The effects of ageing on the perception of speech in noise

Tim Schoof

A dissertation submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
UCL

Department of Speech, Hearing and Phonetic Sciences
Division of Psychology and Language Sciences
University College London

2014

Declaration

I, Tim Schoof, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The primary aim of this thesis is to answer the question why older adults often experience increased difficulties understanding speech in the presence of background noise, even in the absence of hearing impairment. Particular attention is paid to the potential contribution of age-related declines in subcortical auditory processing, as measured by the auditory brainstem response (ABR) and the frequency following response (FFR). In addition, the contribution of potential age-related declines in (spectro)temporal and cognitive processing is explored.

Contrary to expectations, the data showed that older adults who fit strict criteria of normal hearing (thresholds $\leq$ 25 dB HL up to 6 kHz in at least one ear) do not necessarily experience increased difficulties in the perception of speech in the presence of steady-state or amplitude-modulated noise. They only performed more poorly in the presence of two-talker babble. By contrast, older adults with near-normal hearing, with thresholds $\leq$ 25 dB HL up to 4 kHz, did experience increased difficulties understanding speech in steady-state noise.

In line with previous research, the data furthermore showed age-related declines in subcortical auditory processing, as indicated by less robust ABRs and FFRs, working memory, processing speed, and sensitivity to spectro-temporal modulations. By contrast, the normal-hearing older adults showed no age-related declines in behavioural measures of envelope or temporal-fine structure processing.

Taken together, these findings suggest that auditory neuropathy and cognitive declines associated with ageing do not necessarily lead to increased speech in noise difficulties. Normal-hearing older adults did not perform more poorly on a speech in noise task in the presence of steady-state or amplitude-modulated noise, despite declines in subcortical auditory processing, working memory, and processing speed. Moreover, when the older adults with normal or near-normal hearing did experience difficulties understanding speech in noise, this could not be explained by auditory neuropathy, cognitive declines, or reduced spectro-temporal modulation sensitivity.

Audiometric thresholds were the best predictor of individual differences in speech in noise performance.

This thesis furthermore explores the effects of masker modulation type and sentence complexity on the fluctuating masker benefit. The results indicated that contextual information plays a greater role in dip listening than in the perception of speech in steady-state noise.

Lastly, this thesis proposes a new technique to more rapidly collect the FFR. Recording times can significantly be reduced by presenting stimuli continuously, without interstimulus interval, and averaging across a single cycle of the response.

Contents

List of Figures

List of Tables

Acknowledgements

First of all, many thanks to Stuart Rosen, my supervisor, who taught me that the Answer to the Ultimate Question of Life, The Universe, and Everything is 2 kHz, and not, as is commonly believed, 42. Stuart has taught me a lot, not just about The Universe, but also about practical research matters. I could not have done this without him (also, it would not have been nearly as much fun!). I also want to thank my second supervisor, Ifat Yasin, for her support. I have really enjoyed our regular tea and cake meet-ups.

Special thanks goes to Steve Nevard, with whom I have spent many a day calibrating equipment in the basement. We often made the mistake of saying "this should only take an hour." We would usually still be there three days later, with things having gone from bad to worse. Thanks also to Andy Faulkner for always looking out for me.

I also want to thank Josh Bernstein for our fruitful collaboration. Many thanks also to Lina Kubli for checking the click ABR measures. Thanks also goes to Alain de Cheveigné for instigating the Rapid FFR project and useful discussions on numerous topics related to this thesis. A big thank you furthermore goes to Pam Souza, who gave me the idea to recruit participants through the ramblers association. Without this suggestion I would probably still be searching the streets of London for normal-hearing older adults. Thanks also to Erika Skoe for giving me a crash course in recording and analysing auditory brainstem responses.

Many thanks also to the SHaPS PhD students and postdocs, especially Sonia Granlund, Katharine Mair, Yasna Pereira Reyes, Mauricio Figueroa, Joaquín Atria, Wafa Alshangiti, Jieun Song, Albert Lee, Yasu Shinohara, Emma Brint, Lucy Carroll, Lou Stringer, Michèle Pettinato, Outi Tuomainen, and Kurt Steinmetzger, without whom this PhD would have been a very dull affair. I can't imagine what my PhD would have been like without the occasional Twister challenge, Super Mario Kurt championship, or office rowing event. Special thanks goes to Mark Wibrow for chicken

chicken chicken chicken chicken (chicken chicken?). Thanks also to Axelle Calcus, my partner in FFR crime, for all the great Skype sessions. And a big thank you goes to Rebecca Oyekan for her help with data collection.

I am much indebted also to Laura Vale, who kept me sane by plotting regular escapes out of London and taking me on hiking trips along the Thames, the North Downs, and through the Malaysian jungle.

I also want to thank Action on Hearing Loss and Age UK for funding this PhD project. Many thanks also to everyone who participated in this research.

Chapter 1

General introduction

According to estimates by the World Health Organisation (WHO), the proportion of the world's population over 60 years old will increase from 11% to 22%, or from 605 million to two billion, between 2000 and 2050 (WHO, 2012). This dramatic increase in the proportion of older adults that makes up the world's population is the result of major advances in healthcare combined with declining birth rates.

One of the most pressing questions related to population ageing is whether increased longevity will also be accompanied by good health and sustained social engagement, or whether the increased lifespan will instead be associated with longer durations of illness and dependency (NIH and WHO, 2011).

An important aspect of healthy ageing is the ability to successfully communicate with others. Communication difficulties can have a serious impact on everyday life. It can lead to limited access to (healthcare) services, social isolation, and feelings of loneliness and depression.

Older adults often experience increased communication difficulties, however, especially in understanding speech in noisy environments (CHABA, 1988). These difficulties can largely be attributed to age-related hearing loss, or presbycusis (e.g. Humes and Roberts, 1990; Souza and Turner, 1994; Takahashi and Bacon, 1992; Barrenäs and Wikström, 2000; van Rooij and Plomp, 1990; Humes, 1996).

The speech in noise difficulties typically experienced by older adults cannot fully be explained in terms of age-related hearing loss (e.g. Dubno and Morgan, 1984; Helfer and Wilber, 1990; Frisina and Frisina, 1997; Middelweerd et al., 1990). Audiometric thresholds often account for (only) 30-80% of the variability in speech in noise performance in older adults (for a review see Humes and Dubno, 2010). Furthermore, even in the absence of hearing impairment, older adults often experience increased difficulties understanding speech in noise (Dubno et al., 2002, 2003; Helfer and Freyman, 2008).

The central question in this thesis is: what is specific to ageing, independent of hearing loss, that explains why older adults typically experience increased difficulties understanding speech in noise? This is a longstanding question that remains largely unanswered. Most research on the perception of speech in noise by older adults has focussed on hearing impaired individuals (see reviews by Humes et al., 2007; Humes and Dubno, 2010; Houtgast and Festen, 2008). Instead, this thesis focuses on older adults with normal or near-normal hearing.

Three hypotheses are generally put forward to explain the speech in noise difficulties older adults typically experience (CHABA, 1988). First, the peripheral hypothesis suggests that the difficulties can largely be attributed to the effects of hearing impairment, such as reduced audibility and broadened auditory filters, and auditory neuropathy (i.e. neural dysfunction). Second, the central auditory hypothesis suggests an important role for age-related changes in auditory processing along the auditory pathway from the lower brainstem up to the auditory cortex. Third, the cognitive hypothesis suggests that age-related declines in cognitive function likely also affect speech in noise performance for older adults. It is most likely that the increased speech in noise difficulties are attributable to a combination of all three factors.

In this thesis, particular attention is paid to subcortical auditory processing in young and older listeners. Recent animal work has suggested that ageing may lead to neural degeneration even in the absence of a permanent elevation in audiometric thresholds (Kujawa and Liberman, 2009; Sergeyenko et al., 2013). This is often referred to as auditory neuropathy or hidden hearing loss. Auditory neuropathy is typically assessed by measuring click-evoked auditory brainstem responses (ABR). It has been hypothesised that age-related neural degeneration may in part contribute to the increased speech in noise difficulties typically experienced by older adults (Sergeyenko et al., 2013). Furthermore, recent research has suggested an important role for neural speech coding at the level of the brainstem for successful speech in noise perception (Hornickel et al., 2009; Song et al., 2011; Anderson et al., 2011). Neural speech coding at the brainstem can be assessed non-invasively by measuring the frequency following response (FFR), sometimes (depending upon the exact stimuli used) also referred to as the speech ABR, complex ABR, or envelope following response. The FFR reflects sustained synchronous neural firing in the brainstem in response to periodic auditory stimuli, such as pure tones (Moushegian et al., 1973), tonal sweeps (Krishnan and Parkinson, 2000), vowels (Aiken and Picton, 2008), and consonant-vowel syllables (Johnson et al., 2008). Research suggests that more robust FFRs are typically associated with better speech in noise performance (Hornickel et al., 2009; Song et al., 2011; Anderson et al., 2011). The role of age-related auditory neuropathy on speech in noise performance is discussed in more detail in chapters 3 and 4.

This thesis furthermore explores the relative contributions of potential age-related declines in temporal and spectro-temporal auditory processing as well as declines in cognitive function on speech understanding in noise for older adults. A more elaborate discussion of the role of these potential declines can be found in chapters 2 and 4.

The outline of this thesis is as follows. Chapter 2 explores the potential contributions of age-related declines in auditory temporal processing and cognition on the difficulties older adults typically experience understanding speech in noise. Young and older adults with normal hearing, defined as audiometric thresholds $\leq$ 25 dB HL up to 6 kHz (in at least one ear), participated in this study. The same participants also took part in the study reported in chapter 3. This study describes the potential contribution of age-related declines in subcortical speech coding, as measured by the FFR, to any speech in noise difficulties experienced by older listeners. Chapter 4 examines whether older adults are impaired in the ability to detect spectro-temporal modulations. Moreover, it addresses the question whether auditory neuropathy can to some extent explain potential spectro-temporal processing deficits and increased speech in noise difficulties for older adults. The older adults that participated in this study had audiometric thresholds $\leq$ 25 dB HL up to 4 kHz. Chapter 5 describes a study that looks at the effects of sentence complexity and masker modulation type on the fluctuating masker benefit. In addition, it examines whether the supramodal ability to fill in missing information, as measured by the text reception threshold (TRT) task, is a better predictor of speech perception in steady-state or fluctuating maskers. Lastly, chapter 6 describes a new technique to more rapidly collect the FFR. The proposed technique involves presenting stimuli continuously and averaging across a single cycle of the response.

Chapter 2

The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners

2.1 Introduction

As discussed in chapter 1, older adults often experience increased difficulties understanding speech in noisy environments (CHABA, 1988). While this can largely be attributed to an elevation in audiometric thresholds associated with ageing (e.g. Humes and Roberts, 1990; Souza and Turner, 1994; Takahashi and Bacon, 1992; Barrenäs and Wikström, 2000; van Rooij and Plomp, 1990; Humes, 1996), the audiogram alone cannot necessarily fully account for the changes in speech understanding in noise (e.g. Dubno and Morgan, 1984; Helfer and Wilber, 1990; Frisina and Frisina, 1997; Middelweerd et al., 1990). Even in the absence of hearing impairment, for instance, older adults often perform more poorly on a speech in noise task than younger listeners (Dubno et al., 2002, 2003; Helfer and Freyman, 2008). The question that logically follows is: what is specific to ageing, independent of hearing loss, that explains these difficulties? Two possible explanations that have often been suggested are age-related declines in (1) central auditory processing, particularly temporal processing, and (2) cognition (CHABA, 1988). This chapter addresses the question to what extent age-related declines in auditory temporal and cognitive processing can explain difficulties understanding speech in noise by normal-hearing older adults.

2.1.1 Speech perception in noise

One type of masker that seems particularly detrimental to older adults is competing speech (Tun and Wingfield, 1999; Tun et al., 2002; Helfer and Freyman, 2008; Rajan and

Cainer, 2008; Rossi-Katz and Arehart, 2009). Helfer and Freyman (2008), for instance, looked at the perception of syntactically simple sentences (e.g. "The cherries in the bowl are sweet") in the presence of different types of background noise for young adults with normal hearing and older adults with varying degrees of hearing sensitivity. They showed that older adults were more affected by two-talker babble (same and opposite sex to the target) than by steady-state or amplitude-modulated speech-shaped noise. Similarly, Rajan and Cainer (2008) measured speech reception thresholds (SRTs) to the simple BKB sentences (Bench et al., 1979) in the presence of steady-state speech-shaped noise and eight-talker babble for normal-hearing adults (31 participants aged 20 - 69 yrs divided into decade age cohorts). They found a decline in performance in their oldest age cohort (60 - 69 yrs) compared to their younger cohorts only in the presence of the competing talkers.

It has similarly been suggested that older adults benefit less from fluctuations in the masker compared to young adults (Gifford et al., 2007; Stuart and Phillips, 1996; Takahashi and Bacon, 1992; Dubno et al., 2002, 2003; Peters et al., 1998; Grose et al., 2009). Dubno et al. (2002), for example, compared SRTs for syllables in steady-state and 10-Hz square-wave amplitude modulated noise for young and older adults with normal hearing up to 4 kHz. They found that older adults benefited less from the amplitude fluctuations in the masker than the younger listeners. Their analyses indicated that this finding could not entirely be attributed to audibility differences across the groups. Similarly, Peters et al. (1998) showed that older adults with normal hearing up to 4 kHz performed more poorly in speech-envelope modulated (but not steady-state) speech-shaped noise compared to younger listeners when the target materials were the relatively simple sentences from the Hearing in Noise Test (HINT, e.g. "A boy fell from the window"; Nilsson et al., 1994). Using stricter criteria for normal hearing ($\leq$ 15 dB HL up to 6 kHz), Gifford et al. (2007) showed that older adults performed more poorly compared to younger listeners when listening to HINT sentences in the presence of 10-Hz square-wave amplitude modulated speech-shaped noise. While most studies have used relatively simple target materials, such as syllables (e.g. Stuart and Phillips, 1996; Dubno et al., 2002) or simple sentences (e.g. Peters et al., 1998; Gifford et al., 2007), Grose et al. (2009) examined the fluctuating masker benefit (FMB) using more complex materials (IEEE sentences, e.g. "The birch canoe slid on the smooth planks"; Rothauser et al., 1969). They found that older adults with normal hearing up to 4 kHz performed more poorly than younger listeners in the presence of 16 or 32 Hz square-wave modulated noise.

It should be noted that three of the six studies that examined the effects on age on the perception of speech in modulated noise and also looked at speech perception in steady-state noise showed significant group differences in the unmodulated masker (Stuart and Phillips, 1996; Dubno et al., 2002, 2003). It has been suggested, however, that the FMB may partly be dependent on the performance in steady-state noise (Oxenham and Simonson, 2009; Bernstein and Grant, 2009). Bernstein and Grant (2009) showed that FMB differences between normal-hearing and hearing-impaired listeners were smaller when compared with reference to the same steady-state noise SNR, suggesting that hearing-impaired listeners benefit more from fluctuations in the masker than previously thought. Similarly, the apparent age-related reduction in FMB may in part be attributed to differences that also become apparent in steady-state noise.

2.1.2 Temporal processing

These increased difficulties in speech perception in noise may partly result from age-related deficits in auditory temporal processing (e.g. Pichora-Fuller and Souza, 2003; Pichora-Fuller et al., 2007; Frisina and Frisina, 1997; CHABA, 1988; Gordon-Salant, 2005). A useful way to think about auditory temporal processing is in terms of the decomposition of sound in the time domain into a slowly varying envelope (Env) superimposed on a more rapidly varying temporal fine structure (TFS; see Fig. 2.1; Moore, 2008b). Older adults are indeed impaired in both Env and TFS processing, which may both have implications for the perception of speech in noise.

2.1.2.1 Envelope processing

The temporal speech Env, when presented across as few as four frequency bands, provides sufficient cues for successful speech perception in quiet (Drullman et al., 1994; Shannon et al., 1995). Given that Env processing is so important for speech understanding in quiet, it must similarly play an important role in the perception of speech in noise. Env processing may be particularly important for the perception of speech in fluctuating noise maskers. The decreased FMB typically reported for older adults may partly be due to an age-related decline in the ability to track envelope fluctuations in the masker. There is a discrepancy in the literature, however, as to whether older adults indeed have difficulties tracking amplitude modulations. While some behavioural and electrophysiological studies have found age-related differences in temporal envelope processing (Purcell et al., 2004; Leigh-Paffenroth and Fowler, 2006;

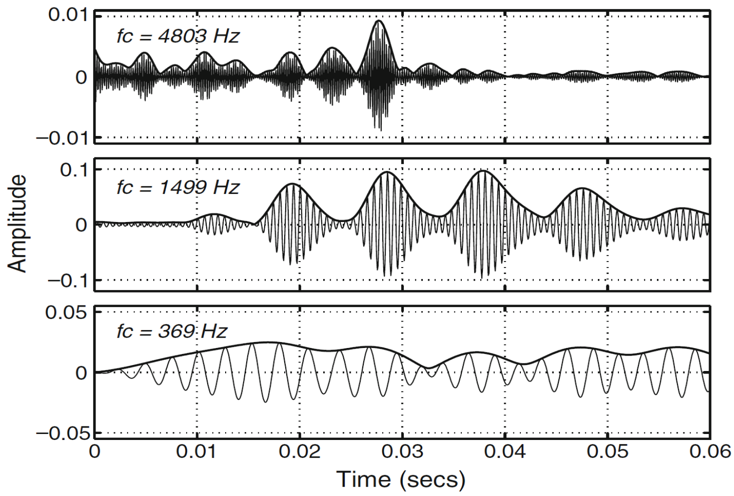


Figure 2.1: Waveforms at the outputs of simulated auditory filters at 369, 1499, and 4803 Hz in response to the sound "en" in "sense". The thick line represents the temporal envelope superimposed on the more rapidly varying temporal fine structure. From Moore (2008b).

He et al., 2008), others have not (Peters and Hall III, 1994; Takahashi and Bacon, 1992; Boettcher et al., 2001). Grose et al. (2009) pointed out that this discrepancy could be the result of differences in modulation rates used in these studies. They hypothesised that age-effects only become apparent at higher modulation rates. Purcell et al. (2004) found, for instance, that the maximum detectable modulation rate was significantly lower for older adults both behaviourally (young 567 Hz, older 264 Hz) and electrophysiologically, assessed by the envelope following response (young 494 Hz, older 294 Hz). However, Takahashi and Bacon (1992) measured sinusoidal AM detection of a broadband noise for modulation rates ranging from 2-1024 Hz but found no age effects for any of the modulation rates.

More consistent support for an age-related decline in Env processing comes from the gap detection literature. Gap detection is generally thought to reflect Env processing because it involves detecting *overall* changes in the amplitude of the more rapidly varying (noise) carrier (Moore, 2008a). Older adults have been shown to have increased gap detection thresholds using a variety of stimuli (He et al., 1999; Schneider and Hamstra, 1999; Schneider et al., 1994; Snell, 1997; Strouse et al., 1998; Snell et al., 2002 but see Moore et al., 1992). Snell (1997), for instance, compared gap detection thresholds for young and older adults with normal hearing up to 4 kHz ($\leq$20 dB HL) for a total of 24 conditions, with differences in overall presentation level, bandwidth of the noise carrier, modulation depth, and background noise, and found age-related declines in all conditions. Similarly, Schneider and Hamstra (1999) found age-related differences using pure tones rather than broadband noise, although group differences were less pronounced for longer stimuli (>250 ms).

Despite the fact that older adults show an age-related decline in temporal Env processing, this may not necessarily be the cause of the decreased FMB in older adults. Instead, declines in temporal Env processing may equally affect the perception of speech in steady-state and fluctuating maskers. For one, the FMB may be reduced in older adults at relatively slow modulation rates (e.g. 10 Hz; Gifford et al., 2007; Dubno et al., 2002, 2003), while age-related declines in Env processing only become apparent at higher modulation rates (e.g. Purcell et al., 2004). Instead, older adults might simply be less able to make use of the information in the gaps. Grose et al. (2009) examined the relative contribution of poor Env processing and reduced speech redundancy to the FMB by measuring speech perception in steady-state and square-wave amplitude-modulated noise at 16 and 32 Hz to time-compressed and unprocessed sentences. The time-compression was applied to degrade the speech cues in the dips of the fluctuating masker. They found that the reduction in FMB at 32 Hz compared to 16 Hz was the same for the two age groups and that the FMB deficit for normal-hearing older adults was smaller for time-compressed than normal-rate speech. This led them to suggest that the reduction in FMB for older listeners is primarily the result of a reduction in the quality of the available speech cues in the dips rather than a reduction in the ability to track the amplitude fluctuations of the masker.

2.1.2.2 TFS processing

A perhaps more compelling theory is that an age-related decline in TFS processing partly explains difficulties understanding speech in noise. It has been argued that while Env information may be sufficient for the perception of speech in quiet (Shannon et al., 1995), TFS may be required to successfully understand speech in the presence of interfering sound sources, particularly in the presence of competing talkers (Moore, 2008b, 2012; Nie et al., 2005; Lorenzi et al., 2006).

Support for an age-related decline in TFS processing comes from a variety of psychophysical measures (Vongpaisal and Pichora-Fuller, 2007; He et al., 2007; Grose and Mamo, 2010; Hopkins and Moore, 2011; Füllgrabe, 2013). Ageing has been shown, for instance, to negatively affect frequency modulation (FM) detection using low carrier frequencies (≤4 kHz) and low modulation rates (≤5 Hz) (He et al., 2007), pitch discrimination using harmonic and inharmonic complex sounds (Vongpaisal and

Pichora-Fuller, 2007; Hopkins and Moore, 2011; Füllgrabe, 2013), and the detection of inter-aural phase or time differences (Grose and Mamo, 2010; Hopkins and Moore, 2011; King et al., 2014).

It has previously been argued that TFS may be important to benefit from amplitude dips in fluctuating maskers (Schooneveldt and Moore, 1987; Lorenzi et al., 2006; Moore, 2008b). The idea that TFS is important for dip listening is in part based on the fact that cochlear implant users (with no access to TFS information in the speech signal) do not appear to be able to benefit from fluctuations in the masker (Nelson et al., 2003). In addition, it has been suggested that hearing impaired listeners are both unable to listen in the dips and impaired in TFS perception (Lorenzi et al., 2006). Lorenzi et al. (2006) showed that the ability of young hearing impaired listeners to understand TFS speech, in which envelope cues were removed while preserving TFS, was correlated with their ability to identify unprocessed speech in fluctuating maskers. It should be pointed out, however, that the perception of TFS speech may rely on Env cues that are reconstructed at the output of peripheral auditory filters (Heinz and Swaminathan, 2009; Zeng et al., 2004; Ghitza, 2001; Gilbert and Lorenzi, 2006) rather than TFS information contained in the signal.

More recently it has been argued that the role of TFS may in fact not be in detecting glimpses, but may instead allow efficient auditory scene analysis and/or spatial release from masking (Moore, 2012; Bernstein and Brungart, 2011). In other words, age-related declines in TFS could be equally important in accounting for difficulties in steady-state noises as well as those that fluctuate. Support for the idea that the ability to use TFS cues is not specifically involved in dip listening comes from Bernstein and Brungart (2011). They measured the perception of noise-vocoded and unprocessed speech in steady-state and four different fluctuating maskers (amplitude-modulated speech-shaped noise, one opposite-sex competing talker, two same-sex competing talkers, and 16-Hz interrupted noise). Noise vocoding removed the TFS information from the speech signal. When the SNR of the steady-state masker was equated for both vocoded and unprocessed speech, the FMB was not significantly different for speech with and without useful TFS information. Similarly, Moore (2012) reports an unpublished study that suggests that normal-hearing older adults may show the same FMB as younger listeners despite impaired TFS processing.

While an age-related decline in TFS processing may not necessarily explain possible declines in FMB, it likely contributes to perceptual difficulties in the presence of competing talkers since TFS information provides cues for sound source segregation based on pitch cues (e.g. Houtsma and Smurzynski, 1990; Smith et al., 2002; Bregman, 1990; Darwin and Carlyon, 1995). This could explain why older adults derive a smaller advantage than young listeners when the competing talker has the opposite, rather than the same, sex to the target talker. It could similarly explain why older adults typically require a larger fundamental frequency separation in concurrent vowel identification tasks (Vongpaisal and Pichora-Fuller, 2007; Summers and Leek, 1998; Snyder and Alain, 2005). Further support for the possible role of a decline in TFS processing in the presence of competing speech comes from a simulation study by Pichora-Fuller et al. (2007). They simulated auditory ageing by temporally jittering the speech signal, thereby disrupting the periodicity cues. In the presence of multi-talker babble, the perception of jittered speech by young normal-hearing listeners was comparable to that of unprocessed speech in noise by older adults.

2.1.3 Cognitive processing

An alternative explanation is that speech in noise difficulties in part arise from an age-related decline in cognitive processing. Ageing is associated with declines in several cognitive abilities that are thought to be important for the perception of speech in noise, such as working memory, attention, and processing speed (Kausler, 1994; Craik and Byrd, 1982; Salthouse, 1996). Older adults may require more cognitive resources, putting higher demands on top-down processing, to interpret the speech signal in the presence of background noise, especially when the input signal is further degraded as a result of auditory temporal processing deficits. Listening effort, which is thought to be an indicator of cognitive load, has indeed been shown to increase with age (Gosselin and Gagné, 2011; Desjardins and Doherty, 2013).

It should be pointed out that while ageing is associated with declines in abilities such as working memory, attention, and processing speed, linguistic knowledge in fact improves with age (for a review see Burke and Shafto, 2008). The size of one's vocabulary, for instance, increases with age (Verhaeghen, 2003). Similarly, the ability to use contextual cues to aid the recognition of speech in noise improves as one gets older (Pichora-Fuller et al., 1995; Dubno et al., 2000; Sheldon et al., 2008; Speranza et al., 2000). Pichora-Fuller et al. (1995) showed that, while older adults required higher SNRs for the perception of speech in the presence of competing talkers, their performance improved more for sentences with high compared to low context. Similarly, Speranza et al. (2000) found that older adults benefitted more from context on a reading task in which sentences were partially masked. While contextual cues may aid the reconstruction of partially masked text or speech, this may not be enough, however, to overcome the age-related difficulties understanding speech in noise or reading masked text. Humes et al. (2007), for instance, found that young participants reading low predictability sentences outperformed older adults reading high predictability sentences when the target was partially masked by dots, suggesting that the benefits of context were not enough to counteract the declines in the ability to fill in missing information.

2.1.3.1 Working memory

Working memory capacity, which refers to the ability to simultaneously store and process task-relevant information (Daneman and Carpenter, 1980; Baddeley, 1986), plays an important role in the perception of speech in noise (Daneman and Merikle, 1996; Rönnberg, 2003; Rönnberg et al., 2010; Akeroyd, 2008). In a review of 20 studies looking at the role of cognition in speech perception in noise, Akeroyd (2008) found that working memory capacity, especially as assessed by the reading span test (Daneman and Carpenter, 1980; Rönnberg et al., 1989), was most predictive of speech perception in noise. A recent model of language understanding (Rönnberg, 2003) indeed theorises a central role for working memory in speech perception in difficult listening situations. The model postulates that working memory is increasingly engaged when the auditory input is more degraded. Listeners with poor working memory are thus expected to be at a disadvantage particularly when the target signal is degraded.

Given that working memory capacity decreases with age (e.g. Van der Linden et al., 1994; Bopp and Verhaeghen, 2005; Craik and Jennings, 1992), it is not unreasonable to assume that a decline in working memory plays an important role in the difficulties older adults experience when understanding speech in noise. While research on the role of working memory in speech perception in noise for normal-hearing older adults is scarce (Pichora-Fuller et al., 1995; Meister et al., 2013), working memory has been shown to play a crucial role in aided and unaided speech perception for older adults with hearing impairment (e.g. Foo et al., 2007; Rudner et al., 2011).

2.1.3.2 Attention

Selective attention, which is best described as the ability to focus on relevant information and ignore irrelevant information, is probably equally important as working memory for successful speech understanding, especially in the presence of competing talkers since this involves the suppression of meaningful competing information. Older adults may be less successful at ignoring competing talkers as a result of an age-related decline in executive function, and more specifically a decline in inhibitory control (Hasher and Zacks, 1988; Hasher et al., 1999).

Inhibitory control is thought to prevent task-irrelevant information from taking up resources that could otherwise be used to process task-relevant information (Hasher and Zacks, 1988). A reduction in selective attention, for example, would thus not necessarily mean that someone would be more likely to report a to-be-ignored message. Instead, reduced selective attention could lead to poorer performance more generally since resources available for processing of the target talker are reduced. It has in fact been argued that age-related declines in working memory may ultimately be the result of declines in attentional resources or inhibitory control (Craik and Byrd, 1982; Hasher et al., 1999), since successful selective attention or inhibitory control likely prevents irrelevant information from entering working memory.

Support for an age-related decline in inhibitory control comes from different measures, such as the Wisconsin card sorting test (Rhodes, 2004) and the Stroop task (e.g. Cohn et al., 1984; Davidson et al., 2003). Individual differences in inhibitory control have furthermore been found to correlate with a measure of distraction by competing speech in older adults. Janse (2009), for example, found that Stroop interference, over and above the effects of hearing loss, predicted performance on a phoneme monitoring task in the presence of a competing talker in a group of older hearing impaired adults.

2.1.3.3 Processing speed

Underlying these age-related changes in working memory and attention may be a decline in processing speed. Salthouse (1985, 1996) argued that age-related declines in cognitive function may be the result of 'cognitive slowing'. A reduction in processing speed means that relevant operations cannot be executed successfully in the time available (limited time mechanism) and that the amount of simultaneously available information required for higher level processing is reduced (simultaneity mechanism).

Processing speed is often measured by substitution tests, which require the participant to match symbols within a certain time limit. Poorer performance on such substitution tests by older adults (Van der Elst et al., 2006; Pronk et al., 2013) has indeed been linked to speech perception in noise difficulties (Pronk et al., 2013; Tun and Wingfield, 1999).

2.1.3.4 Linguistic closure

Another factor thought to be important for speech perception in noise is linguistic closure, a modality aspecific linguistic capacity thought to reflect the ability to fill in missing information (Zekveld et al., 2007). Whether linguistic closure decreases with age, however, remains unclear.

Linguistic closure is often assessed using the Text Reception Threshold (TRT), a visual analogue of the SRT task, in which participants read sentences masked by bars of varying widths. This task was developed, more generally, to assess the extent to which inter-individual differences in speech in noise performance can be attributed to non-auditory factors. While some studies have reported a significant correlation between the TRT and age (Besser et al., 2012; Zekveld et al., 2011), other studies have not (e.g. Zekveld et al., 2007; Besser et al., 2013; Kramer et al., 2009). Support for an age-related decline in linguistic closure comes from a similar task, the Visual SPIN test, in which sentences of low and high predictability were masked by dots. Humes et al. (2007) found that older adults reported fewer correct items than younger adults across different SNRs. Similarly, Speranza et al. (2000) found that older adults required a higher SNR to identify the same number of words as the younger adults.

2.1.4 Relative contribution of auditory and cognitive declines

The relative contribution of auditory and cognitive factors to the perception of speech in noise for older adults has primarily been studied in the context of hearing aid use. These studies have typically found that cognitive factors were at least as important as hearing loss in accounting for the perception of aided speech in noise (see reviews by Humes et al., 2007; Humes and Dubno, 2010; Akeroyd, 2008; Houtgast and Festen, 2008). By contrast, cognitive factors typically only accounted for a small proportion of the variance, if any, of performance in unaided listening situations (Humes and Dubno, 2010; Akeroyd, 2008; Houtgast and Festen, 2008).

It has been argued that the role of cognition may be particularly important in the presence of fluctuating maskers or competing speech as opposed to stationary maskers (Akeroyd, 2008; Rönnberg et al., 2010; Gatehouse et al., 2003; Lunner and Sundewall-Thorén, 2007). Intuitively, the perception of speech in the presence of competing talkers likely requires more cognitive resources because it involves the suppression of meaningful speech. It has indeed been shown that the amount of listening effort, or cognitive load, is larger in the presence of a single talker than in fluctuating noise (Koelewijn et al., 2012, 2014). Koelewijn et al. (2012, 2014) showed that the pupil response for young normal-hearing listeners and hearing impaired listeners was larger when speech was masked by a single talker. In addition, it has been suggested that cognition may also play a more central role in the perception of speech in fluctuating as opposed to steady-state maskers. Several studies looking at the effects of different hearing aid configurations have shown, for instance, that people with high cognitive performance benefit more from fluctuations in the masker than people with lower scores on cognitive tasks (Gatehouse et al., 2003; Lunner and Sundewall-Thorén, 2007). Gatehouse et al. (2003) and Lunner and Sundewall-Thorén (2007) both found larger FMBs for people with high cognitive performance. Similarly, several studies have reported higher correlations between cognitive tasks (working memory and linguistic closure) and speech perception in fluctuating compared to steady-state maskers (e.g. George et al., 2007; Rudner et al., 2008). While this has been interpreted as a tentative suggestion that fluctuating maskers, including competing speech, put a higher demand on top-down cognitive processing than stationary maskers, the correlation coefficients may in fact not have been significantly different (c.f. Akeroyd, 2008; Rönnberg et al., 2010). In other words, the role of cognition may be equally important in steady-state and fluctuating maskers, despite apparent differences in the magnitude of the correlation coefficients.

The aim of this study was to assess why older adults, even in the absence of hearing impairment, experience increased difficulties understanding speech in noise. This study is novel in two ways. Firstly, relatively strict criteria for normal hearing were used (thresholds < 25 dB HL up to 6 kHz). Secondly, while the majority of studies examining the effects of ageing on speech perception in noise have used simple target stimuli, such as syllables (e.g. Stuart and Phillips, 1996; Dubno et al., 2002) or simple sentences (e.g. Peters et al., 1998; Gifford et al., 2007), this study used more complex targets (IEEE sentences; Rothauser et al., 1969)

Speech perception was assessed in the presence of different types of background noise. First, to examine whether normal-hearing older adults indeed benefit less from amplitude fluctuations in the masker, SRTs were measured in steady-state and amplitude-modulated noise (c.f. Gifford et al., 2007; Stuart and Phillips, 1996; Takahashi and Bacon, 1992; Dubno et al., 2002, 2003; Peters et al., 1998; Grose et al., 2009). Second, SRTs were also measured in the presence of two-talker babble since competing speech is both ecologically valid and particularly detrimental for older adults (Tun and Wingfield, 1999; Helfer and Freyman, 2008; Rajan and Cainer, 2008). In addition, various measures of auditory temporal (Env and TFS) and cognitive processing (working memory, attention, processing speed, linguistic closure, and reading skills) were assessed to examine the relative contribution of declines in both domains on speech perception difficulties.

Age-related cognitive declines may be expected to be the primary contributor to increased difficulties in speech perception in noise for normal-hearing older adults as individual differences in cognitive processing appear to be the most important factor explaining aided speech understanding in noise, for hearing impaired older adults, after accounting for differences in audiometric thresholds (see reviews by Humes and Dubno, 2010; Akeroyd, 2008; Houtgast and Festen, 2008). However, cognition may in fact be less important for normal-hearing listeners because the speech signal is less distorted in the absence of hearing loss (due to e.g. reduced audibility and frequency selectivity). The relative contributions of age-related auditory temporal and cognitive processing declines may furthermore depend on the type of background noise, with cognitive processing potentially being more important in more demanding situations such as in the presence of competing talkers.

2.2 Methods

2.2.1 Participants

Nineteen young (19 - 29 years old, mean 23.7 yrs, sd 2.9 yrs, 10 males) and 19 older (60 - 72 years old, mean 64.1 yrs, sd 3.3 yrs, 3 males) monolingual native English speakers participated in this study. All participants had near-normal hearing defined as (air-conducted) pure-tone thresholds of 25 dB HL or better at octave frequencies from 0.25 to 4 kHz in both ears and at 6 kHz in at least one ear (Fig. 2.2). In addition, all participants over the age of 65 had normal cognitive function (scores $\geq$17 MMSE telephone version (Roccaforte et al., 1992)) and normal or corrected-to-normal vision. None of the participants reported a history of language or neurological disorders. All participants signed a consent form approved by UCL Research Ethics Committee and were paid for their participation.

2.2.2 Speech perception in noise

Speech reception thresholds (SRT) were measured for sentences in different types of background noise. The target stimuli were pre-recorded IEEE sentences (Rothauser et al., 1969) produced by a male talker with a standard Southern British accent. Each sentence contained five keywords. The sentences were presented in steady-state speech-shaped noise (SS), speech-shaped noise sinusoidally amplitude modulated at 10 Hz with a modulation depth of 100% (AM), and two-talker babble. The speech-shaped noise used in this study was an approximation of the long-term average spectrum of

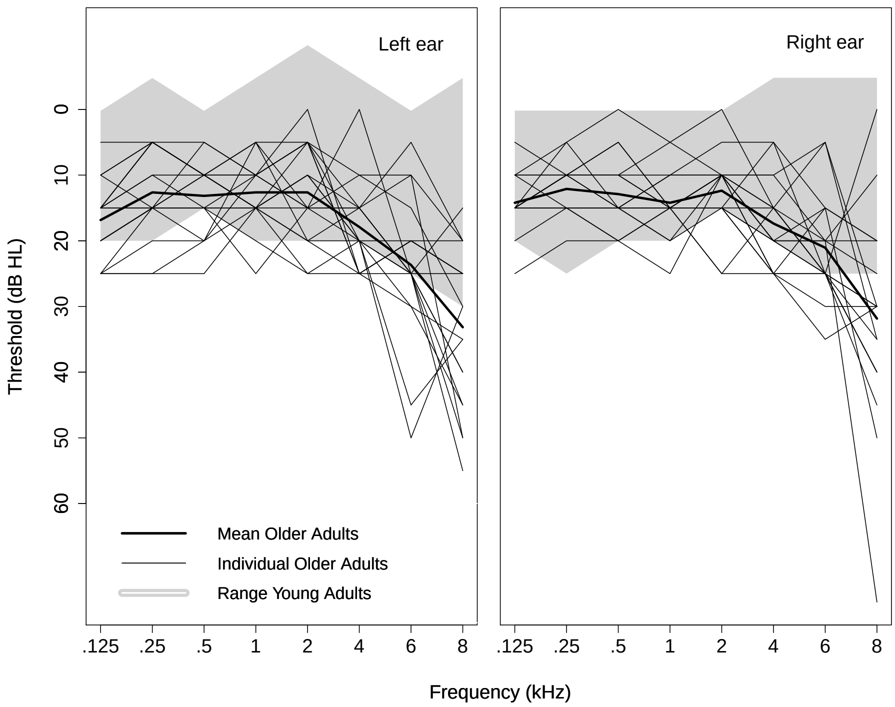


Figure 2.2: Individual audiograms for older adults are plotted for the left and right ear separately. The shaded area represents audiometric thresholds for the younger adults.

speech as described by Rosen et al. (2013). The two-talker babble was composed of connected discourse produced by two male talkers who were different from the target talker (c.f. Rosen et al., 2013). The masker always started 600 ms prior to stimulus onset and was gated on and off across 100 ms.

To rule out possible contributions of differences in audiometric thresholds above 6 kHz, the stimuli were low-pass filtered at 6 kHz using a fourth order Butterworth filter. It should be pointed out, however, that it is possible that the participants still used information processed in the basal region of the cochlea as this region was not masked. In addition, for six older participants with thresholds $>$ 25 dB HL at 6 kHz in one ear the stimuli were spectrally shaped using the National Acoustics Laboratories-Revised (NAL-R) linear prescriptive formula based on their individual thresholds (Byrne and Dillon, 1986).

The participants were seated in a soundproof booth and listened to the stimuli over Sennheiser HD 25 headphones. They were asked to repeat verbatim what they heard. The experimenter scored responses using a graphical user interface (GUI) which showed the five key words. The scoring screen was not visible to the participants and no feedback was provided.

The SNR was varied adaptively following the procedure described by Plomp and Mimpen (1979). The first sentence was presented at an SNR of -10 dB. Until at least three out of five key words were correctly repeated, the SNR was increased by 6 dB on the next presentation. The initial sentence was repeated until at least three out of five keywords were repeated correctly or the SNR reached 30 dB. For each subsequent sentence the SNR increased by 2 dB when zero to two key words were correctly repeated or decreased by the same amount for three to five correct repetitions. The number of trials was fixed at 20, tracking 50% correct.

SRTs for each condition were measured twice. A measurement was repeated, with a different set of sentences, when fewer than three reversals were obtained or when the standard deviation across the final reversals exceeded 4 dB. Thresholds for each run were computed by taking the mean SNR (dB) across the final number of reversals.

Participants were given brief training on the different conditions to familiarise them with the different types of background noise. Practice consisted of five trials and started at 0 dB SNR. The order of conditions in the experiment proper was counterbalanced across participants following a Latin square design. Stimuli were presented binaurally at 70 dB SPL.

2.2.3 Subjective measure of speech perception in noise

Participants were asked to complete section one of the Speech, Spatial, and Qualities of Hearing Scale (SSQ; Gatehouse and Noble, 2004), which addresses listeners' abilities to understand speech in quiet as well as in the presence of different types of noise. Composite scores were calculated for each participant by averaging across all questions.

2.2.4 Temporal processing

Participants completed three tasks that assess temporal processing; gap detection, amplitude modulation (AM) detection and frequency modulation (FM) detection. While the gap and AM detection tasks are concerned with temporal resolution in the envelope domain, the FM detection task assesses processing of TFS. The general procedure was similar for all three tasks. More details on the different tasks are provided below.

In all three tasks, a 3AFC paradigm was used and participants were asked to identify the stimulus that either contained a gap, or was modulated in amplitude or frequency. The duration of the gap or the depth of modulation was varied adaptively following the adaptive three-down, one-up procedure thus tracking 79% (Levitt, 1971).

Thresholds were obtained across two runs. A run was terminated after six reversals or after a maximum of 50 trials. Thresholds were computed by taking the mean gap duration or modulation depth across the last four reversals of each run. Thresholds reported here are the mean across the two runs.

Participants received training on five trials to familiarise themselves with the task. During this brief training they received visual feedback. During the experiment proper no feedback was provided.

Stimuli were presented binaurally over Sennheiser HD 25 headphones at 70 dB SPL. The order of the three tasks was counterbalanced across participants following a Latin square design.

2.2.4.1 Gap detection

Gap detection thresholds were measured using three 3-kHz-wide noises bandpass filtered between 1 and 4 kHz. A relatively wide band of noise was used as this limits the confounding effect of inherent fluctuations of the noise source on gap detection thresholds. The stimuli had a duration of 400 ms with a 10 ms rise-fall time and an interstimulus interval of 500 ms. The bands of noise were generated online at the start of each trial. All three noise bursts were thus based on the same underlying 400 ms section of noise. When a temporal gap was present in the stimulus, it was centred 300 ms after stimulus onset. Gap durations were varied from 0.5 to 7 ms in 20 logarithmic steps. Gaps were created by zeroing the waveform. Since this results in spectral cues that could aid the listener in identifying the presence of a gap, the stimuli were filtered to the required bandwidth after the insertion of the gap using a fourth order Butterworth filter. It should be noted that this procedure causes some temporal smearing of the gap. However, for relatively shallow filters this should not affect gap detection thresholds too much (c.f. Eddins et al., 1992).

The initial gap duration was 7 ms and was decreased after each trial until an error was made. Subsequently, three consecutive correct responses were required to decrease the gap duration, while one incorrect response increased the gap duration. The initial step size was three logarithmic steps and was decreased to two and finally one logarithmic step after each reversal. To prevent the gap duration from decreasing too far below the participant's threshold during the first few runs, the step size was automatically set to one logarithmic step once the gap duration was ≤ 1 ms. A run was repeated when fewer than three reversals were obtained or when the standard deviation across the final reversals exceeded 2 ms.

2.2.4.2 AM detection

As in the gap detection task, AM detection thresholds were measured using three 3-kHz-wide noises bandpass filtered between 1 and 4 kHz. The temporal-modulation transfer function was determined on the basis of AM detection thresholds for five (sinusoidal) AM rates: 10, 20, 40, 80, and 160 Hz. These modulation rates are all multiples of 10 Hz, which is the modulation rate of the masker used in the speech perception in noise task. The duration of the stimuli was 500 ms, which resulted in a whole number of AM cycles in all four conditions. The stimuli had a 10 ms rise-fall time and a 500 ms interstimulus interval. As in the gap detection task, the bands of noise were generated online at the start of each trial, which meant that the three stimuli in each trial were composed of the same noise sample. Amplitude modulation depths varied in 25 steps of 1 dB from -8 to -32 dB for rates up to 80 Hz and from -5 to -29 dB for the 160 Hz modulation rate. Since AM of bandpassed noise produces spectral side bands, the stimuli were filtered using a fourth order Butterworth filter after modulation. It should be noted that this may have reduced the effective modulation depth, especially for higher AM rates, although the filtering used should not have much of an effect (c.f. Eddins, 1999, 1993).

On the initial trial, the modulation depth was set to -8 dB, or -5 dB for the 160 Hz modulation rate, and was decreased after each trial until the participant gave an incorrect response. Subsequently, three consecutive correct responses were required to decrease the AM depth, while one incorrect response increased the AM depth. The initial step size was 6 dB, and was decreased in four steps after each reversal to the final step size of 1 dB. To prevent the AM depth from overshooting the participant's threshold during the initial runs, the step size was automatically set to 1 dB once the AM depth reached $\leq$ -25 dB for modulation rates of 10 and 160 Hz, and $\leq$ -20 dB for modulation rates of 20, 40, and 80 Hz. A run was repeated when fewer than three reversals were obtained or when the standard deviation across the final reversals exceeded 3 dB. The order of conditions was counterbalanced across participants following a Latin square design.

Since the temporal modulation transfer function (TMTF) resembles the form of a low-pass filter (Viemeister, 1979), the AM detection thresholds were fitted with an equation describing the frequency response of a low-pass Butterworth filter using a non-linear least-squares regression (Eddins, 1993):

$$y = 10log_{10}\left(\frac{1}{1+(\alpha f)^2}\right) + c \tag{2.1}$$

where y is the gain of the imputed filter (in dB) and f is the modulation rate in Hz. The inverse of α gives the -3 dB cutoff frequency (TMTF cutoff frequency) and c (the y-intercept) provides a measure of efficiency (AM efficiency). Note that a higher α (i.e. a higher cutoff frequency) and a lower c (i.e. better efficiency) indicate better performance.

2.2.4.3 FM detection

FM detection thresholds were determined using a 1 kHz sinusoidal carrier modulated at 2 Hz. A relatively low carrier frequency and modulation rate were used to ensure participants could only detect FM based on temporal cues (Moore and Sek, 1995, 1996). Frequency modulation depths varied logarithmically between 0.02 and 4.5 dB in 30 steps. The stimuli had a duration of 1 s, which is equal to two FM cycles. The interstimulus interval was set to 500 ms.

On the initial trial the modulation depth was set to 4.5 dB and was decreased after each trial until the listener made an error. Subsequently, three consecutive correct responses were required to decrease the FM depth, while one incorrect response increased the FM depth. The initial step size was three logarithmic steps, and was decreased in three steps after each reversal to the final step size of one logarithmic step. In addition, the step size was automatically set to one logarithmic step once the FM depth reached $\leq$ 0.57 dB to prevent the FM depth from overshooting the participant's threshold during the initial runs. A run was repeated when fewer than 3 reversals were obtained or when the standard deviation across the final reversals exceeded 2 dB.

FM detection thresholds are reported as modulation indices, which is the modulation depth divided by the modulation rate (2 Hz).

2.2.5 Cognitive processing

Cognitive skills were assessed in the visual domain to ensure that auditory factors did not influence these measures.

2.2.5.1 Working memory

A reading span task was used to examine participants working memory capacity (Rönnberg et al., 1989). This task was designed to tax not only information storage and rehearsal (as do, for example, digit span and word span tasks) but also information processing. The reading span task developed by Rönnberg and colleagues is an extension of the task developed by Daneman and Carpenter (1980). Here, participants were asked to read sequences of three to six three-word sentences and judge whether the sentence was semantically sensible or not (e.g.: "The train sang a song", or "The girl brushed her teeth"). At the end of each sequence of sentences, participants were asked to recall either the first or last word of each sentence in the correct order. The typeface of the text was Helvetica with font size 40. Words were presented in black on a grey background at 0.8 s/word. The inter-sentence interval, during which participants are required to make a semantic judgement, was 1.75 s. Participants were given one sequence of three sentences as a practice trial. During the testing phase, participants were presented with three runs of each sequence length (i.e. three to six sentences). The number of correctly remembered words was recorded.

2.2.5.2 Attention

Participants were assessed on the Visual Elevator task, a subtask of the Test of Everyday Attention (TEA; Robertson et al., 1996). It is thought to reflect an ability to switch attention, which is important for understanding speech in noise, especially in the presence of competing talkers. In essence, the participants' task was to count in a certain direction and at a given cue start counting in the opposite direction. The task consists of ten trials. Participants were asked to determine the floor number for each item and complete the task as fast as they could. The responses for each trial and the total time required to complete all ten trials were recorded. The total number of reversals for all correct responses were subsequently recorded. The final score was calculated by dividing the total duration required to complete the task (in seconds) by the total number of reversals for the correct responses.

2.2.5.3 Processing speed

To assess processing speed, participants were asked to complete the Letter Digit Substitution Test (LDST; Van der Elst et al., 2006). Participants were asked to complete the written version of the LDST. They were provided with a key in which the numbers 1 to 9 are each paired with a different letter. The test items, consisting of eight rows of 15 randomised letters, were printed below the key. The letters and digits were printed in font size 14. None of the participants had difficulties reading the items. The participants were asked to replace the letters by the corresponding digits as quickly as possible in sequential order. The first ten items were practice items. After completion of the test items they were given 60 seconds to substitute as many items as possible. The score is the number of correctly substituted items.

2.2.5.4 Text reception threshold

The text reception threshold (TRT) is a visual analogue of the speech reception threshold (SRT), especially in fluctuating noise (Zekveld et al., 2007; Besser et al., 2012). This task was developed to measure the variance in speech perception in noise abilities that are associated with supra-modal cognitive and linguistic skills. In this task sentences that are partly masked by a vertical bar pattern are presented on a computer screen.

As in the speech perception in noise task (measuring SRTs), the target stimuli were IEEE sentences (Rothauser et al., 1969). While the target stimuli were taken from the same corpus, the specific sentences used in the two tasks were different. The participants were seated approximately 50 cm from the screen. The typeface used to present the sentences was Arial, with a font size of 28. The background colour was white, the masked bar pattern was black, and the sentences were presented in red. The participants were asked to read the sentence out loud. The experimenter scored responses using a graphical user interface (GUI) which showed all the words in the sentence. The scoring screen was not visible to the participants and no feedback was provided.

The degree of masking was varied adaptively following the procedure described by Plomp and Mimpen (1979). The first sentence was presented with 16% unmasked text. Until the sentence was correctly repeated, the percentage of unmasked text was increased by 12% on the next presentation. Subsequent sentences were only presented once. When a sentence was correctly repeated, the degree of masking was increased by 6%. Conversely, the degree of masking was decreased by 6% when a sentence was not repeated correctly, thus tracking 50% correct.

TRTs were measured in response to two lists of 20 sentences each. Thresholds for each run were computed by taking the mean percentage unmasked text across sentences 5-20. The thresholds reported here are the mean across the two trials.

2.2.5.5 Reading skills

Given that both the text reception threshold and the reading span tasks rely heavily on reading, participants were assessed on reading ability using the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999). Participants were asked to read out a list of 104 English words as fast as they could. Subsequently, they were asked to do the same for a list of 84 non-words. The words were presented in Arial font size 20. While the first subtask assesses participants' sight reading skills, the second subtask addresses their phonemic decoding efficiency. The TOWRE is aimed at children and normally assesses the number of words that can be correctly identified within 45 seconds. However, to avoid any ceiling effects in adults, participants read out all the words on the list and reading ability was assessed in terms of the time it took them to read the whole list. The score for this task was calculated by dividing the total duration required to complete the task by the number of correctly read items.

2.3 Results

Data points that fell outside the mean $\pm$ 3 sd were considered outliers and excluded from the analyses reported below. In total, ten data points were excluded (data points from the older group were excluded for AM detection threshold at 160 Hz (one), TMTF cut-off (one), AM efficiency (one), TEA (two), TRT (one); data points from the young group were excluded for AM detection threshold at 160 Hz (one), TEA (two), and non-words TOWRE (one)).

Descriptive statistics for all measures as well as confidence intervals for the group differences are summarised in table 2.1.

2.3.1 Audiometric thresholds

While both groups had near-normal hearing, defined as pure-tone thresholds $\leq$25 dB HL up to 4 kHz in both ears and at 6 kHz in at least one ear, their thresholds were significantly different. Independent t-tests indicated that pure-tone averages (PTA) across 0.5 - 4 kHz (all $\leq$25 dB HL) were significantly higher (i.e. worse) by 7 dB HL for the older age group ($t_{(36)}$ = -6.4, $p < 0.001$). This could potentially contribute to any group differences that might exist for the auditory tasks (SRT, gap detection, AM detection, and FM detection; see section 2.3.6).

Analyses were conducted on a PTA across 0.5 - 4 kHz since the auditory tasks in this study, with the exception of the SRT task, did not have energy above 4 kHz. While the materials in the SRT task did contain energy above 4 kHz, stimuli for six older adults were spectrally shaped using the NAL formula to account for audibility differences.

2.3.2 Speech perception in noise

Older adults were expected to perform more poorly (i.e. higher SRTs) in all three background noises. However, the older adults had higher SRTs only in the presence of two-talker babble (Fig. 2.3). A mixed effects model with condition (AM, SS, babble) and group (young, old) as fixed factors and participant and sentence list as random

	Descriptive statistics		
Dependent variable	mean (sd) young	mean (sd) older	CI
SRT SS	-3.7 (1.5)	-3.8 (1.2)	[-1.0, 0.8]
SRT AM	-6.7 (2.4)	-6.4 (1.2)	[-1.0, 1.5]
SRT babble	-1.6 (2.0)	-0.3 (0.8)	[0.4, 2.4]
SSQ	7.5 (1.0)	7.1 (1.2)	[-1.2, 0.3]
AM cutoff	110 (50)	111 (62)	[-36, 38]
AM efficiency	-19.4 (2.0)	-19.5 (1.0)	[-1.2, 0.9]
FM detection	1.0 (0.4)	1.1 (0.3)	[-0.2, 0.3]
Gap detection	4.3 (0.8)	4.6 (1.2)	[-0.4, 0.9]
Working memory	32 (5.6)	24 (5.0)	[-11.6, -4.6]
Attention	4.2 (1.9)	5.0 (1.7)	[-0.4, 2.0]
Processing speed	39 (6.7)	34 (6.6)	[-9.3, -0.3]
TRT	63 (4.3)	63 (3.5)	[-1.8, 3.2]
Reading words	0.49 (0.09)	0.48 (0.07)	[-0.06, 0.05]
Reading non-words	0.74 (0.14)	0.72 (0.15)	[-0.1, 0.08]

Table 2.1: Descriptive statistics (mean and standard deviation) for the young and older adults separately as well as confidence intervals for the group differences are provided for all measures.

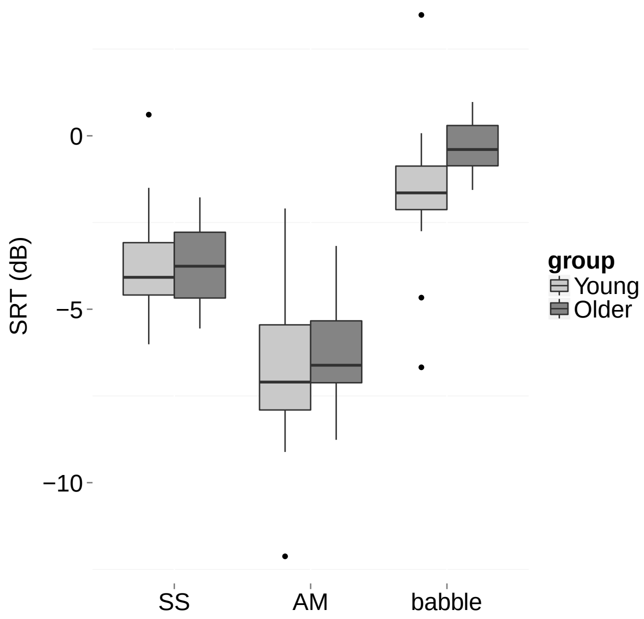


Figure 2.3: Boxplots of speech reception thresholds (SRT, in dB) for young (light grey) and older (dark grey) listeners for SS noise (left), AM noise (middle), and two-talker babble (right).

factors showed a significant interaction between condition and group ($F_{(2,186)} = 5.6$, $p = 0.004$). Post-hoc independent t-tests revealed a significant difference between the two age groups for babble only, with young listeners performing better than older listeners by 1.4 dB ($t_{(36)} = 2.8$, $p = 0.008$, Cohen's $d = 0.9$; all other $p > 0.6$).

Overall, SRTs in AM noise were expected to be lower (i.e. better) compared to SRTs in SS noise, indicative of dip listening. Furthermore, SRTs in babble were expected to be higher (i.e. worse) compared to the two noise maskers (Rosen et al., 2013). Post-hoc independent t-tests indeed revealed a significant dip listening effect, with lower SRTs in AM compared to SS noise ($t_{(37)} = 12.9$, $p < 0.001$, Cohen's $d = 1.4$, mean difference = 2.7 dB). In addition, SRTs in babble were significantly higher compared to the two noise maskers (SS: $t_{(37)} = 8.5$, $p < 0.001$, Cohen's $d = 2.5$, mean difference = 2.6 dB; AM: $t_{(37)} = 16.3$, $p < 0.001$, Cohen's $d = 1.4$, mean difference = 5.3 dB).

While there may be no group differences in SRTs in SS or AM noise and only a small difference in babble, it may be the case that particular older adults experience increased difficulties with one or more of the maskers. To explore these individual differences, we performed a deviance analysis (c.f. Ramus et al., 2003). The SRT scores were converted to z-scores and the deviance threshold was set to 1.65 sd above the mean SRT of the young group. Thus, participants were identified who performed more poorly than the

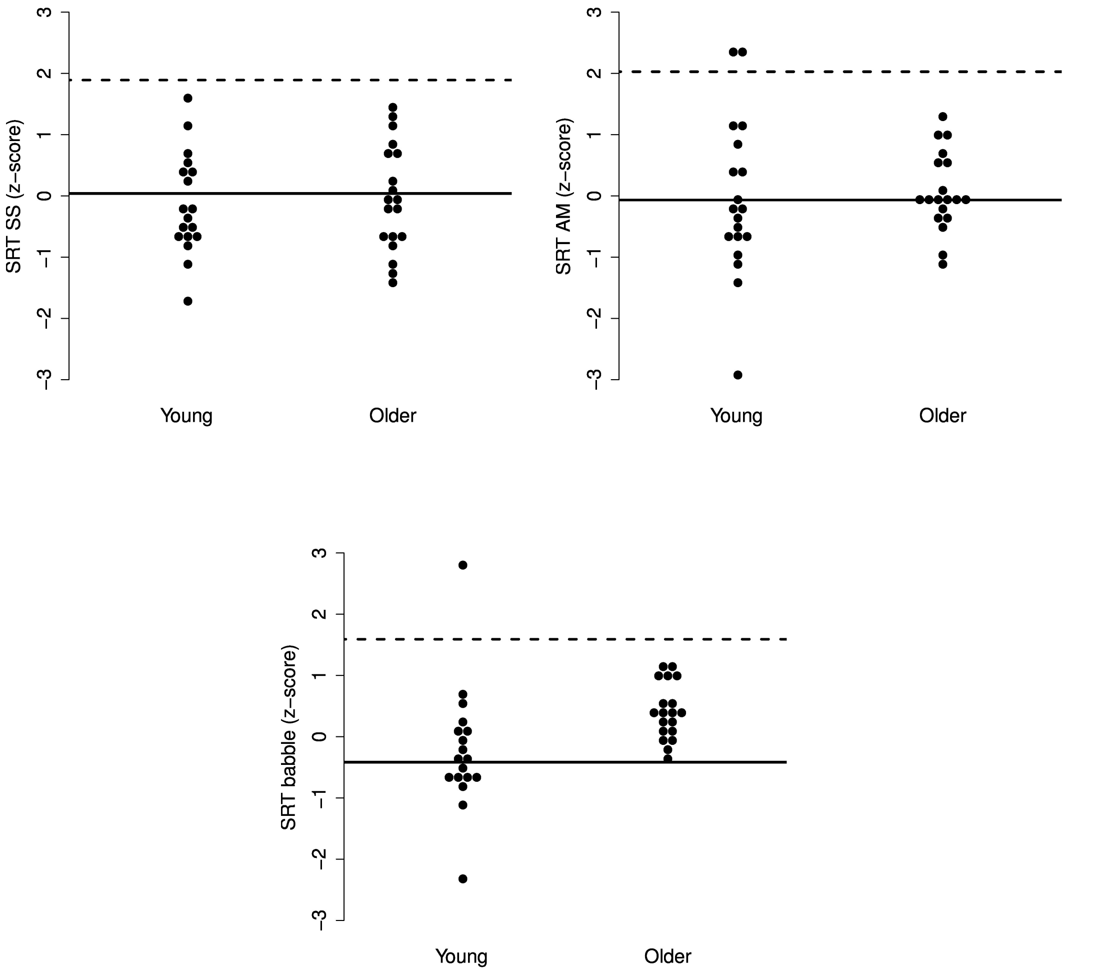


Figure 2.4: Individual z-scores for the SRTs. The solid line indicates the mean for the young adults and the dotted line indicates the deviance threshold (1.65 sd above the mean for the young adults). No deviant older adults were identified for any of the measures.

poorest 5% of a young population.

The results, illustrated in figure 2.4, indicate that none of the older adults performed particularly poorly in any of the maskers. This supports the idea that the normal-hearing older adults do not necessarily experience increased difficulties understanding speech in noise.

2.3.3 Subjective measure of speech perception in noise

While the SRT data showed some group differences (in the presence of two-talker babble only), older adults did not report increased difficulties understanding speech in noise. An independent t-test on the subjective measure of speech perception in noise (SSQ questionnaire) did not reveal a significant difference between the two age groups ($t_{(36)}$

= 1.3, p = 0.2). It should be pointed out, however, that the difference in SRTs in two-talker babble was small (1.4 dB) and that older adults did not perform more poorly in AM and SS noise compared to the young adults.

2.3.4 Temporal processing

While previous studies have reported age-related declines in auditory temporal processing (He et al., 2008; Snell, 1997; Vongpaisal and Pichora-Fuller, 2007; Füllgrabe, 2013; Pichora-Fuller and Schneider, 1992), no support for such a deficit was found in this study. AM, FM, and gap detection thresholds did not differ significantly between the young and older adults (Fig. 2.5, 2.6).

Independent t-tests on the two measures derived from the TMTF (AM efficiency and TMTF cut-off frequency) revealed no significant group differences (AM efficiency: $t_{(35)} = -0.23$, $p = 0.8$; TMTF cut-off: $t_{(35)} = -0.07$, $p = 0.9$).

These findings were supported by a mixed effects model on the AM detection thresholds with rate (10, 20, 40, 80, and 160) and group (young, old) as fixed factors and participant as a random factor. The analysis revealed a significant main effect of rate ($F_{(1,148)} = 220$, $p < 0.001$), due to the fact that the shape of the TMTF resembles a low-pass filter. However, no group or interaction effects were found (group $F_{(1,36)} = 0.4$, $p = 0.5$; interaction $F_{(1,148)} = 1.2$, $p = 0.27$), which means that the AM detection thresholds at the five different rates did not differ between the young and older adults.

Similarly, independent t-tests did not reveal significant differences between the two age groups in terms of FM and gap detection thresholds (FM $t_{(36)} = 0.6$, $p = 0.5$; gap $t_{(36)} = 0.7$, $p = 0.4$).

2.3.5 Cognitive processing

Figure 2.7 show the results for the different cognitive processing tasks. Five independent t-tests were carried out to examine the effect of age on various cognitive skills. The analyses revealed an age-related decline in working memory, as indicated by fewer correctly remembered items on the reading span task ($t_{(36)} = 4.7$, $p < 0.001$, Cohen's d = 1.5). In addition, a significant age-effect was found for processing speed, with older adults performing fewer substitutions on the letter-digit-substitution task ($t_{(36)} = 2.2$, p = 0.04, Cohen's d = 0.7). No age-effects were found for attention ($t_{(32)} = -1.3$, $p = 0.2$), TRT ($t_{(35)} = -0.6$, $p = 0.59$), or reading skills (words $t_{(36)} = 0.3$, $p = 0.8$; non-words $t_{(35)} = 0.4$, $p = 0.7$).

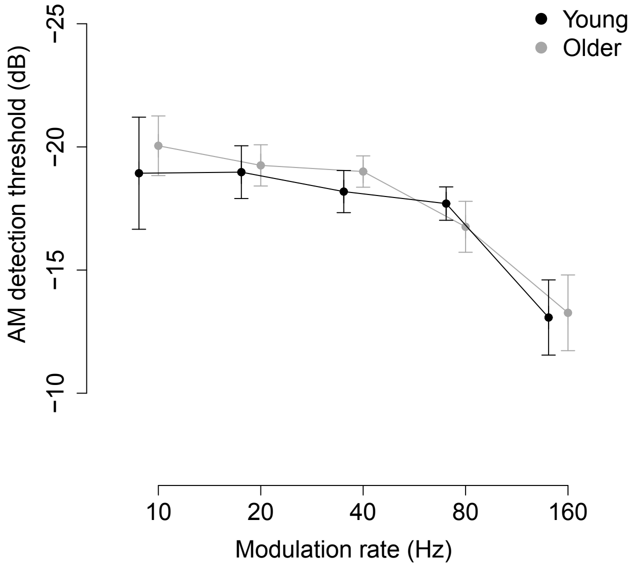


Figure 2.5: Mean AM detection thresholds (dB) as a function of modulation rate (Hz) are shown with error bars representing 95% confidence intervals.

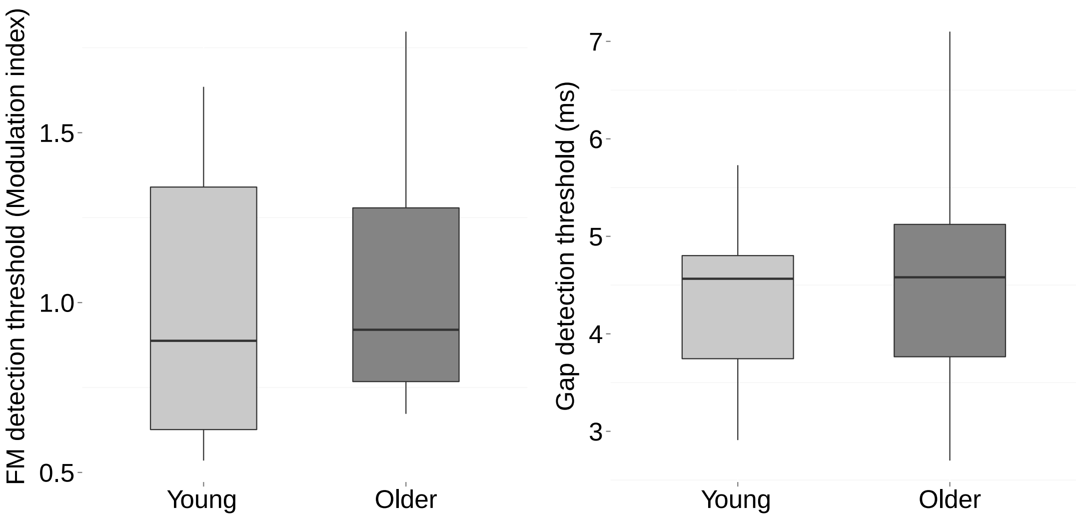


Figure 2.6: FM (left) and gap (right) detection thresholds for young (light grey) and older (dark grey) listeners.

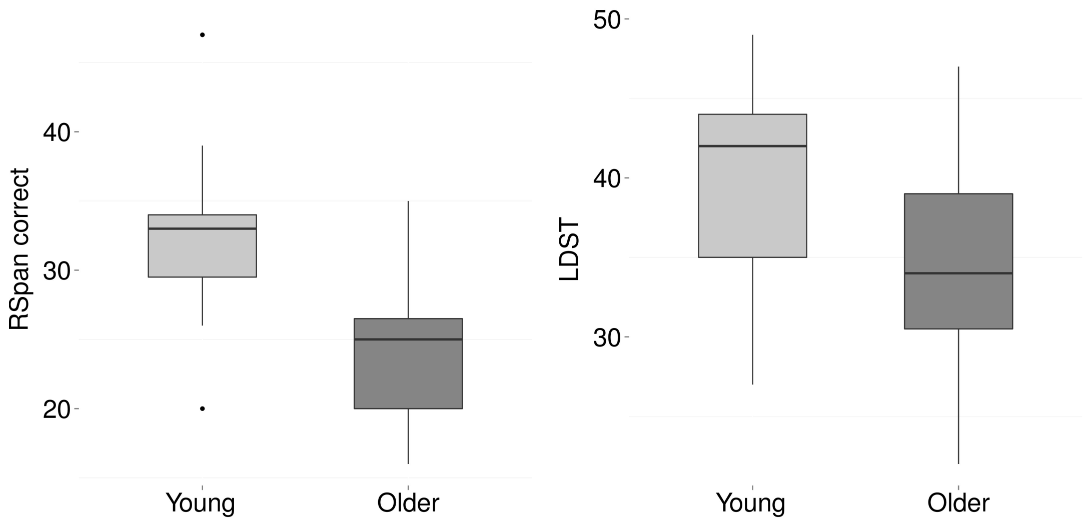


Figure 2.7: Left: Boxplots of the total number of correctly recalled words on the reading span test for young (light grey) and older (dark grey) participants. Right: Performance on the LDST task, reflecting processing speed, for young (light grey) and older (dark grey) participants. Scores are the number of correctly substituted items in 60 seconds.

2.3.6 Predicting speech perception in noise

Of primary interest was the extent to which the various auditory and cognitive measures could predict listeners' abilities to perceive speech in the three noises. The results have so far indicated age-related declines in speech perception in babble (but not SS and AM noise), working memory, and processing speed. In addition, while both groups had near-normal hearing, thresholds for the older adults were significantly higher. These findings indicate that the normal-hearing older adults had no problems understanding speech in SS and AM noise, despite some age-related cognitive declines and slightly higher audiometric thresholds. One of the questions that remains, however, is whether these age-related declines can account for the group difference in SRTs in babble.

Furthermore, the fact that the older adults only experienced increased difficulties understanding speech in two-talker babble, but not in the two noise maskers (SS and AM noise), suggests that the relative contribution of the various auditory and cognitive processes involved in the perception of speech in noise differs depending on the masker type. A question to be answered, then, is which of the auditory and cognitive measures can account for the inter-individual differences in the perception of speech in the presence of babble and noise maskers.

To determine which of the auditory or cognitive measures was predictive of speech understanding in babble and noise maskers, best subsets regression analyses were

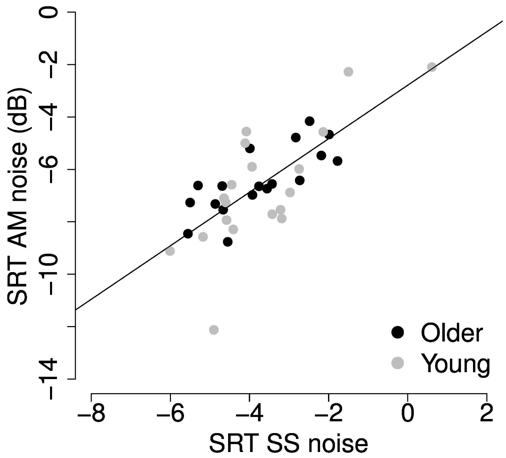


Figure 2.8: Scatter plot of SRTs in SS and AM noise reveals a strong correlation (r = 0.736, p<0.001) between performance in the two noise maskers.

conducted (Hastie et al., 2009). Since the SRTs in AM and SS noise were highly correlated ($r = 0.736$, $p<0.001$, $R^2=0.54$, figure 2.8), the regression was performed on the average of the two.

2.3.6.1 Data reduction

Due to the relatively large number of possible predictors (twelve) given our sample size (38 participants), a principal components analysis (PCA) using varimax rotation with Kaiser normalisation was performed on the cognitive and temporal processing tasks separately to reduce the number of predictors for the regression analysis. Missing data points (see section 2.3) were replaced by the mean. The resulting principal components (PC) were saved as Anderson-Rubin scores to ensure uncorrelated PC scores.

PCA on the six cognitive measures (LDST, reading span, TRT, TEA, and TOWRE words and non-words) resulted in the extraction of two components, following the Kaiser criterion (eigenvalues >1). Together they explained 63% of the variance in the data, with PC1 accounting for 34% and PC2 for 29%. The first PC was interpreted as an overall measure of linguistic closure (c.f. Zekveld et al., 2007) as it mainly reflected the TRT and the two measures of reading ability (TOWRE). The second PC primarily reflected processing speed (LDST) and working memory (reading span). Note that the measure of attention (TEA) did not group clearly with either of the two components.

An initial PCA on the four temporal processing measures (TMTF cut-off frequency, AM efficiency, FM, and gap detection thresholds) suggested the extraction of three PCs;

	PCA Cognitive processing	
	Linguistic closure	Memory and processing speed
LDST	-0.187	**0.763**
Reading span	0.068	**0.857**
TEA	0.356	-0.518
TRT	**0.635**	-0.394
TOWRE words	**0.899**	0.005
TOWRE non-words	**0.811**	-0.117

Table 2.2: Factor loadings for each of the cognitive processing measures. Factor loadings >0.4 are highlighted in bold font.

the two AM detection measures grouped together, but the FM and gap detection scores loaded significantly onto separate components (see table 2.3). Since the latter two components were dominated by a single temporal processing measure, the raw FM and gap detection thresholds were entered into the regression model instead. A subsequent PCA was performed on the two AM detection measures (TMTF cut-off frequency and AM efficiency), which resulted in the extraction of a single component that explained 66% of the variance in the AM detection data (table 2.3).

2.3.6.2 Regression

Following data reduction, the seven possible predictors that were entered into the regression models were; age group, PTA across 0.5-4 kHz, PC linguistic closure, PC memory and processing speed, PC AM detection, FM detection, and gap detection. Note that while individual differences in audiometric thresholds above 4 kHz could also have contributed to differences in SRTs, especially since the stimuli were filtered with a relatively shallow filter, a PTA across 6 - 8 kHz was not included in the regression models as a possible predictor. This is because the NAL-shaping that was applied for some older adults from 6 kHz upwards means the audiometric thresholds do not accurately reflect audibility differences in this region. Best subsets linear regressions were performed for SRTs in babble and noise (averaged across AM and SS) separately. The final models were selected based on the Bayesian Information Criterion (BIC; Schwarz, 1978).

The analyses indicated that SRTs in babble were best predicted by PTA across 0.5-4 kHz and FM detection thresholds ($R^2 = 0.32$, $F_{(2,35)} = 8.3$, $p = 0.001$; see table

	PCA Temporal processing		
PCA across all measures	PC1	PC2	PC3
FM detection	0.004	-0.068	**0.968**
Gap detection	0.011	**0.953**	-0.076
TMTF cutoff frequency	**0.797**	-0.360	-0.295
AM efficiency	**0.824**	0.332	0.264
PCA across AM measures	PC AM detection		
TMTF cutoff frequency	**0.811**		
AM efficiency	**0.811**		

Table 2.3: Factor loadings for each of the temporal processing measures (top) and for the amplitude-modulation detection measures only (bottom). Factor loadings >0.4 are highlighted in bold font.

2.4). Thus, age-related cognitive declines in working memory and processing speed did not in fact predict SRTs in babble. Instead, when audiometric thresholds were accounted for, FM detection thresholds were the primary predictor of SRTs in babble. This would imply that TFS processing in part determines speech understanding in the presence of competing talkers.

SRTs in the noise maskers, by contrast, were best predicted by a model with PTA across 0.5-4 kHz, linguistic closure, and memory and processing speed ($R^2 = 0.32$, $F_{(3,34)} = 5.48$, $p = 0.004$; see table 2.4). The fact that, after controlling for audiometric thresholds, the two cognitive measures, rather than FM detection thresholds, were significant predictors of SRTs in noise suggests that TFS processing might be less important for the perception of speech in noise maskers than in the periodic two-talker babble.

While the results from the best subsets regression analyses appear to suggest that the underlying processes accounting for individual differences in speech perception in two-talker babble and noise maskers is different, this may in fact not be the case. Even though the regression coefficients may be significant in one model but not the other, these differences in significance are in themselves not necessarily significant (Gelman and Stern, 2006). To assess whether the slopes of the predictors in the two models

		Best subsets regression on SRTs					
Dependent variable	Predictors	b	β	SE	p	R^2	CI
SRT babble	PTA 0.5-4 kHz	0.17	0.51	0.047	<**0.001*****	0.2	[0.07, 0.26]
	FM	1.57	0.35	0.64	**0.02***	0.12	[0.28, 2.9]
SRT noise	PTA 0.5 - 4 kHz	0.09	0.31	0.045	**0.045***	0.06	[0.002, 0.18]
	Linguistic closure	0.59	0.4	0.21	**0.008****	0.16	[0.16, 1.02]
	Memory and processing speed	0.5	0.34	0.22	**0.03***	0.1	[0.5, 1.0]

Table 2.4: Results of the best subsets regression analyses on SRTs in babble and the average SRTs across the two noise maskers (i.e. AM and SS noise; * significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$). Note that β refers to the standardised regression coefficient. The R^2 change reflects the proportion of the variance accounted for as predictors are added to the model. 95% confidence intervals (CI) are also provided for the regression coefficients.

were indeed significantly different, a linear regression with the four predictors that were significant in either of the two best subsets regression models (PTA 0.5-4 kHz, FM, PC linguistic closure, PC memory and processing speed) was performed on both SRTs in babble and noise separately (see table 2.5). The results of this regression model are in line with the results of the best subsets regressions, with the same predictors coming out as significant (SRT babble: $R^2 = 0.36$, $F_{(4,33)} = 4.7$, $p = 0.004$; SRT noise maskers: $R^2 = 0.37$, $F_{(4,33)} = 4.839$, $p = 0.003$). Since both models now contained the same predictors, the regression coefficients could be compared. In order to do so, a subsequent linear regression was conducted on both SRTs, with an additional dummy-coded predictor indicating the type of background noise (i.e. babble or noise maskers). The interaction between the dummy variable and the original predictors indicated whether the slopes of the predictors differed depending on the type of background noise. The results did not reveal any significant interactions (see table 2.5), suggesting that even though some measures significantly predicted SRTs in one type of background noise but not the other, the regression coefficients across the models were themselves not significantly different. In other words, there is no support for the claim that the underlying processes involved in the perception of speech in babble and noise maskers are different.

		Linear multiple regression on SRTs					
Dependent variable	Predictors	b	β	SE	p	R^2	CI
SRT babble	PTA 0.5-4 kHz	0.16	0.49	0.05	**0.003****	0.2	[0.06, 0.26]
	Linguistic closure	0.35	0.22	0.23	0.14	0.06	[-0.12, 0.82]
	Memory and processing speed	0.001	0.0008	0.26	0.99	0.02	[-0.53, 0.53]
	FM	1.46	0.32	0.7	**0.045***	0.08	[0.03, 2.9]
SRT noise	PTA 0.5 - 4 kHz	0.115	0.379	0.046	**0.02***	0.06	[0.02, 0.21]
	Linguistic closure	0.55	0.37	0.21	**0.01***	0.16	[0.12, 1.0]
	Memory and processing speed	0.65	0.44	0.24	**0.01***	0.1	[0.17, 1.1]
	FM	0.97	0.24	0.64	0.14	0.04	[-0.3, 2.3]
Interaction with SRT	PTA 0.5-4 kHz	0.023	0.1	0.034	0.5	0.005	[-0.5, 0.09]
	Linguistic closure	-0.1	-0.038	0.16	0.53	0.001	[-0.4,0.2]
	Memory and processing speed	-0.33	-0.12	0.18	0.07	0.017	[-0.7, 0.03]
	FM	0.24	0.1	0.48	0.61	0.001	[-0.7, 1.2]

Table 2.5: Top: Results of the regression analyses on the SRTs in babble and the average SRTs across the two noise maskers (i.e. AM and SS noise) with the four predictors that were significant in either of the two best subsets regression models (PTA 0.5-4 kHz, FM, PC linguistic closure, PC memory and processing speed). Bottom: Results of the regression analysis on both SRT measures with an additional dummy-coded predictor indicating the type of background noise (i.e. babble or noise maskers) assessing whether the slopes of the predictors differed depending on the type of background noise. Significant results are highlighted in bold font (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$). Note that β refers to the standardised regression coefficient. The R^2 change reflects the proportion of the variance accounted for as predictors are added to the model. 95% confidence intervals (CI) are also provided for the regression coefficients.

2.4 Discussion

The aim of this study was to assess why older adults, even in the absence of hearing impairment, typically experience increased difficulties understanding speech in noise. These difficulties are typically attributed to an age-related decline in central auditory processing, particularly in the time domain, and/or a decline in cognitive function (CHABA, 1988). This study examined the relative contribution of age-related declines in both auditory temporal and cognitive processing on the perception of speech in the presence of different noise maskers.

First, it is important to note that the data in fact suggest that older adults with fairly good hearing do not necessarily perform more poorly on a speech in noise task when ecologically valid stimuli are used. Group differences were found only in the presence of two-talker babble but not in steady-state (SS) or fluctuating (AM) noise maskers. These findings are in line with the idea that competing speech is particularly detrimental for older adults (Tun and Wingfield, 1999; Helfer and Freyman, 2008; Rajan and Cainer, 2008). The fact that the older adults performed more poorly only in the presence of two-talker babble, but not the two noise maskers, suggests that these difficulties may be due to increased susceptibility to informational masking (c.f. Freyman et al., 2004). However, it may similarly be attributable to a reduced ability to make use of periodicity cues to successfully segregate the target and masker (Vongpaisal and Pichora-Fuller, 2007; Rossi-Katz and Arehart, 2009).

Contrary to expectations, the data suggest that normal-hearing older adults do not have reduced glimpsing abilities (c.f. Gifford et al., 2007; Stuart and Phillips, 1996; Dubno et al., 2002, 2003; Peters et al., 1998; Grose et al., 2009). It should be noted, however, that the idea that older adults have impaired glimpsing abilities is perhaps somewhat controversial since age-related declines in FMB reported in the literature may in part have been the result of group differences that also became apparent in SS noise (c.f. Bernstein and Grant, 2009; Stuart and Phillips, 1996; Dubno et al., 2002, 2003).

Given that older adults are relatively unimpaired in their perception of speech in noise, could it be that the older adults are similarly unimpaired in terms of auditory temporal and cognitive processing? While an age-related decline in temporal auditory processing is well documented in normal-hearing older adults (e.g. Pichora-Fuller and Souza, 2003; Pichora-Fuller et al., 2007; Frisina and Frisina, 1997; CHABA, 1988; Gordon-Salant, 2005), this study found no decline in either E or TFS processing. However, the fact that AM detection thresholds were not different between young and older adults is likely because age effects only become apparent at higher modulation rates than those assessed in the present study (above about 200 Hz Purcell et al., 2004; Grose et al., 2009). Furthermore, the lack of an age-related increase in gap detection thresholds may be related to the temporal location of the gap. He et al. (1999) only found large age-related declines when the gap was located close to the stimulus onset or offset (at 5 or 95% of the stimulus duration), and when the gap location was random from trial to trial. Consistent with our findings, gaps in the central region of a noise burst were equally detectable by younger and older listeners, even when randomly located. Whatever the exact nature of the deficit in the older listeners found by He et al. (1999) is, it is certainly not a simple deficit in Env processing. Instead, the importance of gap uncertainty suggests a cognitive component. What is perhaps most surprising is the absence of a decline in TFS processing as this has been found using a variety of psychophysical measures (He et al., 2007; Grose and Mamo, 2010; Füllgrabe, 2013). While ageing has been shown to negatively affect frequency modulation (FM) detection using low carrier frequencies ($\leq$4 kHz) and low modulation rates ($\leq$5 Hz) (He et al., 2007), which is thought to be primarily dependent on the neural phase-locking (Moore and Sek, 1995, 1996), we did not replicate this finding.

Similarly, ageing has often been associated with declines in cognitive abilities thought to be important for the perception of speech in noise, such as working memory, attention, and processing speed (Kausler, 1994; Craik and Byrd, 1982; Salthouse, 1996). The current data indeed show declines in both working memory and processing speed. By contrast, however, attentional switching, as measured by the Visual Elevator task (Robertson et al., 1996), was not affected by age. This is somewhat surprising since this task is thought to be similar to the Wisconsin Card Sorting Test (Nelson, 1976; Robertson et al., 1996), which has repeatedly been shown to be negatively affected by age (Rhodes, 2004). Another factor thought to be important for the perception of speech in noise is linguistic closure, which was assessed by the TRT task (Zekveld et al., 2007). The literature is inconclusive as to whether linguistic closure is negatively affected by age. The results from the present study suggest that older adults do not have problems reconstructing partially masked text. This may be because linguistic closure is representative of crystallised intelligence, which does not decline with age, as opposed to fluid intelligence, which does decline with age (Horn and Cattell, 1967). It should furthermore be noted that the absence of any age-related declines in attention, linguistic closure, and perhaps even auditory temporal processing, could in part be attributed to the fact that the older adults who participated in this study were exceptional, if only in the sense that they had good hearing. It has been suggested that age-related cognitive declines are linked to hearing loss (c.f. Lin et al., 2013).

Despite these declines in working memory and processing speed, normal-hearing older adults did not have increased difficulties understanding speech in SS and AM noise. This suggests that cognitive declines associated with ageing do not inevitably lead to speech in noise problems. Furthermore, while the older adults performed worse on the speech perception task in the presence of two-talker babble, this could not be explained by age-related cognitive declines in working memory or processing speed when accounting for differences in audiometric thresholds. In fact, individual differences in SRTs in babble were best predicted by audiometric thresholds and TFS processing, as measured by the FM detection task. It should be noted, however, that since the older adults had higher audiometric thresholds, it is difficult to distinguish between an explanation based on age, and one based on hearing status. The fact that TFS processing, second to audiometric thresholds, was predictive of speech perception in the presence of competing talkers suggests that variability in performance was largely due to differences in abilities to use periodicity cues. However, whether the difficulties in the presence of babble are in fact due to a reduced ability to use periodicity cues in the masker, informational masking, or even reduced glimpsing abilities remains unclear.

While it is tempting to conclude that the underlying processes involved in the perception of speech in babble and noise maskers is different, the current study did not provide sufficient support for this idea. In fact, TFS processing may be equally important for the perception of speech in noise maskers as in the presence of competing speech. Similarly, while cognitive processing was found to be predictive of SRTs in noise maskers, they may be equally important in the presence of babble. Since the predictor coefficients across the two regression models (SRTs in babble and noise maskers) were not significantly different, no conclusions can be drawn regarding differences in underlying processes involved in speech perception in the two interferer types.

In sum, this study set out to determine the relative contribution of age-related declines in auditory temporal and cognitive processing on the perception of speech in different maskers for normal-hearing older adults. The findings can be summarised as follows:

1. Older adults experienced increased difficulties understanding speech only in the presence of two-talker babble.

2. Glimpsing abilities in 10-Hz sinusoidal amplitude modulation noise were not reduced for the normal-hearing older adults.

3. While age-related declines in temporal auditory processing are well documented for older adults, even in the absence of hearing loss, this study failed to identify a decline in either envelope or temporal fine structure processing.

4. Older adults showed cognitive declines in working memory capacity and processing speed. Despite these declines, however, speech perception in steady-state and amplitude-modulated noise was not impaired. Moreover, reduced working memory capacity and processing speed could not explain SRTs in babble beyond differences in audiometric thresholds.

Chapter 3

The role of age-related declines in subcortical auditory processing in speech perception in noise

3.1 Introduction

As discussed in chapter 2, older adults typically experience increased difficulties understanding speech in noisy environments, even in the absence of hearing impairment (Dubno et al., 2002; Helfer and Freyman, 2008). This has often been attributed to an age-related decline in auditory temporal processing (e.g. Pichora-Fuller and Souza, 2003; Pichora-Fuller et al., 2007; Frisina and Frisina, 1997; CHABA, 1988; Gordon-Salant, 2005). These temporal processing deficits, and ultimately the increased difficulties in understanding speech in noise, may be the result of age-related auditory neuropathy, or neural dysfunction (Schmiedt et al., 1996; Makary et al., 2011; Sergeyenko et al., 2013; Zeng et al., 2004), which can occur even in the absence of any permanent elevation in audiometric thresholds (Schmiedt et al., 1996; Sergeyenko et al., 2013; Kujawa and Liberman, 2009; Schuknecht and Woellner, 1953). The study described in this chapter aims to answer the question whether age-related declines in subcortical neural speech coding can in part explain why older listeners, even in the absence of hearing impairment, have more difficulties understanding speech in the presence of different types of background noise than younger listeners. The study described here formed part of the study described in chapter 2.

Auditory neuropathy may come in the form of desynchronisation of neural firing, dysfunctional synaptic transmission, and/or deafferentation (Zeng et al., 2004; Bharadwaj et al., 2014). In particular, a reduction in the number of healthy auditory nerve fibres (ANF) is thought to cause disrupted temporal coding along the auditory pathway (Bharadwaj et al., 2014; Kujawa and Liberman, 2009; Schmiedt et al., 1996; Sergeyenko et al., 2013). ANFs encode sound with great temporal precision, at least at relatively low frequencies, by phase-locking to the stimulating waveform. However, the stochastic nature of the firing of the individual ANFs means that the fidelity of temporal coding improves as multiple fibers fire in response to the auditory stimulus. Indeed, the convergence of multiple ANF inputs has been shown to lead to enhanced temporal precision of neural firing at, for instance, the cochlear nucleus (Joris et al., 1994; Oertel et al., 2000). A loss in the number of healthy ANFs will thus likely lead to reduced temporal precision of neural encoding. Moreover, loss of ANFs may in turn result in a compensatory decrease in γ-aminobutyric acid (GABA) inhibition, which has been suggested to further disrupt the temporal precision of neural coding in older adults (Anderson et al., 2012; Caspary et al., 2008; Walton et al., 1998).

Auditory neuropathy is typically assessed by measuring click-evoked auditory brainstem responses (ABR). The ABR consists of five peaks that occur within 10 ms after stimulus presentation and reflects neural firing from the auditory nerve (waves I and II) up to the inferior colliculus in the brainstem (wave V; Møller, 2007). There is little consensus, however, about how the ABR changes with age because results are often confounded by hearing loss and further complicated by differences related to stimulus parameters (for a review see Tremblay and Burkard, 2007). Often, however, decreased peak amplitudes, particularly for wave I, are reported (Sand, 1991; Burkard and Sims, 2001). Ageing has also been associated with increased wave I and V peak latencies, although support for changes in the I-V interwave interval is less consistent (Tremblay and Burkard, 2007; Konrad-Martin et al., 2012; Burkard and Sims, 2001, 2002). Taken together, these findings support the idea that ageing reduces the number and/or synchrony of ANFs and perhaps also of neurons further along the auditory pathway.

While the transient ABR may be a suitable non-invasive tool to assess neural function from the auditory nerve up to the brainstem, it is a crude measure that does not provide an insight into temporal precision of neural coding of more complex sounds such as speech. Instead, temporal neural coding may be best examined by measuring the frequency following response (FFR; Moushegian et al., 1973; Worden and Marsh, 1968), which reflects sustained synchronous neural firing at the brainstem in response to periodic auditory stimuli. The FFR has been measured, for instance, in response to pure tones (Moushegian et al., 1973), tonal sweeps (Krishnan and Parkinson, 2000), vowels (Aiken and Picton, 2008), and consonant-vowel syllables (Johnson et al., 2008). The FFR is particularly suited to assess temporal neural coding since the upper limit of phase-locking decreases as one ascends along the auditory pathway. The upper limit of phase-locking at the inferior colliculus, the primary generator of the FFR (Sohmer et al., 1977), is about 1.5 - 2 kHz (Krishnan, 2008). FFRs show remarkable fidelity to the input signal. When FFRs in response to speech sounds are converted into an audio signal, for example, they are understandable as speech, albeit lowpass filtered at about 1.5 2 kHz (Galbraith et al., 1995). Furthermore, deficits in subcortical auditory processing may only become apparent in response to more complex periodic stimuli (e.g. Song et al., 2006; Anderson et al., 2012). Anderson et al. (2012) found, for example, that older adults had normal click ABRs but degraded FFRs in response to a consonant-vowel syllable compared to younger listeners.

An additional advantage of the FFR as a measure of auditory temporal processing is that it is objective and requires no cognitive involvement from the participant. By contrast, performance on behavioural measures of auditory temporal processing is likely affected by other non-auditory factors such as motivation, attention, working memory, and fatigue. It has indeed been shown that age-effects are more pronounced for more demanding tasks, such as gap detection with uncertainty in gap location (He et al., 1999; Harris et al., 2010). Since ageing is associated with declines in cognitive functions likely to be involved in behavioural tasks of auditory temporal processing, such as working memory, attention, and processing speed (Kausler, 1994; Craik and Byrd, 1982; Salthouse, 1996), age-related declines in auditory temporal processing may be overestimated when using (more challenging) behavioural tasks. The FFR may thus provide a better, more objective, measure of auditory temporal processing.

However, it remains unclear to what extent the FFR reflects neural encoding of envelope (Env) and temporal fine structure (TFS) information in the stimulus. While it is typically assumed that it is possible to tease apart the Env and TFS information in the response (e.g. Ruggles et al., 2012; Aiken and Picton, 2008; Anderson et al., 2013), the neural Env and TFS as reflected in the FFR may differ from Env and TFS information in the stimulus (c.f. Ghitza, 2001; Shamma and Lorenzi, 2013). FFRs are often recorded to stimuli of both positive and negative polarity. Adding those responses is thought to eliminate the cochlear microphonic and any stimulus artifacts and accentuate the lower-frequency components including phase-locked activity to the envelope (Gorga et al., 1985). The opposite-polarity added FFR in fact likely reflects the recovered envelope, which is introduced through nonlinear cochlear processing (e.g. half-wave rectification at the hair cells; Skoe and Kraus, 2010; Chimento and Schreiner, 1990). However, added opposite-polarity FFRs to pure tones also show a response at double the frequency of the stimulus (Chimento and Schreiner, 1990; Sohmer et al., 1977). It has been argued, however, that for more complex stimuli at higher presentation levels this doubling of frequency components is not necessarily evident in the response (Skoe and Kraus, 2010). FFRs to positive and negative polarity stimuli are also often subtracted to accentuate higher-frequency components, typically assumed to reflect phase-locked activity to the TFS in the stimulus (Aiken and Picton, 2008). Subtracting opposite-polarity responses is thought to be similar to the compound histogram technique used in neurophysiologic studies (Aiken and Picton, 2008; Anderson et al., 1971). Subtraction is thought to remove distortions associated with half-wave rectification in the FFR (for more detail see Aiken and Picton, 2008). However, it is difficult to say with certainty to what extent the FFR components directly reflect energy in the stimulus or result from non-linear distortions along the auditory pathway. Given this uncertainty, interpretation of the FFR is difficult. Therefore, in this study responses to opposite-polarity stimuli were added to eliminate the cochlear microphonic and any potential stimulus artifacts (Gorga et al., 1985; Aiken and Picton, 2008; Skoe and Kraus, 2010). However, no conclusions will be drawn as to whether these FFRs reflect Env or TFS information in the stimulus.

Regardless of what the FFR components actually reflect, the measure does provide an insight into the effects of ageing on subcortical neural processing. In recent years, several studies have shown that FFRs for normal-hearing older adults are less robust compared to younger adults (Clinard et al., 2010; Clinard and Tremblay, 2013; Vander Werff and Burns, 2011; Anderson et al., 2012). Age-related changes in subcortical processing have been shown in response to the syllable /dɑ/ (Vander Werff and Burns, 2011; Anderson et al., 2012; Clinard and Tremblay, 2013). Both Vander Werff and Burns (2011) and Clinard and Tremblay (2013) only found group differences for peaks at the onset and offset of the response. Similarly, Anderson et al. (2012) found increased peak latencies for the older adults only for the onset and formant transition parts of the response, but not for the part of the response reflecting encoding of the steady-state vowel. However, other measures, such as the response-to-response correlation, phase-locking factor, and rms amplitude showed age-effects both for the transition and steady-state portions of the response. Similarly, age-related changes in subcortical processing have been shown in response to pure tones (Clinard et al., 2010; Clinard and Tremblay, 2013; Marmel et al., 2013). Marmel et al. (2013), for example, measured FFRs to pure tones at six different frequencies in the range of 620-720 Hz for participants with a wide range of ages (22 - 77 years old) and audiometric thresholds. They found that age was significantly correlated with FFR synchronisation strength and group delay even when accounting for individual differences in absolute thresholds. Similarly, Clinard et al. (2010) and Clinard and Tremblay (2013) measured FFRs to pure tones at six frequencies around 0.5 and 1 kHz. Based on phase-coherence and response amplitude measures they concluded that FFRs were degraded in the older adults only for stimuli around 1 kHz. They suggested that this may be the result of a reduction in the upper limit of phase locking with age. However, their results were not corrected for multiple comparisons and might not necessarily hold true after Bonferroni correction. More importantly, however, they concluded that the FFRs were degraded in the older adults only for stimuli around 1 kHz on the basis of the level of significance and did not assess the interaction between age group and stimulus frequency directly.

It remains an open question whether these age-related declines in subcortical speech encoding contribute to the increased difficulties older adults experience when understanding speech in the presence of background noise. Several studies have found a relationship between speech in noise perception and the FFR (Hornickel et al., 2009; Song et al., 2011; Anderson et al., 2011). Anderson et al. (2011), for instance, showed that older adults with more robust FFRs, as indicated by a larger spectral magnitude at F0 and a larger rms amplitude of responses both in quiet and noise, performed better on a speech in noise task than older adults with less robust FFRs. Similarly, using structural equation modeling, Anderson et al. (2013) found that both the FFR and cognitive function predicted a large part of the variance in older adults' abilities to perceive speech in noise.

While within-group differences in speech in noise performance may be attributed in part to differences in subcortical auditory processing, it remains unclear whether the FFR can similarly predict differences in speech in noise performance between young and older listeners. It should be noted that a recent study by Anderson et al. (2012) did not convincingly show an age-related degradation of the spectral component at F0, in the FFR in response to /dɑ/, thought to be a particularly important indicator of speech in noise performance (Anderson et al., 2011; Song et al., 2011). While the authors argued that ageing did indeed affect the neural representation at F0, this result was not significant either during the formant transition period or the steady-state vowel period. The group difference only approached significance when the formant transition and steady-state portion were taken together. This might suggest that the increased difficulties in the perception of speech in noise experienced by older adults may not necessarily be the result of disruptions in subcortical neural speech encoding.

Another question that remains open is whether the FFR can predict the benefit a listener derives when perceiving speech in the presence of a fluctuating compared to a steady-state masker. When a masker fluctuates in amplitude over time, it can be expected that the degrading effect of the noise on the FFR will also vary over time. The degree of 'neural release from masking' at the troughs of the fluctuating masker likely depends on the rate of modulation of the masker as well as the neural recovery rate from prior stimulation. At high modulation rates the duration of the trough of the masker may not be long enough for the neurons to fully recover from adaptation effects induced by the peak of the masker. Similarly, listeners with a long recovery time from neural adaptation may not get as much of a neural release from masking at the trough of the masker. This idea is supported by the fact that peak latencies in click ABRs increase with decreasing forward-masking intervals (Lasky and Rupert, 1982; Kramer and Teas, 1982).

It has indeed been shown that normal-hearing older adults show prolonged neural recovery times from forward masking, albeit only for relatively short forward-masking intervals (Walton et al., 1999; Poth et al., 2001; Fujikawa and Weber, 1977; Debruyne, 1986). Walton et al. (1999), for instance, measured click ABRs to 1, 4, and 8 kHz tonebursts in the presence of forward maskers (tonebursts of the same frequency) with different forward-masking intervals. They found increased wave V latencies for older adults using the 4 and 8 kHz tone-bursts for forward-masking intervals of 16 ms or less. Similarly, wave V peak latency shifts in normal-hearing older adults have been found for click ABRs with increasing presentation rate (c.f. Fujikawa and Weber, 1977; Debruyne, 1986, but see Burkard and Sims, 2001). Furthermore, Boettcher et al. (1996) measured ABRs in response to noise bursts separated by a silent gap in young and older gerbils and found increased wave IV latencies for the aged gerbils only when the gap was short. These findings are in line with the idea that Env processing in older adults is only impaired at higher modulation rates (e.g. Purcell et al., 2004). While prolonged neural recovery times are therefore unlikely to contribute to speech perception difficulties in fluctuating noise with relatively slow modulation rates (e.g. 10 Hz; Gifford et al., 2007;

Dubno et al., 2002, 2003), it remains an open question to what extent the FFR is affected by forward masking and how this relates to listeners' abilities to listen in the dips of fluctuating maskers. It likely is a better indicator of the role of Env processing on dip listening than behavioural measures such as gap and AM detection due to the absence of cognitive involvement in the FFR measure.

The primary aim of this study was to determine the role of age-related declines in subcortical auditory processing in the perception of speech in different types of background noise. Speech perception abilities were assessed in steady-state and amplitude-modulated speech-shaped noise, as well as two-talker babble. In addition, click ABRs and FFRs in response to a vowel /ɑ/ in quiet, steady-state, and amplitude-modulated speech-shaped noise were measured. A steady-state stimulus was used to allow for a comparison of the FFR at the peak and trough of the amplitude-modulated masker.

FFRs were expected to be less robust, particularly in the presence of background noise, for the older adults. This anticipated decline in subcortical auditory processing was expected to partly account for increased speech perception in noise difficulties associated with ageing. Furthermore, fluctuations in the masker were expected to become apparent in the FFR, although individual differences in neural release from masking might not necessarily account for differences in dip listening skills between young and older listeners.

3.2 Methods

3.2.1 Participants

Nineteen young (19 - 29 years old, mean 23.7 yrs, sd 2.9 yrs, ten males) and 19 older (60 - 72 years old, mean 64.1 yrs, sd 3.3 yrs, three males) monolingual native English speakers participated in this study. All participants had near-normal hearing defined as (air-conducted) pure-tone thresholds of 25 dB HL or better at octave frequencies from 0.25 to 4 kHz in both ears and at 6 kHz in at least one ear (Fig. 2.2). In addition, all participants over the age of 65 had normal cognitive function (scores $\geq$17 MMSE telephone version (Roccaforte et al., 1992)). None of the participants reported a history of language or neurological disorders. Participants signed a consent form approved by UCL Research Ethics Committee and were paid for their participation.

3.2.2 Speech perception in noise

Speech reception thresholds (SRT) were measured for sentences in different types of background noise. The target stimuli were pre-recorded IEEE sentences (Rothauser et al., 1969) produced by a male talker of standard Southern British English. The sentences were presented in steady-state speech-shaped noise (SS), speech-shaped noise sinusoidally amplitude-modulated at 10 Hz with a modulation depth of 100% (AM), and two-talker babble (see Rosen et al. (2013) for a description of the speech-shaped noise and two-talker babble).

To rule out possible contributions of decreased audiometric thresholds above 6 kHz, the stimuli were low-pass filtered at 6 kHz using a fourth order Butterworth filter. In addition, for six older participants with thresholds > 25 dB HL at 6 kHz in one ear, the stimuli in the relevant ear were spectrally shaped using the National Acoustics Laboratories-Revised (NAL-R) linear prescriptive formula based on their individual thresholds (Byrne and Dillon, 1986).

The SNR was varied adaptively following the procedure described by Plomp and Mimpen (1979), and SRTs were tracked at 50%. Stimuli were presented binaurally at 70 dB SPL. The procedure is described in more detail in chapter 2.

3.2.3 Electrophysiological measures

3.2.3.1 Stimuli

Click ABRs were recorded in response to 2000 sweeps of a 100 μs click with alternating polarity presented monaurally (left and right separately). Stimuli were presented at 70 dB nHL (107.6 dB peSPL) with a repetition rate of 11/s (Campbell et al., 1981). To confirm the reliability of the measures, two click ABRs were measured for each ear at the start of the session.

FFRs were recorded in response to a synthetic vowel /ɑ/, which was created in MATLAB (Mathworks, Natick, MA). The vowel had a fundamental frequency (F0) of 160 Hz (F1: 710, F2: 1200, F3: 2900, F4: 3400 Hz) and a duration of 100 ms. The vowel was tapered on and off across 6.25 ms, which corresponds to one cycle of the F0. FFRs were recorded in response to 3000 sweeps of positive and negative polarities separately. A concern may be that the stimuli were not alternated throughout a run. This could potentially affect the averaged response if there was a between-run change in, for example, the electrode offsets. Stimuli were presented binaurally at 80 dB SPL with a repetition rate of 5/s, corresponding to an interstimulus interval of 100 ms. To minimise contamination by stimulus artifact and the cochlear microphonic, the averaged responses to positive and negative polarities were later added together (Gorga et al., 1985).

FFRs were measured to the vowel in quiet, steady-state speech-shaped noise (SS), and amplitude-modulated speech-shaped noise (AM). In the AM condition, the speech-shaped noise was sinusoidally modulated at 10 Hz with a modulation depth of 100%. The SS and AM maskers were identical to those used in the SRT task described above. The noise was presented continuously for the duration of the condition and had a rise and fall time of 100 ms. The string of stimuli started playing 225 ms after the start of the masker. Within each sweep, stimulus onset was 50 ms after the trigger pulse, which meant that in the AM condition the prestimulus period was always centred at the trough of the AM masker (Fig. 3.1). The SNR was fixed at 7 dB for the SS condition. The level of the AM masker was scaled to give an SNR of 7 dB across a 37.5 ms window centred at the peak of the masker, which corresponds to an SNR of 9.3 dB across the total duration of the AM masker (Fig. 3.2).

FFRs in response to each condition were measured twice. The order of conditions (i.e. quiet, SS, AM) was counterbalanced across participants following a Latin square design.

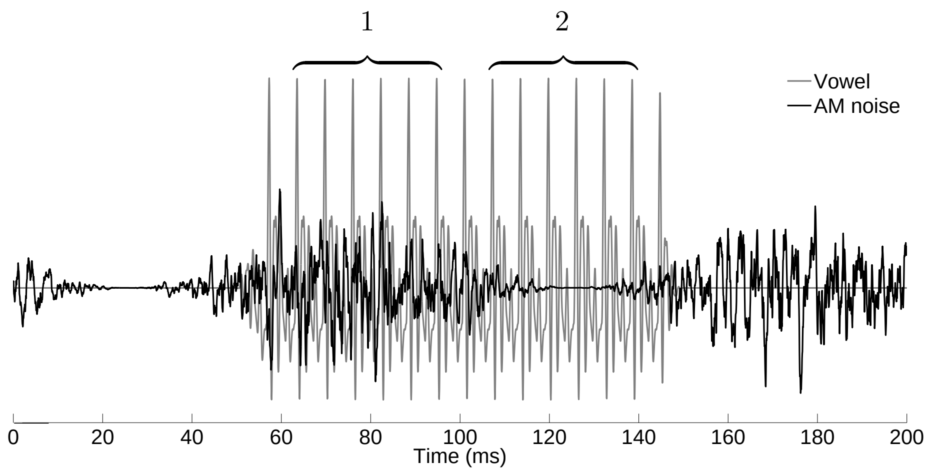


Figure 3.1: The exact position of the vowel /ɑ/, plotted in grey, with respect to the amplitude modulations of the amplitude-modulated speech-shaped noise (AM), plotted in black. The vowel starts at 50 ms and spans a whole AM cycle. Note that the 50 ms prestimulus period is centred at the trough of the AM. The two analysis windows (section 3.2.3.4), indicated by curly brackets, have a duration of 37.5 ms (i.e. six F0 cycles). The first window is centred at the peak of the AM and the second at the trough of the AM. Since the vowel is tapered on and off across 6.25 ms (i.e. one F0 cycle), the first analysis window starts 6.25 ms after stimulus onset and the second analysis window ends 6.25 ms before stimulus offset.

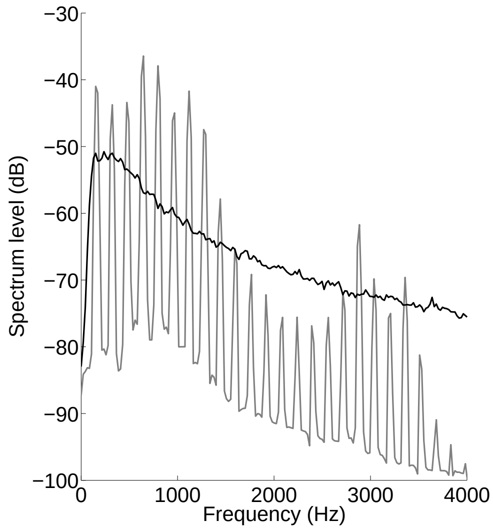


Figure 3.2: Power spectra of the vowel /ɑ/ (grey) and the two maskers (black). The spectral magnitude of the AM is equal to that of the SS when measured across the 37.5 ms window centred at the peak of the AM.

3.2.3.2 Recording parameters

Participants were seated in a reclining chair in an electrically shielded sound proof booth. To promote stillness, participants were asked to close their eyes and told they were allowed to fall asleep. People often encourage their participants to fall asleep to reduce myogenic artifacts (e.g. Aiken and Picton, 2008; Gockel et al., 2011). It is generally believed that the ABR and FFR are largely unaffected by sleep or effects of attention (c.f. Chandrasekaran and Kraus, 2010).

Electrophysiological responses were collected using a BioSemi ActiveTwo system (Amsterdam, The Netherlands). Click ABRs were recorded differentially between Cz and the ipsilateral earlobe. FFRs were collected differentially between Cz and the seventh cervical vertebra (C7). Two additional electrodes, Common Mode Sense (CMS) and Driven Right Leg (DRL), were placed near Pz. In the BioSemi ActiveTwo system these two electrodes replace the ground electrode. Electrode offsets were always <40 mV. Responses were recorded with a sampling rate of 16384 Hz.

Stimuli were generated in MATLAB (Mathworks, Natick, MA). The MATLAB procedure created a string of the required number of stimuli on one channel and a string of triggers on another channel. Both channels were delivered via a computer using an external soundcard (RME FireFace UC, 44.1 kHz) connected to a custom-made trigger box which separated the two channels and simultaneously sent the trigger to the BioSemi machine and the stimulus to electrically shielded ER-3 inserts (Intelligent

Hearing Systems, Miami, FL). This procedure minimised jitter in the trigger times relative to stimulus presentation. If the presentation of the stimulus and the recording of the response are not precisely time locked, and the FFR is subject to even a small amount of jitter, the FFR is degraded when trials are averaged.

3.2.3.3 Preprocessing

The click ABRs were filtered from 100 to 3000 Hz (two second order Butterworth filters, going forwards and backwards, therefore zero phase-shift) and epoched from -14 to 14 ms. FFRs were filtered from 70 to 2000 Hz (two second order Butterworth filters, going forwards and backwards, therefore zero phase-shift) and epoched from -49 to 149 ms. Baseline correction was performed with respect to the prestimulus response. Any sweeps exceeding ± 25 μV were rejected.

When stimulus-to-response correlations of the FFRs (section 3.2.3.4) showed comparable delays for both runs, responses were summed across the two runs. The resulting averages contained, on average, 10408 sweeps (sd 1799). The large variation in the total number of sweeps is mainly due to the fact that responses could not always be summed across the two runs because of missing or noisy data.

3.2.3.4 FFR analysis

Analyses were performed on the whole FFR as well as across two shorter analysis windows. These two shorter analysis windows were of particular interest for the AM condition to assess encoding of the vowel at the peak and trough of the masker. The windows had a duration of 37.5 ms and encompassed exactly six F0 cycles of the vowel (Fig. 3.1). Given that the stimulus was tapered on and off across 6.25 ms, the response to the first and last cycles of the vowel was not taken into account in the analyses.

The onset of the FFR was determined objectively by correlating the stimulus with the response. The stimulus waveform was first band-pass filtered at 70 - 2000 Hz and resampled at 16384 Hz to match the response waveform. Correlation coefficients were determined by shifting the response relative to the stimulus by 3 - 10 ms to find the maximum correlation within this time window. The FFR onset was determined for each individual response and used to determine the time window across which to compute subsequent analyses.

The SNR of the response was determined by dividing the root mean square (rms) amplitude of the response to the vowel by the rms of the prestimulus response. Note that the prestimulus response in the AM condition is centered at the peak of the AM cycle (Fig. 3.1).

Assessing the effect of background noise: Several response properties, computed across the entire FFR, were compared across the three conditions (quiet, SS, AM). Spectral amplitudes were calculated using a fast Fourier transform across 20-Hz wide bins centred at F0 (160 Hz), H2 (320 Hz) and H3 (480 Hz) and taking the peak amplitude within the respective bins. Zero-padding to the sampling rate was applied symmetrically around the response to increase the spectral resolution to 1 Hz. Values reported here are peak amplitudes of the power spectrum, in dB. The rms amplitude of the FFR was computed as an indication of the magnitude of the response. The encoding of pitch information was quantified using an autocorrelation function across a 40-Hz wide analysis window centred at the F0 of the stimulus (160 Hz). The height of the first peak in the autocorrelation function provided a measure of pitch strength (c.f. Krishnan et al., 2005). Stimulus-to-response lags and correlations were computed to provide an overall measure of the robustness of encoding. Cross-correlation coefficients were computed across responses of different conditions to assess the effect of the noise on the robustness of encoding of the vowel. Cross-correlations were computed for quiet to SS, quiet to AM, and AM to SS. Responses were shifted relative to one another across -6 to +6 ms. Similarly, cross-correlations were calculated for responses of the same condition across the two different runs. A Fisher transformation was used to convert the correlation coefficients (r values) to z scores for statistical analyses.

Assessing the effect of amplitude modulations in the masker: To examine the effect of the modulations in the AM masker, spectral amplitudes at F0 (160 Hz), H2 (320 Hz) and H3 (480 Hz), pitch strength, and rms amplitude were compared across the two analysis windows.

3.3 Results

Except where stated otherwise, in the analyses reported below outliers were excluded if they exceeded the mean $\pm$ 3 sd. In total, 19 data points spread across several FFR measures were excluded (FFR quiet: one young and one older adult for stimulus-to-response lag, F0, H2, H3, SNR; FFR SS: one young adult for H3, four young and two older adults for response-to-response correlation; FFR AM: one young and one older adult for response-to-response correlation).

3.3.1 Speech perception in noise

The results of the speech in noise task are displayed in figure 2.3. The figure shows SRTs for both young and older groups in SS noise, AM noise, and two-talker babble.

Older adults were expected to have higher (i.e. worse) SRTs in the presence of all three maskers (AM, SS, two-talker babble). As discussed in chapter 2, however, the older adults only performed worse in the presence of two-talker babble. A mixed effects model with condition (AM, SS, babble) and group (young, old) as fixed factors and participant and sentence list as random factors showed a significant interaction between group and condition ($F_{(2,186)} = 5.6$, $p = 0.004$). Post-hoc independent t-tests revealed a significant difference between the two age groups for SRTs in babble only, with young listeners performing on average 1.4 dB better (i.e. lower SRTs) than older listeners ($t_{(36)} = 2.8$, $p = 0.008$, Cohen's $d = 0.9$; all other $p > 0.6$).

Figure 2.3 illustrates that both groups showed a dip listening effect as indicated by lower (i.e. better) SRTs in AM than SS noise. In addition, SRTs in babble were higher (i.e. worse) compared to the two noise maskers. These findings were supported by independent t-tests. SRTs in AM were on average 2.7 dB lower than in SS noise ($t_{(37)} = 12.9$, $p < 0.001$, Cohen's $d = 1.4$), indicative of dip listening. In addition, SRTs in babble were significantly higher than in SS and AM noise by 2.6 and 5.3 dB respectively (SS: $t_{(37)} = 8.5$, $p < 0.001$, Cohen's $d = 2.5$; AM: $t_{(37)} = 16.3$, $p < 0.001$, Cohen's $d = 1.4$).

3.3.2 Subcortical auditory processing

3.3.2.1 Click ABRs

To assess whether any age-related changes in subcortical auditory processing were evident in the standard click ABR, wave V peak latencies and amplitudes of the response were compared between young and older adults. Two mixed effects models with group (young, old) as a fixed factor and participant as a random factor were conducted. Given the difference in the sex balance between the two age groups (ten young compared to only three older males), models with sex as an additional random factor were also considered. However, model comparisons based on the Akaike Information Criterion (AIC; Akaike, 1974) indicated that the models without sex were a better fit for the data.

The results indicated that the responses were not significantly affected by age (latency $F_{(1,33)} = 0.3$, $p = 0.6$; amplitude $F_{(1,34)} = 1.9$, $p = 0.2$). It should be noted that responses for one young (one ear) and three older adults (both ears for one of the three older adults) were abnormal, defined as having latencies larger than three standard deviations above the mean for the group of young adults (> 6.54 ms, c.f. Campbell et al., 1981). These participants were not excluded from the analyses.

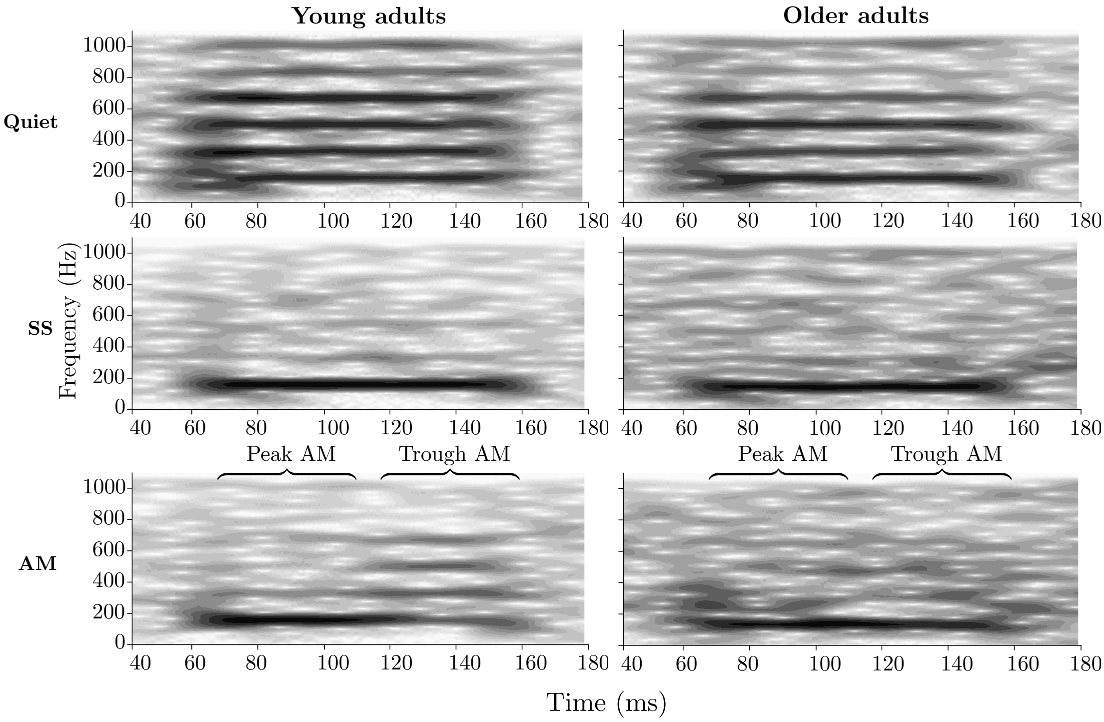


Figure 3.3: Spectrograms of grand averaged FFRs in quiet (top), SS noise (middle), and AM noise (bottom) for young (left) and older (right) listeners.

3.3.2.2 FFRs

The possibility of age-related changes in sustained neural phase-locking to auditory stimuli at the brainstem level was examined by comparing response measures of the FFR (see section 3.2.3.4) for young and older adults. Figure 3.3 shows spectrograms of the grand averaged responses for the two age groups in quiet, AM, and SS noise.

Data from one older participant was excluded from the analyses due to a stimulus artifact in the recordings. While responses with an SNR $\leq$ 1.5 dB are typically excluded from analyses (c.f. Skoe and Kraus, 2010) this rule of thumb was not applied here since the distribution of SNRs was not spread equally across the different groups and conditions. Results from a mixed effects model with condition (quiet, SS noise, and AM noise) and group (young, old) as fixed factors and participant as random factor revealed a significant effect of condition ($F_{(2,68)} = 13$, $p < 0.001$) and group ($F_{(1,35)}$ =5.9, $p = 0.02$) but no interaction ($F_{(2,68)} = 0.3$, $p = 0.7$). Closer examination revealed that SNRs were higher in young compared to older adults, and a planned contrast (quiet vs AM and SS) showed that the SNR in both SS and AM noise was significantly lower than in quiet ($t_{(68)} = 5.1$, $p < 0.001$).

Effects of noise on FFRs: To assess the effects of noise on the FFR, several mixed effects models with condition (quiet, SS noise, and AM noise) and group (young, old) as fixed factors and participant as random factor were conducted. Model comparisons based on the Akaike Information Criterion (AIC; Akaike, 1974) indicated that the models without sex were a better fit for the data. It should be noted that the results were not corrected for multiple comparisons because the comparisons were planned.

Overall, the results show that FFRs are degraded in the presence of background noise (both AM and SS). Furthermore, older adults had less robust responses compared to young adults. However, the FFRs of the older adults were typically not significantly more affected by background noise (i.e. no interactions; except in terms of stimulus-to-response correlations). The results of the mixed effects models are summarised in table 3.1.

The analyses indicated an interaction (condition × group) only in terms of stimulus-to-response correlations. Post-hoc independent t-tests revealed a significant difference between the two age groups for AM and SS noise (AM: $t_{(35)} = 2.5$, $p = 0.02$, Cohen's $d = 0.55$; SS: $t_{(30)} = 2.9$, $p = 0.006$) but not for quiet ($t_{(33)} = -0.3$, $p = 0.8$, Cohen's $d = 0.65$). This would suggest ageing primarily affects the robustness of subcortical encoding of speech in noise (and not quiet). It should be noted, however, that, of the eight measures assessed, an interaction between condition and group was only found in terms of the stimulus-to-response correlation.

The four measures that show a significant main effect of group mostly follow the same pattern with less robust FFRs for the older adults. Figure 3.4 illustrates that the FFRs of older adults are characterised by lower stimulus-to-response correlations, lower spectral magnitude at the second and third harmonic, lower response magnitude, and lower response-to-response correlations. By contrast, stimulus-to-response lags were shorter for the older adults. However, this may simply be because the older group consisted of more women, who tend to have earlier peak latencies than men (Krizman et al., 2012; Jerger and Johnson, 1988).

Planned contrasts were carried out to further examine the effects of condition on the FFR. Two orthogonal contrasts were defined to examine (a) the overall effect of background noise on the FFR (quiet vs. AM and SS) and (b) the effect of amplitude modulation in the masker on the FFR more specifically (AM vs. SS).

The results of the planned contrasts are shown in table 3.2. While the results overall indicate that FFRs are degraded in noise (quiet vs. AM and SS), there does not

	Mixed-effects models		
Measure	Condition	Group	Interaction
Stimulus-to-response	$F_{(2,68)} = 5.7$	$F_{(1,35)} = 5.7$	$F_{(2,68)} = 6.3$
correlation	$\mathbf{p = 0.005}^{**}$	$\mathbf{p = 0.02}^{*}$	$\mathbf{p = 0.003}^{**}$
Stimulus-to-response	$F_{(2,68)} = 50$	$F_{(1,35)} = 0.2$	$F_{(2,68)} = 2.7$
lag	$\mathbf{p< 0.001}^{***}$	$p = 0.7$	$p = 0.07$
Spectral magnitude	$F_{(2,68)} = 1.8$	$F_{(1,35)} = 3.2$	$F_{(2,68)} = 0.7$
at F0	$p = 0.2$	$p = 0.08$	$p = 0.5$
Spectral magnitude	$F_{(2,68)} = 66$	$F_{(1,35)} = 14.5$	$F_{(2,68)} = 2.3$
at 2^{nd} harmonic	$\mathbf{p< 0.001}^{***}$	$\mathbf{p< 0.001}^{***}$	$p = 0.1$
Spectral magnitude	$F_{(2,67)} = 94$	$F_{(1,35)} = 0.3$	$F_{(2,67)} = 1.9$
at 3^{rd} harmonic	$\mathbf{p< 0.001}^{***}$	$p = 0.6$	$p = 0.2$
F0 strength	$F_{(2,68)} = 1.8$	$F_{(1,35)} = 16.7$	$F_{(2,68)} = 1$
	$p = 0.2$	$\mathbf{p< 0.001}^{***}$	$p = 0.4$
rms amplitude	$F_{(2,68)} = 13$	$F_{(1,35)} = 1.28$	$F_{(2,68)} = 0.4$
	$\mathbf{p< 0.001}^{***}$	$p = 0.3$	$p = 0.7$
Response-to-response	$F_{(2,64)} = 7.6$	$F_{(1,35)} = 11.7$	$F_{(2,64)} = 0.1$
correlation	$\mathbf{p = 0.001}^{***}$	$\mathbf{p = 0.001}^{***}$	$p = 0.9$

Table 3.1: Main effects of condition and group as well as interactions for the different FFR measures with significant effects highlighted in bold font (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$).

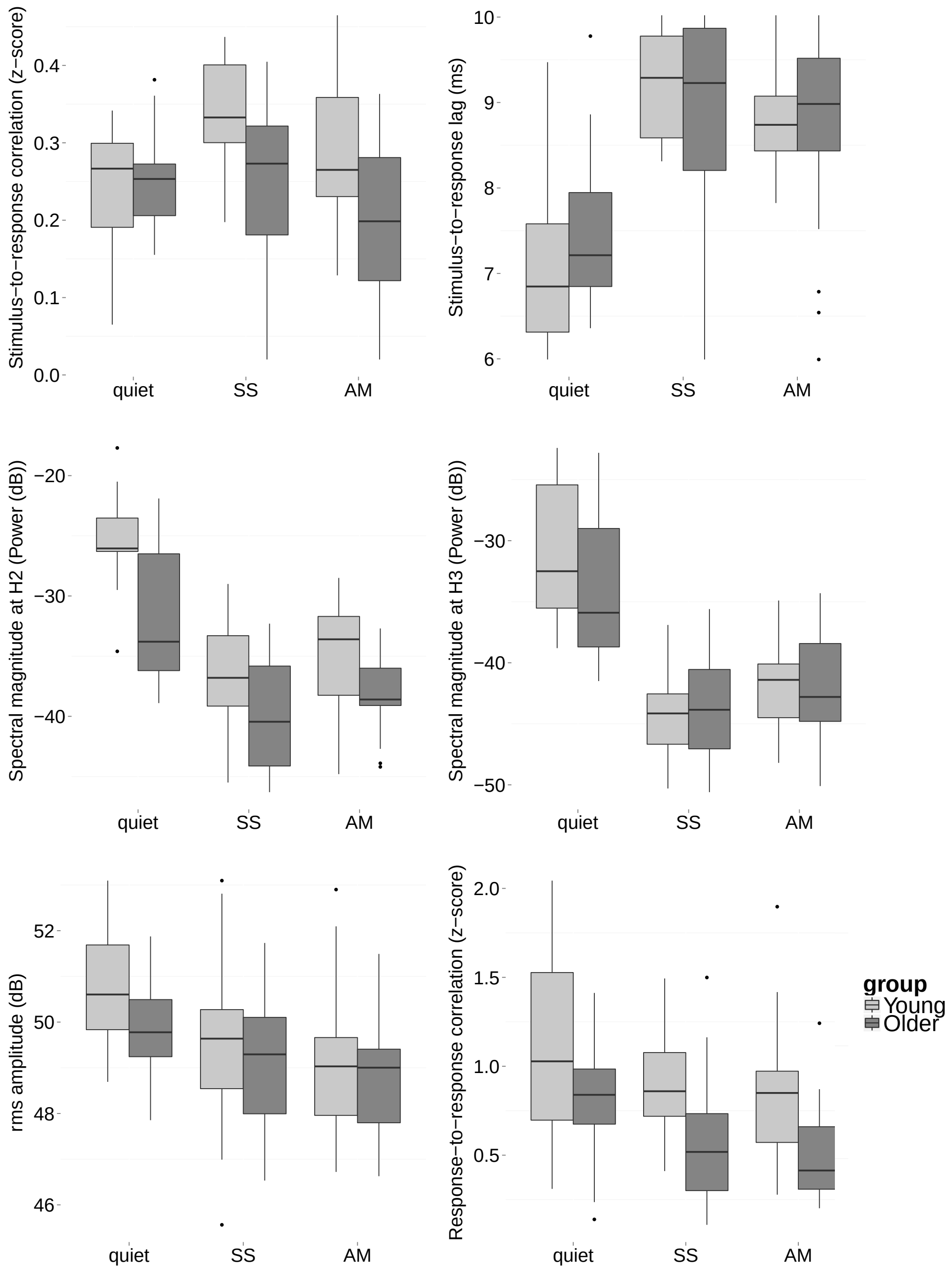


Figure 3.4: Boxplots of response properties of the FFR are shown for young (light grey) and older (dark grey) participants in the three different conditions (quiet, SS noise, and AM noise). The measures plotted here are the stimulus-to-response correlation (Fisher-transformed Pearson's correlations; i.e. z-scores) top-left, stimulus-to-response lag (ms) top-right, spectral magnitude at the second and third harmonics (Power (dB)) in the mid-left and mid-right respectively, rms amplitude (dB) bottom-left, and response-to-response correlations (Fisher-transformed Pearson's correlations; i.e. z-scores) in the bottom-right.

	Planned contrasts	
Measure	Quiet vs. noise	AM vs. SS noise
Stimulus-to-response lag	$t_{(68)}$ = -9.9 **p < 0.001*****	$t_{(68)}$ = 1.1, p = 0.28
Spectral magnitude at 2^{nd} harmonic	$t_{(68)}$ = 11.3, **p < 0.001*****	$t_{(68)}$ = -1.8, p = 0.07
Spectral magnitude at 3^{rd} harmonic	$t_{(67)}$ = 13.5, **p < 0.001*****	$t_{(67)}$ = -2.1, **p = 0.04***
rms amplitude	$t_{(68)}$ =5.0, **p <0.001****	$t_{(68)}$ = 0.89, p = 0.38
Response-to-response correlation	$t_{(64)}$ = 3.8, **p < 0.001*****	$t_{(64)}$ =0.7, p = 0.5

Table 3.2: Results from post-hoc t-tests for the FFR measures that showed a main effect of condition. Significant effects are highlighted in bold font (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$).

tend to be a difference between FFRs in AM and SS noise.

Five of six measures that showed a main effect of condition (all except the stimulus-to-response correlation), showed less robust FFRs in both AM and SS noise compared to quiet. The degrading effect of noise can be characterised by larger stimulus-to-response lags, lower spectral magnitudes at the second and third harmonic, lower rms amplitudes, and lower response-to-response correlations.

No clear trend can be observed in terms of differences between FFRs in AM and SS noise. While stimulus-to-response correlations were lower (i.e. worse) in AM noise, spectral magnitude of the third harmonic was lower (i.e. worse) for SS noise. Moreover, no significant differences between the two noise conditions were found in terms of the stimulus-to-response lag, spectral magnitude of the second harmonic, rms amplitude, or response-to-response correlation.

Effects of amplitude fluctuations in the noise on the FFR: While the analyses described above did not reveal any differences between FFRs in SS and AM noise, this does not necessarily mean that amplitude fluctuations of the masker have no effect on the FFR. To examine the effects of amplitude fluctuations in the noise on the robustness of subcortical speech encoding, response measures of the FFR were compared across two analysis windows, one centred at the peak and one at the trough of the fluctuating masker. Mixed-effects models with window (peak, trough) and group (young, old) as fixed factors and participant as random factor were carried out for four response measures. Model comparisons based on the Akaike Information Criterion (AIC; Akaike, 1974) again indicated that models without sex were a better fit for the data. The results are summarised in table 3.3. Results for the main effect of group are not reported here since only group effects in relative differences in the FFR measures between the peak and trough of the AM noise (i.e. interactions) are of interest here. Again, the results were not corrected for multiple comparisons because the comparisons were planned.

The analyses indicate an overall effect of window, although the pattern of the effect is inconsistent. It would be expected that the response properties are more robust at the trough than at the peak of the masker, similar to the way the FFR is more robust in quiet than in noise. However, while this is true for the spectral magnitudes of the second and third harmonics, which are higher (i.e. better) at the trough of the masker, the spectral magnitude of the fundamental and the rms amplitude are in fact higher (i.e. better) at the peak of the masker. In addition, the FFR did not differ significantly across the two analysis windows in terms of pitch strength.

The fact that there are no interactions for the spectral magnitude of the second and third harmonic suggests that both young and older listeners show a 'neural release from masking' characterised by larger spectral magnitudes at the trough than at the peak of the AM noise. Closer examination of the interactions found for the spectral magnitude at F0 and the rms amplitude indicates that these effects are mainly driven by the young adults. Post-hoc paired t-tests revealed significant differences between the peak and trough of the masker only for the young participants (F0 young: $t_{(18)} = 5.3$, $p < 0.001$, Cohen's $d = 0.9$; old: $t_{(17)} = -0.5$, $p = 0.6$; rms young: $t_{(18)} = 4.6$, $p < 0.001$, Cohen's $d = 0.7$; old: $t_{(17)} = 0.4$, $p = 0.7$). Interestingly, these are precisely the two measures that show a pattern that would not be expected if the FFR was more robustly encoded at the trough than at the peak of the masker.

Data reduction: The analyses above were performed on a large number of FFR measures. However, given that the underlying processes responsible for some of these measures are not clearly understood (c.f. Gockel et al., 2011), the interpretation of the results is not necessarily straightforward. In addition, due to the plethora of possible measures it is often difficult to choose the relevant ones. Interpretation may be further complicated when some of the results contradict each other, which is inevitable when working with large numbers of outcome measures. It is probably more insightful to look at an overall measure of subcortical speech encoding. This overall FFR measure was

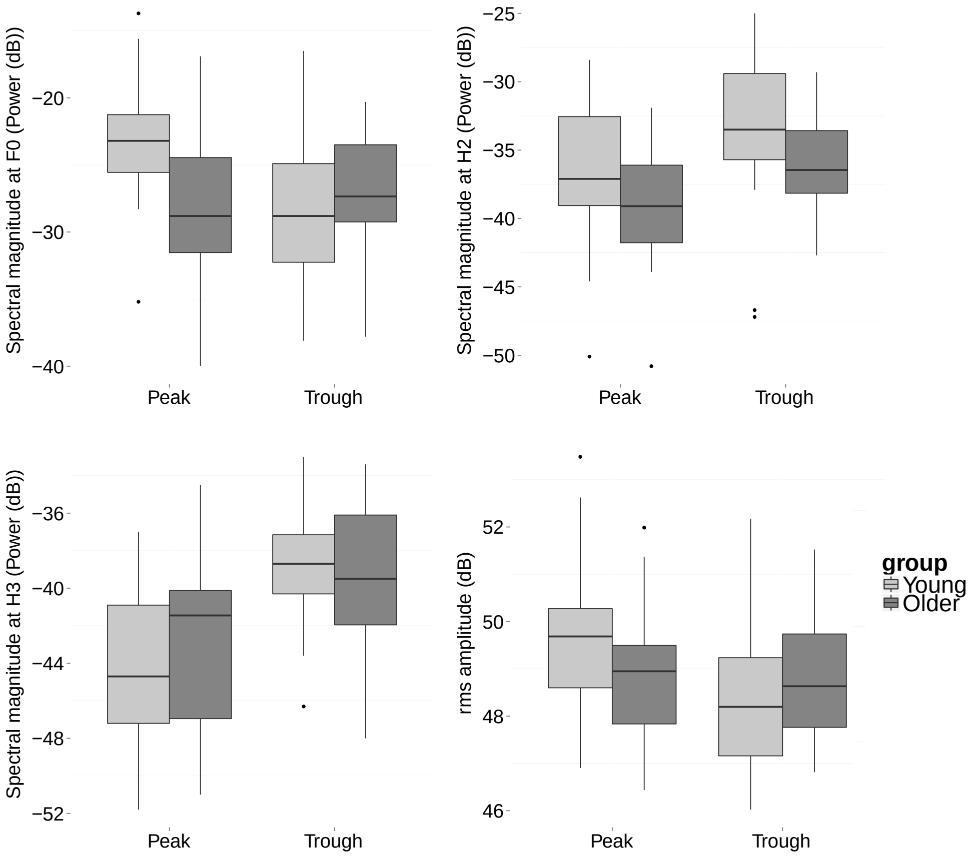


Figure 3.5: Boxplots of response properties of the FFR are shown for young (light grey) and older (dark grey) participants at the peak (left) and trough (right) of the AM noise. The measures plotted here are the spectral magnitude at the fundamental (F0) and second and third harmonics (Power (dB)) in the top-left, top-right, and bottom-left respectively. A boxplot of the rms amplitude (dB) is shown on the bottom-right.

	Mixed-effects models	
Measure	Window	Interaction
Spectral magnitude at F0	$F_{(1,34)} = 5.4$, $\mathbf{p = 0.03}^{*}$	$F_{(1,35)} = 9.5$, $\mathbf{p = 0.003}^{**}$
Spectral magnitude at 2^{nd} harmonic	$F_{(1,35)} = 8.9$, $\mathbf{p = 0.005}^{**}$	$F_{(1,35)} = 0.09$, $p = 0.8$
Spectral magnitude at 3^{rd} harmonic	$F_{(1,35)} = 36.2$, $\mathbf{p < 0.001}^{***}$	$F_{(1,35)} = 1.2$, $p = 0.3$
F0 strength	$F_{(1,35)} = 0.05$, $p = 0.8$	$F_{(1,35)} = 1.6$, $p = 0.2$
rms amplitude	$F_{(1,35)} = 11.8$, $\mathbf{p = 0.002}^{**}$	$F_{(1,35)} = 7.9$, $\mathbf{p = 0.008}^{**}$

Table 3.3: Results from mixed-effects models examining the effects of amplitude modulations in the masker on the FFR. Results show the main effect of window (i.e. peak or trough in the AM masker) and the interaction between window and group for several FFR measures. Significant effects are highlighted in bold font (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$).

computed by means of principal components analysis (PCA).

First, missing data points were imputed using a regression technique with an estimation adjustment based on the residuals. Subsequently, principal components analysis was performed on all FFR measures using varimax rotation with Kaiser normalisation. The resulting principal components (PC) were saved as Anderson-Rubin scores, which ensures the PC scores are uncorrelated.

The percentage of variance explained by each individual component was used as a first guiding principle for the number of PCs to be extracted. This resulted in the initial extraction of three components. Together they explained 43% of the variance in the data, with PC1 accounting for 20%, PC2 for 12%, and PC3 for 11%.

PC1 was interpreted as reflecting subcortical speech encoding in noise (PCnoise), while PC2 was interpreted to reflect FFRs in quiet (PCquiet). The loadings for the third PC were a subset of those of the first PC and was therefore not considered very informative. Consequently, subsequent analyses will only focus on the first two PCs. The factor loadings for these PCs are illustrated in figure 3.6.

The factor loadings for the first two PCs support the results reported above suggesting that the FFRs in quiet and noise are significantly different but that there is not necessarily a difference in terms of the type of background noise (AM vs. SS).

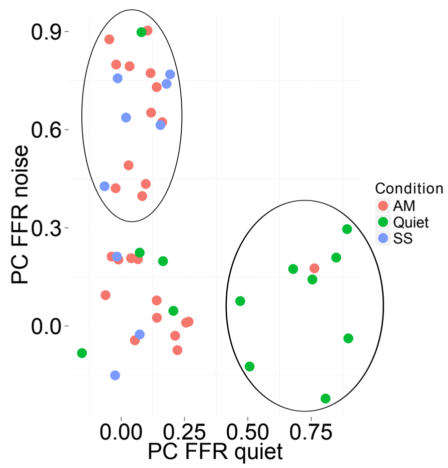


Figure 3.6: Factor loadings for the individual FFR measures on responses in quiet (green), SS noise (blue), and AM noise (red) onto the two extracted principal components (PC FFR noise and PC FFR quiet). The circles group the FFR measures with factor loadings above 0.4 on one of the principal components. This plot illustrates that the PC FFR quiet is dominated by measures on the FFR in quiet, while the PC FFR noise is dominated by measures on the FFR in AM and SS noise.

Furthermore, the finding that the FFR is degraded in older adults is supported by two independent sample t-tests performed on the extracted components. The analyses revealed significant group differences for both the PC of the FFR in quiet and noise (PCnoise: $t_{(35)} = 2.2$, $p = 0.04$, Cohen's $d = 0.9$; PCquiet: $t_{(35)} = 2.7$, $p = 0.01$, Cohen's $d = 0.7$).

Audiometric thresholds do not explain degraded FFRs: While the exact origins of the components of the FFR are not yet entirely clear, it is generally understood that delays imposed by the travelling wave along the basilar membrane play an important role (Don and Eggermont, 1978; Dau, 2003). Moreover, it has been suggested that phase-locked activity to the stimulus primarily stems from neurons at more basal sites (Janssen et al., 1991; Dau, 2003). In other words, both low and high frequency components in the FFR likely result from synchronised neural activity at mid- and high-frequency units and not necessarily from units tuned to the frequency of the stimulating waveform.

Given that basal regions play an important role in the generation of the FFR, differences in audiometric thresholds especially in the higher frequencies may have contributed to a decreased robustness of the FFR in the older adults. However, an elevation in audiometric thresholds above the stimulus frequency may not necessarily affect the FFR if the hearing loss is mainly due to outer hair cell (OHC) dysfunction.

		Best subsets regression: FFR					
Dependent variable	Predictors	b	β	SE	p	R^2	CI
FFR quiet	age group	-0.34	-0.34	0.16	**0.04***	0.12	[-0.66, -0.02]
FFR noise	age group	-0.4	-0.4	0.15	**0.01***	0.17	[-0.7, -0.1]

Table 3.4: Results of the best subsets regression analyses on the FFR principal components (* significant at $\alpha = 0.05$). Note that β refers to the standardised regression coefficient. The R^2 reflects the proportion of the variance accounted for as predictors are added to the model. 95% confidence intervals (CI) for the regression coefficients are also given.

When the hearing loss is mainly the result of OHC, and not inner hair cell, dysfunction, the basilar membrane response may not necessarily be reduced for stimuli in the tail of the tuning curve. Nonetheless, the potential contribution of elevated thresholds in the higher frequencies to the decreased robustness of the FFR in the older adults is worth investigating.

While both groups had near-normal hearing, defined as pure-tone thresholds $\leq$25 dB HL up to 4 kHz in both ears and at 6 kHz in at least one ear, their thresholds were significantly different. Independent t-tests indicated that pure-tone averages (PTA) across 0.5-4 kHz ($\leq$25 dB HL) and 6-8 kHz (some $\geq$25 dB HL) were significantly higher (i.e. worse) for the older age group (500 - 4000 Hz: $t_{(36)} = -6.4$, $p < 0.001$, mean difference 7 dB; 6 - 8 kHz: $t_{(36)} = -8.2$, $p < 0.001$, mean difference 16 dB).

To assess whether the degradation of FFRs in the older adults was driven by an elevation in audiometric thresholds, best subsets regressions were performed on the two principal component measures of the FFR (PC FFR quiet and PC FFR noise) independently with age group and PTAs across 0.5-4 kHz and 6-8 kHz as possible predictors (Hastie et al., 2009). The final models were selected based on the Bayesian Information Criterion (BIC; Schwarz, 1978).

The analyses indicated that both FFR measures in quiet and noise were best predicted by age group alone (see table 3.4). The models that best fit the data did not include either of the PTA measures. This means that degradation of the FFRs in the older adults can be attributed to ageing rather than elevated audiometric thresholds.

3.3.3 Predicting speech perception in noise

The primary aim of this study was to determine whether an age-related decline in subcortical auditory processing could explain difficulties typically experienced by normal-hearing older adults in understanding speech in noise. The older adults only performed worse in the presence two-talker babble but not in the presence of SS or AM noise maskers. Furthermore, while click ABRs were not significantly different between the two age groups, FFRs both in quiet and noise (AM and SS noise) were degraded in older adults. In addition, while both groups had near-normal hearing, thresholds for the older adults were significantly higher. These findings suggest that the older adults had no problems understanding speech in SS and AM noise, despite a decline in subcortical auditory processing and a comparative elevation in audiometric thresholds. The question remains, however, whether age-related changes in subcortical auditory processing can account for the group difference in SRTs in babble.

To answer this question, a best subsets regression analysis was performed on the SRTs in babble with age group, PTA across 0.5-4 kHz, click ABR wave V latency and amplitude, and the two principal component measures of the FFR (PC FFR quiet and PC FFR noise). The analyses suggested that SRTs in babble were best predicted by PTA across 0.5-4 kHz (R^2 = 0.2, $F_{(1,36)}$ = 9.2, p = 0.004, see table 5.4). Thus, age-related declines in subcortical auditory processing did not predict SRTs in babble beyond group differences in audiometric thresholds.

Furthermore, the fact that the older adults only performed worse on the speech in noise task in the presence of two-talker babble, but not in the presence of either of the two noise maskers (SS and AM noise), suggests that the relative contribution of various underlying processes involved in the perception of speech in noise may differ according to masker type. A question that remains, then, is whether individual differences in audiometric thresholds and/or the subcortical auditory processing can account for the variability in SRTs in SS and AM noise. Since the SRTs in AM and SS noise were highly correlated (r = 0.736, $p<0.001$, R^2=0.54, figure 2.8), the best subsets regression was performed on the average of the two, with age group, PTA across 0.5-4 kHz, and the two principal component measures of the FFR (PC FFR quiet and PC FFR noise) as possible predictors.

The results indicated that a model with PTA across 0.5-4 kHz and PC FFR noise best explained the variability in SRTs in the two noise maskers (R^2 = 0.14, $F_{(2,34)}$ = 2.8, p = 0.07, see table 3.5). It should be noted, however, that the FFR measure did not

		Regression on SRTs					
Dependent variable	Predictors	b	β	SE	p	R^2	CI
SRT babble	PTA 0.5-4 kHz	0.14	0.42	0.05	**0.008****	0.2	[0.05, 0.25]
SRT noise	PTA 0.5-4 kHz	0.11	0.35	0.05	**0.042***	0.06	[0.004, 0.2]
	PC FFR noise	0.44	0.29	0.25	0.09	0.08	[-0.08, 0.96]
FMB	Wave V latency	-2.3	-0.5	0.78	**0.006****	0.11	[-3.9, -0.7]
	Wave V amplitude	-6.6	-0.4	2.5	**0.013***	0.16	[-11.8, -1.5]

Table 3.5: Results from best subsets regression analyses on SRTs in babble, the averaged SRT in AM and SS noise, and FMB. Significant effects are highlighted in bold font (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$). Note that β refers to the standardised regression coefficient. The R^2 change reflects the proportion of the variance accounted for as predictors are added to the model. 95% confidence intervals (CI) for the regression slopes are given.

significantly predict SRTs in noise. In other words, individual differences in audiometric thresholds alone accounted for the variability in SRTs in the presence of noise maskers, as was the case for the SRTs in babble.

3.3.4 Predicting FMB

The results described above showed that response properties of the FFR tended to be more robust at the trough than at the peak of the AM masker. A question that remains, however, is whether the amount of neural release from masking could predict the amount of FMB a listener derives. It should be noted that the data did not show any age differences in terms of either the amount of neural release from masking or the amount of FMB. However, this does not mean the two measures could not be correlated.

In order to answer the question whether neural release from masking could predict FMB over and above simpler measures such as audiometric thresholds and the click ABR, a best subsets regression was conducted on the FMB with age group, PTA across 0.5-4 kHz, PC neural masking release, PC FFR quiet, PC FFR noise, and click ABR wave V amplitude and latency as possible predictors.

The results revealed that a model with wave V amplitude and latency best predicted variability in FMB ($R^2 = 0.27$, $F_{(2,31)} = 5.7$, $p = 0.008$, see table 3.5). The relationship between the click ABR measures and the FMB are illustrated in figure 3.7. As expected, longer wave V peak latencies, indicative of increased forward masking (c.f. Walton et al.,

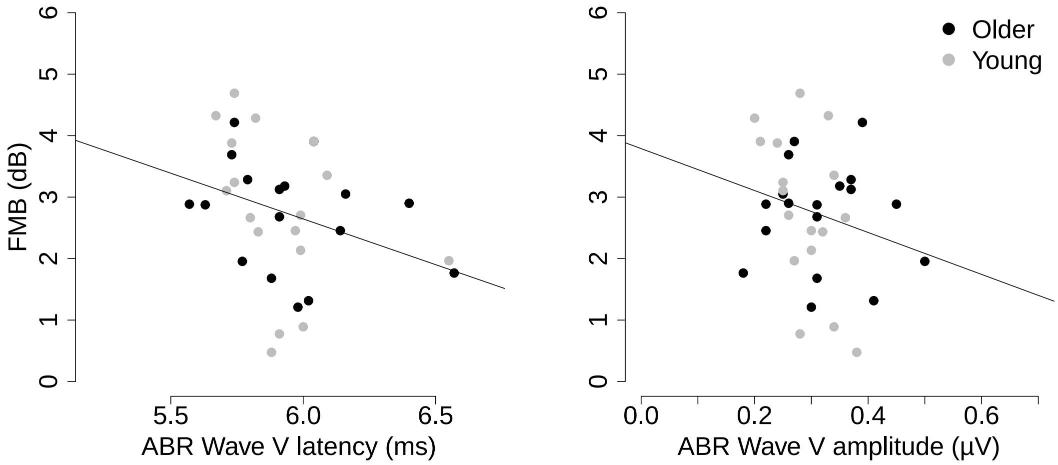


Figure 3.7: Scatter plots illustrating the (significant) relationship between FMB (in dB) and the ABR Wave V latency (in ms, left) and ABR wave V amplitude (in μV, right). Data points for the older adults are in black and for the younger adults in grey. The regression line reflects the correlation between the two measures across both groups.

1999; Poth et al., 2001; Debruyne, 1986; Fujikawa and Weber, 1977; Boettcher et al., 1996), are associated with larger FMBs. However, the direction of the relationship between FMB and wave V amplitude is opposite to what might be expected, with larger wave V amplitudes associated with smaller FMBs. However, it should be noted that the standard error associated with the regression coefficient of wave V amplitude was relatively large (see table 3.5). Nonetheless, these results suggest that the amount of neural masking release, as reflected in the FFR in AM noise, did not predict FMB over and above the simpler click ABR measure.

3.4 Discussion

The primary aim of this study was to determine whether declines in subcortical auditory processing could explain the increased difficulties older adults typically experience in the perception of speech in different types of background noise.

In line with previous research, the data revealed an age-related decline in the robustness of subcortical neural speech encoding that is independent of age-related changes in audiometric thresholds (c.f. Clinard and Tremblay, 2013; Vander Werff and Burns, 2011; Anderson et al., 2012). By contrast, wave V of the click ABR was unaffected by age, supporting the idea that the click ABR may not be sensitive enough to detect decreased precision of neural timing. While Vander Werff and Burns (2011) and Clinard and Tremblay (2013) only found age differences at the onset and offset of the FFR, our data reveal an age-related decline in FFR properties in response to the sustained vowel. These findings are in agreement with Anderson et al. (2012) who found age effects not only for the onset and formant transition period, but also in response to the sustained vowel portion of responses to the syllable /dɑ/. It is important to note that the FFR was not more affected by noise in older than younger listeners. FFRs were less robust in the older group both in quiet and in the presence of SS or AM noise. This indicates that auditory neuropathy may not disrupt subcortical speech encoding more in noise than quiet (c.f. Lopez-Poveda and Barrios, 2013).

It is difficult to determine, however, what this age-related decrease in subcortical speech encoding means in terms of a deficit in temporal processing. It is interesting to note that these same older adults show no declines in behavioural measures of TFS or Env processing (see chapter 2). Their performance was similar to that of younger participants on gap, amplitude modulation, and frequency modulation detection tasks. This discrepancy between the behavioural and electrophysiological data raises the question whether older adults somehow compensate for this decline in neural temporal precision or whether the behavioural measures and the FFR reflect different processes. In order to aid interpretation of FFR data, it is important to improve our understanding of what the components of the FFR reflect, and how this relates to behavioural measures of temporal processing.

It was hypothesised that these declines in subcortical auditory processing would lead to increased difficulties understanding speech in noise. However, the normal-hearing older adults only performed more poorly on the speech in noise task compared to the younger listeners in the presence of two-talker babble. This implies that normal-hearing older adults may be unimpaired in their perception of speech in the presence of AM and SS noise when complex, ecologically valid target stimuli are used, despite declines in subcortical auditory processing. Moreover, the small increase in SRTs in babble could not be explained in terms of the age-related changes in subcortical auditory processing. Instead, individual differences in SRTs both in the presence of noise maskers and two-talker babble were best explained in terms of differences in audiometric thresholds across 0.5-4 kHz. This suggests that declines in the precision of temporal neural coding do not necessarily lead to increased perceptual difficulties of speech in noise.

It is important to consider several possible explanations for the absence of a relationship between the FFR and the performance on a speech in noise task. First, poorer speech in noise performance has often been associated with decreased robustness of F0 encoding (Anderson et al., 2011; Song et al., 2011). However, the present data suggest that ageing does not necessarily lead to a reduction in the FFR spectral magnitude at F0 (see also Anderson et al., 2012). Second, it has been argued that decreased speech encoding relevant for the perception of speech in noise particularly becomes apparent for stimuli with rapidly changing acoustic features (e.g. the formant transition period in /dɑ/; Song et al., 2011; Anderson et al., 2013; Hornickel et al., 2009). The present study used a steady-state vowel, however, in order to be able to assess the effects of amplitude fluctuations in the masker. Could it be that age-related changes in subcortical auditory processing of a dynamically changing stimulus would predict speech in noise performance? Given the reduced robustness of the FFR in response to a steady-state vowel it is reasonable to speculate that the older adults would also show decreased response properties for a rapidly changing stimulus. However, the fact remains that the older adults did not have increased difficulties understanding speech in SS or AM noise. It would therefore be unlikely that a potential group difference in the FFR to, for instance, a formant transition period would relate to the older adults' abilities to understand speech in noise. Lastly, it may simply be that the amount of disruption of neural speech coding was not severe enough to affect speech in noise performance. Speech perception, particularly in the presence of background noise, is a complex process that involves both top-down and bottom-up processes. It may thus be the case that older adults with some form of auditory neuropathy can somehow compensate for the diminished precision of speech coding at the early stages along the auditory pathway.

The data furthermore suggest that normal-hearing older adults do not experience reduced neural release from masking. While older adults show less robust FFRs in AM noise overall, they benefit just as much as younger listeners from amplitude dips in the maskers, with more robust response properties at the trough compared to the peak of the masker. This is perhaps not so suprising since behavioural and ASSR measures have indicated that age-related declines in Env processing only become apparent at higher modulation rates (above about 200 Hz, e.g. Purcell et al., 2004; Grose et al., 2009).

There was no significant relationship between neural masking release and the ability to listen in the dips of the fluctuating noise. This might suggest that individual differences in FMB for relatively slow modulation rates is the result of differences in the ability to use speech cues available in the dips of the masker as opposed to differences in recovery times from neural adaptation (see also Grose et al., 2009). However, the lack of association here could simply be due to the fact that there was limited variability in either FMB or neural masking release across listeners. The amount of FMB was best predicted by the click ABR, with increased wave V peak latencies associated with reduced dip listening skills. Whether neural masking release reflected in the FFR would provide additional information, over and above the click ABR, with regards to dip listening at higher modulaiton rates remains an open question.

In sum, the aim of this study was to determine whether age-related neuropathy could explain the increased difficulties older adults typically experience in the perception of speech in noise. The findings can be summarised as follows:

1. Ageing is associated with a decline in the robustness of subcortical speech encoding, both in quiet and noise, that is independent of age-related changes in audiometric thresholds.

2. Normal-hearing older adults do not show a decrease in neural masking release, as indicated by the differences in the robustness of the FFR at the peak and trough of an AM masker, at slow masker modulation rates (10 Hz).

3. Age-related declines in subcortical neural speech encoding do not necessarily lead to increased difficulties understanding speech in noise.

4. Neural masking release as reflected in the FFR in AM noise does not predict dip listening over and above the click ABR, at least not at relatively slow masker modulation rates.

5. It remains unclear what this disruption in subcortical speech encoding means in the context of age-related declines in temporal processing.

Chapter 4

The effect of auditory neuropathy on sensitivity to spectro-temporal modulations and speech perception in noise for older adults

The study reported in this chapter was conducted in collaboration with Joshua Bernstein. He provided the spectro-temporal modulation (STM) detection task, converted percent-correct scores for the STM task into equivalent-dB values, and calculated speech intelligibility index scores.

4.1 Introduction

As discussed before, the increased speech in noise difficulties typically reported for older adults cannot fully be accounted for by the audiogram alone (e.g. Dubno and Morgan, 1984; Helfer and Wilber, 1990; Middelweerd et al., 1990). It has often been suggested that suprathreshold declines associated with ageing, especially a deficit in auditory temporal processing, may contribute to the speech in noise difficulties (e.g. Pichora-Fuller and Souza, 2003; CHABA, 1988; Gordon-Salant, 2005). One promising measure that may in part explain the age-related changes in speech in noise performance is sensitivity to spectrotemporal modulations (Bernstein et al., 2013; Mehraei et al., 2013).

Speech is characterised by spectrotemporal modulations (i.e. amplitude fluctuations across both time and frequency) as illustrated by the spectrogram in figure 4.1. Just like any waveform can be decomposed into a set of sinusoids, any spectrogram can be decomposed into a set of spectrotemporal ripples, each with a different combination of spectral density (cycles per octave) and temporal modulation rate (Hz).

Speech intelligibility largely depends on spectral densities below about 2 cycles/octave, reflecting the formant structure of speech (Liu and Eddins, 2008; Elliott and Theunissen, 2009), and temporal modulation rates up to about 16 Hz, reflecting the syllable rate of speech (Drullman et al., 1994).

Sensitivity to temporal and spectral modulations are typically studied independently. However, given that spectral and temporal modulations co-occur in speech - think for example of formant transitions - it may be more ecologically valid to assess the combined spectrotemporal modulations in relation to speech perception.

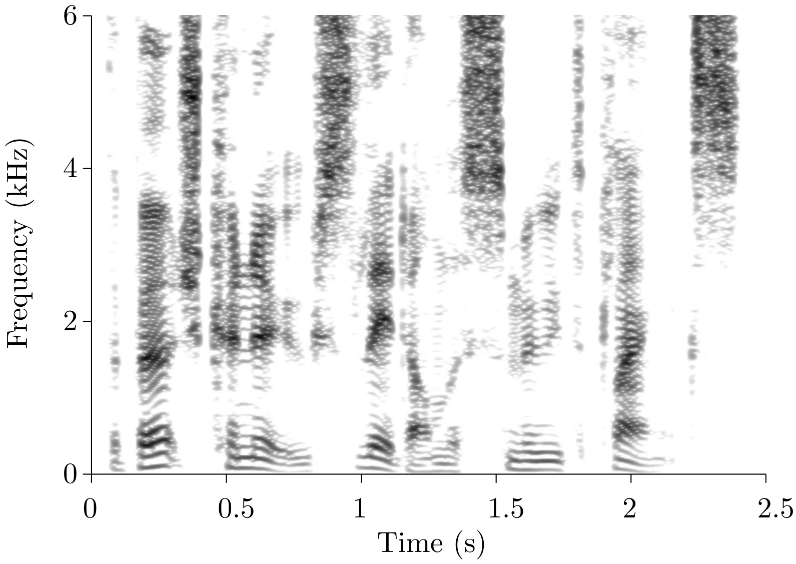


Figure 4.1: Spectrographic representation of the sentence "The birch canoe slid on the smooth planks", illustrating energy modulations across both time and frequency.

Several studies have shown a relationship between spectrotemporal modulation (STM) sensitivity and speech perception in noise, both for normal-hearing and hearing impaired listeners (Chi et al., 1999; Elhilali et al., 2003; Bernstein et al., 2013; Mehraei et al., 2013). Chi et al. (1999) and Elhilali et al. (2003) developed the spectro-temporal modulation index (STMI), which quantifies the degradation of spectrotemporal modulations of a processed signal (e.g. in noise or reverberation) with reference to the modulations present in the unprocessed signal. It is important to note that the STMI is based on the spectrotemporal modulations at the output of a bank of peripheral auditory filters, not on the stimuli themselves. Chi et al. (1999) and Elhilali et al. (2003) showed that the STMI was a good predictor of speech perception by normal-hearing adults in the presence of noise or reverberation.

The success of this STM-based model in predicting speech perception in normal-hearing adults motivated Bernstein et al. (2013) to assess whether STM sensitivity was related to speech perception for hearing impaired individuals. They measured STM sensitivity of a broadband noise for a range of combinations of rates, densities, and directions (i.e. upward or downward moving ripples). They found that hearing loss only affected STM sensitivity for stimuli with spectral densities of 2 or 4 cycles/octave in combination with temporal rates of 4 or 12 Hz. It should be noted that the hearing impaired participants tested by Bernstein et al. (2013) were also a lot older than their normal-hearing controls (mean age controls 44.5 yrs, hearing impaired 75.7 yrs). It is thus possible that the declines in STM sensitivity were the result of ageing rather than hearing impairment. Bernstein et al. (2013) furthermore found that, within a group of hearing impaired listeners, STM sensitivity was correlated with speech perception in steady-state noise, even after accounting for individual differences in the speech intelligibility index (SII).

The current study examined STM sensitivity using a stimulus with a spectral density of 2 cycles/octave and a temporal modulation rate of 4 Hz. If the older adults do indeed show reduced STM sensitivity, this may to some extent be explained by a reduction in frequency selectivity. Poor spectral resolution, as a consequence of broadened auditory filters, will affect the ability to resolve the spectral components of the STM stimulus, especially at high spectral densities. Bernstein et al. (2013) found that, for hearing impaired listeners, detection thresholds for STM of a broadband noise with a spectral density of 2 cycles/octave correlated significantly with a measure of frequency selectivity at 4 kHz. Frequency selectivity accounted for 27.5% of the variability in STM detection thresholds. To minimise the contribution of reduced frequency selectivity on STM detection thresholds as a consequence of hearing loss, the STM stimuli in the current study were limited to 2 kHz, which fell within the range of normal hearing for both young and older adults.

Frequency selectivity may be reduced, however, even when audiometric thresholds are normal (c.f. Moore, 2007). Figure 4.2 shows the equivalent rectangular bandwidths (ERB; an approximation of auditory filter widths) for hearing impaired listeners as a function of audiometric threshold. The figure illustrates that the auditory filters for hearing impaired listeners may be slightly broader compared to normal-hearing listeners, even in their region of normal hearing. It may thus similarly be the case that the older participants in the current study have broader auditory filters in the range of normal hearing (i.e. up to 4 kHz), which may affect both STM sensitivity and speech in noise

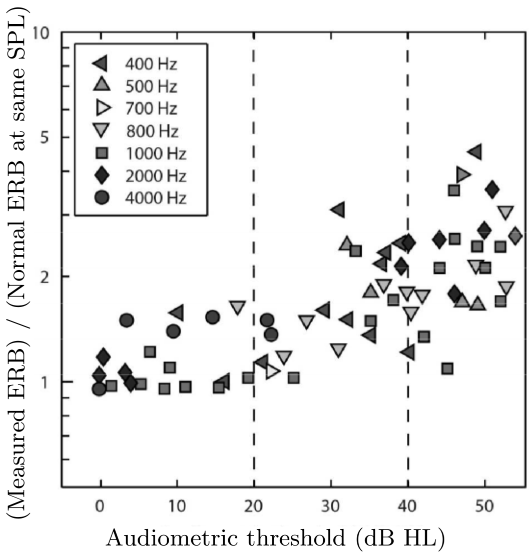


Figure 4.2: Equivalent rectangular bandwidths (ERBs) at different centre frequencies (indicated by different symbols/colours) for hearing impaired listeners as a function of audiometric threshold (in dB HL). The ERB values are plotted relative to the ERBs for young normal-hearing listeners (tested at the same stimulus level). The dotted lines indicate the regions of normal hearing, and mild and moderate hearing loss. From Léger (2012) and Moore (2007)

performance.

If the older adults do indeed have broadened auditory filters, this will likely also affect their ability to understand speech in noise. It has been shown that, for listeners with mild to moderate hearing loss, broader auditory filter widths are associated with poorer speech perception in noise (Festen and Plomp, 1983; Horst, 1987). However, given that hearing impairment is associated with broadened auditory filters, it remains difficult to separate the effects of reduced frequency selectivity by itself from effects of reduced audibility. Perhaps more convincing support for the role of reduced frequency selectivity in the perception of speech in noise comes from simulation studies (Baer and Moore, 1993; Léger, 2012). Baer and Moore (1993) showed, for instance, that the intelligibility of spectrally smeared sentences was significantly reduced for normal-hearing listeners in the presence of background noise. The effects of reduced frequency selectivity on speech perception in noise are best explained by the fact that more noise will pass through broadened auditory filters, thereby effectively reducing the signal-to-noise ratio (SNR) in those filters.

A crucial aspect of successful STM detection is the ability to track envelope (Env) fluctuations. It is unlikely, however, that poor Env processing could explain any reductions in STM sensitivity for older adults in the present study, since a low temporal modulation rate (4 Hz) was used. As discussed in chapter 2, age-related declines in Env processing only become apparent at higher modulation rates (above about 200 Hz; e.g. Purcell et al., 2004; Grose et al., 2009).

Instead, potential age-related reductions in STM sensitivity may in part be the result of a decline in temporal fine structure (TFS) processing associated with ageing (Vongpaisal and Pichora-Fuller, 2007; He et al., 2007; Grose and Mamo, 2010; Hopkins and Moore, 2011; Füllgrabe, 2013). A decline in TFS processing likely affects listeners' abilities to follow the spectral changes over time. The spectral ripples can be described as frequency modulated noisy sinusoids. In this sense the STM task is very similar to frequency modulation detection, which for low carriers and modulation rates reflects TFS processing (Moore and Sek, 1995, 1996). Bernstein et al. (2013) indeed found that STM thresholds (2 cycles/octave, 4 Hz) were significantly correlated with frequency modulation detection thresholds for a 500-Hz carrier modulated at 2 Hz (Bernstein et al., 2013). The ability to detect frequency modulation accounted for 39.9 % of the variability in STM thresholds.

If STM sensitivity to some extent depends on TFS processing, any age-related declines in STM sensitivity may ultimately be the result of auditory neuropathy associated with ageing (Schmiedt et al., 1996; Makary et al., 2011; Sergeyenko et al., 2013; Zeng et al., 2004). In particular, a loss of functional auditory nerve fibres (ANFs) may explain the poorer STM detection thresholds, and ultimately the increased speech in noise difficulties. As discussed in chapter 3, ANFs encode sound with great temporal precision, at least at relatively low frequencies, by phase-locking to the stimulating waveform. Therefore, a loss of healthy ANFs will likely lead to reduced temporal precision of neural coding.

Auditory neuropathy is typically assessed using the click-evoked auditory brainstem response (ABR). Recent animal work has indicated that noise exposure and ageing may lead to neural degeneration, even in the absence of a permanent elevation in audiometric thresholds. Kujawa and Liberman (2009) showed, for example, that noise-exposed mice showed recovery of outer hair cell (OHC) function, but permanent deafferentation of the auditory nerve. While distortion product otoacoustic emissions (DPOAEs), reflecting OHC function, fully recovered, ABR wave I amplitudes were less than half their pre-exposure values 8 weeks post exposure. Similarly, Sergeyenko et al. (2013) looked at DPOAEs and ABRs for mice who were never exposed to high noise levels. They found gradual changes in both DPOAEs and ABRs across the lifespan. However, age-related changes were greater for the click ABR than the DPOAEs, suggesting that while OHC dysfunction may in part contribute to reductions in ABR wave I amplitudes, dysfunctional inner hair cells (IHC) and/or ANF degeneration likely also played an important role (Sergeyenko et al., 2013).

Sergeyenko et al. (2013) furthermore found that the amplitude of wave I, reflecting neural firing at the auditory nerve, was more affected by age than the amplitude of wave V, which is dominated by contributions from the inferior colliculus in the brainstem (Møller, 2007). A similar pattern has been found for humans with tinnitus, with reduced wave I but not wave V amplitudes (Schaette and McAlpine, 2011). Schaette and McAlpine (2011) suggested that deafferentation of the ANFs may lead to increased central gain (i.e. increased neural activity further along the auditory pathway) which might explain the fact that wave V amplitudes did not differ between the groups. It should be noted, however, that the interaction between wave and group was not statistically tested in these human and animal studies. Ageing, and tinnitus, may thus equally affect neural firing at the auditory nerve and the inferior colliculus.

As discussed in chapter 3, it remains unclear how the ABR exactly changes with age in humans since results are often confounded by hearing loss and depend largely on the choices of stimulus parameters (for a review see Tremblay and Burkard, 2007). Often, however, a reduction in wave I amplitude is reported (Sand, 1991; Burkard and Sims, 2001), suggesting a reduction in the number of functioning ANFs, in line with the animal studies reported above.

This study addressed the question whether older adults show reduced STM sensitivity in the range of normal hearing and whether this can predict any speech in noise difficulties. If the older adults do indeed show higher STM detection thresholds compared to the younger group, this would likely reflect a decline primarily in TFS processing. However, given that people may have broadened auditory filters even in the range of normal hearing, which could affect both STM sensitivity and the ability to understand speech in noise, a measure of frequency selectivity was also obtained. Of primary interest, however, was whether any age-related changes in STM detection thresholds and increased difficulties understanding speech in noise could in part be explained by deafferentation from the auditory nerve up to the brainstem, as measured by the click ABR.

4.2 Methods

All tasks in this study were conducted monaurally. Testing took place in an electrically shielded sound-proof booth.

4.2.1 Participants

Eighteen young (11 female, 18-30 yrs, mean 23.3 yrs, sd 4.2 yrs) and 18 older (13 female, 61-78 yrs, mean 67.1 yrs, sd 5.1 yrs) monolingual English speakers participated in this study. All participants had near-normal hearing defined as (air-conducted) pure-tone thresholds $\leq$ 25 dB HL at octave frequencies from 0.25 to 4 kHz in at least one ear. The older adults had varying degrees of sloping hearing loss above 4 kHz. Figure 4.3 shows the audiograms of both groups for the test ear only (young: eight left; older: six left). All participants over the age of 65 had normal cognitive function (scores $\geq$17 MMSE telephone version (Roccaforte et al., 1992)). None of the participants reported a history of language or neurological disorders. Participants signed a consent form approved by the UCL Research Ethics Committee and were paid for their participation.

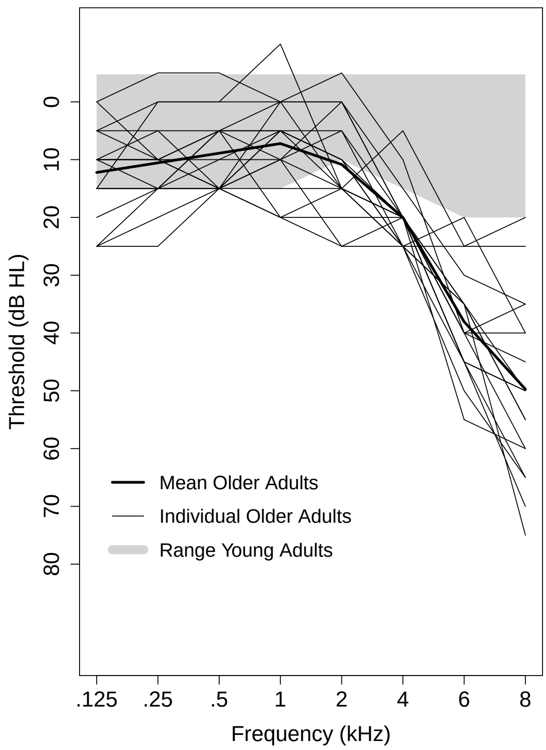


Figure 4.3: Audiometric thresholds for the test ear. Individual audiograms for older adults are plotted. The shaded area represents the range of audiometric thresholds of younger adults.

4.2.2 Behavioural measures

4.2.2.1 Speech perception in noise

Speech reception thresholds (SRTs) were measured for IEEE sentences (Rothauser et al., 1969) in steady-state speech-shaped noise. The target stimuli were pre-recorded IEEE sentences produced by a male talker of standard Southern British English. The speech-shaped noise is described in more detail in Rosen et al. (2013). The masker always started 600 ms prior to stimulus onset and was gated on and off across 100 ms.

Given that the older adults had varying degrees of sloping hearing loss above 4 kHz, the stimuli were spectrally shaped using the National Acoustics Laboratories-Revised (NAL-R) linear prescriptive formula (Byrne and Dillon, 1986). To ensure the stimuli were audible up to at least 6 kHz for all participants, the same amount of gain was applied for everyone, accounting for a maximum audiometric threshold of 55 dB HL at 6 kHz. In addition, due to the large variability in audiometric thresholds at 8 kHz in the older group (20 - 75 dB HL) the stimuli were low-pass filtered at 6 kHz using a fourth order Butterworth filter. The speech and noise files were filtered offline before NAL-shaping was applied. However, given the relatively shallow filter, combined with the spectral shaping, the stimuli may still have been audible above 6 kHz for some participants.

The participants listened to the stimuli over Sennheiser HD 25 headphones. They were asked to repeat what they heard and the experimenter scored the number of correctly repeated key words. No feedback was provided.

The SNR was varied adaptively following the procedure described by Plomp and Mimpen (1979), and SRTs were tracked at 50% correct. The stimuli were presented at 70 dB SPL. The participants were given brief training (five trials, starting at 0 dB SNR) to familiarise themselves with the materials. The procedure is described in more detail in chapter 2.

The SRTs reported here are the mean across three runs. A measurement was repeated, with a different set of sentences, when fewer than three reversals were obtained or when the standard deviation across the final reversals exceeded 4 dB. Thresholds for each run were computed by taking the mean SNR (dB) across the final number of reversals.

4.2.2.2 Spectro-temporal modulation detection

The ability to detect spectro-temporal modulation (STM) of noise was measured for downward-moving STM ripples with a spectral density of 2 cycles/octave and a temporal modulation rate of 4 Hz (see Fig. 4.4). The noise carrier was constructed by summing together 2500 equal-amplitude random-phase sinusoids that were logarithmically spaced across 2.5 octaves (353 - 2000 Hz). Stimuli had a duration of 500 ms and were tapered on and off across 50 ms (cosine ramps).

STM was introduced by adding sidebands to each carrier tone and adjusting the relative phases of the sidebands for each individual carrier. A given modulation sideband was only added to a carrier if its frequency fell within the frequency range of the noise carrier. This was done to reduce the likelihood of listeners using spectral-edge cues which would be present if modulation sidebands fell outside the frequency range of the carrier (c.f. Mehraei et al., 2013). In addition, no unmodulated noise was present above or below the carrier edge frequencies in order to limit the availability of profile-analysis cues that could allow listeners to perform the task by comparing the relative levels of the modulated and unmodulated bands.

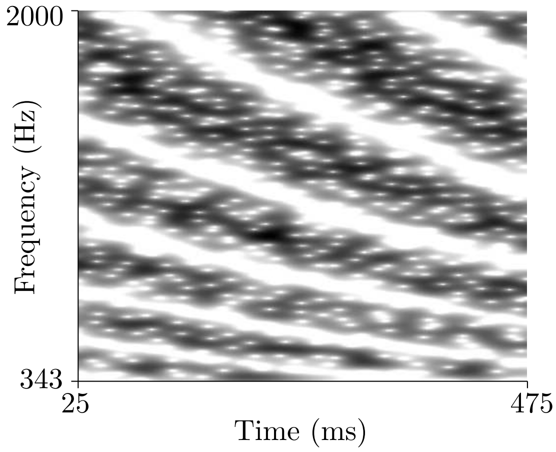


Figure 4.4: Spectrogram of an STM stimulus with a downward-moving ripple with a spectral density of 2 cycles per octave and a temporal modulation rate of 4 Hz at full modulation depth (0 dB). Note that the first and last 25 ms of the stimulus are not captured by the spectrogram due to windowing.

Stimuli were presented to the test ear over Sennheiser HD 25 headphones at an overall level of 85 dB SPL. To rule out any intensity cues the stimuli were roved across a 5 dB range (-2.5 to +2.5). In addition, to ensure the participants were indeed only using their test-ear, an uncorrelated noise was presented in the non-test ear 20 dB below the stimulus level. Visual feedback as to the correctness of the response was provided.

STM detection thresholds were determined using a two-alternative forced-choice task. The interstimulus interval was 100 ms. The modulation depth (in dB; 20log(m), where m is the modulation depth) was varied adaptively using a one-up three-down procedure tracking 79.4% correct (Levitt, 1971). Thus, the modulation depth was decreased following three consecutive correct responses and increased after a single incorrect response. If an incorrect response was given at full modulation depth the subsequent trial was again presented at full modulation depth. The starting modulation depth was 0 dB (i.e. full modulation). The initial step size was 6 dB and was decreased to 4 dB, and finally 2 dB after each reversal. A run was terminated after nine reversals. The STM threshold was calculated across the final six reversals. Thresholds reported here are the average across three runs.

When an incorrect response was given at full modulation more than five times in a single run, the modulation depth was fixed to 0 dB for the remainder of the run. In that case, STM detection thresholds (in dB) could not be recorded and instead a percent-correct score was obtained. In total, the modulation depth was fixed to 0 dB for seven individual runs (one run each for four older adults and all three runs for another).

The percent-correct scores were subsequently converted into equivalent-dB values as follows. First, for each of the 36 participants, the percent correct for each modulation depth was converted to a d' score. These values were fit to a straight line. The average slope across listeners was 0.25 d' units/dB. Subsequently, the percent-correct scores that were to be converted into equivalent-dB values were converted into d' values and compared to the d' value associated with the adaptive-tracking percentage-correct score (79.4% correct). The difference in d' scores was divided by the earlier derived slope (0.25 d' units/dB) to get the equivalent-dB value.

Participants typically received two training runs to familiarise themselves with the task. The first training run was easier than the experiment proper with a spectral density of 1 cycle/octave and a rate of 4 Hz. The second training run was the same as the test condition with ripple stimuli with a spectral density of 2 cycles/octave and a rate of 4 Hz. If a participant did not understand the task, additional training was provided.

4.2.2.3 Frequency selectivity

Masked thresholds were estimated for sinusoidal probe tones of 500 and 4000 Hz in the presence of broadband and notched-noise maskers. These frequencies were used as they spanned the range of normal hearing for the group of older listeners.

The probe level was fixed at 50 dB SPL and the level of the masker varied adaptively following a three-down one-up rule tracking 79.4% (Levitt, 1971). A probe level of 50 dB SPL was used as this was at least 15 dB SPL above threshold for all participants.

A two-alternative forced-choice paradigm with visual feedback was used. On the

	Frequency Selectivity	
Probe	Masker outside edges	Notch edges
500 Hz	100 - 900 Hz	350 - 650 Hz
4000 Hz	800 - 7200 Hz	2800 - 5200 Hz

Table 4.1: Outside edges and notch widths of the maskers used in the frequency selectivity task.

first trial the masker was set to a spectrum level of 30 dB in the notched-noise condition and to a spectrum level of 10 dB in the broadband noise condition. At these levels the probe was clearly audible. The masker level was increased following three consecutive correct responses and decreased after a single incorrect response. The initial step size was 5 dB, decreasing by 1 dB after each reversal to a final step size of 2 dB. The run was terminated after eight reversals at the final step size. Thresholds were computed across the final eight reversals. For each of the four conditions (notch/no notch, 500/4000 Hz) two measurements were obtained. Thresholds reported here are the average across two runs.

The duration of the probe was 360 ms plus 20 ms raised-cosine onsets and offsets. The maskers had a duration of 460 ms plus 20 ms raised-cosine onsets and offsets. The probe was temporally centred in the masker. Stimuli were presented with an interstimulus interval of 200 ms and an inter-trial interval of 600 ms.

The outside edges of the maskers and the notch widths are specified in normalised frequency units relative to the probe tone (f_0) as $(|f - f_0|)/f_0$. Both the notched and broadband noises were placed symmetrically around the probe frequency with outer edges of the noise at $\pm 0.8 \times f_0$. The notch edges were fixed at a normalised value of 0.3 (see table 4.1). The spectral components within the appropriate frequency limits were set to have equal amplitudes while those outside those limits were set to zero. Components within the frequency limits had randomised phases which were uniformly distributed in the range of 0 - 2π radians. An inverse FFT was applied to create the time waveform.

The participants listened to the stimuli through ER-1 insert earphones (Intelligent Hearing Systems, Miami, FL). Stimuli were presented using an external soundcard (RME Babyface UC, 44.1 kHz) connected to the custom made trigger box that was used for the measurement of the click ABRs (see 4.2.3) and an attenuator with zero gain. The trigger box and attenuator functioned as an impedance buffer to drive an appropriate amount of current to the low impedance (10 Ohms) insert earphones. Given that the frequency response of the earphones was not flat, the stimuli were passed through a filter which was the inverse of the frequency response of the earphones to create equal amplitude components at the eardrum, as measured in a Brüel & Kjær ear simulator (Type 2619; Nærum, Denmark).

To provide a simple index of frequency selectivity, a difference measure was computed by subtracting thresholds (of the masker level) in notched noise from those in broadband noise. Larger differences are indicative of better frequency selectivity.

4.2.3 Electrophysiological measures

Click ABRs were recorded in response to 2000 sweeps of a 100 μs click with alternating polarity. Stimuli were presented monaurally via electrically shielded ER-3 inserts (Intelligent Hearing Systems, Miami, FL) at 75 dB nHL (112.6 dB peSPL) with a repetition rate of 11/s (Konrad-Martin et al., 2012). To confirm the reliability of the measures, click ABRs were measured twice.

The stimuli were generated in MATLAB (Mathworks, Natick, MA). In order to eliminate potential jitter from the recordings, a string of the required number of stimuli was created on one channel and a string of corresponding triggers on the other channel. These channels were subsequently delivered via a computer using an external soundcard (RME FireFace UC, 44.1 kHz) connected to a custom-made trigger box which separated the two channels and simultaneously sent the trigger to the BioSemi machine and the stimulus to the insert earphones.

Participants were seated in a reclining chair and allowed to fall asleep. ABRs were collected using a BioSemi ActiveTwo system (Amsterdam, The Netherlands). Responses were recorded differentially between Cz and the ipsilateral earlobe. Two additional electrodes, Common Mode Sense (CMS) and Driven Right Leg (DRL), were placed around Pz. In the BioSemi system, the CMS and DRL electrodes replace the typical ground electrode. Electrode offsets were always $<$40 mV. Responses were recorded with a sampling rate of 16384 Hz.

Click ABRs were filtered from 100 to 3000 Hz (12 dB/octave, zero phase-shift) and epoched from -14 to 14 ms. Baseline correction was performed with respect to the prestimulus response. Any sweeps exceeding $\pm$25 μV were rejected. The click ABRs were subsequently averaged across the two runs.

To examine a possible reduction in the available number of healthy auditory nerve-fibre synapses in the older adults, amplitudes for waves I and V of the click

ABR were determined.

4.3 Results

Several data points were excluded from the analysis reported below. Click ABRs and a measure of frequency selectivity at 500 Hz for one older adult and STM detection thresholds for another older adult could not reliably be determined.

Descriptive statistics for all measures as well as confidence intervals for the group differences are summarised in table 4.2.

	Descriptive statistics		
Dependent variable	mean (sd) young	mean (sd) older	CI
SRT	-4 (0.9)	-2.3 (1.2)	[-2.2, -0.8]
STM	-12.1 (1.4)	-8.7 (3.9)	[-5.4, -1.5]
ABR wave I amplitude	0.17 (0.09)	0.09 (0.08)	[0.02, 0.13]
ABR wave V amplitude	0.38 (0.1)	0.34 (0.1)	[-0.04, 0.12]
Frequency selectivity at 0.5 kHz	19.0 (3.3)	19.3 (7.9)	[-4.5, 4.0]
Frequency selectivity at 4 kHz	25.0 (3.3)	24.7 (4.5)	[-2.3, 3.0]

Table 4.2: Descriptive statistics (mean and standard deviation) for the young and older adults separately as well as confidence intervals for the group differences are provided for all measures.

4.3.1 Audiometric thresholds

Summary measures for various combinations of frequencies in the pure-tone audiogram (PTA) were calculated and are summarised in table 4.3.

As illustrated by figure 4.3, the older group had significantly higher thresholds above 4 kHz. While both groups had normal hearing up to 4 kHz (pure-tone thresholds $\leq$25 dB HL), the older group also had higher thresholds within this region of normal hearing (see table 4.3). These differences in audiometric thresholds could potentially explain any age-related differences found for any of the behavioural and electrophysiological measures.

Given that the STM stimuli only contained energy up to 2 kHz, a PTA measure was computed across 0.25 - 2 kHz. This measure was used as a predictor in the regression analyses reported below (section 4.3.6).

Furthermore, since the stimuli used in the SRT task may have been audible above 6 kHz for some but not all participants, Speech Intelligibility Index scores (SII) were

	Independent t-tests PTA measures			
Dependent variable	df	t	p	mean difference
PTA 0.25 - 2 kHz	34	3.4	**0.002****	6 dB
PTA 0.5 - 4 kHz	34	5.3	$<$ **0.001*****	8 dB
PTA 0.5 - 6 kHz	34	8.8	$<$ **0.001*****	13 dB
PTA 6 - 8 kHz	34	12.8	$<$ **0.001*****	39 dB
SII	17.1	4.2	$<$ **0.001*****	0.006

Table 4.3: Results of independent t-tests comparing young and older adults on different combinations of frequencies in the PTA (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha < 0.001$).

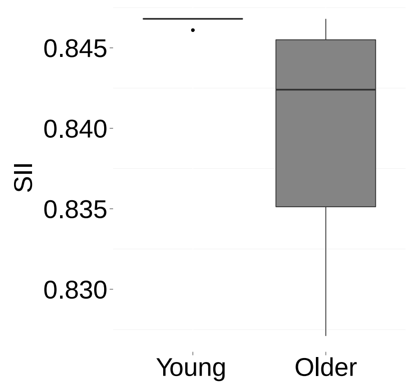


Figure 4.5: Boxplots of the Speech Intelligibility Index (SII) scores for young (left) and older adults (right), calculated for the sentence materials (in quiet) used in the speech in noise task.

computed for the processed sentences in quiet. Figure 4.5 shows that the variability in SII scores was incredibly small, with scores ranging from 0.82 - 0.84 (for a measure ranging from 0 to 1). Nonetheless, the scores differed significantly between the two groups, with slighly lower SIIs for the older adults. It is important to note, however, that the two groups were not homogeneous in terms of their variances (see Fig. 4.5). Therefore, a PTA measure was computed across 0.5 - 6 kHz that was used in the regression analyses on the SRTs instead (section 4.3.6).

4.3.2 Speech perception in noise

To examine whether the older adults did indeed have difficulties understanding speech in noise, an independent t-test was conducted. The results showed a significant difference

($t_{(34)}$ = 4.4, p <0.001, Cohen's d = 1.46) between the two groups, with higher (i.e. worse) SRTs for the older group by 1.5 dB (see Fig. 4.6).

4.3.3 Spectro-temporal modulation detection

Older adults were expected to perform more poorly on the STM task. This was confirmed by an independent t-test which revealed a significant difference ($t_{(19.7)}$ = 3.5, p = 0.003, Cohen's d = 1.15) between the age groups, with higher (i.e. worse) modulation depth thresholds for the older group by 3.5 dB (see Fig. 4.6). Note that the degrees of freedom were adjusted because the Levene's test for equality of variances indicated that the two groups were not homogeneous in terms of their variances.

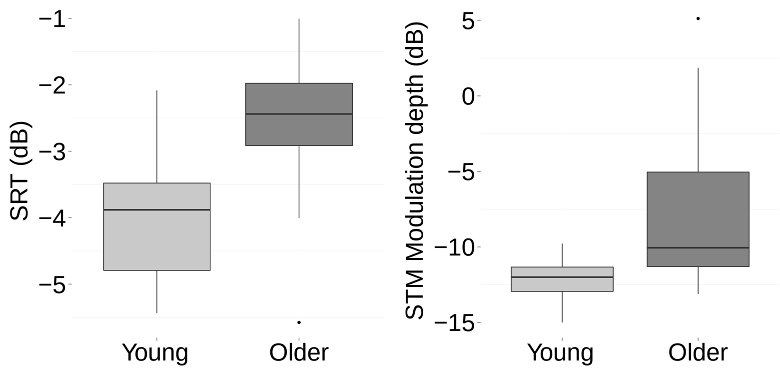


Figure 4.6: Boxplots of the speech reception thresholds (left) and spectrotemporal modulation detection thresholds (right) for the young and older adults.

4.3.4 Frequency Selectivity

In order to assess whether older adults had decreased frequency selectivity, a mixed effects model with group and frequency as fixed factors and listener as a random factor was conducted. The analyses revealed a main effect of frequency ($F_{(1,32)}$ =34.8, p< 0.001), with better frequency selectivity at 4 kHz than at 500 Hz as indicated by larger difference scores (notch minus no-notch) at 4 kHz. However, no significant main effect of group ($F_{(1,35)}$ = 0, p = 0.99) and no interaction ($F_{(1,32)}$ = 0.14, p = 0.7) were found. This implies that the older adults did not have broadened auditory filters in the range of normal hearing up to 4 kHz compared to the younger group (see Fig. 4.7).

4.3.5 Click ABR

Figure 4.8 shows the grand averaged ABRs for the young and older adults. To assess potential age-related reductions in ABR wave amplitudes, thought to reflect deafferentation of nerve fibres from the auditory nerve up to the lateral lemniscus, a mixed-effects model with age group (young, older) and wave (I, V) as fixed factors and

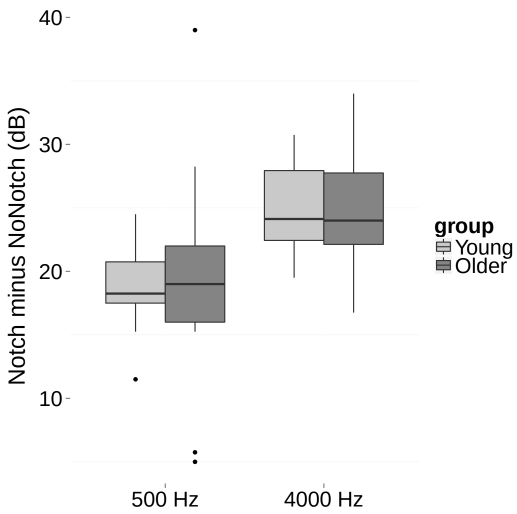


Figure 4.7: Boxplots of the difference measure (masked thresholds in notched noise minus broadband noise) of frequency selectivity for young (light grey) and older (dark grey) adults.

listener as a random factor was conducted.

The analysis revealed a significant effect of group ($F_{(1,33)} = 88.7$, $p = 0.03$, Cohen's $d = 0.47$), with lower amplitudes for the older adults. However, there was no interaction between group and wave ($F_{(1,33)} = 0.6$, $p = 0.43$), suggesting that ageing did not affect the amplitude of one wave more than another. Insofar as reduced wave amplitudes reflect fewer nerve fibres, these findings suggest that age-related deafferentation of nerve fibres affects neural processing from the auditory nerve up to the brainstem.

The question remains, however, whether the reduced amplitude of the click ABR is a consequence of ageing or hearing loss at the higher frequencies. To answer this question, a best subsets regression analysis was performed on the wave I and wave V amplitudes separately (Hastie et al., 2009). Age group, PTA across 0.5 - 4 kHz (region of normal hearing), and PTA across 6 - 8 kHz (hearing loss in the older group) were entered into the analyses as possible predictors.

Wave I appears to be best predicted by hearing thresholds up to 4 kHz, while wave V was best explained by PTA across 6 - 8 kHz, after accounting for age group (see table 4.4). It is tempting to conclude that the reduced amplitudes of waves I and V in the older group are attributable to different factors. However, the fact that one predictor comes out as significant in one model but not the other, does not necessarily mean that the regression coefficients are themselves significantly different between the two models. In other words, hearing thresholds across 0.5 - 4 and 6 - 8 kHz may equally predict wave I and wave V amplitudes, especially since the two PTA measures are highly correlated

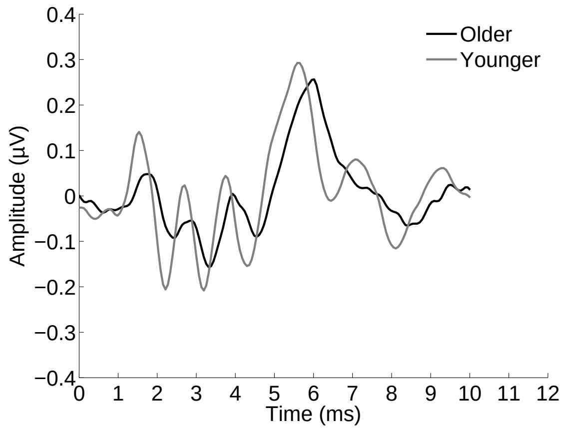


Figure 4.8: Click ABRs for young (grey) and older adults (black), averaged across all participants in the respective groups.

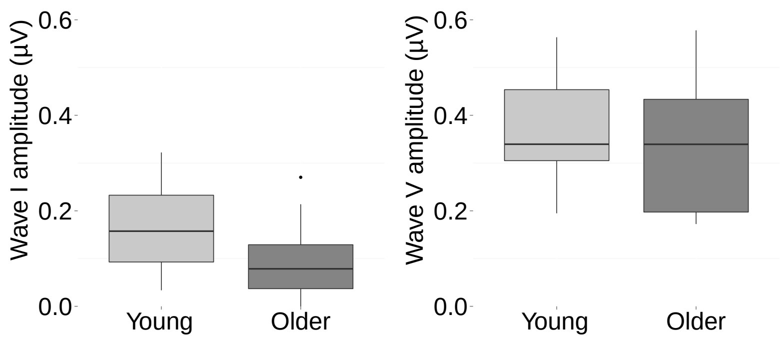


Figure 4.9: Boxplots of the amplitudes of click-evoked auditory brainstem response waves I (left) and V (right) for young and older adults.

with one another ($r = 0.7$, $p < 0.001$).

To answer the question whether hearing thresholds in the different frequency regions indeed differently affected wave I and V amplitudes, a multiple linear regression analysis was carried out on both ABR measures, with a dummy-coded predictor indicating the wave (i.e. wave I or V).

The results revealed a significant interaction for the PTA measure across 0.5 - 4 kHz, indicating that the slopes of this predictor were different for the two ABR waves. The best subsets analyses (table 4.4) would suggest that hearing thresholds up to 4 kHz can partly explain individual differences in wave I but not wave V amplitudes. Furthermore, the analyses revealed no significant interaction between PTA 6 - 8 kHz and ABR wave (table 4.5), suggesting that audiometric thresholds in the higher frequencies are equally important (or unimportant) for the two ABR response measures.

		Best subsets regression: click ABR					
Dependent variable	Predictors	b	β	SE	p	R^2	CI
Wave I amplitude	PTA 0.5-4 kHz	-0.007	-0.47	0.002	**0.004****	0.23	[-0.1, -0.002]
Wave V amplitude	age group	-0.17	-0.72	0.09	0.07	0.03	[-0.007, 0.2]
	PTA 6 - 8 kHz	-0.006	-0.97	0.002	**0.016***	0.16	[-0.01, -0.001]

Table 4.4: Results of the best subsets regression analyses on the STM and SRT data; * significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$). Note that β refers to the standardised regression coefficient. The R^2 reflects the proportion of the variance accounted for as predictors are added to the model. 95% confidence intervals (CI) for the regression coefficients are also given.

		Multiple linear regression: click ABR					
Dependent variable	Predictors	b	β	SE	p	R^2	CI
Interaction with wave	PTA 0.5-4 kHz	0.009	0.57	0.003	**0.006****	0.38	[-7.3, 0.01]
	PTA 6-8 kHz	0	0.04	0.001	0.8	0	[-2.4, 0.004]

Table 4.5: Results of the linear regression analysis on the ABR data, assessing whether PTA across 0.5-4 and 6-8 kHz differently affected wave I and V amplitudes (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$). Note that β refers to the standardised regression coefficient. The R^2 reflects the proportion of the variance accounted for as predictors are added to the model. 95% confidence intervals (CI) for the regression coefficients are also given.

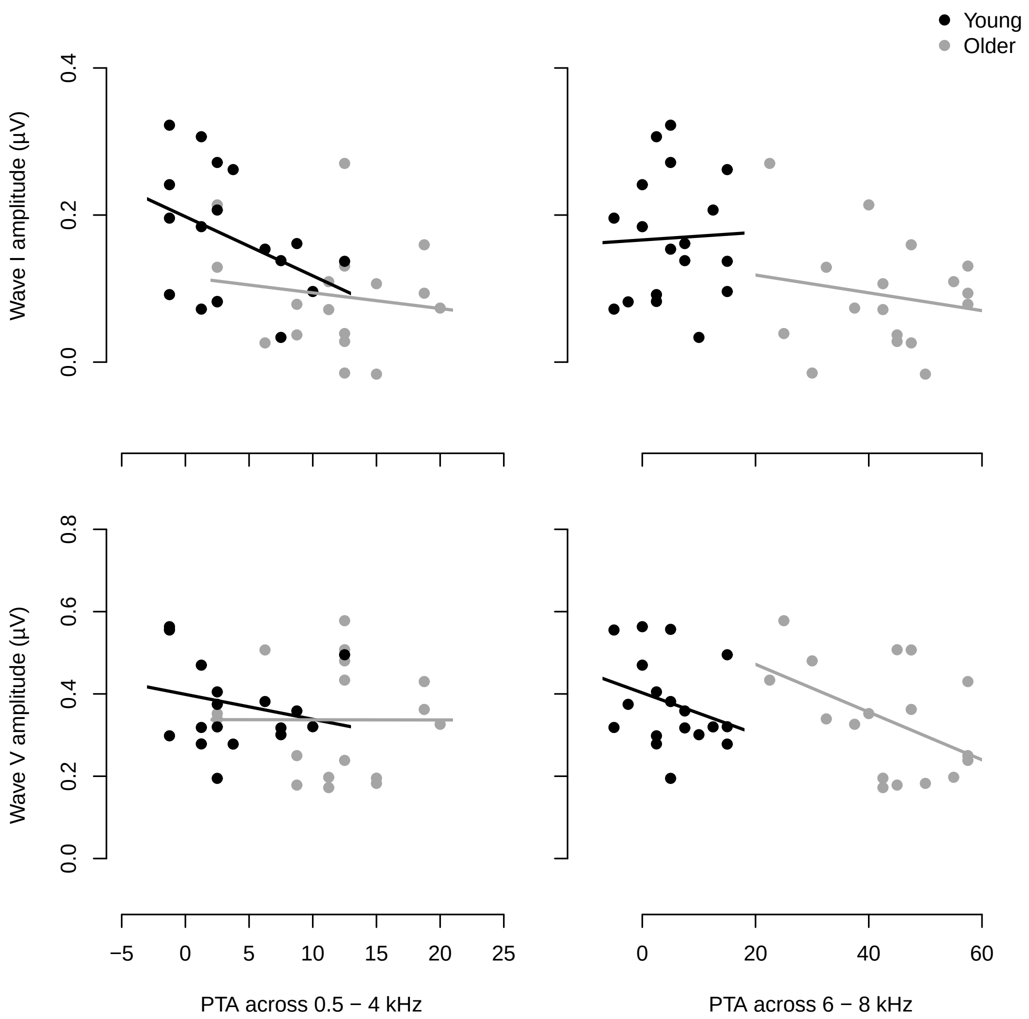


Figure 4.10: Scatter plots of ABR wave I and V amplitudes with PTA across 0.5 - 4 kHz (region of normal hearing) and 6 - 8 kHz (hearing loss in the older group), with young adults in black and older adults in grey.

These analyses show that it is difficult to determine whether the observed group differences in the ABR wave amplitudes are the result of differences in audiometric thresholds up to 4 kHz or above, or ageing. Since ageing is associated with elevated audiometric thresholds, it is difficult to tease apart the potential effects of ageing independent of age-related hearing loss.

4.3.6 Predicting age-related changes in SRTs and STM detection

Of primary interest was the question whether age-related changes in speech perception in noise and STM detection could be the result of a decrease in synchronous neural firing along the auditory pathway from the auditory nerve up to the brainstem. The results so far showed an age-related decline in both SRTs and STM detection thresholds as well as a reduction in neural firing from the auditory nerve up to the brainstem, as indicated by decreased wave I and wave V amplitudes.

Predicting STM detection thresholds: To answer the question whether the elevated STM thresholds for the older group were best explained in terms of deafferentation, over and above higher audiometric thresholds and reduced frequency selectivity, a best subsets regression analysis was performed (Hastie et al., 2009). Wave I and wave V amplitudes, PTA across 0.25 - 2 kHz, as well measures of frequency selectivity at 0.5 and 4 kHz were entered into the analyses as potential predictors. The best model was selected based on the Bayesian Information Criterion (BIC; Schwarz, 1978).

The analyses suggested that STM thresholds were best predicted by PTA across 0.25 - 2 kHz and frequency selectivity at 500 Hz, and not ABR wave amplitudes (see table 4.6 and Fig. 4.11). It should be noted, however, that the relationships between the STM detection thresholds and the PTA and frequency selectivity measures were mainly driven by one older adult (see Fig. 4.11). The relationships would not hold if this participant was removed from the data set.

Predicting speech perception in noise: Similarly, a best subsets regression was performed on the SRT data, with wave I and wave V amplitudes, PTA across 0.5 - 6 kHz, and measures of frequency selectivity at 0.5 and 4 kHz as possible predictors.

The results showed that SRTs were also best predicted by individual differences in audiometric thresholds (see table 4.6 and Fig. 4.11). Again, this suggests that the click ABR did not account for the variability in SRTs.

Individual differences: The group differences in STM detection thresholds and SRTs could not be explained satisfactorily by individual differences beyond the audiogram.

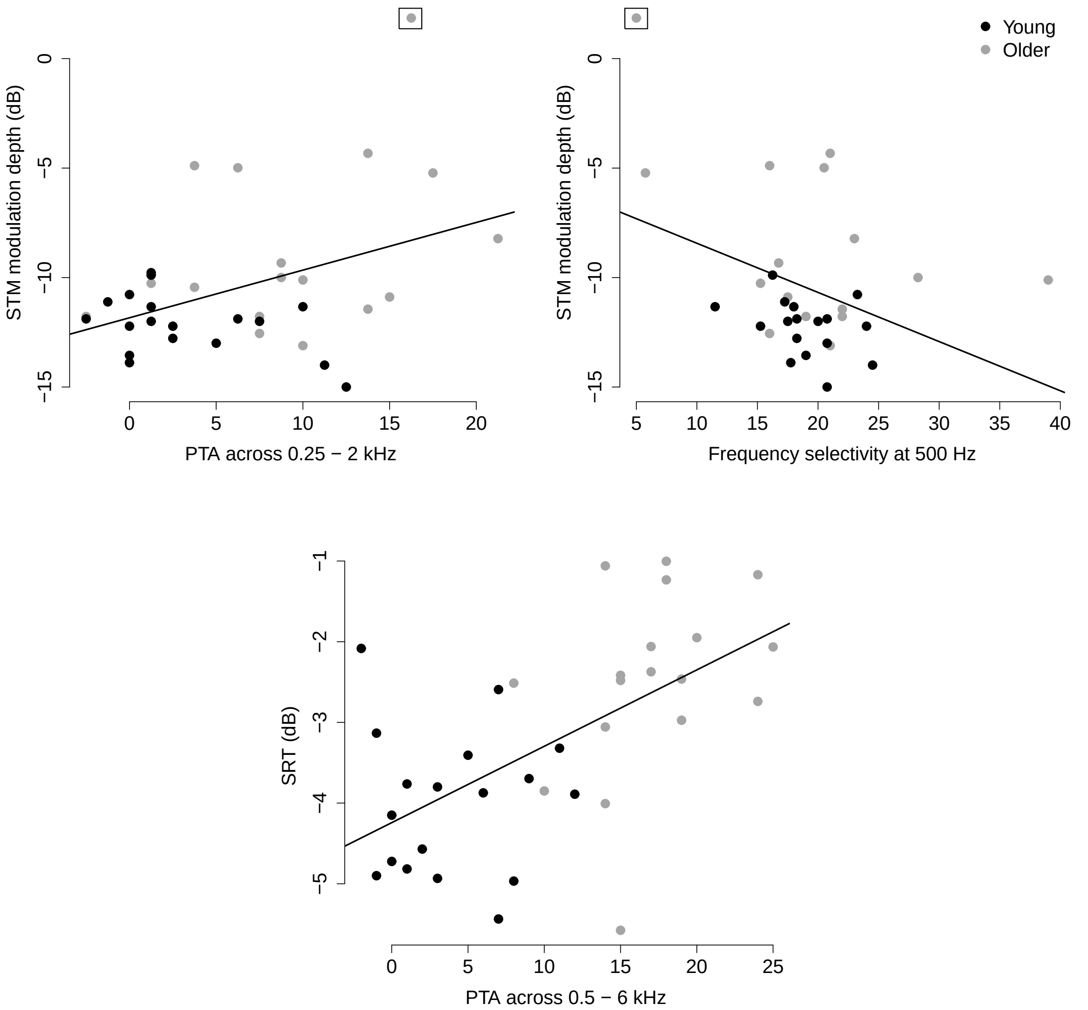


Figure 4.11: Scatter plots showing the relationship between the STM detection thresholds PTA across 0.25 - 2 kHz (top left) and frequency selectivity at 500 Hz (top right), and the relationship between SRTs and PTA across 0.5 - 6 kHz (bottom). Data points for the young adults are shown in black and for the older adults in grey. The two data points (from the same older adult) that drive the relationship between the STM detection thresholds and the PTA and frequency selectivity measures are indicated by a square.

		Best subsets regression: STM and SRT					
Dependent variable	Predictors	b	β	SE	p	R^2	CI
STM	PTA 0.25 - 2 kHz	0.22	0.39	0.08	**0.01***	0.17	[0.05, 0.4]
	Frequency selectivity at 500 Hz	-0.21	-0.36	0.9	**0.02***	0.13	[-0.4, -0.03]
SRT	PTA 0.5 - 6 kHz	0.09	0.59	0.02	<**0.001*****	0.35	[0.05, 0.14]

Table 4.6: Results of the best subsets regression analyses on the STM and SRT data (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$). Note that β refers to the standardised regression coefficient. The R^2 reflects the proportion of the variance accounted for as predictors are added to the model. 95% confidence intervals (CI) on the regression coefficients are also provided.

It could be, however, that *some* of the older adults who perform particularly poorly on either of those two tasks also perform more poorly on some of the other tasks.

To further explore some of these individual differences in the data set, a deviance analysis was performed (c.f. Ramus et al., 2003). First, all the individual measures were converted into z-scores on the basis of the measures from the group of young adults. The deviance threshold was then set to 1.65 sd above (for the SRT and STM measures) or below (for the ABR and frequency selectivity measures) the mean for the young group. In other words, the participants were identified who performed more poorly than the poorest 5% of a young population.

The results, illustrated in figure 4.12, indicate that two older adults had abnormally small ABR wave I amplitudes (E04 and E10). One of these participants (E10) also had a deviant ABR wave V amplitude, as well as abnormal STM and SRT scores. It is worth pointing out that the other participant (E04) with an abnormally small ABR wave I amplitude did not show especially reduced STM sensitivity or speech in noise performance.

The results furthermore show that one older participant (E19) had particularly reduced frequency selectivity, as a result of lower (i.e. poorer) thresholds in notched noise, both at 0.5 and 4 kHz, despite normal audiometric thresholds in that range (PTA 0.5 - 4 kHz: 18.75 dB HL). This participant also had the poorest STM sensitivity and showed abnormal speech in noise performance. This is in line with the idea that broadened auditory filters affect both STM detection and the ability to understand speech in noise (c.f. Léger, 2012).

Overall, however, the results show that while a relatively large group of older adults performed particularly poorly on the STM and SRT tasks, only a few older adults also had abnormal click ABRs or particularly reduced frequency selectivity. This is in line with the findings reported above, which showed that neither frequency selectivity nor ABR amplitudes predict poor STM sensitivity or speech in noise performance.

4.4 Discussion

This study set out to determine whether older adults have reduced STM sensitivity and whether this can to some extent explain the increased difficulties older adults typically experience understanding speech in noise. Furthermore, of primary interest was whether age-related changes in STM detection thresholds and SRTs could partly be attributed to deafferentation from the auditory nerve up to the brainstem, as measured by the click ABR.

The data indicate that STM sensitivity in the range of normal hearing indeed declines with age. STM detection thresholds for the older group were on average 3.5 dB higher than for the young adults. This age-related reduction in STM sensitivity is comparable to that reported for (older) hearing impaired individuals (about 3 - 5 dB; Bernstein et al., 2013; Mehraei et al., 2013).

It remains unclear why older adults perform more poorly on the STM detection task. Given that ageing has often been associated with a decline in TFS processing (Vongpaisal and Pichora-Fuller, 2007; He et al., 2007; Grose and Mamo, 2010; Hopkins and Moore, 2011; Füllgrabe, 2013), it can be speculated that the decreased STM sensitivity for the older adults may at least in part be attributed to a deficit in TFS processing. However, since TFS sensitivity was not assessed directly, this remains an open question. Instead, the reduced STM sensitivity may be the result of impaired envelope processing or frequency selectivity. It is unlikely, however, that impaired envelope processing contributes to poor STM sensitivity here since older adults typically only have problems detecting relatively high modulation rates (e.g. Purcell et al., 2004; Grose et al., 2009). The results showed that reduced frequency selectivity may explain reduced STM sensitivity at least for some older adults, but only those with particularly poor frequency selectivity.

The data furthermore showed that the older adults experienced increased difficulties understanding speech in noise. SRTs were higher (i.e. poorer) for the older group, albeit

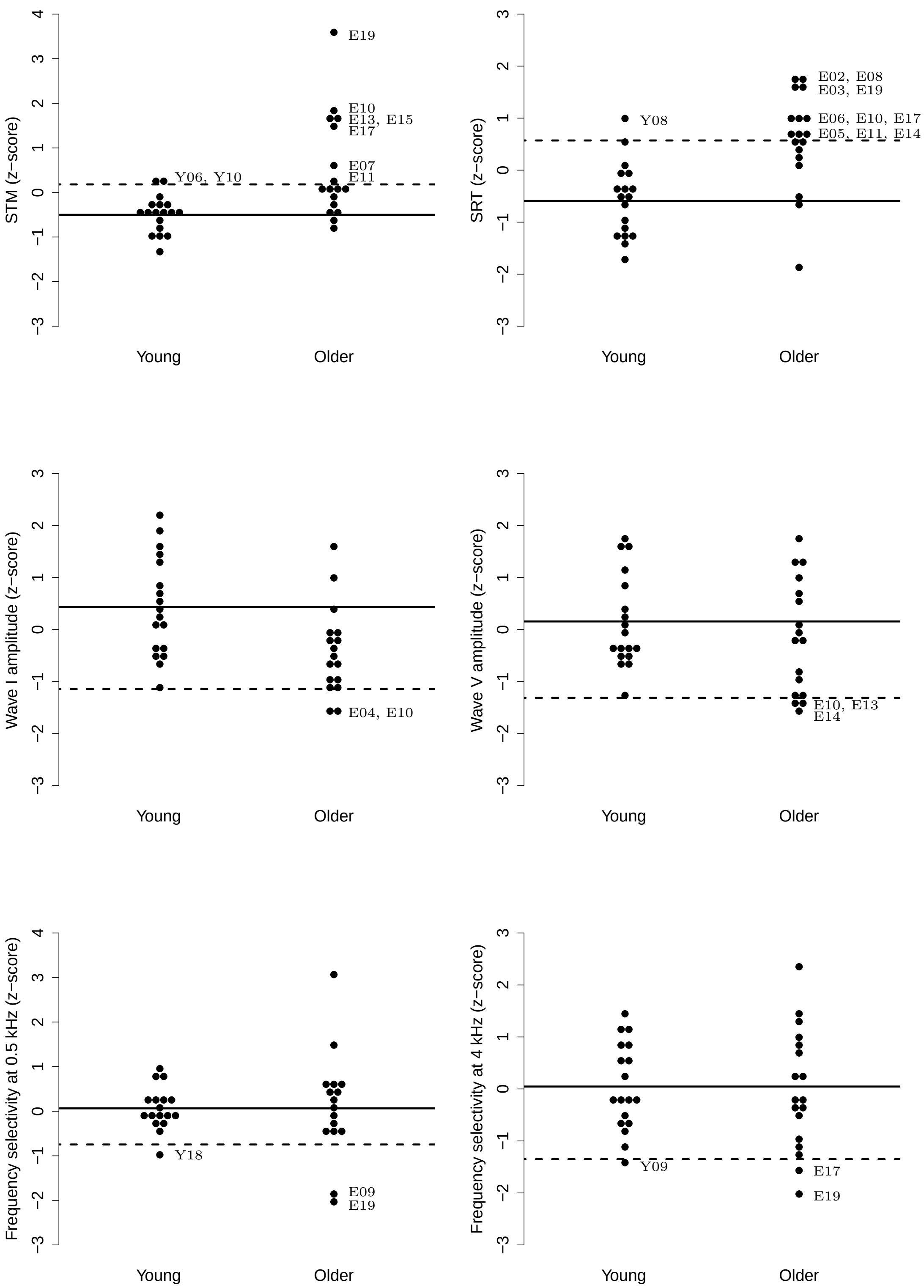


Figure 4.12: Individual z-scores for all measures. The solid line indicates the mean for the young adults and the dotted line indicates the deviance threshold (1.65 sd above or below the control mean). Deviant individuals are identified for both groups.

only by 1.5 dB. These increased speech in noise difficulties could not be explained, however, in terms of reduced STM sensitivity. Instead, individual differences in SRTs were best explained by differences in audiometric thresholds. This finding appears to contradict reports by Bernstein et al. (2013); Mehraei et al. (2013), who found that STM sensitivity accounted for 40 - 80 % of the variance within a group of hearing impaired listeners in SRTs in steady-state noise (using the same target stimuli) after taking differences in the SII into account. It should be noted, however, that the group difference in SRTs was small (1.5 dB). It may be the case that STM detection thresholds would predict inter-individual differences in SRTs over and above the audiogram for people who perform more poorly on a speech in noise task. Nonetheless, these findings suggest that listeners may be relatively unimpaired in speech in noise performance, despite a reduction in STM sensitivity. Further research is therefore required to assess the use of an STM-based model for to predict speech in noise performance, at least for people with relatively good hearing.

The central question was whether a loss of nerve fibres could explain these age-related changes in STM sensitivity and speech in noise performance. The data indeed show that the click-evoked ABR is reduced, suggesting a loss of functioning neurons from the auditory nerve up to the inferior colliculus (Kujawa and Liberman, 2009; Schaette and McAlpine, 2011; Sergeyenko et al., 2013). While research on ageing mice has found that the amplitude of wave I was more reduced than the amplitude of wave V (Sergeyenko et al., 2013), the present data revealed that both waves I and V were significantly reduced. This suggests a loss of neurons not only at the auditory nerve, but also at the inferior colliculus (c.f. Møller, 2007). This would imply no or limited increased central gain as a result of deafferentation at the auditory nerve (c.f. Schaette and McAlpine, 2011).

These results appear to contradict the findings reported for mice (Sergeyenko et al., 2013). However, it should be noted that Sergeyenko et al. (2013) did not statistically test the interaction between the ABR wave (I and V) and age group (young and older). It could thus potentially be the case that the age-related amplitude reductions for waves I and V were not significantly different from one another. This is unlikely, however, since Sergeyenko et al. (2013) also found age-related reductions in the wave V/wave I ratio. The present results appear to be in line with human data reported by Konrad-Martin et al. (2012) who measured click ABRs for 131 participants across the adult lifespan with a range of audiometric thresholds. They report age-related reductions for both waves I and V, with the largest absolute reduction found for wave V.

It should be noted that it remains unclear to what extent the reductions in ABR wave amplitudes reported here are the result of ageing or age-related hearing loss. Since clicks are presented at a high level (here 70 dB nHL) and by definition have a broadband spectrum, a large part of the basilar membrane contributes to the generation of the ABR. Don and Eggermont (1978) indeed showed, by presenting clicks in various high-pass noise maskers, that neurons along nearly the whole length of the basilar membrane are involved in the generation of the click ABR, especially for wave I although perhaps less so for wave V. It is thus likely that high-frequency hearing loss in part contributed to the reduced ABR wave amplitudes.

It was hypothesised that this loss of functioning nerve fibres, as indicated by reduced ABR wave amplitudes, could in part explain age-related reductions in STM sensitivity. The results indicated, however, that STM sensitivity could not be predicted by either ABR wave I or wave V amplitudes. Moreover, while deviance analysis indicated that two older adults had abnormally small ABR wave I amplitudes, only one of them also performed particularly poorly on the STM detection task. Inter-individual differences in STM detection thresholds were instead best predicted by audiometric thresholds. The results furthermore seem to suggest that people with particularly reduced frequency selectivity will likely also have especially poor STM sensitivity.

The finding that the click ABR could not predict STM detection thresholds could imply two things. First, the poor STM sensitivity for the older adults may not necessarily be the result of an age-related decline in TFS processing. Second, the reduction in the ABR amplitude may not necessarily cause a decline in the ability to process TFS information. The latter is perhaps the most likely given that a decline in TFS processing has repeatedly been shown to be associated with ageing (Vongpaisal and Pichora-Fuller, 2007; He et al., 2007; Grose and Mamo, 2010; Hopkins and Moore, 2011; Füllgrabe, 2013, but see chapter 2) and TFS processing has been suggested to be important for STM detection (Bernstein et al., 2013). However, these questions cannot be answered on the basis of the current data set since no measure of TFS processing was obtained.

It has furthermore been suggested that a loss of healthy neurons from the auditory nerve up to the inferior colliculus may to some extent explain speech in noise problems for older adults above and beyond the pure-tone audiogram (Kujawa and Liberman, 2009; Sergeyenko et al., 2013). The results show, however, that individual differences in SRTs could not be predicted by ABR wave I and wave V amplitudes. As already mentioned, SRTs were best predicted by audiometric thresholds across 0.5 - 6 kHz. Furthermore, the deviance analysis indicated that while some older adults with reduced ABR wave amplitudes performed particularly poorly on the speech in noise task, others with similarly deviant ABR amplitudes did not show especially poor SRTs. It is perhaps not surprising that the click ABR could not predict speech in noise performance as the latter is a complex process that relies on both bottom-up and top-down processing. The older adults may have been able to compensate to some extent for the deafferentation from the auditory nerve up to the brainstem.

In sum, the first aim of this study was to determine whether reduced STM sensitivity can in part explain increased speech in noise difficulties for older adults. The second aim was to examine whether a loss of functioning neurons from the auditory nerve up to the brainstem can explain reduced STM sensitivity and speech in noise performance. The results can be summarised as follows:

1. Older adults with normal hearing up to 4 kHz show reduced STM sensitivity in the range of normal hearing.

2. Age-related declines in STM sensitivity do not account for increased speech in noise problems. Older adults are relatively unimpaired in speech in noise performance despite reduced STM sensitivity.

3. Reduced click ABR wave I and wave V amplitudes do not predict age-related changes in STM sensitivity or the ability to understand speech in noise.

Chapter 5

The importance of context for dip listening

5.1 Introduction

One important factor that allows young normal-hearing listeners to understand speech in a considerable amount of background noise is their ability to 'listen in the dips' of fluctuating maskers. Their performance improves relative to steady-state maskers when the masker fluctuates in amplitude over time (e.g. Miller and Licklider, 1950; Wilson and Carhart, 1969; Festen and Plomp, 1990). The amount of release from masking can be anywhere from a few dB to as much as 20-30 dB depending on the target stimuli used and the temporal characteristics of the masker.

The results of the speech in noise task reported in chapters 2 and 3 only revealed a small fluctuating masker benefit (FMB) of on average 2.7 dB. This raised the question whether different stimulus parameters would have elicited a larger FMB. The primary aim of the present study was to explore the effects of sentence complexity and masker modulation type on the FMB.

It is well documented that the FMB largely depends on the characteristics of the fluctuating masker. It has been shown, for example, that the rate of modulation greatly affects the FMB. The modulation rate that results in the largest FMB, however, is dependent on the stimulus material, with relatively fast modulation rates (10-32 Hz) for sentences and slower modulation rates (1-10 Hz) for spondees and mono-syllables resulting in the largest masking release (c.f. Gustafsson and Arlinger, 1994; Miller and Licklider, 1950; Wilson and Carhart, 1969). Similarly, it has been found that larger modulation depths result in larger FMBs (Howard-Jones and Rosen, 1993; Wilson and Carhart, 1969).

The proportion of the speech signal that is relatively unaffected by the masker is clearly an important factor in successful speech in noise perception. Noise maskers modulated by the envelope of single or multiple talkers, for example, have been shown to result in a smaller release from masking compared to sinusoidally (SAM) or square-wave (sqAM) amplitude-modulated maskers (e.g. Bacon et al., 1998). This can largely be explained by the fact that, in comparison to SAM or sqAM maskers, speech-envelope modulated maskers contain glimpses that are often shallower, shorter in duration, and less predictable in terms of their temporal location. It can similarly be expected that, for a given modulation depth and rate, sqAM leads to a larger FMB than SAM as it leaves a larger proportion of the speech signal unaffected by the masker.

The FMB likely depends not only on the characteristics of the fluctuating masker, however, but also on the complexity of the target signal and the redundancy of the information available in the amplitude dips of the masker. In other words, it may in part depend on the availability of contextual cues in the dips of the masker. Speech perception in noise in general indeed improves when the listener can make use of contextual information (Miller et al., 1951; Kalikow et al., 1977). It remains unclear, however, to what extent the FMB depends on the availability of contextual cues. It has been shown that the FMB for the perception of words is independent of manipulations of response-set size (Bernstein and Brungart, 2011; Bernstein et al., 2012), suggesting that context does not affect the FMB. However, sentential context may aid the ability to fill in missing information and integrate glimpses of the target signal.

A secondary aim was to examine whether linguistic closure, the supramodal linguistic ability to fill in missing information, is equally important for the perception of speech in steady-state and fluctuating maskers. It has been suggested that top-down cognitive processing may be particularly important for the perception of speech in fluctuating maskers (Akeroyd, 2008; Rönnberg et al., 2010). Successfully integrating the information extracted from glimpses of the speech signal contained in the amplitude dips of the fluctuating masker may rely more heavily on the ability to fill in missing information. Higher correlations have indeed been found for SRTs in fluctuating than steady-state noise with a measure of linguistic closure, as measured by the Text Reception Threshold (TRT; George et al., 2007; Zekveld et al., 2007; Besser et al., 2012). However, despite apparent differences in correlation coefficients for the different SRT measures, the correlation coefficients may in fact not have been significantly different. In other words, the ability to fill in missing information may be equally important for the perception of speech in steady-state and fluctuating maskers. Linguistic closure may be important for the perception of speech in steady-state noise because glimpses are to some extent available even in steady-state maskers since speech itself varies in level, causing the SNR to vary.

This study examined the effects of sentence complexity and masker modulation type on the FMB by comparing the FMB for BKB and IEEE sentences (Rothauser et al., 1969; Bench et al., 1979) in sqAM and SAM noise. The FMB was expected to be larger for the BKB sentences (e.g. The clown had a funny face) since they are not only shorter, but are also less complex and contain more contextual information than the IEEE sentences (e.g. The birch canoe slid on the smooth planks). In addition, the importance of the ability to fill in missing information, as measured by the TRT, for the perception of speech in steady-state and amplitude-modulated maskers was assessed.

5.2 Methods

5.2.1 Participants

Twenty normal-hearing participants (age range 18 - 47 yrs, mean 26 yrs, 11 female) took part in this study. All were native British English speakers. Normal hearing was defined as pure-tone thresholds of 20 dB HL or better at octave frequencies between 250 and 8000 Hz. None of the participants reported a history of language or neurological disorders. Participants signed a consent form approved by UCL Research Ethics Committee and were paid for their participation.

5.2.2 Speech reception threshold

Speech reception thresholds (SRTs) were measured for sentences in steady-state speech-shaped noise (SS) that matched the long-term average spectrum of the sentence materials, 10-Hz square-wave amplitude-modulated speech-shaped noise (sqAM), and 10-Hz sinusoidally amplitude-modulated speech-shaped noise (SAM; see Rosen et al. (2013) for a description of the speech-shaped noise). The modulation depth was 100% for both modulated maskers. Two different types of sentence materials were used; IEEE and BKB sentences (Rothauser et al., 1969; Bench et al., 1979). The sentences were read by the same male Southern British English speaker. IEEE sentences each contained five key words, while the shorter BKB sentences contained three key words each. The masker always started 500 ms prior to the target sentence and was gated on and off across 100 ms.

Stimuli were presented binaurally at 70 dB SPL using an external soundcard (RME Babyface). The participants were seated in a soundproof booth and listened to the stimuli over Sennheiser HD 650 headphones. They were asked to repeat what they heard and the experimenter scored correctly repeated key words using a graphical user interface (GUI). The scoring screen was not visible to the participants and no feedback was provided.

The SNR was varied adaptively following the procedure described by Plomp and Mimpen (1979). The first sentence was presented at an SNR of -10 dB. Unless all key words were correctly repeated, the SNR was increased by 6 dB on the next presentation. The initial sentence was repeated until all key words were repeated correctly or the SNR reached 18 dB. The step size was decreased to 4 dB on the following sentence. For each subsequent sentence, the SNR increased by 2 dB when zero to two (IEEE) or zero to one (BKB) key words were correctly repeated or decreased by the same amount for three to five (IEEE) or two to three (BKB) correct repetitions, thus tracking 50% in both cases. The number of trials was fixed to twenty.

SRTs for each condition were measured twice. A measurement was repeated when fewer than three reversals were obtained or when the standard deviation across the reversals exceeded 4 dB at the minimum step size. Thresholds for each run were computed by taking the mean SNR (dB) across the final number of reversals. The SRTs reported here are the mean across the two runs.

Participants were given brief training on sentences in SS, followed by sentences in sqAM, with one of the blocks using IEEE sentences and the other block using BKB sentences. Practice consisted of three sentences and started at 0 dB SNR. The order of conditions in the experiment proper was counterbalanced across participants following a Latin square design.

5.2.3 Text reception threshold

The text reception threshold (TRT) is a visual analogue of the SRT (Zekveld et al., 2007; Besser et al., 2012). This task was developed to measure the variance in speech perception in noise abilities that is associated with modality general cognitive and linguistic skills. In this task, sentences that are partly masked by a vertical bar pattern are presented on a computer screen. The degree of masking varies from trial to trial by varying the width of the bars.

As in the speech perception in noise task (measuring SRTs), the target stimuli were both IEEE and BKB sentences (Rothauser et al., 1969; Bench et al., 1979). While the target stimuli were taken from the same corpus, the specific sentences used in the two tasks were different. The participants were seated approximately 50 cm from the screen.

They were asked to read the sentence out loud. The experimenter scored responses using a GUI which showed all the words in the sentence. The scoring screen was not visible to the participants and no feedback was provided.

The degree of masking was varied adaptively following the procedure described by Plomp and Mimpen (1979), tracking 50% correct. The first sentence was presented with 16% unmasked text. Until the sentence was correctly repeated, the percentage of unmasked text was increased by 12% on the next presentation. Subsequent sentences were only presented once. When a sentence was correctly repeated, the degree of masking was increased by 6%. Conversely, the degree of masking was decreased by 6% when a sentence was not repeated correctly.

TRTs were measured in response to two lists of twenty sentences each for each sentence type (BKB and IEEE). Data were not included when TRTs across two trials differed by more than 8%. The thresholds reported here are the mean of the two trials and are expressed as percentage of unmasked text.

5.3 Results

The primary aim of this study was to determine whether the amount of fluctuating masker benefit (FMB) differs according to sentence type (IEEE and BKB) and type of amplitude modulation of the masker (square wave and sinusoidal). In addition, the effect of different sentence materials on the TRT was assessed, as well as the relationship between the TRT and SRT.

Outliers were excluded from the analyses if they exceeded the mean $\pm$ 3 sd. Based on this criterion, three outliers were excluded from the SRT data (two for BKB sqAM, one for IEEE sqAM). Two TRT data points for the IEEE sentences were excluded because the thresholds across the runs differed by more than 8%.

5.3.1 Speech reception thresholds

Table 5.1 shows the mean SRTs (and standard deviations) for the two sentence materials (BKB and IEEE) in SS, sqAM, and SAM noise.

The FMB was calculated by subtracting SRTs in sqAM and SAM from SRTs in SS noise for each sentence type separately. It has been suggested, however, that the FMB may partly be dependent on the SRT in SS noise (Oxenham and Simonson, 2009; Bernstein and Grant, 2009). If this is indeed the case, a potential difference in FMB in terms of sentence type could not necessarily be attributed to differences in the usefulness of glimpses but could in part result from inherent differences in sentence difficulty.

SRTs (dB) - Mean (sd)					
BKB			IEEE		
SS	sqAM	SAM	SS	sqAM	SAM
-6.0 (0.9)	-19.2 (3.2)	-12.6 (1.6)	-4.9 (1.1)	-15.4 (2.8)	-10.1 (1.7)

Table 5.1: SRTs for the two different sentence materials (BKB and IEEE) in steady-state (SS), square-wave amplitude-modulated (sqAM), and sinusoidally amplitude-modulated noise (SAM).

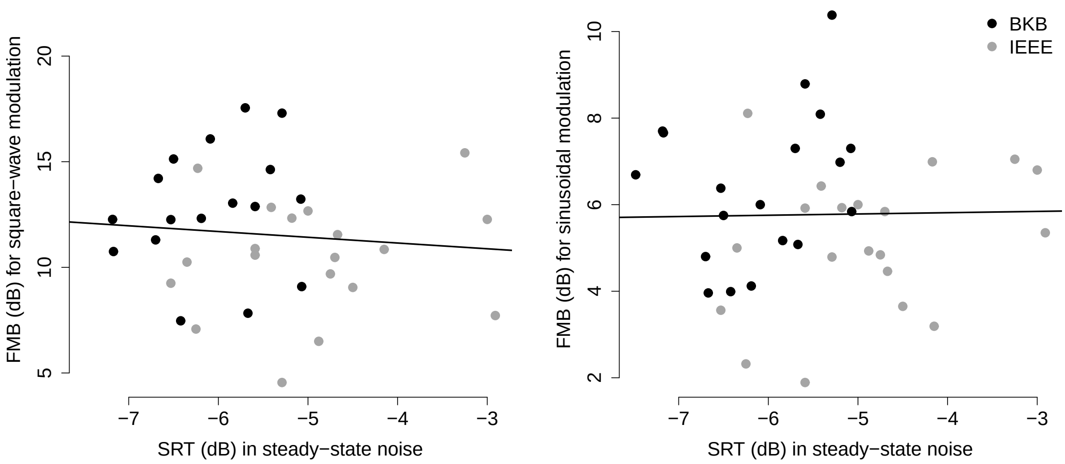


Figure 5.1: Scatter plots illustrating the relationship between the FMB and the SRT in steady-state noise for BKB (black) and IEEE (light grey) sentences, for square-wave (left) and sinusoidally (right) amplitude-modulated noise. Note the different scales on the y-axes.

To examine whether FMB across sentence type could be compared directly, a paired t-test on SRTs in SS was conducted. The results revealed a significant difference between BKB and IEEE sentences ($t_{(18)} = -5.4$, $p< 0.001$, Cohen's d = 1.2, mean difference = 1 dB) suggesting a potential confound of the baseline SRT at which the FMB is estimated. To examine whether FMB was in fact dependent on the baseline SRT, two Pearson's correlations were performed on SRTs in SS and FMB for sqAM and SAM (pooled across the two sentence materials). The results (see Fig. 5.1) showed that the FMB for sqAM and SAM was not dependent on the SRT in SS noise (sqAM: r = -0.095, p = 0.58; SAM: r = 0.018, p = 0.92), which meant a direct comparison of FMB across sentence materials was justified.

Figure 5.2 shows the amount of masking release derived using sqAM and SAM maskers for BKB and IEEE sentences. The figure indicates that listeners benefited

	Descriptive statistics		
Dependent variable	mean (sd) BKB	mean (sd) IEEE	CI
TRT	56 (4.2)	61 (3.4)	[3.5, 7.0]
FMB	9.8 (4.4)	7.7 (3.5)	[5.3, 8.1]
	mean (sd) sqAM	mean (sd) SAM	CI
FMB	11.8 (3.4)	5.9 (1.9)	[5.2, 6.9]

Table 5.2: Descriptive statistics (mean and standard deviation) for the TRT and FMB measures for the BKB and IEEE sentences (top) and masker types (bottom) separately. Confidence intervals for the differences between the sentences (top) or masker types (bottom) are also provided.

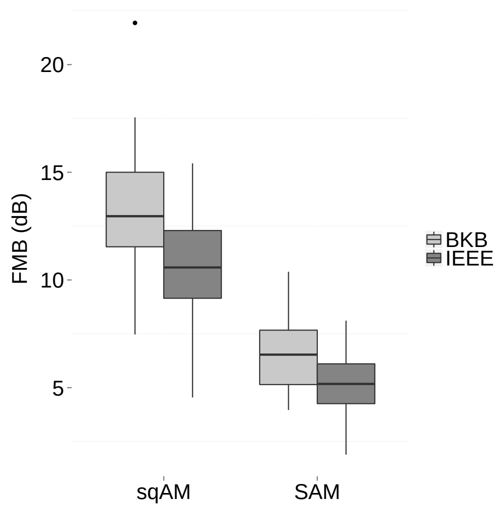


Figure 5.2: Boxplots of the FMB for square-wave (left) and sinusoidally (right) amplitude-modulated speech-shaped noise. Results for BKB and IEEE sentences are shown in light and dark grey respectively.

more from square-wave modulations than sinusoidal modulations in the masker, as indicated by a larger masking release for the sqAM condition. In addition, listeners experienced a larger masking release for the BKB sentences, compared to the longer and more complex IEEE sentences.

This pattern was confirmed by a mixed effects model with sentence material and masker type as fixed effects and listener as a random effect. The model showed a main effect of masker ($F_{(54,1)} = 238$, $p < 0.001$, Cohen's $d = 2.7$) and a main effect of sentence type ($F_{(54,1)} = 29$, $p < 0.001$, Cohen's $d = 1.2$). There was no significant interaction ($F_{(54,1)} = 3$, $p = .07$), which means that the difference in masking release in terms of type of modulation was not more prominent in one sentence type or another.

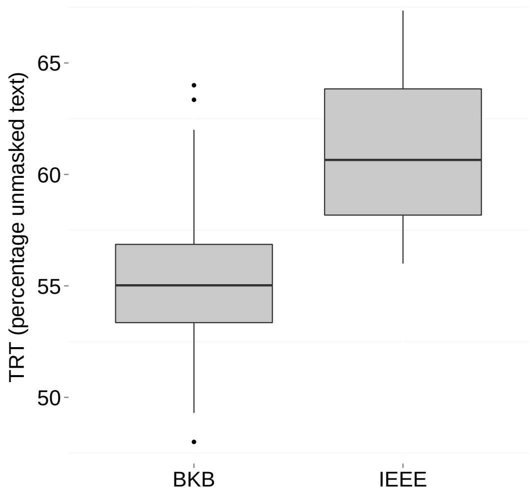


Figure 5.3: Boxplots of TRTs (percent unmasked text) for BKB (left) and IEEE (right) sentences. Note that lower TRT values indicate better performance.

5.3.2 Text reception thresholds

To assess whether TRTs were different across sentence type, a paired t-test was conducted. The results showed that participants could read the BKB sentences with significantly less unmasked text available to them than the longer and more complex IEEE sentences ($t_{(17)} = 6.3$, $p < 0.001$, Cohen's $d = 1.5$). The results are illustrated in figure 5.3.

5.3.3 Relationship between TRT and SRT

Given the relatively large number of possible correlations that could be performed to assess the relationship between the TRT and SRT measures, principal components analyses (PCA) were performed on the SRT and TRT data separately to reduce the number of measures. The PCA used varimax rotation with Kaiser normalisation and missing data points were replaced by the mean. The Kaiser criterion (eigenvalues >1) was used to determine the number of PCs to be extracted. The resulting principal components (PC) were saved as Anderson-Rubin scores.

PCA on the SRT data (SRTs for BKB and IEEE sentences in SS, sqAM, and SAM noise) resulted in the extraction of two components. The PCs were interpreted as reflecting SRTs in AM noise (PC SRT AM) and SS noise (PC SRT SS). Together the two components accounted for 76.3 % of the variance in the data, with PC SRT AM accounting for 44.6% and PC SRT SS accounting for 31.6 % of the variance (see table 5.3).

		Principal Components Analysis	
		PC SRT AM	PC SRT SS
SRT	BKB SS	0.006	**0.86**
	BKB sqAM	**0.75**	**-0.49**
	BKB SAM	**0.84**	0.16
	IEEE SS	0.21	**0.88**
	IEEE sqAM	**0.91**	0.05
	IEEE SAM	**0.73**	0.36
		PC TRT	
TRT	BKB	**0.83**	
	IEEE	**0.83**	

Table 5.3: Factor loadings for each of the SRT (top) and TRT (bottom) measures. Factor loadings >0.4 are highlighted in bold font.

Similarly, a PCA on the TRT data was performed. As shown in table 5.3, the analysis resulted in the extraction of a single TRT measure, accounting for 68.8 % of the variability in the data.

The PCs were used as a guiding principle for data reduction only. Following the PCAs, simple averages were computed for the SRTs in AM and SS noise and the TRT. These averages were used in the subsequent analyses.

To answer the question whether TRTs could predict listeners' abilities to understand speech in different types of background noise, two simple linear regressions were performed on the averaged SRT measures in SS and AM noise with the averaged TRT measure as a predictor. The results suggest that performance on the TRT task was significantly correlated only with people's abilities to understand speech in fluctuating noise but not in steady-state noise (see table 5.4 and Fig. 5.4). However, the relationship went in the opposite direction from what would normally be expected. Poorer performance on the TRT task was associated with better SRTs.

It should be noted that, while the correlations appear to suggest that TRTs are associated with SRTs in one type of noise but not the other, this difference cannot directly be assessed through significance levels in two separate analyses (Gelman and Stern, 2006). To assess whether the regression coefficients in the two models were indeed

		Regression on SRTs				
Dependent variable	Predictor	b	SE	p	R^2	CI
SRT SS	TRT	0.12	0.07	0.1	0.16	[-0.26, 0.27]
SRT AM	TRT	-0.42	0.16	$\mathbf{0.02}^{*}$	0.35	[-0.75, -0.08]
Interaction with SRT	TRT	0.08	0.004	$<\mathbf{0.001}^{***}$	0.9	[0.07, 0.08]

Table 5.4: Top: Results of simple linear regression analyses on the averaged SRTs in steady-state (SRT SS) and amplitude-modulated noise (SRT AM). Bottom: Results of the multiple regression analysis on both SRT measures with an additional dummy-coded predictor indicating the type of background noise (i.e. steady-state or amplitude-modulated noise) assessing whether the slopes of the predictor (TRT) differed depending on the type of background noise. The R^2 here reflects the proportion of the variance accounted for as the interaction (TRT $\times$ SRT dummy) is added to the model. Significant effects are highlighted in bold font (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$). 95% confidence intervals (CI) for the regression coefficients are provided.

significantly different, a multiple linear regression was performed on both SRT measures with an additional dummy-coded predictor indicating the type of background noise (i.e. SS or AM). Table 5.4 shows that the interaction between the dummy-coded variable and the TRT significantly predicted SRTs, supporting the idea that the relationship between TRT and SRT differs depending on the masker type.

5.4 Discussion

This study set out to explore the effects of sentence complexity and masker modulation type on the FMB. A secondary aim was to assess whether the supramodal linguistic ability to fill in missing information is equally important for the perception of speech in steady-state and fluctuating maskers.

First, the data showed a larger FMB for BKB than IEEE sentences. Given that the BKB sentences contain more contextual information, this finding implies that the availability of contextual cues leads to a larger FMB. Furthermore, since sentence complexity affected SRTs in fluctuating noise to a greater extent than in steady-state noise, it can be concluded that contextual cues are more beneficial for the perception of speech in fluctuating than stationary maskers. It should be pointed out, however, that these effects may simply be due to a difference in complexity across sentence types *per se* instead of differences specifically in the degree of available contextual information.

The data furthermore reveal a larger FMB for sqAM than SAM maskers. This can

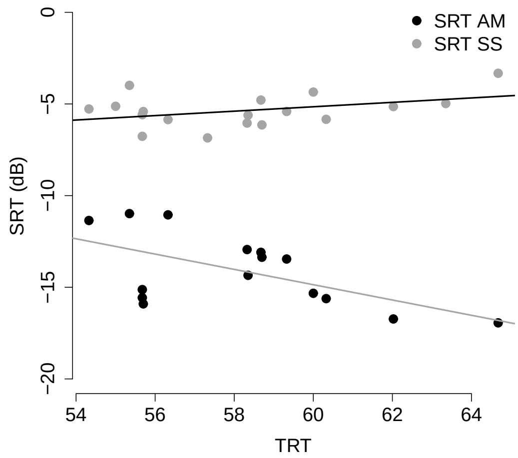


Figure 5.4: Scatter plot illustrating the relationship between the averaged TRT and SRT measures, with SRTs in steady-state noise in grey and SRTs in amplitude-modulated noise in black. Note that better (i.e. lower) SRTs are associated with poorer (i.e. higher) TRTs.

simply be explained by the fact that, at the same modulation rate and depth, sqAM results in a larger proportion of the speech signal being unaffected by the masker. However, the advantage of sqAM over SAM is not more prominent for one sentence type or another. In other words, a relative increase in the size of a glimpse does not aid the use of contextual information.

Lastly, the supramodal ability to fill in missing information, as measured by the TRT, correlated significantly with the perception of speech in amplitude-modulated but not steady-state noise. However, the relationship was going in the wrong direction, with better linguistic closure associated with poorer SRTs. This contradicts previous studies which indeed reported higher correlations with the TRT for SRTs in fluctuating compared to steady-state maskers, but with poor TRTs associated with poor SRTs (e.g. George et al., 2007; Besser et al., 2012). These findings cannot readily be explained. One possibility may be that the two TRT runs collected for each participant were too variable, leading to a poor estimate of linguistic closure ability.

In sum, the results from this study showed that sentence complexity affects speech perception in fluctuating noise to a greater extent than in steady-state noise. These findings suggest that the ability to use contextual information plays a greater role in dip listening than in the perception of speech in a steady state noise. It remains unclear, however, whether linguistic closure skills play a more important role in steady-state or modulated maskers.

Chapter 6

Rapid FFR: A new technique to rapidly collect the frequency following response

6.1 Introduction

Frequency following responses (FFRs) are typically recorded to stimuli with a duration of 40 - 170 ms and sometimes even up to as much as 2 seconds (e.g. Russo et al., 2004; Dajani et al., 2005). Since the acquisition of a robust response requires approximately 3000 repetitions of the stimulus (Skoe and Kraus, 2010; Jeng et al., 2011), recording the FFR can take quite some time. The recording time for the FFRs described in chapter 3, for example, was ten minutes per polarity. This resulted in a session of almost three hours (including breaks). It can be quite taxing, especially for older adults, to sit still for such a long time. This chapter therefore proposes a new technique to more quickly collect the FFR. This technique, the rapid FFR, significantly reduces recording times by presenting the stimulus continuously (i.e. without interstimulus interval) and averaging across a single cycle of the response.

An important question that needs to be answered is how many cycles are sufficient for the rapid FFR to appear above the noise floor. Jeng et al. (2011) showed that response properties of the FFR, such as pitch strength and signal-to-noise ratio, change exponentially with increasing numbers of sweeps. The number of required sweeps (or cycles of the response) can be calculated post-hoc according to the following formula:

$$A_n = A_\infty(1 - e^{-n/\tau}) - A_{noise} \tag{6.1}$$

where A is a response property of the FFR, n is the number of sweeps (or cycles in the case of the rapid FFR), A_∞ indicates the asymptotic amplitude of the fitted curve, τ is refers to the number of sweeps required to reach 63% of the asymptotic amplitude, and A_{noise} indicates the size of the response measure A when the number of sweeps equals one.

A larger number of sweeps, or response cycles, may be required to reliably record the rapid FFR than the standard FFR. Since the rapid FFR involves prolonged stimulus presentation, it is not unlikely that there is a reduction in neural firing phase-locked to the stimulus over time. Neural adaption may thus affect how the response properties of the rapid FFR develop with increasing numbers of sweeps.

The aim of this study was to determine whether the FFR gives the same results when acquired with the rapid as with the standard technique. More specifically, the question was whether the rapid FFR is equally sensitive to inter-individual differences as the standard FFR. First, however, the stimulus duration required to collect a robust response was estimated.

6.2 Methods

6.2.1 Participants

Sixteen young normal-hearing listeners (20 - 29 yrs, 9 female) participated in this study. Normal hearing was defined as pure-tone thresholds of 20 dB HL or better at octave frequencies between 250 and 8000 Hz and a normal click ABR, defined as wave V latency within 3 sd around the mean for a 100 μs click at 70 dB nHL (107.6 dB peSPL) with a repetition rate of 11/s (c.f. Campbell et al., 1981). None of the participants reported a history of neurological disorders. Participants were paid for taking part and gave informed consent approved by the UCL/UCLH Committee on the Ethics of Human Research.

6.2.2 Recording parameters

6.2.2.1 Stimuli

FFRs were recorded in response to a harmonic complex (sawtooth) with a fundamental frequency (F0) of 128 Hz. The F0 was a divisor of the sampling rate at which the FFRs were collected (16384 Hz), thus ensuring that the exact timing of each F0 cycle in the continuous EEG signal could be determined with one sample point precision. This degree of precision is required when averaging across single, continuously presented F0 cycles.

Two versions of the stimulus were created: one with a duration of 54 ms (seven F0 cycles) for the standard FFR measurement, and one with a duration of approximately 58 s (7503 F0 cycles) for the rapid FFR measurement. In both cases the stimulus was tapered on and off across 7.8125 ms (one F0 cycle).

6.2.2.2 Procedure

Participants were seated in a reclining chair in an electrically shielded soundproof booth and were allowed to fall asleep. FFRs were collected using a BioSemi ActiveTwo system (Amsterdam, The Netherlands). FFRs were referenced using a vertical electrode montage (Cz - seventh cervical vertebra, C7), which is thought to reflect neural activity from the rostral brainstem (i.e. the inferior colliculus and lateral lemniscus; Stillman et al., 1978; Galbraith, 1994). Two additional electrodes, Common Mode Sense (CMS) and Driven Right Leg (DRL), were used. In the BioSemi ActiveTwo system these two electrodes replace the ground electrode. Electrode offsets were always $<$40 mV and responses were recorded with a sampling rate of 16384 Hz.

The stimuli were generated in MATLAB (Mathworks, Natick, MA). The MATLAB procedure created stimuli on one channel and triggers on another channel. Both channels were delivered via a computer with an external sound card (RME FireFace, UC) connected to a custom-made trigger box. This trigger box separated the two channels and simultaneously sent the stimulus to electrically shielded ER-3 inserts (Intelligent Hearing Systems, Miami, FL) and the trigger to the EEG machine.

Responses were recorded using two different methods; the standard and rapid techniques. Following the standard recording technique, FFRs were recorded in response to 1500 sweeps of positive and negative polarities separately. Stimuli were presented with a repetition rate of 10/s, corresponding to an interstimulus interval of 45 ms. Following the rapid FFR technique, FFRs were recorded to the continuous stimulus presented in positive and negative polarity separately. Stimuli were presented binaurally at 80 dB SPL.

FFRs were measured twice using both recording techniques. The order of conditions (i.e. standard and rapid FFR) was alternated across participants.

6.3 Estimating required stimulus duration

In order to estimate the number of cycles required to collect a robust FFR using the rapid technique, the discriminability of the spectral components of the FFR with respect to the noise floor was determined for eight stimulus durations ranging from 500 to 7500 F0 cycles in 1000-cycle steps. To allow for a comparison with the standard technique, the extent to which the spectral components of the standard FFR appeared above the noise floor was also calculated.

The spectral noise floor at F0 and its subsequent harmonics was computed, using a permutation technique, by randomising the timing of the epochs and averaging across epochs by sampling with replacement. The same epoch could thus be included in the average more than once and individual epochs could overlap. To get a good estimate of the noise floor, this process was repeated 500 times, giving a distribution of the spectral noise floor of the response. Similarly, distributions were also computed for the spectral components of the FFR itself. In this case, the timing of the epochs was phase-locked to the stimulating waveform.

The following steps were undertaken for each participant, condition (i.e. standard and rapid FFR), and run separately:

1. The continuous EEG signals recorded in response to the positive and negative polarity stimuli were band-pass filtered between 70 and 2000 Hz (12 dB/octave, zero phase-shift).

2. For the standard FFR, 1500 100-ms epochs were sampled with replacement from the continuous EEG signal. For the rapid FFR, the total duration of the EEG signal was first capped at 500 to 7500 F0 cycles, in 1000-cycle steps. Subsequently, 500 to 7500 epochs with a duration of 7.8125 ms (one F0 cycle), in 1000-epoch steps, were drawn with replacement from the EEG signal. The number of epochs was always the same as the total number of F0 cycles contained in the capped EEG signal. The timing of the epochs, for both the standard and rapid FFRs, was either phase-locked to the stimulus (signal) or randomised (noise).

3. The standard FFR epochs were baseline corrected with respect to the 15 ms pre-stimulus response. No baseline correction was performed for the rapid FFR epochs since the stimuli were presented continuously and therefore no pre-stimulus response (i.e. baseline) was present. Any epochs exceeding ± 25 μV were rejected.

4. The epochs were averaged. The resulting FFRs in response to the positive and negative polarity stimuli were subsequently both added and subtracted. Adding polarities enhances the lower frequency components of the response and eliminates the stimulus artifact and any contributions from the cochlear microphonic (Gorga et al., 1985). Subtracting polarties, on the other hand, accentuates the higher frequency components (Aiken and Picton, 2008). While the opposite-polarity responses were only added and not subtracted in the study described in chapter

3, responses in the present study were also subtracted to fully evaluate the new technique.

5. Spectral magnitudes were calculated for the signal and noise FFRs using a fast Fourier transform (FFT) at F0 and its subsequent two harmonics for the added response and at the fourth and fifth harmonics for the subtracted response. The spectral magnitudes for the signal FFR recorded using the standard technique were calculated across the part of the response locked to the stimulus only (not including the first and last F0 cycles as these were tapered) and not across the interstimulus interval. FFTs for the standard noise FFR were computed across a section of the FFR with the same duration as the standard signal FFR (i.e. 39 ms). FFTs for the rapid FFR were calculated on the single cycle.

6. Steps 2 - 5 were repeated 500 times to obtain a distribution of the estimated spectral magnitudes for the signal and the noise floor.

Subsequently, d' scores were calculated as a measure of the discriminability of the signal with respect to the noise floor for the spectral magnitudes at the fundamental and the second and third harmonic for the added-polarity FFR and at the fourth and fifth harmonics for the subtracted-polarity FFR. The d' scores indicated the separation between the signal and the noise distributions and were computed using the following formula:

$$d' = \frac{\mu_s - \mu_n}{\sqrt{0.5 \times (\sigma_s^2 + \sigma_n^2)}} \tag{6.2}$$

where μ_s, μ_n, σ_s, and σ_n are the mean and standard deviation of the signal and noise distributions respectively. The d' scores were then averaged across the two runs.

Next, an exponential curve was fit to the change in d' scores with increasing numbers of sweeps or F0 cycles, following equation 6.1.

Two approaches were taken to estimate the number of sweeps, or F0 cycles, required to collect a robust FFR using the rapid technique. First, for each participant and spectral magnitude, the point along the curve was found that corresponded to the d' score for the standard FFR. In other words, the stimulus duration for the rapid FFR was found for which the spectral components were above the noise floor to the same extent as for the standard FFR. Second, the number of F0 cycles was determined that corresponded to 63% of the asymptotic amplitude of the fitted curve (i.e. τ).

Increasing the number of F0 cycles further would only lead to a small improvement in the discriminability of the spectral components of the response.

6.4 Results

6.4.1 Fitted curves

Figure 6.1 shows 95% confidence intervals of d' scores for the rapid FFR as a function of stimulus duration. Confidence intervals are plotted for the spectral magnitude at F0, and the second and third harmonics (H2 and H3) for the added-polarity FFR, and the fourth and fifth harmonic (H4 and H5) for the subtracted-polarity FFR.

The figure illustrates that for the spectral magnitudes at F0 and H2, and perhaps to some extent H3, the discriminability of the signal above the noise floor increases exponentially with increasing stimulus duration, as indicated by increasing d' scores. On the other hand, the d' scores for the spectral magnitudes at H4 and H5, derived from the subtracted-polarity FFRs, do not appear to increase as the stimulus duration was increased. The figure furthermore indicates that the discriminability of spectral components above the noise floor decreases with increasing frequency.

Exponential curves were fit to the rapid FFR d' scores as a function of stimulus duration for each individual and spectral magnitude separately. The curves were a good fit for F0 and H2, and to some extent H3, as indicated by relatively low residual standard errors (see table 6.1). d' scores increased exponentially and typically, although not always, reached plateau. The exponential function could not always be fit to the data, however, especially for the subtracted FFR measures. The curve could only be fit to the spectral magnitudes at H4 and H5 for four and three participants respectively. In total, curves could not successfully be fit to the data for 30 out of 80 measures.

Descriptive statistics for the residual standard error (i.e. the goodness of fit of the fitted curve) are shown in table 6.1. Since the curve could only be fit for a small number of participants for the subtracted-polarity FFR, descriptive statistics are only provided for the added-polarity FFR measures.

6.4.2 Estimate of required stimulus duration

Comparison to d' standard FFR: The point along the curve that corresponded to the d' score for the standard FFR could only be determined for 11 out of 50 fitted curves (F0 = 3, H2 = 6, H4 = 2). The mean estimate of the required stimulus duration for these ten measures was 7119 F0 cycles. The standard deviation, however, was large (sd = 13017 cycles). Estimates ranged from -14903 to 37087 F0 cycles. The negative

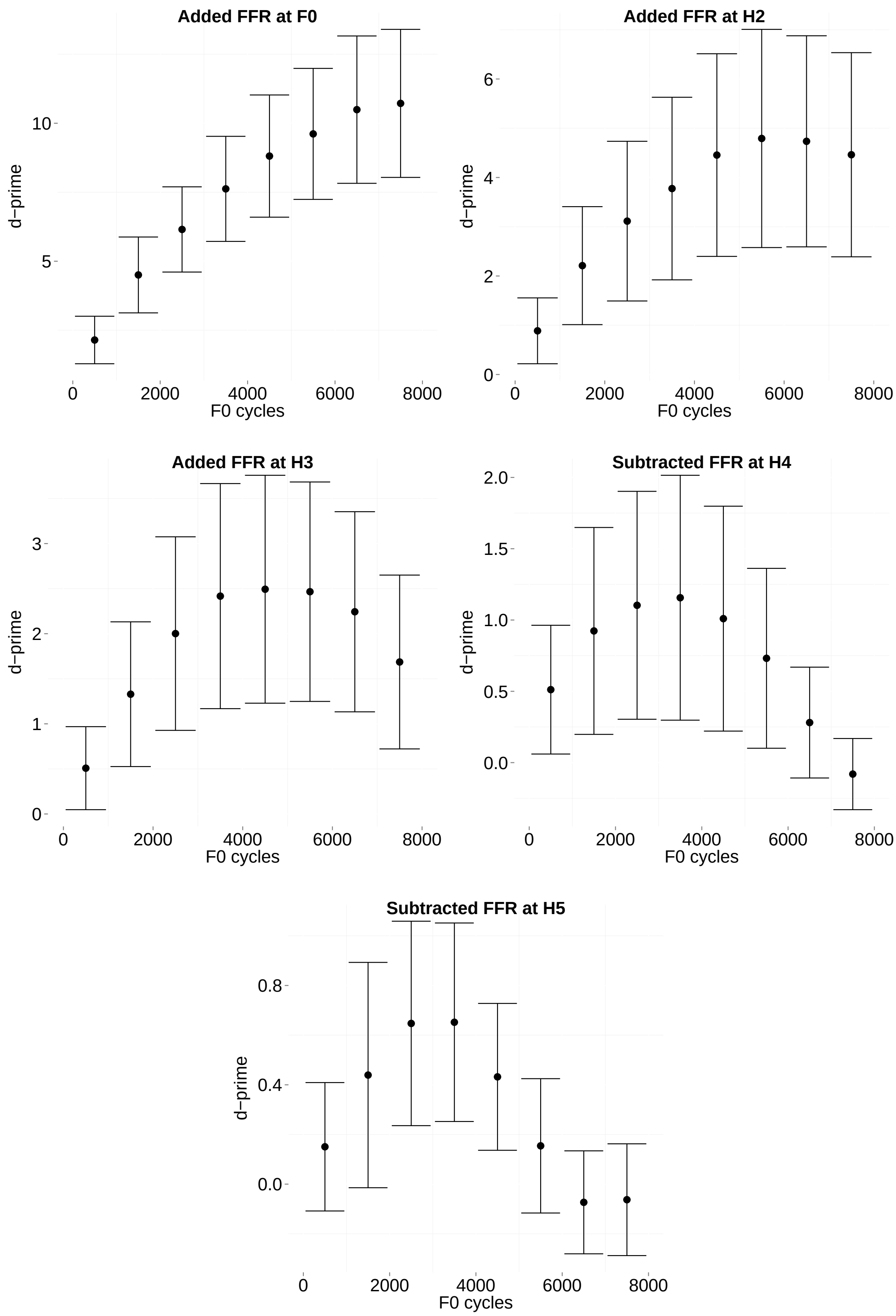


Figure 6.1: 95% confidence intervals of the d' scores as a function of stimulus duration (number of F0 cycles) for the rapid FFR at F0, and the second and third harmonics for the added-polarity response and at the fourth and fifth harmonic for the subtracted-polarity response. Note the different scales along the y-axes.

	Descriptive statistics: Goodness of fit	
Dependent variable	mean (sd)	n
Spectral magnitude at F0	0.39 (0.18)	16
Spectral magnitude at H2	0.36 (0.16)	16
Spectral magnitude at H3	0.35 (0.18)	11

Table 6.1: Descriptive statistics for the residual standard error, a measure of the goodness of fit of the fitted exponential curve. n indicates the number of participants for whom a curve was fit.

	Required F0 cycles: 63% asymptotic amplitude	
Dependent variable	mean (sd)	n
Spectral magnitude at F0	4420 (1922)	16
Spectral magnitude at H2	4470 (3805)	16
Spectral magnitude at H3	2202 (1014)	11

Table 6.2: Descriptive statistics for the estimated required number of F0 cycles of the stimulus to collect a robust rapid FFR. The estimates were based on the 63% asymptotic amplitude of the fitted curves. n indicates the number of participants for whom a curve was fit.

estimates were derived from noisy data, as indicated by low d' scores for the measures in question.

63% asymptotic amplitude of fitted curve: Perhaps a better estimate of the required stimulus duration can be provided based on the 63% asymptotic amplitude of the fitted curve. Descriptive statistics for the estimated required number of F0 cycles is given in table 6.2. 95% confidence intervals for the estimates are plotted in figure 6.2. Note that data estimates are only provided for the added-polarity FFR measures since the exponential curve was a bad fit for most subtracted-FFR measures.

The results suggest that a reasonable stimulus duration to collect the rapid FFR is 4500 F0 cycles. This duration corresponds to the average number of F0 cycles that corresponds to the 63% asymptotic amplitude of the fitted curve (see figures 6.1 and 6.2). It should be noted, however, that this estimate is subject to a high degree of inter-individual variability. While for some participants a shorter stimulus may be

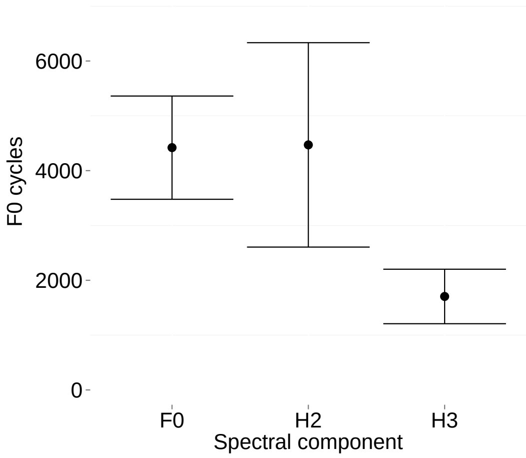


Figure 6.2: 95% confidence intervals for the estimated required stimulus duration based on the 63% asymptotic amplitude of the fitted curves for F0, and the second and third harmonics.

sufficient, for other participants a longer stimulus may be required to collect a robust rapid FFR.

6.4.3 Components above the noise floor

An important question is whether the spectral components of the rapid FFR, collected in response to a stimulus with 4500 F0 cycles, are above the noise floor. To answer this question, a single estimate was computed for each participant (i.e. not using a permutation technique) for each of the spectral magnitudes at F0 and its subsequent harmonics for the 4500-cycle FFR. These values were subsequently compared to the distributions of spectral magnitudes for the noise floor estimates in order to calculate a percentile. Spectral magnitudes above the 95^{th} percentile were considered to be significantly above the noise floor (i.e. α <0.05).

The results, summarised in table 6.3, show that only the spectral magnitudes at F0 and H2 were typically above the noise floor. Spectral magnitudes at H3, and particularly H4 and H5, however, were not significantly above the noise floor for most participants (see Fig. 6.3).

6.5 FFR analyses

The central question of this study was whether the rapid FFR is equally sensitive to inter-individual differences as the standard FFR. Since the analyses reported above suggest that a stimulus duration of 4500 F0 cycles is sufficient to collect rapid FFRs with spectral components above the noise floor, at least for the first two harmonics,

	Descriptive statistics: Percentiles	
Dependent variable	mean (sd)	n
Spectral magnitude at F0	99.9 (0.14)	15
Spectral magnitude at H2	94.4 (10.5)	11
Spectral magnitude at H3	83.5 (19.7)	7
Spectral magnitude at H4	32.0 (25.7)	0
Spectral magnitude at H5	37.8 (18.1)	0

Table 6.3: Descriptive statistics for the percentiles, indicating whether the response to a stimulus with 4500 F0 cycles was above the noise floor. n refers to the number of participants (out of 16) for whom the spectral magnitude of the response was above the 95^{th} percentile of the distribution of spectral magnitudes of the noise floor.

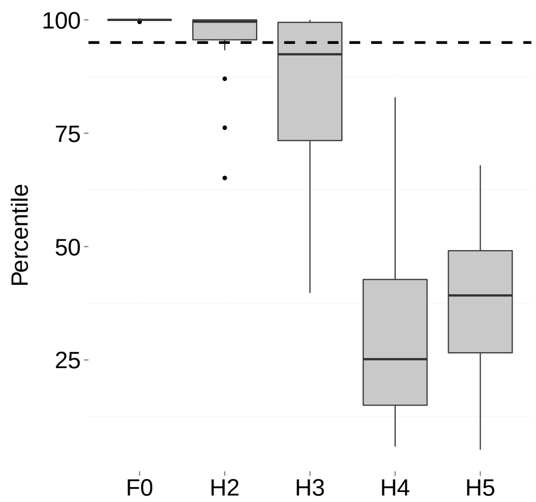


Figure 6.3: Boxplots of the percentiles of the 4500-cycle FFR for F0 and its subsequent harmonics. The dashed line indicates the 95^{th} percentile.

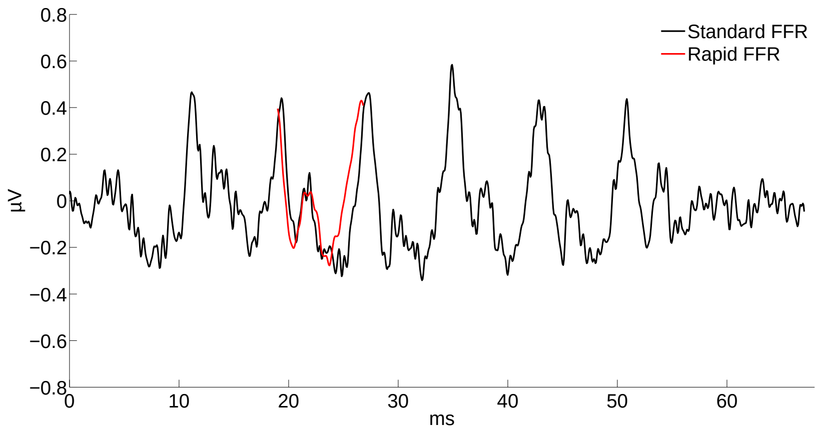


Figure 6.4: A typical rapid FFR (red) is superimposed on a standard FFR (black) from the same participant. Note that the rapid FFR is shifted in time so it overlaps with a cycle of the standard FFR.

rapid FFRs were averaged across 4500 F0 cycles per polarity.

All analyses were performed on the single cycle of the rapid FFR and on 5 cycles of the standard FFR (i.e. not taking the two tapered F0 cycles into account). Averages reflecting both added and subtracted polarities were analysed, unless stated otherwise. The precise timing of the standard FFR was identified by means of a stimulus-to-response correlation on the average (added polarities only).

Response-to-response correlations were computed across the two runs that were collected for each recording technique (added polarities only). A Fisher transformation was used to convert the correlation coefficients (r values) to z scores for statistical analyses. The following analyses were performed on responses averaged across the two runs. Spectral magnitudes were calculated at F0 and its subsequent two harmonics for the added responses and at the fourth and fifth harmonic for the subtracted responses. The rms amplitude of the FFRs (added polarities only) was computed as a measure of the overall magnitude of the response.

6.6 Results

A typical rapid FFR, superimposed on a standard FFR from the same participant, is shown in figure 6.4. This figure suggests a high degree of similarity between the rapid and standard FFRs. Boxplots illustrating the spectral composition of the standard and rapid FFRs are shown in figure 6.5.

The extent to which the response properties of the rapid and standard FFRs are the same may not necessarily be important or informative. A more important question is whether the rapid FFR is as sensitive to inter-individual differences in subcortical

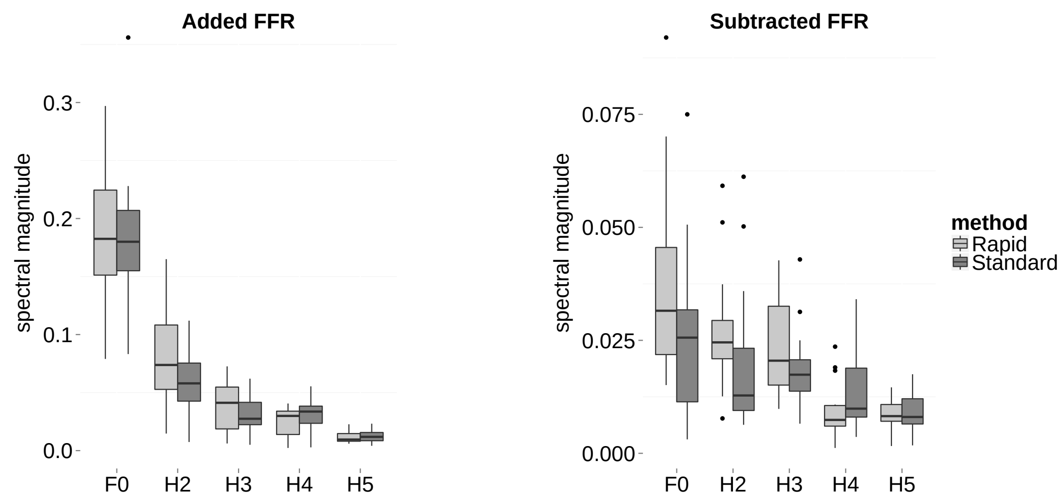


Figure 6.5: Boxplots of the spectral magnitudes at F0 and its subsequent harmonics for the added and subtracted-polarity FFRs, with measurements from the rapid FFR in light grey and the standard FFR in dark grey. Note the different scales along the y-axis for the added and subtracted polarity FFTs.

auditory processing as the standard FFR.

To answer this question, Pearson's correlations between the rapid and standard FFRs were performed on commonly used outcome measures: response-to-response correlation, rms amplitude, and spectral magnitudes at F0, H2, and H3 for the added-polarity FFR and spectral magnitudes at H4 and H5 for the subtracted-polarity FFR. The results were not corrected for multiple comparisons because the correlations were planned.

The analyses showed that, when polarities were added, there was a significant correlation between the rapid and standard FFR in terms of spectral magnitude at F0, H2, H3, and rms amplitude (but not response-to-response correlation). When polarities were subtracted, however, neither of the outcome measures (H4 and H5) was significantly correlated across methods. The results are summarised in table 6.4 and scatter plots for the measures are shown in figures 6.6 and 6.7.

It should be noted that, while these results appear to suggest that the measurement techniques largely correlate in terms of the added-polarity but not the subtracted-polarity FFR, this difference cannot directly be assessed through significance levels (Gelman and Stern, 2006). To assess whether the correlation coefficients for the added and subtracted FFRs were indeed significantly different, a linear regression was performed on the added and subtracted FFRs separately. Furthermore, a linear

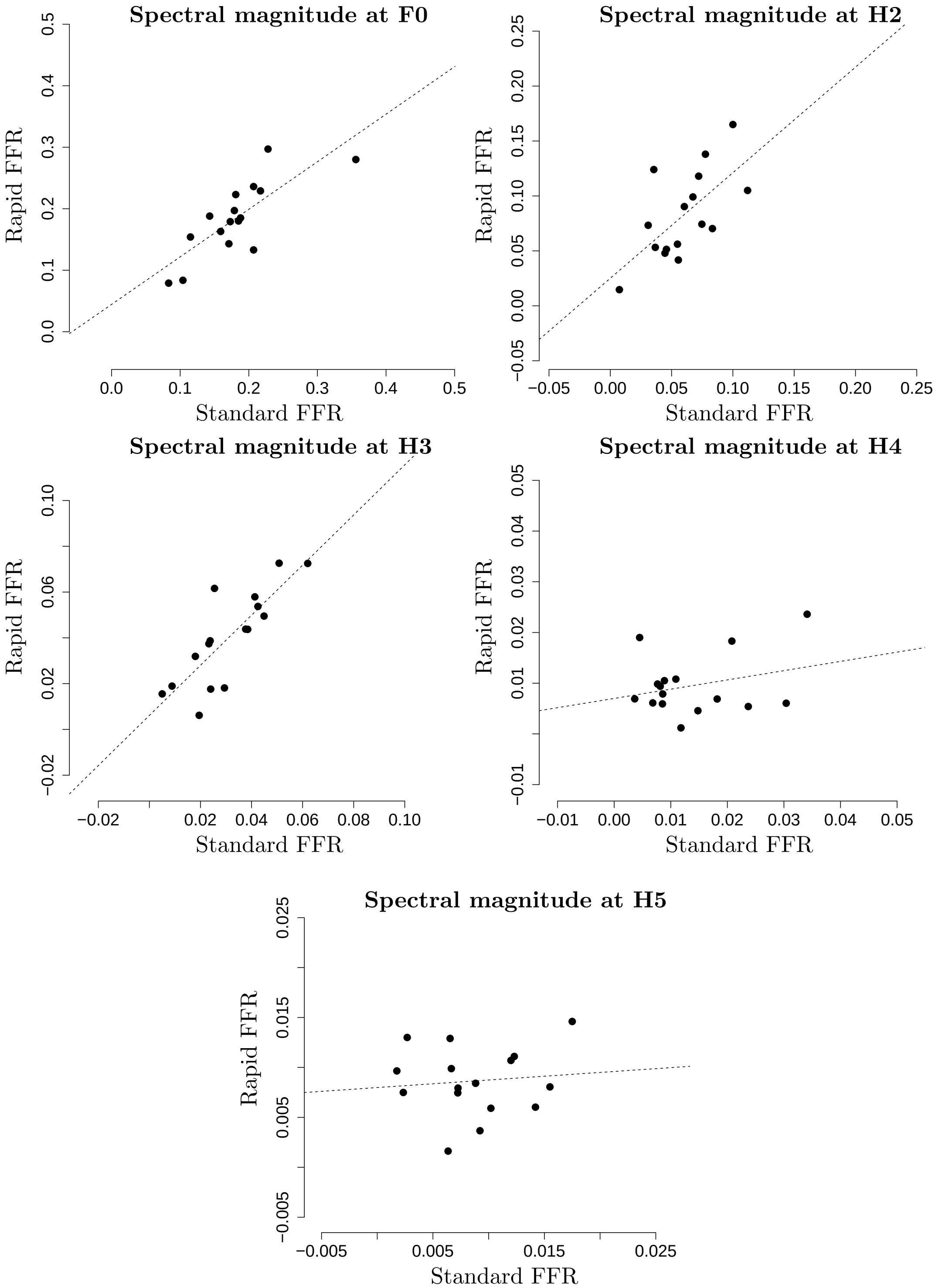


Figure 6.6: Scatter plots illustrating the relationship between response measures from the standard and rapid FFR. The relationship is illustrated for the spectral magnitude at F0, and the second and third harmonics for the added-polarity FFRs, and the fourth and fifth harmonics for the subtracted-polarity FFRs.

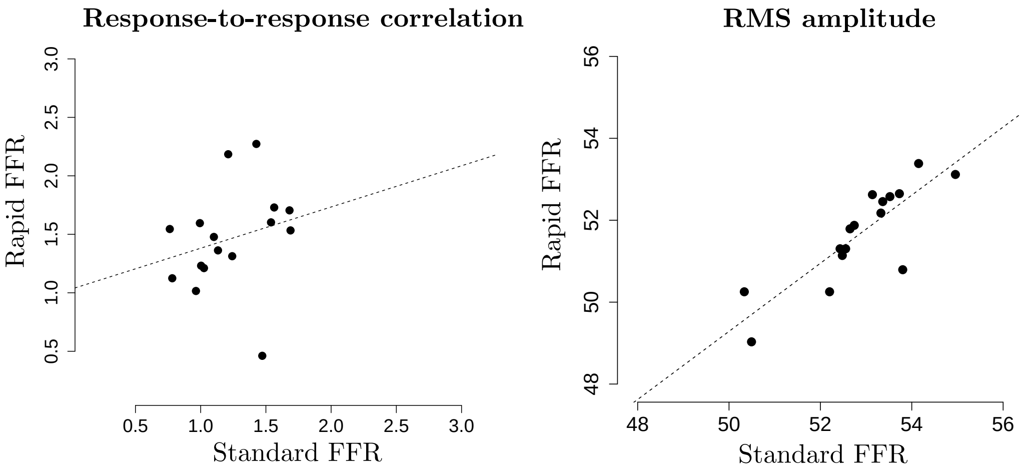


Figure 6.7: Scatter plots illustrating the relationship between response measures from the standard and rapid FFR. The relationship is illustrated for the response-to-response correlation and rms amplitude for the added-polarity FFRs.

Correlations	
	Added polarity
Spectral magnitude at F0	r = 0.79, **p = 0.001*****
Spectral magnitude at H2	r = 0.65, **p = 0.007****
Spectral magnitude at H3	r = 0.81, **p < 0.001*****
RMS amplitude	r = 0.84, **p < 0.001*****
Response-to-response correlation	r = 0.25, p = 0.4
	Subtracted polarity
Spectral magnitude at H4	r = 0.28, p = 0.3
Spectral magnitude at H5	r = 0.10, p = 0.7

Table 6.4: Correlations between response measures of the standard and rapid FFRs. Correlations are provided for the spectral magnitude at F0, and the second and third harmonics, the rms amplitude, and the response-to-response correlation for the added-polarity FFRs, and spectral magnitudes at the fourth and fifth harmonics for the subtracted-polarity FFRs.

	Linear regression models				
Model	b	SE	p	R^2	CI
Added-polarity FFR	1.1	0.05	$<\mathbf{0.001}^{***}$	0.86	[1.0, 1.2]
Subtracted-polarity FFR	0.16	0.11	0.2	0.07	[-0.07, 0.4]
Interaction with polarity	1.1	0.04	$<\mathbf{0.001}^{***}$	0.87	[1.05, 1.2]

Table 6.5: Results of linear regression analyses assessing the relationship between the standard and rapid FFRs for the added and subtracted-polarity measures, as well as the interaction with polarity (i.e. added or subtracted). Significant results are highlighted in bold font. (* significant at $\alpha = 0.05$, ** significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$). 95% confidence intervals (CI) are also provided for the regression coefficients.

regression was conducted on both the added and subtracted FFR measures with a dummy-coded predictor indicating whether polarities were added or subtracted.

Table 6.5 shows that the interaction between the dummy-coded variable and the standard FFR outcome measures significantly predicted the rapid FFR outcome measures. This supports the idea that the two measurement techniques largely correlate in terms of the added-polarity but not the subtracted-polarity FFR.

6.7 Discussion

This study showed that a robust added-polarity FFR can be collected more rapidly by presenting the stimulus continuously (i.e. without interstimulus interval) and averaging across a single cycle of the response. Recording times of the FFR were reduced from approximately 2.5 minutes for the standard technique to only 35 seconds using the rapid technique.

The results indicated that the extent to which the spectral components of the added-polarity FFR were above the noise floor increased exponentially with increasing stimulus duration. This is in line with findings by Jeng et al. (2011), who showed a similar pattern of change for response measures of the standard FFR. Exponential functions could not always be fit to the data, however, especially not for the subtracted-polarity FFR measures. It is unclear why the discriminability of the spectral magnitudes at the fourth and fifth harmonics, derived from the subtracted-polarity FFR, did not increase with increasing stimulus duration. One possible explanation may be that this is due to a change in group delay over time, particularly affecting the higher frequency components.

Based on the fitted curves it was estimated that 4500 F0 cycles (per polarity) were required to collect a rapid FFR with spectral components above the noise floor. This is three times more than the 1500 sweeps that are typically sufficient for the standard FFR. One possible explanation for this finding is that the prolonged stimulus presentation, characteristic of the rapid FFR technique, reduces neural firing phase-locked to the stimulus over time. Future research should address the potential effects of neural adaptation on the rapid FFR.

It should be noted that the estimate of the required stimulus duration was subject to a high degree of inter-individual variability, with standard devations of 2000 - 4000 F0 cycles. The estimate was based on the 63% asymptotic amplitude of the fitted curve with the idea that further increasing the stimulus duration would only lead to a small improvements in the discriminability of the spectral components of the FFR. Future research should derive a better estimate of the required stimulus duration to collect a robust response by testing a larger sample.

The results furthermore showed that the response measures of the added-polarity FFR, but not subtracted-polarity FFR, correlated significantly across the two acquisition methods. This suggests that the rapid technique provides a reliable measure in terms of the added-polarity FFR, but cannot currently reliably be used to assess higher frequency spectral components thought to reflect temporal fine structure. Future research is required to further develop the rapid FFR techinque to allow for the reliable acquisition of subtracted-polarity FFR.

The fact that the subtracted-FFR measures did not correlate significantly across measures can probably be attributed to the fact that the fourth and fifth harmonics were not significantly above the noise floor. It is unlikely, however, that increasing the stimulus duration would have led to a more reliable measure of the fourth and fifth harmonics.

The results of this study could have major implications for future research as the rapid FFR can reduce recording times significantly. This means that it becomes easier to test a larger number of conditions within one session and to test participants who have difficulty sitting still for a long time. Furthermore, reducing the acquisition time of the FFR may ultimately also have implications for clinical audiology settings. Reducing recording times for the FFR may make it more likely for the measure to be included in audiological test batteries.

Chapter 7

General discussion

7.1 The effects of ageing on speech perception in noise

The primary aim of this thesis was to examine why older adults, even those with normal or near-normal hearing, typically experience increased difficulties understanding speech in the presence of background noise. The main focus was on the potential contribution of age-related auditory neuropathy, as indicated by declines in the auditory brainstem response (ABR) and frequency following response (FFR), to these speech in noise problems. Furthermore, the role of potential age-related declines in (spectro)temporal processing and cognition was examined.

7.1.1 Speech perception in noise

Contrary to expectations, the data indicated that normal-hearing older adults do not necessarily experience increased difficulties understanding speech in noise when complex, ecologically valid target stimuli are used. Chapters 2 and 3 showed that older adults with normal hearing, following fairly strict criteria (audiometric thresholds <25 dB HL up to 6 kHz in at least one ear and up to at least 4 kHz in the other ear), did not perform more poorly in steady-state or amplitude-modulated speech-shaped noise compared to younger listeners. The older adults thus also did not show a reduced fluctuating masker benefit (FMB).

These findings appear to contradict previous reports in the literature that showed age-related declines in the FMB (Gifford et al., 2007; Stuart and Phillips, 1996; Takahashi and Bacon, 1992; Dubno et al., 2002, 2003; Peters et al., 1998; Grose et al., 2009). Although, as noted in chapter 2, the idea that older adults benefit less from fluctuations in the masker is perhaps somewhat controversial since age-related declines in the FMB reported in the literature may in part have been the result of group differences that were also apparent in steady-state noise (c.f. Bernstein and Grant, 2009; Stuart and Phillips, 1996; Dubno et al., 2002, 2003).

One possible explanation for the absence of an age-related FMB decline may be the fairly strict criteria for normal hearing used in the study reported in chapters 2 and 3. Normal hearing is often defined as thresholds within 20 or 25 dB HL up to 4 kHz (e.g. Takahashi and Bacon, 1992; Peters et al., 1998; Dubno et al., 2002, 2003; Grose et al., 2009 but see Stuart and Phillips, 1996; Gifford et al., 2007). Speech in noise performance for older adults with hearing loss above 4 kHz was examined in the study reported in chapter 4. While the FMB was not assessed in this study, the results did show that the older adults performed more poorly in the presence of steady-state speech-shaped noise compared to young listeners, although it should be noted that the group difference was small (1.7 dB). An independent t-test confirmed that the SRTs for the older adults with hearing loss above 4 kHz were indeed higher (i.e. poorer) than for the older adults who participated in the study reported in chapters 2 and 3 (t_{35} = 3.2, $p = 0.003$, mean difference 1.2 dB, Cohen's d = 0.7). This finding suggests that even elevated thresholds above 4 kHz, which are often still considered 'normal hearing', could contribute to increased speech in noise difficulties. However, it remains unclear whether this difference in SRTs can indeed be attributed to differences in audiometric thresholds, or whether it is attributable to other differences between the groups (e.g. age, temporal processing, cognitive function). Nonetheless, this suggests that future studies investigating the effects of ageing, independent of hearing loss, on the perception of speech in noise should use strict criteria of normal hearing.

The data furthermore showed that normal-hearing older adults perform more poorly only in the presence of competing speech. While no group differences were found in steady-state or amplitude-modulated noise, SRTs in two-talker babble were higher (i.e. poorer) for the older adults by 1.4 dB. This is in line with previous research that suggests that competing speech is particularly detrimental for older adults (Tun and Wingfield, 1999; Helfer and Freyman, 2008; Rajan and Cainer, 2008). It remains unclear, however, why two-talker babble is more detrimental than noise maskers. Future research is required to determine whether these increased difficulties are due to increased susceptibility to informational masking (c.f. Freyman et al., 2004), reduced ability to make use of periodicity cues to successfully segregate the target and masker (Vongpaisal and Pichora-Fuller, 2007; Rossi-Katz and Arehart, 2009), or even reduced ability to listen in the dips of the speech masker (Dubno et al., 2002; Gifford et al., 2007).

7.1.2 Subcortical auditory processing

Of particular interest in this thesis was the contribution of age-related declines in subcortical auditory processing to the perception of speech in noise. In line with previous research, the data showed declines in the ABR and FFR for the older adults (Konrad-Martin et al., 2012; Vander Werff and Burns, 2011; Anderson et al., 2012; Clinard and Tremblay, 2013). Chapter 4 showed that older adults with normal-hearing up to 4 kHz had reduced ABR wave amplitudes, thought to reflect a loss of functioning nerve fibres from the auditory nerve up to the inferior colliculus at the brainstem. Similarly, chapter 3 showed that FFRs in response to the vowel /ɑ/ in quiet, and steady-state and amplitude-modulated speech-shaped noise were less robust in the older compared to the younger adults. While a simulation study by Lopez-Poveda and Barrios (2013) suggested that auditory neuropathy would likely affect subcortical speech coding more in noise than in quiet, the FFR was not more affected by noise in the older listeners. Similarly, the older adults did not show a reduced 'neural masking release', as reflected in the FFR in the presence of amplitude-modulated noise. This is probably because age-related declines in Env processing only become apparent at higher modulation rates (above about 200 Hz, e.g. Purcell et al., 2004; Grose et al., 2009).

It remains difficult, however, to tease apart the effects of ageing and high-frequency hearing loss on these declines in subcortical neural coding. This is especially the case for the declines in the ABR found for the older adults with hearing loss above 4 kHz. As mentioned in chapter 4, nerve fibres along nearly the whole length of the basilar membrane are involved in the generation of the click ABR (Don and Eggermont, 1978). It is thus likely that high-frequency hearing loss in part contributed to the reduced ABR wave amplitudes. Regression analyses could not reliably determine whether the declines in the ABR were the result of ageing or differences in audiometric thresholds up to 4 kHz or above between the young and older adults, not least because these measures are so highly inter-correlated.

Similarly, high-frequency hearing loss could also contribute to decreased robustness of the FFR for older adults. It has been suggested that the FFR primarily stems from neurons at more basal sites along the cochlea (Janssen et al., 1991; Dau, 2003). In other words, the FFR likely results from synchronised neural activity from mid- and high-frequency units and not necessarily from units tuned to the frequency of the stimulating waveform. It is unlikely, however, that hearing losses above 6 kHz played a major role in the decreased robustness of the FFR observed for the older adults that participated in the study described in chapter 3. In particular, if the elevated audiometric thresholds in the high frequency range were mainly due to outer hair cell dysfunction, the FFR may not necessarily have been affected. The basilar membrane response may not necessarily be significantly affected for stimuli in the tail of the tuning curve. Indeed, regression analyses indicated that the FFR in quiet and in noise was best predicted by age group alone, suggesting that the degradation of the FFR was not due to elevated audiometric thresholds, either across 0.5-4 or 6-8 kHz.

Future studies investigating the effects of ageing, independent of hearing loss, on subcortical auditory processing should rule out the contribution of high-frequency hearing loss to the ABR and FFR. High-pass noise masking is thought to eliminate the contribution of basal areas to both responses (Don and Eggermont, 1978; Janssen et al., 1991; Dau, 2003). This method could thus ensure that the response was generated only by neurons along the basilar membrane within the region of normal hearing.

7.1.3 Temporal processing

This thesis furthermore explored the role of age-related declines in auditory temporal and spectro-temporal processing on the perception of speech in noise. Contrary to expectations, the normal-hearing older adults that participated in the study reported in chapter 2, did not show any age-related declines in Env or TFS processing. Age-related declines in temporal processing are well documented for normal-hearing older adults (e.g. Pichora-Fuller and Souza, 2003; Pichora-Fuller et al., 2007; Frisina and Frisina, 1997; CHABA, 1988; Gordon-Salant, 2005). However, the criteria for normal hearing are often less stringent than those used in chapter 2, with thresholds within 20 or 25 dB HL up to only 4 kHz often qualifying as normal hearing (He et al., 2007, 2008; Grose et al., 2009; Grose and Mamo, 2010; Füllgrabe, 2013). In fact, it could be argued that the older adults who participated in this study were exceptional, if only in the sense that they had good hearing. This could perhaps explain the absence of any age-related temporal processing declines in this study.

The older adults with normal hearing up to only 4 kHz, who participated in the study described in chapter 4, revealed a decline in spectro-temporal processing. STM detection thresholds were significantly higher for the older adults. However, whether this decline can be attributed to a decline in TFS processing remains unclear. Bernstein et al. (2013) found that STM detection thresholds were correlated with a measure of TFS processing, namely FM detection for a 500-Hz carrier modulated at 2 Hz. A limitation in this study is that no measure of TFS processing was obtained.

7.1.4 Auditory neuropathy and temporal processing

One topic that deserves further attention is the relationship between subcortical auditory processing, as measured by the ABR and FFR, and behavioural measures of auditory temporal processing, and TFS processing in particular. As discussed in chapters 3 and 4, auditory nerve fibres encode sound with great temporal precision, at least at low frequencies, by phase-locking to the auditory stimulus. Due to the stochastic nature of the firing patterns of auditory nerve fibres, temporal coding improves as multiple fibres fire in response to the stimulating waveform. Conversely, a reduction in the number of functioning nerve fibres can thus be expected to lead to reduced temporal precision of neural coding. The data presented in this thesis, however, do not reveal a relationship between behavioural measures of TFS processing and subcortical auditory processing.

No support was found for the idea that poor STM sensitivity could in part be attributed to auditory neuropathy, as measured by the click ABR. While the older adults who participated in the study described in chapter 4 showed declines in both STM sensitivity and ABR wave amplitudes, a best subsets regression indicated that the click ABR could not predict STM thresholds over and above individual differences in audiometric thresholds. As already mentioned, however, the STM declines may not necessarily be the result of impaired TFS processing. The lack of association between the two measures may furthermore be explained by the fact that the ABR reflects neural firing along almost the entire basilar membrane, whereas STM sensitivity was measured only up to 2 kHz.

Furthermore, there is no support for an association between the FFR and FM detection. On the one hand, normal-hearing older adults showed declines in subcortical auditory processing, as indicated by less robust FFRs. On the other hand, however, these same older adults did not have increased FM detection thresholds for a 500-Hz carrier modulated at 2 Hz, a measure thought to reflect phase-locking (Moore and Sek, 1995, 1996). Indeed, the principal component measures of the FFR in quiet and noise did not correlate significantly with FM detection thresholds (FFR quiet: $r = 0.057$, $p = 0.7$; FFR noise: $r = -0.008$, $p = 0.9$). One possible explanation for the lack of association between these measures is that the older adults were able to compensate for the declines in subcortical auditory processing. Performance on behavioural measures of auditory temporal processing are likely affected by non-auditory factors such as motivation, attention, and working memory (c.f. He et al., 1999; Harris et al., 2010). An alternative explanation is that the FFR measures may reflect subcortical processing of envelope rather than TFS information, since the responses to the positive and negative-polarity stimuli were added and not subtracted (Aiken and Picton, 2008; Skoe and Kraus, 2010). The added-polarity FFR would thus not necessarily correlate with a behavioural measure of TFS processing.

To get a better understanding of the relationship between the FFR and behavioural measures of TFS processing, it is imperative to improve our understanding of what the FFR actually reflects. Interdisciplinary research, combining human FFR measures with modelling work, should examine to what extent the FFR components directly reflect envelope and TFS information in the stimulus or result from non-linear distortions along the auditory pathway.

7.1.5 Cognitive processing

Since the perception of speech in noise is a complex process, involving both bottom-up and top-down processing, the effects of cognitive declines on speech in noise understanding were also examined. In line with previous research, age-related declines in working memory and processing speed were found (e.g. Craik and Jennings, 1992; Salthouse, 1996). By contrast, the older adults did not perform more poorly on tasks of attentional switching and linguistic closure compared to the younger participants.

The literature is inconclusive as to whether linguistic closure, or the ability to read masked text, declines with age (e.g. Besser et al., 2012; Kramer et al., 2009). As suggested in chapter 2, older adults may not experience increased difficulties reconstructing partially masked text because this skill is representative of crystallised intelligence, which does not decline with age (Horn and Cattell, 1967).

The absence of an age-related decline in attentional switching is perhaps more surprising given reports in the literature of declines in inhibitory control, a capacity thought to prevent task-irrelevant information from taking up resources that could otherwise be used to process task-relevant information (Hasher and Zacks, 1988; Hasher et al., 1999). One possible explanation for this finding is that the test of attention used in this thesis, the Visual Elevator task (Robertson et al., 1996), is not the best measure of attentional switching. In future, a better validated task of attention should be sought.

7.1.6 Predicting speech in noise performance

An important finding in this thesis is that normal-hearing older adults experience relatively few difficulties in the perception of speech in noise, despite declines in subcortical auditory processing, working memory, and processing speed. This suggests that auditory neuropathy and cognitive declines associated with ageing do not necessarily lead to increased speech in noise problems. Moreover, when the older adults did experience increased speech in noise difficulties, this could not be explained by auditory neuropathy, cognitive declines, or reduced STM sensitivity. Instead, audiometric thresholds were the best predictor of speech in noise performance.

Contrary to expectations, auditory neuropathy, as measured by the ABR and FFR, does not necessarily lead to increased perceptual difficulties of speech in noise. It is perhaps not so surprising that a reduction in the amplitude of the click ABR does not lead to speech in noise problems since the ABR is a crude measure that does not reflect neural coding of more complex sounds such as speech. The FFR is thought to be a better measure of subcortical auditory processing important for speech perception. Several studies have indicated a relationship between the FFR and a listener's ability to understand speech in noise, with more robust FFRs corresponding to better speech in noise performance (Hornickel et al., 2009; Song et al., 2011; Anderson et al., 2011). One likely explanation for the lack of association between the FFR and speech in noise performance in this thesis is that the disruption of subcortical speech coding was not severe enough to cause increased speech in noise difficulties. Understanding speech in noise is a complex process and older adults may be able to compensate for declines in subcortical auditory processing.

Age-related cognitive declines were similarly expected to lead to increased speech in noise difficulties for normal-hearing older adults. Cognitive declines have repeatedly been shown to be the most important factor explaining aided speech understanding in noise for hearing impaired older adults, after accounting for differences in audiometric thresholds (see reviews by Humes and Dubno, 2010; Akeroyd, 2008; Houtgast and Festen, 2008). However, the data showed that cognitive declines do not necessarily lead to speech in noise problems for normal-hearing older adults. This is perhaps best explained by the fact that the speech signal is less distorted in the absence of hearing loss and temporal processing deficits, which may put a lower demand on top-down cognitive processing. An alternative explanation is that the observed cognitive declines were not severe enough to affect the ability to understand speech in noise.

Furthermore, the fact that STM sensitivity does not predict speech in noise performance beyond the audiogram is not in line with previous studies that have found that STM sensitivity accounts for 40 - 80 % of the variance in SRTs (Bernstein et al., 2013; Mehraei et al., 2013). It should be noted, however, that these findings were for a group of hearing impaired older adults, who likely perform more poorly on a speech in noise task. The group difference in SRTs in the study reported in chapter 4 was only 1.7 dB. Nonetheless, this finding highlights the need for further research into the applicability of an STM-based model of speech perception in noise.

It could be argued that a possible limitation of the ageing studies presented in this thesis is that the sample sizes were relatively small (n = 18 - 19). However, the sample sizes were limited by resources since that the majority of people over the age of 60 have some degree of hearing loss, especially at the high frequencies (Mościcki et al., 1985; Cruickshanks et al., 1998). In fact, the sample sizes in this thesis are larger than many similar studies (e.g., n = 10 in Dubno et al., 2002; Gifford et al., 2007; Grose et al., 2009). Moreover, the studies were able to detect group differences in SRTs of 1.4 - 1.7 dB which could be argued to be as small as is meaningful.

It could similarly be argued that the normal-hearing older adults who took part in the study described in chapters 2 and 3 were exceptional, if only in the sense that they had very good hearing for their age. Their cognitive function and FFRs may similarly have been exceptional compared to the more typical older population. The age-related declines in cognition and subcortical speech coding may not have been severe enough to affect their speech in noise performance. This argument would not hold true, however, for the participants who took part in the study described in chapter 4 as they had more typical hearing thresholds, at least with reference to the literature (e.g. Takahashi and Bacon, 1992; He et al., 2007; Grose et al., 2009).

7.2 The importance of context for dip listening

This thesis furthermore explored the effects of masker modulation type and sentence complexity on the FMB. In addition, the role of linguistic closure on speech perception in steady-state and amplitude-modulated noise was assessed.

This study was motivated by the finding that even young listeners only obtained a small FMB of on average 2.7 dB when IEEE sentences (Rothauser et al., 1969) were used as targets and the masker was sinusoidally amplitude-modulated (see chapters 2 and 3). This raised the question whether simpler materials, such as the BKB sentences (Bench et al., 1979), and square-wave amplitude modulation would have elicited a larger FMB.

The data indeed showed a larger FMB for the BKB than the IEEE sentences.

Since these sentence materials are simpler and contain more contextual information, this finding suggests that the availability of contextual cues facilitates a larger FMB. The data furthermore showed a larger FMB for the square-wave compared to the sinusoidally amplitude-modulated masker.

It is unlikely, however, that the failure to replicate the finding that ageing is associated with a decline in the FMB (Gifford et al., 2007; Stuart and Phillips, 1996; Takahashi and Bacon, 1992; Dubno et al., 2002, 2003; Peters et al., 1998; Grose et al., 2009) can in part be attributed to the choice of stimulus parameters, even though the benefit for the young listeners was so small to begin with. Age-related declines in the FMB have indeed been reported using simpler sentence materials (Peters et al., 1998; Gifford et al., 2007 but see Grose et al., 2009) and square-wave amplitude-modulated maskers (Dubno et al., 2002, 2003; Gifford et al., 2007). However, the ability to use contextual information in fact improves with age (Pichora-Fuller et al., 1995; Dubno et al., 2000; Sheldon et al., 2008; Speranza et al., 2000). It is thus unlikely that an age-related decline in the FMB would have become apparent using simpler target sentences. Similarly, it is unlikely that square-wave modulation of the masker would lead to an age-related FMB decline when in fact a larger proportion of the target is unaffected by square-wave than sinusoidally amplitude-modulated maskers.

The data furthermore showed that linguistic closure, as measured by the TRT, correlated significantly with speech perception in amplitude-modulated but not steady-state noise. However, better linguistic closure was associated with poorer SRTs. In other words, the correlation was going in the opposite direction than expected (e.g. George et al., 2007; Besser et al., 2012). One possible explanation for this finding, as suggested in chapter 5, is that the two TRT runs collected for each participant were too variable, leading to a poor estimate of linguistic closure ability. It remains an open question whether linguistic closure is more important for the perception of speech in steady-state or fluctuating maskers.

7.3 Rapid collection of the FFR

Last but not least, a new technique was proposed that allows for a more rapid collection of the FFR by presenting the stimulus continuously (i.e. without interstimulus interval) and averaging across a single cycle of the response. Chapter 6 showed a reduction in recording times from approximately 2.5 minutes for the standard technique to only 35 seconds using the rapid technique.

The study suggested that the rapid technique can be used to collect the added-polarity, but not the subtracted-polarity FFR. Response measures largely correlated across the standard and rapid techniques for the added but not the subtracted FFRs. Future research is therefore required to further develop the rapid FFR techinque and extend it to subtracted-polarity FFRs.

It was estimated that 4500 F0 cycles (per polarity) were required to collect a robust FFR using the rapid technique. This estimate was subject to a high degree of inter-individual variability, however, with standard devations of 2000 - 4000 F0 cycles. A limitation of this study is that a relatively small sample size (n = 16) was used. A larger sample would likely have led to a more accurate estimate of the required stimulus duration.

Nonetheless, the data appeared to suggest that a larger number of sweeps is required to collect a robust FFR using the rapid compared to the standard technique. One possible explanation for this finding is that neural adaptation may affect the development of the FFR with increasing stimulus duration. Future research is required to examine the potential contribution of neural adaptation to the rapid FFR.

The results of this study could have major implications for future research. A significant reduction in recording times means that a larger number of conditions can be recorded within the same amount of time. In addition, it will be easier to measure subcortical auditory function for difficult to test populations or in clinical settings.

7.4 Conclusions

In sum, this thesis set out to determine why older adults, even in the absence of hearing impairment, typically experience increased difficulties understanding speech in the presence of background noise. In addition, a study was conducted which assessed the effects of masker modulation type and sentence complexity on the FMB. This study also examined the role of linguistic closure in the perception of speech in steady-state and fluctuating maskers. Furthermore, a new technique was developed that more rapidly collects the FFR. The findings of this thesis can be summarised as follows:

1. Older adults who fit strict criteria of normal hearing do not necessarily have temporal processing deficits or increased difficulties understanding speech in the presence of noise maskers.

2. Auditory neuropathy and age-related cognitive declines do not inevitably lead to increased speech in noise problems.

3. Auditory neuropathy does not necessarily lead to temporal processing deficits.
4. Contextual information is important for dip listening.
5. The FFR can be collected more rapidly by presenting the stimulus continuously, without interstimulus interval, and averaging across a single cycle of the response.

Bibliography

Aiken, S. J. and Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, 245(1-2):35–47.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

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. *International Journal of Audiology*, 47:S53–S71.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (1971). Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects. *Journal of the Acoustical Society of America*, 49(4):1131–1139.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience*, 32(41):14156–14164.

Anderson, S., Parbery-Clark, A., Yi, H.-G., and Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and hearing*, 32(6):750–7.

Anderson, S., White-Schwoch, T., and Parbery-Clark, A. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, 300:18–32.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech, Language, and Hearing Research*, 41:549–563.

Baddeley, A. (1986). *Working memory.* Oxford University Press, Cambridge, England.

Baer, T. and Moore, B. C. J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. *Journal of the Acoustical Society of America*, 94(3):1229–1241.

Barrenäs, M. L. and Wikström, I. (2000). The influence of hearing and age on speech recognition scores in noise in audiological patients and in the general population. *Ear and Hearing*, 21(6):569–577.

Bench, J., Kowal, A., and Bamford, J. (1979). The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology*, 13(3):108–112.

Bernstein, J. G. W. and Brungart, D. S. (2011). Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio. *Journal of the Acoustical Society of America*, 130(1):473–488.

Bernstein, J. G. W. and Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 125(5):3358–3572.

Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., and Leek, M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology*, 24(4):293–306.

Bernstein, J. G. W., Summers, V., Iyer, N., and Brungart, D. S. (2012). Set-size procedures for controlling variations in speech-reception performance with a fluctuating masker. *Journal of the Acoustical Society of America*, 132(4):2676–89.

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

Besser, J., Zekveld, A. A., Kramer, S. E., and 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*, 55:194–209.

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., and Shinn-Cunningham, B. G. (2014). Cochlear neuropathy and the coding of supra-threshold sound. *Frontiers in Systems Neuroscience*, 8(Article 26):1–18.

Boettcher, F. A., Mills, J. H., Swerdloff, J. L., and Holley, B. L. (1996). Auditory evoked potentials in aged gerbils: responses elicited by noises separated by a silent gap. *Hearing Research*, 102(1-2):167–78.

Boettcher, F. A., Poth, E. A., Mills, J. H., and Dubno, J. R. (2001). The amplitude-modulation following response in young and aged human subjects. *Hearing Research*, 153:32–42.

Bopp, K. L. and Verhaeghen, P. (2005). Aging and verbal memory span: a meta-analysis. *The journals of gerontology. Series B, Psychological sciences and social sciences*, 60(5):P223–233.

Bregman, A. S. (1990). *Auditory scene analysis.* MIT Press, Cambridge, MA.

Burkard, R. F. and Sims, D. (2001). The human auditory brainstem response to high click rates: aging effects. *American Journal of Audiology*, 10(2):53–61.

Burkard, R. F. and Sims, D. (2002). A comparison of the effects of broadband masking noise on the auditory brainstem response in young and older adults. *American Journal of Audiology*, 11:13–22.

Burke, D. M. and Shafto, M. A. (2008). Language and Aging. In Craik, F. I. M. and Salthouse, T. A., editors, *The handbook of aging and cognition*, pages 373–443. Psychology Press, New York, NY.

Byrne, D. and Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, 7(4):257–265.

Campbell, K., Picton, T. W., Wolfe, R. G., Maru, J., Baribeau-Braun, J., and Braun, C. (1981). Auditory potentials. *Sensus*, 1:21–31.

Caspary, D. M., Ling, L., Turner, J. G., and Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *The Journal of Experimental Biology*, 211:1781–1791.

CHABA (1988). Speech understanding and aging. *Journal of the Acoustical Society of America*, 83(3):859–895.

Chandrasekaran, B. and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology*, 47(2):236–246.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *Journal of the Acoustical Society of America*, 106(5):2719–2732.

Chimento, T. and Schreiner, C. (1990). Selectively eliminating cochlear michrophonic contamination from the frequency-following response. *Electroencephalography and clinical Neurophysiologogy*, 75:88–96.

Clinard, C. G. and Tremblay, K. L. (2013). Aging degrades the neural encoding of simple and complex sounds in the human brainstem. *Journal of the American Academy of Audiology*, 24(7):590–599.

Clinard, C. G., Tremblay, K. L., and Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hearing Research*, 264(1-2):48–55.

Cohn, N. B., Dustman, R. E., and Bradford, D. C. (1984). Age-related decrements in Stroop Color Test performance. *Journal of Clinical Psychology*, 40(5):1244–1250.

Craik, F. I. M. and Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In Craik, F. I. M. and Trehub, S., editors, *Aging and cognitive processes*, pages 191–211. Plenum, New York, NY.

Craik, F. I. M. and Jennings, J. (1992). Human Memory. In Craik, F. I. M. and Salthouse, T. A., editors, *Handbook of aging and cognition*, pages 51–110. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Cruickshanks, K., Wiley, T., Tweed, T., Klein, B., Klein, R., Mares-Perlman, J. A., and Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. *American Journal of Epidemiology*, 148(9):879–886.

Dajani, H. R., Purcell, D., Wong, W., Kunov, H., and Picton, T. W. (2005). Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, 52(9):1614–1618.

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal Of Verbal Learning And Verbal Behavior*, 19:450–466.

Daneman, M. and Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychonomic bulletin & review*, 3(4):422–433.

Darwin, C. J. and Carlyon, R. P. (1995). *Auditory grouping.* Academic Press, London.

Dau, T. (2003). The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *Journal of the Acoustical Society of America*, 113(2):936–950.

Davidson, D. J., Zacks, R. T., and Williams, C. C. (2003). Stroop interference, practice, and aging. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition*, 10(2):85–98.

Debruyne, F. (1986). Influence of age and hearing loss on the latency shifts of the auditory brainstem response as a result of increased stimulus rate. *Audiology*, 25(101-106).

Desjardins, J. L. and Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear and Hearing*, 34(3):261–272.

Don, M. and Eggermont, J. J. (1978). Analysis of the click-evoked brainstem potentials in man unsing high-pass noise masking. *Journal of the Acoustical Society of America*, 63(4):1084–1092.

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America*, 95(2):1053–64.

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *Journal of the Acoustical Society of America*, 107(1):538–546.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *Journal of the Acoustical Society of America*, 111(6):2897–2907.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2003). Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing. *Journal of the Acoustical Society of America*, 113(4):2084–2094.

Dubno, R. and Morgan, E. (1984). Effects of age and mild hearing loss on speech recognition in noise literature. *Journal of the Acoustical Society of America*, 76(1):87–96.

Eddins, D. A. (1993). Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region. *Journal of the Acoustical Society of America*, 93(1):470 – 479.

Eddins, D. A. (1999). Amplitude-modulation detection at low- and high-audio frequencies. *Journal of the Acoustical Society of America*, 105(2):829–837.

Eddins, D. A., Hall, J. W., and Grose, J. H. (1992). The detection of temporal gaps as a function of frequency region and absolute noise bandwidth. *Journal of the Acoustical Society of America*, 91(2):1069–1077.

Elhilali, M., Chi, T., and Shamma, S. A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication*, 41(2-3):331–348.

Elliott, T. M. and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS computational biology*, 5(3):1–14.

Festen, J. M. and Plomp, R. (1983). Relations between auditory functions in impaired hearing. *Journal of the Acoustical Society of America*, 73(2):652–662.

Festen, J. M. and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, 88(4):1725–1736.

Foo, C., Rudner, M., Rönnberg, J., and Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology*, 18:618–631.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America*, 115(5):2246–2256.

Frisina, D. R. and Frisina, R. D. (1997). Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hearing Research*, 106(1-2):95–104.

Fujikawa, S. M. and Weber, B. A. (1977). Effects of increased stimulus rate on brainstem electric response (BER) audiometry as a function of age. *Journal of the American Audiology Society*, 3(3):147–150.

Füllgrabe, C. (2013). Age-dependent changes in temporal-fine- structure processing in the absence of peripheral hearing loss. *American Journal of Audiology*, 22:313–315.

Galbraith, G., Arbagey, P., Comerci, N., and Rector, P. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*, 6:2363–2367.

Galbraith, G. C. (1994). Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalography and clinical Neurophysiologogy*, 92:321–330.

Gatehouse, S., Naylor, G., and Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, 42:S77–S85.

Gatehouse, S. and Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, 43:85–99.

Gelman, A. and Stern, H. (2006). The difference between significant and not significant is not itself statistically significant. *The American Statistician*, 60(4):328–331.

George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *Journal of the Acoustical Society of America*, 121(4):2362 – 2375.

Ghitza, O. (2001). On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *Journal of the Acoustical Society of America*, 110(3):1628–1640.

Gifford, R. H., Bacon, S. P., and Williams, E. J. (2007). An examination of speech recognition in a modulated background and forward masking in young and older listeners. *Journal of Speech, Language, and Hearing Research*, 50(4):857–864.

Gilbert, G. and Lorenzi, C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. *Journal of the Acoustical Society of America*, 119(4):2438–2444.

Gockel, H. E., Carlyon, R. P., Mehta, A., and Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology*, 12(6):767–782.

Gordon-Salant, S. (2005). Hearing loss and aging: New research findings and clinical implications. *The Journal of Rehabilitation Research and Development*, 42:9.

Gorga, M., Abbas, P., and Worthington, D. (1985). Stimulus calibration in ABR measurements. In Jacobsen, J., editor, *The auditory brainstem response*, pages 49–62. College-Hill, San Diego.

Gosselin, P. A. and Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*, 54:944–958.

Grose, J. H. and Mamo, S. K. (2010). Processing of temporal fine structure as a function of age. *Ear and Hearing*, 31(6):755–760.

Grose, J. H., Mamu, S. K., and Hall III, J. W. (2009). Age effects in temporal envelope processing: speech unmasking and auditory steady state responses. *Ear and Hearing*, 30(5):568–575.

Gustafsson, H. A. and Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *Journal of the Acoustical Society of America*, 95(1):518–29.

Harris, K. C., Eckert, M. A., Ahlstrom, J. B., and Dubno, J. R. (2010). Age-related differences in gap detection: effects of task difficulty and cognitive ability. *Hearing Research*, 264(1-2):21–29.

Hasher, L. and Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. In Bower, G., editor, *The Psychology of Learning and Motivation*, pages 193–225. Academic Press, New York, NY, 22 edition.

Hasher, L., Zacks, R. T., and May, C. P. (1999). Inhibitory control, circadian arousal, and age. In Gopher, D. and Koriat, A., editors, *Attention and performance XVII*, pages 653–675. MIT Press, Cambridge, MA.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* Springer-Verlag.

He, N., Horwitz, A. R., Dubno, J. R., and Mills, J. H. (1999). Psychometric functions for gap detection in noise measured from young and aged subjects. *Journal of the Acoustical Society of America*, 106(2):966–978.

He, N., Mills, J. H., Ahlstrom, J. B., and Dubno, J. R. (2008). Age-related differences in the temporal modulation transfer function with pure-tone carriers. *The Journal of the Acoustical Society of America*, 124(6):3841–3849.

He, N., Mills, J. H., and Dubno, J. R. (2007). Frequency modulation detection: effects of age, psychophysical method, and modulation waveform. *Journal of the Acoustical Society of America*, 122(1):467–477.

Heinz, M. G. and Swaminathan, J. (2009). Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech. *Journal of the Association for Research in Otolaryngology*, 10:407–423.

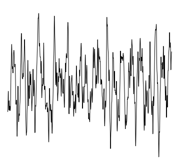
Helfer, K. S. and Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and hearing*, 29(1):87–98.

Helfer, K. S. and Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech and Hearing Research*, 33(1):149–55.

Hopkins, K. and Moore, B. C. J. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America*, 130(1):334–49.

Horn, J. I. and Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26:107–129.

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., and Kraus, N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):13022–13027.



Horst, W. (1987). Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects. *Journal of the Acoustical Society of America*, 82(3):874–885.

Houtgast, T. and Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology*, 47(6):287–295.

Houtsma, A. J. M. and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America*, 87(1):304 – 310.

Howard-Jones, P. and Rosen, S. (1993). The perception of speech in gluctuating noise. *Acustica*, 78:258–272.

Humes, L. E. (1996). Speech understanding in the elderly. *Journal of the American Academy of Audiology*, 7(3):161–167.

Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., and Strauser, L. E. (2007). Auditory speech recognition and visual text recognition in younger and older adults: similarities and differences between modalities and the effects of presentation rate. *Journal of speech, language, and hearing research*, 50(2):283–303.

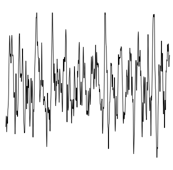
Humes, L. E. and Dubno, J. R. (2010). Factors affecting speech understanding in older adults. In Gordon-Salant, S., Frisina, R., Popper, A. N., and Fay, R. R., editors, *The Aging Auditory System*, pages 211–258. Springer, New York, NY.

Humes, L. E. and 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.

Janse, E. (2009). Processing of fast speech by elderly listeners. *Journal of the Acoustical Society of America*, 125(4):2361–2373.

Janssen, T., Steinhoff, H.-J., and Böhnke, F. (1991). Zum entstehungsmechanismus der frequentzfolgpotentiale. *Oto-Rhino-Laryngologia Nova*, 1:16–25.

Jeng, F.-C., Hu, J., Dickman, B., Lin, C.-Y., Lin, C.-D., Wang, C.-Y., Chung, H.-K., and Li, X. (2011). Evaluation of two algorithms for detecting human



frequency-following responses to voice pitch. *International Journal of Audiology*, 50:14–26.

Jerger, J. and Johnson, K. (1988). Interactions of age, gender, and sensorineural hearing loss on ABR latency. *Ear and Hearing*, 9(4):168–176.

Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., and Kraus, N. (2008). Brainstem encoding of voiced consonant-vowel stop syllables. *Clinical Neurophysiology*, 119(11):2623–2635.

Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. (1994). Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. *Journal of neurophysiology*, 71(3):1022–1036.

Kalikow, D., Stevens, K., and Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61(5):1337–1351.

Kausler, D. H. (1994). *Learning and memory in normal aging.* Academic Press, San Diego.

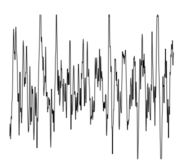
King, A., Hopkins, K., and Plack, C. J. (2014). The effects of age and hearing loss on interaural phase difference discrimination. *Journal of the Acoustical Society of America*, 135(1):342–351.

Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2014). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *Journal of the Acoustical Society of America*, 135(3):1596–1606.

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., and Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology*, pages Article ID 865731, 11 pages.

Konrad-Martin, D., Dille, M. F., McMillan, G., Griest, S., McDermott, D., Fausti, S. a., and Austin, D. F. (2012). Age-related changes in the auditory brainstem response. *Journal of the American Academy of Audiology*, 23:18–35.

Kramer, S. and Teas, D. (1982). Forward masking of auditory nerve (N1) and brainstem (wave V) response in humans. *Journal of the Acoustical Society of America*, 72(3):795 – 803.



Kramer, S. E., Zekveld, A. A., and Houtgast, T. (2009). Measuring cognitive factors in speech comprehension: the value of using the Text Reception Threshold test as a visual equivalent of the SRT test. *Scandinavian Journal of Psychology*, 50(5):507–515.

Krishnan, A. (2008). Frequency-Following Response. In Burkard, R. F., Don, M., and Eggermont, J. J., editors, *Auditory evoked potentials: Basic principles and clinical application*, chapter 15, pages 313–333. Lippincott Williams & Wilkins, Philadelphia.

Krishnan, A. and Parkinson, J. (2000). Human frequency-following response: Representation of tonal sweeps. *Audiology & Neuro-Otology*, 5:312–321.

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25(1):161–8.

Krizman, J., Skoe, E., and Kraus, N. (2012). Sex differences in auditory subcortical function. *Clinical Neurophysiology*, 123(3):590–597.

Kujawa, S. G. and Liberman, M. C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *The Journal of Neuroscience*, 29(45):14077–14085.

Lasky, R. E. and Rupert, A. L. (1982). Temporal masking of auditory evoked brainstem responses in human newborns and adults. *Hearing Research*, 6(3):315–34.

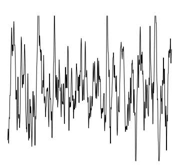
Léger, A. (2012). *Investigation of speech processing in frequency regions where absolute thresholds are normal for hearing-impaired listeners.* PhD thesis.

Leigh-Paffenroth, E. and Fowler, C. G. (2006). Amplitude-modulated auditory steady-state responses in younger and older listeners. *Journal of the American Academy of Audiology*, 17:582–597.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49:467–477.

Lin, F., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H. N., Ferrucci, L., and Simonsick, E. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, 173(4):293–299.

Liu, C. and Eddins, D. A. (2008). Effects of spectral modulation filtering on vowel identification. *Journal of the Acoustical Society of America*, 124(3):1704–1715.



Lopez-Poveda, E. a. and Barrios, P. (2013). Perception of stochastically undersampled sound waveforms: a model of auditory deafferentation. *Frontiers in Neuroscience*, 7(Article 124):1–13.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49):18866–18869.

Lunner, T. and Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology*, 18:604–617.

Makary, C. a., Shin, J., Kujawa, S. G., Liberman, M. C., and Merchant, S. N. (2011). Age-related primary cochlear neuronal degeneration in human temporal bones. *Journal of the Association for Research in Otolaryngology*, 12(6):711–717.

Marmel, F., Linley, D., Carlyon, R. P., Gockel, H. E., Hopkins, K., and Plack, C. J. (2013). Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *Journal of the Association for Research in Otolaryngology*, 14(5):757–766.

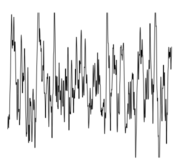
Mehraei, G., Leek, M. R., Gallun, F. J., and Bernstein, J. G. W. (2013). Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility. *Manuscript submitted for publication.*

Meister, H., Schreitmüller, S., Grugel, L., Beutner, D., Walger, M., and Meister, I. (2013). Examining speech perception in noise and cognitive functions in the elderly. *American Journal of Audiology*, 22:310–312.

Middelweerd, M., Festen, J. M., and Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology*, 29:1–7.

Miller, G. A., Heise, G. A., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of experimental psychology*, 41(5):329–335.

Miller, G. A. and Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 22(2):167 – 173.



Møller, A. R. (2007). Neural generators for auditory brainstem evoked potentials. In *Auditory evoked potentials: Basic principles and clinical application.*, pages 336–354.

Moore, B. C. (2007). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. Wiley-Interscience, Chichester, England.

Moore, B. C. (2008a). *An introduction to the psychology of hearing*. Emerald Group Publishing Ltd, Bingley, UK, 5 edition.

Moore, B. C. (2008b). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, 9(4):399–406.

Moore, B. C. (2012). The importance of temporal fine structure for the intelligibility of speech in complex backgrounds. In T. Dau, J.C. Dalsgaard, M. J. and Poulsen, T., editors, *Speech Perception and Auditory Disorders*, number August 2011, pages 21–32. Centertryk A/S, Denmark.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (1992). Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss. *The Journal of the Acoustical Society of America*, 92(4 Pt 1):1923–1932.

Moore, B. C. and Sek, A. (1996). Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. *The Journal of the Acoustical Society of America*, 100(4 Pt 1):2320–2331.

Moore, B. C. J. and Sek, A. (1995). Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation). *Journal of the Acoustical Society of America*, 97(4):2468–2478.

Mościcki, E. K., Elkins, E. F., Baum, H. M., and McNamara, P. M. (1985). Hearing loss in the elderly: an epidemiologic study of the Framingham Heart Study Cohort. *Ear and hearing*, 6(4):184–90.

Moushegian, G., Rupert, A., and Stillman, R. D. (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiologogy*, 35:665–667.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, 12(4):313–324.

Nelson, P. B., Carney, A. E., and Nelson, D. A. (2003). Understanding speech in modulated interference : Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*, 113:961–968.

Nie, K., Stickney, G., and Zeng, F.-G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE transactions on bio-medical engineering*, 52(1):64–73.

NIH and WHO (2011). Global health and aging. Technical report.

Nilsson, M., Soil, S. D., and Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the acoustical society of america*, 95(2):1085–1099.

Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22):11773–11779.

Oxenham, A. J. and Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *Journal of the Acoustical Society of America*, 125(1):457–68.

Peters, R. W. and Hall III, J. W. (1994). Comodulation masking release for elderly listeners with relatively normal audiograms. *Journal of the Acoustical Society of America*, 96:2674–2682.

Peters, R. W., Moore, B. C., and Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, 103(1):577–87.

Pichora-Fuller, M. K. and Schneider, B. A. (1992). The effect of interaural delay of the masker on masking-level differences in young and old adults. *Journal of the Acoustical Society of America*, 91(4 Pt 1):2129–35.

Pichora-Fuller, M. K., Schneider, B. A., and 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.

Pichora-Fuller, M. K., Schneider, B. A., Macdonald, E., Pass, H. E., and Brown, S. (2007). Temporal jitter disrupts speech intelligibility: a simulation of auditory aging. *Hearing Research*, 223:114–121.

Pichora-Fuller, M. K. and Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology*, 42(2):S11–S16.

Plomp, R. and Mimpen, A. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18:43–52.

Poth, E. A., Boettcher, F. A., Mills, J. H., and Dubno, J. R. (2001). Auditory brainstem responses in younger and older adults for broadband noises separated by a silent gap. *Hearing Research*, 161(1-2):81–6.

Pronk, M., Deeg, D. J. H., Festen, J. M., Twisk, J. W., Smits, C., Comijs, H. C., and Kramer, S. E. (2013). Decline in older persons' ability to recognize speech in noise: the influence of demographic, health-related, environmental, and cognitive factors. *Ear and Hearing*, 34(6):722–732.

Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *Journal of the Acoustical Society of America*, 116(6):3581–3593.

Rajan, R. and Cainer, K. E. (2008). Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. *Neuroscience*, 154(2):784–795.

Ramus, F., Rosen, S., Dakin, S., Day, B., Castellote, J., White, S., and Frith, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*, 126(4):841–865.

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card sorting test: a meta-analytic review. *Psychology and aging*, 19(3):482–94.

Robertson, I. H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1996). The structure of normal human attention: The Test of Everyday Attention. *Journal of the International Neuropsychological Society*, 2(6):525–534.

Roccaforte, W. H., Burke, W. J., Bayer, B. L., and Wengel, S. P. (1992). Validation of a telephone version of the mini-mental state examination. *Journal of the American Geriatrics Society*, 40:697–702.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *International Journal of Audiology*, 42(Suppl 1):S68–S76.

Rönnberg, J., Lyxell, B., Arlinger, S., and Kinnefors, C. (1989). Visual evoked potentials: relation to adult speechreading and cognitive function. *Journal of Speech and Hearing Research*, 32:725–735.

Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. A. (2010). When cognition kicks in: working memory and speech understanding in noise. *Noise & Health*, 12(49):26326–26329.

Rosen, S., Souza, P., Ekelund, C., and Majeed, A. A. (2013). Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *Journal of the Acoustical Society of America*, 133(4):2431–2443.

Rossi-Katz, J. and Arehart, K. H. (2009). Message and Talker Identification of the Competing Message. *Journal of Speech, Language, and Hearing Research*, 52:435–454.

Rothauser, E. H., Chapman, N. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. *IEEE transactions on audio and electroacoustics*, 17:225–246.

Rudner, M., Foo, C., Sundewall-Thorén, E., Lunner, T., and Rönnberg, J. (2008). Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. *International Journal of Audiology*, 47 Suppl 2:S91–8.

Rudner, M., Rönnberg, J., and Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *Journal of the American Academy of Audiology*, 22(3):156–167.

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current biology*, 22(15):1417–1422.

Russo, N., Nicol, T., Musacchia, G., and Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical neurophysiology*, 115(9):2021–2030.

Salthouse, T. A. (1985). *A theory of cognitive aging.* Amsterdam, The Netherlands.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3):403–428.

Sand, T. (1991). BAEP amplitudes and amplitude ratios: relation to click polarity, rate, age and sex. *Electroencephalography and Clinical Neurophysiology*, 78(4):291–296.

Schaette, R. and McAlpine, D. (2011). Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *The Journal of neuroscience*, 31(38):13452–7.

Schmiedt, R. A., Mills, J. H., and Boettcher, F. A. (1996). Age-related loss of activity of auditory-nerve fibers. *Journal of Neurophysiology*, 76(4):2799–2803.

Schneider, B. A. and Hamstra, S. J. (1999). Gap detection thresholds as a function of tonal duration for younger and older listeners. *Journal of the Acoustical Society of America*, 106(1):371–80.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (1994). Gap detection and the precedence effect in young and old adults. *Journal of the Acoustical Society of America*, 95(2):980–991.

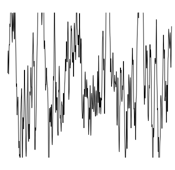
Schooneveldt, G. P. and Moore, B. C. J. (1987). Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotonic versus dichotic presentation of the flanking band. *Journal of the Acoustical Society of America*, 82(6):1944 – 1956.

Schuknecht, H. and Woellner, R. (1953). Hearing losses following partial section of the cochlear nerve. *The Laryngoscope*, 6:441–465.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Sergeyenko, Y., Lall, K., Liberman, M. C., and Kujawa, S. G. (2013). Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *The Journal of Neuroscience*, 33(34):13686–13694.

Shamma, S. and Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *The Journal of the Acoustical Society of America*, 133(5):2818–2833.



Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234):303–304.

Sheldon, S., Pichora-Fuller, M. K., and Schneider, B. a. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *Journal of the Acoustical Society of America*, 123(1):489–499.

Skoe, E. and Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear and Hearing*, 31(3):302–324.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876):87–90.

Snell, K. B. (1997). Age-related changes in temporal gap detection. *Journal of the Acoustical Society of America*, 101(4):2214–2220.

Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (2002). Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity. *Journal of the Acoustical Society of America*, 112(2):720–727.

Snyder, J. S. and Alain, C. (2005). Age-related changes in neural activity associated with concurrent vowel segregation. *Cognitive Brain Research*, 24(3):492–499.

Sohmer, H., Pratt, H., and Kinarti, R. (1977). Sources of frequency following responses (FFR) in man. *Electroencephalography and Clinical Neurophysiology*, 42(5):656–664.

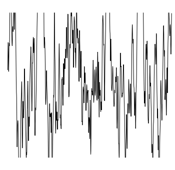
Song, J. H., Banai, K., Russo, N. M., and Kraus, N. (2006). On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiology & Neuro-Otology*, 11(4):233–241.

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (2011). Perception of speech in noise: neural correlates. *Journal of Cognitive Neuroscience*, 23(9):2268–2279.

Souza, P. E. and Turner, C. W. (1994). Masking of speech in young and elderly listeners with hearing loss. *Journal of Speech and Hearing Research*, 37(3):655–661.

Speranza, F., Daneman, M., and Schneider, B. A. (2000). How Aging Affects the Reading of Words in Noisy Backgrounds. *Psychology and Aging*, 15(2):253–258.

Stillman, R. D., Crow, G., and Moushegian, G. (1978). Components of the frequency-following potential in man. *Electroencephalography and Clinical Neurophysiology*, 44:438–446.



Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (1998). Temporal processing in the aging auditory system. *The Journal of the Acoustical Society of America*, 104(4):2385–2399.

Stuart, A. and Phillips, D. P. (1996). Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusis listeners. *Ear and Hearing*, 17(6):478–489.

Summers, V. and Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research*, 41:1294–1306.

Takahashi, G. A. and Bacon, S. P. (1992). Modulation Detection , Modulation Masking , and Speech Understanding in Noise in the Elderly. *Journal of Speech and Hearing Research*, 35:1410–1421.

Torgesen, J. K., Wagner, R. K., and Rashotte, C. A. (1999). Test of word reading efficiency (TOWRE).

Tremblay, K. L. and Burkard, R. F. (2007). The aging auditory system: Confounding effects of hearing loss on AEPs. In *Auditory evoked potentials: Basic principles and clinical application.*, pages 403–427.

Tun, P. A., O'Kane, G., and Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, 17(3):453–467.

Tun, P. A. and Wingfield, A. (1999). One voice too many: adult age differences in language processing with different types of distracting sounds. *The journals of gerontology. Series B, Psychological sciences and social sciences*, 54(5):P317–P327.

Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., and Jolles, J. (2006). The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. *Journal of Clinical and Experimental Neuropsychology*, 28(6):998–1009.

Van der Linden, M., Brédart, S., and Beerten, A. (1994). Age-related differences in updating working memory. *British journal of psychology*, 85(Pt 1):145–152.

van Rooij, J. C. and Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *The Journal of the Acoustical Society of America*, 88(6):2611–2624.

Vander Werff, K. R. and Burns, K. S. (2011). Brain stem responses to speech in younger and older adults. *Ear and Hearing*, 32(2):168–180.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, 18(2):332–339.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *Journal of the Acoustical Society of America*, 66(5):1364–1380.

Vongpaisal, T. and Pichora-Fuller, M. K. (2007). Effect of age on F0 difference limen and concurrent vowel identification. *Journal of Speech, Language, and Hearing Research*, 50:1139–1156.

Walton, J., Orlando, M., and Burkard, R. (1999). Auditory brainstem response forward-masking recovery functions in older humans with normal hearing. *Hearing Research*, 127(1-2):86–94.

Walton, J. P., Frisina, R. D., and O'Neill, W. E. (1998). Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse. *The Journal of Neuroscience*, 18(7):2764–2776.

WHO (2012). Aging and life course: Interesting facts about ageing. Retrieved July 2014, from http://www.who.int/ageing/about/facts/en/.

Wilson, R. H. and Carhart, R. (1969). Influence of pulsed masking on the threshold for spondees. *Journal of the Acoustical Society of America*, 46(4):998–1010.

Worden, F. G. and Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and clinical neurophysiology*, 25(1):42–52.

Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T., and Houtgast, T. (2007). The development of the text reception threshold test: a visual analogue of the speech reception threshold test. *Journal of Speech, Language, and Hearing Research*, 50(3):576–584.

Zekveld, A. A., Kramer, S. E., and 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.

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.-Y., and Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues. *Journal of the Acoustical Society of America*, 116(3):1351–1354.