Speech Perception in Autism Spectrum Disorder: Susceptibility to Masking and Interference

Katharine Rachel Mair

Speech, Hearing & Phonetic Sciences

UCL

A thesis submitted for the degree of Doctor of Philosophy

November 2013

Declaration

I, Katharine Mair, 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.

Katharine Mair

Abstract

High-functioning adults with autism spectrum disorder (ASD) frequently report difficulties with speech perception in background noise, which cannot be explained simply by an impairment in peripheral hearing or structural language ability. In spite of the apparent prevalence of this problem, however, only a handful of studies so far have evaluated speech reception thresholds (SRTs) in this group under controlled conditions, and then only with a limited range of (mainly non-speech) masking sounds. Results have indicated relatively minor deficits in some types of background noise, but not in others: at most, ASD listeners have required the signal-to-noise ratio to be around 3 dB more favourable than matched controls to report speech with 50% accuracy.

This thesis describes a series of sentence in noise tasks, in which SRTs were measured for a far wider range of speech and non-speech masking sounds than before. Two groups of normally-hearing young adult participants completed the experiments: one group who had received a clinical diagnosis of Asperger's syndrome, and the other a group of neurotypical controls matched for age, verbal IQ and non-verbal IQ. The key results were that a substantial proportion of the ASD group ('ASD+') performed consistently within the normal range across the majority of target sentence types and masking conditions, whereas around half ('ASD-') tended to perform significantly poorly. When these two subgroups' results were analysed separately, it became clear that the ASD- listeners showed much greater deficits in SRT with speech maskers than with non-speech. Moreover, they were particularly affected by masking speech with similar perceptual and linguistic features to the target material, showing deficits of up to 15 dB in these conditions. This pattern of performance is strongly suggestive of an impairment in auditory stream segregation and/or selective attention to speech in a subset of high-functioning adults with ASD.

Acknowledgements

The research described here would simply not have been possible without the help of my participants, all of whom were remarkably generous with their time and patient in the face of multiple auditory tests. Their comments and insights have helped to shape this thesis. The work was funded by an ESRC studentship.

My primary supervisor Stuart Rosen has been an inspiration throughout, and his depth of knowledge, great good humour and willingness to discuss new ideas have made working with him a unique and unforgettable experience. I am also very grateful to John Swettenham, my second supervisor, and to his former PhD student Anna Remington, for many helpful discussions and for sharing their pool of participants with ASD. Andrew Faulkner has been so generous with advice and encouragement over the years that he has often seemed like an unofficial third supervisor, and his guidance on key theoretical and methodological issues has been invaluable. On the technical side, I would have been lost without Steve Nevard and Dave Cushing who have provided outstanding support, alongside much-needed common sense, in moments of equipment failure. Thanks also to Breege and all at Chandler House library, and to Dom at the RNID library, for assistance in tracking down obscure books and documents, and for being willing to pilot listening tasks in their spare time.

I have been helped through the PhD by more colleagues and friends at UCL than I can possibly list individually, but I would not have survived the process without the wisdom, kindness and friendship of Sophie Gates, Gaston Hilkhuysen, Tim Schoof and Yasna Pereira. Special thanks also to Tim Green, Michele Pettinato, Kayoko Yanigasiwa, Chierh Cheng, Young Shin Kim, Jeong Sug Keong, and all fellow inhabitants of Room 326 at Chandler House over the past few years. Thank you to Volker Dellwo for help with Praat and to Muhsin for looking after my participants in reception and keeping me cheerful.

Jan Eaton has been a calm voice of reason throughout the write-up and I cannot thank her enough for her warmth, kindness and encouragement. Likewise, thanks to Dede Rose and all friends at East London Late Starters Orchestra for keeping me sane, cheering me on, and tolerating some very bad cello playing when work cut into music practice time.

Last, but not least, thank you to my parents - I could not have done it without you!

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Speech perception in everyday life: a challenge for adults with ASD

High-functioning adults with autism spectrum disorder (ASD) are often eloquent in describing the effects their atypical sensory/perceptual style can have on the way they interpret the world around them. It is not uncommon for these individuals to feel overloaded and exhausted by the amount of sensory information in typical educational, workplace and social settings (Grandin, 2006; Williams, 2008; Tammett, 2012). Whilst visual processing in ASD has attracted the lion's share of research over the past few decades (for reviews, see Behrmann, Thomas, & Humphreys, 2006; Simmons et al., 2009), evidence is fast accumulating that auditory perception may also be abnormal (O'Connor, 2012). One particularly problematic aspect of auditory processing in ASD, and the one which forms the subject-matter of this thesis, is vividly illustrated by the following anecdote from Tony Attwood's clinical experience, recounted in his *Complete Guide to Asperger's Syndrome*:

> 'I remember a child with Asperger's syndrome who was in an open-plan classroom that comprised two classes. The teacher of his class was reading out a maths test while the teacher in the other class was reading out a spelling test. When his teacher marked his test paper, she noted he had written the answers to both tests.'
> (Attwood, 2006, p. 221)

Inability to 'screen out' background noise, and particularly background speech, in order to focus on one particular talker is commonly reported in the ASD literature, and the difficulties this causes for communication and social interaction can be distressing. The effects on social participation are significant as well, especially in early adulthood and at university where social events tend to take place in noisy, multi-talker environments

(Madriaga, 2010). The following extract is from a post which appeared on a discussion board for people with Asperger's syndrome, and which received many responses from people with similar problems:

> 'Effectively doubling my social ineptness in loud pubs and clubs (i [*sic*] know that it's loud for everyone, but lean toward them or speak in their ear and they get what you're saying) - It doesn't matter if someone shouts in my ear or leans toward me to speak, the background noise of music, people shouting/talking, and whatever they are trying to say to me all jumble together to a sort of 'white noise'.
>
> Being with a group of friends, and sitting in the middle, one conversation going on to my left and one to my right, no excessive background noise, but i can follow and participate in a singular conversation, but when two are going at once i can't follow either, so i usually just sit in silence.' (Newg, 2008)

Personal anecdotes such as these highlight the very real nature of the difficulties faced even by adults with ASD who have relatively strong language skills. In spite of the clear importance of this problem, however, we have remarkably little evidence regarding the possible underlying causes, and very few research studies have so far even addressed the issue. The one key study in this field was conducted by Alcántara and colleagues, who demonstrated that a group of adolescents and young adults with high-functioning ASD did indeed perform worse than matched neurotypical controls on tests of sentence perception in noise under laboratory conditions (Alcantara, Weisblatt, Moore, & Bolton, 2004). In that study, moreover, the ASD group's deficit seemed highly dependent on the nature of the masking sounds involved. They did not perform significantly worse than the controls with a steady background noise masker, nor with a noise masker which contained spectral dips, which implies that their cochlear function, and in particular frequency resolution, was not impaired. They did perform worse however, when the noise masker was amplitude modulated so that the overall intensity fluctuated at a relatively slow rate, and also when background speech was presented as a masker. This pattern of results hints at the possibility of something more than a purely peripheral auditory deficit, but it is not possible to be much more specific on the basis of that particular study. Interestingly, although the group effects in the latter two conditions were significant, the actual differences in mean speech reception threshold (SRT; i.e. signal-to-noise ratio at which sentences were reported with 50% accuracy) were fairly small: only 2 - 3.5 dB. Although this may be sufficient to have a measurable impact on everyday conversation, it hardly seems to reflect the high degree of impairment and frustration expressed in subjective reports.

A similar observation applies to a study by Groen and colleagues, who presented single words (rather than sentences) in similar types of background noise to those used by Alcántara, but not including a background speech masker (Groen et al., 2009). Across four noise conditions (steady pink noise, fluctuating pink noise, steady spectral ripple noise, and fluctuating spectral ripple noise) the largest group difference in SRT between children with high-functioning ASD and controls was only just over 1 dB, and that was with the steady pink noise, where the ASD group were actually the better performers. Unsurprisingly, there was no main effect of group in the statistical analysis, and the only significant finding (of a complex interaction between group, spectral detail of the noise and temporal detail of the noise) seems to be driven almost entirely by results from the first two conditions involving the pink noise with unchanging spectral content. In these the ASD group performed slightly better, on average, than controls with steady noise (as described) and rather worse when the noise fluctuated. This meant that they appeared to show a deficit of around 2 dB in 'release from masking' (improvement in performance when a noise fluctuates rather than remaining steady). Since over 1 dB of this apparent deficit actually derives from slightly enhanced performance in the baseline steady noise condition, it once again hardly seems to explain poor performance in everyday life. These are the only two published auditory studies to have looked in any detail at the effects of masking sounds on speech processing in high-functioning ASD, and the results fall somewhat short of supplying a convincing account of what makes it so difficult for these listeners to understand speech in noisy environments. The need for more research in this area is self-evident, and much can be learnt about potential contributing factors - both at a theoretical and at a methodological level - from work in other developmental disorders.

1.2 Speech perception difficulties in developmental disorders

Speech is a complex auditory stimulus and our ability to perceive and understand it relies on a highly interactive hierarchy of processing abilities, which range from the transduction of sound waves in the cochlea all the way up to recognition and interpretation of the intended message in the cortex. Clearly, there is potential for a processing task as intricate as this to be undermined by deficits in any of the sensory or higher-level cognitive skills involved. Not only that, but the speech signal itself is often of poor quality, whether due to characteristics of the source speaker (such as an unfamiliar accent or an upper respiratory tract infection), the level of masking noise within a given auditory scene, or the mode of transmission (e.g. phones, radios, tannoys, etc., which can all distort the signal).

It is unsurprising, therefore, that a variety of clinical populations experience difficulty with speech perception, and even more so in the presence of background noise. The most obvious case is cochlear hearing impairment, in which damage to the inner ear structures means the speech signal is poorly represented at even the earliest stages of the processing chain (Moore, 2007). Individuals with developmental language disorders, on the other hand, must have normal audiometric thresholds to meet diagnostic criteria, but in these groups atypical speech processing seems to occur somewhat further up the chain (Bishop, 1997; Badecker, 2008). Unfortunately, even in the case of cochlear hearing impairment, where the physical locus of the core deficit is relatively well-defined, impaired speech perception stems from more than just a straightforward decrease in audibility. Reduced frequency and temporal resolution, and poor transmission of temporal fine structure, also affect hearing impaired listeners' ability to discriminate effectively between sounds and to track auditory objects (Moore, 1987). Thus the ability of these individuals to perceive speech in everyday settings may be worse than could be predicted from a simple audiogram. This serves as a reminder that even relatively early stages of the peripheral / subcortical auditory processing which encode basic features of a sound rely on the integrity of an intricate system.

When it comes to language disorders, the situation is even more complicated. Deficits in phonological processing are a key feature of developmental language impairment, but it is often suggested that specific language impairment (SLI) and specific reading disorder (SRD) may actually stem from difficulties with lower-level non-speech auditory processing - in particular from impaired perception of rapidly changing stimuli (see, for example, Tallal & Piercy, 1973; Witton et al., 1998). Whilst numerous studies have shown that non-speech auditory deficits do frequently co-occur with language impairments, however, the fact that this is not always the case means the evidence for a bottom-up causal link is equivocal at best (Ramus et al., 2003; Rosen, 2003; 2006). In reality, although it seems intuitive to assume that performance on non-speech psychoacoustic tests of frequency or temporal resolution might correlate with speech perceptual skills, in normally-hearing listeners this assumption is not well-supported by the evidence (for a discussion of this issue see Moore, 2012). On the other hand, of course, domain-specific theories of SLI and dyslexia which argue that the core deficit is linguistic (e.g. van der Lely, 2005) struggle to explain poor performance on psychoacoustic tasks such as gap detection, as well as slow reaction times on other non-linguistic cognitive (rather than perceptual) tasks (Miller, Kail, Leonard, & Tomblin, 2001).

Recent advances in the controversial field of auditory processing disorder have provided thought-provoking data on the relationship between 'pure' auditory processing and listening abilities. The term 'auditory processing disorder' (APD) is used to describe a developmental (or acquired) impairment in the processing of auditory information within the central nervous system, which results in poor identification, discrimination, and ordering of sounds, and increased susceptibility to masking - in the absence of a peripheral hearing impairment (ASHA, 2005; BSA, 2011). Unlike SLI and SRD, it is not suggested that APD is - at root- a language disorder and, on current criteria, deficits in non-speech auditory processing must be present for a diagnosis, but speech processing is also affected and is typically the presenting difficulty (Rosen, Cohen, & Vanniasegaram, 2010). Poor speech perception in noise appears to be one of the more common symptoms (Dawes, Bishop, Sirimanna, & Bamiou, 2008; Lagace, Jutras, Giguere, & Gagne, 2011), although not all studies confirm this (Ferguson, Hall, Riley, & Moore, 2011). There also appears to be a high degree of comorbidity between APD, SLI and SRD (Dawes & Bishop, 2009; Sharma, Purdy, & Kelly, 2009). In one (small-scale) study based on the medical notes of a group of children diagnosed with APD, 25% also had a diagnosis of SRD, and 13% had comorbid language impairment (Dawes et al., 2008). A larger study of children with suspected APD found even more overlap: on a battery of listening tests and language assessments, 45% showed evidence of both APD, language impairment and reading difficulties (Sharma et al., 2009). This again seems to reflect both the complexity of the interactions between non-speech auditory processing and language processing, and the difficulties faced by those attempting to define and diagnose impairments affecting these functions. Many of the most widely-administered assessments for APD, such as the SCAN-C for instance (Hind, 2007), actually involve speech stimuli, making it very difficult to distinguish whether poor performance might be due to an underlying language disorder or to auditory processing difficulties more generally.

Moore and colleagues tested the hypothesis that children's performance on individual psychoacoustic tasks might in fact have more to do with supra-modal cognitive abilities, including the ability to follow task-instructions, than on auditory processing abilities themselves (Moore, Ferguson, Edmondson-Jones, Ratib, & Riley, 2010). Rather than focusing on a group comparison between children diagnosed with APD and matched controls, they carried out a large population study in school-age children. In order to derive relatively pure measures of auditory processing ability (specifically frequency and temporal resolution), whilst controlling for more cognitive skills more generally, they calculated difference scores based on individuals' thresholds for pairs of listening tasks which were virtually identical in terms of their procedural demands (e.g. backward masking with 0 ms masker onset delay or backward masking with 50 ms masker onset delay). They found that whereas thresholds on the individual listening tasks were moderately correlated with cognitive measures such as non-verbal IQ and digit span and also with speech perception (non-words) in noise, the derived measures were not. They also made the striking observation that children who scored poorly on the individual listening tasks (those in the bottom 5% of the sample) showed significantly more variable performance within those tasks (measured by examining the detail of each child's track from the adaptive staircase procedure) than did children who scored normally. Taking this variability to reflect fluctuating attention to the task, the authors suggested that what is currently described as 'auditory processing disorder' may actually be a more general cognitive (or even specifically attentional) impairment which affects auditory processing. In a factor analysis, cognitive test scores and the variability measure accounted for far more of the variance in speech-in-noise perception and in parent-reported communication difficulty than did any of the auditory processing measures.

A thorough exploration of these issues goes beyond the scope of this introduction, but even in overview there is much to be learned both from the way the relationship between speech/language processing and non-speech auditory processing presents clinically, and from the approaches which have been taken to evaluating this relationship scientifically. There is a clear difficulty in isolating measures of speech processing from more general perceptual and cognitive processes even when language processing itself is - arguably - the core area of impairment. Likewise, when auditory processing per se is thought to be impaired, language and cognitive skills appear to play a more important role in psychoacoustic task performance than abilities such as temporal or frequency resolution. This is important, not least because both language impairment and atypical non-speech auditory processing are also found in ASD. APD also shows substantial rates of comorbidity with autism spectrum disorder (ASD): 9% in Dawes' study (Dawes et al., 2008), and 13 % in a more recent paper (Ferguson et al., 2011). Even amongst children diagnosed with APD but not ASD there appear to be high levels of autistic features (Dawes & Bishop, 2009).

The defining area of impairment in ASD is not (currently) thought to be either language processing or auditory processing, however, and there is a characteristic pattern of strengths and weaknesses within these areas which seems to distinguish ASD from SLI and APD as much as it resembles them. These strengths and weaknesses themselves raise questions about exactly how poor speech perception in noise might arise in ASD.

1.3 Autism spectrum disorder: definition and core features

The term 'autism spectrum disorder' refers to an early-onset neurodevelopmental condition which is currently diagnosed on the basis of deficits in three core areas of behaviour: impaired social interaction, impaired communication and repetitive behaviours/interests (American Psychiatric Association, 1994; World Health Organization, 1992). A recent UK population study indicated the prevalence of ASD to be 116.1 per 10 000 (i.e. around 1.2 %; Baird et al., 2006), which is broadly in line with other recent studies and suggests that ASD is far more common than was thought 50 years ago (for a review, see Fombonne, 2009). In spite of advances in our understanding of the genetics and neuroscience of autism (for reviews see Geschwind, 2011 and Minshew & Williams, 2007), a purely objective method for determining whether an individual has an ASD, ready for use in everyday clinical practice, still appears to be some way off (although see, for example, Skafidas et al., 2012). Instead, a decision must be made as to whether an individual meets the standard behavioural criteria outlined in ICD 10 or DSM IV. To reach this decision clinicians rely heavily on a combination of parent/teacher reports of developmental history and everyday functioning, and structured observation. A variety of published assessments (including screening instruments) are used to support the process, and amongst those viewed as the 'gold-standard' for ASD diagnosis are the ADI-R for parent interviews and the ADOS-G for behavioural observation (Lord et al.,2000; Lord, Rutter, & Lecouteur, 1994). Whilst these assessments are designed for administration by professionals, with intensive training to ensure inter-rater reliability, the fact remains that in rating the severity of a behavioural impairment an element of personal judgement is being made. Since autism is a highly heterogeneous condition, the lack of truly objective measures of severity for each behavioural dimension is as problematic for researchers as it is for clinicians.

Cognitive ability varies widely amongst those diagnosed with an ASD: in one recent study, 55% of a large sample of UK children diagnosed with ASD had IQ < 70, and 3% had an IQ > 115 (Charman et al., 2011). Moreover, whilst communication - encompassing pragmatic and discourse level language skills - is by definition impaired in ASD, structural language skills (i.e. phonology, syntax and morphology) can fall anywhere on the scale from non-verbal through to the level of a published author (for example Daniel Tammett or Temple Grandin). Recent work indicates that around 14-20% of children with autism have not achieved functional speech by age 9 (Lord, Risi, & Pickles, 2004), whilst earlier estimates had put the percentage of non-verbal autistic adults even higher, at around 50% (for a review see Pickett, Pullara, O'Grady, & Gordon, 2009). At the other extreme, meanwhile, are highly able individuals, typically diagnosed with Asperger's syndrome (a subtype of ASD in the DSM-IV system), who by definition have shown no significant delay in language or cognitive development. Asperger's syndrome appears to be much rarer than typical childhood autism, with the ratio of autism to Asperger's syndrome based on prevalence being approximately 3 or 4 to 1 (Fombonne, 2009).

At the time of writing, both the definition and diagnosis of autism spectrum disorder is in a state of flux, as the American Psychiatric Association prepares to release a radically revised set of guidelines in early 2013. The draft DSM-5 proposes a number of major changes, including the introduction of a single diagnostic category - 'Autism Spectrum Disorder' (ASD) - to replace all the sub-types, including Asperger's Syndrome, which previously fell under the umbrella of Pervasive Developmental Disorders in DSM-IV. Also proposed is a shift from the three core domains of impairment described above to two (Social communication and interaction and Restricted/repetitive behaviours and interests). One effect of this is to take the focus off language impairment as a distinct core element in the diagnosis of an ASD, and move away from its use in making a binary distinction between autism and Asperger's syndrome. Clearly, however, language skills will continue to form an important part of the clinical picture and will in many cases affect how core diagnostic features of ASD are manifested (Lord & Jones, 2012).

1.4 Language processing in high-functioning ASD: implications for speech in noise

High-functioning individuals with ASD, as described previously, may be able to produce and understand complex sentences, and often also have a wide vocabulary. It is amongst these more able individuals that the deficits in pragmatic language ability, often considered to be a hallmark of ASD, are particularly evident. These include over-literal interpretation of language, lacking appreciation of a speaker's intention, failure to use contextual information in comprehension (e.g. to interpret ambiguous messages), poor judgement of what a listener needs to be told, and a tendency to one-sided monologues rather than conversational turn-taking (Bishop, 2003; Jolliffe & Baron-Cohen, 1999; Jolliffe & Baron-Cohen, 2000; Loukusa & Moilanen, 2009). Pragmatic difficulties of this nature have a considerable impact on an individual's ability to communicate successfully in everyday life, and could well be an important factor contributing to impaired ability to converse in background noise. Pragmatic impairments in comprehension are often assessed by presenting messages whose meaning is deliberately ambiguous, but even when the original linguistic content of a sentence suggests only one possible interpretation, background sounds may interfere with the transmission of that sentence and render it ambiguous to the listener. Many words have a number of close phonological neighbours, often with only one phoneme to differentiate them, and often with only one phonetic feature such as place of articulation to differentiate between phonemes. If acoustic cues to such features are masked by competing environmental sounds, then a listener will be forced to make a best guess at the intended word from a choice of two or more possibilities. In this situation, use of context (linguistic and social), world knowledge and judgement about the speaker's likely intention will be just as important for effective communication as they are in the case of interpreting deliberate jokes or sarcasm.

When a conversation takes place face-to-face, visual articulatory cues are also available to complement the phonological information in an auditory signal. Tone of voice, accompanying gestures and facial expressions may all also support a speaker's intended meaning. Whilst neurotypical adults are highly-skilled in the use of these cues, and may rely on them more heavily when an acoustic signal is degraded, the same cannot be said for those with ASD. Atypical eye-gaze, joint attention and non-verbal communication are hallmarks of ASD (Frith, 2003). Moreover, a considerable number of studies have shown that face-processing itself is atypical in ASD, with evidence of reduced attention to faces, poor face discrimination and recognition, and impaired processing of facial expressions (for a review, see Dawson, Webb, & McPartland, 2005). It appears that people with ASD may be employing different processing strategies when viewing faces, and that this may reflect a tendency towards reduced configural or global processing (Behrmann et al., 2006). They may actually be better than neurotypical controls at recognising isolated facial features and, in reversal of the normal pattern, may be better at recognising the lower half of a face (including the mouth) in isolation than the upper half (Langdell, 1978). Unfortunately, however, this does not appear to contribute to enhanced audio-visual speech perception. There is evidence that high-functioning children with ASD show less benefit than controls from audio-visual over purely auditory speech presentation, that individuals with ASD are less susceptible to the McGurk effect, and that lip-reading skills may be poor in ASD (Mongillo et al., 2008 ; Smith & Bennetto, 2007; Williams et al., 2004).

Taken in conjunction with pragmatic language impairment and poor non-verbal communication skills, an inability to take full advantage of visual speech cues could be expected to have a significant impact on speech comprehension in noisy backgrounds. In fact, one might think that these three factors alone could be sufficient to explain the subjective difficulties with speech in noise described at the start of this chapter. However, the results of Alcántara's study of *unimodal* auditory speech perception in noise suggest that this is not the whole story. In the absence of visual speech cues, and in the absence of normal conversational setting or gestural adjuncts to speech, the high-functioning ASD group still performed worse than controls (Alcantara et al., 2004). It seems likely, therefore, that atypical processing of the auditory speech signal itself may be at least as important as these other factors.

Structural language skills - particularly articulation and phonology - are usually considered to be an area of relative strength in high-functioning ASD, but speech sound distortions actually seem to be surprisingly common in this group (Cleland, Gibbon, Peppe, O'Hare, & Rutherford, 2010; Shriberg et al., 2001). Around 30-40% of people with high-functioning ASD may be affected to some degree (which is much higher than the expected norm), although each individual may make only a small number of speech errors. Shriberg suggested that the consistent changes in phoneme manner or place of articulation that many of his ASD group were prone to making (e.g. sibilant lateralisation) may reflect a failure of fine-tuning in the speech production system to the native language environment (Shriberg et al., 2001). He related this to the data from infant studies in which autistic children have failed to show normal preference for their mother's speech over cafeteria noise (Klin, 1991). Since Shriberg's study, Kuhl and colleagues have conducted an experiment in pre-school children which seems to support his hypothesis, at least in part, and suggests a similar pattern may also apply to phonological *input* processing (Kuhl, Coffey-Corina, Padden, &

Dawson, 2005). In this experiment, the autistic participants as a group showed a significant preference for listening to a non-speech analogue (sinewave speech tracking the formants of the original) over infant-directed speech, whereas the controls as a group showed no significant preference in either direction. This was broadly in line with Klin's observation (1991). However, event-related potentials (ERPs) to a 'wa'/'ba' phoneme contrast were also recorded by Kuhl, using the oddball paradigm, and whilst the autistic group as a whole showed reduced mismatch-negativity response (MMN) to this contrast relative to controls, there was also a significant difference between two subgroups of these ASD children. The minority (7/27 autistic children) who had preferred speech to non-speech in the previous task produced ERPs which were more similar to those of controls, with a measurable MMN to the syllable change, whereas the rest of the group (who had preferred the non-speech) produced highly abnormal waveforms. This does seem to imply a relationship between listening preference and phonological development in children with ASD.

Detailed behavioural studies of phonological processing specific to high-functioning ASD are surprisingly scarce, but two recent experiments have investigated the perception of non-native phonemes in this group (Constantino et al., 2007; Depape, Hall, Tillmann, & Trainor, 2012). The phoneme discrimination tasks used in these studies are based on the fact that whereas very young infants appear to be highly sensitive to all possible speech contrasts, regardless of language, exposure to the native language environment during the first year of life gradually leads to some loss of sensitivity to contrasts which are not meaningful in the native language. This makes it harder to discriminate between two perceptually similar sounds which fall within the same phonemic category than between two which fall on either side of a phoneme boundary. DePape's study involved two pairs of Zulu phonemes - one pair which should be more easily discriminated by English listeners because the phonemes map onto separate English phonemic categories, and a second pair which should be more difficult because they map onto the same English category. Using a 3-interval, 2-alternative (ABB or AAB) forced choice paradigm, listeners were asked to identify whether the first or last sound differed from the other two on each trial. Although there was no overall group effect, the interaction between group and phoneme condition (two-category or single-category) did reach significance. There appeared to be a very slight tendency for this group of children (11-18 year olds) with high-functioning ASD to categorise the two-category phonemes less accurately than controls, and the one-category phonemes more accurately. Two things must be emphasized about this finding, however. Firstly, the group differences within individual conditions were not significant, and secondly, there was no necessity for this task to be performed as a phonemic one at all. Listeners could just as easily perform well by focusing purely on the acoustic detail of what they heard, rather than thinking of the stimuli as speech. The authors' argument that their results indicate weaker specialisation for native phonemic categories in ASD is not, therefore, entirely convincing. Their comment that the data may reflect enhanced attention to the low-level detail of sounds makes considerably more sense, and is in line with current theories about non-speech auditory processing in high-functioning ASD (see below). Since an earlier study by Constantino had used only a single-category pair of Mandarin phonemes, and also found no group difference in performance (Constantino et al., 2007), the evidence for reduced perceptual warping to native language phoneme categories in ASD is very limited.

Electrophysiological studies have proved to be a rather more informative measure of speech sound processing in ASD, as the study by Kuhl et al. (2005) described earlier suggests. There is a considerable body of evidence that ASD in general is associated with atypical ERPs to speech sounds (for recent reviews see Groen, Zwiers, van der Gaag, & Buitelaar, 2008; Haesen, Boets, & Wagemans, 2011). Rather less work has investigated speech processing specifically in high-functioning ASD, but here a broadly similar pattern of abnormal responses has also been found. Magnetoencephalography (MEG), ERP, and functional magnetic resonance imaging (fMRI) studies have all indicated that the neurotypical tendency for left-hemisphere dominance for language is reversed in high-functioning ASD (Boddaert et al., 2003; Frye & Beauchamp, 2009; Jansson-Verkasalo et al., 2003; Lepisto et al., 2006; Lepisto, Nieminen-von Wendt, von Wendt, Naatanen, & Kujala, 2007). One suggestion, based on similar observations in SRD, is that this overuse of the right hemisphere may be compensating for dysfunctional left hemisphere processing centres (Frye & Beauchamp, 2009). Atypical laterality appears to be dependent on the precise stimuli involved, and may be specific to complex auditory stimuli only, as the pattern does not appear to be replicable with pure tones (Jansson-Verkasalo et al., 2003; 2005).

Lepisto and colleagues (2006) investigated auditory discrimination in children with high-functioning ASD, using an ERP / MMN paradigm and presenting repeating vowel sounds or repeating non-speech stimuli matched to the vowels (sinewave analogues with four components matching the F0 and first three formants of the vowels). MMN was measured for infrequent deviant sounds which varied from the repeating standards either in terms of pitch, duration or in terms of the vowel itself. In addition to atypical right-hemisphere dominance, N4 responses were diminished for the speech sounds in these autistic children, but not for the non-speech. There was also more parietal activity relative to controls in the MMN for a pitch change in the speech but not in the non-speech, and prolonged MMN latencies for a vowel change in both the speech and matched non-speech-vowel analogue. When the speech and matched non-speech contained a duration change, MMN amplitudes were reduced, and in a separate behavioural test of duration discrimination the ASD group were significantly less accurate and showed a tendency (only marginally significant) towards responding more slowly than controls. Finally, the ASD group showed diminished P3a responses to vowel changes, speech pitch changes, and speech duration changes, but

P3a for the matched non-speech sounds appeared normal. The P3a is associated with involuntary attentional orienting to auditory change, and previous studies have also indicated that P3a for speech stimuli is abnormal in lower functioning autism as well (Ceponiene et al., 2003), although apparently only in cases where no specific instruction has been given to attend to the stimuli (Whitehouse & Bishop, 2008). A similar pattern of enhanced MMN for pitch and atypical P3a orienting responses, particularly for speech stimuli, also appears to apply to adults with high-functioning ASD, although enhanced (rather than reduced) responses to duration change were found in one study (Lepisto et al., 2007).

Many of these findings are also similar to data gathered from a lower-functioning group with ASD in an earlier study carried out by the same Finnish research group (Lepisto et al., 2005). There are some differences, though, which imply more speech-specific and less widespread deficits in auditory processing at the higher-functioning end of the spectrum. P3a, for instance was atypical for non-speech as well as for speech stimuli in the lower-functioning study, additional exogenous components of the waveform were abnormal (not just N4), and right-hemisphere dominance was not observed in this group. The overall picture these ERP studies tend to suggest is that across the autistic spectrum, discrimination of auditory features which are key for speech perception - duration and formant structure (as in the vowel change) - is less efficient than in the neurotypical population. Furthermore, in high-functioning individuals, cortical responses to speech stimuli seem to be more abnormal than to matched non-speech stimuli. Parents of autistic children seem to share some of these abnormalities, although mothers and fathers show slightly different patterns and MMN for feature discrimination is atypical only in fathers (Jansson-Verkasalo et al.,2005). It seems possible that a combination of less efficient phonological processing and atypical attentional orienting might not lead to any obvious deficit in phonological processing under optimal conditions, but that these factors might well contribute to reported difficulties in background noise. Furthermore, electrophysiological responses to pitch change seem to be enhanced across the spectrum, in spite of the fact that at least two behavioural studies have suggested that frequency discrimination is superior only amongst lower-functioning individuals (Bonnel et al., 2010; Jones et al., 2009). This area of enhanced performance / hyper-reactivity to pitch may also have implications for speech in noise.

1.5 Pitch and enhanced local processing in ASD: implications for speech input

Exceptional pitch perception seems like something which should be advantageous to listeners - and indeed it may well be beneficial for some aspects of music processing (for reviews see Molnar-Szakacs & Heaton, 2012; Ouimet, Foster, Tryfon, & Hyde, 2012). It does seem possible, nevertheless - as Lepisto suggests - that enhanced pitch processing might also actually contribute to the troublesome symptoms of auditory hyper-sensitivity and abnormal reactions to sound which are a hallmark of ASD (Baranek, David, Poe, Stone, & Watson, 2006; Kern et al., 2006; Lepisto et al., 2006). When it comes to making use of the pitch information in speech, there is also little sign that heightened perceptual sensitivity is of much practical benefit.

Pitch information in speech directly reflects the fundamental frequency of a talker's vocal fold vibrations. As such it conveys important cues to talker identity, alongside the pitch movement (intonation) cues to utterance function (e.g. question versus statement) and to emotional content. Comprehension of grammatical prosody seems to be normal in ASD (Chevallier, Noveck, Happe, & Wilson, 2009), but it appears that the ability to recognise specific emotions conveyed by intonation is often impaired even in high-functioning ASD (Korpilahti et al.,2007; Rutherford, Baron-Cohen, & Wheelwright, 2002). As with the pragmatic language deficits discussed earlier, this impairment of affective prosody could pose more of a problem for speech comprehension when the phonological content is degraded by background noise and all types of contextual information become important for reconstructing the intended meaning.

The ability to recognise familiar voices and to discriminate between unfamiliar ones is a prerequisite for focusing attention on one speaker in a multi-talker environment, and this appears to be relatively unimpaired in ASD (Boucher, Lewis, & Collis, 2000). Unfortunately, however, once a target speaker has been identified it is also necessary to track their utterance as an auditory stream, distinct from competing voices, throughout its duration - and again vocal pitch is an important cue (Shinn-Cunningham, 2008). Little research has addressed auditory stream segregation in ASD so far, but one ERP study suggests that non-speech streaming based on pitch may be less efficient in high-functioning ASD (Lepisto et al., 2009). The authors of that paper do not propose a direct link between this finding and their earlier ERP data on enhanced MMN to pitch change (Lepisto et al., 2006), but it seems feasible that in the standard streaming task paradigm they used, this enhanced change detection might actively work against the formation of coherent auditory streams. The tone stimuli used in their task were presented sequentially, so one might imagine that any enhanced response to a local change in tone frequency could potentially conflict with the percept of similarity which drives the grouping of same-frequency tones across the sequence as a whole. In a more typical listening situation with two competing streams of speech, the same problem could apply. Glimpses of one talker's voice (and pitch) are likely to occur during lower intensity dips in the other voice, meaning a similar enhanced response to sudden pitch change - should it exist - might also undermine speech stream segregation.

Behavioural evidence does also suggest that the pitch information in speech may capture attention more amongst ASD listeners than the in the neurotypical population. One study using a quasi-open-format task, in which participants had the option to respond either to the linguistic content or intonation contour of sentence stimuli, found that children with ASD made significantly more intonation responses than did controls (35% of trials compared to 6%), although content responses did still predominate, as was the pattern for controls (Jarvinen-Pasley, Pasley, & Heaton, 2008). Perceptual responses from the ASD group were also significantly more accurate. Ploog and colleagues carried out a substantially more complex experiment and found a similar tendency in lower-functioning ASD, although in this study the ASD group did not show any overall attentional preference for content over prosody whilst controls continued to prefer content (Ploog, Banerjee, & Brooks, 2009).

This atypical pattern of attention to perceptual detail in speech processing can be linked directly to theories of Weak Central Coherence (WCC) and Enhanced Perceptual Functioning (EPF) in ASD (Frith, 2003; Mottron, Dawson, Soulieres, Hubert, & Burack, 2006). These theories both suggest that there is a cognitive style which is characteristic of autism whereby attention to local detail is enhanced, although they take rather different standpoints with regard to what happens at the global processing level. Methods used to investigate local and global processing have tended to use visual stimuli, often involving direct conflict between information at the two levels (e.g. Navon figures - see Plaisted, Swettenham, & Rees, 1999), and ASD groups seem to be less affected than controls by conflicting global level information when attending to local level detail. WCC suggests this is due to an impairment of the global level, whereas EPF focuses purely on heightened sensitivity to lower-level perceptual features. Fewer studies have investigated the interaction between auditory local and global processing - not least because 'local' and 'global' levels seem harder to define and manipulate in this modality - but it certainly appears that there may be a similar absence of global interference on local auditory processing for musical stimuli in high-functioning ASD (Foxton et al., 2003). Moreover, the signs of reduced auditory stream segregation, already discussed, would seem to reflect a deficit in global level auditory processing.

Aside from pitch processing, however, there is little behavioural evidence of enhanced auditory perception at the local level, particularly if we take 'local' auditory processing to mean the discrimination or recognition of acoustic features such as intensity, spectral shape, and temporal detail. There is a tendency in the literature to equate hyper-reactivity and hypersensitivity to sound in ASD (usually assessed by measures such as sensory questionnaires, which rely on parent report) with enhanced auditory processing. However, it is arguable whether this can be taken as valid evidence since it relies purely on inference regarding the causes of aversive reaction to sound. When direct psychacoustic measures of auditory discrimination ability were employed, one key study found no sign of atypical performance for intensity, vocal timbre or non-vocal timbre discrimination either in children with autism or Asperger's syndrome, even though the autistic group did show enhanced pitch discrimination (Bonnel et al., 2010). Likewise, in a much larger study involving 72 adolescents with ASD, there were no signs of atypical intensity or duration discrimination, whilst frequency discrimination was again enhanced in around one fifth of the group (Jones et al.,2009). In fact, some of the behavioural data is actually suggestive of local level auditory *deficits* - in children, at least - in areas such as frequency resolution (Plaisted, Saksida, Alcantara, & Weisblatt, 2003) and temporal processing (Alcantara, Cope, Cope, & Weisblatt, 2012 ; Alcantara et al., 2004; White et al., 2006).

The picture is further confused by evidence from electrophysiological studies. As already discussed, Lepisto et al. (2006) observed reduced MMN in children for duration changes (of

86 ms) in the non-speech as well as the vowel stimuli, which might relate to a developmental delay in temporal aspects of auditory processing. By contrast, however, there was some evidence of enhanced MMN for duration changes in the analogous *adult* study, specifically at frontal electrodes, although it is interesting that this enhancement did not extend to the matched non-speech stimuli (Lepisto et al., 2007). A second adult study using only non-speech stimuli (harmonic complexes) did find enhanced MMN for duration change (65 ms) and for deviant stimuli containing a 10 ms gap, further supporting the suggestion that the development of efficient temporal processing may only be delayed in high-functioning children with ASD. This study also investigated discrimination of intensity and location (based on an interaural time delay cue) and found no significant abnormality at all (neither deficit nor enhancement). All the behavioural data for atypical temporal processing reviewed here comes from child or adolescent studies, so it is possible that enhanced temporal processing in high-functioning adults could become apparent in future research. It is less than easy, however, to predict what effects - if any - enhanced temporal processing might have on speech perception, particularly in background noise.

1.6 Summary

Impaired speech perception in noise affects even high-functioning adults with ASD, but as yet we know little about the reasons why. High-level pragmatic language and communication deficits are core features of ASD, but in laboratory tests where the contribution of these higher-level skills to speech comprehension in noise is minimised, performance remains poor. Signs of atypical language processing can be found even at relatively early stages of phonological input processing, and seem to affect even individuals who have no history of language delay and who perform within the normal range on formal language assessment. Likewise, a number of aspects of non-speech auditory processing appear abnormal.

Whilst work in the fields of SLI and APD suggests that poor speech perception in noise can rarely be attributed directly to deficits in non-speech auditory perceptual skills, the unusual profile of listening skills in ASD presents a rather different picture. Although behavioural evidence suggests that pitch processing may only be enhanced in lower-functioning ASD, electrophysiological studies have shown increased pitch discrimination responses in high-functioning listeners as well. In addition, it appears that although the maturation of auditory temporal processing may be delayed in ASD, by adulthood the ERPs of high-functioning individuals may also indicate enhanced discrimination of temporal detail. As discussed, it is plausible to think that enhanced responses to auditory feature change - and to pitch change in particular - might interfere with the formation of coherent speech streams when more than one person is speaking at the same time. There is little published research on auditory stream segregation in ASD, but ERP data does suggests this ability may be impaired.

Finally, a convincing body of evidence suggests that attentional orienting to speech is abnormal in both children and adults with ASD. This is perhaps the most significant finding of all. Neurotypical adults can attend to one talker in the presence of another relatively easily, but if even automatic, exogenous attentional orienting to speech in quiet is atypical in ASD one might imagine this could lead to problems under less advantageous listening conditions. Recent work in APD has certainly highlighted the importance of attentional processing as a core factor in listening task performance. Coupled with the possibility that auditory stream segregation abilities may also be weak, it seems highly likely that atypical auditory attention in ASD could cause increased susceptibility to speech-on-speech masking. The first experiment described here explores this hypothesis.

Chapter 2 Speech-on-speech masking in ASD

2.1 Introduction

When investigating why reports of poor speech comprehension in noise are so common in high-functioning ASD, there are numerous factors to consider. Only a handful of published behavioural studies, though, have yet explored the problem directly using controlled tests of speech in noise (Alcantara et al., 2004; Groen et al., 2009). Moreover, these studies have concentrated on only a narrow range of background noise conditions, and - possibly as a result of that - have found relatively small impairments in speech in noise which do not seem to reflect the severity of the reported difficulties. The addition of masking sounds to a speech perception task often involves more than simply making a target stimulus less audible, and auditory and cognitive abilities which are not essential for listening in quiet conditions may have a significant impact on performance in noise (Schneider, Li, & Daneman, 2007). Different types of masking sound challenge different combinations of listening skills, and there is reason to believe that by focusing on mainly non-speech maskers, previous studies may have uncovered only one component of the impairment in autism.

The first experiment in this area was conducted by Alcántara and colleagues (2004), who investigated speech reception thresholds (SRTs; the signal-to-noise ratio at which speech perception is 50% correct) in a small group of high-functioning young participants with autism. Results indicated that, whereas SRTs in steady-state noise and in noise with regularly-spaced spectral dips were within the normal range, the ASD group's performance was significantly poorer than controls' (by around 2 to 3.5 dB) in conditions where the background contained amplitude fluctuations over time. These deficits were very similar for both a non-speech modulated noise condition and masking speech from an opposite-gender talker. Alcántara's data has some interesting implications. Firstly, the fact that these ASD listeners performed normally in steady-state noise, and in noise with spectral dips, implies that they were well able to meet the cognitive and linguistic demands of the task. As described in Chapter 1, difficulties with speech perception in noise are not peculiar to ASD and have also been studied in specific language impairment (SLI) and specific reading disorder (SRD). In those disorders, however, SRTs in all types of noise - including steady-state noise - tend to be impaired (Ziegler, Pech-Georgel, George, & Lorenzi, 2009; Ziegler, Pech-Georgel, George, & Lorenzi, 2011). There are several potential reasons for this, but one hypothesis is that phonological representations in SLI/SRD may be less robustly encoded. Typical adult listeners benefit from redundancy in the speech signal - more than one acoustic cue coding for each phoneme - so that even when some cues are obscured by noise, sufficient information is transmitted for correct identification. Less-detailed phonological representations in SLI/SRD, however, could heighten the impact of any lost acoustic cues. In the light of this, although there is some evidence of atypical phonological processing in high-functioning ASD (see Chapter 1), the fact that speech perception in steady-state noise appears normal suggests that stored phonological representations, at least, are not impaired to the same degree as in SLI or SRD.

Since a diagnosis of Asperger's syndrome by definition implies no developmental language delay, differences in the pattern of performance between ASD and SLI are perhaps unsurprising, but they do suggest that factors beyond the purely linguistic may affect SRTs in ASD. Normal performance in two of the four background noise conditions presented by Alcántara, however, also implies that sustained top-down attention to the task is not the core problem either. Furthermore, unimpaired speech perception in steady-state noise is important on yet another level, because it suggests normal functioning of the auditory periphery. At least one study has suggested that auditory filters may be abnormally broad in autism (Plaisted et al., 2003), but if that had been the case with ASD listeners in Alcántara's 2004 study, it is highly unlikely that their SRTs in steady noise would have been unaffected.

The ASD group's ability to perform normally in some listening conditions of Alcántara's study is a key finding, as it seems to point away from at least some of the possible factors underlying the reported difficulties. A further implication - based on the observation that ASD listeners were only impaired in background noises containing amplitude modulations - is that autism is associated with a deficit in what is often referred to as 'dip listening'. It is well-established that normally-hearing neurotypical adult listeners' SRTs tend to be better in noises that are amplitude modulated than in steady-state noise (Miller, 1947; Miller & Licklider, 1950; Howard-Jones & Rosen, 1993). This release from masking occurs because at moments when the intensity of the background dips down to a low level, the target speech can be glimpsed at a more favourable signal-to-noise ratio. It is apparent, therefore, that at least two separable skills or abilities are involved in dip listening: the ability to process temporal detail in the auditory scene accurately, and the ability to reconstruct a spoken message from brief glimpses. Unfortunately, without conducting a separate non-speech test of auditory temporal processing in the same group of listeners, it is impossible to be sure whether the impairment Alcántara's ASD listeners showed for speech in fluctuating backgrounds was due to a more general deficit in encoding auditory temporal information, or to difficulty in reconstructing the fragmented speech signal. Alcántara's target speech material consisted of simple everyday sentences (Adaptive Sentence Lists; MacLeod & Summerfield, 1990), and there is evidence that the processing of linguistic context may be atypical in ASD (Happe, 1997; Jolliffe & Baron-Cohen, 1999; Lopez & Leekam, 2003; Henderson, Clarke, & Snowling, 2011). It certainly seems possible that any reduced tendency to process the sentences for meaning in ASD might contribute to weaker dip listening performance, but direct evidence of this is lacking. The existence of a study which hinted at impaired perception of single words in fluctuating noise (Groen et al., 2009; see below) implies that sentence-level deficits may not be the sole factor. The evidence supporting a more purely perceptual deficit in temporal processing, however, is similarly equivocal.

Alcántara and colleagues themselves, in a recent study, found no evidence of a deficit in auditory temporal resolution in ASD, although their data did indicate reduced temporal processing *efficiency* (Alcantara et al., 2012). Relating these results to the data on speech perception in amplitude-modulated noise is not straightforward, however. A clear deficit in auditory temporal resolution in ASD would suggest that the auditory system is in some way sluggish, encoding rapid amplitude changes in sounds less accurately, and presumably therefore reducing or blurring the dips in the background noise. This does not appear to be what is happening in ASD, though, as the only impairment Alcántara found was in temporal processing efficiency. The term 'efficiency' in this context refers to the depth of the amplitude fluctuations that have to be added to a noise before listeners can discriminate it from a steady noise without regular fluctuations. The ASD group needed deeper fluctuations than controls, regardless of the speed at which the fluctuations occurred. The trouble here is that merely detecting rapidly-occurring amplitude modulation in a forced-choice psychoacoustic task is very different from listening to speech during the dips in a background noise, and the notion of efficiency is hard to map accurately from one situation to the other. Unpublished results from a study which investigated nonsense-syllable perception, in noise which was modulated at varying rates, suggests that any effect of reduced temporal processing efficiency in ASD may be marginal at best, given that small group differences in release from masking with modulations (at all modulation rates) did not reach significance (Fullgrabe, Alcantara, & Weisblatt, 2008). By contrast, Groen and colleagues (2009) did find significantly reduced release from masking in children with ASD for real words in fluctuating versus steady noise, but here the group mean difference in release from masking was only around 2 dB - an even smaller deficit than Alcántara (2004) found for sentences in noise. The lack of any significant group difference in either masking condition itself, however means these results are not particularly convincing.

Fullgrabe et al.'s (2008) study of nonsense-syllables in noise - with only 6 participants in each group - may have lacked power, but it is also important in being the only experiment of the three in which the effects of word- and sentence-level linguistic ability were minimised. The lack of a significant result, therefore, in what was arguably the purest measure of the effects of temporal processing in ASD on speech perception in noise - especially when taken alongside the fairly small deficits in dip listening found by Alcántara (2004) and Groen (2009) - seems to indicate that reduced temporal processing efficiency is only a part of the explanation for the difficulties this group experiences.

Concurrent speech and background noise of any type mingle to form a complex auditory scene, in which a listener must segregate overlapping sounds into separate streams of information and focus attention on the target signal (Shinn-Cunningham, 2008). The precise details of the target and background stimuli involved, and in particular their relative similarity - linguistic and affective as much as acoustic - will determine the degree to which these processes of segregation and attention are challenged. The types of non-speech masking noise used in the studies described above are typically quite different in perceptual quality from the target speech, and are therefore relatively easy to segregate out. This means that speech reception thresholds in these non-speech masking conditions tend to be dominated by energetic masking (i.e. they reflect the signal-to-noise ratio at the output of the auditory periphery in a fairly predictable way, Rhebergen & Versfeld, 2005). In everyday life, however, the auditory scene often contains a number of talkers, and the background speech may be considerably more difficult to segregate and screen out from the target signal. This can lead to 'informational masking', in which it is not just the acoustic overlap between two messages which affects target intelligibility, but interference from the background speech at a more central processing level - where it may be confused with the target or divert attention away from it. Whilst deficits in bottom-up auditory processing and/or top-down linguistic processing may both contribute to reduced speech perception in noise in ASD, therefore, it is also possible that auditory streaming and/or selective attention processes may be involved. In fact, there are good theoretical reasons to suggest that increased susceptibility to informational masking in ASD is quite probable.

As discussed in Chapter 1, cognitive theories concerned with perceptual processing style in ASD have highlighted a characteristic which seems highly relevant to speech perception in noise. Investigations into 'weak central coherence' (Frith & Happe, 1994) suggest that individuals with ASD may have a superior ability, or bias, to focus on perceptual detail without interference from global or gestalt-level information, in auditory as well as visual tasks (Foxton et al., 2003; Shah & Frith, 1983). The mechanism underlying these differences in performance is still not fully understood, but it no longer seems likely that there is a straight-forward deficit at the global processing level (Plaisted et al., 1999; Mottron, Peretz, & Menard, 2000). More recently, therefore, in an attempt to shift the focus away from the idea of global-level impairment in ASD, superior detection and discrimination of simple, low-level stimuli in some tasks have been interpreted as evidence for enhanced perceptual functioning in ASD (EPF; Mottron et al., 2006). So far, however, it appears that only a subset of people with ASD show such enhancements on auditory tasks. Pitch perception has received far more attention than any other aspect of psychoacoustics in ASD, but even in this case not all individuals show enhanced performance, and it is notable that those who do tend to have a diagnosis of autism rather than Asperger's syndrome (Bonnel at al., 2010, Heaton, Williams, Cummins, & Happe, 2008).

We may not fully understand why people with ASD are more able to resist gestalt-level interference on visual tasks like the Embedded Figures Test or Navon hierarchical figures, but these findings are robust and point to a possible difference in the way incoming perceptual information is structured and/or prioritised, as well as raising the question of whether low-level featural processing may actually be superior. Such differences in auditory information processing style might well affect how concurrent speech signals are organised into separate streams and how attention is allocated. Moreover there is evidence that ASD may also be associated with increased susceptibility to interference from irrelevant distractors in visual tasks (Burack, 1994; Remington, Swettenham, Campbell, & Coleman, 2009). Under Lavie's load theory of selective attention (2005), when perceptual load increases beyond a certain point (e.g. when the number of items in a visual array during a search task exceeds the limits of perceptual capacity) task-irrelevant distractor items cease to pass through the attentional filter and are excluded from perceptual processing at an early stage. A study by Remington and colleagues, based on this theory, has shown that adults with ASD continue to show distractor interference effects at higher set sizes than neurotypical control participants, suggesting that they may have increased perceptual capacity (Remington et al., 2009). This finding has yet to be replicated by other groups, but an association between perceptual capacity and autistic traits in the neurotypical population (Bayliss & Kritikos, 2011) is clearly consistent with Remington's results. Whilst there is a limit to how far the results of a simple visual search task might be applicable to speech perception in noise, not least because of the far higher cognitive demands of the latter, if increased perceptual capacity were to apply in the auditory modality it might also be expected to lead to increased informational masking.

Probably the most compelling reason to suppose that speech-on-speech masking might be atypical in autism, however, is the special status of speech itself as a stimulus. As reviewed in Chapter 1, a wealth of evidence points to the fact that people with ASD respond abnormally to social stimuli - including speech - from earliest childhood (see for example Klin, 1991; Kuhl et al., 2005; Whitehouse & Bishop, 2008). Even into adulthood, patterns of cortical activation to vocal stimuli appear to be atypical, although responses to environmental sounds are not significantly different to those of controls (Gervais et al., 2004). Recall of the vocal stimuli, but not the environmental sounds, at the end of Gervais et al.'s fMRI study was also impaired relative to that of controls.

Deficits in social interaction are core to ASD, and this seems to encompass a lack of expertise in processing social stimuli. In a multi-talker environment this lack of expertise, or reduced automaticity, in speech and voice processing could make it more difficult to focus attention on a single target speech stream. Even an apparent strength, such as enhanced perception of pitch information in speech, could actually draw attention away from the target message itself. There is also evidence from at least one study that children with Asperger's syndrome may not be as proficient as their neurotypical peers at perceptually separating simultaneous non-speech sound sources (Lepisto et al., 2009). There is good reason, then, to predict that speech-on-speech masking may be particularly impaired in ASD. Surprisingly, though, whilst Alcántara's (2004) study did include a speech masker condition, the ASD group's deficit was relatively small and barely differed from their impairment in modulated noise. Perhaps the key point about Alcántara's speech condition, however, is that it involved a female background talker masking a male target, and this two-talker opposite-gender listening situation is usually very easy for neurotypical adults (Brungart, 2001). On average the fundamental frequencies of male and female voices differ by around an octave, and this is a strong cue to stream segregation (Darwin, Brungart, & Simpson, 2003). Speech tends to become a much more effective masker, however, as the number of talkers in a scene increases, or as the target and masker talkers become more similar (Brungart, Simpson, Ericson, & Scott, 2001). Bearing in mind that listening to speech 'in noise' over headphones in a laboratory eliminates many of the additional visual distractions and incidental sounds we experience in more realistic listening environments, it may well be that this task was simply not sufficiently challenging to reveal an underlying difficulty with speech stream segregation or selective attention in Alcántara's ASD listeners.

Our aim in the current study was to explore speech-on-speech masking in ASD in greater depth. For this purpose we based our method on the Coordinate Response Measure (Bolia, Nelson, Ericson, & Simpson, 2000) - a listening task which was specifically designed to explore informational masking effects, which uses highly familiar vocabulary, and which minimises the importance of sentence context. We used this task to test SRTs in a range of different background noise conditions, including masking speech with varying degrees of similarity to the target material, predicting that our ASD participants would show greater impairment with speech than non-speech maskers, and increasing impairment with increasing target-masker similarity. We hoped that by using the mCRM as the main task, the effect of any between-group differences in pragmatic or structural language ability would be kept to a minimum. In addition, however, to investigate how performance in the mCRM might relate to performance with more semantically and syntactically challenging speech material, SRTs in a core set of masking conditions were also measured for target sentences somewhat more like those encountered in everyday life.

2.2 Method

2.2.1 Participants

The study involved two groups of participants: 13 high-functioning young adults with ASD (10 male, 3 female) and 14 neurotypical controls (10 male, 4 female). All ASD participants had received an independent clinical diagnosis of Asperger's Syndrome, and all completed the Autism Diagnostic Observation Schedule - Generic (Lord et al., 2000) during their visit to confirm that they met the cut-off score of 7 for a classification of autism spectrum disorder on that assessment (group mean ADOS score (total) = 10.6, std. dev. 3.2). As an additional screening measure, all participants also completed the Autism Spectrum Quotient questionnaire (ASQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). There was a significant difference between the scores of the control group (mean ASQ 13.7, std dev. 7.6), and those of the ASD group (mean ASQ 35.2, std dev. 8.3), ($t (25) = -7.04$, $p < 0.001$), indicating that ASD participants were considerably higher in traits associated with the autistic spectrum than were controls. Groups were matched for age, and for verbal and non-verbal intelligence using the short form of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), (see Table 2.1). Participants reported no history of language impairment, other than the pragmatic language difficulties typically associated with ASD. Prior to the experimental tasks, all listeners passed an audiometric screening test at ≤ 20 db HL, indicating normal sensitivity to pure tones at octave frequencies from 0. 25 – 8 kHz.

Table 2.1 Participant age and WASI IQ data

	Age	**Verbal t-score (vocabulary)**	**Performance t-score (matrices)**	**FSIQ***
Control Mean (Std. dev.)	24.3 (5.7)	65.5 (5.4)	60.3 (5.2)	124.1 (8.7)
ASD Mean (Std. dev.)	25.5 (5.6)	64.4 (11.9)	60.8 (8.2)	123.5 (13.8)

* FSIQ = Abbreviated full-scale IQ score (based on 2 subtests only)

There is evidence that musical training may be associated with better performance on speech-in-noise tasks. None of our participants reported more than one or two years of entry-level music lessons while they were at school.

Participants with ASD were recruited by advertising on the National Autistic Society website and through online communities. At the time of the experiment, 7 were in full- or part-time education, 4 had jobs, and 2 were unemployed. Control participants were recruited through advertisements in and around UCL: 11 were students, 2 were in work, and 1 was unemployed. All experiments in this study were approved by the UCL ethics committee and were conducted according to their guidelines. Participants were given written information sheets and an oral explanation. They were informed that they could withdraw from the study at any time, and were only included once they had given written consent. All participants received reimbursement for their time (£7 per hour), and ASD participants travelling from outside London also received travel expenses.

All participants in the clinical group had originally been diagnosed with Asperger's Syndrome and none reported any history of language delay. Given the proposed move to remove Asperger's syndrome as a diagnostic category from DSM-5, however, and the lack of evidence for a clear distinction between Asperger's syndrome and high-functioning autism, the term 'ASD' will be used to refer to this group here (following the example of Chevallier et al., 2009).

2.2.2 Listening tests

Participants completed two main tests of speech perception in noise.

i) A modified version of the Coordinate Response Measure

The modified Coordinate Response Measure was developed to be a more child-friendly version of the original Coordinate Response Measure (Bolia et al., 2000), but is equally appropriate for use with adults. Sentence stimuli consist of a carrier phrase taking the form 'Show the dog where the [colour] [number] is'. The target colour and number words are varied randomly across trials (possible colours: black, blue, green, pink, red or white; possible numbers: 1 – 9, excluding the bisyllabic seven). On each trial, listeners are instructed to select the colour-number combination they heard, from an on-screen array of 6 coloured grids, each containing the 8 possible digits. The linguistic content of the task is very simple, therefore, but since a response only counts as correct when both target colour and number are reported accurately, there is only around a 2% chance of guessing successfully on each trial.

The computerised test uses a 1-up 1-down adaptive tracking procedure to estimate a speech-reception threshold (SRT) for each condition - the signal-to-noise ratio (SNR) at which an individual participant's responses are 50% correct. In each block of trials, the first sentence was presented at +20 dB SNR and the initial step size was 10 dB. Step size then decreased linearly over the first two reversals to either 2 dB (condition 1), 3 dB (conditions 2, 7, 8), or 4 dB (conditions 4, 5, 6). These final step sizes were based on pilot data, and take into account the fact that the slope of the average psychometric function tends to be slightly steeper for steady noise than for other background sounds. The maximum number of trials in each block was restricted to 30, but the software automatically terminated the block when 6 reversals at the final step size had been obtained. In practice, this tended to occur within the first 15-20 trials. The SRT was calculated from the mean of the SNRs at the 6 final reversal points.

A key feature of this task is that competing mCRM sentences (or 'distractors') can be presented as masking material simultaneously with the targets. Listeners hear two pairs of colour-number stimuli (e.g. 'Show the dog where the green three is' paired with 'Show the pig where the blue eight is'), and must report the correct pair, with the only linguistic cue coming from the animal name in the carrier sentence. In this case, the demands on auditory stream segregation and selective attention are far higher than in the standard case of speech-on-speech masking, where the structure and content of the masker overlaps less with the target material. Target sentences in this study always began with 'Show the dog...', and were always spoken by the same female talker with a southern English accent. Masking sentences were spoken by talkers with varying degrees of similarity to the target, as described below, and could include any one of 5 different animal names selected randomly on each trial (pig, cat, cow, duck, or sheep). Maskers never contained the same colour or number as the target on any given trial.

There were eight main test conditions, in each of which the mCRM targets were presented against a background of one of the following maskers:

1) steady speech-spectrum noise
2) amplitude-modulated speech-spectrum noise
3) Japanese male talker
4) task-irrelevant English male talker (sentences from another corpus)
5) task-relevant noise-vocoded English speech (mCRM distractor sentences, based on condition 6)
6) task-relevant English male talker (mCRM distractor sentences)
7) task-relevant English female talker (mCRM distractor sentences)
8) mCRM distractor sentences spoken by the same female as the targets

These maskers were selected with the main aim of varying the potential for informational masking across conditions, whilst broadly controlling for listeners' ability to take advantage of dip listening. The steady noise in condition 1 represented the 'purest' form of energetic masking and was therefore a baseline measure. This steady noise also acted as the carrier signal in condition 2, but here it was amplitude-modulated with the envelope fluctuations of randomly-selected mCRM distractor sentences. mCRM envelopes were extracted by full-wave rectification, and smoothing with a low-pass filter (2nd order Butterworth filter, 30 Hz cut-off). Differences in performance between condition 2 and condition 1, therefore, represent listeners' ability to take advantage of dip-listening.

The remaining maskers all consisted of speech, selected to represent increasing levels of informational masking. The Japanese male speech background (condition 3) was created by concatenating sentences from the HINT corpus (Shiroma, Iwaki, Kubo, & Soli, 2008), and presenting a randomly-selected excerpt from this sequence on each trial. The task-irrelevant English male speech (condition 4) was created in a similar way by concatenating stimuli from the BKB corpus (Bench, Kowal, & Bamford, 1979): simple sentences using high-frequency vocabulary (e.g. 'The clown had a funny face'). Japanese morae have a less elaborate phonetic structure than English syllables, which led to a perceptually salient difference in the natural speech rate of the Japanese and English maskers. To reduce this effect, both maskers were manipulated in PRAAT (Boersma, 2001) to have approximately the same number of phonemes per second: the English speech rate was increased by 4.5%, and the Japanese rate reduced by 4.5%. Whilst these two maskers were similar, therefore, in spectro-temporal properties, and also in being irrelevant to the target sentences, they differed from each other in terms of their intelligibility to our English listeners (none of whom reported any experience of learning Japanese).

The background speech in conditions 5-8 consisted of mCRM sentence distractors, as already described, with increasing levels of perceptual similarity to the targets. All included distractor colour and number words which could potentially be confused with the targets, but for normal adult listeners target-masker confusion errors are known to be less common with opposite gender target/masker combinations (as in condition 6) than with same gender pairings (as in condition 7) (Brungart, 2001). We included condition 5, in which the mCRM stimuli from condition 6 were noise vocoded, in order to explore the effect of distractor sentences which were even more perceptually distinct from the female target than natural male speech. 16 channels were used in the vocoding process, so that the speech was fully intelligible at normal listening levels (Friesen, Shannon, Baskent, & Wang, 2001), but it lacked most of the original pitch information and had a hoarse whispery quality. Signal processing was based on the method of Shannon (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Notably, a 30 Hz low-pass smoothing filter was used in the envelope extraction so as to minimise any voice-pitch-related fluctuations. At the other extreme from the vocoded background, in condition 8 we minimised the perceptual difference between target and masker speech by using the same female talker for both. This renders the task extremely challenging even for the most competent listeners.

ii) Standard sentence in noise task (SIN)

The second listening task used sentences from the ABC corpus as target material (an updated version of the IEEE sentence set; IEEE, 1969), and as such was more similar to the speech in noise task employed by Alcántara et al. (2004). Each sentence contains 5 key words (e.g. 'A large size in shoes is hard to sell'), and participants score a point for each word they score correctly. Once again, the SNR is varied adaptively to establish an SRT.

Target sentences were spoken by a female Southern English talker, and were presented in three masking noise conditions:

1) steady speech-spectrum noise (as in Task 1, condition 1)

2) amplitude-modulated speech-spectrum noise (using the envelope fluctuations of the speech in condition 3)

3) male English talker (as in Task 1, condition 4)

2.2.3 Procedure

ASD participants proved relatively difficult to recruit, and in many cases travelled from outside London to take part. In order to maximise productivity, therefore, several different lines of enquiry were conducted in parallel, so that all the data described in chapters 2 and 3 was collected during one main block of testing sessions (Experiment 1), and all the data described in subsequent chapters was collected during a second main block (Experiment 2). This was less than ideal, but testing sessions were planned carefully to ensure several long breaks, and shorter ones where necessary. Participants were encouraged to say if they were feeling tired and were observed for any signs of fatigue. For both main experiments, time spent actually completing the listening tasks varied slightly for each person, but was usually around 2 hours in total. The abbreviated WASI took an additional 30 minutes to administer, after the listening tests were complete. In addition to this, most listeners took one hour-long break and at least one 30 minute break, and many reported that they enjoyed completing the listening tasks. In most cases, the ASD participants completed the ADOS on a different day.

All listening tests described in Chapters 2 and 3, therefore, were completed on a single day, spread over several sessions and interspersed with breaks. The experiment was conducted in a sound-protected booth on a Samsung Q45 laptop running Matlab (MathWorks), and the stimuli were presented diotically via Sennheiser HD-25 II supra-aural headphones, at a sampling rate of 22050 Hz. The order in which the two tasks (mCRM and SIN) were presented was balanced across participants, and within-task the condition order was randomised (but matched across groups). Following a short practice session at the start of each test, two blocks of trials were completed consecutively for each condition, and in cases where a listener's two SRTs for a given mCRM condition differed by 5 dB or more, a third block was completed. The final SRT in each condition used in the analysis is a median of the 2-3 values recorded. During the mCRM , participants saw an on-screen message at the start of each block, reminding them to listen to the female target saying 'Show the dog', and ignore the noise or man/woman talking in the background. They also received on-screen feedback (correct/incorrect) during the mCRM, but not during the SIN.

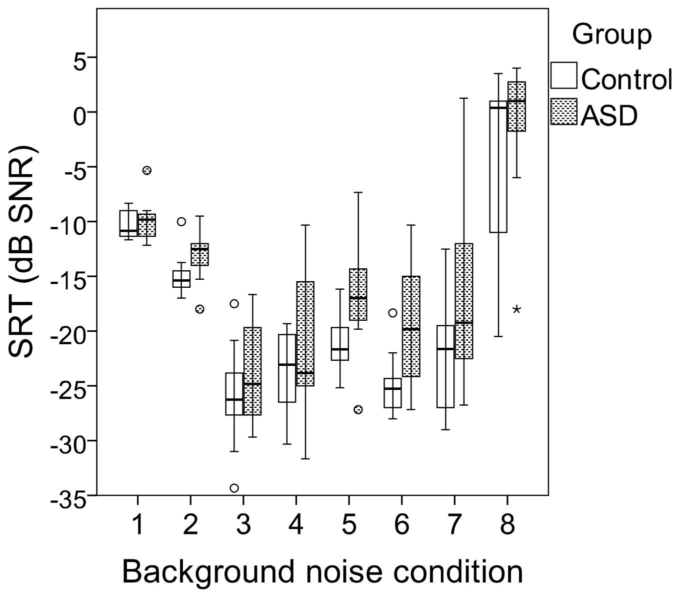
2.3 Results

2.3.1 Modified Coordinate Response Measure

The performance of the control listeners across the 8 mCRM conditions generally followed the pattern established by previous studies. In listening conditions where energetic masking was maximal (with steady noise and AM noise), SRTs were higher (i.e. worse) than in most of the speech-masking conditions (see Figure 2.1). An improvement in the mean SRT of around 5 dB between steady and AM noise (conditions 1 & 2) shows that control participants were able to make use of glimpses of the target material when the masking noise contained low-energy dips. A small (2 dB) increase in SRT going from male CRM distractors to female (conditions 6 and 7) indicates that the same-gender target and masker speech combination was more challenging than the opposite-gender combination. When the same female talker was used for the target sentences and the background distractor material (condition 8), SRTs were even worse for some controls than in the steady non-speech noise, and the wide variability in performance across listeners is typical of situations involving high levels of informational masking.

A repeated measures ANOVA (2 groups x 8 conditions) with SRT as the dependent variable, indicated significant main effects of condition (F (7, 175) = 94.5, $p < 0.001$, partial η^2 0.79) and group (F (1. 25) = 8.7, $p = 0.007$, partial η^2 0.26), but no group x condition interaction. Taken as a single group, the ASD listeners clearly tended to have somewhat higher (i.e. worse) SRTs than controls, irrespective of condition (see Figure 2.1). However, in many conditions variances were not homogeneous across group and, particularly when English speech was used as a masker (conditions 4-8), some ASD individuals performed extremely poorly, whilst others appeared to perform well within the normal range.

Figure 2.1 mCRM speech reception thresholds

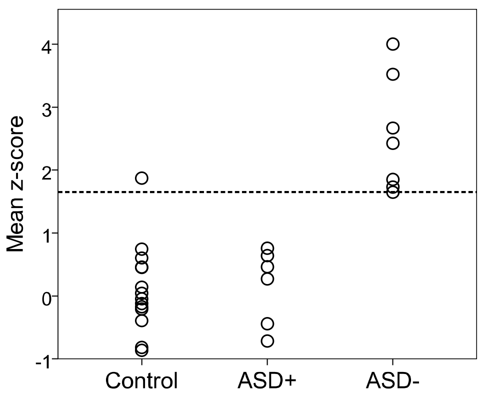


Background noise conditions:

1	Steady noise
2	AM noise
3	Japanese speech
4	English speech
5	Vocoded CRM
6	Male CRM
7	Female CRM
8	Target female CRM

To investigate this variability within the ASD group further, we took a multiple-case studies approach and calculated z-scores for each listener in every mCRM and SIN task condition, based on the mean SRT and standard deviation of the control listeners (in a two-stage procedure, excluding control outliers, $z \geq 1.65$ or $z \leq -1.65$, from the final mean) (see Ramus et al., 2003). Mean z-scores for each listener across all 11 conditions were then computed, and these provided further evidence of a bimodal distribution within the ASD group (see Figure 2.2). Six of the group (referred to as 'ASD+') tended to perform well within the normal range for most types of masker (range of ASD+ mean z-scores -0.72 to + 0.76; mean = 0.16, s.d. = 0.60), whereas 7 others ('ASD-') had indeed performed significantly poorly (z-scores ranging from +1.65 to +4.00; mean = 2.55, s.d. = 0.92). By comparison, only one control listener had a mean z-score greater than +1.65 (mean z = 1.87). ASD listeners who performed poorly in the mCRM also performed poorly in the SIN task (see Figure 2.3), and their performance within both tasks was poor across conditions (see Figure 2.4 and Figure 2.6). A univariate ANOVA and post hoc tests with listeners' mean SRT (across both tasks / 11 conditions) as the dependent variable confirmed a significant difference between ASD- and both other groups (all $p < 0.001$). Given this clear difference in performance between these two subsets of ASD listeners, they were treated as two separate groups for the remainder of the analyses. Although $z \geq 1.65$ is typically used as a cutoff point for clinically significant impairment, for these analyses we chose $z \geq 1.64$ because the best-performing ASD- listener's mean z-score was in fact just marginally under +1.65 (1.647).

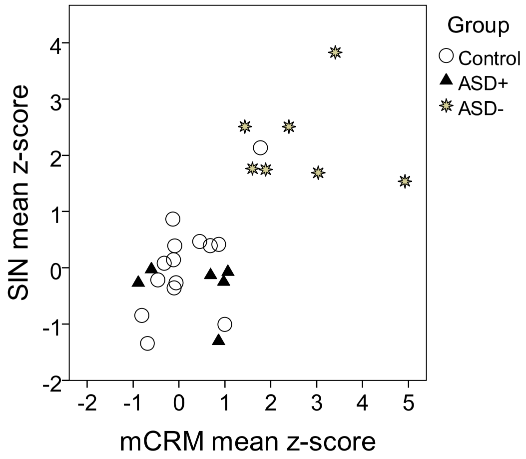
Figure 2.2 Individual mean z-scores (across all 11 conditions of mCRM and SIN)



Higher z-scores (i.e. higher SRTs) indicate worse performance

- - - - - indicates z = +1.65

Figure 2.3 Scatterplot: relationship between mean z-scores in mCRM and SIN tasks



The mCRM repeated measures ANOVA was rerun, this time with 3 groups and 8 conditions. This revealed a significant group effect (F1, 24 = 17.57, $p < 0.001$, partial η^2 0.59) and a significant group x condition interaction ($F(14, 188) = 2.61$, $p = 0.008$, partial η^2 0.18,with Huynh-Feldt correction). Tukey's HSD indicated that ASD- performed significantly differently from both controls and ASD+ (both $p < 0.001$), whereas there was no significant difference between controls and ASD+. We explored the group x condition interaction with separate univariate ANOVAs for each condition and further post hoc tests (see Table 2.2 and Table 2.3, overleaf).

There were only three masking conditions in which there was no clear difference in SRT across groups (see Table 2.2) - the steady noise (condition 1), the Japanese male speech (condition 3) and the CRM distractors spoken by the target female (condition 8). Although post hoc testing for the steady noise condition indicated a difference between ASD+ and ASD- which approached significance, there was in fact only one ASD- listener who performed outside the normal range (see Figure 2.4, plot i).

Table 2.2 Speech reception thresholds across mCRM conditions

Condition	1 Steady noise	2 AM noise	3 Japanese male speech	4 Irrelevant English male speech	5 Vocoded CRM distractors	6 Male CRM distractors	7 Female CRM distractors	8 CRM target talker
Control Mean (dB)	-10.3	-15.0	-26	-23.4	-21.1	-25.2	-22.9	-4.5
ASD+ Mean (dB)	-11.3	-14.4	-25.5	-25.8	-21.1	-23.4	-23.3	-3.3
ASD- Mean (db)	-9.0	-11.4	-22.7	-16.8	-14.7	-16.6	-7.8	-1.8
Univariate Anova p = (df 2, 24)	0.017	0.001*	0.27	0.002*	0.003*	< 0.001*	< 0.001*	0.18

Table 2.3 Results of post-hoc testing (Tukey's HSD): between-group differences (mean SRT)

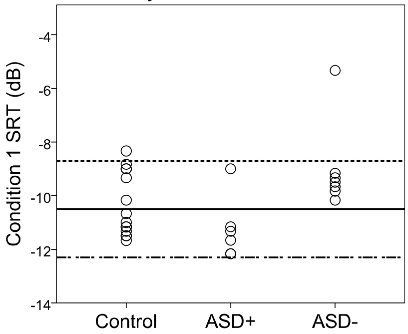
Condition	Control / ASD- Difference in mean SRT (dB)	Control / ASD- p =		ASD+ / ASD- Difference in mean (SRT dB)	ASD+ / ASD- p =
1	1.3	0.09		2.3	0.014
2	3.6	<0.001*		3	0.012
3	3.3	0.25		2.8	0.51
4	6.6	0.006*		9	0.004*
5	6.4	0.002*		6.4	0.10
6	8.6	<0.001*		6.8	0.004*
7	15.1	<0.001*		15.5	0.001*
8	2.7	0.16		1.5	0.42

* indicates significance at p < 0.01

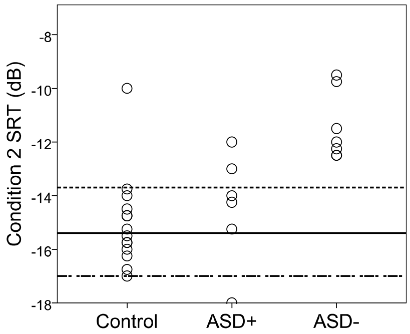
Even with the Japanese speech masker three of the seven ASD- listeners still performed within the normal range (Figure 2.4, plot iii). The results of these two conditions (steady noise and Japanese masker) imply several things: firstly, that with a steady noise masker the ASD- group experienced similar amounts of energetic masking to controls, and therefore their auditory peripheral filtering is not atypical; secondly, that they were able to cope with the basic cognitive and linguistic demands of the task procedure; thirdly that - in spite of poor performance with AM noise - they were not significantly impaired in all other types of fluctuating masker, such as the Japanese speech.

Figure 2.4 mCRM task: individual speech reception thresholds (showing outliers)

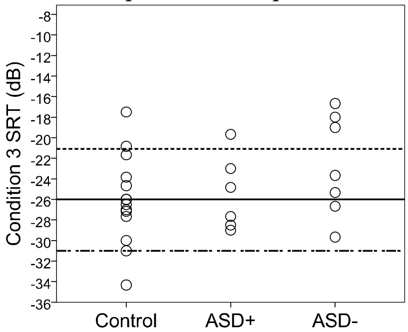
i) Condition 1: steady noise masker



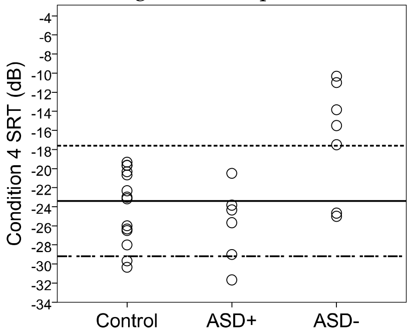
ii) Condition 2: AM noise masker



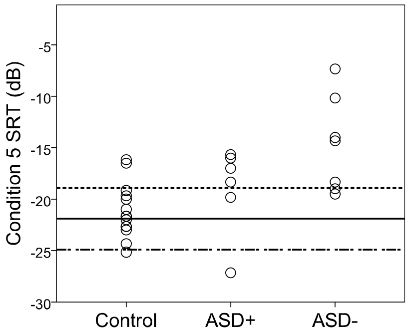
iii) Condition 3: Japanese male speech



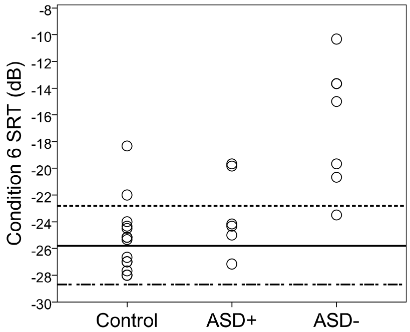
iv) Condition 4: English male speech



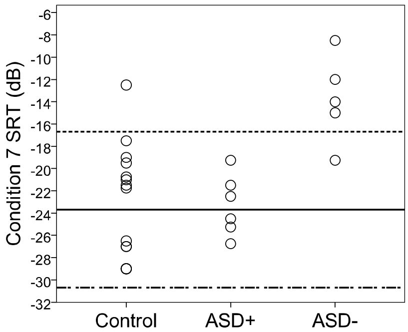
v) Condition 5: Vocoded mCRM distractors



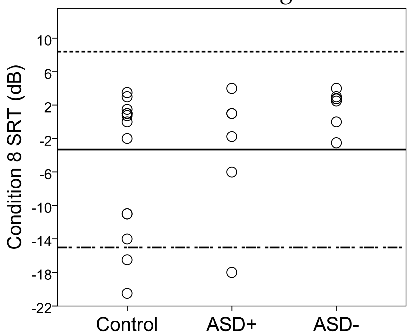
vi) Condition 6: Male mCRM distractors



vii) Condition 7: Female mCRM distractors



viii) Condition 8: Same talker target & distractors



_____ control mean (excl. outliers)

- - - - - - z = + 1.65 (poor performance)

--- --- --- z = -1.65 (good performance)

'Normal' ASD- performance in condition 8, however (see Figure 2.4, plot viii) - where the task was so difficult that even most control participants performed extremely poorly - is more difficult to interpret. Although the ASD- listeners were all within the normal range, the scatterplot indicates that, even in this condition, a handful of control and ASD+ listeners were able to perform fairly well - with SRTs of around -14-19 dB - whereas the best-performing ASD- listener's SRT was only -3 dB.

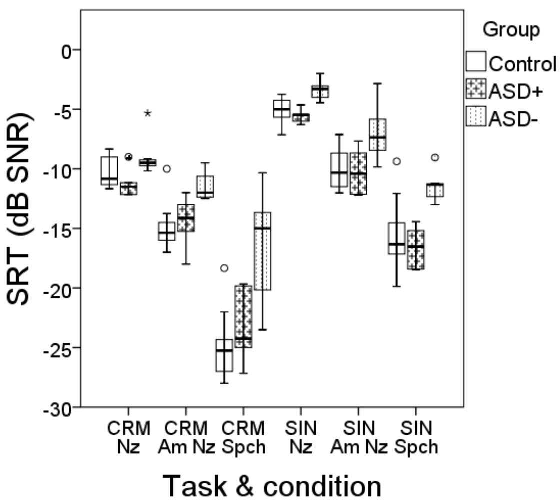
Whereas the ASD- listeners were able to perform normally in the steady noise and Japanese speech masker conditions, and showed no major deficit in the most difficult identical target/masker talker condition, for all other conditions univariate ANOVAs showed a robust group effect (all $p < 0.001$). Post hoc testing indicated that the ASD- group performed significantly poorly in all but the 3 conditions already mentioned, and were particularly impaired with masking noises involving intelligible English speech (Tukey's HSD comparing control with ASD-, all $p < 0.01$: conditions 4, 5, 6, 7). Moreover, the difference in mean SRT between ASD- and controls increased as the background speech became more similar to that of the target material, except when the masker talker was identical to the target (see Table 2.3). With vocoded CRM distractors, the difference between control and ASD- mean SRTs is already large (approaching 7 dB) compared to the deficit in amplitude-modulated noise (4 dB), but with the female CRM distractor sentences - which are perceptually very similar to the targets - the difference grows to 15 dB. Whilst the differences in SRT between ASD+ and ASD- tended to be rather smaller and in many cases did not reach significance, they also followed a similar pattern (see Table 2.3).

ASD- seem to perform most differently from other listeners with unmodified English speech as a masker, but the content of this background speech appears to have less of an impact than talker similarity. For ASD-, mean SRTs with unrelated male speech background and the male CRM sentence backgrounds were virtually indistinguishable (both around -17 dB) , whereas the shift from male to female CRM distractors raised (worsened) their SRT by almost 9 dB. This effect of changing the CRM masker gender was only 2 dB for controls, and less than 1 dB for ASD+.

2.3.2 SIN task results

Control participants' SRTs for this task (see Figure 2.5) tended to be higher than in the equivalent mCRM conditions, which can be explained by the more complex target material and higher cognitive demands of the task. Otherwise, however, the typical pattern of improvement from steady noise to AM noise, and from AM noise to background speech pertained. Data was again analysed with 3 participant groups (Control, ASD+ and ASD-) based on mean z-scores across both tasks/all conditions, as for the mCRM analysis. A repeated-measures ANOVA (3 groups x 3 conditions) showed significant effects of condition ($F_{2, 48} = 202.4$, $p < 0.001$, with Huynh-Feldt correction) and group ($F_{1, 24} = 22.3$, $p < 0.001$, partial η^2 0.65), but no interaction ($p = 0.18$). Tukey's HSD indicated that ASD- again performed significantly poorly compared to both controls and ASD+ (both $p < 0.001$).

Figure 2.5   Boxplots showing subgroup SRTs across similar CRM and SIN conditions

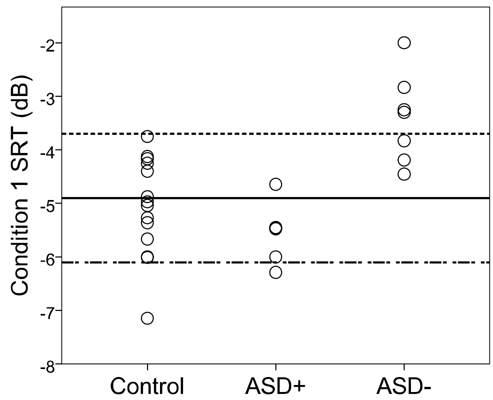


Although the group x condition interaction was not significant, there did seem to be a similar trend to that in the mCRM data, with the ASD- participants tending to be more impaired with the speech masking than with non-speech noise: no ASD- listener performed within the normal range in condition 3 (see Figure 2.6, iii). The difference between their mean SRT (-11.5 dB) and that of the controls (-15.8 dB) in this condition was only 4 dB, however, which is much more in line with the size of their deficit in modulated noise (2.9 dB) in this task than it was in the mCRM (deficit with AM noise 3.3 dB, deficit with unrelated talker 6.6 dB). In the mCRM, controls showed a 10 dB improvement between the AM noise masker and the male CRM sentences, whereas in the SIN the improvement from AM noise to speech was only 6 dB. ASD- improved by only 5 dB in the mCRM, and by 4 dB in the SIN. Overall, both the pattern of control performance and the size of the ASD- group's impairment in our SIN task were more in line with Alcántara et al.'s (2004) results, than were our results from the mCRM.

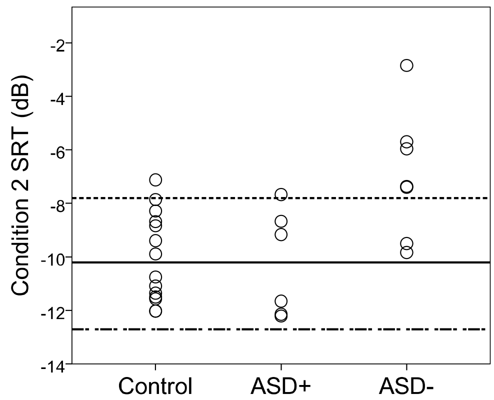
The SIN task takes longer to administer than the mCRM and the procedure is more taxing, so we were unable to include as many background noise conditions here. To investigate whether there was a higher order interaction between task/target material, group and masking condition, therefore, we focused only on the 3 conditions which occurred in both tasks: steady noise, AM noise and male background speech (see Figure 2.5). The AM noise in mCRM was derived from the male CRM distractor sentences so Condition 3 was the speech condition we included for that task. The results of a multivariate ANOVA (within-subjects factors: 2 tasks and 3 conditions; between-subjects factor: 3 groups) showed the expected main effects of target material, masking condition and group (all significant $p < 0.001$). There was no interaction between target material and group, but the interactions between masking condition and group ($p = 0.001$) and between target material and masking condition ($p = 0.002$, partial η^2 0.24) were both significant. There was also a marginally significant 3-way task x condition x group interaction ($p = 0.035$, partial η^2 0.20). Given the small sample sizes involved here, these results cannot be viewed as reliable, but they do appear to reflect our earlier observation that patterns of performance, particularly in the background speech condition, appeared to differ across tasks. Separate 2x3 ANOVAs for each masking condition on its own, entering task (CRM or SIN) and group as factors, showed a significant task x group interaction only for the speech masker condition.

Figure 2.6 SIN task: individual speech reception thresholds (showing outliers)

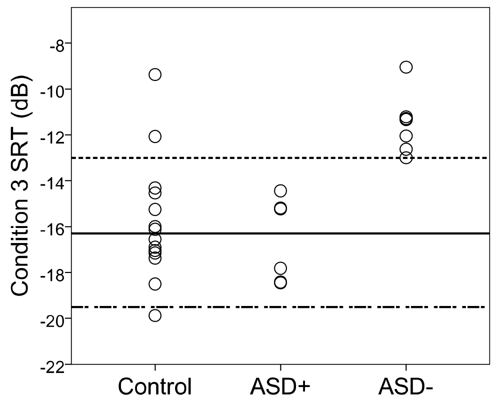
i) Condition 1: Steady noise masker



ii) Condition 2: AM noise masker



iii) Condition 3: English male speech masker



- - - - - - z = + 1.65 (poor performance)

______ control mean

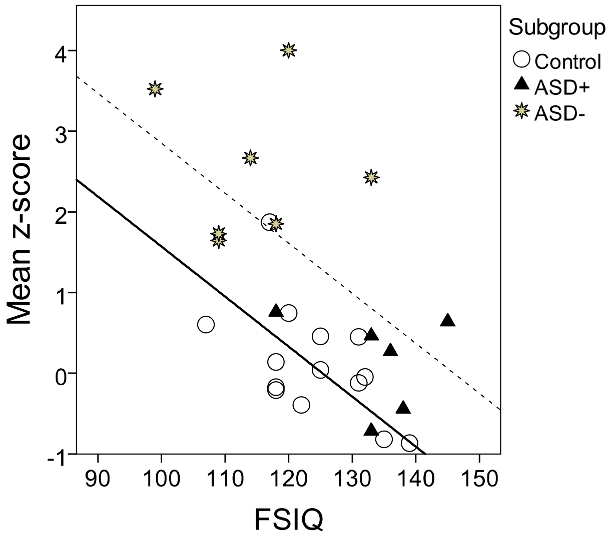
--- --- --- z = -1.65 (good performance)

2.3.3 IQ as a factor in listening task performance

The 2 main control and ASD groups were initially matched for age and IQ, but the split between ASD+ and ASD- subgroups took place a posteriori so we could not actively match the participants. The final 3 groups did not differ significantly in terms of age, but the results of a univariate Anova indicated a significant group difference in full-scale IQ score ($F_{2, 24} = 7$, $p = 0.004$). Tukey's HSD indicated that this was driven by a significant difference between ASD- (mean FSIQ 115, s.d. 11) and ASD+ (mean FSIQ 134, s.d. 9) ($p = 0.003$). Neither ASD group differed significantly from controls (mean FSIQ 124, s.d. 9).

FSIQ also appeared to be an important factor in our listeners' performance with speech in noise (see Figure 2.7). The overall correlation between FSIQ and mean z-score collapsed across listeners was significant: $r = -0.55$, $p = 0.003$. We used a regression model to explore the relative contributions of group (control or ASD) and FSIQ to mean listening test z-score, with group x FSIQ score interaction as an additional predictor. This model was highly significant ($F_{3, 23} = 10.0$, $p < 0.001$) . The group x FSIQ interaction was not significant ($p = 0.50$), however, indicating that the way in which listening performance varies with FSIQ is indistinguishable between groups. The interaction term was removed from the model, therefore and the final results showed that both group ($p = < 0.001$, partial η^2 0.44) and FSIQ ($p = 0.001$, partial $\eta^2 = 0.39$) were highly significant factors in individual listening performance. The power of this final model itself was high: partial $\eta^2 = 0.56$, $p < 0.001$. The intercept values showed that the ASD group's z scores were around 1.3 units worse than control listeners', and performance for both groups improved by 0.06 z with every 1 point increase in FSIQ (i.e. almost a 1 point improvement in z-score with 15 IQ points, or one standard deviation in listening performance for every one IQ standard deviation).

Figure 2.7 Scatterplot showing effects of group and FSIQ on listening performance



Lines indicate regression of mean z on FSIQ for control group (solid line, y = 0.062 * x + 7.77) and ASD group (dashed line, y = 0.062 * x + 9.05) based on the GLM described above.

ASD+/ASD- data points are shown separately for information only.

The implications of this regression model are that, whilst IQ affects speech in noise performance to the same extent for both groups, the ASD listeners have an additional deficit which causes even some of those at the top end of the range to perform considerably less well than controls with equivalent IQ scores.

2.4 Discussion

We used a modified version of the Coordinate Response Measure to investigate speech reception thresholds in ASD across a much wider range of masking conditions than previously reported, and several aspects of our results are striking. The first key observation is that only a subset of our ASD listeners showed impaired speech intelligibility in noise in any condition, whereas a substantial number performed within the normal range across nearly all maskers (including amplitude-modulated noise). Secondly, participants in our low-performing ASD- group showed deficits for speech-on-speech masking which were far in excess of their impairment for speech in fluctuating noise, but only when the background speech was in a language they understood. Thirdly, the ASD- listeners' SRTs were increasingly impaired as the target and masker speech became more similar: their mean SRT increased (worsened) by 9 dB with the change from an opposite-gender to a same-gender background talker, whereas the control mean SRT changed by only 2 dB across the same two conditions.

Whilst previous work on speech in noise in ASD has highlighted a possible deficit in dip listening amongst this group (Alcántara et al, 2004; Groen et al., 2009), our results imply that this may play only a relatively small part in everyday listening difficulties. In fact, with Japanese speech in the background even the ASD- listeners' SRTs were not significantly different from those of controls, suggesting that if there is a difficulty with dip listening it is highly dependent on the type of masker involved. At least for our lower-performing ASD adults, far larger deficits with intelligible background speech than with unintelligible speech or fluctuating noise, and deficits which increased with increasing target-masker speech similarity, seem to suggest that impaired auditory selective attention to speech may also be an important factor.

One of the benefits of the mCRM as a method for estimating listening skills in noise is that it is relatively quick to administer, with each block of trials taking only 2-3 minutes to complete. This allowed us to investigate more background noise conditions, and in particular more combinations of simultaneous talkers, than is typically possible with more complex target sentences. Further work is needed, however, to establish whether target-masker talker similarity affects SRTs for other types of target speech in the same way as for the mCRM. One aspect of our data suggests that perceptual differences between talkers might not be quite so important for performance with more complex targets, and that is the fact that increased susceptibility to speech maskers in ASD appeared far more clearly with the mCRM than with the SIN. Although we continued to see a clear overall split between high- and low-performing listeners in the SIN, and the ASD- group's SRTs were again significantly worse than controls, we did not find a significant interaction between group and masking condition with this more complex target material. In this respect, our SIN results were closer than our mCRM results to what Alcántara et al. (2004) found for sentences in noise.

Our ASD- listeners' mean SRT in the SIN background speech condition was only 4 dB worse than that of controls - a considerably smaller deficit than with the same speech masker in the mCRM where the group mean difference was 6.6 dB. Furthermore, this impairment with speech masking in the SIN was much more in line with these listeners' deficit with fluctuating noise (3 dB), and also with the impairments found by Alcántara et al. for speech masking in their study (around 3 dB at most). Whilst bearing in mind the possible effects of the small sample sizes involved, the mCRM does appear to highlight difficulties with speech-on-speech masking in ASD more dramatically than other sentence tasks. What is somewhat surprising, though, is that it does so not just in its typical configuration - which was designed to maximise informational masking - whereby two competing CRM sentences are presented simultaneously, but is almost as effective with task-irrelevant speech in the background. None of our groups of listeners showed more than a 2.5 dB change in mean SRT going from the irrelevant background speech condition to CRM distractors. In fact, for the ASD- group the effect was negligible (0.2 dB) and it was actually the ASD+ group (who had performed even better than controls with the irrelevant speech masker) who showed the biggest difference between conditions 4 and 6 (2.4 dB). This tends to suggest that either the mCRM target material itself, or some other feature of the task procedure, may have been quite important in determining the distractor interference effects we observed.

A smaller speech masker effect for the ASD- listeners was one difference between the SIN and mCRM results, but there is a second important difference which may provide a clue to interpreting the overall pattern of performance across tasks. This was that our ASD- listeners showed significantly worse SRTs than controls in the SIN steady noise condition, but not in the mCRM steady noise. Admittedly this difference between mean SRTs was only small (1.5 dB), but 4 of the 7 ASD- listeners' SRTs were outside the normal range, whereas only one performed significantly poorly with steady noise in the mCRM, and neither Alcántara et al (2004) nor Groen et al. (2009) found significant group differences in steady noise. The obvious explanation for this apparent discrepancy in results with steady noise backgrounds is that it arises from differences in the complexity of the target material. Groen et al. presented single words, thus eliminating all need for syntactic and contextual processing, and Alcántara's stimuli were ASL sentences, which contain only 3 key words (compared to our ABC sentences' 5) and consist of rather simpler vocabulary, based on child language samples (e.g. ASL: *He hit his head*; ABC: *The boy's canoe slid on the smooth planks*). Not only do the ABC sentences tax speech input processing more heavily, they also clearly place heavier demands on working memory and speech output when it comes to giving a response, which may well explain why some ASD- listeners could perform normally with steady state noise in the mCRM, but not in the SIN. Accounting for their heightened susceptibility to speech-on-speech masking in the mCRM is rather more complicated, but it may be that the lower processing demands of the target material may have played a role there too.

There are several ways to think about how increased susceptibility to interference from intelligible distractor speech might arise in ASD, and why this interference might be more evident in the mCRM than in the SIN task, but there is no way to be certain from our current study what the underlying factors actually are. There are two potential explanations which seemed plausible a priori, but were not supported by our data. At the most basic level speech is a social stimulus, and so it seems possible that the sound of two simultaneous talkers might - on a purely perceptual level - be sufficiently unpleasant to some ASD listeners that it could cause distraction from the task. This does not seem to have been the case, though, since even our ASD- listeners showed no significant deficit in SRT when the background speech consisted of unintelligible Japanese. Apparently, substantial interference from background speech only occurred when the speech could be understood. However, as already mentioned, a defining feature of the mCRM is that two (or more) highly similar mCRM sentences can be presented simultaneously, meaning that both the target and the distractor speech contain colour and number words. This situation maximises informational masking, as all the messages reaching a listener are potentially task-relevant. The listener must then identify the target sentence on the basis of the talker and the animal word in the carrier phrase, form this sentence into a coherent perceptual stream separate from the distractor, focus selective attention on it and inhibit their response to the competing message. As expected, our control listeners appeared to achieve this with relative ease, at least with the opposite-gender distractor material, whereas the ASD- group's mean SRT in this condition was more than 8 dB worse. However, this group performed almost as poorly when the background speech consisted of irrelevant sentences as with mCRM distractors, and this suggests that distractor key word intrusions - in the sense of a specific difficulty with inhibiting a task-relevant competing colour/number response - were not the sole source of error for them in the English male masker conditions. One problem with presenting the mCRM material adaptively is that - particularly when two groups' SRTs differ as widely as this - there is insufficient data at any given SNR to compare error patterns in a meaningful way. Nevertheless, a difference of less than 1 dB in SRTs between irrelevant masking speech and CRM distractors suggests that colour or number intrusions did not seriously affect performance, at least with the opposite gender background talker. The dramatic shift in SRT with the change from opposite- to same-gender CRM distractors, however, could indicate an increase in such intrusions, but without an irrelevant female masker condition to compare, it is impossible to be sure.

Even in the absence of a clear-cut distractor intrusion effect amongst the ASD- listeners, heightened susceptibility to masking by English background speech, but not by Japanese speech, does seem to point to atypical attentional processes. Any intelligible background speech reaching central processing could, after all, compete for linguistic and working memory processing capacity, even if it is not in direct semantic competition with the target. Remington et al. (2009) have suggested that ASD may be associated with increased perceptual capacity - at least in visual search tasks of the type they used - causing distractor interference to continue occurring under conditions of higher perceptual load than for controls. Clearly, a visual perceptual load experiment is very different from an investigation of speech perception in noise, not least because the key dependent variable in the former is reaction time for a simple target, rather than accuracy in reporting a relatively complex target as in the latter case. There is no sensible way to compare the demands of these two tasks. Nevertheless, if in a given listening situation such as the mCRM, perceptual load exceeded capacity for control listeners but not for some ASD listeners, it seems possible that background speech material might not be processed by control listeners (early attentional selection), whereas in the ASD group it might reach higher linguistic processing centres where it could compete with or distract from the target material. A divided attention condition in the mCRM would be one way to investigate this, and without that it is impossible to say whether the ASD listeners were more able than controls to report the content of the background speech. However, a bigger problem with using increased perceptual capacity in ASD to account for our results is that neurotypical adults also experience distractor interference in the CRM (albeit to a lesser extent than our clinical group) - especially when listening to two similar talkers (Brungart, 2001). This suggests that in fact a two-talker listening situation in the mCRM does not exceed normal perceptual capacity.

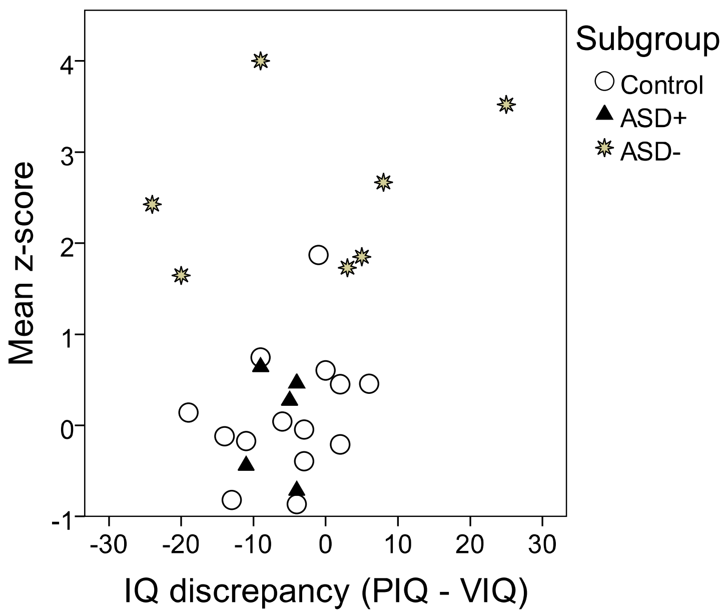
Load theory of selective attention, however, also includes a cognitive control component which becomes important for response accuracy when perceptual load is low enough for both competing stimuli to be perceived, and this might well have been a factor influencing our ASD listeners' performance. Cognitive load operates on selective attention in the opposite way to perceptual load: so when perceptual load increases, distractor interference is less likely, but when perceptual load is low and the cognitive load increases, distractor interference becomes more likely (Lavie, Hirst, de Fockert, & Viding, 2004). Increased susceptibility to speech masking in ASD could simply reflect a deficit in cognitive control. One way this might happen would be if processing speech in itself somehow placed a heavier cognitive load on listeners with ASD than neurotypical controls, but that seems unlikely given that their deficits were so much greater in the mCRM than in the (superficially at least) more demanding SIN. On the other hand, a number of studies have pointed to atypical central executive functioning in ASD, which may be underpinned by differences in frontal lobe development and functional connectivity between the frontal and other cortical/subcortical regions (Hill, 2004). If cognitive control were weaker overall in these listeners we would expect to see atypical distractor interference, especially under conditions of low perceptual load. Increased deficits with speech maskers in the mCRM, then, could even be interpreted as resulting from it presenting a lower perceptual load than the SIN. The mCRM target words are from a relatively small, closed set and there are only two in each sentence. They can be discriminated from each other largely on the basis of vowel differences alone, and vowels are the phonemes least vulnerable to energetic masking. This could, perhaps, be viewed as constituting relatively low perceptual load. Keywords in the ABC sentences however, come from an infinitely open set (as far as the participant is concerned), and so all the phonetic information is potentially valuable, as is the syntactic and semantic relationship between words. On the face of it, therefore, this target material presents a higher perceptual load. It may be that this higher load is sufficient to exceed perceptual capacity and thereby suppress distractor speech at an earlier processing stage, thus helping to reduce or eliminate informational masking.

The dramatic effect of target-masker speech similarity on the ASD- group's SRTs is consistent with the hypothesis that they have poor cognitive control. If the top-down ability to direct and sustain attention to task-relevant stimuli is impaired in the presence of distractors, it seems likely that bottom-up differences in the perceptual distance between those distractors and the target material will have a big impact on performance. There is nothing in our data, however, to rule out the opposite hypothesis, which is that auditory stream segregation processes are less efficient in ASD, and that the problems with cognitive control are driven from the bottom-up. Independent evidence for a specific deficit in auditory stream segregation in ASD is less abundant than for a more general impairment in executive function, but at least one study measuring event-related potentials has suggested that cortical activity associated with passive auditory streaming processes may be reduced in children with Asperger's syndrome (Lepisto et al., 2009). By contrast, a single behavioural study which required participants to make an active judgement (sequential tone delay detection) found no significant difference in performance between children with AS and age-matched controls across several different frequency separations between tone streams (Fullgrabe, 2009). Since both these experiments used pure tone stimuli, and involved detecting a fairly low-level change in a single acoustic feature (intensity or timing), neither, perhaps, can tell us much about stream separation abilities with more complex stimuli, or their effects on speech perception. ASD listeners have actually shown superior performance in at least one study involving the discrimination of target melodies embedded in distractors, and this has been interpreted as a sign of enhanced stream segregation (Bouvet, Donnadieu, Mottron, & Valdois, 2011). Even this more challenging test remains very different from speech perception in noise, however, as the target melody was presented initially in isolation and then embedded in the distractor. Listeners' task was to judge whether the embedded melody matched the target or not, and it may well be that if we had asked for a similar same/different judgement for sentences with speech masking our listeners would have performed normally. There is very little evidence that simple stimuli like pure tones are processed atypically in ASD, whereas converging evidence from behavioural and electrophysiological studies, as outlined in the introduction, suggests that both speech and voice processing is abnormal from early childhood onwards. Given that, even when presented in isolation, speech and voice stimuli seem to be processed differently by ASD listeners, it would perhaps be surprising to find that streaming processes operate normally when these sounds are presented concurrently.

Our ASD listeners' pattern of performance across conditions raises as many questions as it answers, and further work is needed to elucidate the mechanisms underlying both heightened susceptibility to speech maskers and increased target-masker similarity effects. One of the biggest effects we observed, however, was that individual performance across the ASD group was highly variable. Such individual variability is not unusual in studies of this kind, particularly when it comes to a condition as heterogeneous as autism, and it can often reflect variation in factors not directly related to a specific task or task condition, such as symptom severity, general attention level or motivation. Slightly surprising, therefore, was that in the current experiment this variability was restricted mainly to listening conditions involving English background speech, and was particularly striking with the same-gender target and masker (condition 7): ASD mean SRT -16 dB, standard deviation 9; control mean -23 dB, std. dev. 5. As discussed, the speech masker conditions of the mCRM seem to load more heavily on cognitive control than do standard sentences in noise, which may explain both the amount of variability we saw in SRTs, and the reason why Alcántara et al. (2004) did not record anything similar amongst their autistic participants. In non-speech psychoacoustic tasks, individual variability amongst adults is often viewed as a hallmark of informational masking (see for example, Neff & Dethlefs, 1995; Durlach, Mason, Shinn-Cunningham, Arbogast, Colburn, & Kidd, 2003), whereas in tests of speech-on-speech masking individual variability, along with overall susceptibility to informational masking itself, is high during childhood but tends to reduce by early adulthood (Wightman, Kistler, & O'Bryan, 2010). Prefrontal maturation may be one reason for this increasing resistance to speech maskers, and this again demonstrates the importance of top-down attentional processes in speech masking situations. It also highlights the fact that speech perception in a multi-talker environment is a highly-skilled task, and that experience may play a role. As speech as a stimulus appears to be less intrinsically attractive or motivating in autism, it may be that expertise in processing develops to a lower degree.

We found that full-scale IQ was an important factor in all our listeners' performance, but it also appeared to explain at least part of the apparent difference between our high- and low-performing ASD participants. This is in line with a number of studies suggesting that exceptionally high IQ may be a protective factor in ASD, and that it may even support the development of executive skills (Kalbfleisch & Loughan, 2012). In fact Kalbfleisch & Loughan provide more specific evidence that IQ discrepancy favouring verbal IQ (i.e. higher verbal than performance IQ scores) may be associated with some aspects of executive function, including working memory and inhibition, which may be important for speech in noise. We only used the brief version of the abbreviated WASI, so our data is not well-suited to measuring IQ discrepancy, but nevertheless it is striking that our ASD+ listeners generally showed quite small IQ discrepancy, and that none had performance t-scores higher than their verbal scores. By contrast, the ASD- participants showed more extreme discrepancy at both ends of the scale and performance IQ exceeded verbal IQ for more than half of them (see Figure 2.8).

Figure 2.8 Scatterplot: distribution of IQ discrepancy relative to listening test performance



In conclusion, our results are strongly suggestive of increased susceptibility to speech-on-speech masking in high-functioning autism, and the large deficits our ASD- listeners showed with speech distractors in the laboratory seem commensurate with subjective reports of poor speech intelligibility in everyday life. More work is needed in order to clarify what underlies this impairment, but reduced cognitive control, either alone or in combination with more specific deficits in processing efficiency or auditory scene analysis, seems to be implicated. IQ may act as a buttress to executive skills, allowing the most able ASD adults to resist informational masking, but even these individuals commented on difficulties in natural listening environments. In real life, potentially distracting information assails listeners in every modality, and it is important to remember that the auditory attentional problems we have observed under experimental conditions may well be compounded during normal communication by sensory sensitivities, pragmatic difficulties, and the demands of social interaction.

Chapter 3 Informational masking in the mCRM: further investigations

3.1 Introduction

The eight mCRM masker conditions described in Chapter 2 were a core set, selected with the aim of investigating one main hypothesis: that listeners with ASD would show heightened susceptibility to speech-on-speech masking. Since the mCRM had not been used with autistic listeners prior to this, a second important aim was to establish how their performance on the mCRM would relate to their performance with more complex target sentence material. The results presented in chapter 2 indicate that at least a subset of high-functioning adults with ASD perform poorly across many different types of masker, and that they do indeed seem to be particularly susceptible to intelligible speech maskers - especially when the target and masker talkers are similar. Moreover, these speech masker deficits seemed to be magnified in the mCRM, when compared to results of the SIN task in which the greater complexity of the target material and the overall cognitive demands may have been additional factors. Given how little is known about speech masking in ASD, therefore, the mCRM appears to be a time-efficient listening task with which to investigate a number of more speculative hypotheses about the nature of the deficit, whilst to some extent limiting the influence of top-down linguistic processes. The work described in this chapter was conceived with the aim of exploring further how adults with ASD may differ from neurotypical listeners in the way they process multi-talker speech environments. Cues to speech stream segregation, such as masker pitch content and target-masker spatial configuration, were manipulated, and further aspects of informational masking such as error patterns and performance with more than one speech masker were investigated.

3.1.1 Error patterns and the effects of an adaptive procedure

Higher numbers of masker intrusions in the CRM (error responses which include the colour and/or number from a simultaneously-presented masker sentence) are often taken to reflect higher levels of informational masking, so this can be a useful measure. Standard experiments using the CRM (e.g. Brungart, 2001) tend to use the method of constant stimuli and present multiple trials at a number of pre-designated SNRs, which are the same for every listener. In his speech masker condition, for instance, Brungart (2001) presented a total of 2000 trials to each listener, at 10 different SNRs ranging from -12 to +15 dB in 3 dB steps (compared to around 30- 45 trials per listener/condition in the current experiment, using adaptive tracking). This means that there is plenty of data at each SNR to perform an error analysis. When using an adaptive testing procedure for the mCRM, however, as in Experiment 1 here, speech reception thresholds (SRTs) can be estimated from relatively short blocks of trials, allowing measurements of speech intelligibility across a relatively wide variety of masking conditions. The drawback of this adaptive approach, however, is that the range of signal-to-noise ratios at which the target is presented during a block of trials, and the number of times it may be presented at any given SNR during that block, varies from listener to listener depending on individual performance. Although collecting data at fixed SNRs in all the masking conditions of this experiment would have taken too long to be feasible, we took one key condition - the female target with male mCRM masker - to compare more directly how groups performed at each of a small number of SNRs. One prediction was that heightened susceptibility to informational masking might lead to increased numbers of masker intrusion errors amongst the ASD listeners.

A further benefit of presenting at least one condition at fixed SNRs was to gain some insight into whether the SRTs estimated from the adaptive procedure were indeed accurate. One assumption of adaptive test procedures is that there is a monotonic relationship between stimulus 'strength' (here, SNR) and performance (Levitt, 1971; Leek, 2001). However, Brungart's work suggests that the psychometric function for single-talker maskers in the CRM may - sometimes - contain a plateau at around 0 to -10 dB SNR, reflecting the use of intensity cues to support auditory stream segregation (Brungart, 2001). If there had been a similar plateau affecting our listeners' performance across SNRs - and particularly if there had been a difference between our two groups in this respect (because, for instance, ASD listeners used streaming cues differently from controls) - it seems possible that the SRTs estimated from reversals in the adaptive track might be highly inaccurate. If ASD participants used streaming cues less effectively, they might even perform sufficiently poorly at these mid SNRs for the tracking procedure to stall at this level, even though they might actually have performed better at a lower SNR had they reached it. For this reason, we presented a substantial number of trials with the fixed procedure at -20 dB, as well as at 0 dB and -6 dB. Furthermore, the adaptive blocks did not include catch trials (occasional trials at 'easy' SNRs, to monitor overall attention to the task), mainly from concern that these might be more distracting or disruptive to ASD listeners than controls. Presenting a block of trials at +6 dB allowed us to explore performance at this easy level more directly.

3.1.2 Gender of the target talker

Another potential confounding factor which could have contributed to group differences in performance with speech maskers - albeit highly unlikely - was the gender of the target talker. The same female target talker was used throughout the mCRM in this experiment, mainly to simplify the procedure for our participants. Again, this diverges from the paradigm adopted by Brungart, whose target talker on any given trial in a block could be either male or female, selected from a pool of 8 sets of target recordings (4 male, 4 female). There is no reason to suppose that listeners with ASD might perform worse with a female target than a male, or even just perform worse with this particular target talker than any other, but to be sure that this feature of the testing was not crucial, the opposite-gender masker configuration was reversed. The male talker who was the mCRM masker in condition 6 spoke the target sentences for a new condition (condition 10), and the female talker who was usually the target became the mCRM masker. Results of this condition were expected to follow the same pattern as those with the original opposite-gender masker, with the ASD- listeners continuing to show a substantial deficit relative to controls.

3.1.3 Multi-talker speech masking

The remainder of the new conditions manipulated aspects of the target/masker speech combination which are known to affect neurotypical listeners' performance in multi-talker environments. Firstly, whilst adult listeners are usually extremely good at resisting interference from a single opposite-gender speech masker, target intelligibility decreases dramatically with two or more speech maskers (Miller, 1947; Simpson & Cooke, 2005; Brungart et al., 2001). The two-talker masker situation is particularly interesting because SRTs can be as much as 8-10 dB higher with a two-talker masker than with a single-talker, whereas the change in SRT when going from 2 to 3 or 4 maskers is much smaller (see, for example, Cullington & Zeng, 2008). Brungart has also shown that the relationship between SNR and performance appears to become more monotonic with a two-talker masker, and argues from this that listeners may be unable to use intensity differences between the target and two maskers to boost performance (Brungart et al., 2001). Cullington and Zeng (2008) demonstrated that whilst performance with a two-talker masker can be better than with a steady noise, higher numbers of masking talkers lead to performance which is actually worse than with noise. They interpret this as indicating high levels of informational masking with 3 and 4 talker maskers, which is partially mitigated in the case of a two-talker masker by the fact that enough of all the sentences is intelligible to allow listeners to use grammatical and semantic context to help segregate the streams. Noticeably, in Brungart's study performance even with the one- and two-talker maskers is worse than with speech-modulated noise, which probably reflects the lack of helpful contextual cues in the CRM. If our ASD- listeners' poor performance with a single-talker masker arose mainly from an inability to use these intensity cues for auditory stream separation, one might predict that their performance in a two-talker masker might more closely resemble that of controls. Alternatively, as the potential for informational masking appears to be higher with a two-talker masker, their performance might be even weaker relative to controls.

3.1.4 Spatial release from masking

Whilst increasing the number of background talkers in an auditory environment has a detrimental effect on intelligibility amongst normal listeners, increasing the spatial separation between a target and masker talker has the opposite effect on performance. More than one factor contributes to this spatial release from masking. Firstly, there is release from energetic masking. Early studies demonstrated that this can improve SRTs by up to 10 dB in cases where the target is located beside one ear and the masker faces the opposite ear, and over 8 dB when target and masker are separated by 90° azimuth in the horizontal plane. This occurs partly because of interaural time differences (a binaural effect) and partly because of the 'better ear' advantage (a monaural effect), in which the SNR for a spatially-separated target and masker will be more favourable at the ear which is further from the masker thanks to acoustic head shadow (Dirks & Wilson, 1969). A second benefit of spatial separation, however, is release from informational masking. Freyman, for example, created an illusory spatial separation between target and masker using the precedence effect, which minimised the better ear effect. In this scenario, listeners' SRTs for a speech-masker condition improved with perceived separation from the target, whereas SRTs in the equivalent non-speech noise masker condition showed no change (Freyman, Helfer, Mccall, & Clifton, 1999). A number of follow-up studies were conducted, manipulating various acoustic and linguistic features of the stimuli to separate out informational and energetic masking effects, all of which indicated that release from informational masking can be a significant factor in improved SRTs when a speech-masker is spatially separated from the target (Freyman, Balakrishnan, & Helfer, 2001). Given the complex nature of the binaural listening situation, there is certainly a range of potential reasons for ASD listeners to perform differently from controls. Nevertheless, if poor performance in speech masker conditions of the mCRM genuinely reflects heightened susceptibility to informational masking, significant improvements in SRT when the target and masker are spatially separated might well be expected - perhaps even in excess of those shown by neurotypical listeners.

3.1.5 Masking of speech by music

Speech is only one type of masker which is commonly encountered in everyday life, and it is perhaps the one that seems most likely to be problematic for listeners with ASD because of its social and affective content. Another sound which frequently forms an auditory background to conversation, however, is music, and at least one study has suggested that its masking properties can be similar to those of a single interfering talker (Rhebergen, Versfeld, & Dreschler, 2008). Its dynamic spectro-temporal properties mean that music resembles speech on an acoustic level, but it also has its own hierarchical syntactic structure - and, arguably, semantics - which resemble those of speech and which appear to recruit similar cortical processing areas (Patel, 2003; Koelsch, 2011). This leads to the suggestion that the effects of background music may involve informational as well as energetic masking, since it may compete for resources at a more central level. Research by Shi has demonstrated that the effects of 'scrambling' a masker (i.e. cutting it into short sections and rearranging their order to destroy the hierarchical structure) are similar for music and for speech: sentence SRTs are worse with the scrambled version than the original masker (Shi & Law, 2010). These results were interpreted as demonstrating that listeners use the hierarchical structure of the original music and speech masker to make online predictions about what is coming next, to aid perceptual stream segregation. When the structure is degraded, stream segregation is impaired.

The resemblances between music and speech as maskers in terms of dip listening, stream separation and informational masking suggest that both might present similar challenges to susceptible ASD listeners. However, non-vocal music does not contain intelligible English words, and therefore does not compete for resources at a linguistic level. Whilst this might make music less distracting than speech, autism has also been associated with enhanced pitch perception, as already discussed, and it therefore seems possible that a music masker could, in itself, be more inherently distracting for this reason. There is no published research on speech intelligibility with music maskers in ASD, so a music masker was included in the current study on a purely exploratory basis.

3.1.6 Monotone speech

Two final conditions of our investigation considered aspects of the target speech itself. Fundamental frequency contours play an important role in speech intelligibility, as can be demonstrated by the fact that when the contour is artificially flattened to remove all pitch movement, speech perception in noise becomes less accurate (Binns & Culling, 2007; Laures & Weismer, 1999; Laures & Bunton, 2003; Watson & Schlauch, 2008). One reason for this is thought to be that intonation contours help direct listeners' attention to key content words in a sentence, effectively cueing them in to the stressed words (Cutler & Foss, 1977). Autistic listeners may be less able to interpret the affective information conveyed through vocal intonation (Peppe, McCann, Gibbon, O'Hare, & Rutherford, 2007; Rutherford et al., 2002), but at least some aspects of grammatical prosody appear to be relatively intact (Chevallier et al., 2009). This suggests that flattening the intonation contour of target speech might have the same detrimental effect on speech reception thresholds as for neurotypical listeners. On the other hand, if pitch-based speech stream segregation is in some way impaired in a subset of individuals with ASD, it might be that maintaining one steady fundamental frequency for the target, and a different steady fundamental frequency for the masker, could actually simplify the task for them by removing the need to track dynamic pitch movement.

3.1.7 Degraded target speech in the absence of a masker

Noise-vocoded speech, which was used as a masker in Chapter 2 (Condition 5), has greatly reduced pitch information, as well as losing much of its spectral detail. Even with as few as 4 spectral channels, however, it can be highly intelligible to most neurotypical listeners when presented in quiet (Shannon et al., 1995). We wondered whether smaller differences in SRT between ASD- and controls in the vocoded speech masker condition (mean difference 6.4 dB) than with the natural speech masker (mean difference 8.6 dB) might have been due to the former being less intelligible to the autistic listeners than to controls (rather than because it was easier to separate speech streams in this condition, as hypothesised). To check this, we measured the intelligibility of noise-vocoded speech in quiet, but with fewer than the original 20 channels in order to avoid ceiling effects.

Perhaps more importantly, this condition also gave us the opportunity to investigate whether the effect of reducing the number of spectral channels would be the same for autistic listeners as for controls. To this end we tested speech intelligibility with 8, 6 and 4 spectral channels. Many of the ASD- listeners performed poorly in nearly all types of background noise. Presenting target speech that is itself degraded, therefore, but in the absence of any other auditory signal, was a means to investigate whether speech perception abilities in this group are robust to a reduction in phonetic cue transmission when the auditory scene is less complex.

3.2 Method

3.2.1 New mCRM conditions

Data for this set of 8 mCRM conditions was collected during the same testing session as the 8 conditions described in Chapter 2, so the participants and overall procedure were identical. Of the new conditions described here, some used the main adaptive tracking method to find an SRT (50% correct) and some tested only at fixed SNRs (conditions 15 and 16, see below). The order in which the main 16 mCRM conditions were presented was randomised across listeners (although matched across the control and ASD groups, as described in Chapter 2), but the sets of fixed conditions were kept together so as to avoid confusing participants. The fixed SNR conditions (15 a, b, c and d), therefore, always appeared together, and in order of increasing difficulty, as did the vocoded speech conditions (16 a, b, and c). Unless otherwise specified, all targets were mCRM sentences with the same female talker as for Chapter 2.

The 8 new conditions making up experiment 1.2 were as follows:

Table 3.1 New mCRM conditions (Experiment 1.2)

mCRM condition	**Description**
Condition 9	Music masker
Condition 10	Male target with female mCRM masker
Condition 11	Monotone condition with male mCRM masker
Condition 12	2 simultaneous male mCRM maskers
Condition 13	Female mCRM masker with virtual location 0°
Condition 14	Female mCRM masker with virtual location 90°
Condition 15	Male mCRM masker at 4 fixed signal-to-noise ratios: a) +6 dB SNR b) 0 dB c) -6 dB d) -20 dB
Condition 16	Noise-vocoded mCRM sentences in quiet, with varying number of channels: a) 8 channels b) 6 channels c) 4 channels

Stimuli are described in more detail below.

i) Music masker

The music masker (condition 9) consisted of an extract from a harpsichord melody (Francesco Geminiani's *Allegro Assai* in F Major). This was partly selected because it is in the Western musical tradition, but is relatively obscure so our participants were unlikely to be familiar with it. The chosen extract has a clear melodic line throughout, and a regular fast beat (approximately 115 beats per minute), which meant silent gaps were kept to a minimum. Where pauses longer than 50 ms did occur they were removed (using CoolEdit 96, Syntrillium Corp). For each trial in this condition a randomly-selected excerpt of the appropriate length was presented as a masker. In having a strong pitch contour, this masker was more similar to a speech distractor than either the basic speech-spectrum noise or amplitude-modulated noise, but like the noises it did not directly engage linguistic input processes.

ii) Reversing the gender of the target talker

In condition 10, distractor sentences were from the female talker who was the target in all other mCRM conditions. The targets in this condition were recordings from the same male talker who provided the distractor sentences in condition 6.

iii) Monotone stimuli

Monotone versions of the female target and male mCRM masker sentences were created for condition 11 using Praat (Boersma, 2001). A median pitch for each sentence was measured, and each talker's median pitch across sentences was calculated. The female target's median was 203 Hz, and the male masker's median was 125 Hz. All pitch points in each sentence were then adjusted to match the original talker's median value.

iv) Two-talker maskers

For Condition 12, maskers always consisted of two masking mCRM sentences presented together. One talker was the same male talker as originally used in condition 6, and the second was also a male speaker with a Southern British English accent. The software automatically ensured that none of the masker colour or numbers matched the target, or each other.

v) Spatial separation of target and masker

Head-related transfer functions (HRTFs) were used in Conditions 13 and 14 to give a percept of sound source location in virtual space. Stimuli were still presented through headphones, but the original mCRM sentence recordings were synthesized with the binaural HRTFs to simulate the effects of pinna transfer functions, interaural time difference and interaural level difference which would occur if the speech emanated from a particular location in space. HRTFs were measured in an anechoic chamber on a KEMAR-type mannequin (which simulates the head and torso), by rotating the mannequin relative to the source in the horizontal plane. These HRTFs were then digitally convolved with the original sentences before presentation. In Condition 13, both the target and masker sentences had a virtual location of 0° azimuth (i.e. in front of the listener), and in Condition 14 the target remained at 0° whilst the masker had a virtual location of 90° (i.e. on the listener's right-hand side). The masker in these conditions was the same set of female mCRM distractor sentences as used for condition 7 in Chapter 2.

vi) Testing at fixed SNRs

In Condition 15, listeners completed several blocks of trials with the male mCRM masker. Each block contained 20 trials at a single, unchanging SNR: a) +6 dB, b) 0 dB, c) -6 dB, or d) -20 dB. Listeners continued to receive on-screen feedback about their response accuracy but there was no adaptive tracking procedure: the level of task difficulty varied across blocks rather than within them.

vii) Noise-vocoded speech

A similar testing format as in the previous condition applied in Condition 16, but here there was no masking sound and the noise-vocoded target sentences were presented at a single fixed intensity (approximately 65 dB SPL over 0.1-5.0 kHz). Signal processing followed the same steps as for the vocoded masker sentences described in Chapter 2, based on Shannon et al. (1995). The vocoded maskers had 16 spectral channels, and speech processed with this much spectral information is typically highly intelligible to normally-hearing adults when there is no background noise. To make the vocoded targets in Condition 16 more challenging, therefore, lower numbers of spectral channels were used. There were three blocks of trials, with decreasing numbers of spectral channels across blocks: a) 8 channels (10 trials), b) 6 channels (15 trials), and c) 4 channels (20 trials). Listeners were still expected to perform relatively well with 8 and 6 channel vocoded speech, and ceiling effects were therefore likely. These blocks had to be included because ASD listeners' performance with spectrally-degraded speech could not be reliably predicted a priori, but to save time fewer trials were completed in these conditions.

3.3 Results

3.3.1 ASD subgroups

Participants were allocated to the original ASD subgroups on the basis of their combined performance across the 8 core mCRM conditions and 3 SIN conditions. As demonstrated in Chapter 2, most participants' performance was highly correlated across these two sets of listening tasks. There was one ASD listener, however, whose overall mean z-score for those 11 conditions (1.73) appeared to be highly affected by his SRTs in the SIN conditions (mean z for 8 mCRM conditions = 1.44, mean z for 3 SIN conditions = 2.51).

Figure 3.1 shows listeners' mean z-scores across the grand total of 14 mCRM adaptive conditions (8 conditions described in chapter 2, and 6 from this chapter), and it is clear that when SIN conditions are excluded from the, calculation this listener's mean z-score (1.21) is well below the cut-off for inclusion in the ASD- group (z ≥ 1.64).   Since here we focus only on mCRM conditions, and a split between ASD+ and ASD- seems to occur most naturally between this listener and the next weakest performer (mean z = 1.76), he was recoded to ASD+ for the following analyses.

As Figure 3.2 demonstrates, there was a strong relationship between listeners' mean z-scores from the 8 core mCRM conditions described in chapter 2 and mean z-scores from the 6 conditions described in Chapter 3: Pearson's $r = 0.91$, $p < 0.001$. Most of the conditions discussed in Chapter 3 contained a speech masker, and the ASD- subgroup (now n = 6) continued to perform poorly compared to the other listeners (see Figure 3.3 and Table 3.2)

Figure 3.1 Scatterplot showing mean z-scores for all 14 adaptive mCRM conditions (with subgroups as allocated in chapter 2)

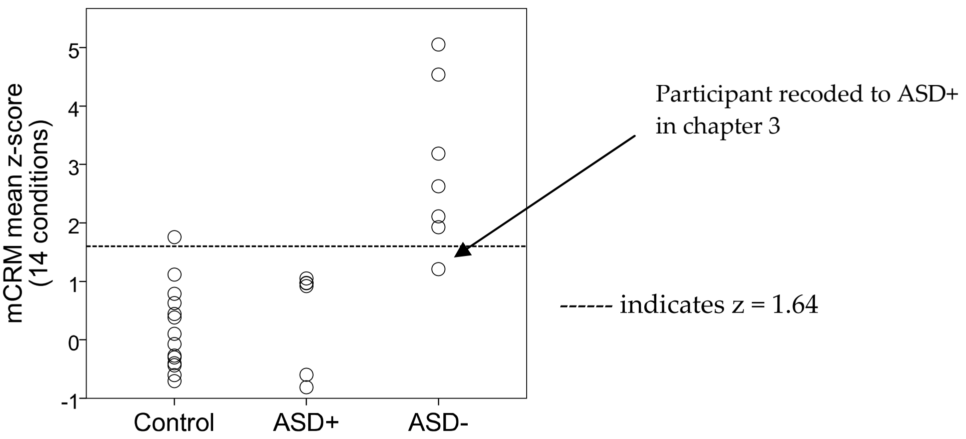


Figure 3.2 Scatterplot showing relationship between performance in 8 core mCRM conditions (Exp 1.1) and 6 new adaptive conditions (Exp 1.2)

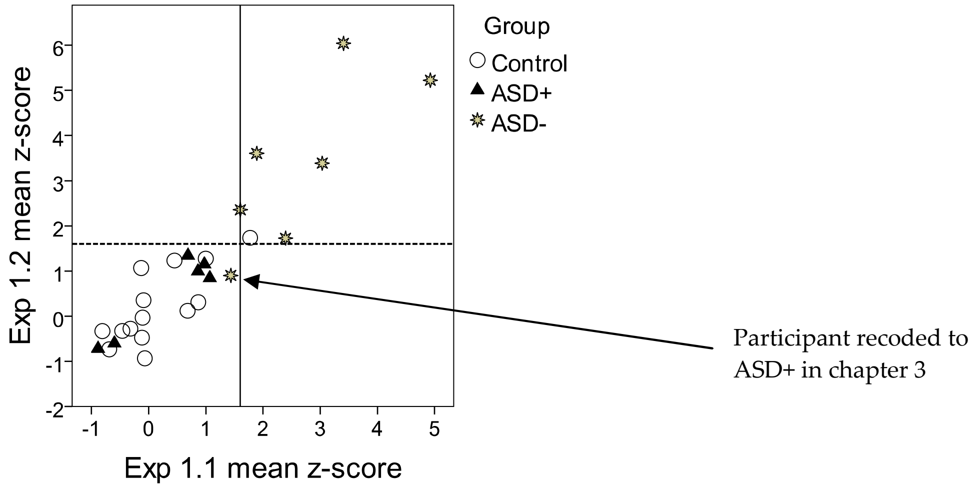
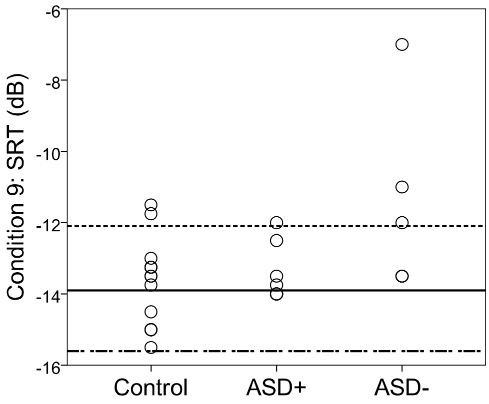
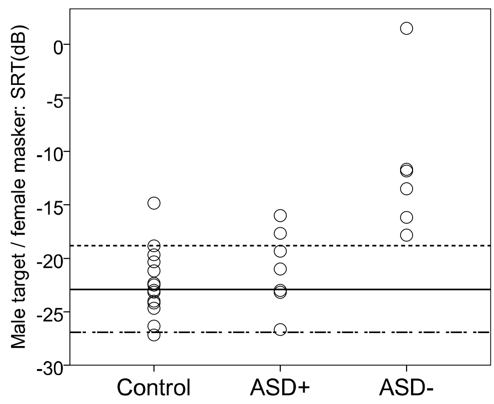


Figure 3.3 Scatterplots showing SRTs: mCRM adaptive conditions (with outliers)

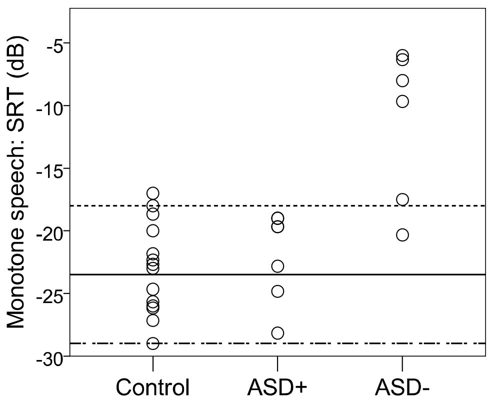
i) Condition 9: Music masker



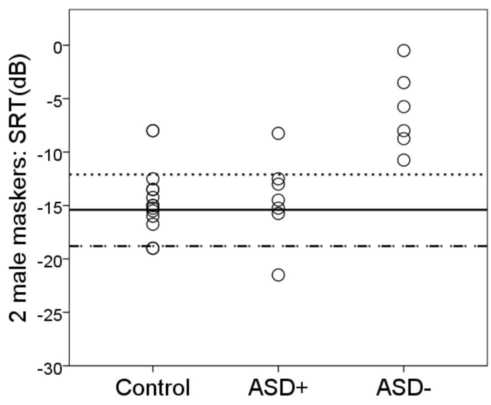
ii) Condition 10: Male target + female masker



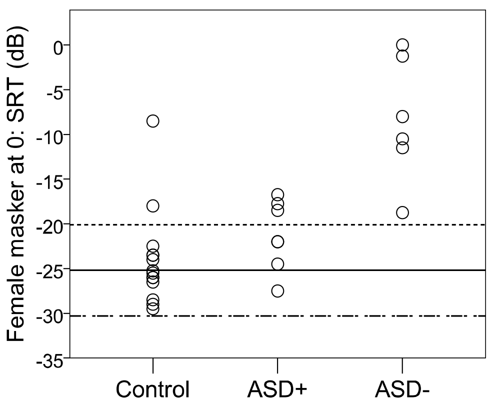
iii) Condition 11: Monotone target and masker



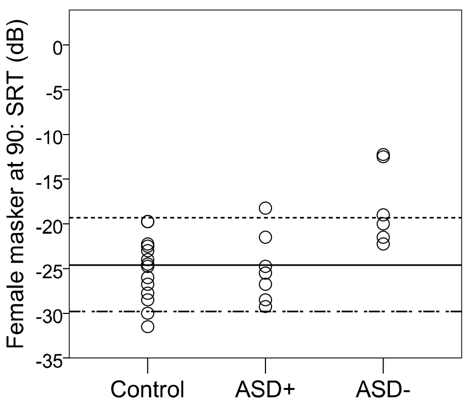
iv) Condition 12: Two male mCRM maskers



v) Condition 13: Target + masker at 0°



vi) Condition 14: Target at 0° + masker at 90°



_____ control mean (excl. outliers)

-------- z = + 1.65 (poor performance)

--- --- --- z = -1.65 (good performance)

Table 3.2 Descriptive statistics: mCRM adaptive conditions

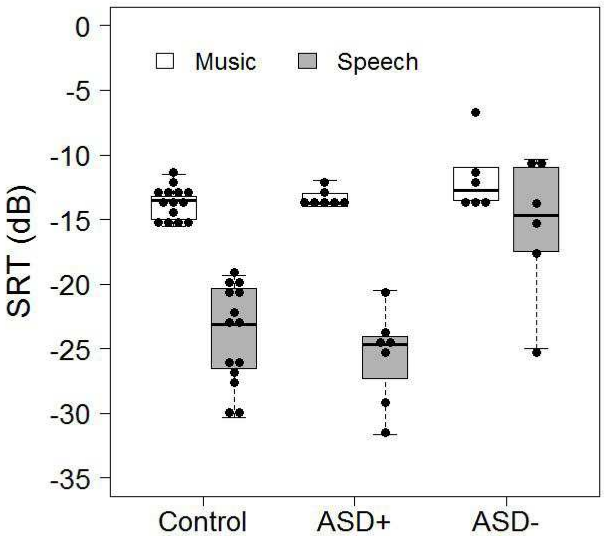
	Control mean SRT (s.d.)	ASD+ mean SRT (s.d.)	ASD- mean SRT (s.d.)
Condition 9 Music masker	-13.7 (1.2)	-13.4 (1.8)	-11.8 (2.5)
Condition 4 Irrelevant male speech	-23.9 (3.8)	-25.7 (3.7)	-15.5 (5.4)
Condition 10 Male target + female mCRM masker	-22.3 (3.2)	-21.0 (3.6)	-11.6 (6.9)
Condition 6 Female target + male mCRM masker	-25.2 (2.6)	-23.0 (2.9)	-16.0 (4.8)
Condition 11 Monotone: male mCRM masker	-23.0 (3.6)	-21.9 (3.5)	-11.3 (6.1)
Condition 12 2 male mCRM maskers	-14.4 (3.3)	-14.4 (4.0)	-6.2 (3.8)
Condition 13 Female mCRM masker 0	-24.0 (5.4)	-21.3 (3.9)	-8.3 (7.0)
Condition 14 Female mCRM masker 90	-25.1 (3.7)	-24.9 (3.9)	-17.9 (4.4)

3.3.2 Music Masker (Condition 9)

As Table 3.2 shows, the control and ASD+ groups' speech reception thresholds tended to be higher, and individual variability in performance lower, with the music masker than with any single-talker speech masker (with music masker, control mean = -13.7, s.d. 1.2). In fact their SRTs most closely resembled those in condition 12 with two masking talkers (control mean = -14.4), which might reflect similar energetic masking demands and dip-listening opportunities across these two conditions. As Figure 3.3 i) shows, half the ASD- group were significant outliers in this condition, which was considerably more than with the steady noise masker in chapter 2, but fewer than with mCRM speech maskers (for comparison see Figure 3.3 iii - v).

Our primary aim was to consider how performance with a music masker compared to that with a single speech masker, since in both these cases there is a single pitch contour but only the speech should directly engage language processes. For this purpose, we chose the irrelevant male speech masker (condition 4) from Chapter 2 as our main point of comparison (see Figure 3.4). A mixed ANOVA (2 conditions x 3 groups) showed significant main effects of condition ($F_{1, 24} = 95.2$, $p < 0.001$) and group ($F_{2, 24} = 13.8$, $p < 0.001$, and a significant group x condition interaction ($F_{2, 24} = 6.8$, $p = 0.005$, partial η^2 0.36). The ASD- group performed significantly worse overall than the control and ASD+ groups (Tukey's HSD, both $p < 0.001$), but there was no significant difference between controls and ASD+ ($p = 0.77$).

Figure 3.4 Boxplots showing SRTs for the Music and Irrelevant male speech maskers (with individual data points)



All three groups had higher SRTs with the music masker than with the male speech masker (see Table 3.2), but whereas the difference in mean SRT between these conditions was more than 10 dB for controls and ASD+ listeners, it was less than 4 dB for ASD-. Furthermore, the group effect in a univariate ANOVA, taking music masker SRTs as the dependent variable, only just reached significance ($F_{2, 24} = 8.6$, $p = 0.045$, partial $\eta^2 = 0.23$). Tukey's HSD showed a significant difference between ASD- and controls (mean difference in SRT 1.95 dB, $p = 0.04$), but not between ASD- and ASD+ (mean difference 1.64 dB, $p = 0.15$). The ASD-group's mean SRT appeared to be strongly affected by a single outlier whose SRT was 4 dB worse than that of any other listener. When this outlier was excluded from the analysis, the mean difference between controls and ASD- was less than 1dB, and the univariate ANOVA showed no significant effect of group ($p = 0.24$). By comparison, as already described in Chapter 2, the group effect in a univariate ANOVA for the male speech masker was highly significant ($p < 0.001$), and ASD- performed significantly differently from both other groups (Tukey's HSD, both $p < 0.001$). All but one of the ASD- listeners had a z-score > 1.65. The mean differences in SRT with the speech masker were also substantially larger than with the music masker: 10 dB between ASD- and controls.

Figure 3.4 shows that the ASD+ listeners' SRTs with the music masker tended to be slightly higher (worse) than the control mean (although still within the normal range), which is similar to the way they performed with an amplitude-modulated noise masker (see Chapter 2, Figure 2.4. ii). Unlike in the amplitude-modulated noise condition, however, half of the ASD- listeners actually performed within the normal range with music. All had been significant outliers with AM noise.

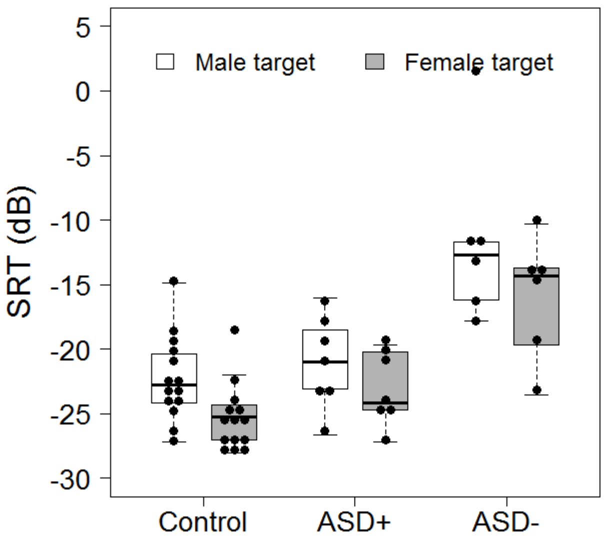
These results suggest that if enhanced pitch perception exists in ASD, it does not lead to increased interference/distraction effects from all maskers with strong pitch contour, and in fact it may even aid intelligibility amongst the lower-performing listeners at least where non-speech maskers are concerned. Whatever allows control and ASD+ listeners to perform better with the speech than the music masker does not seem to work in the same way for ASD- listeners.

3.3.3 Gender of the target talker (Conditions 10 and 6)

The gender of the target talker did appear to affect performance (see Figure 3.5). A mixed ANOVA (2 conditions x 3 groups) showed a significant main effect of condition (male target with female masker or female target with male masker), $F_{1, 24} = 13.9$, $p = 0.001$. All three groups had slightly lower SRTs with the female target and male masker, and mean differences between the 2 conditions ranged from 2 - 4 dB. The standard configuration throughout the experiment was to have a female target, so better SRTs may simply reflect greater familiarity with this talker. As expected, the group effect was also significant ($F_{2, 24} = 20.2$, $p < 0.001$), with ASD- performing worse than both other groups across the two conditions (Tukey's HSD, both $p < 0.001$). There was, however, no significant group x condition interaction ($p = 0.57$). This indicates that target gender had the same effect on all three groups, and is not likely to have been a confounding factor in the remainder of the results.

Figure 3.5 Boxplots showing the effect of target gender on SRT

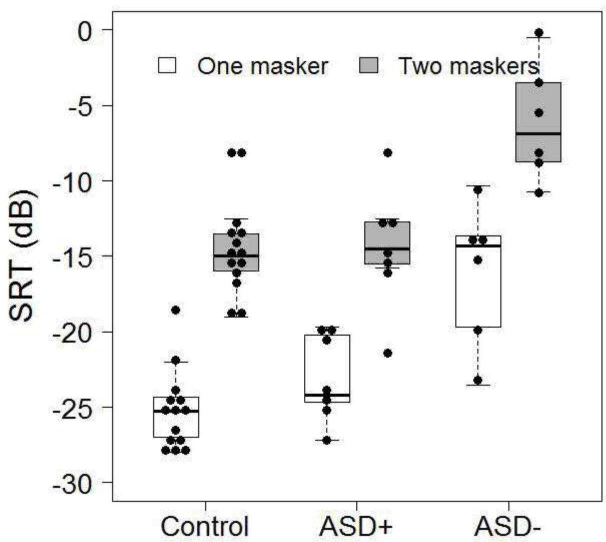
(with opposite-gender speech masker in both cases)



3.3.4 Number of masking talkers (Conditions 12 and 6)

A mixed ANOVA analysing performance across two conditions (one male mCRM masker or two male mCRM maskers) showed significant main effects of condition (F 1, 24 = 166.0, $p < 0.001$) and group (F 2, 24 = 19.8, $p < 0.001$), but no significant interaction ($p = 0.42$). Overall, ASD- performed significantly worse than both controls and ASD+ (Tukey's HSD, both $p < 0.001$), and all groups performed worse with the two-talker masker than with the single-talker (see Figure 3.6; mean differences ranged from around 9-11 dB).

Figure 3.6 Boxplots showing the effect of a second masking talker on SRT



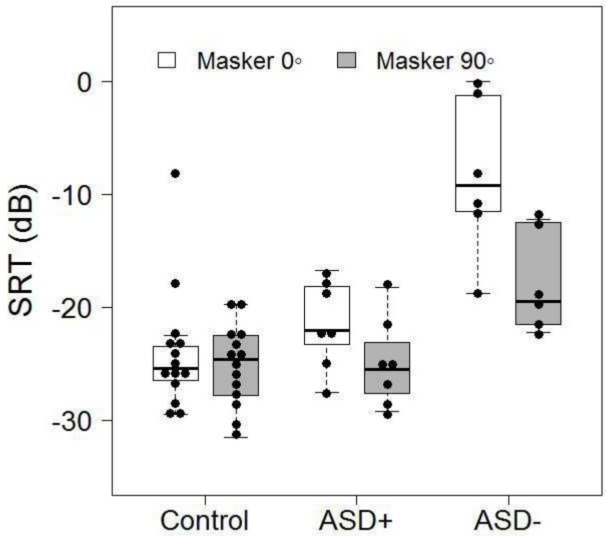
Since the difference between conditions was so similar across groups, it is tempting to attribute the effect primarily to increased energetic masking and reduced glimpsing, rather than to increased informational masking from the additional competing talker. This is impossible to prove with the current data. What is clear, though, is that the size of the ASD- group's deficit with the two-talker masker (mean difference ASD- and controls = 8 dB) does not increase beyond what was observed with a single masking talker (8.6 dB). Whatever raises (worsens) SRTs when a second speech masker is added, therefore, seems to affect ASD and control listeners in the same way.

3.3.5 Spatial release from masking (Conditions 13 and 14)

A mixed (2 x 3) ANOVA showed highly significant main effects of condition (female masker at 0° or at 90°; $F_{1, 24} = 25.6$, $p < 0.001$) and group ($F_{2, 24} = 16.8$, $p < 0.001$), and a significant group x condition interaction ($F_{2, 24} = 7.3$, $p = 0.003$). The ASD- listeners performed significantly worse overall than both other groups (Tukey's HSD both $p \leq 0.001$).

Figure 3.7 Boxplots showing the effect of perceived target-masker spatial separation

(In both conditions the target was at 0° azimuth)



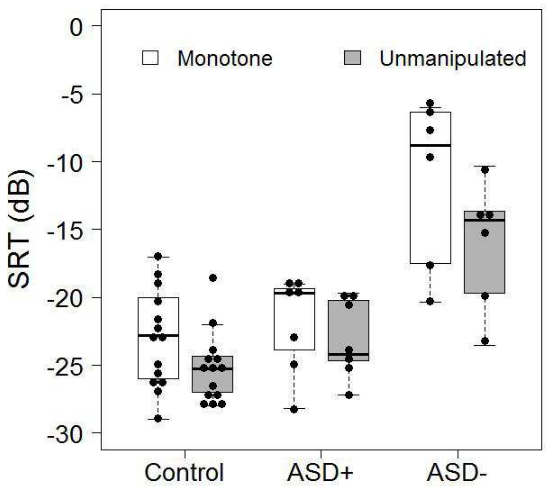
Post-hoc t-tests exploring the effect of condition for each group indicated that both the ASD+ and ASD- groups experienced significant spatial release from masking (ASD+ $t(6) = 3.31$ and ASD- $t(5) = 3.41$, both $p = 0.02$), but that controls showed no effect of condition ($p = 0.35$) (see Figure 3.7). The mean difference between conditions was 3.6 dB for ASD+, and 9.6 dB for

ASD-. What is perhaps most surprising about these results is that on average the control listeners improved by only 1 dB with the spatial separation. It seems probable that this arose from a ceiling effect, since at very low SNRs the target must eventually become inaudible. Certainly the two control participants who were outliers when target and masker were collocated, and therefore had room to improve with the masker at 90° before hitting a ceiling, did improve substantially in this second condition.

3.3.6 Monotone target and masker (Conditions 11 and 6)

A mixed ANOVA (2 conditions x 3 groups, including the un-manipulated female target + male masker condition for comparison) showed significant main effects of condition ($F_{1, 24} = 24.1$, $p < 0.001$) and group ($F_{2, 24} = 18.7$, $p < 0.001$), but an interaction which only approached significance ($p = 0.06$). Tukey's HSD again indicated that ASD- listeners' SRTs were higher than both other groups' (both $p < 0.001$).

Figure 3.8 Boxplots showing the effect of flattening fundamental frequency contours (for female target and male mCRM masker)



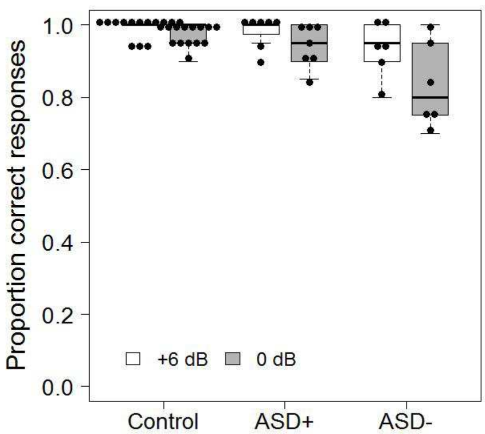
Control performance was rather more variable with the monotone speech (see Figure 3.8), presumably because the unnatural percept was distracting to some listeners. Separate post-hoc t-tests for each group showed that both the control and ASD- groups' SRTs were significantly higher with monotone speech than the original recordings, but that ASD+ were not significantly affected by this manipulation ($p = 0.41$). The mean difference between conditions was 2.2 dB for controls ($t(13) = 4.18$, $p = 0.001$) and 4.7 dB for ASD- ($t(5) = 3.58$, $p = 0.02$).

3.3.7 Performance patterns with target and masker at fixed SNRs (Condition 15)

To establish whether performance with the fixed SNR procedure (entire blocks of trials at a given SNR) followed broadly the same pattern as with the adaptive procedure, the proportion of error responses made by each listener at each SNR (+6, 0, -6, and -20 dB) was calculated (see Figure 3.9 and Table 3.3).

Figure 3.9 Boxplots showing mCRM performance at fixed SNRs, with male masker (including individual data points)

i) + 6 dB and 0 dB



ii) -6 dB and -20 dB

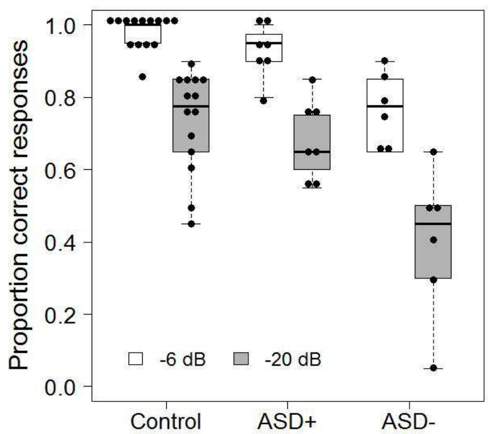


Table 3.3 Performance (proportion correct) for each group at each SNR (dB)

	Group	Mean proportion correct	Std. Deviation	Total N errors	Total N trials
SNR +6	Control	.99	.02	3	280
	ASD+	.98	.04	3	260
	ASD-	.93	.08	8	260
SNR 0	Control	.98	.03	7	280
	ASD+	.94	.06	8	260
	ASD-	.83	.12	20	260
SNR -6	Control	.98	.04	7	280
	ASD+	.93	.07	10	260
	ASD-	.77	.10	28	260
SNR -20	Control	.74	.14	74	280
	ASD+	.68	.11	45	260
	ASD-	.40	.21	72	260

A mixed ANOVA (4 conditions x 3 groups) showed significant main effects of condition ($F_{3, 72} = 85.2$, $p < 0.001$, with Huynh-Feldt correction) and group ($F_{2, 21} = 26.1$, $p < 0.001$), and a significant interaction ($F_{6, 72} = 4.1$, $p = 0.009$). ASD- performed significantly worse overall than the other two groups (Tukey's HSD, both $p < 0.001$).

The group x condition interaction was explored using separate univariate ANOVAs to compare across groups at each SNR (see Table 3.3). Even at +6 dB SNR, the ASD- group made more errors than controls, but the group difference did not quite reach significance after an alpha correction ($F_{2, 24} = 3.68$, $p = 0.04$, corrected $\alpha = 0.01$), and Tukey's HSD indicated that whilst the difference between ASD- and Controls was approaching significance($p = 0.03$), there was no difference between ASD- and ASD+ ($p = 0.16$). At the three remaining SNRS, the group effects were all highly significant (all $p \leq 0.001$), and ASD- performed significantly worse than both other groups whereas there was no difference between ASD+ and controls.

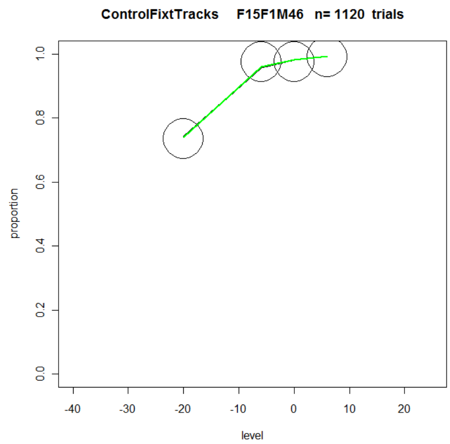
These results seem broadly in line with the highly significant deficit in mean SRT that the ASD- group showed with the adaptive testing procedure for this same combination of female target and male masker. It is striking, however, that even at +6 dB SNR, when the target should have been highly audible, the ASD- group tended to make more errors than controls. Closer examination reveals that 4 of the 6 ASD- participants made one or more errors at this SNR, but half the total number of errors here (4/8) were made by a single participant. This individual also made more errors overall than any other participant across the fixed SNR blocks of trials (30 errors out of 80 trials, summed across the four SNRs), and was by far the weakest performer across the 14 conditions of the mCRM which were tested adaptively (mean z-score = 5.05). The fact that this participant made a substantial number of errors even when the target was considerably more intense than the masker (and well above his SRT for this condition when tested adaptively: - 10 dB), suggests that his poor performance may have involved a more general deficit in attentional control or task motivation.

3.3.8 Psychometric functions

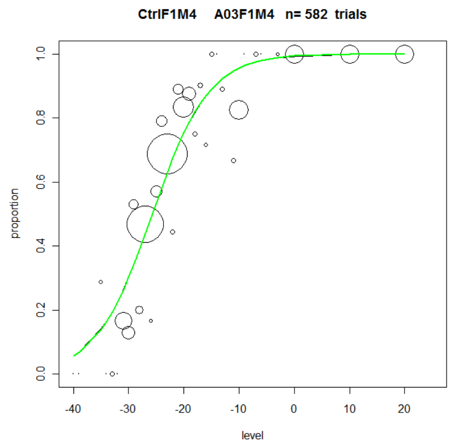
Figure 3.10 shows aggregated group psychometric functions from the fixed SNR blocks of trials and from the adaptive blocks. Allowing for the lower number of SNRs included in the fixed procedure, and the smaller number of trials per SNR in the adaptive procedure, the pairs of functions for each group look very similar. There is nothing to suggest that the large difference in mean SRT between the ASD- and control groups which we found using the adaptive procedure resulted from measurement error. In fact, as Figure 3.11 (based on the adaptive data) shows, and in agreement with the statistical analysis of the fixed data described in the preceding section, the ASD- group appear to perform more poorly than controls at almost every SNR tested. There is also no real sign of non-monotonicity in the function for any group, and therefore no reason to suppose the assumptions for adaptive testing were not met.

Figure 3.10 Group aggregated psychometric functions with male masker

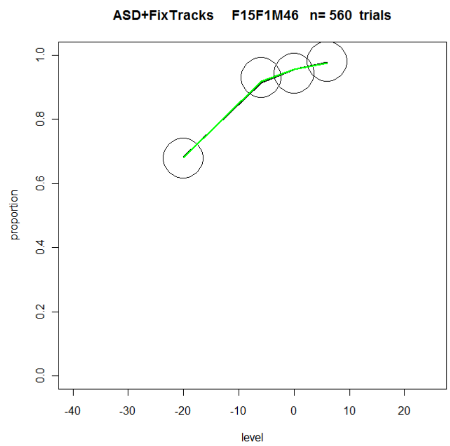
i) Control: fixed procedure



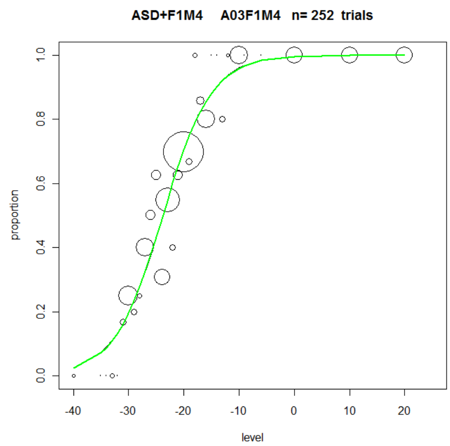
ii) Control: adaptive procedure



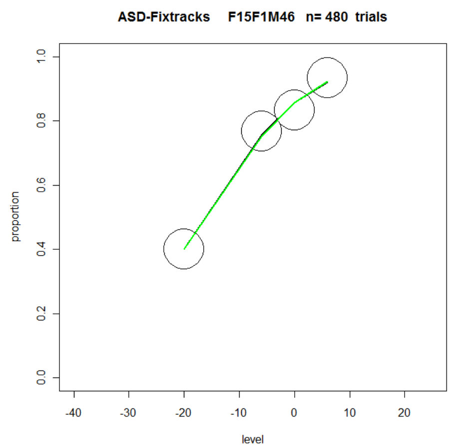
iii) ASD+ : fixed procedure



iv) ASD+ : adaptive procedure



v) ASD- : fixed procedure



vi) ASD- : adaptive procedure

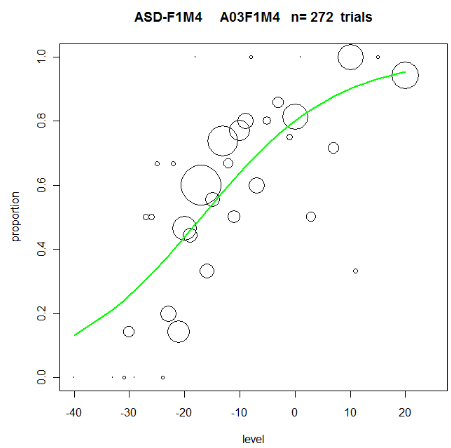
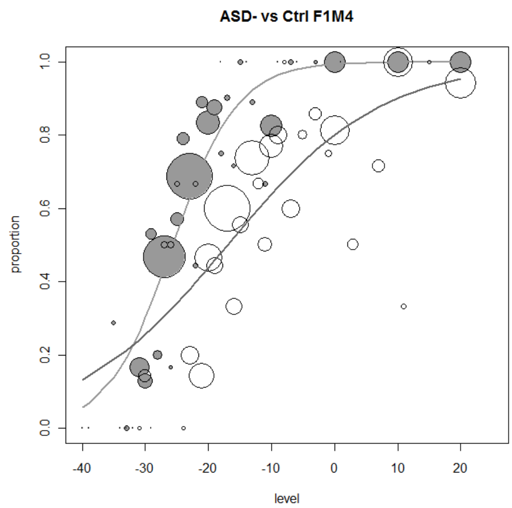


Figure 3.11 Comparison of Control and ASD- psychometric functions (male masker)

(from adaptive procedure)



3.3.9 Error Analysis

Trials from the fixed procedure which contained an error were extracted, and split into two groups: those with colour errors and those with digit errors. (Trials which involved both colour and number errors were included once in each group.) For each group and SNR, the masker colours presented on each trial were then cross-tabulated with listeners' colour responses, and the same for digits. This allowed us to calculate the number of intrusion errors – trials on which the response colour or digit matched the masker. As can be seen in Table 3.4, all three groups tended to make more errors involving colours than digits.

An exact binomial test was used to compute whether the intrusion rate for each group at each SNR was significantly above the level expected by chance (0.20 for colour errors, and 0.14 for digits). At +6 dB the total numbers of errors for colours and for digits made by each group was too low to be particularly informative, and this difficulty affected the digit error data even at lower SNRs as well. When *colour* errors occurred at 0 dB SNR, however, all groups showed masker intrusion rates significantly higher than chance, and this was also the case at - 6 dB, except for the control group (alone) where the rate was lower and did not meet significance. This is broadly in line with Brungart's (2001) data, and suggests that - for our control group, at least - intrusion errors were most likely to occur when no intensity cues were available to aid target/masker stream segregation.

Table 3.4 Colour and digit errors: proportion of masker intrusions

	COLOUR	ERRORS			DIGIT	ERRORS	
SNR	+6	+6	+6		+6	+6	+6
Group	Control	ASD+	ASD-		Control	ASD+	ASD-
N Intrusions	2	2	2		1	2	0
N errors	2	2	8		1	2	1
% Intrusions	1*	1*	0.25		1	1*	0
SNR	0	0	0		0	0	0
	Control	ASD+	ASD-		Control	ASD+	ASD-
N Intrusions	6	4	9		1	1	10
N errors	6	5	14		2	4	12
% Intrusions	1**	0.8*	0.64**		0.5	0.25	0.83**
SNR	-6	-6	-6		-6	-6	-6
	Control	ASD+	ASD-		Control	ASD+	ASD-
N Intrusions	2	5	20		2	2	5
N errors	6	7	25		3	3	7
% Intrusions	0.33	0.71*	0.80**		0.67	0.67	0.71**
SNR	-20	-20	-20		-20	-20	-20
	Control	ASD+	ASD-		Control	ASD+	ASD-
N Intrusions	20	8	17		12	7	20
N errors	55	33	60		30	17	44
% Intrusions	0.36**	0.24	0.28		0.40**	0.41**	0.45**

* $p < 0.05$ ** $p < 0.001$

In terms of colour intrusion rate, the only SNR at which the three groups looked as though they might differ from each other was -6 dB, where both the ASD groups continued to make significant numbers of intrusion errors and controls did not. Given the theoretical importance of the shift between the 0 dB SNR condition (maximum informational masking) and the -6 dB condition (intensity cues to reduce informational masking, with minimum loss of target audibility), one hypothesis was that the effects of SNR might vary across our groups. Whilst the results of separate logistic regression analyses for colour intrusion rates at each SNR, showed no significant effect of group (Control or ASD-) even for the -6 dB condition ($p = 0.096$), a logistic regression for colour intrusion rate using both group and SNR (0 dB or -6 dB) as categorical predictors found the interaction term to be highly significant ($p = 0.005$). This result is particularly striking in light of the lack of power in this error analysis, and adds weight to the idea that ASD- listeners may be impaired at speech stream segregation.

As mentioned, lack of power was an even more pervasive problem across SNRs when it came to digit errors. In line with the colour data, though, the ASD- group made significantly more masker intrusion errors than expected by chance at -6 dB, which again suggests that they may not have been using intensity cues to separate speech streams as efficiently as the other listeners. All three groups made substantial numbers of digit errors at -20 dB, and here the intrusion rate for all three groups was significantly higher than chance (all $p < 0.002$).

3.3.10 Noise-vocoded speech with varying numbers of channels (Condition 16)

In order to investigate how groups performed with noise-vocoded speech which varied in the number of spectral channels it contained (8, 6 or 4 channels), listeners' performance (proportion correct) was calculated for each condition (see Figure 3.12 and Table 3.5).

A mixed ANOVA (3 groups x 3 vocoding conditions) showed significant main effects of condition ($F_{2, 48} = 137.1$, $p < 0.001$, with Huynh-Feldt correction) and group ($F_{2, 24} = 4.7$, $p = 0.02$). The group x condition interaction was also significant ($F_{2, 24} = 2.90$, $p = 0.04$). As Figure 3.12 shows, all three groups showed a decrease in accuracy with decreasing numbers of speech channels. Univariate ANOVAs exploring group performance for each vocoding condition separately showed no group differences for 8 channel (p = 0.83) or 6 channel (p = 0.63) vocoded speech, but a significant difference with 4 channels (p = 0.006). In the 4 channel condition, ASD- performed significantly worse than ASD+ (p = 0.006) and the difference from controls also approached significance (p = 0.02, corrected α 0.01).

Figure 3.12 Boxplots showing response accuracy for noise-vocoded targets (no masker)

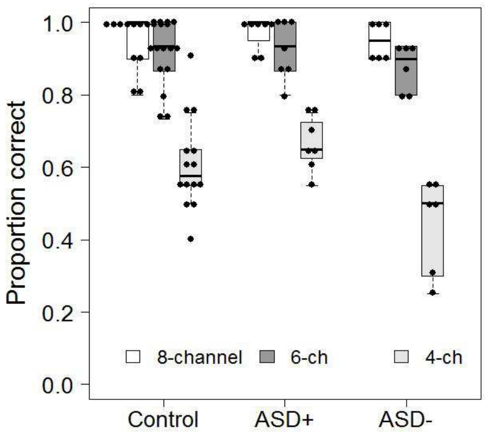


Table 3.5 Group performance (proportion correct) for noise-vocoded speech

	Group	Mean	Std. Deviation
8 channel vocoding	Control	.96	.08
	ASD+	.97	.05
	ASD-	.95	.05
6 channel vocoding	Control	.90	.09
	ASD+	.92	.08
	ASD-	.88	.07
4 channel vocoding	Control	.61	.13
	ASD+	.66	.07
	ASD-	.44	.13

At first sight, these results seem to suggest that - in addition to being highly susceptible to interference from background speech - some autistic listeners have speech processing networks which are less robust to other forms of signal degradation as well. Again, however, caution is warranted, as numbers are small and one of the two most extreme ASD- outliers with 4-channel noise-vocoding was the individual who had performed very poorly in other conditions as well. When this listener was excluded there was no significant group effect or interaction. Overall, therefore, whilst this study presents relatively strong evidence of a deficit amongst a subset of ASD listeners when there are speech *maskers*, evidence that speech processing is impaired at a more basic level is more doubtful. With a serious deficit in either auditory temporal processing on the one hand, or phonological representations on the other, larger and more widespread impairments with spectrally-degraded speech might have been expected.

3.4 Discussion

The results of these new mCRM conditions support the hypothesis that a subset of high-functioning adults with ASD are more susceptible than neurotypical listeners to informational masking. The poor SRTs of the ASD- group with intelligible speech maskers, as described in Chapter 2, do not appear to be due to any specific characteristics of the female target speaker used in most of the conditions. These listeners performed equally poorly - on average 10 dB worse than controls - regardless of whether they heard a female target talker with a male masker, or a male target talker with a female masker. They also performed significantly worse than controls at a range of SNRs when the method of constant stimuli was used (0 dB, -6 dB and -20 dB), rather than the adaptive tracking procedure which had been employed throughout the rest of this study. At -20 dB, for example, the control listeners were still giving almost 75% correct responses, whereas the ASD- group were scoring only 40% correct. This is well in line with the fact that the ASD- group's mean SRT when measured adaptively (tracking 50% correct) in the equivalent male mCRM masker condition was -16 dB. It does not, therefore, appear as though the adaptive testing procedure was in itself a confounding factor. These listeners appear to suffer excessive interference from opposite-gender masking speech, regardless of the measurement procedure used.

A high incidence of masker intrusions in the CRM can also be interpreted as a sign of informational masking - particularly at relatively favourable SNRs - and it had been hoped that using the method of constant stimuli would have enabled us to compare intrusion error rates across groups. Unfortunately, however, the low numbers of listeners involved, the low numbers of trials per listener (owing to time constraints), and the overall low numbers of errors made by control and ASD+ listeners meant that statistical power was insufficient for a truly meaningful comparison between groups, particularly for digits.

At low SNRs, where the masker is considerably more audible than the target, intrusion errors are hardly surprising. Brungart (2001) found, however, that substantial numbers of intrusion errors occurred across the entire range of SNRs he tested, and even for an opposite-gender masker. Although he did not report any statistical analysis of the error data in this paper, he made the observation that masker intrusions occurred more often than expected by chance. Our control data for colour errors seemed to show a very similar pattern of high intrusion rates at +6 dB and 0 dB SNR. However, at -6 dB and -20 dB masker intrusions occurred in only around one third of control listeners' error responses. Brungart's error graphs indicate at least 75% intrusion errors at -6 and -12 dB (which was the lowest SNR he tested). Although our control participants made only 6 errors in total at -6 dB, they made 55 errors at -20 dB, which is the level at which intrusions are most likely. It is surprising, therefore, that only 36% of these were intrusions. This proportion was significantly higher than chance, but still much lower than expected from Brungart's data. The most probable reason for this is that Brungart's CRM loads rather more heavily on informational masking than our modified version, because he varied the target talker from trial to trial. Stimulus uncertainty is a well-known factor in informational masking (Durlach, Mason, Kidd, Arbogast, Colburn, & Shinn-Cunningham, 2003;Freyman, Helfer, & Balakrishnan, 2007). It seems likely that with our slightly less challenging CRM task, any intensity difference between target and masker - even an unfavourable one in terms of overall SNR - enabled our control listeners to stream out the masking speech with relative ease.

By comparison, however, although between-group differences within each SNR condition did not reach significance, the ASD- group did appear to respond quite differently from controls to changes in SNR across the mid-range. Whereas control listeners made the same overall number of colour errors at -6 dB as at 0 dB, their rate of intrusions went down (2/6 at -6 dB, compared to 6/6 at 0 dB). The ASD- listeners, by contrast, showed a big leap in the overall number of errors (from 14 at 0 dB to 25 at -6 dB), as well as an increase in the rate of intrusions, from 64% intrusions at 0 dB to 80% at -6 dB. The limited amount of data here makes it difficult to draw any hard conclusions, but this pattern of results is certainly not inconsistent with the idea that the ASD- listeners were less successful at using intensity differences between target and masker at this moderate (-6 dB) SNR to separate the two speech streams. At -20 dB the intrusion rates appeared far more similar across groups, perhaps because this SNR provided a sufficiently robust streaming cue to support the ASD- listeners.

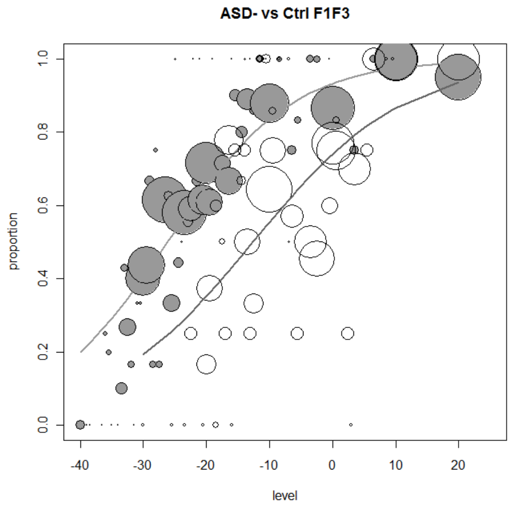
A basic assumption of the adaptive tracking procedure for measuring SRTs is that the underlying relationship between SNR and performance is monotonic. Brungart (2001), however, reported this not to be the case, particularly for same-gender and same-talker speech maskers. In these cases, the psychometric function changes steeply from 10 dB to 0 dB SNR, and then appears to plateau, with little change in performance at SNRs between 0 and -10 dB. Similar results have also been reported by other groups for a same-talker masker, although these studies found a slight dip in performance at around 0 dB, with accuracy actually improving again at slightly lower SNRs (Dirks & Bower, 1969; Egan, Carterette, & Thwing, 1954). Even for the opposite-gender talker, Brungart found a psychometric function which - whilst not exactly non-monotonic - had an extremely shallow slope across the entire range of SNRs he measured, and looked nothing like the functions for the non-speech maskers. Performance dropped from 100% at +15 dB only to around 80% at -12 dB.

The adaptive procedure used in the mCRM for the current study would have been inappropriate in any conditions where the underlying psychometric function was similarly non-monotonic, so, as well as facilitating an error analysis, the fixed SNR data provided an opportunity to consider this factor. The main benefit of using an adaptive procedure with large initial step sizes is that it rapidly moves listeners away from the level where performance is relatively easy and focuses a larger number of trials on the stimulus levels in close proximity to the estimated SRT. Unfortunately, this meant that in the adaptive conditions of the current study, control listeners completed relatively few trials at SNRs around the level of Brungart's reported plateau. Our blocks of fixed SNR trials compensated for this, and it is clear from Figure 3.10 i) and ii) that there was no plateau or dip in the control psychometric function. If anything, it looks more monotonic than suggested by Brungart, because rather than a very shallow slope in performance between

+10 and -12 dB SNR, our control listeners performed almost perfectly from +10 to -10 dB with a relatively steep drop in performance with SNR below that. Again, this difference between the current data set and that of Brungart may reflect the lower informational masking load of the modified CRM, or a number of other differences between the tasks (e.g. how levels are controlled, and the specific properties of the talkers involved).

There appeared to be slightly more of a shallow slope in the ASD+ fixed procedure data than that of controls (although not in the adaptive data -see Figure 3.10, iii and iv), suggesting that perhaps these listeners may have been marginally less adept at using intensity cues for streaming at mid-range SNRs. This occurred well above their 50% threshold level, however, and evidently did not affect their overall SRTs in the adaptive procedure since those did not differ significantly from controls'. Meanwhile, it is actually the ASD- group's psychometric functions which perhaps bear the closest resemblance to Brungart's data for an opposite-gender talker, because there is much a clearer overall slope in performance going from +6 to -10 dB SNR (see Figure 3.10 v and vi). This tends to suggest, yet again, that this subgroup may be unusually susceptible to informational masking. The key point, however, is that this shallow - but still relatively monotonic - slope is not inconsistent with the use of an adaptive procedure. There is no hint of non-monotonicity in the functions from either test procedure. Time did not allow for a fixed-SNR test procedure to be carried out with the female masker condition as well as the male, but to check that that this also met the assumption of monotonicity, group aggregated psychometric functions were also calculated from the adaptive procedure data (see Figure 3.13). The data here - especially for the ASD- group - is rather more noisy, but again it is evident that they performed less accurately than controls at all SNRs below 10 dB.

Figure 3.13 Comparison of Control and ASD- group aggregated psychometric functions (female masker; adaptive procedure)



Many of the conditions described in this chapter were designed to expand on the findings of Chapter 2, by exploring performance with a single opposite-gender masking talker in greater depth. The evidence from error analyses and psychometric functions in the male masking condition, taken in conjunction with significantly raised SRTs, is highly suggestive of a deficit in auditory stream separation in the ASD- group. Furthermore, the results of our spatialised listening conditions also tend to lend weight to this hypothesis. In a same-gender masking condition - which involves increased informational masking, according to Brungart, and in which our ASD- listeners showed an even greater deficit relative to controls - the introduction of a strong (virtual) spatial cue to stream separation led to substantial improvements in their performance. Control listeners were already performing so well with this same-gender masker that their SRTs (mean -24 dB) appeared to be approaching a level where they were probably limited purely by the audibility of the target. This group showed no significant improvement with the introduction of a 90° virtual separation between target and masker. ASD- listeners, on the other hand, improved by almost 10 dB when the masker was perceptually shifted to the right, although their mean SRT remained significantly worse than that of controls. It is impossible to be sure, of course, whether this effect is due predominantly to release from energetic masking owing to improved SNRs at the better ear, or whether improved auditory stream separation was also involved. The latter, however, does seem a strong possibility.

Whilst all six ASD- listeners were significant outliers when target and masker were collocated, half the group performed within the normal range when target and masker were spatially-separated (see Figure 3.3, v and vi). Of the 14 masking conditions tested in this study, there were only 3 others in which half or more of the ASD- group performed within the normal range: these were the steady noise, the music and the Japanese masker, all of which were conditions presumed to involve only minimal informational masking. Furthermore the ASD- listeners' mean SRT with the spatially-separated female masker (-18 dB) was actually very close to their mean SRT with the standard (diotic) irrelevant male masker reported in chapter 2 (-17 dB). Segregating two collocated female speech streams, as our control listeners did with relative ease, relies heavily on the processing of talker characteristics (i.e. cues specific to a social stimulus), whereas segregating two spatially-separated female speech streams involves the use acoustic cues (intensity and timing differences) which are equally relevant to non-social stimuli. Spatial separation between target and masker may therefore have allowed ASD- listeners to overcome the additional impairment (6.5 dB) which they showed with a same-gender over an opposite-gender masker. This was seemingly not sufficient, however, to compensate for their overall deficit with any type of intelligible speech masker; hence SRTs with a spatially-separated female masker closely resemble those with a collocated male masker.

An additional point of interest regarding the spatial listening conditions is that there is some evidence suggesting that auditory spatial attention may be impaired in autism (Teder-Salejarvi, Pierce, Courchesne, & Hillyard, 2005). There was nothing to support this notion in the current study. Shifting the masker by 90° should have relatively strong effects both on SNR at the better ear and on binaural cues, however, so perhaps a deficit in auditory spatial attention would have only have become evident with a smaller target-masker separation.

The final parts of this study considered the importance of pitch and spectral detail in speech stimuli, and how this might affect intelligibility for listeners with ASD. One hypothesis, which was not borne out by the data, was that music - being similar to speech in terms of complex hierarchical organisation, and involving highly salient pitch information - might be equally distracting as a masker to ASD listeners. In fact, however, there was very little difference between groups' SRTs in the music-masking condition, and nothing like the large deficit ASD- listeners showed with a speech masker. This seems to suggest that it is not pitch movement in a masker, per se, which is distracting to this group. The ASD- listeners' performance, in terms of SRTs, was actually far more similar across the speech and music maskers than the other groups', with a difference in mean SRT of only 3.5 dB between conditions for ASD- (as compared to improvements in excess of 10 dB with the speech masker over the music for both other groups). This implies that rather than having a deficit in resisting interference from any masker involving pitch, the deficit involves some other aspect of speech as a masker.

The idea that fundamental frequency movement in the speech stimuli might in some way be distracting to listeners with ASD was also explored by flattening the intonation contour of both target and masker stimuli in the opposite-gender masking condition. Again, the results did not bear out this hypothesis. Rather than improving intelligibility for the ASD-participants, flattening the intonation contours resulted in slightly worse SRTs for all three groups and the ASD- group continued to show the same deficit in SRT relative to the other listeners. Binns and Culling (2007) have demonstrated that in conditions similar to this, the reduction in performance with flattened intonation contours is mainly due to loss of target intelligibility rather than to properties of the masker. It is interesting, therefore, that the interaction between group and condition in the current study was approaching significance. The increase in mean SRT with flattened contours was only 2.5 dB greater for the ASD-group than controls, but still this difference could be interpreted as implying that these listeners may actually rely even more than controls on the intonation contour for target intelligibility, at least in the presence of masking speech.

The only conditions in this study which did not involve any type of masker, measured intelligibility for vocoded speech targets with varying numbers of spectral channels in quiet. With the low numbers of channels used here (8, 6 and 4), not only is voice pitch information almost completely lost, but spectral detail is smeared and listeners must rely heavily on temporal envelope cues. Given the impaired performance of ASD- listeners when speech information is degraded by the presence of other sounds, and the idea outlined above that they may be slightly more vulnerable to the loss of intonation contours than other listeners, it seemed likely that performance with noise-vocoded speech might also be poor. In fact, however, the only significant difference between groups was with the 4-channel targets, and this appeared to be due to only two ASD- listeners. Overall, therefore, speech processing in the absence of masking sounds appears relatively normal, even when that speech signal is severely degraded, and even for most of the weaker-performing ASD listeners. This suggests, again, that it may be the processing of simultaneous speech streams, rather than loss of acoustic/phonetic detail from target speech itself, which is a key factor in everyday speech processing difficulties, at least for some autistic adults.

Chapter 4 Fluctuating maskers, pitch and selective attention

4.1 Introduction

4.1.1. Variable performance across different types of fluctuating masker

One of the more thought-provoking findings of Experiment 1 was that ASD- listeners, in spite of performing extremely poorly across many of the mCRM masker conditions, appeared to experience far less interference from the Japanese speech and the music masker than from English speech maskers. On average, their performance with these unintelligible maskers was much more akin to that of controls. With the music masker, the difference between groups only just reached significance ($p = 0.045$) and rested mainly on the poor performance of a single extreme ASD- outlier, whilst with the Japanese masker there was no significant difference at all between ASD- and control listeners. In spite of this, at an individual level around half of the ASD- listeners still had significantly elevated SRTs ($z \geq 1.65$) in these conditions: 3/7 were outliers with the Japanese masker, and 3/6 with the music. By way of comparison, the entire ASD- group performed significantly poorly with modulated noise, whilst there was only one ASD- outlier in the steady noise masking condition. This suggests that modulated *noise* may in some way present more of a challenge to ASD- listeners than other types of fluctuating masker (although it is still not nearly as problematic as intelligible speech).

Whilst the Japanese speech and music maskers in Experiment 1 varied in amplitude over time in a similar way to the modulated noise, and thus allowed for brief glimpses of the target speech, they both also differed from modulated noise in containing strong pitch information and dynamic spectral variation. The ASD- group's pattern of performance across these maskers, therefore, seems to imply that although a deficit in dip listening might not be uncommon in ASD (as indicated by poor performance with modulated noise), its impact on target speech intelligibility appears to be reduced (in at least some listeners) when a masking sound contains pitch and dynamic spectral detail, as well as dips. This mitigating effect, however, appears only in the absence of intelligible content in the masker.

There is certainly very little reason to suppose that the ability to process spectral detail in masking sounds might be impaired in ASD: Alcántara included a non-speech noise masker containing spectral 'dips' (but no pitch information) in his study of speech perception in high-functioning autism, and found no significant difference in SRT between the clinical group and matched controls (Alcántara et al., 2004). Like controls, the autistic group also performed better with the spectral dip masker than with steady noise. Spectral dips alone, however, did not appear to be sufficient to compensate for the deficit in performance when a noise masker also contained temporal dips, since Alcántara's HFA group did perform significantly poorly in that combined condition. The key aspect of this data with respect to the current study is that unlike our Japanese and music maskers, Alcántara's masker with spectral and temporal dips was based on a noise carrier and therefore lacked periodicity. There would have been no percept of a dynamic pitch contour.

If spectral dips in a masker are not sufficient to compensate for impaired processing of temporal dips (as Alcántara's data suggests), the next important consideration is what role, if any, masker pitch information might play in ASD listeners' speech reception thresholds. Evidence from neurotypical listeners indicates that voice fundamental frequency (F0) is an important factor in auditory stream segregation (Binns & Culling, 2007), and moreover that perception of acoustic temporal fine structure (TFS) - which contributes to pitch processing - may also support dip listening in amplitude modulated maskers (Gnansia, Jourdes, & Lorenzi, 2008; Hopkins & Moore, 2009). Much of the work investigating this role of temporal fine structure has centred on listeners with hearing impairment and/or cochlear implants, for whom access to TFS (in the target speech, as well as in any masking sound) is reduced or absent altogether. There is evidence not only that the ability to benefit from fluctuations in a background noise is greatly reduced in these listeners, but also that release from masking (i.e. improvement in speech identification performance with AM noise maskers relative to steady noise) is directly correlated with their ability to use TFS for speech identification in quiet (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006).

This pattern of performance in hearing-impaired listeners seems to suggest that pitch in target speech can be important for auditory scene analysis, and when it comes to speech in a fluctuating masker, perhaps pitch acts as a strong cue to the presence of speech in the 'dips' (Lorenzi et al., 2006). Whilst amplitude modulating a non-speech masking noise does allow for useful glimpses of the target speech, it also makes the temporal envelopes of target and masker more perceptually similar to each other. This may explain why pitch in the target can be important for differentiating it from the masker, and why listeners with weaker pitch representation gain less benefit from dip listening. Target intonation contours also appear to cue attention to key words in a sentence (Binns & Culling, 2007), which could support dip listening in terms of knowing *when* to focus most listening effort.

All of the target material used in Experiment 1 here contained periodic TFS. Moreover, the ASD listeners in Experiment 1 (see chapter 3 of this thesis) showed a similar response to controls when the target speech intonation contour was flattened: their SRTs were slightly worse than when target material had a natural F0, and if anything the effect was slightly larger in the ASD- subgroup. This implies that these listeners were making relatively normal use of the target pitch contours, which seems predictable in light of the idea that pitch processing may, if anything, actually be enhanced in ASD, and that processing of grammatical prosody is intact. If they are extracting and using the target TFS detail successfully, there seems little reason to suppose ASD- listeners would not also be using this information to support dip listening.

The question remains, therefore: why do ASD- participants appear to perform more similarly to control listeners when a fluctuating masker contains periodicity than when it consists of random noise. It has been suggested that when two competing vowels are presented, the auditory system relies heavily on harmonic structure in the masking vowel (rather than the target) to segregate it and 'cancel' it out (de Cheveigne, McAdams, Laroche, & Rosenberg, 1995; de Cheveigne, Kawahara, Tsuzaki, & Aikawa, 1997; Deroche & Culling, 2011). This theory of harmonic cancellation has been called into question by other groups, however, (see for example, Vestergaard & Patterson, 2009), and even if it is important for identifying simultaneous vowels which are precisely concurrent, this is a very different task from reporting sentences in fluctuating maskers. Binns and Culling's study (2007) suggests that for neurotypical listeners, altering the intonation contour of masking speech has much less impact on target intelligibility than the F0 of the target sentence itself. Similarly, Apoux and Healy (2011) have found that disrupting the TFS of a speech masker has almost no effect on target intelligibility. This seems to imply that masker periodicity is typically of relatively little importance, at least for speech-on-speech masking, in normal listeners. It might thus be reasonable to assume that periodicity in non-speech maskers would be similarly irrelevant.

It may be that this is another case where the atypical ASD pattern of performance reflects reduced proficiency in processing of the target material - evidently not in terms of processing target intonation, but possibly in terms of maintaining selective attention to speech. Some types of non-speech masker - such as modulated noise, for example - could be more inherently distracting than others. A serious barrier to any attempt at interpreting the possibility of reduced deficits with periodic, unintelligible maskers such as the Japanese speech or music used in Experiment 1, however, is that they differed from the other fluctuating maskers presented (AM noise on the one hand, and intelligible speech on the other) in a variety of additional ways. Long-term average spectra and temporal envelopes were not matched across these different masking sounds, nor was F0 itself. Any or all of these things could have been a factor in the results, and so for Experiment 2, a new set of masker stimuli was created, all based on a single male speech masker. This allowed us to investigate in a more controlled way the patterns of response in the ASD group to masker pitch information, envelope fluctuations and intelligibility. Based on their performance with the Japanese and music masker, the hypothesis was that deficits in ASD performance would be smaller when a fluctuating non-speech masker containing periodicity was presented, compared to when there was an aperiodic fluctuating masker or intelligible masking speech itself.

4.1.2 Selective attention to speech

A second very striking aspect of the results of Experiment 1 was the evidence that introducing a spatial separation between target and masking speech greatly reduced ASD- listeners' deficits in speech reception (see Chapter 3). On average, this subgroup experienced 9.6 dB release from masking when the masker was separated from the target by 90°, and even the ASD+ listeners showed an improvement of 3.6 dB. Furthermore, whilst all ASD- listeners were significant outliers (based on z-score) when target and masker were collocated, half of them performed within the normal range with spatial separation (and one further individual was only just on the borderline). These significant effects occurred even though controls showed virtually no detectable improvement in SRT between conditions.

Since this occurred in a listening condition where informational masking is presumed to be high (same-gender female masker), it appears that auditory scene analysis and/or selective attention can be boosted by a strong spatial cue (which, unlike stream segregation based on the differences between two female talkers, does not require a skill specific to speech perception).

As a group, the ASD- listeners did still perform significantly worse than controls in the spatially-separated condition, however, which raised the question of whether there might be any listening conditions under which performance with a speech masker could be unimpaired. Aside from the spatial conditions, in which head-related transfer functions were used to induce a sound localisation percept, all the target and masker material in Experiment 1 was presented diotically. The method of presenting target and masking stimuli *dichotically*, however, can be used as a means of exploring interference from masking speech material when there is minimal overlap between masker and target signals in the auditory periphery - in other words, almost pure informational masking. A recent study using the CRM found that whilst neurotypical adults show relatively small changes in monaural target speech recognition with the addition of an opposite-ear speech masker (increases of around 4 dB in SRT), children can experience quite high levels of interference (up to 20 dB; Wightman et al., 2010). This suggests that it takes time to develop adult-like proficiency in selective attention to speech. Interestingly, data from Wightman's study also supported earlier findings that developmental improvements in SRT with opposite-ear masking speech follow a more rapid course than those with same-ear masking speech (Leech, Aydelott, Symons, Carnevale, & Dick, 2007). This could indicate that these two target-masker listening configurations involve somewhat different processing mechanisms.

The second strand of Experiment 2 presented here, therefore, was designed to investigate whether ASD- listeners would continue to show significantly poor performance with an opposite-ear masker, or whether this degree of separation between target and masker would be sufficient for them to achieve normal performance. Comparing performance in a baseline condition (monaural target and no masker) with SRTs for opposite-ear masking speech should indicate whether susceptibility to informational masking by speech is still apparent in ASD even when energetic masking is minimised. Furthermore, neurotypical adults tend to be relatively unaffected by an opposite-ear masker when speech is presented alone to the target ear, but when performance with an *ipsilateral* speech masker (presented monaurally) is compared to the same configuration with additional opposite-ear masking, significant effects on intelligibility can be observed (equivalent to a 5 dB shift in SRT; Wightman & Kistler, 2005). This effect is somewhat reduced, however, in young children - who show much higher levels of masking from ipsilateral speech on its own. One might imagine that if ASD- listeners in the current study were already experiencing high levels of informational masking from a single same-ear masker, like children they might also show smaller decrements in performance with the addition of a second masking talker in the opposite-ear. To explore this idea, Experiment 2 included conditions with ipsilateral speech masking and ipsilateral plus contralateral speech masking.

4.2 Method

4.2.1 Participants

Nine of the original adults with ASD who took part in Experiment 1 were involved in Experiment 2, along with four of the original control participants. This involved a full new set of listening tests, as described in Chapters 4 - 5, which were administered around one year after the first round of data collection. A further 6 new control listeners and one new ASD listener were recruited for Experiment 2, using the same recruitment methods as described in Chapter 2. In total the two groups each contained 10 participants, all male except for one female in the ASD group. As before, control participants reported no history of language or communication problems, and no other neurological disorder. The newly-recruited ASD participant had a confirmed diagnosis of Asperger's Syndrome, and his score on the ADOS was consistent with this. All new recruits to this study completed the WASI (abbreviated) and ASQ, as described in Chapter 2. All participants in Experiment 2 also completed the British Picture Vocabulary Scales II (BPVS; Dunn, Dunn, Whetton, & Burley, 1997) as an additional measure of receptive vocabulary. This is an assessment primarily designed for use with children, but which can also be used with adults (see Dunn et al., 1997 and Hazan, Messaoud-Galusi, Rosen, Nouwens, & Shakespeare, 2009). Standardisation data for adults is based on only a relatively small sample, however, many of whom were speakers of English as a second language, so raw scores only are presented here.

T-tests indicated there were no significant differences between the two groups in terms of age, performance IQ (WASI Performance t-score) or vocabulary knowledge (BPVS raw score): all $p > 0.05$ (see Table 4.1). As expected there was a significant difference between control and ASD listeners' scores on the Autism Spectrum Quotient ($t(18) = 9.44$, $p < 0.001$).

Table 4.1 Participant group statistics: IQ, ASQ and age

	Group	N	Mean	Std. Deviation
Performance t-score	Control	10	61.9	4.2
	ASD	10	62.2	7.0
BPVS raw score	Control	10	155.5	6.9
	ASD	10	149.2	12.3
ASQ	Control	10	13.0	4.1
	ASD	10	36.1	6.6
Age	Control	10	25.9	4.5
	ASD	10	26.2	5.4

The experiment was conducted with the approval of UCL's Ethics Committee and in accordance with their guidelines, as previously described.

4.2.2 Stimuli

i) AM and F0 masker conditions (Conditions 1-5)

Targets for this set of conditions were the female mCRM sentences exactly as described in previous chapters, and all target and masker stimuli were presented diotically. There were 5 masker conditions designed to differentiate the effects of fundamental frequency content and amplitude modulation on performance (see Table 4.2).

Table 4.2 AM and F0 masker conditions (presented diotically, with diotic targets)

Condition 1	Steady speech spectrum noise
Condition 2	Amplitude-modulated speech spectrum noise ('AM noise')
Condition 3	A continuous intonation contour, based on male speech ('F0')
Condition 4	Amplitude-modulated intonation contour ('AM F0')
Condition 5	Male speech masker

The basis of all 5 of these maskers was English male speech, unrelated to the mCRM targets (condition 5). Several short passages recorded by a single talker were taken from the

EUROM corpus (Chan et al., 1995), and concatenated to make a 30 second sound file. The EUROM passages each consist of 5 short thematically-related sentences with everyday subjects, such as a phonecall to an orders department or a description of a dinner menu. The selected passages contained no colour, number or animal words, in order to minimise masker similarity to the mCRM target material. Any long pauses (more than 100 ms) between sentences were removed from the final sound file.

The four remaining non-speech maskers were all filtered so as to have the same long-term average spectrum as the speech masker described above. Conditions 1 and 2 consisted of a white noise, filtered to have the same spectrum as the speech. In condition 1 the noise had a steady amplitude, and for condition 2 it was also amplitude modulated with the envelope fluctuations of the original speech. For this purpose, the speech envelope was extracted by full-wave rectification and smoothing with a low-pass filter (2nd order Butterworth filter, 30 Hz cut-off).

Conditions 3 and 4 were based on a pulse train carrier which approximated the intonation contour of the natural speech masker. The pitch contour of the speech was sampled in PRAAT (Boersma & Weenink, 2010) (sampling frequency = 1000, minimum fundamental frequency = 70 Hz). The pitch contour was then interpolated through unvoiced and silent periods using piecewise cubic Hermite interpolation in logarithmic frequency. The end result was a continuous pitch contour following the same rise-fall pattern of the original speech. Perceptually, this has a faintly musical, 'buzzing' quality. The contour was presented without amplitude modulation for condition 3, and was modulated with the speech envelope fluctuations for condition 4.

ii) Monaural targets with monotic/dichotic maskers

To explore the effects of presenting different material to each ear, the female mCRM targets in the remaining six conditions were presented monaurally (and the target ear - right or left - balanced across listeners and groups). Listeners were told in advance which the target ear would be, and for each listener the target ear remained the same across all six conditions. In some conditions, masking speech was presented to the same ear and in others to the opposite ear. The similarity of the masking material to the target was also varied, as described in the table below:

Table 4.3 Monotic/dichotic masker conditions (presented with monaural female targets)

Condition 6	Baseline: no masker
Condition 7	Male speech (task irrelevant): same ear as targets
Condition 8	Male speech (task irrelevant): opposite ear to targets
Condition 9	2 male maskers: 1 same ear, 1 opposite ear
Condition 10	Female speech (mCRM masker): same ear as targets
Condition 11	Female speech (mCRM masker): opposite ear to targets

A baseline condition, in which targets were presented in the absence of any masker, was included here to provide an estimate of SRTs in a situation when no masking was involved at all (Condition 6). Although expressed as an SRT for convenience in comparing results across conditions, this value represents the minimum level of the target speech which supports threshold performance levels. This could be used for comparison with the opposite-ear speech masker condition, which should cause no energetic masking, but might introduce masking at higher processing levels. The male speech masker in conditions 7 and 8 was the same one as in the previous section, consisting of concatenated EUROM passages. To create a dichotic masking condition (Condition 9), this masker was presented to one ear (the target ear) and masking speech from a second male talker was presented to the opposite ear. This second speech masker was also created by concatenating EUROM passages, but using a different talker and a different selection of passages from those of the first masker).

The female masker conditions (10 and 11) involved mCRM sentences, selected so that the masker colour and number were always different from those of the target sentence on each trial. In all the other conditions of this experiment, the masker on each trial was an excerpt from single relatively long sound file, and the target sentence onset was delayed by 200 ms relative to the masker. In order to achieve this same 200 ms of 'warning noise' using the relatively short mCRM masker recordings, on each trial two masker sentences were concatenated and then a randomly-selected excerpt of the required length was played out (masker length = 200 ms + target sentence length). This meant that, unlike in the previous experiment, there was an onset cue to support auditory stream segregation with this female mCRM masker material. It also meant that target and masker key words would not overlap in time on every trial. For condition 10, the female masker was presented to the same ear as the target, and for condition 11 to the opposite ear.

4.2.3 Procedure

The mCRM task procedure itself was identical to that described in Chapter 2, and ran, as before on a Samsung Q45 laptop, but this time using an external Edirol Cakewalk UA-1SG USB sound card. Participants were seated in a sound-protected booth and stimuli were presented via Sennheiser HD25 II closed supra-aural headphones. Listeners reported the target colour and number on each trial using a mouse to select from an on-screen response grid, and were given feedback (correct/incorrect) on each trial. An adaptive tracking procedure was used to estimate the speech reception threshold (SRT; 50% correct) in each condition. The starting step size was 10 dB and final step size 2 dB in all cases. In

Experiment 1 (Chapters 2 and 3), the intensity of both masker and target changed from trial to trial in order to achieve the desired SNR whilst maintaining a constant overall output level. For experiment 2, however, the masker intensity was always fixed (at around 65 dB SPL) and only the target intensity varied from trial to trial.

The AM/F0 masker conditions and the monotic/dichotic conditions were completed as two separate sessions, and the order of these sessions was balanced across listeners. Within the sessions, listeners completed 2 consecutive blocks of trials for each masker condition, and a mini-block of 4 practice trials at the start of each new condition. Condition order within-session was randomised across listeners. Participants were also given a demonstration of the task, including examples of all the maskers, and completed a short practice session prior to starting the task proper, in order to familiarise them with the procedure. Once a participant had completed the mCRM sessions, their SRTs were reviewed and in cases where these differed by 5 dB or more across the 2 blocks of a condition, a third block was completed. The median SRT for each listener / condition was entered into the main analysis.

The mCRM test sessions were conducted as part of the larger round of data collection which made up Experiment 2, and which included the listening task described in Chapter 5 (Interrupted speech). Additional pilot tests, not further reported in this thesis due to methodological flaws, were also conducted as part of this experiment, including a test of environmental sound detection and a working memory task. All listeners completed the listening tests before the working memory task and BPVS (plus WASI, where required), and the order of the listening tests (mCRM (this chapter), Interrupted Speech (See chapter 5) and environmental sound detection (not reported)) was balanced across listeners and matched across groups. Most participants completed all sections of Experiment 2 in a single day, with two main sessions (approximately 2 hours of listening tests and 1 hour of cognitive tests) interspersed with regular breaks. None reported finding this unduly arduous.

4.3 Results

4.3.1 Effects of masker amplitude modulation and frequency contour

Data from two of the ASD participants had to be excluded from the analysis, owing to an error with the stimuli, thus reducing the final sample size of the ASD group to 8 for these 5 conditions.

As Figure 4.1 and Table 4.4 show, the control group's SRTs followed the expected pattern. Steady noise was the most effective masker (mean SRT= -10.1 dB) and, from this baseline, mean SRTs improved substantially with both the F0 masker (mean difference in SRT between conditions 1 and 3 = 5.3 dB) and the amplitude modulated noise (mean difference between conditions 1 and 2 = 7.4 dB). The amplitude modulated noise would have offered clearer opportunities for dip listening than the F0 masker, which accounts for the enhanced release from masking in the former condition. Adding amplitude fluctuations to the F0 masker itself improved SRTs by an additional 3.6 dB over the original unmodulated F0 condition, but this AM F0 masker still affected target intelligibility more than the natural speech masker. Controls' mean SRT with the natural speech masker was 1.3 dB better than with the AM F0 masker.

Figure 4.1 Effects of masker amplitude modulation and frequency contour on group SRTs

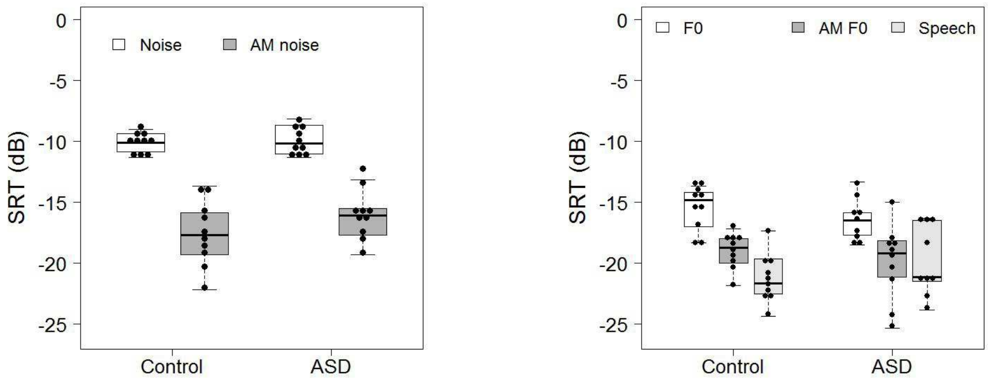


Table 4.4   Descriptive statistics: amplitude modulation and frequency contour conditions

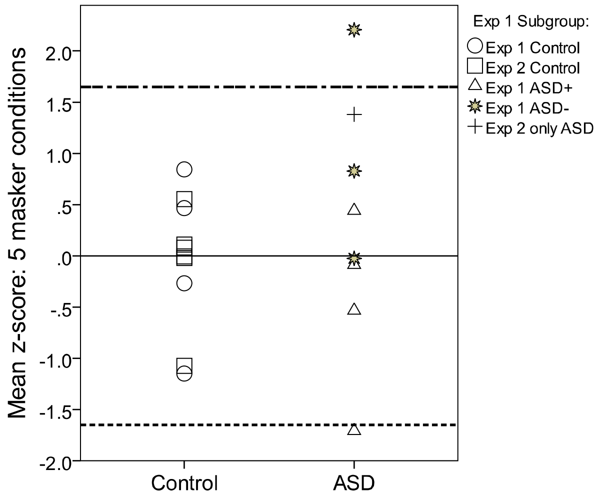
	Group			
	Control		ASD	
	Mean SRT (dB)	Std.dev.	Mean SRT (dB)	Std. dev.
Condition 1: Steady noise	-10.1	0.8	-9.9	1.2
Condition 2: AM noise	-17.5	2.7	-16.0	2.2
Condition 3: Male F0	-15.4	1.8	-16.5	1.7
Condition 4: AM male F0	-19.0	1.4	-19.9	3.0
Condition 5: Male speech	-21.3	2.0	-19.8	2.9

A mixed ANOVA (2 groups x 5 masking conditions) showed a significant effect of condition ($F_{4, 64} = 86.0$, $p < 0.001$, with Huynh-Feldt correction), but no significant effect of group ($p = 0.31$), and an interaction between group and condition that only approached significance ($F_{4, 84} = 2.1$, $p = 0.09$). The small sample sizes here clearly led to a lack of power, but even so the largest mean differences between groups - which occurred with the AM noise and male speech maskers - were only 1.5 dB.

In spite of this, there did seem to be a slight trend towards atypical patterns of performance amongst the ASD group. ASD listeners appeared, on average, to respond slightly differently from controls to some of the stimulus manipulations outlined above. They showed virtually no difference in performance between the AM F0 masker and the natural speech masker (mean difference between conditions = 0.1 dB, compared to 1.3 dB for controls), and improved slightly less than controls with the AM noise over steady noise (control mean improvement = 7.4 dB, ASD mean improvement = 6.1 dB). However, they improved slightly *more* than controls with the F0 masker over the steady noise (control mean improvement = 5.3 dB, ASD mean improvement = 6.6 dB). In fact, the ASD listeners' mean SRTs were around 1 dB better than controls' for both the F0 and AM F0 maskers, even though their performance with the natural speech masker remained worse. Given that there are reasons to believe such a pattern might well occur in ASD across this combination of maskers (i.e. evidence elsewhere of poor dip listening with fluctuating maskers, and of enhanced pitch perception), it would be interesting to see whether a future study with more power might detect a significant group x condition interaction.

The results of Experiment 1 had led to an expectation that some ASD listeners would perform within the normal range across these new listening conditions, and that some would perform extremely poorly. Once again, therefore, participants' z-scores were calculated for each condition, based on the control mean and standard deviation (corrected for outliers). The results of this analysis were surprising. Firstly, as Figure 4.2 shows, only one listener from the ASD group had a mean z-score > 1.65 (mean z = 2.21, averaged across 5 conditions). There was little resemblance to the bimodal distribution of ASD listeners' mean z-scores that was observed in Experiment 1. Part of the explanation for this may lie in the fact that only 4 of the original 6 listeners who were in the poor-performing ASD-subgroup returned for this second experiment, and data from one of these was subsequently excluded from the analysis of these conditions due to the stimulus error. Nevertheless, this still left 3 of the original ASD- listeners, and yet only one of them performed significantly poorly across the masker conditions described here. The other two were well within the normal range.

Figure 4.2 Scatterplot showing mean z-scores for the 5 AM/F0 masker conditions

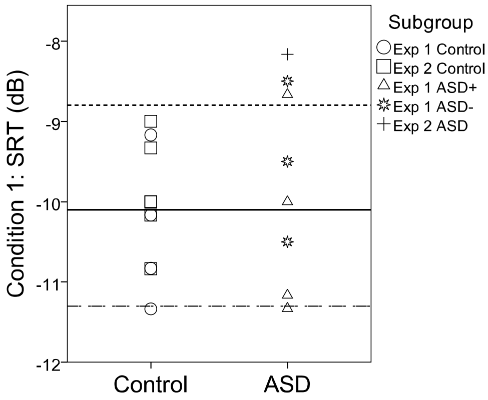


Solid line indicates z = 0
Dashed lines indicate z = + 1.65 (poor performance) and z = - 1.65 (good performance)

Of the five masking conditions described here, only one involved intelligible speech, and it appears that this low proportion of speech maskers may have been a key factor contributing to relatively 'normal' performance (in terms of mean z-scores) across the entire ASD group. Figure 4.3 shows how listeners' SRTs were distributed relative to the control mean SRT (corrected for control outliers) in each individual condition. A total of four ASD listeners, including two of the three from the original ASD- group, were markedly poor performers with the male speech masker (condition 5) and had z-scores ranging from 2.27 to 3.63. There was a clear split in SRTs between this group and the remainder of the ASD participants who were within the normal range.

Figure 4.3 Scatterplots showing SRTs in dip listening conditions, by group and subgroup

i) Condition 1: Steady noise

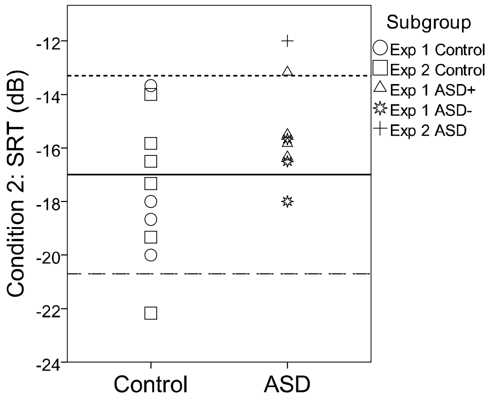


Key:

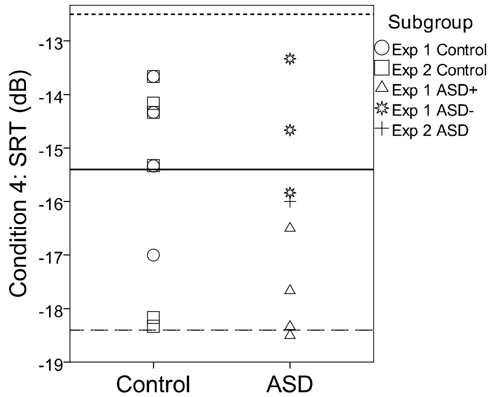
———	control mean
- - - - - -	z = +1.65 (poor performance)
--- --- ---	z = -1.65 (good performance)

Exp 1 control = also participated in Exp 1
Exp 2 control = only participated in Exp 2

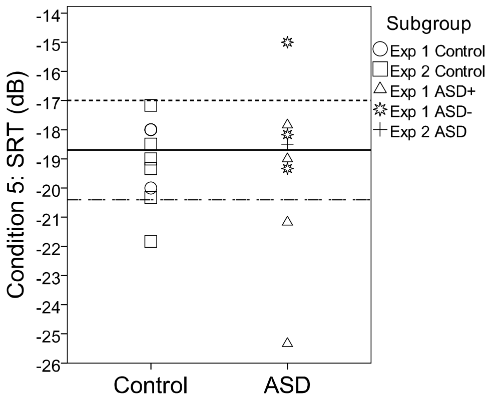
ii) Condition 2: Modulated noise



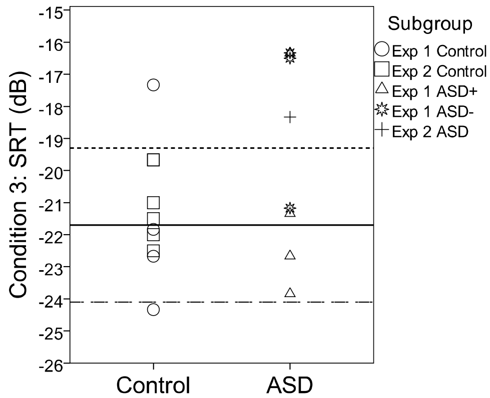
iii) Condition 3: F0 contour



iv) Condition 4: AM F0 contour



v) Condition 5: Natural speech masker



Looking across the remaining maskers (all non-speech), however, the outliers are far less clear-cut as a group. Firstly, there were no significant outliers at all with the F0 contour masker, and only one with the AM F0, which suggests that these masking conditions involving pitch but no language may have been less challenging on some level, even for the weaker performers. Secondly, although there were 3 outliers with steady noise and 2 with AM noise, their deficits tended to be considerably smaller than with the speech masker (range of outlying z-scores = 1.85 -2.50 for steady noise, and 1.73 - 2.25 for AM noise). This is broadly in line with the relatively small deficits for non-speech maskers shown by ASD-listeners in Experiment 1.

Figure 4.3 also shows which participants had previously been involved in Experiment 1, and which were newcomers for Experiment 2. Since more than half the control group were new to Experiment 2 it seemed possible that their relative inexperience compared to ASD listeners might have been a confounding factor. There is no indication, however, that controls who had experience from Experiment 1 performed better than the newcomers in Experiment 2, as confirmed by results of a mixed ANOVA (2 control groups x 5 conditions) which found no significant effect of group ($p = 0.91$) and no group x condition interaction ($p = 0.88$). The sole new recruit to the ASD group did perform significantly poorly with steady noise, AM noise and the speech masker, but was well within the normal range in the two non-speech F0 conditions, which again suggests that familiarity with the task was probably not a major factor in his performance.

Only the steady noise and natural speech masker conditions in the current experiment were sufficiently similar to conditions in the Experiment 1 mCRM for a direct comparison between SRTs to be valid. Taking control ($N = 6$) and ASD ($N = 7$) listeners separately, and using paired t-tests to compare performance in Experiment 1 Condition 1 with Experiment 2

Condition 4 (both steady noise) and Experiment 1 Condition 4 with Experiment 2 Condition 5 (both male speech unrelated to the mCRM targets), there were no significant differences in SRT across experiment for either group in either pair of conditions (all $p > 0.05$). Since around a year had elapsed between the two experiments, this suggests that performance amongst both groups of listeners was relatively stable across time.

4.3.2 Monaural targets with monotic/dichotic speech maskers

Further difficulties with the software meant that 3 ASD listeners did not complete the full set of 5 dichotic mCRM conditions. One participant failed to complete condition 9 (two dichotic male maskers), leaving a sample size of 9, and three omitted conditions 10 and 11 (female masker conditions), reducing the sample size to 7 for those conditions. Although these five dichotic conditions were conceived as a coherent set, they were analysed in three separate subsets of conditions in order to make the most of the existing data.

All listeners completed a core set of 3 conditions: the female mCRM targets presented monotically in quiet (i.e. a baseline measure with no masker; Condition 6), the monotic mCRM targets with a male masker presented to the same ear (Condition 7), and the monotic mCRM targets with the male masker in the opposite ear (Condition 8). Figure 4.4 shows that control listeners' SRTs with the opposite-ear masker tended not to be quite as low (good) as with no masker at all, which suggests that some informational masking was still taking place. The mean difference between these conditions was 4.3 dB. There was, however, an improvement of 16.7 dB in mean SRT between the same-ear and opposite-ear masker conditions (see Table 4.5).

Taken as a single group, the ASD participants showed a broadly similar pattern of SRTs, although the mean differences between the no masker and opposite-ear masker conditions (2.6 dB) and between the same-ear and opposite-ear masker conditions (15.3 dB) were both slightly smaller than for controls. Mean SRTs for both the no-masker and opposite-ear masker conditions were also a little higher than in the control group (see Table 4.5). A mixed ANOVA (2 groups x 3 conditions) showed no significant effect of group, however, (F 1, 18 = 1.1, p = 0.31) and no group x condition interaction. The main effect of condition alone was highly significant (F 2, 36 = 262.4, p < 0.001.

Figure 4.4 Boxplots showing SRTs for monotic mCRM female target with male masker

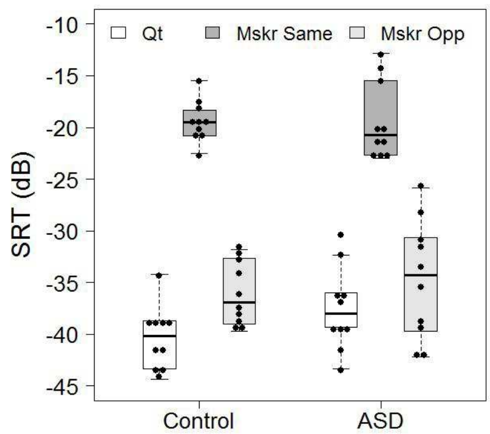
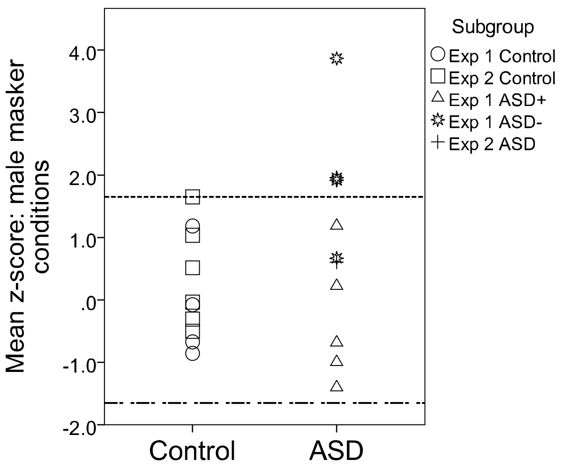


Table 4.5 Descriptive statistics: monotic female mCRM target and male masker

	Group	Mean	Std. Deviation	N
Condition 6:	Control	-40.4	3.1	10
No masker	ASD	-37.4	4.0	10
Condition 7:	Control	-19.4	2.0	10
Same ear male masker	ASD	-19.5	3.8	10
Condition 8:	Control	-36.1	3.2	10
Opposite ear male masker	ASD	-34.8	5.8	10

Although differences in SRT between the ASD group and control participants did not appear, on average, to be significant, there was clearly considerable variability in performance within the ASD group itself. By contrast with the non-speech conditions described in the previous section, z-scores based on control performance for the two speech-masker conditions here indicated that a number of the ASD listeners once again performed consistently poorly across the 3 conditions - much as observed in Experiment 1. As Figure 4.5 shows, 3 ASD listeners had mean z-scores greater than 1.65 (z = 1.92, 1.95 and 3.87), and all of these had been in the original experiment 1 ASD- group. Only a single one of the listeners who had previously been ASD- produced a mean score for these dichotic conditions which was within the normal range.

Figure 4.5 Scatterplot showing listeners' mean z-score across 3 mCRM conditions



Key:

- - - - - - - z = +1.65 (poor performance)
--- --- --- z = -1.65 (good performance)

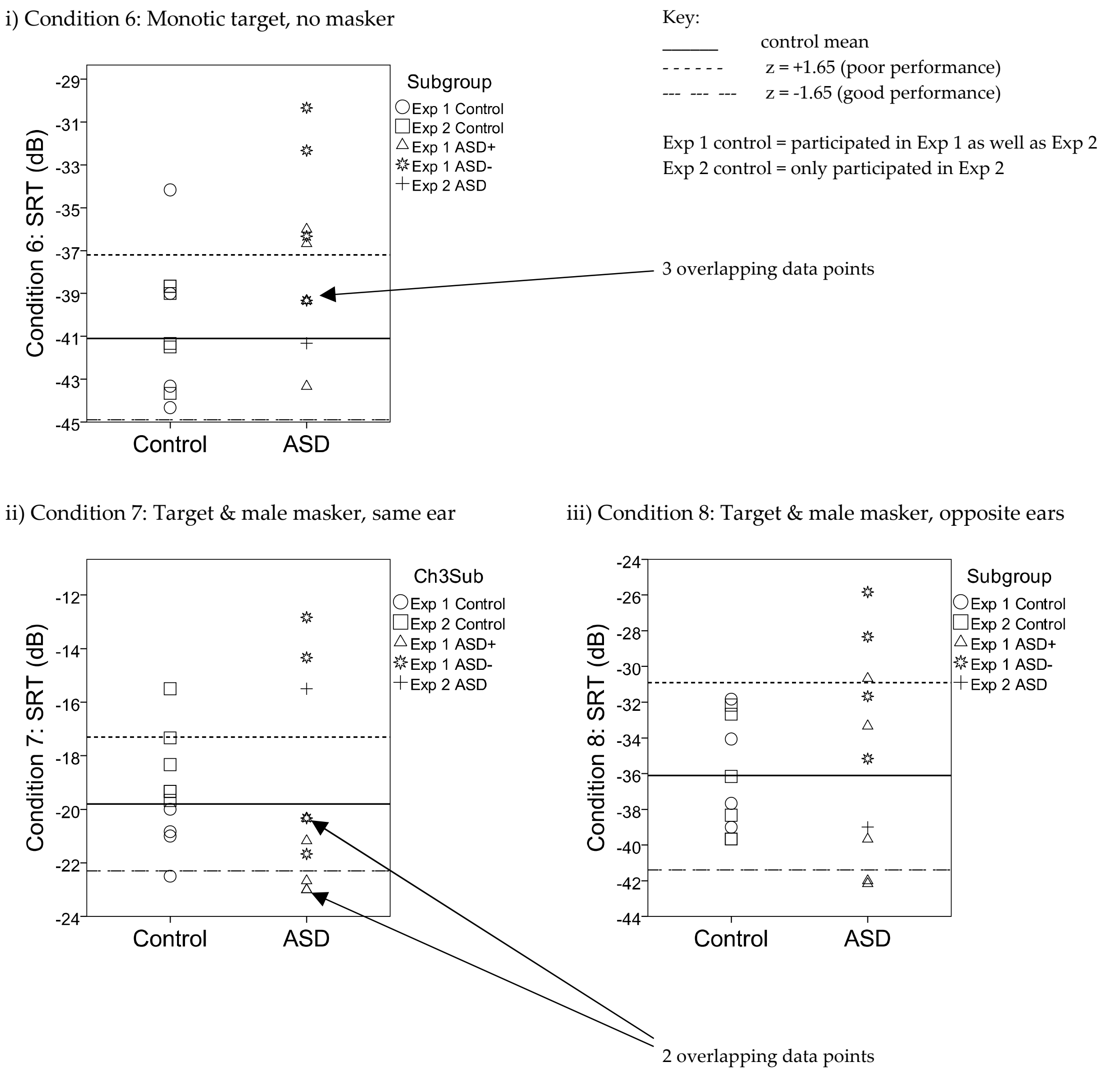
Exp 1 control = participated in Exp 1 and Exp 2
Exp 2 control = only participated in Exp 2

z-scores were averaged across 3 conditions:

Condition 6: no masker
Condition 7: same-ear male masker
Condition 8: opposite-ear male masker

The condition here which seemed to reveal the split between ASD+ and ASD- most clearly was where target and masker were presented to the same ear (see Figure 4.6, ii). This target-masker configuration would have maximised energetic masking as well as informational masking, and the distribution of SRTs amongst the ASD group is clearly bimodal: three listeners had z-scores well above the norm (z = 2.84, 3.60, and 4.59), whilst the remainder of the group actually tended to perform better than the control mean. With the opposite-ear masker, however, the ASD scores were spread more evenly across the range. Not only that, but there was almost no overlap between poor-performers in this condition, and poor- performers with the same-ear masker. Only a single listener was a significant outlier in both, and two listeners who had actually performed relatively well in the same-ear condition (z = -0.34 and z = -1.22) performed outside the normal range for the opposite-ear masker (z = 1.72 and z = 2.45).

Figure 4.6 Scatterplots showing SRTs in dichotic mCRM conditions, by group & subgroup



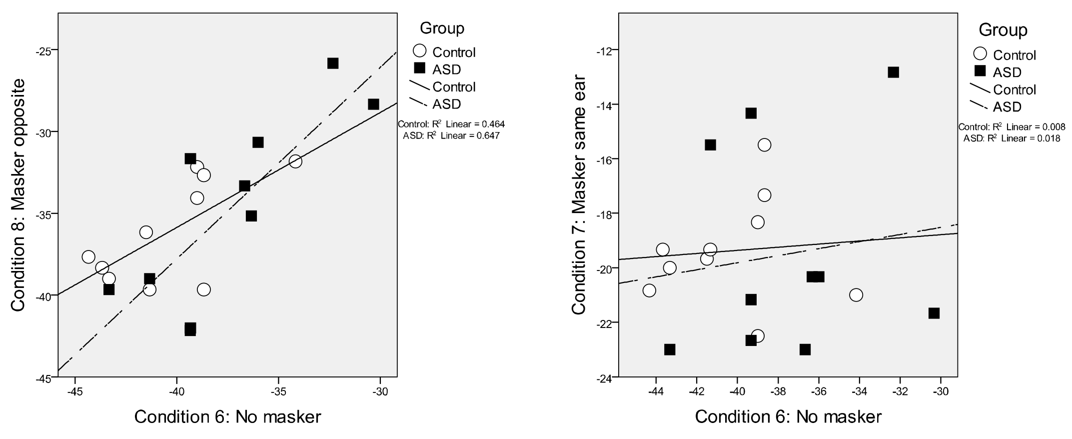
What is particularly striking, however, is the fact that in the absence of any masker, 5 of the ASD listeners - half the group - performed significantly poorly (Figure 4.6, i). All listeners had passed an audiometric screening test, so the speech should have been equally 'audible' to them in the literal sense. What this result may reflect, then, is a basic difference in speech processing efficiency amongst some ASD listeners. ASD- listeners in Experiment 1 had also shown reduced intelligibility for 4-channel vocoded speech targets, again with no masker, but in that case presented at a comfortable listening level, so it appears that perhaps some ASD listeners require a fuller set of acoustic cues to be available for accurate speech perception.

SRTs with the opposite-ear masker appeared to be highly dependent on performance with the targets in quiet, particularly amongst the ASD listeners, whereas performance with the same-ear masker appeared to bear little relationship to this more basic measure of speech processing efficiency (see Figure 4.7).

Figure 4.7 Scatterplots showing the relationship between SRTs with no masker and

i) SRTs with opposite-ear masker and ii) SRTs with same-ear masker

i) SRT in quiet x SRT masker opposite ear ii) SRT in quiet x SRT masker same ear

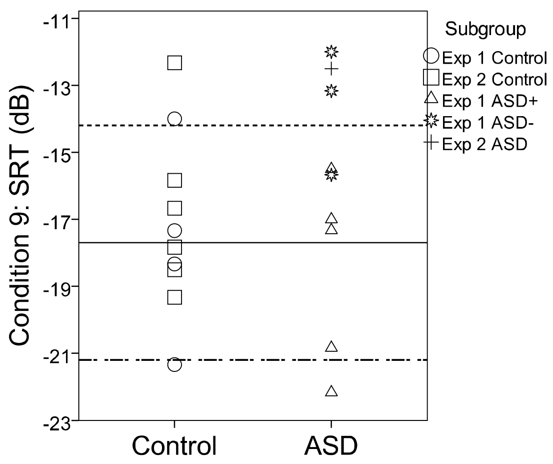


Control: Pearson's R = 0.68, p = 0.03*
ASD: Pearson's R = 0.80, p = 0.005*

Control: Pearson's R = 0.09, p = 0.81
ASD: Pearson's R = 0.14, p = 0.71

One ASD listener did not complete Condition 9, in which two male maskers were presented dichotically, so this data was analysed separately (see Figure 4.8). Once again the between-groups difference in mean SRT was small (Control mean SRT = -17.1 dB, ASD mean = -16.2 dB) and did not reach significance ($t(17) = 0.6$, $p = 0.53$). These thresholds were around 2-3 dB higher than with a single same-ear male masker, which is much less than the 9-11 dB increase observed going from 1 to 2 male maskers when target and maskers were all presented diotically (see Chapter 3).

Figure 4.8 Scatterplot showing individual SRTs: dichotic male maskers (Condition 9)



Key:

______ control mean

- - - - - - z = +1.65 (poor performance)

--- --- --- z = -1.65 (good performance)

Exp 1 control = participated in Exp 1 and Exp 2

Exp 2 control = only participated in Exp 2

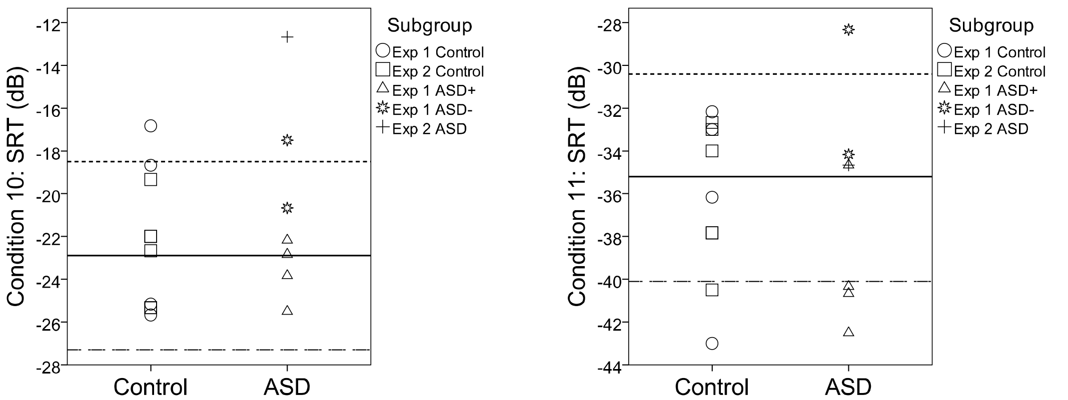
Three ASD listeners again performed significantly poorly in this condition ($z > 1.65$), and these were the exactly the same three (2 from the original ASD- group, and the newcomer to Experiment 2) who were also outliers with a single same-ear male masker. Two of them were also outliers with the female same-ear masker (see below), but only one continued to perform significantly poorly with the opposite-ear maskers and in the no-masker condition.

Finally, in conditions 10 and 11, where the masker presented in same-ear and opposite-ear configurations was changed from male (opposite gender) to female (same gender), there continued to be some extreme ASD outliers (see Figure 4.9), even though for these conditions the ASD sample size was only 7, and only 2 of these were members of the original ASD- group. Again, one of the original ASD- listeners performed well within the normal range on these conditions, whilst the other continued to perform poorly. Interestingly, the new ASD participant - who had not taken part in Experiment 1 - performed significantly poorly with the same-ear female masker, as he did with the same-ear male masker, and dichotic male maskers, but not with either of the opposite-ear maskers nor in the no-masker condition.

Figure 4.9 Scatterplots showing SRTs: female masker conditions (same/opposite ear)

i) Condition 10: Target & female masker, same ear

ii) Condition 11: Target & female masker, opposite ears



Key:
______ control mean
- - - - - - z = +1.65 (poor performance)
--- --- --- z = -1.65 (good performance)

Exp 1 control = participated in Exp 1 as well as Exp 2
Exp 2 control = only participated in Exp 2

Table 4.6 Descriptive statistics: mCRM female masker conditions (same ear / opposite ear)

	Group	Mean	Std. Dev.	N
Condition 10: Same ear Female masker	Control	-22.3	3.2	10
	ASD	-20.7	4.4	7
Condition 11: Opposite ear Female masker	Control	-36.0	3.7	10
	ASD	-36.5	4.9	7

4.4 Summary of results

Whilst it was unfortunate that a number of missing data points made the analysis more complicated than might otherwise have been the case, this study both confirms and extends the findings of Experiment 1. The key points are as follows:

- When non-speech maskers are matched to their speech counterparts as closely as possible in terms of amplitude modulation and fundamental frequency content, speech reception deficits amongst even the lowest-performing ASD listeners tend to be small.
- Almost the entire ASD group performed within the normal range for non-speech maskers based on a pitch contour and in fact, on average, their SRTs were slightly better than controls'. There were more poor-performing outliers when maskers consisted of noise with no regular pitch content.
- When intelligible speech was used as a masker, challenging selective attention more directly, a clear group of low-performing listeners became apparent, just as in Experiment 1.
- Presenting a speech masker to the opposite ear from the target sentences appeared to affect some ASD listeners considerably more than others, and poor-performance in this condition appeared to reflect a different type of impairment from that involved in the same-ear target-masker configuration. There was little overlap between outliers in the same-ear and opposite-ear masking conditions.
- Strong correlation between SRTs in quiet and SRTs with the opposite-ear masker suggests that performance in the latter condition may be quite directly dependent on basic speech processing efficiency, or even attention/vigilance more generally, whereas with a same-ear masker other factors such as auditory stream segregation and selective attention become important.

4.5 Discussion

First and foremost, the data set presented in this chapter is reassuring, in that it broadly replicates the pattern of performance observed in Experiment 1. A range of masking conditions was presented, and once more, whilst the speech reception thresholds of some ASD listeners were consistently within the normal range, a subset of this group performed significantly poorly relative to controls. Once again, the largest deficits in terms of SRT tended to occur in conditions involving speech maskers, as demonstrated by the fact that a low-performing ASD subgroup was most evident in the second set of conditions (6-11) with monaural targets and monaural/dichotic maskers, and in condition 5 with the diotic natural speech masker.

When *non-speech* maskers were closely matched to speech maskers in terms of spectrum and F0, the majority of ASD participants' SRTs were within the normal range, just as they had been in Experiment 1 with Japanese speech and music maskers. In non-speech masking conditions consisting of aperiodic noise, however, ASD outliers were once more in evidence, although as in Experiment 1 their deficits were rather smaller than those for speech maskers. It is striking that 3 ASD listeners were outliers here even with the steady noise masker (condition 1), whereas there were there was only 1 outlier in the equivalent condition of Experiment 1. This is hard to explain and, given the small sample sizes involved, may not be meaningful. It does not appear that this reflects any overall inability to cope with the cognitive demands of the task, as this time around all listeners performed normally in a different condition (which, like steady noise, loads heavily on energetic masking): the F0 contour.

The low numbers of participants involved here and relatively small-scale deficits for non-speech masking conditions meant that there was no justification for defining ASD subgroups based on conditions 1-5. It may be that with a larger sample size, however, a trend towards a subgroup who are abnormally susceptible to aperiodic maskers might be revealed. The evidence here is highly suggestive that noise-based maskers are more problematic for at least some ASD listeners than non-speech maskers with a strong intonation contour. Given the evidence of auditory hyperreactivity and aversions to specific sounds such as vacuum cleaners, flushing toilets, and food blenders in ASD (Koegel, Openden, & Koegel, 2004), it may be that some of the participants simply found the aperiodic maskers unpleasant to listen to and that this affected their performance.

Viewing these results in conjunction with the Experiment 1 data, evidence for a 'pure' deficit in dip listening with fluctuating maskers is fairly equivocal. Such a deficit would have to be highly masker-specific since there was only a single outlier in the Experiment 2 AM F0 condition, which was actually fewer than the 3 outliers with steady noise (as well as being less than for AM noise, with 2). Since the F0 contour makes a masker more perceptually similar to the target speech (and hence presumably slightly harder to segregate from it), it is difficult to see how ASD listeners could be using it to support dip listening in a positive way. Perhaps the reason for better performance with F0-based maskers in ASD, whether fluctuating or not, is simply that they are inherently less distracting than aperiodic ones. The pattern of normal performance with steady noise and deficits in AM noise, as described by Alcántara et al. (2004) and observed in Experiment 1 here, was not clearly replicated in Experiment 2. This may purely be a reflection of the small sample size, yet again, but even if it were a consistent pattern of listening ability in ASD, it seems possible that is less reflective of a deficit in dip listening than of the fact that an unpleasant sound can become even more unpleasant (and hence more distracting) when it is switched on and off in an irregular manner. The auditory system responds most strongly to novelty, and it appears from the

MMN data described in Chapter 1 that responses to auditory change may be even stronger in ASD than in the neurotypical population.

The second group of masking conditions described in this chapter focused on speech maskers alone, and the effect of varying the target-masker listening configuration across ears. It was clear from Experiment 1 that a subgroup of ASD listeners are highly susceptible to speech-on-speech masking, and that a strong cue to selective attention such as moving the masker to a different spatial location from the target led to large reductions in SRT. Even with virtual spatial separation, however, there is a substantial amount of energetic masking going on, and so in the current experiment targets were presented monaurally with an opposite-ear speech masker to investigate performance under conditions where the masking is primarily informational.

Interestingly, although the hypothesis was that a subgroup of ASD listeners might show atypical levels of informational masking with the opposite-ear masker, the situation did not appear to be as simple as this. Half the ASD group actually performed significantly poorly in the baseline condition, with monaural targets and no masker at all (Condition 6), and in fact rather fewer of them (3/10) were outliers with opposite-ear masking (male; Condition 8). A separate analysis of performance by subgroup would not have been valid, but there is little evidence here that the 3 ASD- listeners (defined by mean z-score across conditions 6-8; see Figure 4.5) were, in fact, experiencing substantially more informational masking with the opposite-ear talker than were controls. Their mean shift in SRT between the baseline and condition 8 was 5 .4 dB, and for controls the shift was 4.2 dB. The 7 ASD+ listeners showed an even smaller shift of 1.8 dB. The pattern of performance with an opposite-ear speech masker also looked rather different than with a same-ear masker: only a single listener was an outlier in both.

Conditions 6-8 involved a male (opposite-gender) masker and speech content completely unrelated to the mCRM targets. A very similar pattern occurred with the female mCRM masking material, however, in spite of this being far more perceptually and linguistically similar to the target speech. There was only one ASD outlier in Condition 11 (opposite-ear female masker), and the listeners who were ASD- for conditions 6-8 showed similar shifts in SRT to controls (around 4 dB, on average, going from baseline to Condition 11). Of the three outliers with the same-ear female mCRM masker (Condition 10), two performed completely normally in Condition 11.

These findings are entirely consistent with the idea that a key factor affecting at least some ASD listeners is difficulty with auditory stream segregation and/or selective attention. Thus, when there is the strongest possible cue to target-masker segregation and selective attention (monaural opposite-ear presentation), even low-performing ASD listeners show typically low levels of informational masking (as quantified by the shift in SRT between baseline and opposite-ear masking). The same-ear target-masker data (e.g. Condition 7) continues to show a very distinct low-performing ASD subgroup like that in Experiment 1. Furthermore, when a second (opposite-ear) masking talker is added to this same-ear configuration (see Condition 9), the ASD- listeners tend, as predicted, to show rather smaller shifts in SRT than the other groups - and particularly the ASD+ listeners. Control mean SRTs worsened by 2.2 dB, ASD+ by 3.6 dB, and ASD- by only 1 dB. Whilst these figures are clearly marginal, it would be interesting to explore the effect of a second masking talker in a larger ASD sample. One might hypothesise that slightly increased susceptibility to a second masking talker could well occur if ASD+ listeners are exerting a high degree of top-down control to perform well in this task with a single masker .

On top of the evidence of a small ASD- subgroup who seem to be vulnerable to interference from background speech in typical listening configurations, it is very striking in Experiment 2 that so many of the ASD group performed poorly in the no-masker condition, and that a number remained outliers for the opposite-ear masker (when defined simply in terms of the SRT itself in that condition, rather than the shift between conditions). This is suggestive of an impairment which has as much to do with processing of the target as it does with any masking material involved, and which may be more prevalent in high-functioning ASD than the difficulties with stream segregation / selective attention found in same-ear masking speech conditions. It has already been suggested, on the basis of performance with vocoded speech in Chapter 3, that ASD listeners may be less proficient at processing degraded target speech - possibly due to the reduction in available acoustic-phonetic cues - and poor performance with speech at the threshold of audibility seems to reinforce this. Poor SRTs for speech in quiet and with an opposite-ear speech masker here in Experiment 2 may reflect the same type of reduced processing efficiency. The low levels of overlap between outliers in same-ear and outliers in opposite-ear / quiet conditions, therefore, suggests that impairments in more than one aspect of speech processing may affect everyday listening ability in ASD.

Chapter 5 Perception of interrupted speech: silence and noise

5.1 Introduction

As demonstrated in previous chapters, a subset of listeners with ASD appears to suffer from unusually degraded speech perception under a wide range of background noise conditions. A feature of the speech intelligibility deficit in this group is for performance to be far more atypical with some types of masking sound than with others, and this pattern of performance across various listening conditions has suggested some potential underlying factors. SRTs tend to differ far more from those of controls in conditions where intelligible speech maskers are presented than with similar non-speech maskers, and this effect is magnified when the target and masking talker have similar vocal characteristics. This strongly suggests that an impairment of auditory stream segregation and/or selective attention may be involved. Furthermore, amplitude-fluctuating background sounds disrupt performance more than steady noise maskers and, interestingly, the presence of pitch information in such fluctuating sounds seems to go some way towards alleviating the difficulty.

Two of the most likely underlying causes of poor SRTs in fluctuating noise are either a deficit in bottom-up temporal resolution or a deficit in top-down linguistic processing. As previously discussed, independent evidence of a non-speech deficit in auditory temporal resolution is limited, and the experiments described here also tend not to favour such an explanation. In Experiment 1, our ASD- group's performance was near-normal under at least some masking conditions involving fluctuating sounds (most notably the Japanese masking talker and the music masker) and a specific investigation of AM effects in

Experiment 2 showed little evidence of a low-performing subgroup at all (see Chapter 4). Whilst small sample sizes may have played a role in the latter case, it still seems likely that any effect of poor temporal resolution in ASD plays only a relatively small role in speech perception difficulties, particularly since three listeners in the AM/F0 investigation did still appear to be extreme outliers in the single condition involving interfering speech. Taking into account the fact that nearly all ASD listeners performed within the normal range when a *non-speech* fluctuating masker contained pitch information (the amplitude-modulated intonation contour in Chapter 4), it seems probable that somewhat higher-level aspects of speech processing in fluctuating background noise are more problematic than is temporal resolution per se. There may also, of course, be more than one component contributing to a top-down difficulty.

Perhaps the most obvious potential source of a top-down processing deficit in ASD is that these listeners may simply be less good at using context to reconstruct the intended message from a degraded speech signal. Pragmatic deficits form a core feature of the communication profile in autism, and in everyday listening contexts these are likely to become particularly important when it comes to disambiguating a half-heard comment on the basis of such factors as relevance to the current topic, probable intentions of the speaker, real-world knowledge, etc. The speech in noise tasks used in this set of experiments, however, involve target sentences which are thematically unrelated to each other and are isolated from any conversation-level context, thus minimising the effects of any deficit in higher-level pragmatics. Nevertheless, in Experiment 1 the ASD- subgroup did show impairment in performance across all masking conditions involving the standard sentence in noise target material (including the steady background noise masker) whereas when the more formulaic mCRM targets were presented, more masker-specific deficits were revealed.

This effect of target sentence type may well indicate that a deficit in using sentence-level context to recover the message is an important factor in everyday listening difficulties.

In the case of the mCRM, the form of the target sentences was fixed throughout the task, with the carrier phrase 'Show the dog where the [colour] [number] is' presumably providing few cues to the key words themselves beyond relatively small phonetic effects at the word boundaries. Performance of the ASD- subgroup with some types of AM masker in this simpler task was still significantly poor compared to controls, though, so any linguistically-based deficit in reconstructing the key words from glimpsed fragments here must surely have been occurring almost entirely at the word level. The possibility that a phonological processing deficit may affect at least some high-functioning autistic adults is, to say the least, somewhat unexpected, given that a diagnosis of Asperger's syndrome (which was the clinical diagnosis of all the participants involved here) specifically excludes any language delay. Nevertheless, it is not impossible that mild language deficits in some of these listeners could have been overlooked during development. The evidence from ERP studies reviewed in Chapter 1 suggests that at a neurophysiological level, at least, speech processing is atypical in ASD. It is also, perhaps, feasible that whilst phonological representations in the traditional sense may be intact, there may be some sort of a difference in the way phonetic cues are weighted which affects intelligibility only with degraded signal quality. The observation in experiment 1 that ASD- listeners performed significantly worse than controls when noise-vocoded mCRM target speech contained only a small number of channels might seem to support this hypothesis.

There is another possible explanation, however, which concerns the idea that a fluctuating noise itself creates some sort of distraction, reducing performance either by drawing ASD- listeners' attention away from the target material or by affecting their ability to fuse glimpsed fragments of the target sentence into a coherent auditory stream. One way to explore this possibility is by considering how ASD listeners perform in the absence of any background noise, when target sentence material is interrupted by silent gaps. As early as the 1950s researchers were interested in this form of degraded speech, and in a key paper Miller and Licklider (1950) described a series of studies into the effects of the frequency and duration of silent interruptions on word intelligibility. For neurotypical listeners, both these factors appear to be important, as is their interaction with the speech rate. When the duration of audible speech within a given period (also referred to as speech time fraction or speech duty cycle) is held equal (e.g. 0.5, where half of each cycle is audible speech and half silence) and only the interruption frequency is varied, performance tends to be poorer at low interruption frequencies (< 10Hz) than at higher ones (10-100 Hz), and worst at around 1 Hz. The monosyllabic words used by Miller and Licklider were around 600 ms long, so the duration of an interruption with frequency 1 Hz and speech duty cycle 0.5 would be 500 ms - enough for almost the entire target word to be completely obliterated should the interruption coincide with it. At higher frequencies, with the duty cycle at 0.5 there would be several glimpses of each word, which is apparently sufficient for near-perfect intelligibility. One question this raises in the current context, is whether ASD listeners perform in a similar way to controls when presented with interrupted speech of this kind. Deficits in either auditory stream fusion (across silent gaps) or in top-down linguistic processes might both lead to a difference in the threshold speech duty cycle required for adequate target perception.

A second interesting feature of research into interrupted speech in neurotypical adults is what Miller and Licklider (1950) referred to as the 'picket fence effect' - more frequently called the phonemic restoration effect - whereby introducing noise into the gaps between bursts of interrupted speech gives the illusion of continuous speech. Miller and Licklider's analogy refers to seeing a landscape through a picket fence - the fence-posts obscure the view at regular intervals, but the landscape is perceived as continuing behind them. There are mixed reports as to whether this perceptual effect leads to any measurable improvement in interrupted speech intelligibility - Miller and Licklider themselves found no significant effect with the addition of noise - but more recent work suggests that phonemic restoration is more likely to occur with target material consisting of sentences or connected speech, rather than single words. Powers and Wilcox (1977), for instance, presented sentences interrupted at a number of different interruption frequencies, with a speech duty cycle of 0.5, and found significantly higher intelligibility when interruptions were filled with noise than when they were silent, across the rates tested. Further investigations by Verschuure and Brocaar (1983) also demonstrated significant improvements in interrupted sentence intelligibility with noise, and noted that the largest improvements occurred for listeners who were relatively poor in the silent interruptions condition. Referring back to work by Huggins (1964) the authors suggest that the process of interrupting a speech signal leads to distortion at the onset and offset of the 'on' sections, and that this distortion may itself be perceived as (inaccurate) speech information. On this account, the addition of noise improves intelligibility by masking the misleading distortion.

Mechanisms underlying the phonemic restoration effect in neurotypical listeners are evidently not yet fully understood; nonetheless it remains interesting to explore whether ASD listeners show a similar pattern of performance to controls across the two types of interrupted speech. If noise in itself is particularly aversive or distracting for some ASD listeners, it is possible that filling silent interruptions with noise might actually have a detrimental effect on their performance. In the study described here, therefore, sentence reception thresholds were measured with both silent and noisy interruptions, and the relationship of interrupted speech performance to performance with speech in modulated background noise was also investigated.

5.2 Method

Listeners completed a test of sentence intelligibility with two main conditions:

Condition 1 sentences with silent interruptions (square gated at 4 Hz, with 5 ms ramps)

Condition 2 sentences interrupted by noise
(as in 1, but adding noise to fill the silent gaps).

In these conditions, the duty cycle of the speech (i.e. the ratio of speech duration to the interruption in each period) was varied adaptively to establish a quasi 'speech reception threshold' (SRT-dc) for each condition.

A subset of listeners also completed condition 3, in which sentences were presented in background noise which was square-wave modulated at 4 Hz, with the speech duty cycle fixed at 0.3. Here, the signal-to-noise ratio was varied adaptively to find a standard SRT.

5.2.1 Stimuli

Target stimuli for all three conditions were sentences recorded by a female talker of Southern British English, and were taken from lists 6-10 of the ABC corpus. These lists were chosen to avoid repetition of those used in Experiment 1, and the lists were counterbalanced across conditions and subjects within Experiment 2 in attempt to control for any difference in sentence difficulty. Sentences contain 5 key words and listeners scored a point for each word correctly repeated. The speech-shaped noise used in conditions 2 and 3 was the same as that presented in Experiment 1, with a long-term average spectrum approximating Byrne's measurements for combined male and female voices (Byrne et al., 1994).

5.2.2 Participants

The same 20 participants previously described in chapter 4 also completed the tasks described here: 10 young adults with ASD, and 10 age- and IQ-matched neurotypical controls. Nine of the ASD listeners and four of the controls had also taken part in Experiment 1. For ease of reference, group age and IQ descriptives are repeated in Table 5.1 below.

Table 5.1 Participant group statistics: IQ, ASQ and age

	Group	N	Mean	Std. Deviation
Performance t-score	Control	10	61.9	4.2
	ASD	10	62.2	7.0
BPVS raw score	Control	10	155.5	6.9
	ASD	10	149.2	12.3
ASQ	Control	10	13.0	4.1
	ASD	10	36.1	6.6
Age	Control	10	25.9	4.5
	ASD	10	26.2	5.4

5.2.3 Procedure

Participants completed the tasks described here as a discrete session within Experiment 2 (which also involved two further sessions consisting of the mCRM listening tests described in Chapter 4 and the cognitive tests in Chapter 6). The order in which sessions were completed was counterbalanced across participants, as was the order of conditions within each session.

Listeners were once again seated in a sound-protected booth and stimuli were presented on a Samsung Q45 laptop, using an external Edirol Cakewalk UA-1SG USB sound card, via Sennheiser HD25 II closed supra-aural headphones. Two blocks of sentences (20 sentences per block) were presented in each condition. Listeners were requested to repeat back each sentence as accurately as possible to the researcher seated nearby. Scores were manually entered onto the computer, and an adaptive tracking procedure was used to estimate the speech reception threshold (SRT-dc or SRT; 50% correct) in each condition.

In the two interrupted speech conditions, the duty cycle of the speech was varied to estimate an SRT-dc. This was implemented in Matlab (Mathworks Inc.), by manipulating a modulating wave on each trial to achieve the desired duty cycle. Starting with a duty cycle of 0.9 (i.e. 225 ms of speech and 25 ms of silence or noise in each 250 ms period), the initial step size was 0.12 (or 30 ms), reducing to a final step size of 0.04 (10 ms). For Condition 2, where noise was added to the interruptions, the SNR was 0 dB throughout. The phase of the modulating wave was varied randomly from trial to trial. Figure 5.1 shows spectrograms of a sample sentence with silent and noisy interruptions. In Condition 3 (sentences in background noise), the speech duty cycle was held constant at 0.3 (equivalent to 75 ms of uninterrupted speech in each period, and 175 ms of speech in background noise), and the SNR was varied adaptively (initial step size 10 dB and final step size 2 dB).

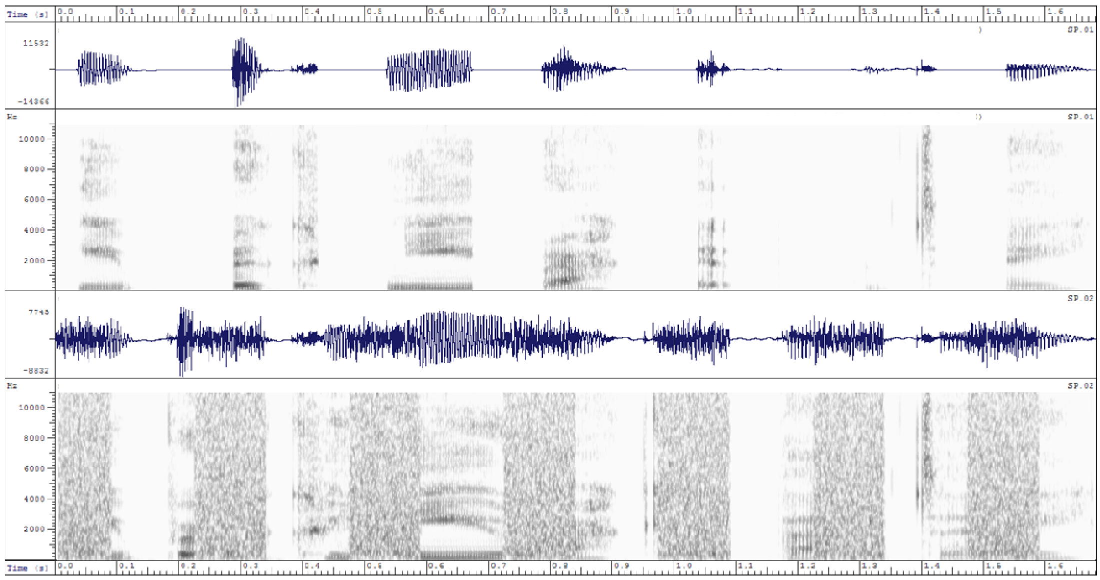
A mean SRT calculated from the two blocks per condition completed by each listener was entered into the final analysis.

Figure 5.1 Waveform & wideband spectrograms of an interrupted sentence

i) silent interruptions ii) noisy interruptions

i)

ii)



'He broke a new racquet that day' Speech duty cycle in both cases = 0.54

5.3 Results

5.3.1 Interrupted speech (Conditions 1 and 2)

Control listeners did not show any significant improvement in speech intelligibility (in terms of duty cycle required for 50% correct performance) with noise-filled interruptions compared to silent interruptions (see Figure 5.2 and Table 5.2). Unexpectedly, their mean thresholds in these two conditions were almost identical: 0.45 with silence and 0.44 with noise. These data do not, therefore, appear to indicate that perceptual restoration was taking place with the addition of noise. One possibility is that our method of adaptively varying the duty cycle may have obscured any such effect, and that comparing performance accuracy across conditions at a fixed duty cycle, as is typical in the literature, might have been more informative. Pilot data had indicated, however, that listeners vary so widely in their ability to decode interrupted speech that it would have been difficult to find a single duty cycle which would suit all listeners without leading to substantial floor or ceiling effects.

Figure 5.2 Group boxplots and individual data points for the interrupted speech task

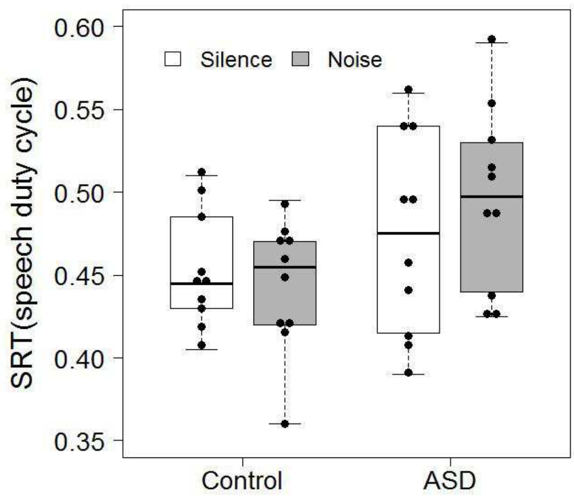


Table 5.2 Descriptive statistics: Interrupted speech

	Group	Mean SRT (duty cycle)	Std. Deviation	N
Condition 1 Silent interruptions	Control	.45	.03	10
	ASD	.47	.06	10
Condition 2 Noisy interruptions	Control	.44	.04	10
	ASD	.50	.06	10

On average, the ASD group tended to perform slightly worse than controls across both interrupted speech conditions, but a mixed ANOVA (2 conditions x 2 groups) showed a group effect which was only approaching significance (F 1, 18 = 3.45, p = 0.08). There was also no significant effect of condition (F 1, 18 = 0.52, p = 0.48) and no condition x group interaction (F 1, 18 = 2.81, p = 0.11). The experiment clearly lacked power, but as with previous listening tasks there was more individual variability amongst the ASD group than controls, and it was again evident that some ASD listeners were significantly impaired whilst others performed within the normal range. For interest, z-scores based on the control mean and standard deviation (corrected for control outliers - one in each condition) were calculated, and it appeared that once more a small subset of ASD participants were significant outliers (see Figure 5.3).

Figure 5.3 Boxplots showing z-scores averaged across 2 interrupted speech conditions

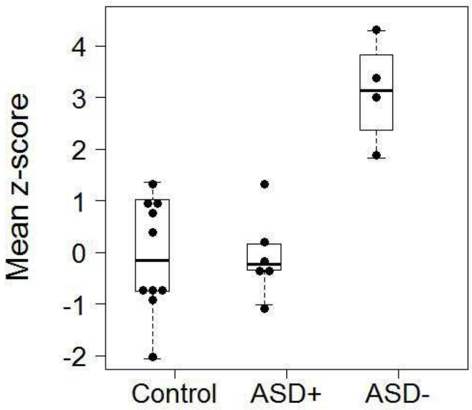
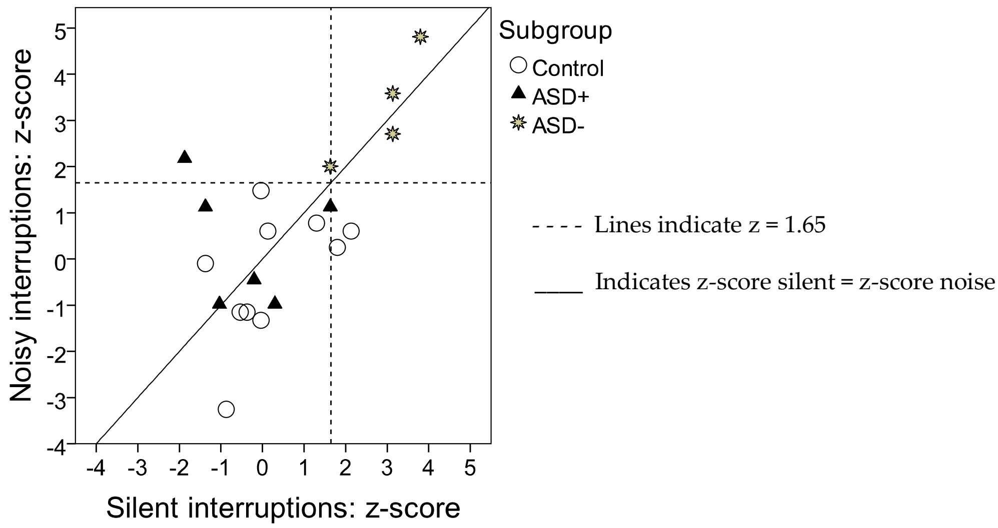


Table 5.3 Descriptive statistics: interrupted speech performance by subgroup

	Subgroup	Mean	Std. Deviation	N
Condition 1: Silent interruptions SRT (duty cycle)	Control	.45	.03	10
	ASD+	.43	.04	6
	ASD-	.53	.03	4
Condition 1: Noisy interruptions SRT (duty cycle)	Control	.44	.04	10
	ASD+	.46	.04	6
	ASD-	.55	.03	4
Mean z-score across interrupted speech conditions	Control	-.06	1.11	10
	ASD+	-.04	.80	6
	ASD-	3.10	1.03	4

Taking a mean z-score averaged across the 2 interrupted conditions there were four low-performing ASD listeners ('ASD-') , all with scores above a cut-off of 1.65 (range of z-scores 1.82 - 4.31). However, whilst all four members of this new ASD- subgroup scored well outside the normal range in the noisy interruptions condition (all $z > 2$), they were not all bivariate outliers (see Figure 5.4), since one scored just within the normal range for silent interruptions ($z = 1.63$). All the others had z-scores > 3 with silent interruptions.

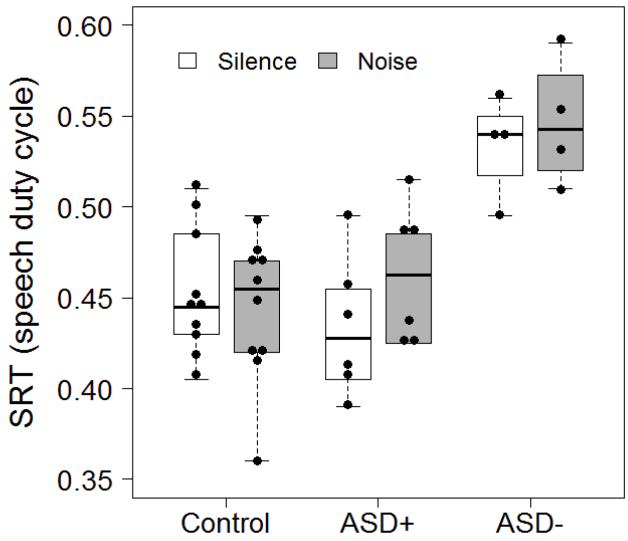
Figure 5.4 Scatterplot showing z-scores for the interrupted speech conditions



Three of the ASD- listeners, had also belonged to the original ASD- group in Experiment 1, and the fourth was a newcomer for Experiment 2, which suggests the presence of a genuine deficit in speech perception - at least where it comes to degraded speech signals - amongst some high-functioning autistic adults. On that basis, whilst running further analyses on the same small data set here was not really statistically acceptable, a mixed ANOVA (3 subgroups x 2 interrupted speech conditions) was carried out purely for exploratory purposes. The results should be treated with extreme caution until they can be replicated, but this did confirm a significant main effect of subgroup ($F_{2, 17} = 15.4$, $p < 0.001$), with

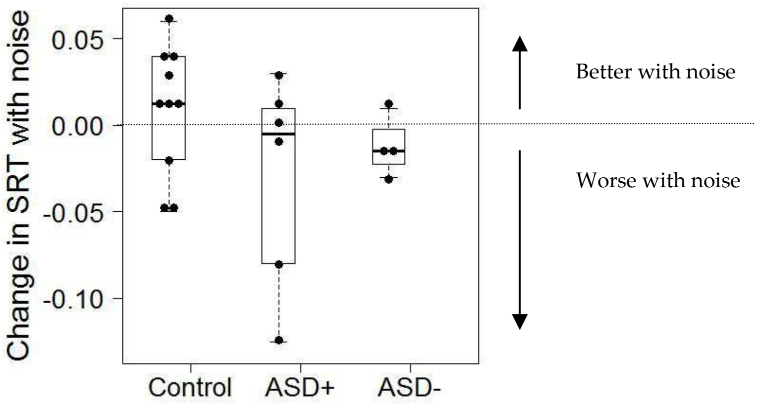
ASD- performing significantly more poorly than ASD+ as well as control listeners (Tukey's HSD, both $p < 0.001$) (see Figure 5.5). There was, however, still no significant effect of condition ($p = 0.30$) or interaction between subgroup and interruption condition ($p = 0.24$).

Figure 5.5 Subgroup boxplots showing individual data points for interrupted speech



It is clear from the above that our listening task was not effective for measuring perceptual restoration effects in interrupted speech, if indeed any such effects were occurring. Closer examination of the difference scores does suggest that there may have been a slight trend amongst control listeners towards better performance with noisy interruptions than silent ones. 7 of the 10 controls performed better with noise than silence, whereas amongst the ASD listeners the opposite seemed to apply and 6 of them performed worse with noise whilst one showed no change. Overall, only 4/10 ASD listeners improved with noise, and only 1/4 of the ASD- group did so. These numbers are small, but the general trend remains interesting - not least because the ASD- listeners' difference scores were amongst the smallest measured here, whereas two of the ASD+ participants,who had performed particularly well with silent interruptions showed much the largest increases in threshold with noise (see Figure 5.6).

Figure 5.6 Boxplots showing difference in threshold duty cycle (Silent - Noisy Interruptions)



5.3.2 Speech in AM background noise (Condition 3)

Subgroups based on the interrupted speech z-scores were also used to explore the standard SRT data from the test of speech in AM background noise (Condition 3). Only 8 ASD listeners completed this condition, so the sample size was 5 for ASD+ and 3 for ASD-. Perhaps unsurprisingly, the univariate ANOVA showed a highly significant effect of group ($F_{2, 15} = 9.63$, $p = 0.002$). Once again ASD- performed significantly worse than both other groups (both $p < 0.01$), whilst there was no significant difference between ASD+ and controls (Tukey's HSD, $p = 0.50$) (see Figure 5.7 and Table 5.4).

Figure 5.7 Boxplots showing subgroups' SRTs for sentences in background noise

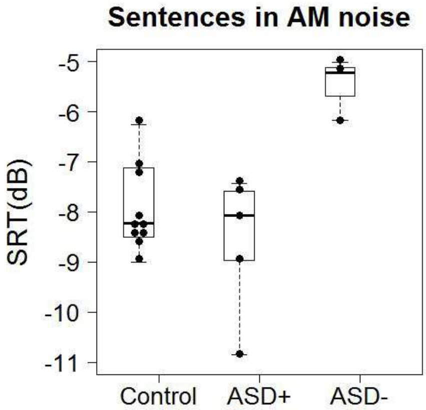
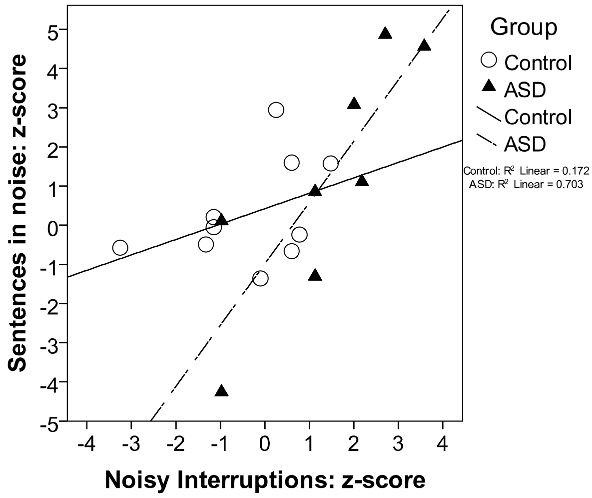


Table 5.4 Sentences in AM noise: Descriptive statistics

	Subgroup	Mean	Std. Deviation	N
Sentences in AM noise SRT (dB)	Control	-7.94	0.85	10
	ASD+	-8.58	1.41	5
	ASD-	-5.47	0.61	3
Sentences in AM noise (z-scores)	Control	0.29	1.33	10
	ASD+	-0.70	2.20	5
	ASD-	4.17	0.96	3
Sentences with noisy interruptions (z-scores)	Control	-0.33	1.40	10
	ASD+	0.50	1.41	5
	ASD-	2.77	0.79	3

The relationship between performance on the interrupted speech condition with noise (Condition 2) and the sentences in AM background noise (Condition 3) appeared to be relatively strong for most listeners (see Figure 5.8). When all participants who completed condition 3 were included in the analysis (ASD n = 8; Control n = 10), the correlation between the two conditions reached significance for the ASD group (Spearman's rho = 0.91, $p = 0.002$, $r^2 = 0.70$), but not for the Controls (Spearman's rho = 0.32, $p = 0.37$, $r^2 = 0.17$). This apparent difference in the pattern of relationships across task may well, however, reflect the wider spread of scores and higher number of extreme outliers amongst the ASD group, which makes it difficult to give a meaningful interpretation of these results.

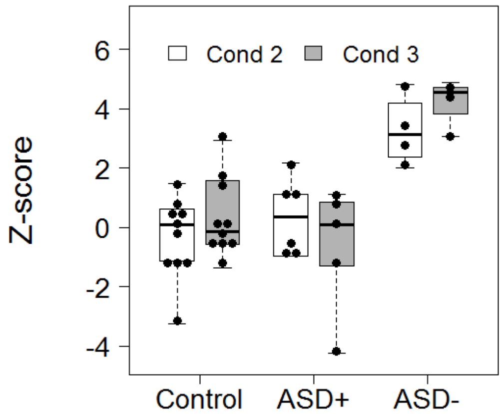
Figure 5.8 Scatterplot showing the relationship between interrupted speech and speech in noise



A mixed ANOVA (2 conditions x 2 main groups) showed no significant main effects of condition (AM noise / Noisy interruptions) or group, and no significant condition x group interaction. Repeating this mixed ANOVA with the 3 subgroups (Control n = 10, ASD+ n = 5, ASD- n = 3), however, did show a marginally significant condition x subgroup interaction ($F_{2, 15} = 3.68$, $p = 0.05$) as well as the expected main effect of subgroup ($F_{2, 15} = 9.73$, $p = 0.002$). Again it is difficult to draw any real conclusion from such small samples, but the interaction does appear to be driven by a slightly smaller deficit in the ASD- group for speech with noisy interruptions (Condition 2) than for speech in AM background noise (Condition 3) (see Figure 5.9 and Table 5.4). One factor contributing to this was that the ASD- listener who performed worst with noisy interruptions did not complete the speech in

AM noise condition. Nevertheless the other 3 ASD- listeners' z-scores were all substantially worse (by around 1 z) for speech in noise than speech with noisy interruptions. This seems to imply that difficulties in reconstructing speech from glimpsed fragments, and difficulties with perceiving speech in simultaneous noise, may have an additive impact on speech intelligibility in adverse conditions for at least some ASD listeners.

Figure 5.9 Boxplots comparing z-scores for speech with noisy interruptions and speech in background noise



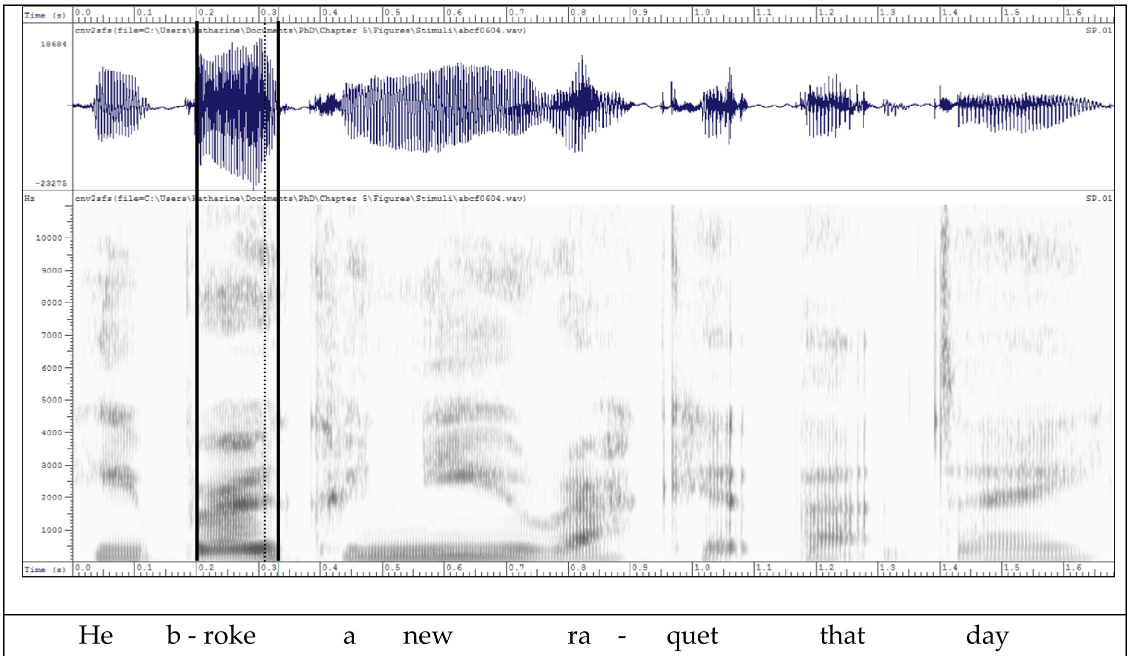
Cond 2: Noisy interruptions

Cond 3: AM background noise

5.4 Discussion

Probably the most important aspect of the results described above is the evidence that, for some ASD adults, sentence intelligibility is significantly below the norm even when the signal is degraded by means other than adding a masking sound. The ASD- subgroup in this study required significantly longer segments of audible speech and shorter silent gaps in each 250 ms period, in order to achieve 50% accuracy. Their mean SRT-dc for speech with silent interruptions occurring at a rate of 4 Hz was 0.53 - equivalent to four 132.5 ms glimpses of audible speech per second. This adds up to around 20 ms more speech per glimpse (or 80 ms more per second) than for controls, whose mean SRT-dc was 0.45 (i.e. four 112.5 ms glimpses per second). Figure 5.10 shows a full (uninterrupted sentence) with markers to indicated a randomly-selected 132.5 ms 'glimpse'. In this glimpse, almost the entire voiced portion of the word 'broke' is audible.

Figure 5.10 An uninterrupted sentence with markers to indicate mean ASD- glimpse length



Solid lines indicate the boundaries of a 132.5 ms glimpse (based on ASD- mean SRT-dc)
Dashed line indicates the end of a 112.5 ms glimpse (based on control mean SRT-dc)

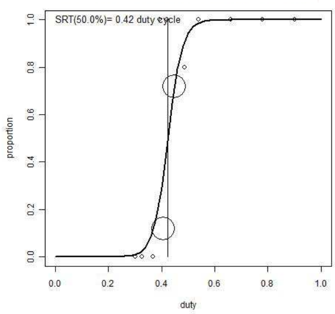
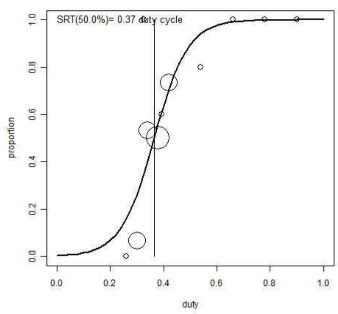
Whilst a difference of 20 ms in length does not seem much, and the same phonemes remain audible in a 112.5 ms glimpse, the dashed marker in Figure 5.10 indicates how the ends of the formant transitions going into the plosive /k/ are truncated in this shorter glimpse. This gives a considerably more disjointed percept, as well as meaning that an acoustic cue to the word-final consonant is lost. What is impossible to tell from the current study, however, is whether the ASD- listeners needed longer glimpses in order to help them fuse the sentence together across interruptions, or whether they needed more phonetic information than controls in order identify key words correctly. It would certainly be interesting to investigate how this group might perform with interrupted single words as in Miller and Licklider's original study.

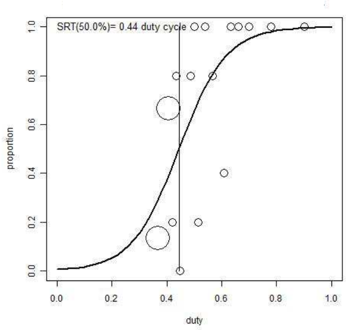
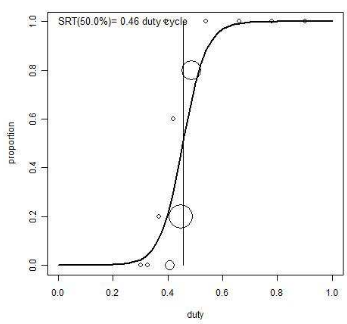
It was disappointing to find little evidence that control listeners experienced phonemic restoration when noise was added to the silent interruptions. If the main factor affecting intelligibility with silent interruptions for the ASD- group had been a difficulty with auditory stream fusion across the gaps, and if the task had been optimised for measuring phonemic restoration, one might even have expected to see large increases in their performance with the addition of noise. Unfortunately, in the absence of a clear effect amongst controls, little can be concluded from the fact that most of the ASD- listeners actually had worse thresholds when noise was added.

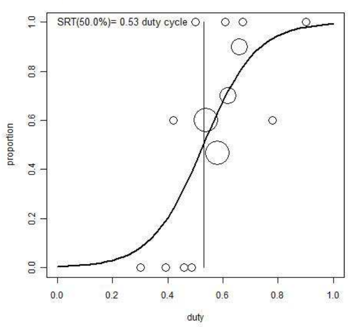
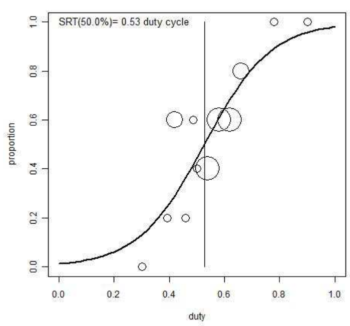
In spite of the fact that there is no precedent in the literature for using an adaptive method to establish an SRT for interrupted speech in terms of duty cycle, there are some indications from these results that the basic approach was not entirely unfit for the current purpose. Amongst the control group there was a reasonable spread of SRT-dc for the silent interruptions condition (range of SRT-dc: 0.41 - 0.51). Moreover, with noisy interruptions the best-performing control listener had an SRT which was considerably lower (SRT-dc = 0.36) and the worst-performing ASD- listener's SRT was considerably higher (SRT-dc = 0.59). This indicates both that the task was sensitive to individual variability and that floor and ceiling effects were not a serious problem. Furthermore, when psychometric functions are estimated from the data, there is no evidence of any non-monotonicity (see Figure 5.11, overleaf).

Figure 5.11 Examples of individual psychometric functions from each group

<table>
<tr><td></td><td>Silent interruptions</td><td>Noisy interruptions</td></tr>
<tr><td>Control listener

Better with noise</td><td></td><td></td></tr>
<tr><td>ASD+ listener

Worse with noise</td><td></td><td></td></tr>
<tr><td>ASD- listener

No change with noise</td><td></td><td></td></tr>
<tr><td></td><td>x-axis shows: duty cycle
y-axis shows: % correct</td><td>y-axis reference line shows duty cycle at which listener scored 50% correct</td></tr>
</table>

At least one other group of researchers has used almost identical sentence material to investigate interrupted speech perception in the past (Nelson & Jin, 2004). Whilst Nelson and Jin's study did not use an adaptive procedure, and only a limited number of fixed speech duty cycles was presented, sentence intelligibility (percent correct) was measured at a range of interruption rates which included 4 Hz, as in the current experiment. At 4 Hz, and with a speech duty cycle of 50% (the closest to our control group's mean SRT-dc of 0.45), normally-hearing listeners appeared to be scoring around 40% correct on average (see figure 6, Nelson & Jin, 2004). Since the current control group was able to score 50% correct with a duty cycle which was, on average, slightly lower than the 0.50 used by Nelson, it would appear that our listeners performed rather better. The difference is relatively small, however, and one explanation may be that our listeners were somewhat more experienced with the procedure and with listening tests more generally. Nelson and Jin only included participants who had not previously been involved in similar studies of theirs, whereas many participants in the study described here had also been involved in Experiment 1, which presented sentences from the same corpus using a very similar procedure to investigate SRTs in background noise. Experience does appear to be a factor which can affect the intelligibility of speech with silent interruptions, as Verschuure and Brocaar (1983) described modest learning effects amongst their participant group. A more recent training study has also shown similar effects for interruptions containing noise (Benard & Baskent, 2013).

This may even go some way towards explaining why there was little evidence of phonemic restoration in the current study, because amongst Verschuure and Brocaar's participants, it was those with no prior experience of psychophysical testing who showed marked improvements with the addition of noise to the interruptions.

One final possible factor in the lack of phonemic restoration effects here centres on a relatively small detail of the stimuli, which unfortunately may have proven to be crucial. In several of the other studies described earlier, when noise was added to the interruptions it was substantially louder than the speech. Nelson and Jin (2004) used - 8 dB SNR, and Verschuure and Brocaar (1983) used -15 dB SNR, whereas here our SNR for noisy interruptions was 0 dB. Verschuure also comments, on the basis of unreported pilot data, that a 'full continuity perception' is obtained at SNRs below -10 dB. Powers and Wilcox (1977) measured across a range of SNRs from +9 to -15 dB, and did find significant improvements with noise at many SNRS, including one in which the speech level was as high as + 3 dB SNR. Notably, however, the effects at lower SNRs were considerably larger, and the difference at 0 dB SNR did not reach significance. Returning to Miller and Licklider's picket fence analogy, it seems possible that a clearly perceptible difference in amplitude between the signal (landscape) and noise (fence-posts) is an important factor in supporting auditory stream segregation and thereby plays a role in the continuity illusion.

Two of the listeners in the current study who showed something resembling the opposite of a phonemic restoration effect - by far the largest increases (worsening) of SRT-dc when noise was added to the interruptions - belonged to the ASD+ group. In fact, they were actually the highest performers of all with silent interruptions. Verschuure and Brocaar (1983) found a similar tendency amongst their neurotypical listeners for the highest-performers with silent interruptions to be adversely affected by the addition of noise. However, in their discussion they focused more on the fact that it was the worst performers who showed the largest improvements. They accounted for this by saying that good performers with silent interruptions were already making maximum use of the available speech information and therefore could improve no further with noise (since this introduced no new information), whereas noise in some way enabled the poor performers to make better use of the available information. This raises two questions regarding the current set of results, neither of which are easily answered. Firstly, why is it that noise does not appear to help the ASD- listeners in a similar way, even though they are clearly low-performers with silent interruptions? And secondly, why exactly, even in Verschuure's study, did the high-performers with silent interruptions actually get worse when the noise was added? Could it be that when a high level of executive control is being exerted in order to deal with a degraded signal, any type of extraneous sound interferes with top-down control? This might explain the tendency of our ASD group as a whole to perform worse with noise: all of them - whether successfully or not - may need to exert more conscious effort than neurotypical listeners to deal with an interrupted speech signal, and the introduction of noise may undermine top-down control.

The main finding of this experiment was that a subset of the ASD group performed poorly with interrupted speech, even when no noise was present. This is significant because, in the absence of any simultaneous noise, the glimpses of the speech signal should have been equally audible to all listeners. A deficit with silent interruptions probably then depends upon an inability to fuse together the glimpsed fragments into a coherent stream, or upon a deficit in linguistic processing which could be occuring at any level from that of the phoneme up to the sentence. A priori, both of these explanations seem quite probable. Theories of weak central coherence in ASD and the evidence of a detail-focused visual processing style suggest that auditory stream fusion might function differently in these listeners, whilst differences in electrophysiological response to speech sounds suggest that aspects of language processing may be atypical even in high-functioning adults. Although speech reception thresholds for speech interrupted by noise appeared to be highly-related to those with a modulated background noise, it is interesting that the ASD- listeners tended to be even more impaired in the latter situation. This seems to imply not only that they have difficulty taking full advantage of the glimpses of speech they hear, even when those glimpses are fully audible, but also that in typical everyday listening situations where there is often likely to be some level of noise simultaneous with the speech, the glimpses themselves become even less useable.

Chapter 6 Conclusion

The main aim of the research described in this thesis was to investigate factors contributing to atypical speech perception in autism spectrum disorder. As outlined in Chapter 1, even the highest-functioning adults, with normal IQ and no history of language delay, report problems with speech comprehension in background noise. Such difficulties appear to be a common symptom of high-functioning ASD, and their relationship to the core areas of impairment in this condition is likely to be complex. It seems highly probable that difficulties with communication and social interaction are likely to play a role in poor speech reception in noise, and also that poor speech reception in noise is likely to further undermine attempts at communication and social interaction. This vicious circle can have very real effects on social participation and psychosocial well-being, causing affected individuals actively to avoid gatherings where they are likely to encounter high levels of background noise (Madriaga, 2010). In spite of the importance of this subject, however, little research has so far been undertaken to explore the nature and underlying causes of impaired speech perception in noise. Two main studies have been published: one looking at sentence perception in noise (Alcantara et al., 2004) and one investigating single word perception in noise (Groen et al., 2009), both of which were discussed in Chapters 1 and 2. These studies were alike in only presenting target stimuli in a fairly limited range of background sounds, and in focusing mainly on non-speech maskers. Both revealed relatively mild deficits in speech reception thresholds amongst their high-functioning ASD listeners.

The series of listening tasks described in this thesis was designed to measure SRTs for more than one type of target sentence material, and across a much broader range of masking conditions and listening configurations than previously. Perception of target speech which was itself degraded, through loss of spectral detail or interruptions to the signal, in the *absence* of any masking sound, was also assessed. As discussed in Chapter 1, the literature on auditory processing in ASD suggests a number of hypotheses as to why processing speech in noise might be particularly difficult for this group, and so the work described here was primarily exploratory in nature. It was hoped that by measuring performance across a number of different sets of masking sounds, it would become possible to identify which - if any - specific type(s) of masker cause unusually high levels of interference in ASD, and from there to formulate somewhat less speculative hypotheses about the underlying causes. To some extent this objective was successfully achieved.

Alcántara and colleagues (2004) were the only group previously to have measured speech reception thresholds in a speech masking condition, and then only with one type of speech masker (alongside 3 non-speech masking noise conditions). In Experiment 1 here, a total of 6 main types of speech masker were presented as core mCRM conditions (see Chapter 2): foreign language speech, task-irrelevant opposite-gender English speech, vocoded English speech, task-relevant opposite-gender English, task-relevant same-gender speech and task-relevant same-talker speech. Two non-speech noise maskers were also included in this first set of conditions. Although Alcántara did find an overall main effect of group in his statistical analysis, and the difference in mean SRT between the ASD and control groups for the speech masker was also significant, this difference was only small (approximately 3 dB), and very similar to the between-group difference with a non-speech fluctuating noise. The results of Experiment 1 here, by contrast, presented a very different profile.

Amongst the ASD group in Experiment 1 (unlike controls) there was an extremely wide spread of SRTs in most of the speech masking conditions. This type of variability - whilst not entirely unexpected when testing participants with a disorder as heterogeneous as ASD -

was far in excess of anything observed by Alcántara. It was also a little surprising, given that this was a study focusing purely on high-functioning adults amongst whom full-scale IQ scores ranged from 99 -145. When individual SRTs were explored more closely using scatterplots, it became clear that around half of the Experiment 1 ASD listeners were in fact performing fairly consistently within the normal range across conditions, whereas a second subgroup tended to have extremely poor SRTs across the board. By calculating z-scores for each condition, based on the control mean (corrected for control outliers), it was confirmed that whilst many members of this ASD- subgroup performed within the normal range (z = +1.65 to -1.65) in steady noise and one or two other conditions (see below), when the masker was intelligible English speech, the majority were significant outliers. Moreover, when the masking speech was extremely similar to the targets (i.e. competing female mCRM sentences), their deficits became even more striking, with mean differences between ASD- and controls as large as 15 dB. Impairments in SRT on this scale seem, at face value, far more commensurate with the difficulties subjectively reported by adults with ASD than the smaller-scale deficits reported in either of the previous studies (Alcántara et al., 2004; Groen et al., 2009).

This novel finding of a bimodal distribution in the performance of ASD listeners, predominantly for conditions involving intelligible speech maskers, points to the strong likelihood of an attentional component to the difficulties this high-functioning group report with everyday listening. Further to that, mean z-score across listening conditions was strongly correlated with full-scale IQ, in both the ASD and control groups. A regression model predicting mean z-score on the basis of group and FSIQ found that both were highly significant factors, each accounting for similar amounts of the variance in mean z. The interaction between group and FSIQ was not, however, significant. One might speculate, therefore, that whilst all the ASD listeners here could have a greater underlying susceptibility to interference from masking speech than controls, IQ may act as a protective factor allowing the most able individuals (perhaps through a combination of deliberate listening strategies and top-down control) to perform within the normal range. Participants were not individually matched for IQ across groups, although on average the difference between groups was not significant. Nevertheless, as Figure 2.7 demonstrated, the mean z-scores of some ASD+ listeners even at the very top end of the IQ scale were considerably worse than those of controls with similar IQ.

Experiment 1 was not originally planned with a multiple-case studies approach in mind, and the statistical analyses would have been more straight-forward - and possibly more convincing - if it had. The ASD subgroups were defined a posteriori, on the basis of their mean z-scores across all 11 conditions described in Chapter 2, of which 8 involved the modified Coordinate Response Measure, and 3 were from the Sentence in Noise task. This meant that when the main 3 (subgroup) x 8 (mCRM condition) ANOVA was carried out, although the z-scores which contributed to grouping were based on more than just those 8 conditions, nevertheless there was a degree of overlap. Ideally, the groups would have been defined on the basis of one or more measures, selected in advance, which could then have been analysed independently of the main experimental conditions. In this respect, the data from a fresh set of listening conditions completed by the same group of listeners in Experiment 1.2 (described in Chapter 3) was encouraging. The original ASD- group continued to perform significantly poorly across many of the new mCRM masking conditions, and once again the most extreme outliers appeared in the conditions with masking speech. Not all the members of this ASD- group returned to participate in Experiment 2, but where they did, nearly all of them also performed significantly poorly with speech maskers. The ASD+ group appeared similarly stable.

Sample sizes in these studies were small (N=13 at best), and it is possible that there was something highly unusual about this particular selection of ASD participants. Whilst the subgroups appeared to remain stable across conditions and tasks over a period of a year, it could be that the individuals in this particular sample happened to cluster at two extremes of performance purely by chance. For this reason, it would be highly desirable to replicate the study with a larger sample, and perhaps to allocate the ASD participants to subgroups on the basis of one key condition such as the same-gender masking talker (which revealed the largest deficits in SRT here). It would also be useful to gather more detailed data on specific aspects of cognition which might contribute to the difference between ASD subgroups. Working memory, for example seems to play an important role in speech in noise tasks like the mCRM (which does not require a spoken response), as participants could often be observed to repeat back the stimulus to themselves, seemingly to help select out or reconstruct the correct key word. Unfortunately, a working memory test included as part of Experiment 2, but not reported here, had ceiling effects that made the data uninterpretable.

In light of the theory of Enhanced Perceptual Functioning in ASD (Mottron et al.,2006), and the evidence of atypical pitch processing in at least some listeners on the spectrum, it would also be extremely interesting to explore whether increased susceptibility to interference from speech maskers might correlate with pitch discrimination or pitch labelling ability. Behavioural evidence suggests that enhanced pitch processing tends to occur more commonly in lower-functioning individuals (Bonnel et al., 2010), and whilst all the listeners in the current study were high-functioning, nevertheless susceptibility to masking did appear to correlate with IQ. Given that the subject of pitch processing in ASD is far from perfectly understood as yet, it seems possible that individual differences in this area might relate to different patterns of performance with masked speech. As previously discussed, one idea about how this might occur is that hyper-reactivity to a relatively low-level auditory feature such as pitch could in some way function to undermine auditory selective attention to a more complex global-level target such as speech. The main research so far to evaluate pitch processing and speech in noise in ASD has involved electrophysiological rather than behavioural measures, and considered pitch processing of speech stimuli only (Russo, Nicol, Trommer, Zecker, & Kraus, 2009). In this experiment, ASD children's auditory brainstem response (ABR) actually showed degraded representation of the frequency content in speech (in quiet), and they showed significantly greater degradation of the speech ABR than controls when speech was presented in background noise. The relationship between non-speech pitch processing in ASD and speech intelligibility in noise has yet to be investigated.

The decision to use the mCRM task in these experiments was made partly on the basis that it minimises cognitive and linguistic demands on the participant, but also because it maximises informational masking (Brungart, 2001). The fact that ASD- listeners performed particularly poorly with *speech* maskers in this task, and worst of all with same-gender mCRM masking speech - whilst showing far smaller deficits with *non-speech* maskers - is a strong indication that they were unusually susceptible to informational masking. This could reflect either inefficient auditory stream segregation, impaired auditory selective attention, or both. When a strong spatial cue to stream segregation was introduced, deficits in SRT for the same-gender masker were greatly reduced (see Chapter 3), and when target and masker were presented monaurally to opposite ears (perhaps the strongest cue to auditory selective attention there could be) even low-performing ASD participants showed similar (small) amounts of informational masking to controls (see Chapter 4). All of this suggests that informational masking is likely to be a key factor in everyday listening difficulties.

This pattern of abnormal susceptibility to speech-on-speech masking makes sense in terms of what we know about atypical patterns of attention to speech from early infancy onwards in ASD, as reviewed in Chapter 1. One particularly relevant piece of evidence comes from the ERP study carried out by Whitehouse and Bishop (2008), in which the P3a orienting response for speech stimuli was missing in ASD children in a passive listening task *unless* they were specifically directed to attend to the speech. It seems very likely that if automatic orienting to speech is so atypical during language development, the whole process of selective attention to speech may be vulnerable in adulthood. Another possibility is that the increased perceptual capacity observed in some studies of *vision* in high-functioning ASD (Remington et al., 2009) may also apply in the auditory domain, contributing to higher levels of interference from background speech. An interesting direction for future research would be to look at distractor processing across modalities.

In addition to the evidence of two distinct patterns of listening ability within the ASD group, which suggests there is a deficit in selective attention to speech amongst around half their number, this research produced two further key findings. The first concerns the hypothesis, initially proposed by Alcántara, that difficulties with speech perception in noise may primarily reflect impaired dip listening in ASD, due to a temporal processing deficit. This idea was based on the evidence of significantly poor performance for conditions involving amplitude-modulated maskers in his 2004 study. The data described here, however, do not provide particularly strong support for this theory. In Experiment 1 there were several conditions involving fluctuating maskers in which very few ASD listeners performed significantly poorly, and in which the mean difference in SRT between the ASD- subgroup and controls either did not reach significance (Japanese speech masker) or was only marginally significant, and driven mainly by one listener (music masker).

When SRTs for amplitude modulated maskers were explicitly investigated in Experiment 2, it once again appeared that nearly all ASD listeners were able to perform relatively normally with at least one type of fluctuating masker (the amplitude modulated F0 contour), whilst even with amplitude modulated noise only two performed just outside the normal range (see Chapter 4). The temporal envelopes of the non-speech fluctuating maskers in Experiment 2 were all based on the same speech masker, and in that respect offered similar opportunities for dip listening. It is a little difficult on this basis, therefore, to argue that temporal processing efficiency might cause deficits with fluctuating maskers. Since there does, however, appear to be a trend towards poor performance with AM noise maskers in ASD - looking across the results here and in Alcántara's work - it seems possible that some other aspect of fluctuating aperiodic maskers may be to blame. One idea, based on studies of auditory hyper-reactivity in ASD (e.g. Koegel et al., 2004) is that these sounds may be inherently more unpleasant to listen to, and therefore more distracting, for listeners with ASD. Another possibility is that, whilst the speech heard during dips in a masker may be just as audible to ASD listeners as to controls, they may be less able to reconstruct the original signal from short glimpses. This would imply a linguistic deficit rather than a sensory or perceptual one.

The mCRM is an unusual task with regard to dip listening, since the carrier phrase is identical across all sentences and therefore provides no semantic or grammatical cues to the target key word. From this perspective, although the task was extremely useful for minimising the effect of any individual differences in language ability, it did lack ecological validity. Whilst the bulk of the data presented here was gathered using the mCRM, it is obviously important to consider whether similar effects would occur with more typical sentence targets. The final part of Experiment 2, therefore, used more ordinary sentence materials (the Sentence in Noise task) to explore ASD listeners' ability to understand speech from short glimpses (see Chapter 5). In the baseline condition here, rather than presenting target sentences with a simultaneous masker, the speech was periodically interrupted by silent gaps. This meant that there was no energetic masking or distraction by a background sound, with the main challenge being to reconstruct the fragmented message. Once again, there were some clear outliers amongst the ASD group, who required longer glimpses of the speech than controls in order to reach 50% correct performance. This does tend to suggest that deficits for speech with fluctuating maskers may well involve an underlying lack of efficiency in language processing, even for some high-functioning individuals with ASD. Whether this reduced efficiency in the interrupted speech task was due to poor top-down linguistic processing mechanisms or to a difficulty with binding speech fragments into a coherent stream is rather harder to say.

Some other aspects of the data in this thesis also point towards the possibility of speech processing in ASD being atypical even in the absence of masking material. Poor performance with spectrally-degraded noise-vocoded target material (Chapter 3) and poor performance with low-intensity mCRM targets (near the threshold of audibility; Chapter 4) may suggest that at least a subset of ASD listeners may require a more complete set of acoustic-phonetic cues for speech identification than neurotypical adults. On the other hand, poor attentional orienting to speech - just as much as a deficit in language processing per se - could also result in poor intelligibility of these degraded target signals. Both possibilities seem quite plausible, based on the literature outlined earlier. However, the ASD- group's low response accuracy with the 4-channel noise-vocoded speech (which was presented at a clearly audible level throughout) was heavily influenced by one outlier, whereas SRTs for the target material with no masker (measured by adaptively varying the signal intensity) were more consistently poor. Five out of the overall total of 10 ASD listeners in the latter condition were significant outliers. One might think that differences in attention to speech might be more important for very quiet stimuli than for spectrally-degraded (but nevertheless fully audible) stimuli, which suggests once again that attention to speech is atypical in ASD, and, moreover, that it can affect performance independently of auditory streaming difficulties.

Given the questions raised in this discussion, and the variability observed within the ASD group, further research involving a full battery of language assessments and/or neurophysiological measures of phonological processing, alongside speech in noise testing (with speech maskers) could be extremely informative. There is clearly a need to evaluate the relative contributions of language processing ability and selective attention to the intelligibility of masked speech in ASD. Given the similar difficulties with speech in noise which are characteristic of APD, it would also be interesting to explore how patterns of performance across different speech and non-speech maskers might vary between listeners with ASD and APD. It seems possible that difficulties with speech stream segregation and selective attention to a target talker might be relatively specific to ASD, due to the core deficits with social stimuli in this condition. It also seems likely that the intrinsic attention to listening tasks which was found by Moore and colleagues (2010) to be atypical in the lowest-performing children in their study of non-speech processing, might be relatively preserved in high-functioning ASD. There was little evidence in the research presented here of response variability. This aspect of the data has not been analysed in great detail, but in the mCRM, listeners were required to complete an additional block of trials in any condition where SRTs in the first two (compulsory) blocks differed by more than 4 dB. There was no significant difference between groups in the number of third blocks required. Moreover the slopes of the psychometric functions estimated from the adaptive tracking procedure tended to be at least as steep, if not steeper in the ASD group, particularly in the non-speech masking conditions, which would not be the case if they had responded more erratically than controls within the blocks themselves. Given the serious questions surrounding the differential diagnosis of APD, better understanding of how these patterns of listening ability vary across developmental condition can only be helpful.

The pattern of results across tasks and conditions in this thesis suggests that auditory selective attention, speech stream segregation and language ability may all be important factors contributing to reduced speech intelligibility under challenging listening conditions in ASD. A substantial proportion of the high-functioning participants with ASD involved in these experiments appeared to be particularly susceptible to interference from masking speech, whereas their deficits with non-speech maskers were considerably less severe. Given that communication in everyday life very often takes place against a background of other talkers, it is unsurprising that difficulties with speech perception in noise are commonly reported in ASD. High IQ appears to protect against these difficulties at least to some extent, as measured under carefully controlled laboratory conditions. It may not be sufficient, however, to compensate for atypical speech processing in more challenging real-life listening environments, where environmental sounds mingle with speech, talkers move about, sounds reflect off multiple surfaces, and a wealth of information is being processed in other sensory modalities at the same time. One final, and more practical, goal for future research would be to investigate whether specific training in selective attention to a target talker with masking background speech - perhaps alongside training with interrupted speech signals - might contribute to more satisfactory speech perception in everyday life.

Reference List

Alcantara, J., Cope, T. E., Cope, W., & Weisblatt, E. Auditory temporal-envelope processing in high-functioning children with autism spectrum disorder. *Neuropsychologia*, (in press).

Alcantara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry, 45,* 1107-1114.

American Psychiatric Association (1994). *Diagnostic & Statistical Manual of Mental Disorders*. (4 ed.) Washington DC: APA.

Apoux, F. & Healy, E. W. (2011). Relative contribution of target and masker temporal fine structure to the unmasking of consonants in noise. *Journal of the Acoustical Society of America, 130,* 4044-4052.

ASHA (2005). (Central) auditory processing disorders. American Speech-Language-Hearing Association Working group on auditory processing disorders [On-line]. Retrieved from: http://www.asha.org/docs/html/tr2005-00043.html#sec1.3

Badecker, W. (2008). Speech perception following focal brain injury. In D.B.Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 524-545). Oxford: Blackwell Publishing.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet, 368,* 210-215.

Baranek, G. T., David, F. J., Poe, M. D., Stone, W. L., & Watson, L. R. (2006). Sensory Experiences Questionnaire: discriminating sensory features in young children with autism, developmental delays, and typical development. *Journal of Child Psychology and Psychiatry, 47*, 591-601.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5-17.

Bayliss, A. P. & Kritikos, A. (2011). Brief Report: Perceptual Load and the Autism Spectrum in Typically Developed Individuals. *Journal of Autism and Developmental Disorders, 41,* 1573-1578.

Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., & Minshew, N. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia, 44,* 110-129.

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: visual processing in autism. *Trends in Cognitive Sciences, 10,* 258-264.

Benard, M. R. & Baskent, D. (2013). Perceptual learning of interrupted speech. *Plos One, 8,* e58149.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-hearing Children. British Journal of Audiology 13, 108-112.

Binns, C. & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *Journal of the Acoustical Society of America, 122,* 1765-1776.

Bishop, D. V. M. (1997). *Uncommon understanding development and disorders of language comprehension in children*. Hove, East Sussex, England: Psychology Press.

Bishop, D. V. M. (2003). *Autism and specific language impairment: categorical distinction or continuum?* Chichester: John Wiley & Sons Ltd.

Boddaert, N., Belin, P., Chabane, N., Poline, J. B., Barthelemy, C., Mouren-Simeoni, M. C., Brunelle, F., Samson, Y., & Zilbovicius, M. (2003). Perception of complex sounds: Abnormal pattern of cortical activation in autism. *American Journal of Psychiatry, 160,* 2057-2060.

Boersma, P. (2001). Praat: a system for doing phonetics by computer. Glot International 5[9/10], 341-345.

Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.1.43) [Computer software].

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *Journal of the Acoustical Society of America, 107,* 1065-1066.

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., Burack, J. A., & Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia, 48,* 2465-2475.

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41,* 847-857.

Bouvet, L., Donnadieu, S., Mottron, L., & Valdois, S. (2011). Stream segregation in autism: an auditory embedded figures task? International Meeting for Autism Research Abstract Book [On-line]. Retrieved from: http://imfar.confex.com/imfar/2011/webprogram/Paper8940.html

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109,* 1101-1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America, 110*, 2527-2538.

BSA (2011). Position statement: auditory processing disorder. British Society of Audiology [On-line]. Retrieved from: http://www.thebsa.org.uk/images/stories/docs/BSA_APD_PositionPaper_31March11_FINAL.pdf

Burack, J. A. (1994). Selective Attention Deficits in Persons with Autism - Preliminary Evidence of An Inefficient Attentional Lens. *Journal of Abnormal Psychology, 103,* 535-543.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., Elkholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., & Ludvigsen, C. (1994). An International Comparison of Long-Term Average Speech Spectra. *Journal of the Acoustical Society of America, 96*, 2108-2120.

Ceponiene, R., Lepisto, T., Shestakova, A., Vanhala, R., Alku, P., Naatanen, R., & Yaguchi, K. (2003). Speech-sound-selective auditory impairment in children with autism: They can perceive but do not attend. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 5567-5572.

Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G. et al. (1995). EUROM - a spoken language resource for the EU. In *Eurospeech '95* (pp. 867-870). Madrid, Spain.

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine, 41*, 619-627.

Chevallier, C., Noveck, I., Happe, F., & Wilson, D. (2009). From acoustics to grammar: Perceiving and interpreting grammatical prosody in adolescents with Asperger Syndrome. *Research in Autism Spectrum Disorders, 3*, 502-516.

Cleland, J., Gibbon, F. E., Peppe, S. J. E., O'Hare, A., & Rutherford, M. (2010). Phonetic and phonological errors in children with high functioning autism and Asperger syndrome. *International Journal of Speech-Language Pathology, 12,* 69-76.

Constantino, J. N., Yang, D., Gray, T. L., Gross, M. M., Abbacchi, A. M., Smith, S. C., Kohn, C. E., & Kuhl, P. K. (2007). Clarifying the associations between language and social development in autism: A study of non-native phoneme recognition. *Journal of Autism and Developmental Disorders, 37*, 1256-1263.

Cullington, H. E. & Zeng, F. G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *Journal of the Acoustical Society of America, 123,* 450-461.

Cutler, A. & Foss, D. J. (1977). Role of Sentence Stress in Sentence Processing. *Language and Speech, 20,* 1-10.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America, 114,* 2913-2922.

Dawes, P. & Bishop, D. (2009). Auditory processing disorder in relation to developmental disorders of language, communication and attention: a review and critique. *International Journal of Language & Communication Disorders, 44,* 440-465.

Dawes, P., Bishop, D. V. M., Sirimanna, T., & Bamiou, D. E. (2008). Profile and aetiology of children diagnosed with auditory processing disorder (APD). *International Journal of Pediatric Otorhinolaryngology, 72,* 483-489.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology, 27,* 403-424.

de Cheveigne, A., Kawahara, H., Tsuzaki, M., & Aikawa, K. (1997). Concurrent vowel identification .1. Effects of relative amplitude and F-0 difference. *Journal of the Acoustical Society of America, 101,* 2839-2847.

de Cheveigne, A., McAdams, S., Laroche, J., & Rosenberg, M. (1995). Identification of Concurrent Harmonic and Inharmonic Vowels - A Test of the Theory of Harmonic Cancellation and Enhancement. *Journal of the Acoustical Society of America, 97,* 3736-3748.

Depape, A. M. R., Hall, G. B. C., Tillmann, B., & Trainor, L. J. (2012). Auditory Processing in High-Functioning Adolescents with Autism Spectrum Disorder. *Plos One, 7*.

Deroche, M. L. D. & Culling, J. F. (2011). Voice segregation by difference in fundamental frequency: Evidence for harmonic cancellation. *Journal of the Acoustical Society of America, 130,* 2855-2865.

Dirks, D. D. & Bower, D. R. (1969). Masking Effects of Speech Competing Messages. *Journal of Speech and Hearing Research, 12,* 229-&.

Dirks, D. D. & Wilson, R. H. (1969). The Effect of Spatially Separated Sound Sources on Speech Intelligibility. *Journal of Speech and Hearing Research, 12,* 5-38.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, G. (1997). *British picture vocabulary scale*. (2nd ed ed.) Windsor: NFER-Nelson.

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *Journal of the Acoustical Society of America, 113,* 2984-2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America, 114,* 368-379.

Egan, J. P., Carterette, E. C., & Thwing, E. J. (1954). Some Factors Affecting Multi-Channel Listening. *Journal of the Acoustical Society of America, 26,* 774-782.

Ferguson, M. A., Hall, R. L., Riley, A., & Moore, D. R. (2011). Communication, Listening, Cognitive and Speech Perception Skills in Children With Auditory Processing Disorder (APD) or Specific Language Impairment (SLI). *Journal of Speech Language and Hearing Research, 54*, 211-227.

Fombonne, E. (2009). Epidemiology of Pervasive Developmental Disorders. *Pediatric Research, 65*, 591-598.

Foxton, J. M., Stewart, M. E., Barnard, L., Rodgers, J., Young, A. H., O'Brien, G., & Griffiths, T. D. (2003). Absence of auditory 'global interference' in autism. *Brain, 126*, 2703-2709.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America, 109*, 2112-2122.

Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2007). Variability and uncertainty in masking by competing speech. *Journal of the Acoustical Society of America, 121*, 1040-1046.

Freyman, R. L., Helfer, K. S., Mccall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America, 106*, 3578-3588.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America, 110*, 1150-1163.

Frith, U. (2003). *Autism: Explaining the Enigma*. (2 ed.) Oxford: Blackwell Publishing.

Frith, U. & Happe, F. (1994). Autism - Beyond Theory of Mind. *Cognition, 50*, 115-132.

Frye, R. E. & Beauchamp, M. S. (2009). Receptive Language Organization in High-Functioning Autism. *Journal of Child Neurology, 24*, 231-236.

Fullgrabe, C. (2009). Role of pitch-discrimination in sequential auditory stream segregation by individuals with Asperger's syndrome: Poster presented at International Meeting for Autism Research. International Meeting for Autism Research Abstract Book [On-line]. Retrieved from: https://imfar.confex.com/imfar/2009/webprogram/Paper4755.html

Fullgrabe, C., Alcantara, J., & Weisblatt, E. (2008). Release from masking in speech perception by children with Asperger's syndrome. International Meeting for Autism Research Abstract Book [On-line]. Retrieved from: https://imfar.confex.com/imfar/2008/webprogram/Paper2808.html

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthelemy, C., Brunelle, F., Samson, Y., & Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience, 7*, 801-802.

Geschwind, D. H. (2011). Genetics of autism spectrum disorders. *Trends in Cognitive Sciences, 15*, 409-416.

Gnansia, D., Jourdes, V., & Lorenzi, C. (2008). Effect of masker modulation depth on speech masking release. *Hearing Research, 239,* 60-68.

Grandin, T. (2006). *Thinking in Pictures: And Other Reports from my Life with Autism*. London: Bloomsbury.

Groen, W. B., van Orsouw, L., ter Huurne, N., Swinkels, S., van der Gaag, R. J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact Spectral but Abnormal Temporal Processing of Auditory Stimuli in Autism. *Journal of Autism and Developmental Disorders, 39,* 742-750.

Groen, W. B., Zwiers, M. P., van der Gaag, R. J., & Buitelaar, J. K. (2008). The phenotype and neural correlates of language in autism: An integrative review. *Neuroscience and Biobehavioral Reviews, 32,* 1416-1425.

Haesen, B., Boets, B., & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders, 5,* 701-714.

Happe, F. G. E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology, 15,* 1-12.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech Perception Abilities of Adults With Dyslexia: Is There Any Evidence for a True Deficit? *Journal of Speech Language and Hearing Research, 52,* 1510-1529.

Heaton, P., Williams, K., Cummins, O., & Happe, F. (2008). Autism and pitch processing splinter skills: A group and subgroup analysis. *Autism, 12,* 203-219.

Henderson, L. M., Clarke, P. J., & Snowling, M. J. (2011). Accessing and selecting word meaning in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 52,* 964-973.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review, 24,* 189-233.

Hind, S. (2007). Survey of care pathway for auditory processing disorder. *Audiological Medicine, 4,* 12-24.

Hopkins, K. & Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *Journal of the Acoustical Society of America, 125,* 442-446.

Howard-Jones, P. A. & Rosen, S. (1993). The Perception of Speech in Fluctuating Noise. *Acustica, 78*, 258-272.

Huggins, A. W. F. (1964). Distortion of Temporal Pattern of Speech - Interruption + Alternation. *Journal of the Acoustical Society of America, 36,* 1055-&.

IEEE (1969). Ieee Recommended Practice for Speech Quality Measurements. *Ieee Transactions on Audio and Electroacoustics, AU17*, 225-&.

Jansson-Verkasalo, E., Ceponiene, R., Kielinen, M., Suominen, K., Jantti, V., Linna, S. L., Moilanen, I., & Naatanen, R. (2003). Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. *Neuroscience Letters, 338,* 197-200.

Jansson-Verkasalo, E., Kujala, T., Jussila, K., Mattila, M. L., Moilanen, I., Naatanen, R., Suominen, K., & Korpilahti, P. (2005). Similarities in the phenotype of the auditory neural substrate in children with Asperger syndrome and their parents. *European Journal of Neuroscience, 22,* 986-990.

Jarvinen-Pasley, A., Pasley, J., & Heaton, P. (2008). Is the linguistic content of speech less salient than its perceptual features in autism? *Journal of Autism and Developmental Disorders, 38,* 239-248.

Jolliffe, T. & Baron-Cohen, S. (1999). A test of central coherence theory: linguistic processing in high-functioning adults with autism or Asperger syndrome: is local coherence impaired? *Cognition, 71,* 149-185.

Jolliffe, T. & Baron-Cohen, S. (2000). Linguistic processing in high-functioning adults with autism or Asperger's syndrome. Is global coherence impaired? *Psychological Medicine, 30,* 1169-1187.

Jones, C. R. G., Happe, F., Baird, G., Simonoff, E., Marsden, A. J. S., Tregay, J., Phillips, R. J., Goswami, U., Thomson, J. M., & Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia, 47,* 2850-2858.

Kalbfleisch, M. L. & Loughan, A. R. (2012). Impact of IQ Discrepancy on Executive Function in High-Functioning Autism: Insight into Twice Exceptionality. *Journal of Autism and Developmental Disorders, 42,* 390-400.

Kern, J. K., Trivedi, M. H., Garver, C. R., Grannemann, B. D., Andrews, A. A., Savla, J. S., Johnson, D. G., Mehta, J. A., & Schroeder, J. L. (2006). The pattern of sensory processing abnormalities in autism. *Autism, 10,* 480-494.

Klin, A. (1991). Young Autistic Childrens Listening Preferences in Regard to Speech - A Possible Characterization of the Symptom of Social Withdrawal. *Journal of Autism and Developmental Disorders, 21,* 29-42.

Koegel, R. L., Openden, D., & Koegel, L. K. (2004). A systematic desensitization paradigm to treat hypersensitivity to auditory stimuli in children with autism in family contexts. *Research and Practice for Persons with Severe Disabilities, 29,* 122-134.

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews, 8,* 89-105.

Korpilahti, P., Jansson-Verkasalo, E., Mattila, M. L., Kuusikko, S., Suominen, K., Rytky, S., Pauls, D. L., & Moilanen, I. (2007). Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of Autism and Developmental Disorders, 37,* 1539-1549.

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental Science, 8,* F1-F12.

Lagace, J., Jutras, B., Giguere, C., & Gagne, J. P. (2011). Speech perception in noise: Exploring the effect of linguistic context in children with and without auditory processing disorder. *International Journal of Audiology, 50,* 385-395.

Langdell, T. (1978). Recognition of Faces - Approach to Study of Autism. *Journal of Child Psychology and Psychiatry, 19,* 255-268.

Laures, J. S. & Bunton, K. (2003). Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions. *Journal of Communication Disorders, 36,* 449-464.

Laures, J. S. & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech Language and Hearing Research, 42,* 1148-1156.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9,* 75-82.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology-General, 133,* 339-354.

Leech, R., Aydelott, J., Symons, G., Carnevale, J., & Dick, F. (2007). The development of sentence interpretation: effects of perceptual, attentional and semantic interference. *Developmental Science, 10,* 794-813.

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics, 63,* 1279-1292.

Lepisto, T., Kuitunen, A., Sussman, E., Saalasti, S., Jansson-Verkasalo, E., Nieminen-von Wendt, T., & Kujala, T. (2009). Auditory stream segregation in children with Asperger syndrome. *Biological Psychology, 82,* 301-307.

Lepisto, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Naatanen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research, 1066,* 147-157.

Lepisto, T., Nieminen-von Wendt, T., von Wendt, L., Naatanen, R., & Kujala, T. (2007). Auditory cortical change detection in adults with Asperger syndrome. *Neuroscience Letters, 414,* 136-140.

Lepisto, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Naatanen, R., & Kujala, T. (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical Neurophysiology, 117,* 2161-2171.

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467-&.

Lopez, B. & Leekam, S. R. (2003). Do children with autism fail to process information in context? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44,* 285-300.

Lord, C. & Jones, R. M. (2012). Annual Research Review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 53,* 490-509.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30,* 205-223.

Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autistic spectrum disorders. In M.L.Rice & S. F. Warren (Eds.), *Developmental language disorders: from phenotypes to etiologies* (pp. 7-30). Mahwah, New Jersey: Erlbaum.

Lord, C., Rutter, M., & Lecouteur, A. (1994). Autism Diagnostic Interview-Revised - A Revised Version of A Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders. *Journal of Autism and Developmental Disorders, 24,* 659-685.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & 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,* 18866-18869.

Loukusa, S. & Moilanen, I. (2009). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. *Research in Autism Spectrum Disorders, 3,* 890-904.

MacLeod, A. & Summerfield, A. Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. *British Journal of Audiology, 24,* 29-43.

Madriaga, M. (2010). 'I avoid pubs and the student union like the plague': Students with Asperger Syndrome and their negotiation of university spaces. *Childrens Geographies, 8,* 39-50.

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech Language and Hearing Research, 44,* 416-433.

Miller, G. A. (1947). The Masking of Speech. *Psychological Bulletin, 44,* 105-129.

Miller, G. A. & Licklider, J. C. R. (1950). The Intelligibility of Interrupted Speech. *Journal of the Acoustical Society of America, 22,* 167-173.

Minshew, N. J. & Williams, D. L. (2007). The new neurobiology of autism - Cortex, connectivity, and neuronal organization. *Archives of Neurology, 64,* 945-950.

Molnar-Szakacs, I. & Heaton, P. (2012). *Music: a unique window into the world of autism*. Oxford: Blackwell Science Publishing.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*, 1349-1358.

Moore, B. C. J. (1987). Psychophysics of Normal and Impaired Hearing. *British Medical Bulletin, 43*, 887-908.

Moore, B. C. J. (2007). *Cochlear hearing loss physiological, psychological and technical issues*. (2nd ed. ed.) Chichester: Wiley.

Moore, D. R. (2012). Listening difficulties in children: Bottom-up and top-down contributions. *Journal of Communication Disorders, 45*, 411-418.

Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., Ratib, S., & Riley, A. (2010). Nature of Auditory Processing Disorder in Children. *Pediatrics, 126*, E382-E390.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36*, 27-43.

Mottron, L., Peretz, I., & Menard, E. (2000). Local and global processing of music in high-functioning persons with autism: Beyond central coherence? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*, 1057-1065.

Neff, D. L. & Dethlefs, T. M. (1995). Individual-Differences in Simultaneous Masking with Random-Frequency, Multicomponent Maskers. *Journal of the Acoustical Society of America, 98*, 125-134.

Nelson, P. B. & Jin, S. H. (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America, 115*, 2286-2294.

Newg (2008). Hearing everything, and inability to filter background noise. WrongPlanet.net [On-line]. Retrieved from: www.wrongplanet.net_postt61882

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews, 36*, 836-854.

Ouimet, T., Foster, N. E. V., Tryfon, A., & Hyde, K. L. (2012). *Auditory-musical processing in autism spectrum disorders: a review of behavioral and brain imaging studies*. OXFORD: BLACKWELL SCIENCE PUBL.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*, 674-681.

Peppe, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech Language and Hearing Research, 50*, 1015-1028.

Pickett, E., Pullara, O., O'Grady, J., & Gordon, B. (2009). Speech Acquisition in Older Nonverbal Individuals With Autism A Review of Features, Methods, and Prognosis. *Cognitive and Behavioral Neurology, 22*, 1-21.

Plaisted, K., Saksida, L., Alcantara, J., & Weisblatt, E. (2003). Towards an understanding of the mechanisms of weak central coherence effects: experiments in visual configural learning and auditory perception. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 358,* 375-386.

Plaisted, K., Swettenham, J., & Rees, L. (1999). Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40,* 733-742.

Ploog, B. O., Banerjee, S., & Brooks, P. J. (2009). Attention to prosody (intonation) and content in children with autism and in typical children using spoken sentences in a computer game. *Research in Autism Spectrum Disorders, 3,* 743-758.

Powers, G. L. & Wilcox, J. C. (1977). Intelligibility of Temporally Interrupted Speech with and Without Intervening Noise. *Journal of the Acoustical Society of America, 61,* 195-199.

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

Remington, A., Swettenham, J., Campbell, R., & Coleman, M. (2009). Selective Attention and Perceptual Load in Autism Spectrum Disorder. *Psychological Science, 20,* 1388-1393.

Rhebergen, K. S. & Versfeld, N. J. (2005). A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *Journal of the Acoustical Society of America, 117,* 2181-2192.

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2008). Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing. *Ear and Hearing, 29,* 169-175.

Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics, 31,* 509-527.

Rosen, S., Cohen, M., & Vanniasegaram, I. (2010). Auditory and cognitive abilities of children suspected of auditory processing disorder (APD). *International Journal of Pediatric Otorhinolaryngology, 74,* 594-600.

Russo, N., Nicol, T., Trommer, B., Zecker, S., & Kraus, N. (2009). Brainstem transcription of speech is disrupted in children with autism spectrum disorders. *Developmental Science, 12,* 557-567.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders, 32,* 189-194.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18,* 559-572.

Shah, A. & Frith, U. (1983). An Islet of Ability in Autistic-Children - A Research Note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 24,* 613-620.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science, 270,* 303-304.

Sharma, M., Purdy, S. C., & Kelly, A. S. (2009). Comorbidity of Auditory Processing, Language, and Reading Disorders. *Journal of Speech Language and Hearing Research, 52,* 706-722.

Shi, L. F. & Law, Y. (2010). Masking effects of speech and music: Does the masker's hierarchical structure matter? *International Journal of Audiology, 49,* 296-308.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12,* 182-186.

Shiroma, M., Iwaki, T., Kubo, T., & Soli, S. (2008). The Japanese Hearing in Noise Test. *International Journal of Audiology, 47,* 381-382.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and Prosody Characteristics of Adolescents and Adults With High-Functioning Autism and Asperger Syndrome. *Journal of Speech, Language, and Hearing Research, 44,* 1097-1115.

Simmons, D. R., Robertson, A. E., Mckay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research, 49,* 2705-2739.

Simpson, S. A. & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America, 118,* 2775-2778.

Skafidas, E., Testa, R., Zantomio, D., Chana, G., Everall, I. P., & Pantelis, C. (2012). Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. *Mol Psychiatry*.

Smith, E. G. & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry, 48,* 813-821.

Tallal, P. & Piercy, M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. *Nature, 241,* 468-469.

Tammett, D. (2012). *Born on a blue day: A memoir of Asperger's and an extraordinary mind*. London: Hodder & Stoughton Ltd.

Teder-Salejarvi, W. A., Pierce, K. L., Courchesne, E., & Hillyard, S. A. (2005). Auditory spatial localization and attention deficits in autistic adults. *Cognitive Brain Research, 23,* 221-234.

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: insight from Grammatical-SLI. *Trends in Cognitive Sciences, 9,* 53-59.

Verschuure, J. & Brocaar, M. P. (1983). Intelligibility of Interrupted Meaningful and Nonsense Speech with and Without Intervening Noise. *Perception & Psychophysics, 33*, 232-240.

Vestergaard, M. D. & Patterson, R. D. (2009). Effects of voicing in the recognition of concurrent syllables. *Journal of the Acoustical Society of America, 126,* 2860-2863.

Watson, P. J. & Schlauch, R. S. (2008). The Effect of Fundamental Frequency on the Intelligibility of Speech With Flattened Intonation Contours. *American Journal of Speech-Language Pathology, 17*, 348-355.

Wechsler, D. (1999). Wechsler abbreviated scale of intelligence (WASI). Hove, Psychological Corporation.

White, S., Frith, U., Milne, E., Rosen, S., Swettenham, J., & Ramus, F. (2006). A double dissociation between sensorimotor impairments and reading disability: A comparison of autistic and dyslexic children. *Cognitive Neuropsychology, 23,* 748-761.

Whitehouse, A. J. & Bishop, D. V. M. Do children with autism 'switch off' to speech sounds? An investigation using event-related potentials. *Developmental Science,* (in press).

Wightman, F. L. & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *Journal of the Acoustical Society of America, 118,* 3164-3176.

Wightman, F. L., Kistler, D. J., & O'Bryan, A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *Journal of the Acoustical Society of America, 128*, 270-279.

Williams, D. (2008). *The jumbled jigsaw: An insider's approach to the treatment of autism spectrum 'fruit salads'*. London: Jessica Kingsley.

Williams, J. H. G., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Research in Developmental Disabilities, 25,* 559-575.

Witton, C., Talcott, J. B., Hansen, P. C., Richardson, A. J., Griffiths, T. D., Rees, A., Stein, J. F., & Green, G. G. R. (1998). Sensitivity to dynamic auditory and visual stimuli predicts nonword reading ability in both dyslexic and normal readers. *Current Biology, 8,* 791-797.

World Health Organization (1992). *International Classification of Diseases: Diagnostic criteria for Research*. Geneva: WHO.

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science, 12,* 732-745.

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology, 110,* 362-372.