the driving factor it might be expected for all cognitive abilities to be effected and therefore performance in all perceptual tasks would be directly linked to task difficulty. However, here this may not be the case because the relationship between Central Executive ability and SiN intelligibility did not vary depending on listening condition. Although it should be noted all the SiN perception tests used here were comparable in difficulty.

Further age group differences were seen in Phonological Loop ability. Interestingly, in this group higher Phonological Loop abilities were associated with poorer SiN intelligibility. This highlighting two things, first that older listeners may employ additional or different cognitive resources to perceive the target signal; and second that this difference in cognitive engagement can be most detrimental when there are intelligible background distractors.

Additionally, based on the supplementary analysis in the older adults, Episodic Buffer ability appeared to play a greater role compared to hearing sensitivity in energetic masking. In contrast, in informational masking, Episodic Buffer ability and hearing sensitivity appeared to be equally important. This indicates that in the presence of intelligible distractors hearing sensitivity plays a greater role and can lower SiN perception regardless of cognitive ability. Therefore, it appears that younger and older listeners deploy different listening strategies and these strategies depend on SiN listening condition and hearing sensitivity.

These findings have possible implications for hearing intervention and highlight an explanation as to why hearing restored by a hearing aid cannot necessarily fully restore SiN perception, i.e., if there is cognitive decline and/or other contributing common factors. In this light, there is growing evidence and call for auditory and cognitive training to be administered alongside hearing intervention (Ferguson et al., 2015; Rudner, 2016; Tremblay et al., 2016; Yu et al., 2017). Furthermore, if sensory deprivation hypothesis accounts for even a proportion age-related declines this would support an argument for intervention as early as possible. Targeting intervention at an earlier stage would potentially reduce exposure to cognitive declines relating to prolonged cognitive compensation due to an impoverished sensory input. However, further work is needed in this area.

In summary, the overall evidence indicates a complex role of cognitive abilities for SiN perception performance, whereby performance differs in somewhat different way depending on age, hearing sensitivity and SiN listening condition. This thesis highlights the importance of systematically accounting for each aspect rather than drawing generalised conclusions regarding the role of cognition for SiN perception.

Appendix

Chapter two

Cognitive test by sub-domain	Description of test procedure
1) ATTENTION	
Alerting	
Integrated Visual and Auditory continuous performance test of attention (IVA+): Auditory Attention Quotient (AAQ) subtest (Sandford et al., 2004)	The numbers, 1 and 2, are presented aurally or visually. The task is to respond only presentations of '1' in either domain and to ignore presentation of the number '2'.
Test of Everyday Attention (TEA) task 7: Telephone search dual task (Robertson et al., 1994)	Visual search for key symbols in a telephone directory while simultaneously counting strings of tones presented by a tape recorder. The combined performance on sub-tests 6 and 7 gives a measure of divided attention - a 'dual task decrement'.
Orienting	
Test of Everyday Attention (TEA) task 6: Telephone search (Robertson et al., 1994)	Visual search for key symbols in a telephone directory page.
2) EXECUTIVE PROCESSES	
Set-Shifting	
TMT B: Trail making test - B (Reitan, 1958)	The task is to connect, using a pencil, encircled numbers and letters in alternating order.
TMT B/A: Trail making test B/A (Reitan, 1958)	Ratio between parts TMT B and TMT A (B: A), B-A difference score may also be used - ratio and difference score proposed to attempt to partial out contribution of Processing Speed and motor speed. TMT A is described under Processing Speed.
The connections test (Salthouse, 2000)	Similar to TMT B.
Inhibitory Control	
Auditory Hayling task (Burgess et al., 1996)	This test consists of two parts: firstly to give a verbal response to complete the final word of a sentence. The second part is to complete a sentence using a nonsense word, suppressing the predictable word.
Proactive interference (Kane et al., 2000)	Recall of lists of related or unrelated word lists (e.g. semantically or phonologically similar/dissimilar) after a distractor task (letter-number recall).
Simon task (Burle et al., 2005)	Green and red circles are presented visually on a screen. The coloured circles can appear on either side of the screen, however participants are given left and right arrow response keys which correspond to the colour of the circle and not the position on the screen.
Stroop test - original (Stroop, 1935), computerized (Jesse et al., 2012)	The test contains two parts: firstly, a neutral condition, a series of X's printed in different colour inks. Secondly, an incongruent condition, a list of colour names printed in a different colour ink to the word they represent, e.g., the word GREEN written in RED ink. The task in both conditions is to read out loud the colour of the ink, and in the incongruent task to try to discount the meaning of the word. Stroop interference is calculated by subtracting the reading time neutral condition from the reading time of the incongruent condition.
Visual distraction test (May, 1999)	Identification of a target word based on commonalities between three visually presented cue words. Cue words are presented with and without additional (leading or misleading) distractor words.
Working Memory	
Auditory Working Memory (Woodcock et al., 2001)	Lists of verbally presented words and numbers, the task is to reorder the sequence, citing the words and then the numbers.

Backward digit recall (Wechsler, 1997)	Recall of verbally presented numbers in reverse serial order.
Digit ordering (Cooper et al., 1991)	Recall of verbally presented digit sequences in ascending order.
Letter-number sequence (re-ordering) (Gold et al., 1997; Wechsler, 2008)	Recall of verbally presented letter and number strings in ascending/alphabetical order.
Letter memory test (Morris et al., 1990)	Lists of consonants presented one at a time. The task is to recall the previous four letters (beginning once four letters have been presented) in correct serial order. List lengths vary between 5, 7, 9 and 11 letters.
Listening span test (Daneman et al., 1980)	Sentence lists are presented aurally, the task is to decide if the final word was predictable. A letter is presented visually with each sentence and the participant must recall the letter sequence in correct serial order after each sentence block.
Numbers reversed (Woodcock et al., 2001)	See backward digit recall
Operation span (Unsworth et al., 2005)	Recall of lists of words and solution of simple math problems. First a math problem is displayed visually, e.g., Is (10+4)/2 = 8?, participants read the problem out loud then give their response with a button press 'yes' or 'no'. Next a word is displayed for a short amount of time to be read aloud. Then the process repeats giving word lists of increasing lengths, the task is to recall each word list in the correct serial order at the end of each trial.
Paced auditory serial addition test (Gronwall, 1977)	A random sequence of numbers (1-9) is presented aurally. The task is to add consecutive pairs of numbers such that each number is added to the number directing proceeding it. The response is prompted when the number list has been fully presented.
Reading span test (Andersson et al., 2001; Besser et al., 2013; Carroll et al., 2015; Daneman et al., 1980; Rönnberg et al., 1989)	The task is to read out loud lists of visually presented sentences and to remember the last word of each sentence for later recall. See (Conway et al., 2005) for review article on scoring methods.
Size comparison span (Sorqvist et al., 2010)	Size-comparison sentences (e.g. is x bigger than a y?) are presented visually. To which a 'yes' or 'no' response is given. After each sentence a to-be-remembered word (semantically similar to final word in sentence) is presented for later recall.
Visual letter monitoring (Gatehouse, 2003)	Identification of ten consonant-vowel-consonant words embedded within an 80-letter sequence displayed visually on a computer screen.
Colorado Assessment test: Visual Working Memory subtest (Davis et al., 1998)	A screen is displayed with eight boxes. The boxes light up individually to give a unique sequence, the task is to repeat back the sequence. The test has both forward and reverse conditions. The test is similar to the electronic Corsi block tapping test.
3) MEMORY	
Episodic Memory	
Cognitive Spare Capacity Examination (CCSE) word recall (Jacobs et al., 1977)	Delayed word recall of four item lists.
Forward digit recall (Wechsler, 1997)	Recall of verbally presented numbers in correct serial order.
Letter-number sequence (serial recall) (Gold et al., 1997)	Recall of verbally presented sequences of letters and numbers in correct serial order.
Memory for words (Woodcock et al., 2001) Word list recall (Cervera et al., 2009; Schuchardt et al., 2006)	Recall of verbally presented unrelated word lists in correct serial order (similar to forward digit recall).
Verbal learning and memory test (Helmstaedter et al., 1990) Word list memory (Morris et al., 1989)	Recall of as many words as possible from a verbally presented list. On subsequent trials the participant is reminded of word missed from the previous trial - the trails repeat until all words are recalled.
4) INTELLIGENCE	
Fluid IQ	

Matrix reasoning (Wechsler, 1999)	Selection of one image, from a choice of five, to complete a matrix displaying images with a logical pattern.
Crystalized IQ	
Lexical decision test (Carroll et al., 2016)	Decision task as to whether or not visually presented monosyllabic words are real meaningful words or pseudo-words.
Mill Hill vocabulary scale (Raven et al., 1982)	Identification of the correct synonym of a target word in a six alternative multiple choice format.
Nelson-Denny reading test (Brown et al., 1981)	Eight short passages are read followed by 36 multiple choice questions based on the eight passages, within 20 minutes.
Peabody vocabulary test (Bell et al., 2001)	Visual presentation of four pictures and simultaneous auditory presentation a target word. The task is to select the picture which best matches the target word.
Rhyme verification task (Johnston et al., 1986)	A pair of words is displayed visually, the task is to decide if the word pair rhyme or not.
Verbal ability (Stenbäck et al., 2015)	Lists of five words are visually displayed. The task is to select two words in each list which are antonyms.
Vocabulary test (Wechsler, 1981)	Words are visually presented and the task is to give a definition of each word in turn.
Word vocabulary test (Snijders et al., 1983)	Similar to Mill Hill vocabulary scale, but in Dutch language and within a five alternative multiple choice format.
Wortschatztest (Schmidt et al., 1992)	Visual presentation of rows of words, each containing six words: five of which are pseudo-words and one an existing word. The task is to select the one existing word in each row.
5) PROCESSING SPEED	
Processing Speed	
Digit symbol substitution test (Wechsler, 1981, 1997)	The task is to copy symbols that are paired with geometric shapes or numbers in a set sequence.
Letter digit substitution test (Jolles et al., 1995; Wechsler, 1997)	Similar to the digit symbol substitution test.
Trail making test –A (TMT A) (Reitan, 1958)	The connection of, using a pencil, encircled numbers in numerical order displayed on a sheet of paper.

Supplementary Figure 2. 1 – Cognitive test descriptions

Description of all cognitive tests used by the reviewed studies, categorized into cognitive domains and sub-domains

2.2a	**Score: (✖, ?, ✓ or N/A)**
Risk of Bias	
Did the authors include a sample size justification?	
If any participant data are excluded from the analysis is a clear justification given?	
Were all the outcome measures in the methods included in the results?	
Were there any conflicts of interest? I.e., is the study funded or conducted by a body with vested interests in the results?	

2.2b

✖ = High risk of bias (not enough information to make a judgement (Q1-3) or clear conflict of interest (Q4))
? = Unclear (incomplete information or not reported)
✓ = Low risk of bias (appropriate use and sufficient information (Q1-3) or no conflict of interest (Q4))
N/A = Not applicable (no participant data are excluded (Q2))

Supplementary Figure 2. 2 – Risk of bias assessment

Checklist for risk of bias assessment. 2a: Questions assessing risk of bias; 2b: score key for all questions

Study Demographics	Speech-in-noise Tests	Cognitive Tests	Sub-domain	Domain
(Anderson et al., 2013) N=120, 55-79 years, normal hearing-to-moderate hearing loss (HL)	Words in >2-talker babble (non-adaptive) Sentences in >2-talker babble (non-adaptive) Sentences in unmodulated noise (adaptive)	Auditory attention quotient of IVA+ (Sandford et al., 2004)	Alerting	Attention
		Memory for words (Woodcock et al., 2001)	Episodic Memory	Memory
		Auditory Working Memory (Woodcock et al., 2001)	Working Memory	Executive Processes
(Besser et al., 2012) N=55, 18-78 years, normal hearing-to-mild HL	Sentences in unmodulated noise (adaptive) Sentences in modulated noise (adaptive)	Reading span test (Andersson et al., 2001)	Working Memory	Executive Processes
		Letter digit substitution test (Jolles et al., 1995)	Processing Speed	Processing Speed
(Carroll et al., 2016) N=22, 18-35 years, normal hearing-to-mild HL	Sentences in unmodulated noise (non-adaptive)	Reading span test (Carroll et al., 2015)	Working Memory	Executive Processes
		The Lexical decision test (Carroll et al., 2016)	Crystalized IQ	Intelligence
		Wortschatztest (Schmidt et al., 1992)	Crystalized IQ	Intelligence
		Peabody vocabulary test (Bell et al., 2001)	Crystalized IQ	Intelligence
(Cervera et al., 2009) N=28, 19-25 & 55-65 years, normal hearing-to-moderate HL	Consonants in unmodulated noise (non-adaptive)	Word list recall (Cervera et al., 2009)	Episodic Memory	Memory
		Digit ordering (Cooper et al., 1991)	Working Memory	Executive Processes
(Ellis et al., 2014) N=24, age 24 (mean) years (age range not available), normal hearing-to-mild HL	Sentences in ≤2-talker babble (non-adaptive)	Proactive interference (Kane et al., 2000)	Inhibitory Control	Executive Processes
(Gordon-Salant et al., 2015) N=24, 18-26 & 65-80 years, normal hearing-to-moderate HL	Words in >2-talker babble (non-adaptive)	Digit symbol substitution test (Wechsler, 1997)	Processing Speed	Processing Speed
(Gordon-Salant et al., 2016) n=53, 18-25 & 61-75 years, normal hearing-to-mild HL	Words in >2-talker babble (adaptive) Sentences in >2-talker babble (adaptive)	Listening span test (Daneman et al., 1980)	Working Memory	Executive Processes
		Paced auditory serial addition test (Rao et al., 1989)	Working Memory	Executive Processes
		Reading span test (Rönnberg et al., 1989)	Working Memory	Executive Processes
		Letter digit substitution test (Wechsler, 1997)	Processing Speed	Processing Speed
(Heinrich et al., 2015)	Sentences in modulated noise (non-adaptive)	Test of Everyday Attention subtest 6	Orienting	Attention

N=44, 50-74 years, normal hearing-to-moderate HL		(Robertson et al., 1994)		
		Test of Everyday Attention subtest 7 (Robertson et al., 1994)	Alerting	Attention
		Backward digit recall (Wechsler, 1997)	Working Memory	Executive Processes
		Visual letter monitoring (Gatehouse, 2003)	Working Memory	Executive Processes
		Reading span test (Daneman et al., 1980)	Working Memory	Executive Processes
		Forward digit recall (Wechsler, 1997)	Episodic Memory	Memory
		Matrix reasoning (Wechsler, 1999)	Fluid IQ	Intelligence
(Heinrich & Knight, 2016) N=30, 62-85 years, normal hearing-to-moderate HL	Words in modulated noise (non-adaptive) Sentences in modulated noise (non-adaptive)	Stroop test (Stroop, 1935)	Inhibitory Control	Executive Processes
		Letter number sequencing (Wechsler, 1997)	Working Memory	Executive Processes
		Reading span test (Daneman et al., 1980)	Working Memory	Executive Processes
		Mill hill vocabulary scale (Raven et al., 1982)	Crystalized IQ	Intelligence
		Nelson-Denny reading test (Brown et al., 1981)	Crystalized IQ	Intelligence
(Helfer et al., 2014) N=30, 45-85 years, normal hearing-to-moderate HL	Sentences in ≤2-talker babble (non-adaptive) Sentences in unmodulated noise (non-adaptive)	Letter-number sequence serial recall (Gold et al., 1997)	Episodic Memory	Memory
		Letter-number sequence re-ordering (Gold et al., 1997)	Working Memory	Executive Processes
		Connections test (Salthouse, 2000)	Set-Shifting	Executive Processes
		Stroop test (Jesse et al., 2012)	Inhibitory Control	Executive Processes
(Janse, 2012) N=39, 65-83 years, normal hearing-to-moderate HL	Phonemes in ≤2-talker babble (non-adaptive)	Stroop test (Stroop, 1935)	Inhibitory Control	Executive Processes
(Koelewijn et al., 2014) n=32, 31-76 years, normal hearing-to-moderate HL	Sentences in unmodulated noise (adaptive) Sentences in ≤2-talker babble (adaptive)	Stroop test (Stroop, 1935)	Inhibitory Control	Executive Processes
		Listening span test (Daneman et al., 1980)	Working Memory	Executive Processes
		Size comparison span (Sorqvist et al., 2010)	Working Memory	Executive Processes
(Meister, Schreitmuller, Grugel, Beutner, et al., 2013) N=12, 58-79 years, normal hearing-to-mild HL	Sentences in unmodulated noise (adaptive) Sentences in modulated noise (adaptive)	Verbal learning and memory test (Helmstaedter et al., 1990)	Episodic Memory	Memory

	Sentences in >2-talker babble (adaptive)			
(Meister, Schreitmuller, Grugel, Ortmann, et al., 2013) N=26, 18-27 & 58-79 years, normal hearing-to-mild HL	Sentences in ≤2-talker babble (non-adaptive)	Verbal learning and memory test (Helmstaedter et al., 1990)	Episodic Memory	Memory
(Parbery-Clark et al., 2009) N=31, mean age 23±3 SD years (age range not available), normal hearing-to-mild HL	Sentence in unmodulated noise (adaptive) Sentence in >2-talker babble (non-adaptive)	composite score of Auditory Working Memory and Numbers reversed (Woodcock et al., 2001)	Working Memory	Executive Processes
(Parbery-Clark et al., 2011) N=37, age 45-65 years, normal hearing-to-mild HL	Words in >2-talker babble (non-adaptive) Sentence in unmodulated noise (adaptive) Sentence in >2-talker babble (non-adaptive)	composite score of Auditory Working Memory and Numbers reversed (Woodcock et al., 2001)	Working Memory	Executive Processes
		Colorado Assessment test: Visual Working Memory subtest (Davis et al., 1998)	Working Memory	Executive Processes
(Rönnberg et al., 2014) N=20, 28-42 years, normal hearing-to-mild HL	Sentences in unmodulated (non-adaptive) Sentences in modulated noise (non-adaptive) Sentences in >2-talker babble (non-adaptive)	Reading span test (Rönnberg et al., 1989)	Working Memory	Executive Processes
		Letter memory test (Morris et al., 1990)	Working Memory	Executive Processes
(Slater et al., 2016) N=54, 18-35 years, normal hearing-to-mild HL	Sentences in >2-talker babble (non-adaptive)	Auditory Working Memory (Woodcock et al., 2001)	Working Memory	Executive Processes
(Stenbäck et al., 2015) N=36, 18-22 & 61-79 years, normal hearing-to-mild HL	Sentences in modulated noise (adaptive)	Auditory Hayling task (Burgess et al., 1996)	Inhibitory Control	Executive Processes
		Reading span test (Rönnberg et al., 1989)	Working Memory	Executive Processes
(Surprenant, 2007) N=75, 30-80 years, normal hearing-to-mild HL	Syllables in unmodulated noise (non-adaptive)	Rhyme verification task (Johnston et al., 1986)	Crystalized IQ	Intelligence
		Operation span (Unsworth et al., 2005)	Working Memory	Executive Processes
(Tun et al., 1999) N=36, 18-22 & 61-79 years, normal hearing-to-moderate HL	Sentences in ≤2-talker babble (non-adaptive) Sentences in >2-talker babble (non-adaptive) Sentences in unmodulated noise (non-adaptive)	Vocabulary test (Wechsler, 1981)	Crystalized IQ	Intelligence
		Digit symbol substitution test (Wechsler, 1981)	Processing Speed	Processing Speed

(Uslar et al., 2013) N=20, mean age 24±2SD years (age range not available), normal hearing-to-mild HL	Sentences in unmodulated noise (adaptive)	word list recall (Schuchardt et al., 2006)	Episodic Memory	Memory
(Veneman et al., 2013) N=15, 20-28 & 66-78 years, normal hearing-to-mild HL	Sentences in >2-talker babble (non-adaptive)	Visual distraction test (May, 1999)	Inhibitory Control	Executive Processes
(Zekveld et al., 2011) N=76, 19-31 & 46-73 years, normal hearing-to-mild HL	Sentences in unmodulated noise (adaptive)	Letter digit substitution test (Jolles et al., 1995)	Processing Speed	Processing Speed
		Word vocabulary test (Snijders et al., 1983)	Crystalized IQ	Intelligence
(Zekveld et al., 2014) N=24, age 22±2.8 SD years (age range not available), normal hearing-to-mild HL	Sentences in ≤2-talker babble (adaptive)	Reading span test (Besser et al., 2013)	Working Memory	Executive Processes
		Size comparison span (Sorqvist et al., 2010)	Working Memory	Executive Processes
		Letter memory test (Morris et al., 1990)	Working Memory	Executive Processes
		Trail making test B-A difference (Reitan, 1958)	Set-Shifting	Executive Processes

Supplementary Figure 2. 3 – Summary of studies included in the meta-analysis

Summary of included studies including: participant demographics include number of participants, age range, and Hearing Loss (HL) categorization. The identity of speech tests was characterized in terms of target stimulus (Phonemes, words, sentences) and masker (≤2 talker babble, >2 babble, modulated noise, unmodulated noise). Cognitive tests lists cognitive tests used in each study.
SiN and cognitive tests are only included in this table if they were eligible for analysis under the criteria of this review. Allocation of tests to cognitive domains and sub-domains is described in the main text. SD=standard deviation.

Chapter three

QUESTIONNAIRE

Participant code	Date of Birth	Age	Gender
Years of Education in... School: Post-secondary:	Profession or subject of study (student):		

Have you ever lost consciousness? had a head injury? had a serious car accident?	 Yes Yes Yes	 No No No
Do you take any medication on a regular basis? Which?		
Do you suffer from a neurological disorder (e.g., MS, Parkinson’s, ALS, Stroke, etc..)?		

Are you colour blind?	Are you dyslexic?

Do you have hearing problems? Which ear?	Yes Left / Right / Both	No

Have you had a hearing test before?	Yes	No	Year
Have you ever had an ear infection ear discharge ear operation glue ear	 Yes Yes Yes Yes	 No No No No	 Year Year Year Year

Do you suffer from tinnitus? For episodes lasting longer than 5 minutes: Which ear? What does it sound like?	Yes No L/R both in head Ring / hum / whistle / buzz / hiss / other

Is English the language you use most often? If No, which language do you use most often?	Yes	No
Do you use other languages on a regular basis? Which?	Yes	No
Is English the first language you learned? If No, which language did you learn first?	Yes	No

Which other languages do you speak? To what extent?	Basic / fluent / native-like
Which of those languages did you learn as a child?	
How often do you use English (in percent) to communicate with Family? Friends? Work colleagues?	
How often do you use other languages (in percent) to communicate with Family? Friends? Work colleagues?	
Have you participated in hearing experiments before?	Yes No

Supplementary Figure 3. 1 – Medical questionnaire

Medical questionnaire

British English Sematic Sentence Test (BESST) stimulus list		
	Sentence pairs	
#	*High Predictability (HP)*	*Low Predictability (LP)*
1	Sam looked at the clock to check the time	Sam walked round the room to find the time
2	The teachers work at the school	The speakers went to the school
3	Daisy wore a helmet on her head	Daisy saw a feather on her head
4	Alice read a chapter of her book	Alice chose an actor for her book
5	Helena was staying in the guest room	Everyone was dressing in the best room
6	Tom rode his bike because he didn't have a car	Ann told a joke because he didn't have a car
7	Kate knew his face but not his name	Kate saw his card but not his name
8	I climbed the mountain to see the view	I promised my aunt she'd see the view
9	We'll never get there at this rate	He's always had it at this rate
10	The visitors knocked on the door	The shopkeepers looked at the door
11	The soldiers fought hard in the war	The farmers took part in the war
12	Ellen tried to speak but she had lost her voice	Ellen saw her face but didn't know her voice
13	Beth could afford it since they dropped the price	Beth was delighted since they changed the price
14	The car forced the cyclist off the road	The man forced the children off the road
15	The football captain encouraged his team	The dancing teacher encourage his team
16	James made a deposit at the bank	James went in a taxi to the bank
17	The actors performed the play on stage	It isn't allowed to walk on stage
18	The sailors were happy at the sight of land	The builders were angry at the price of land
19	The farmer was ploughing the field	The jockey was judging the field
20	The patient spent aged lying in bed	The teacher spent ages thinking in bed
21	Tim bought some grapes at the shop	Tim fought his friends at the shop
22	Ella went to the salon to dye her hair	Ella wanted the singer to change her hair
23	The apples are growing on the tree	The students are rowing by the tree
24	Chris went to the garden to water his plants	Chris went to the centre to order is plants
25	The homeless people live on the street	The youngest children sat on the street
26	Paul returned the shirt to get a bigger size	Jess revealed the plan to get the bigger size
27	Bill went to the movies to see a film	Bill wanted a reason to see a film
28	They wait anxiously to hear the news	They wanted desperately to hide the news
29	Every night the prisoners were locked in their cell	Every night the rabbits were left in their cell
30	The car broke down so they walked the last mile	The man left low so he skipped the last mile
31	In the house there is carpet on the floor	In the park there is blossom on the floor
32	Her trendy brother had a great sense of style	The working mother had a strong sense of style
33	Beth went to the stables to feed her horse	Beth though of the lady to keep her horse
34	Hannah played fetch with her dog	Hannah stayed in with her dog
35	Church windows are made from stained glass	Fresh linen is used on stained glass
36	The singer hit a very high note	The mother held a very long note
37	Robby was in danger so he shouted for help	Robby finished working so he waited for help
38	The captain sailed the ship across the sea	The chaplain saw the hill across the sea

39	Beth led the performance checked by the signs	Beth took the diversion directed by the signs
40	The captain scored the final goal	The doctor watched the final goal
41	Jack thought the bats would suck his blood	Jack thought the mark was not his blood
42	The popstar sold her story to the press	The doctor sent her letter to the press
43	At the end of the meal Joseph paid the bill	At the end of the day Nina left the bill
44	Salmon is one of the biggest types of fish	Simon is one of the keenest cooks of fish
45	He knew it was her birthday so he sent a card	He saw it was a problem so he left a card
46	The Earth and the planets go around the sun	The boys and Janet know about the sun
47	Tim went to rehab to stop using drugs	Tim wanted Peter to stop using drugs
48	The bones from the Turkey were boiled to make stock	The drones from the office were told to take stock
49	The night was lit by the moon and the stars	The wife was sick of the gloom of the stars
50	Sam was the lead singer of the band	Sam was the last runner in the band
51	The compass pointed due North	The actress found a new North
52	The British sprinter ran the race	The pretty singer ran the race
53	The best man was preparing jokes for his speech	The old man was expecting cash for his speech
54	Carl liked to have biscuits with a cup of tea	Carl had to read textbooks with a cup of tea
55	Zoe bought a new dress to wear to the ball	Zoe had a blue rag to wipe on the ball
56	Jess was under oath to tell the truth	Jess was on the way to tell the truth
57	The tree blew down in the wind	The man fell down in the wind
58	The blind man had lost his sight	The kind man had kept his sight
59	Hannah's dog ran away when left off the lead	Hannah's mum was upset when pushed of the lead
60	Stella milked a cow on her farm	Stella built a house on her farm
61	Jack tried to call her on the phone	Jack cried to call her to the phone
62	Alice chased the rabbit down the hole	Alice pushed the rabbit down the hole
63	Sam writes the lyrics for the songs	Sam checks the library for the songs
64	The children played on the swings in the park	The women stayed with the King in the park
65	The couple were married at the church	The people were carried to the church
66	The captain threw the anchor off the boat	The chaplain knew the answer was the boat
67	Once the paint had dried they added another coat	When they added the tin they added another coat
68	The ship's crew were stood on deck	The man's shoe was stuck on deck
69	Only cold water was running from the tap	Many odd noises were coming from the tap
70	The angry driver beeped her horn	The angry writer gripped her horn
71	The matador was stood in the field with the bull	The Labrador was stuck in the field with the bull
72	The cricket player hits the ball with the bat	The fitness trainer pick the boy with the bat
73	Tom threw the rubbish in the bin	Tom grow the flowers in the bin
74	The bird was collecting branches for a nest	The girl was arranging drawings of a nest
75	The race cars collided in a crash	The program concluded with a crash
76	The farmer sent Joe to milk the cow	The father went home to see the cow
77	The cat was chasing the mouse	The man was facing the mouse
78	The farmer sowed the seeds	The doctor chose the seeds
79	The dog has wagged its tail	The man had grabbed its tail

80	The bird flapped its wing	The girl touched its wing
81	George forgot to pay his landlord all the rent	George forgot to give his sister all the rent
82	The thunder was rumbling during the storm	The youngster was crying during the storm
83	Helen went to their pond to feed the ducks	Helen went to their land to feed the ducks
84	The actor read his lines from the script	The doctor set his mind on the script
85	Brenda tap stubbed her toe	Brenda had rubbed her toe
86	Bob can't post his letter without a stamp	Bob won't show his answer without a stamp
87	The gambler placed the bet	The angler faced the bet
88	The surfer was happy to catch a wave	The worker was happy to watch a wave
89	They could rob the safe because they knew the code	They should sale the sale because they knew the code
90	Laura would have got lost if she didn't have a map	Laura should have come last if she didn't have map
91	They noticed the fires when they smelt the smoke	They covered the tires when they smelt the smoke
92	They couldn't write the note without the pen	They wouldn't leave the house without a pen
93	Ella's brother had disappeared without a trace	Ella's mother had volunteered without a trace
94	The parents worried about their child	The lawyers worried about their child
95	The pirate wore a patch on his eye	The pilot saw a mark on his eye
96	The shop offered mass reductions on the sale	The man witnessed mass consumption in the sale

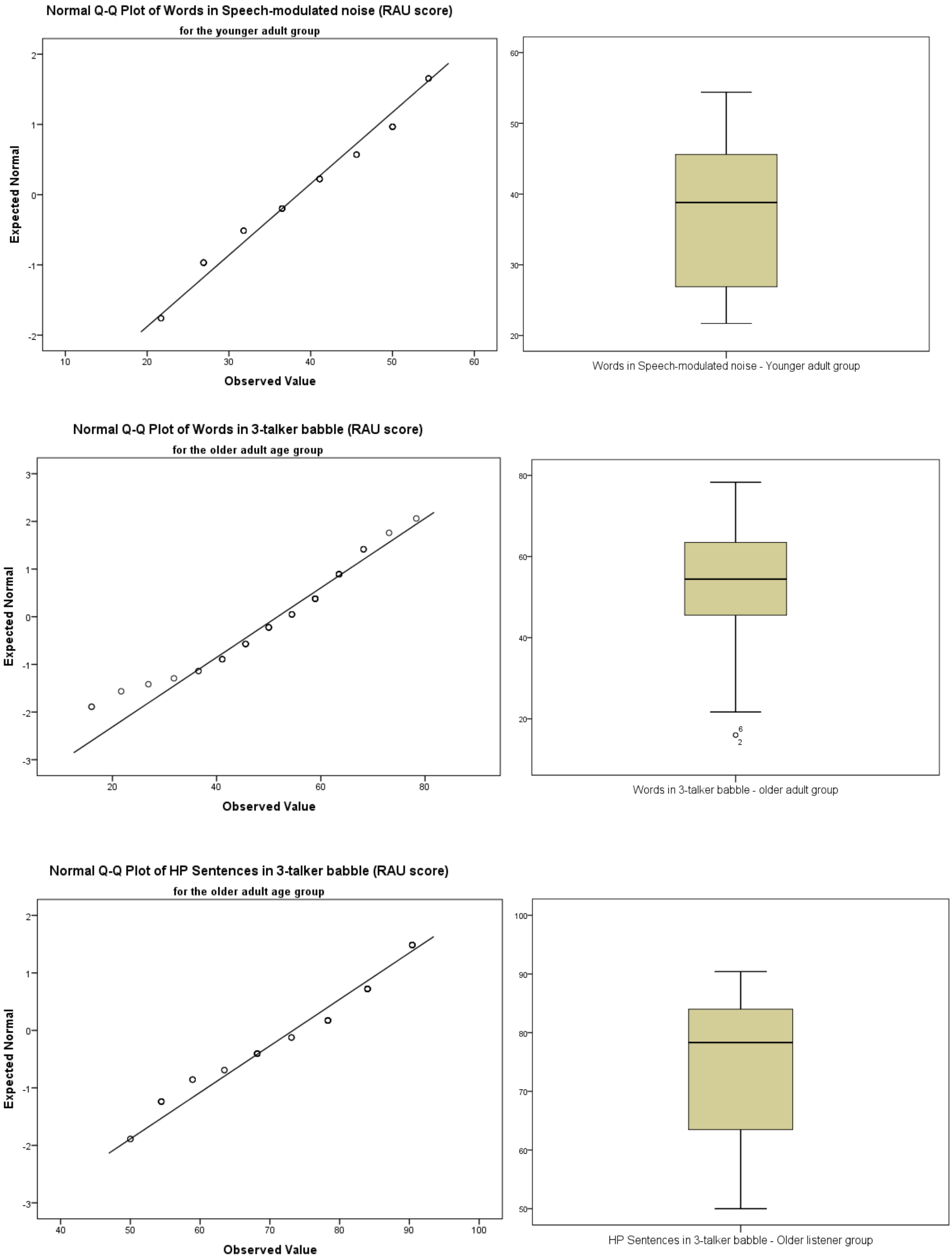
Supplementary Figure 3. 2 – Sentence stimulus list

British English Sematic Sentence Test (BESST) stimulus list. Only one sentence from each sentence pair was used in the association study in chapter three – greyed sentences were not used.

Word stimulus list			
#	Word	#	Word
1	act	29	law
2	ape	30	light
3	bit	31	list
4	board	32	loss
5	care	33	lot
6	change	34	man
7	chum	35	mirth
8	club	36	part
9	comb	37	plan
10	cork	38	pod
11	course	39	point
12	cup	40	pun
13	date	41	race
14	death	42	rest
15	face	43	risk
16	fact	44	role
17	foal	45	scheme
18	food	46	set
19	friend	47	sill
20	germ	48	snob
21	group	49	space
22	hand	50	tack
23	heart	51	tax
24	hog	52	thaw
25	house	53	thing
26	hump	54	top
27	lapse	55	wife
28	lark	56	work

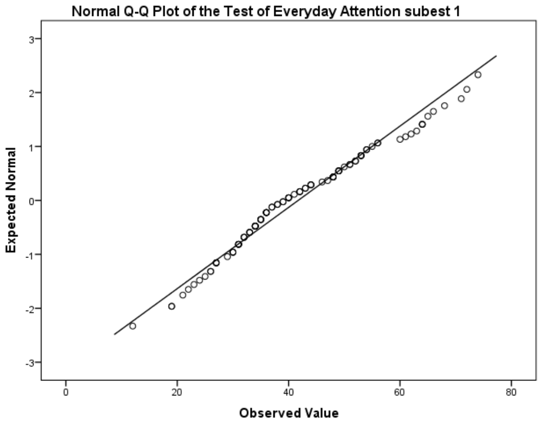
Supplementary Figure 3. 3 – Word stimulus list

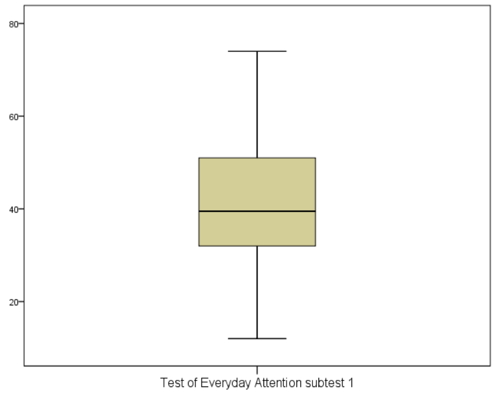
Monosyllabic word list

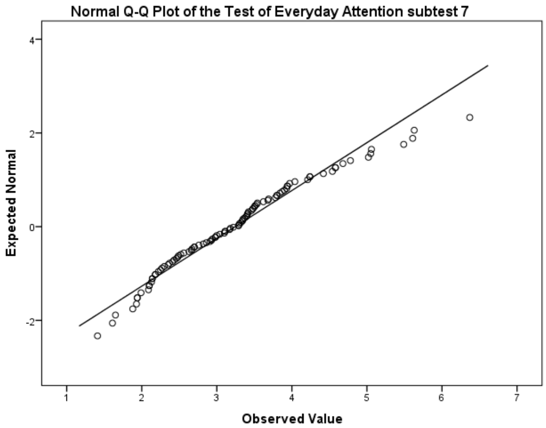


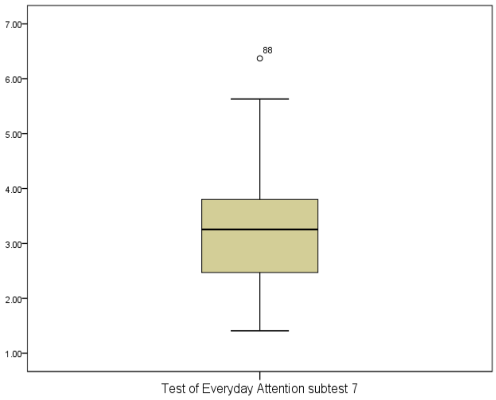
Supplementary Figure 3. 4 – Summary of normality assessment for SiN *perception* test variables

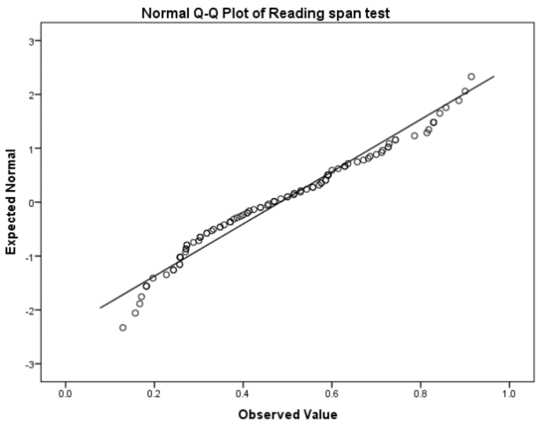
Q-Q and box plots for SiN perception tests which had a significant (<0.05) Shapiro-Wilk statistic – Words in Speech-modulated noise (younger adults) and HP sentences in 3-talker babble (older adults).

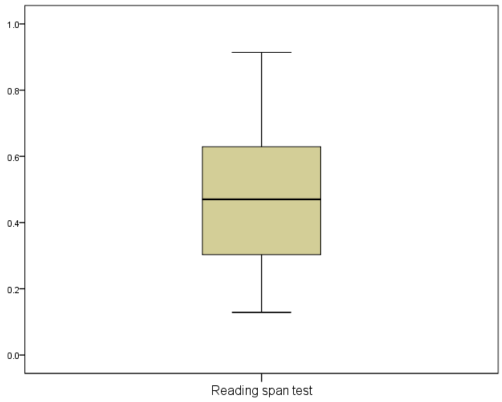

Normal Q-Q Plot of the Test of Everyday Attention subest 1
Expected Normal
Observed Value

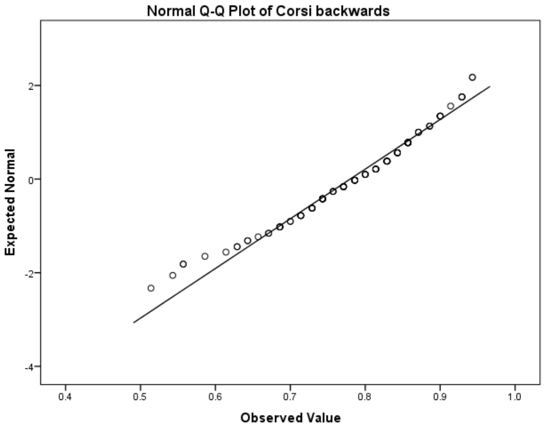

Test of Everyday Attention subtest 1

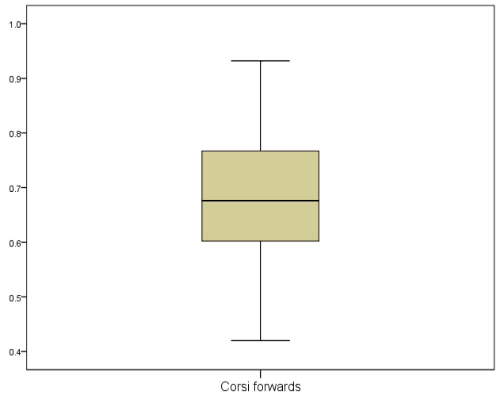

Normal Q-Q Plot of the Test of Everyday Attention subtest 7
Expected Normal
Observed Value

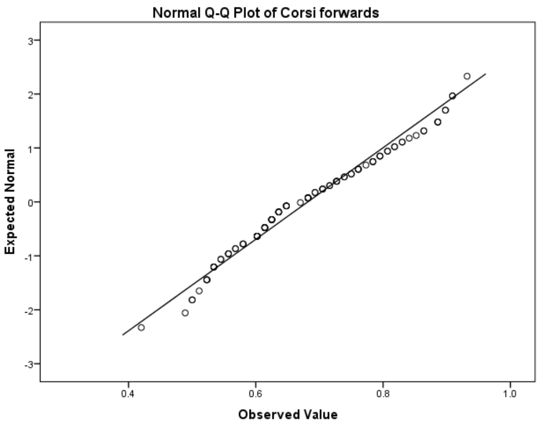

88
Test of Everyday Attention subtest 7

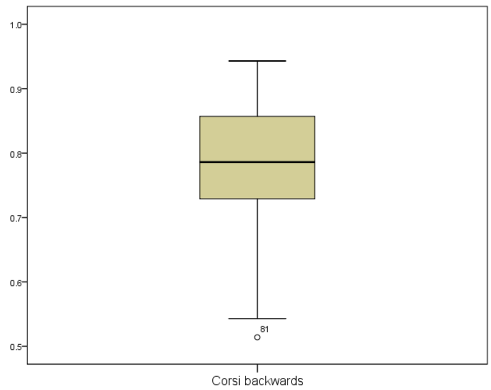

Normal Q-Q Plot of Reading span test
Expected Normal
Observed Value

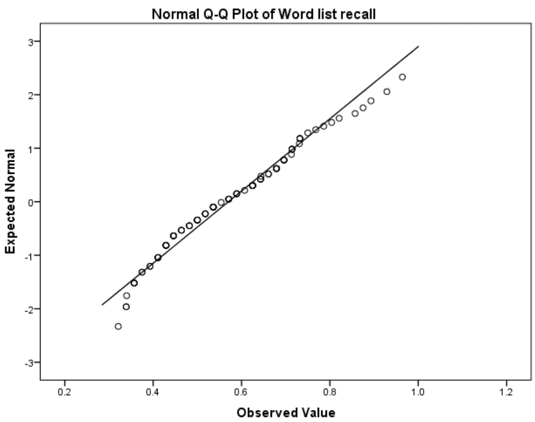

Reading span test

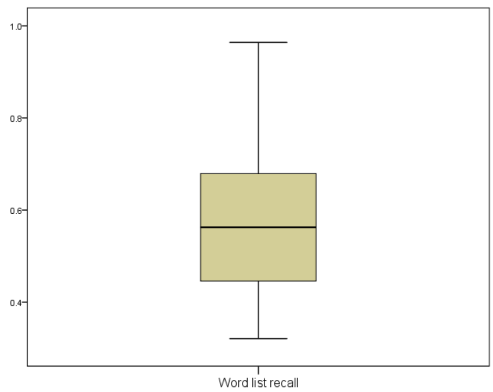

Normal Q-Q Plot of Corsi backwards
Expected Normal
Observed Value

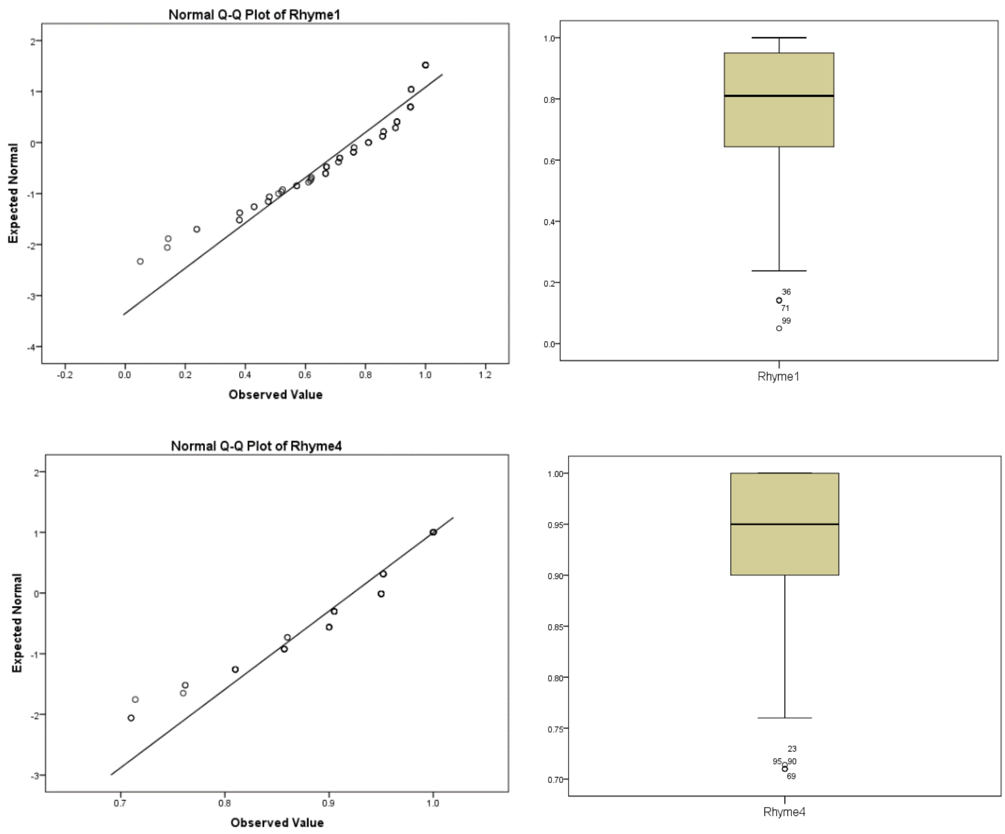

Corsi forwards


Normal Q-Q Plot of Corsi forwards
Expected Normal
Observed Value


81
Corsi backwards


Normal Q-Q Plot of Word list recall
Expected Normal
Observed Value


Word list recall



Supplementary Figure 3. 5 – Summary of normality assessment for cognitive test variables

Q-Q and box plots for cognitive tests which had a significant (<0.05) Shapiro-Wilk statistic – TEA1, TEA7, Reading span test, Corsi backwards, Corsi forwards, Word list recall, Rhyme1, and Rhyme4

Correlation coefficients (r-values) between cognitive test variables

	2	3	4	5	6	7	8	9	10	11	12	13
1	**-0.68**	**-0.65**	-0.08	**0.23**	0.14	**0.39**	**0.29**	**0.48**	**0.47**	**0.48**	-0.16	-0.14
2		**0.78**	0.19	-0.10	-0.17	**-0.31**	**-0.24**	**-0.49**	**-0.43**	**-0.35**	0.05	0.01
3			0.03	-0.01	**-0.20**	**-0.24**	-0.15	**-0.41**	**-0.34**	**-0.22**	0.08	0.03
4				**-0.24**	**-0.25**	**-0.28**	-0.18	-0.04	-0.15	-0.18	**-0.26**	-0.19
5					**0.41**	**0.48**	**0.47**	0.15	**0.27**	**0.63**	0.15	0.14
6						**0.46**	**0.40**	**0.22**	**0.25**	**0.43**	0.12	0.12
7							**0.56**	**0.28**	**0.48**	**0.52**	0.06	0.09
8								**0.23**	**0.23**	**0.45**	0.10	0.04
9									**0.56**	**0.37**	-0.10	-0.16
10										**0.55**	0.00	-0.11
11											0.01	0.02
12												**0.24**

Supplementary Figure 3. 6 – Correlation matrix of cognitive test variables

Correlation matrix (Pearson's correlation coefficients, two-tailed, not multiple comparison corrected) of cognitive test variables: 1=TEA1, 2=TEA6, 3=TEA7, 4=Stroop test, 5=Reading span test, 6=Letter-number sequencing, 7=Digit Span Backward, 8=Digit Span Forward, 9=Corsi Span Backward, 10=Corsi Span Forward, 11=Word list recall, 12=Rhyme 1, 13=Rhyme4. R-values highlighted in bold font are significant to the level of p<0.05

A) Full model for Baddeley sub-domain analysis – both age groups

SiN_Intell ~ Masker + Target + Age_group + PTA + CE + EB + PL + Masker*Target + Masker*Age_group + Masker*PTA + Masker*CE + Masker*EB + Masker*PL + Target*Age_group + Target*PTA + Target*CE + Target*EB + Target*PL + Age_group*PTA + Age_group*CE + Age_group*EB + Age_group*PL + PTA*CE + PTA*EB + PTA*PL + CE*EB + CE*PL + EB*PL + Masker*Target*Age_group + Masker*Target*PTA + Masker*Target*CE + Masker*Target*EB + Masker*Target*PL + Masker*Age_group*PTA + Masker*Age_group*CE + Masker*Age_group*EB + Masker*Age_group*PL + Masker*PTA*CE + Masker*PTA*EB + Masker*PTA*PL + Masker*CE*EB + Masker*CE*PL + Masker*EB*PL + Target*Age_group*PTA + Target*Age_group*CE + Target*Age_group*EB + Target*Age_group*PL + Target*PTA*CE + Target*PTA*EB + Target*PTA*PL + Target*CE*EB + Target*CE*PL + Target*EB*PL+ Age_group*PTA*CE + Age_group*PTA*EB + Age_group*PTA*PL + Age_group*CE*EB + Age_group*CE*PL + Age_group*EB*PL + PTA*CE*EB + PTA*CE*PL + PTA*EB*PL + CE*EB*PL + Age_group*Masker*Target*PTA + Age_group*Masker*Target*CE + Age_group*Masker*Target*EB + Age_group*Masker*Target*PL + Age_group*Masker*PTA*CE + Age_group*Masker*PTA*EB + Age_group*Masker*PTA*PL + Age_group*Masker*CE*EB + Age_group*Masker*CE*PL + Age_group*Masker*EB*PL + Age_group*Target*PTA*CE + Age_group*Target*PTA*EB + Age_group*Target*PTA*PL + Age_group*Target*CE*EB + Age_group*Target*CE*PL + Age_group*Target*EB*PL + Age_group*PTA*CE*EB + Age_group*PTA*CE*PL + Age_group*PTA*EB*PL + Age_group*CE*EB*PL + (1 + Masker + Target + Age_group | Participant)

B) Full model for Baddeley sub-domain analysis – age groups separately

SiN_Intell ~ Masker + Target + PTA + CE + EB + PL + Masker*Target + Masker*PTA + Masker*CE + Masker*EB + Masker*PL + Target*PTA + Target*CE + Target*EB + Target*PL + PTA*CE + PTA*EB + PTA*PL + CE*EB + CE*PL + EB*PL + Masker*Target*PTA + Masker*Target*CE + Masker*Target*EB + Masker*Target*PL + Masker*PTA*CE + Masker*PTA*EB + Masker*PTA*PL + Masker*CE*EB + Masker*CE*PL + Masker*EB*PL + Target*PTA*CE + Target*PTA*EB + Target*PTA*PL + Target*CE*EB + Target*CE*PL + Target*EB*PL + PTA*CE*EB + PTA*CE*PL + PTA*EB*PL + CE*EB*PL + (1 + Masker + Target| Participant)

C) Full model for Diamond sub-domain analysis – both age groups

SiN_Intell ~ Masker + Target + Age_group + PTA + IC + WM + Masker*Target + Masker*Age_group + Masker*PTA + Masker*IC + Masker*WM + Target*Age_group + Target*PTA + Target*IC + Target*WM + Age_group*PTA + Age_group*IC + Age_group*WM + PTA*IC + PTA*WM + IC*WM + Masker*Target*Age_group + Masker*Target*PTA + Masker*Target*IC + Masker*Target*WM + Masker*Age_group*PTA + Masker*Age_group*IC + Masker*Age_group*WM + Masker*PTA*IC + Masker*PTA*WM + Masker*IC*WM + Target*Age_group*PTA + Target*Age_group*IC + Target*Age_group*WM + Target*PTA*IC + Target*PTA*WM + Target*IC*WM + Age_group*PTA*IC + Age_group*PTA*WM + Age_group*IC*WM + PTA*IC*WM + Age_group*Masker*Target*PTA + Age_group*Masker*Target*IC + Age_group*Masker*Target*WM + Age_group*Masker*PTA*IC + Age_group*Masker*PTA*WM + Age_group*Masker*IC*WM + Age_group*Target*PTA*IC + Age_group*Target*PTA*WM + Age_group*Target*IC*WM + Age_group*PTA*IC*WM + (1 + Masker + Target + Age_group | Participant)

D) Full model for Diamond sub-domain analysis – age groups separately

SiN_Intell ~ Masker + Target + PTA + IC + WM + Masker*Target + Masker*PTA + Masker*IC + Masker*WM + Target*PTA + Target*IC + Target*WM + PTA*IC + PTA*WM + IC*WM + Masker*Target*PTA + Masker*Target*IC + Masker*Target*WM + Masker*PTA*IC + Masker*PTA*WM + Masker*IC*WM + Target*PTA*IC + Target*PTA*WM + Target*IC*WM + (1 + Masker + Target | Participant)

Supplementary Figure 3. 7 – Full mixed linear models for Baddeley and Diamond model analyses, for both age group combined (A: Baddelely and C: Diamond) and age groups separately (B: Baddely and D: Diamond).

	coefficient (β)	SE (β)	t	p
(Intercept)	53.9	2.49	21.68	<.001
Masker				
3-talker babble (Vs. speech-modulated noise)	10.0	3.02	3.30	.001
Target				
LP Sentences (Vs. HP sentences)	-24.7	1.98	-12.49	<.001
Words (Vs. HP sentences)	-16.1	2.11	-7.60	<.001
Age group				
Older adults (Vs. younger adults)	45.6	5.05	9.04	<.001
PTA	.1	.30	.45	.653
Central Executive	-0.6	1.27	-.47	.640
Episodic Buffer	2.1	.95	2.21	.029
Phonological Loop	.1	1.24	.07	.944
Masker*Target				
3-talker babble:LP sentences	8.5	2.20	3.89	<.001
3-talker babble:words	4.9	2.20	2.25	.025
Target*Age group				
LP sentences:Older adults	-3.4	2.33	-1.46	.147
Words:Older adults	-9.3	2.56	-3.66	<.001
Age group*Central Executive				
Older adults:Central Executive	-9.3	4.30	-2.16	.033
PTA*Central Executive	.4	.17	2.15	.034
Age group*Phonological Loop				
Older adults:Phonological Loop	1.7	1.76	.95	.346
Masker*Phonological Loop				
3-talker babble:Phonological Loop	2.5	2.01	1.25	.214
Age group*PTA				
Older adults:PTA	-1.4	.38	-3.64	<.001
Masker*PTA				
3-talker babble:PTA	-.8	.46	-1.71	.090
Masker*Age group				
3-talker babble:Older adults	-22.9	5.99	-3.82	<.001
Masker*Age group*PTA				
3-talker babble:Older adults:PTA	1.5	.51	3.00	.003
Masker*Age group*Phonological Loop				
3-talker babble:Older adults:Phonological Loop	-7.0	2.89	-2.43	.017

Random effects

Groups	Name	Variance	Standard deviation
Participant	*(intercept)*	97.0	9.8
	Target: LP sentences	15.4	3.9
	Target: Words	42.7	6.5
	Masker: 3-talker babble	123.9	11.1
	Residual	120.6	11.0

Supplementary Figure 3. 8 – Summary of fixed effects coefficients and random effects for the linear mixed model for the Baddeley model sub-domains for younger and older adults

Note: the data points in the models are SiN intelligibility levels (expressed in Rationalized Arcsine Units (RAUs)) for each participant in each SiN condition

Predictor	coefficient (β)	SE (β)	t	p
(Intercept)	55.7	2.81	19.83	<.001
Masker				
3-talker babble (Vs. speech-modulated noise)	5.0	2.69	1.87	.065
Target				
LP Sentences (Vs. HP sentences)	-25.2	2.62	-9.62	<.001
Words (Vs. HP sentences)	-16.6	2.74	-6.06	<.001
PTA	-.2	.38	-.56	.578
Central Executive	-.1	1.15	-.13	.901
Episodic Buffer	1.7	1.06	1.59	.117
Phonological Loop	-2.8	2.58	-1.10	.278
Masker*Target				
3-talker babble:LP sentences	10.0	2.79	3.58	<.001
3-talker babble:words	7.2	2.79	2.59	.010
Target*PTA				
LP Sentences:PTA	-.1	.37	-.34	.738
Words:PTA	-.2	.40	-.50	.622
Target*Phonological Loop				
LP Sentences:Phonological Loop	-1.4	2.48	-.55	.584
Words:Phonological Loop	5.2	2.68	1.94	.056
PTA*Phonological Loop	.9	.40	2.26	.028
Target*PTA*Phonological Loop				
LP sentences:PTA:Phonological Loop	.1	.38	.18	.856
Words:PTA:Phonological Loop	-1.2	.41	-2.85	.006

Random effects

Groups	Name	Variance	Standard deviation
Participant	*(intercept)*	130.1	11.4
	Target: LP sentences	37.1	6.1
	Target: Words	51.1	7.2
	Masker: 3-talker babble	166.4	12.9
	Residual	97.4	9.9

Supplementary Figure 3. 9 - Summary of fixed effects coefficients and random effects of the linear mixed model for the Baddeley model sub-domains for younger adults only

Note: the data points in the models are SiN intelligibility levels (expressed in Rationalized Arcsine Units (RAUs)) for each participant in each SiN condition

Predictor	coefficient (β)	SE (β)	t	P
(Intercept)	92.4	5.72	16.16	<.001
Masker				
3-talker babble (Vs. speech-modulated noise)	1.9	6.63	.28	.780
Target				
LP Sentences (Vs. HP sentences)	-23.9	1.86	-12.81	<.001
Words (Vs. HP sentences)	-22.9	1.86	-12.31	<.001
PTA	-1.0	.26	-3.75	<.001
Central Executive	-2.2	1.82	-1.19	.241
Episodic Buffer	.5	6.81	.079	.937
Phonological Loop	1.6	1.36	1.15	.252
Masker*Phonological Loop				
3-talker babble:Phonological Loop	-4.2	1.63	-2.57	.011
Masker*Episodic Buffer				
3-talker babble: Episodic Buffer	14.5	8.08	1.79	.074
Masker*PTA				
3-talker babble:PTA	.2	.31	.55	.582
PTA*Episodic Buffer	.1	.30	.24	.808
Masker*PTA*Episodic Buffer				
3-talker babble:PTA:Episodic Buffer	-.7	.36	-2.09	.038

Random effects			
Groups	Name	Variance	Standard deviation
Participant	*(intercept)*	22.8	4.8
	Residual	173.7	13.2

Supplementary Figure 3. 10 - Summary of fixed effects coefficients and random effects of the linear mixed model for the Baddeley model sub-domains for older adults only

Note: the data points in the models are SiN intelligibility levels (expressed in Rationalized Arcsine Units (RAUs)) for each participant in each SiN condition

Predictor	coefficient (β)	SE (β)	t	p
(Intercept)	55.1	2.44	22.60	<.001
Masker				
3-talker babble (Vs. speech-modulated noise)	9.5	3.07	3.10	.002
Target				
LP Sentences (Vs. HP sentences)	-24.7	1.99	-12.42	<.001
Words (Vs. HP sentences)	-16.1	2.12	-7.59	<.001
Age group				
Older adults (Vs. younger adults)	39.5	4.37	9.04	<.001
PTA	-.1	.29	-.41	.684
Inhibitory Control	-.6	1.02	-.60	.552
Working Memory	.0	0.75	.01	.994
Masker*Target				
3-talker babble:LP sentences	8.5	2.20	3.87	<.001
3-talker babble:words	4.9	2.20	2.24	.026
Target*Age group				
LP sentences:Older adults	-3.4	2.35	-1.45	.150
Words:Older adults	-9.3	2.56	-3.65	<.001
Age group*PTA				
Older adults:PTA	-1.0	.32	-2.97	.004
Masker*PTA				
3-talker babble:PTA	-.8	.47	-1.75	.084
Masker*Age group				
3-talker babble:Older adults	-24.6	6.06	-4.05	<.001
Masker*Age group*PTA				
3-talker babble:Older adults:PTA	1.6	.52	3.09	.003

Random effects

Groups	Name	Variance	Standard deviation
Participant	*(intercept)*	109.0	10.4
	Target: LP sentences	16.1	4.0
	Target: Words	42.0	6.5
	Masker: 3-talker babble	134.5	11.6
	Residual	121.5	11.0

Supplementary Figure 3. 11 - Summary of fixed effects coefficients and random of the linear mixed model for the Diamond model sub-domains for younger and older adults

Note: the data points in the models are SiN intelligibility levels (expressed in Rationalized Arcsine Units (RAUs)) for each participant in each SiN condition

Predictor		coefficient (β)	SE (β)	t	p
(Intercept)		57.3	2.54	22.51	<.001
Masker					
3-talker babble (Vs. speech-modulated noise)		5.0	2.71	1.85	.067
Target					
LP Sentences (Vs. HP sentences)		-25.5	2.17	-11.73	<.001
Words (Vs. HP sentences)		-17.2	2.34	-7.36	<.001
PTA		-.4	.21	-2.06	.044
Inhibitory Control		-.1	.87	-.10	.918
Working Memory		-.3	1.20	-.25	.804
Masker*Target					
3-talker babble:LP sentences		10.0	2.85	3.50	.001
3-talker babble:words		7.2	2.85	2.53	.012
Random effects					
Groups	Name	Variance	Standard deviation		
Participant	*(intercept)*	156.6	12.5		
	Target: LP sentences	32.0	5.7		
	Target: Words	70.5	8.4		
	Masker: 3-talker babble	163.2	12.8		
	Residual	101.7	10.1		

Supplementary Figure 3. 12 - Summary of fixed effects coefficients and random effects of the linear mixed model for the Diamond model sub-domains for younger adults only

Note: the data points in the models are SiN intelligibility levels (expressed in Rationalized Arcsine Units (RAUs)) for each participant in each SiN condition

Predictor	coefficient (β)	SE (β)	t	P
(Intercept)	93.2	3.89	23.98	<.001
Masker				
3-talker babble (Vs. speech-modulated noise)	-5.6	5.73	-.97	.337
Target				
LP Sentences (Vs. HP sentences)	-23.9	1.73	-13.81	<.001
Words (Vs. HP sentences)	-22.9	1.73	-13.28	<.001
PTA	-1.0	.17	-6.27	<.001
Inhibitory Control	.6	4.27	.15	.885
Working Memory	-2.1	1.76	-1.18	.245
PTA*Working Memory	.0	.18	.26	.794
Masker*Working Memory				
3-talker babble:Working Memory	8.4	6.57	1.28	.207
Masker*PTA				
3-talker babble:PTA	.5	.25	1.89	.065
Masker*PTA*Working Memory				
3-talker babble:PTA:Working Memory	-.6	.27	-2.01	.049

Random effects

Groups	Name	Variance	Standard deviation
Participant	*(intercept)*	28.5	5.3
	Masker: 3-talker babble	86.9	9.3
	Residual	149.2	12.2

Supplementary Figure 3. 13 - Summary of fixed effects coefficients and random effects of the linear mixed model for the Diamond model sub-domains for older adults only

Note: the data points in the models are SiN intelligibility levels (expressed in Rationalized Arcsine Units (RAUs)) for each participant in each SiN condition

Chapter four

British English Sematic Sentence Test (BESST) stimulus list	
#	*Low Predictability (LP)*
1	Sam walked round the room to find the time
2	The speakers went to the school
3	Daisy saw a feather on her head
4	Alice chose an actor for her book
5	Everyone was dressing in the best room
6	Ann told a joke because he didn't have a car
7	Kate saw his card but not his name
8	I promised my aunt she'd see the view
9	He's always had it at this rate
10	The shopkeepers looked at the door
11	The farmers took part in the war
12	Ellen saw her face but didn't know her voice
13	Beth was delighted since they changed the price
14	The man forced the children off the road
15	The dancing teacher encourage his team
16	James went in a taxi to the bank
17	It isn't allowed to walk on stage
18	The builders were angry at the price of land
19	The jockey was judging the field
20	The teacher spent ages thinking in bed
21	Tim fought his friends at the shop
22	Ella wanted the singer to change her hair
23	The students are rowing by the tree
24	Chris went to the centre to order is plants
25	The youngest children sat on the street
26	Jess revealed the plan to get the bigger size
27	Bill wanted a reason to see a film
28	They wanted desperately to hide the news
29	Every night the rabbits were left in their cell
30	The man left low so he skipped the last mile
31	In the park there is blossom on the floor
32	The working mother had a strong sense of style
33	Beth though of the lady to keep her horse
34	Hannah stayed in with her dog
35	Fresh linen is used on stained glass
36	The mother held a very long note

37	Robby finished working so he waited for help
38	The chaplain saw the hill across the sea
39	Beth took the diversion directed by the signs
40	The doctor watched the final goal
41	Jack thought the mark was not his blood
42	The doctor sent her letter to the press
43	At the end of the day Nina left the bill
44	Simon is one of the keenest cooks of fish
45	He saw it was a problem so he left a card
46	The boys and Janet know about the sun
47	The bucket helped Paula hold a lot of weight
48	The presenter was showing an interesting talk
49	Tim wanted Peter to stop using drugs
50	The drones from the office were told to take stock
51	The wife was sick of the gloom of the stars
52	Sam was the last runner in the band
53	The actress found a new North
54	The pretty singer ran the race
55	The old man was expecting cash for his speech
56	Carl had to read textbooks with a cup of tea
57	Zoe had a blue rag to wipe on the ball
58	Jess was on the way to tell the truth
59	The boss praised the gambler for his crime
60	We started at the signal for the train
61	The man fell down in the wind
62	The kind man had kept his sight
63	Hannah’s mum was upset when pushed of the lead
64	Stella built a house on her farm
65	Jack cried to call her to the phone
66	Alice pushed the rabbit down the hole
67	Sam checks the library for the songs
68	The women stayed with the King in the park
69	The people were carried to the church
70	The chaplain knew the answer was the boat
71	When they added the tin they added another coat
72	The man’s shoe was stuck on deck
73	Many odd noises were coming from the tap
74	The angry writer gripped her horn
75	The Labrador was stuck in the field with the bull
76	The fitness trainer pick the boy with the bat
77	Tom grow the flowers in the bin

78	The girl was arranging drawings of a nest
79	The program concluded with a crash
80	The father went home to see the cow
81	The man was facing the mouse
82	The doctor chose the seeds
83	The man had grabbed its tail
84	The girl touched its wing
85	George forgot to give his sister all the rent
86	The youngster was crying during the storm
87	Helen went to their land to feed the ducks
88	The doctor set his mind on the script
89	Brenda had rubbed her toe
90	Bob won't show his answer without a stamp
91	The angler faced the bet
92	The worker was happy to watch a wave
93	They should sale the sale because they knew the code
94	Laura should have come last if she didn't have map
95	They covered the tires when they smelt the smoke
96	They wouldn't leave the house without a pen
97	Ella's mother had volunteered without a trace
98	The lawyers worried about their child
99	The pilot saw a mark on his eye
100	The man witnessed mass consumption in the sale
101	Simon agreed his time for the test
102	Liam's only son gave them the match
103	I must have a pen because I need to make a call
104	Simon looked at the things in the fire
105	I'll leave you a massage in the post
106	Jayne judged the brothers as they prepared for a play
107	John had lied when he said he didn't have a will
108	Paul wanted to of stayed at a better wage
109	Sam's Dad was depressed since they wouldn't delay his loan
110	Ella knows her friend would never try to make her laugh
111	The singer was offended by the prize
112	Greg saw old books at his feet
113	He won't receive the score without a key

Supplementary Figure 4. 1 – Sentence stimulus list

British English Sematic Sentence Test (BESST) stimulus list. Only low predictability sentences were used in the dual-task study in chapter four.

References

Agus, T. R., Akeroyd, M. A., Gatehouse, S., & Warden, D. (2009). Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *J Acoust Soc Am, 126*(4), 1926-1940. doi: 10.1121/1.3205403

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol, 47 Suppl 2*, S53-71. doi: 10.1080/14992020802301142

Anderson, S., Parbery-Clark, A., Yi, H. G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear Hear, 32*(6), 750-757. doi: 10.1097/AUD.0b013e31822229d3

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear Res, 300*, 18-32. doi: 10.1016/j.heares.2013.03.006

Andersson, U., Lyxell, B., Rönnberg, J., & Spens, K. E. (2001). Cognitive correlates of visual speech understanding in hearing-impaired individuals. *J Deaf Stud Deaf Educ, 6*(2), 103-116. doi: 10.1093/deafed/6.2.103

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scand J Psychol, 50*(5), 371-384. doi: 10.1111/j.1467-9450.2009.00753.x

Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology, 80*(1), 23-34. doi: 10.1016/j.biopsycho.2008.01.015

Atkins, D., Best, D., Briss, P. A., Eccles, M., Falck-Ytter, Y., Flottorp, S., . . . Grp, G. W. (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal, 328*(7454), 1490-1494. doi: 10.1136/bmj.328.7454.1490

Bach, M. (2007). The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefes Arch Clin Exp Ophthalmol, 245*(7), 965-971.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn Sci, 4*(11), 417-423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory.* (Vol. Vol. VIII, pp. 47-90). New York: New York: Academic Press.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12*(1), 12-21. doi: Doi 10.1037/0882-7974.12.1.12

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

Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the Peabody Picture Vocabulary Test-Third Edition and Wechsler Adult Intelligence Scale-Third Edition with university students. *J Clin Psychol, 57*(3), 417-422.

Ben-David, B. M., & Schneider, B. A. (2009). A Sensory Origin for Color-Word Stroop Effects in Aging: A Meta-Analysis. *Aging Neuropsychology and Cognition, 16*(5), 505-534.

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise--a review. *Trends Amplif, 17*(2), 75-93. doi: 10.1177/1084713813495459

Besser, J., Zekveld, A. A., Kramer, S. E., Rönnberg, J., & Festen, J. M. (2012). New Measures of Masked Text Recognition in Relation to Speech-in-Noise Perception and Their Associations With Age and Cognitive Abilities.

Journal of Speech Language and Hearing Research, 55(1), 194-209. doi: 10.1044/1092-4388(2011/11-0008)
Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a Test of Speech-Perception in Noise. *Journal of Speech and Hearing Research, 27*(1), 32-48.
Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., . . . et al. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology, 35*(6), 662-672. doi: 10.1002/ana.410350606
Bockstael, A., Samyn, L., Corthals, P., & Botteldooren, D. (2018). Presenting and processing information in background noise: A combined speaker-listener perspective. *J Acoust Soc Am, 143*(1), 210. doi: 10.1121/1.5020799
Boothroyd, A. (1968). Statistical theory of the speech discrimination score. *J Acoust Soc Am, 43*(2), 362-367.
Brannstrom, K. J., Zunic, E., Borovac, A., & Ibertsson, T. (2012). Acceptance of Background Noise, Working Memory Capacity, and Auditory Evoked Potentials in Subjects with Normal Hearing. *Journal of the American Academy of Audiology, 23*(7), 542-552. doi: 10.3766/jaaa.23.7.6
Broadbent, D. (1958). *Perception and Communication*. Oxford, England: OUP.
Brown, J. I., Bennett, J. M., & Hanna, G. (1981). *The Nelson-Denny reading test*. Chicago, IL: Riverside.
Brunetti, R., Del Gatto, C., & Delogu, F. (2014). eCorsi: implementation and testing of the Corsi block-tapping task for digital tablets. *Front Psychol, 5*, 939. doi: 10.3389/fpsyg.2014.00939
BSA. (2011). Recommended Procedure - Pure tone air conduction and bone conduction threshold audiometry with and without masking.
BSA. (2013). *Recommended Procedure for Tympanometry*. Reading: British Society of Audiology.
Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia, 34*(4), 263-272. doi: 10.1016/0028-3932(95)00104-2
Burle, B., van den Wildenberg, W. P. M., & Ridderinkhof, K. R. (2005). Dynamics of facilitation and interference in cue-priming and Simon tasks. *European Journal of Cognitive Psychology, 17*(5), 619-641. doi: 10.1080/09541440540000121
Cabeza, R., Grady, C. L., Nyberg, L., McIntosh, A. R., Tulving, E., Kapur, S., . . . Craik, F. I. M. (1997). Age-related differences in neural activity during memory encoding and retrieval: A positron emission tomography study. *Journal of Neuroscience, 17*(1), 391-400.
Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends Cogn Sci, 2*(7), 247-253. doi: 10.1016/S1364-6613(98)01189-9
Carroll, R., Meis, M., Schulte, M., Vormann, M., Kiessling, J., & Meister, H. (2015). Development of a German reading span test with dual task design for application in cognitive hearing research. *Int J Audiol, 54*(2), 136-141. doi: 10.3109/14992027.2014.952458
Carroll, R., Warzybok, A., Kollmeier, B., & Ruigendijk, E. (2016). Age-Related Differences in Lexical Access Relate to Speech Recognition in Noise. *Frontiers in Psychology, 7*, 16. doi: 10.3389/fpsyg.2016.00990
Centre for Research and Dissemination. (2009). *Systematic Reviews*. Centre for Research and Dissemination, University of York.
Cervera, T. C., Soler, M. J., Dasi, C., & Ruiz, J. C. (2009). Speech recognition and working memory capacity in young-elderly listeners: effects of hearing sensitivity. *Can J Exp Psychol, 63*(3), 216-226. doi: 10.1037/a0014321

CHABA. (1988). Speech understanding and aging. *J Acoust Soc Am, 83*(3), 859-895.
Chan, R. C. K., Lai, M. K., & Robertson, I. H. (2006). Latent structure of the Test of Everyday Attention in a non-clinical Chinese sample. *Archives of Clinical Neuropsychology, 21*(5), 477-485. doi: 10.1016/j.acn.2006.06.007
Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with 2 Ears. *Journal of the Acoustical Society of America, 25*(5), 975-979.
Ciorba, A., Bianchini, C., Pelucchi, S., & Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clinical Interventions in Aging, 7*, 159-163. doi: 10.2147/Cia.S26059
Claessen, M. H. G., van der Ham, I. J. M., & van Zandvoort, M. J. E. (2015). Computerization of the Standard Corsi Block-Tapping Task Affects Its Underlying Cognitive Concepts: A Pilot Study. *Applied Neuropsychology-Adult, 22*(3), 180-188.
Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.
Cooper, J. A., Sagar, H. J., Jordan, N., Harvey, N. S., & Sullivan, E. V. (1991). Cognitive impairment in early, untreated Parkinson's disease and its relationship to motor disability. *Brain, 114 (Pt 5)*, 2095-2122. doi: 10.1093/brain/114.5.2095
Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*, 819B.
Costello, A. B., & Osborne, J. (2005). Best practicies in exploratory factor analysis: four recommmantions for getting the most from you analysis. *Practical Assessment Research and Evaluation, 10*(7).
Cragg, L., & Nation, K. (2010). Language and the development of cognitive control. *Top Cogn Sci, 2*(4), 631-642. doi: 10.1111/j.1756-8765.2009.01080.x
Craik, F. I. M., & Tulving, E. (1975). DEPTH OF PROCESSING AND RETENTION OF WORDS IN EPISODIC MEMORY. *Journal of Experimental Psychology-General, 104*(3), 268-294. doi: 10.1037/0096-3445.104.3.268
Dalton, D. S., Cruickshanks, K. J., Klein, B. E. K., Klein, R., Wiley, T. L., & Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *Gerontologist, 43*(5), 661-668. doi: DOI 10.1093/geront/43.5.661
Daneman, M., & Carpenter, P. A. (1980). Individual-Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi: 10.1016/S0022-5371(80)90312-6
Davis, H., & Keller, F. (1998). *Colorado assessment test manual, Colorado Assessment Tests*. Colorado Springs.
Desjardins, J. L. (2016). The Effects of Hearing Aid Directional Microphone and Noise Reduction Processing on Listening Effort in Older Adults with Hearing Loss. *J Am Acad Audiol, 27*(1), 29-41. doi: 10.3766/jaaa.15030
Desjardins, J. L., & Doherty, K. A. (2013). Age-Related Changes in Listening Effort for Various Types of Masker Noises. *Ear and Hearing, 34*(3), 261-272.
Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis Testing Using Factor Score Regression: A Comparison of Four Methods. *Educational and Psychological Measurement, 76*(5), 741-770. doi: 10.1177/0013164415607618
Diamond, A. (2013). Executive functions. *Annu Rev Psychol, 64*, 135-168. doi: 10.1146/annurev-psych-113011-143750

DiDonato, R. M., & Surprenant, A. M. (2015). Relatively effortless listening promotes understanding and recall of medical instructions in older adults. *Frontiers in Psychology, 6*, 20.

Doumas, M., Smolders, C., & Krampe, R. T. (2008). Task prioritization in aging: effects of sensory information on concurrent posture and memory performance. *Exp Brain Res, 187*(2), 275-281. doi: 10.1007/s00221-008-1302-3

Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in hearing, 21*, 2331216517744675. doi: 10.1177/2331216517744675

Ellis, R. J., & Rönnberg, J. (2014). Cognition and speech-in-noise recognition: the role of proactive interference. *J Am Acad Audiol, 25*(10), 975-982. doi: 10.3766/jaaa.25.10.6

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23. doi: 10.1111/1467-8721.00160

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *Psychology of Learning and Motivation: Advances in Research and Theory, Vol 44* (Vol. 44, pp. 145-199). San Diego: Elsevier Academic Press Inc.

Ferguson, M. A., & Henshaw, H. (2015). Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. *Frontiers in Psychology, 6*. doi: 10.3389/Fpg.2015.00556

Ferguson, M. A., Kitterick, P. T., Chong, L. Y., Edmondson-Jones, M., Barker, F., & Hoare, D. (2017). Hearing aids for mild and moderate hearing loss in adults. *The Cochrane Collaboration*(9; CD012023), 1-46. doi: 10.1002/14651858.CD012023.pub2

Fletcher, R. (1998). *The City University Colour Vision Test* (Third ed.). Windsor: Keeler Ltd.

Fraser, S., Gagne, J. P., Alepins, M., & Dubois, P. (2010). Evaluating the Effort Expended to Understand Speech in Noise Using a Dual-Task Paradigm: The Effects of Providing Visual Speech Cues. *Journal of Speech Language and Hearing Research, 53*(1), 18-33. doi: 10.1044/1092-4388(2009/08-0140)

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am, 115*(5 Pt 1), 2246-2256. doi: 10.1121/1.689343

Füllgrabe, C., Moore, B. C. J., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience, 6*, 25.

Füllgrabe, C., & Rosen, S. (2016a). Investigating the Role of Working Memory in Speech-in-noise Identification for Listeners with Normal Hearing. *Adv Exp Med Biol, 894*, 29-36. doi: 10.1007/978-3-319-25474-6_4

Füllgrabe, C., & Rosen, S. (2016b). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Front Psychol, 7*, 1268. doi: 10.3389/fpsyg.2016.01268

Gagne, J. P., Besser, J., & Lemke, U. (2017). Behavioral Assessment of Listening Effort Using a Dual-Task Paradigm: A Review. *Trends in hearing, 21*. doi: 10.1177/2331216516687287

Gatehouse, S. (2003). Rehabilitation: identification of needs, priorities and expectations, and the evaluation of benefit. *Int J Audiol, 42 Suppl 2*, 2S77-83. doi: 10.3109/14992020309074648

Gold, J. M., Carpenter, C., Randolph, C., Goldberg, T. E., & Weinberger, D. R. (1997). Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. *Arch Gen Psychiatry, 54*(2), 159-165. doi: 10.1001/archpsyc.1997.01830140071013

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *J Exp Psychol Learn Mem Cogn, 22*(5), 1166-1183. doi: 10.1037/0278-7393.22.5.1166

Gordon-Salant, S., & Cole, S. S. (2016). Effects of Age and Working Memory Capacity on Speech Recognition Performance in Noise Among Listeners With Normal Hearing. *Ear Hear, 37*(5), 593-602. doi: 10.1097/AUD.0000000000000316

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36*(6), 1276-1285.

Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech Language and Hearing Research, 40*(2), 423-431.

Gordon-Salant, S., & Fitzgibbons, P. J. (2001). Sources of age-related recognition difficulty for time-compressed speech. *Journal of Speech Language and Hearing Research, 44*(4), 709-719. doi: 10.1044/1092-4388(2001/056)

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Cohen, J. I. (2015). Effects of age and hearing loss on recognition of unaccented and accented multisyllabic words. *J Acoust Soc Am, 137*(2), 884-897. doi: 10.1121/1.4906270

Gosselin, P. A., & Gagne, J. P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology, 50*(11), 786-792. doi: 10.3109/14992027.2011.599870

Grady, C. L., & Craik, F. I. M. (2000). Changes in memory processing with age. *Curr Opin Neurobiol, 10*(2), 224-231. doi: Doi 10.1016/S0959-4388(00)00073-8

Gronwall, D. M. (1977). Paced auditory serial-addition task: a measure of recovery from concussion. *Percept Mot Skills, 44*(2), 367-373. doi: 10.2466/pms.1977.44.2.367

Hasher, L., & Zacks, R. T. (1979). Automatic and Effortful Processes in Memory. *Journal of Experimental Psychology-General, 108*(3), 356-388. doi: 10.1037/0096-3445.108.3.356

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Front Syst Neurosci, 8*, 35. doi: 10.3389/fnsys.2014.00035

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London: Academic Press.

Heinrich, A., Henshaw, H., & Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Front Psychol, 6*, 782. doi: 10.3389/fpsyg.2015.00782

Heinrich, A., Henshaw, H., & Ferguson, M. A. (2016). Only behavioral but not self-report measures of speech perception correlate with cognitive abilities. *Frontiers in Psychology, 7*(576). doi: 10.3389/fpsyg.2016.00576

Heinrich, A., & Knight, S. (2016). The Contribution of Auditory and Cognitive Factors to Intelligibility of Words and Sentences in Noise. *Adv Exp Med Biol, 894*, 37-45. doi: 10.1007/978-3-319-25474-6_5

Heinrich, A., Knight, S., Young, M., Moorhouse, R., & Barry, J. G. (2014). *Assessing the interplay of linguistic factors and background noise characteristics to successful speech in noise listening*. Paper presented at the BSA.

Helfer, K. S., & Freyman, R. L. (2014). Stimulus and listener factors affecting age-related changes in competing speech perception. *Journal of the Acoustical Society of America, 136*(2), 748-759.

Helmstaedter, C., & Durwen, H. F. (1990). The Verbal Learning and Retention Test. A useful and differentiated tool in evaluating verbal memory performance. *Schweiz Arch Neurol Psychiatr, 141*(1), 21-30.

Higgins, J. P., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., . . . Cochrane Statistical Methods, G. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*, d5928. doi: 10.1136/bmj.d5928

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat Med, 21*, 1539-1558.

Humes, L. E. (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *J Acoust Soc Am, 112*(3 Pt 1), 1112-1132.

Humes, L. E., & Roberts, L. (1990). SPEECH-RECOGNITION DIFFICULTIES OF THE HEARING-IMPAIRED ELDERLY - THE CONTRIBUTIONS OF AUDIBILITY. *Journal of Speech and Hearing Research, 33*(4), 726-735. doi: 10.1044/jshr.3304.726

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., & Lee, L. (1994). Factors Associated with Individual-Differences in Clinical Measures of Speech Recognition among the Elderly. *Journal of Speech and Hearing Research, 37*(2), 465-474.

Jacobs, J. W., Bernhard, M. R., Delgado, A., & Strain, J. J. (1977). Screening for organic mental syndromes in the medically ill. *Ann Intern Med, 86*(1), 40-46.

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition, 19*(6), 741-758. doi: 10.1080/13825585.2011.652590

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *Q J Exp Psychol (Hove), 67*(9), 1842-1862. doi: 10.1080/17470218.2013.879391

Jerger, J. (1992). Can age-related decline in speech understanding be explained by peripheral hearing loss? *J Am Acad Audiol, 3*(1), 33-38.

Jerger, J., Jerger, S., & Pirozzolo, F. (1991). Correlational Analysis of Speech Audiometric Scores, Hearing-Loss, Age, and Cognitive-Abilities in the Elderly. *Ear and Hearing, 12*(2), 103-109. doi: 10.1097/00003446-199104000-00004

Jesse, A., & Janse, E. (2012). Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners. *Language and Cognitive Processes, 27*(7-8), 1167-1191. doi: 10.1080/01690965.2011.620335

Johnson, E. E., & Dillon, H. (2011). A comparison of gain for adults from generic hearing aid prescriptive methods: impacts on predicted loudness, frequency bandwidth, and speech intelligibility. *J Am Acad Audiol, 22*(7), 441-459. doi: 10.3766/jaaa.22.7.5

Johnston, R. S., & McDermott, E. A. (1986). SUPPRESSION EFFECTS IN RHYME JUDGMENT TASKS. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 38*(1), 111-124. doi: 10.1080/14640748608401587

Jolles, J., Houx, P., van Boxel, M., & Ponds, R. (1995). *Maastricht Aging Study: Determinants of Cognitive Aging*. Maastricht, The Netherlands: Neuropsych Publishers.

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

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval.

J Exp Psychol Learn Mem Cogn, 26(2), 336-358. doi: 10.1037/0278-7393.26.2.336
Kemtes, K. A., & Allen, D. N. (2008). Presentation modality influences WAIS Digit Span performance in younger and older adults. *J Clin Exp Neuropsychol, 30*(6), 661-665. doi: 10.1080/13803390701641414
Kessels, R. P., van den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. *Assessment, 15*(4), 426-434. doi: 10.1177/1073191108315611
Kessels, R. P., van Zandvoort, M. J., Postma, A., Kappelle, L. J., & de Haan, E. H. (2000). The Corsi Block-Tapping Task: standardization and normative data. *Appl Neuropsychol, 7*(4), 252-258.
Kessels, R. P. C., Overbeek, A., & Bouman, Z. (2015). Assessment of verbal and visuospatial working memory in mild cognitive impairment and Alzheimer's dementia. *Dement Neuropsychol, 9*(3), 301-305. doi: 10.1590/1980-57642015DN93000014
Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). REDUCING INFORMATIONAL MASKING BY SOUND SEGREGATION. *Journal of the Acoustical Society of America, 95*(6), 3475-3480. doi: 10.1121/1.410023
Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *Journal of the Acoustical Society of America, 122*(1), 418-435. doi: 10.1121/1.2743154
Knight, S., & Heinrich, A. (2017). Different Measures of Auditory and Visual Stroop Interference and Their Relationship to Speech Intelligibility in Noise. *Frontiers in Psychology, 8*, 21. doi: 10.3389/fpsyg.2017.00230
Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2014). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *J Acoust Soc Am, 135*(3), 1596-1606. doi: 10.1121/1.4863198
Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International journal of otolaryngology, 2012*, 865731. doi: 10.1155/2012/865731
Lin, F. R., Thorpe, R., Gordon-Salant, S., & Ferrucci, L. (2011). Hearing Loss Prevalence and Risk Factors Among Older Adults in the United States. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences, 66*(5), 582-590. doi: 10.1093/gerona/glr002
Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: a strong connection. *Psychol Aging, 9*(3), 339-355.
Luis, C. A., Keegan, A. P., & Mullan, M. (2009). Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *Int J Geriatr Psychiatry, 24*(2), 197-201. doi: 10.1002/gps.2101
MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. *Br J Audiol, 24*(1), 29-43.
MacPherson, A., & Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. *Trends Hear, 18*. doi: 10.1177/2331216514537722
Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203-243. doi: 10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978. doi: 10.1080/01690965.2012.705006

May, C. P. (1999). Synchrony effects in cognition: the costs and a benefit. *Psychon Bull Rev, 6*(1), 142-147. doi: 10.3758/BF03210822

Meister, H., Schreitmuller, S., Grugel, L., Beutner, D., Walger, M., & Meister, I. (2013). Examining Speech Perception in Noise and Cognitive Functions in the Elderly. *American Journal of Audiology, 22*(2), 310-312. doi: 10.1044/1059-0889(2012/12-0067)

Meister, H., Schreitmuller, S., Grugel, L., Ortmann, M., Beutner, D., Walger, M., & Meister, I. G. (2013). COGNITIVE RESOURCES RELATED TO SPEECH RECOGNITION WITH A COMPETING TALKER IN YOUNG AND OLDER LISTENERS. *Neuroscience, 232*, 74-82.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100. doi: 10.1006/cogp.1999.0734

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ, 339*, b2535. doi: 10.1136/bmj.b2535

Morris, & Jones, D. M. (1990). Memory Updating in Working Memory - the Role of the Central Executive. *British Journal of Psychology, 81*, 111-121. doi: 10.1111/j.2044-8295.1990.tb02349.x

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., . . . Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*(9), 1159-1165.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699.

Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw Hill.

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., . . . Kramer, S. E. (2017). Effects of Hearing Impairment and Hearing Aid Amplification on Listening Effort: A Systematic Review. *Ear and Hearing, 38*(3), 267-281. doi: 10.1097/Aud.0000000000000396

Olsen, W. O. (1998). Average Speech Levels and Spectra in Various Speaking/Listening Conditions: A Summary of the Pearson, Bennett, & Fidell (1977) Report. *Am J Audiol., 7*(2), 21-25.

Pals, C., Sarampalis, A., van Rijn, H., & Baskent, D. (2015). Validation of a simple response-time measure of listening effort. *Journal of the Acoustical Society of America, 138*(3), El187-El192. doi: 10.1121/1.4929614

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician Enhancement for Speech-In-Noise. *Ear and Hearing, 30*(6), 653-661. doi: 10.1097/AUD.0b013e3181b412e9

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise. *Plos One, 6*(5), 8. doi: 10.1371/journal.pone.0018082

Peelle, J. E., Troiani, V., Grossman, M., & Wingfield, A. (2011). Hearing loss in older adults affects neural systems supporting speech comprehension. *Journal of Neuroscience, 31*(35), 12638-12643. doi: 10.1523/JNEUROSCI.2559-11.2011

Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Journal of the Acoustical Society of America, 103*(1), 577-587. doi: 10.1121/1.421128

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience, Vol 35, 35*, 73-89. doi: 10.1146/annurev-neuro-062111-150525

Pichora-Fuller, M. K. (2003a). Cognitive aging and auditory information processing. *International Journal of Audiology, 42*, S26-S32.

Pichora-Fuller, M. K. (2003b). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42*, S59-S67. doi: 10.3109/14992020309074625

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., . . . Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing, 37*, 5S-27S. doi: 10.1097/aud.0000000000000312

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How Young and Old Adults Listen to and Remember Speech in Noise. *Journal of the Acoustical Society of America, 97*(1), 593-608.

Picou, E. M., & Ricketts, T. A. (2014). The Effect of Changing the Secondary Task in Dual-Task Paradigms for Measuring Listening Effort. *Ear and Hearing, 35*(6), 611-622.

Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual Consequences of "Hidden" Hearing Loss. *Trends in hearing, 18*, 1-11. doi: 10.1177/2331216514550621

Plummer, P., & Eskes, G. (2015). Measuring treatment effects on dual-task performance: a framework for research and clinical practice. *Frontiers in Human Neuroscience, 9*.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol, 88*(5), 879-903. doi: 10.1037/0021-9010.88.5.879

Pollack, I. (1975). AUDITORY INFORMATIONAL MASKING. *Journal of the Acoustical Society of America, 57*, S5-S5. doi: 10.1121/1.1995329

Posner, M. I., & Petersen, S. E. (1990). THE ATTENTION SYSTEM OF THE HUMAN BRAIN. *Annu Rev Neurosci, 13*, 25-42. doi: 10.1146/annurev.neuro.13.1.25

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Q J Exp Psychol, 20*(3), 241-248. doi: 10.1080/14640746808400158

Randell, M. (2017). Overview of the UK population: July 2017. 2018

Rao, S. M., Leo, G. J., Haughton, V. M., St Aubin-Faubert, P., & Bernardin, L. (1989). Correlation of magnetic resonance imaging with neuropsychological testing in multiple sclerosis. *Neurology, 39*(2 Pt 1), 161-166. doi: 10.1212/01.wnl.0000406006.96282.eb

Raven, J. C., Raven, J., & Court, J. H. (1982). *Mill Hill Vocabulary scale*. Oxford: Oxford Psychologists Press.

Reitan, R. M. (1958). VALIDITY OF THE TRAIL MAKING TEST AS AN INDICATOR OF ORGANIC BRAIN DAMAGE. *Perceptual and Motor Skills, 8*, 271-276. doi: 10.2466/pms.1958.8.3.271

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current Directions in Psychological Science, 17*(3), 177-182. doi: DOI 10.1111/j.1467-8721.2008.00570.x

Roberts, K. L., & Allen, H. A. (2016). Perception and Cognition in the Ageing Brain: A Brief Review of the Short- and Long-Term Links between Perceptual and Cognitive Decline. *Front Aging Neurosci, 8*, 39. doi: 10.3389/fnagi.2016.00039

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention: TEA*. Bury St. Edmunds, UK: Thames Valley Test Company.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *International Journal of Audiology, 42*, S68-S76. doi: 10.3109/14992020309074626

Rönnberg, J., Lunner, T., Zekveld, A., Sorqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front Syst Neurosci, 7*, 31. doi: 10.3389/fnsys.2013.00031

Rönnberg, J., Lyxell, B., Arlinger, S., & Kinnefors, C. (1989). Visual Evoked-Potentials - Relation to Adult Speechreading and Cognitive Function. *Journal of Speech and Hearing Research, 32*(4), 725-735. doi: 10.1044/jshr.3204.725

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology, 47*, S99-S105. doi: 10.1080/14992020802301167

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise & health, 12*(49), 263-269. doi: 10.4103/1463-1741.70505

Rönnberg, N., Rudner, M., Lunner, T., & Stenfelt, S. (2014). Memory performance on the Auditory Inference Span Test is independent of background noise type for young adults with normal hearing at high speech intelligibility. *Frontiers in Psychology, 5*, 11. doi: 10.3389/fpsyg.2014.01490

Rudner, M. (2016). Cognitive Spare Capacity as an Index of Listening Effort. *Ear and Hearing, 37*, 69S-76S. doi: 10.1097/aud.0000000000000302

Rudner, M., & Lunner, T. (2014). Cognitive Spare Capacity and Speech Communication: A Narrative Overview. *Biomed Research International*. doi: 10.1155/2014/869726

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology, 54*(1-3), 35-54.

Sandford, J., & Turner, A. (2004). *IVA+ plus: Integrated visual and Auditory Continuous Performance Test administration manual*. Richmond, VA: Brain Train (2004).

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective Measures of Listening Effort: Effects of Background Noise and Noise Reduction. *Journal of Speech Language and Hearing Research, 52*(5), 1230-1240. doi: 10.1044/1092-4388(2009/08-0111)

Schmidt, K.-H., & Metzler, P. (1992). *Wortschatztest [Vocabulary test (WST)]*. Weinheim: Beltz.

Schneider, B. A., Avivi-Reich, M., & Daneman, M. (2016). HOW SPOKEN LANGUAGE COMPREHENSION IS ACHIEVED BY OLDER LISTENERS IN DIFFICULT LISTENING SITUATIONS. *Experimental Aging Research, 42*(1), 40-63. doi: 10.1080/0361073x.2016.1108749

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Schoof, T., & Rosen, S. (2015). High sentence predictability increases the fluctuating masker benefit. *Journal of the Acoustical Society of America, 138*(3), El181-El186. doi: 10.1121/1.4929627

Schuchardt, K., Kunze, J., Grube, D., & Hasselhorn, M. (2006). Working memory deficits in children with low calculating and spelling performance. *Zeitschrift Fur Padagogische Psychologie, 20*(4), 261-268. doi: 10.1024/1010-0652.20.4.261

Scialfa, C. T. (2002). The role of sensory factors in cognitive aging research. *Can J Exp Psychol, 56*(3), 153-163.

Sengpiel, A. Sengpielaudio. Retrieved 28/02/2017, 2017, from http://www.sengpielaudio.com/TableOfSoundPressureLevels.htm

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cognitive Processing, 17*(1), 79-87. doi: 10.1007/s10339-015-0740-7

Smoorenburg, G. F. (1992). Speech Reception in Quiet and in Noisy Conditions by Individuals with Noise-Induced Hearing-Loss in Relation to Their Tone Audiogram. *Journal of the Acoustical Society of America, 91*(1), 421-437. doi: 10.1121/1.402729

Snijders, J. T., Luteijn, F., & Verhage, F. (1983). *Groninger Intelligentie Test (GIT). English: Groningen Intelligence Test*. Lisse, The Netherlands: Swets Test Publishers.

Somberg, B. L., & Salthouse, T. A. (1982). Divided attention abilities in young and old adults. *J Exp Psychol Hum Percept Perform, 8*(5), 651-663.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychol Aging, 14*(3), 458-472. doi: 10.1037/0882-7974.14.3.458

Sorqvist, P., Ljungberg, J. K., & Ljung, R. (2010). A sub-process view of working memory capacity: Evidence from effects of speech on prose memory. *Memory, 18*(3), 310-326. doi: 10.1080/09658211003601530

Spearman, C. (1904). "General intelligence " objectively determined and measured *American Journal of Psychology 15*, 201-292. doi: 10.2307/1412107

Stenbäck, V., Hallgren, M., Lyxell, B., & Larsby, B. (2015). The Swedish Hayling task, and its relation to working memory, verbal ability, and speech-recognition-in-noise. *Scandinavian Journal of Psychology, 56*(3), 264-272. doi: 10.1111/sjop.12206

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. doi: 10.1037/0096-3445.121.1.15

Studebaker, G. A. (1985). A RATIONALIZED ARCSINE TRANSFORM. *Journal of Speech and Hearing Research, 28*(3), 455-462.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., & Gwaltney, C. A. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. *Journal of the Acoustical Society of America, 105*(4), 2431-2444. doi: 10.1121/1.426848

Surprenant, A. M. (2007). Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging Neuropsychology and Cognition, 14*(2), 126-143. doi: 10.1080/13825580701217710

Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. Hove: Psychology Press.

Taljaard, D. S., Olaithe, M., Brennan-Jones, C. G., Eikelboom, R. H., & Bucks, R. S. (2016). The relationship between hearing impairment and cognitive function: a meta-analysis in adults. *Clinical Otolaryngology, 41*(6), 718-729. doi: 10.1111/coa.12607

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some Factors Underlying Individual Differences in Speech Recognition on PRESTO: A First Report. *Journal of the American Academy of Audiology, 24*(7), 616-634. doi: 10.3766/Jaaa.24.7.10

Treisman, A. M. (1964). Selective Attention in Man. *Br Med Bull, 20*, 12-16.

Tremblay, K. L., & Backer, K. C. (2016). Listening and Learning: Cognitive Contributions to the Rehabilitation of Older Adults With and Without Audiometrically Defined Hearing Loss. *Ear and Hearing, 37*, 155S-162S. doi: 10.1097/aud.0000000000000307

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381-403). New York and London: Academic Press Inc.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychol Aging, 17*(3), 453-467.

Tun, P. A., Williams, V. A., Small, B. J., & Hafter, E. R. (2012). The Effects of Aging on Auditory Processing and Cognition. *American Journal of Audiology, 21*(2), 344-350. doi: 10.1044/1059-0889(2012/12-0030)

Tun, P. A., & Wingfield, A. (1999). One voice too many: Adult age differences in language processing with different types of distracting sounds. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 54*(5), P317-P327. doi: 10.1093/geronb/54B.5.P317

Tun, P. A., Wingfield, A., & Stine, E. A. (1991). Speech-processing capacity in young and older adults: a dual-task study. *Psychol Aging, 6*(1), 3-9.

Unsworth, N., Heitz, R. R., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505. doi: 10.3758/BF03192720

Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., & Kollmeier, B. (2013). Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test. *Journal of the Acoustical Society of America, 134*(4), 3039-3056. doi: 10.1121/1.4818760

van Rooij, J. C., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *J Acoust Soc Am, 88*(6), 2611-2624. doi: 10.1121/1.399981

van Rooij, J. C., & Plomp, R. (1992). Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion. *J Acoust Soc Am, 91*(2), 1028-1033. doi: 10.1121/1.402628

Veneman, C. E., Gordon-Salant, S., Matthews, L. J., & Dubno, J. R. (2013). Age and Measurement Time-of-Day Effects on Speech Recognition in Noise. *Ear and Hearing, 34*(3), 288-299. doi: 10.1097/AUD.0b013e31826d0b81

Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The Relationship between High-Frequency Pure-Tone Hearing Loss, Hearing in Noise Test (HINT) Thresholds, and the Articulation Index. *Journal of the American Academy of Audiology, 23*(10), 779-788. doi: 10.3766/jaaa.23.10.4

Wayne, R. V., & Johnsrude, I. S. (2015). A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Research Reviews, 23*, 154-166. doi: 10.1016/j.arr.2015.06.002

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corp.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale - Third Edition (WAIS-III)*. San Antonio, TX: Harcout Assessment.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV)*. San Antonio, TX: Harcourt Assessment.

WHO. (1992). The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.

Wilson, B. S., Tucci, D. L., Merson, M. H., & O'Donoghue, G. M. (2017). Global hearing health care: new findings and perspectives. *Lancet, 390*(10111), 2503-2515. doi: 10.1016/S0140-6736(17)31073-5

Wingfield, A. (1996). Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. *J Am Acad Audiol, 7*(3), 175-182.

Wingfield, A., Tun, P. A., Koh, C. K., & Rosen, M. J. (1999). Regaining lost time: Adult aging and the effect of time restoration on recall of time-

compressed speech. *Psychology and Aging, 14*(3), 380-389. doi: 10.1037//0882-7974.14.3.380
Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.
Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage, 25*(3), 1002-1015. doi: 10.1016/j.neuroimage.2004.12.013
Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94.
Yu, L. D., Rao, A., Zhang, Y., Burton, P. C., Rishiq, D., & Abrams, H. (2017). Neuromodulatory Effects of Auditory Training and Hearing Aid Use on Audiovisual Speech Perception in Elderly Individuals. *Frontiers in Aging Neuroscience, 9*, 8. doi: 10.3389/fnagi.2017.00030
Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News, 2*(3), 7-10.
Zekveld, A. A., George, E. L., Kramer, S. E., Goverts, S. T., & Houtgast, T. (2007). The development of the text reception threshold test: a visual analogue of the speech reception threshold test. *J Speech Lang Hear Res, 50*(3), 576-584. doi: 10.1044/1092-4388(2007/040)
Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response. *Ear and Hearing, 32*(4), 498-510. doi: 10.1097/AUD.0b013e31820512bb
Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in neuroscience, 8*, 88. doi: 10.3389/fnins.2014.00088